Protein Unfolding and Aggregation Under Hydrodynamic Flow

Alexander David Page

The University of Leeds

School of Molecular and Cellular Biology

Astbury Centre for Structural Molecular Biology

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

December 2023

Declaration

I confirm that the work submitted is my own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

Firstly, my thanks go to my project supervisors, Professors David Brockwell and Sheena Radford. Their enthusiasm and seemingly endless fountain of ideas has always been appreciated and a source of inspiration.

Thanks also to the sponsors for this project, BBSRC Collaborative Training Partnership and AstraZeneca. Particularly, I must extend my thanks to several members of the AstraZeneca team; Christopher Lloyd and Carolina Orozco were integral in the collaboration, and made my time in the AZ labs at Cambridge a valuable and enjoyable experience. Additionally, thanks to Isabelle Sermadiras, for helping with the project at such short notice.

My heartfelt thanks go to the ‘Squishers’ team; Drs Leon Willis, Samantha Lawrence and Ioanna Panagi, a small but close-knit group that has been fantastic to be a part of from day one. Especially I must thank my good friend Leon for his endless ideas, enthusiasm, and support throughout – and for demonstrating firsthand what it is to be an exemplary scientist.

Thanks to the Radford/Brockwell lab as a whole, it has been a friendly and diverse environment full of collaboration. I am yet to find a member of the group who doesn’t spring to help if they can offer expertise. Especially, my thanks got to Sophie Cawood and Nasir Khan who both work tirelessly to keep the lab running smoothly and offer their help with any issue. Nasir, I am going to miss the curries and bourbon biscuits I seemed to always get away with – and of course your sense of humour.

Finally, my thanks go to my parents and brother, for their constant encouragement and interest in my work from start to finish. Thanks also go to all the friends and family that have asked and been interested in this project, and for the motivation and support when needed.

Abstract

Monoclonal antibodies (mAbs) are a class of therapeutic that has seen enormous success in recent years, in both commercial value and their broad spread of applications. Despite this, antibody-based therapeutics are hindered by the issues of conformational stability and protein aggregation. It is thought that hydrodynamic flow fields encountered during manufacture are a major cause of aggregation in mAbs, but the mechanism by which this occurs is poorly understood.

Here, an assay has been developed to probe for conformational changes under hydrodynamic flow; transient flow-induced conformational change may facilitate exposure of aggregation-prone regions and subsequent oligomerisation. An extensional flow device (EFD) capable of generating defined flow fields, with a focus on extensional flow, has been used to stress the folding model protein RNase H*. Using a panel of cysteine-containing variants of this protein, a fluorescence-based labelling assay has been used in the EFD to probe for unfolding in a site-specific manner.

This work demonstrates that hydrodynamic flow induces protein unfolding in a manner that appears dictated by the structural features present in RNase H*; residues present in the hydrophobic core of RNase H* are better protected from unfolding than those that are more peripheral. Differences in unfolding propensity are reflected in the diverse aggregation propensities observed in the RNase H* cysteine variants, suggesting that the EFD-induced aggregation observed in RNase H* variants is a result of flow-induced conformational changes. The insight into the structural features that determine flow-induced unfolding may help elucidate the mechanism of flow-induced aggregation, and assist in structure-guided design of more ‘stable’ therapeutics.

Additionally, the effects of EFD-induced flow fields were analysed in a panel of 27 mAbs, to characterise the response of more pharmacologically relevant molecules. The mAb panel was also exposed to a set of 9 other assays, to ascertain how EFD-induced aggregation relates to parameters widely employed in biopharmaceutical development, and if this selection of assays could be indicative of a candidate biopharmaceutical’s stability under accelerated stability (AS) conditions. It was found that the combination of assays was unable to predict for AS performance over 4 weeks at 40 °C, indicating that AS probes aspects of biopharmaceutical stability not reflected in other assays. EFD-induced aggregation performance did not correlate with any other assays, suggesting that flow-induced aggregation propensity is a unique metric that should be considered reflective of specific stresses encountered in a biopharmaceutical’s lifecycle.

Contents

Declaration III
Acknowledgements IV
Contents VI
List of Figures XII
List of Tables XVI
List of Abbreviations XVII
1 Introduction 2
1.1 Protein Folding and Unfolding 2
1.1.1 Protein Folding 2
1.1.2 Protein Unfolding 6
1.1.3 Thermodynamic Stability 7
1.1.4 Mechanical Stability 8
1.1.5 Comparability of Protein Stability Measurements 8
1.2 Protein Aggregation 9
1.2.1 Aggregation Mechanism 9
1.2.2 Aggregation-Prone Regions 10
1.2.3 Consequences of Protein Aggregation 11
1.3 Biopharmaceuticals 12
1.3.1 Aggregation in Biopharmaceuticals 13
1.3.2 Causes of Aggregation in Biopharmaceuticals 14
1.3.3 Developability Studies in Biopharmaceuticals 16
1.3.4 Initial Developability Measurements 17
1.3.5 Accelerated Stability 18
1.3.6 *In Silico* Predictors of Protein Aggregation 19
1.3.7 Use of Developability Parameters for Protein Engineering 20
1.3.8 Use of Developability Parameters for Protein Stability Predictions 21
1.4 Hydrodynamic Flow 22
1.4.1 Magnitudes of Hydrodynamic Forces 23
1.4.2 The Mechanism of Flow-Induced Protein Aggregation 24
1.4.3 The Reported Effects of Hydrodynamic Flow on Protein Conformation 29
1.4.4 Studies with the Extensional Flow Device (EFD) 30
1.5 Studying the Effects of Hydrodynamic Flow on Protein Unfolding 31
1.5.1 Methods to Study Hydrodynamic Flow 31
1.6 Model Proteins for Flow-Induced Unfolding Studies 34

1.6.1 RNase H 34
1.6.2 RNase H* Folding 34
1.6.3 RNase H* Unfolding 36
1.7 Aims 37
1.8 Results Chapters Overview 38
2 Thesis Methods 40
2.1 Materials 40
2.1.1 Equipment 40
2.1.1.1 Molecular Biology 40
2.1.1.2 Protein Expression and Purification 40
2.1.1.3 Protein Purity Quantification 40
2.1.1.4 Protein Characterisation 41
2.1.1.5 Hydrodynamic Flow Experiments 41
2.1.1.6 Analytical Chromatography 41
2.1.1.7 Other Equipment 42
2.1.2 Software 42
2.1.3 Chemicals 42
2.1.4 Buffers 44
2.1.5 Other Materials 45
2.1.5.1 Pre-prepared Reagent Mixes and Miscellaneous 45
2.2 Methods 45
2.2.1 Molecular Biology 45
2.2.1.1 Design of RNase H* Cysteine Variants 45
2.2.1.2 Site-Directed Mutagenesis 46
2.2.1.3 Plasmid DNA Purification 47
2.2.2 Protein Expression and Purification 48
2.2.2.1 Plasmid Transformation for Protein Expression in *E. coli* 48
2.2.2.2 Cell Lysis by Sonication 48
2.2.2.3 Heparin Affinity Chromatography 49
2.2.2.4 Ion Exchange Chromatography 49
2.2.2.5 Size-Exclusion Chromatography (SEC) 50
2.2.2.6 Lyophilisation 50
2.2.3 Protein Purity Quantification 50
2.2.3.1 SDS-PAGE 50
2.2.3.2 Densitometry 52
2.2.3.3 Protein Quantification by Spectrophotometry 53
2.2.4 Protein Characterisation 54

2.2.4.1 Intact LC-TOF-MS 54
2.2.4.2 Circular Dichroism Spectroscopy (CD) 54
2.2.4.3 Intrinsic Fluorescence Spectroscopy 55
2.2.4.4 Equilibrium Unfolding for Thermodynamic Stability Measurements 55
2.2.5 Hydrodynamic Flow Experiments 56
2.2.5.1 Extensional Flow Device Assembly 56
2.2.5.2 Analytical Reverse-Phase HPLC (RP-HPLC) 58
2.2.5.3 Analytical Size Exclusion Chromatography-HPLC (HP-SEC) 59
2.2.5.4 Analysis of Protein Surface Coverage by Spectroscopic Ellipsometry 60
2.2.6 Quantification of Protein Aggregation 61
2.2.6.1 Analysis of RNase H* Aggregation 61
2.2.6.2 Kinetic Modelling of RNase H* Aggregation 61
2.2.7 Flow-Induced Unfolding 62
2.2.7.1 Thermolysin-mediated Proteolysis Assay 62
2.2.7.2 RNase H* Enzyme Activity Assay 62
2.2.7.3 IAEDANS Labelling in Urea 63
2.2.7.4. IAEDANS Labelling in Arginine 64
2.2.7.5 IAEDANS Labelling in Phosphate Buffer 64
2.2.8 Analysis of Biopharmaceutical Developability 66
2.2.8.1 Preparation of Biopharmaceutical Samples 66
2.2.8.2 Analysis of Flow-Induced Aggregation 67
2.2.8.3 Accelerated Stability Analysis 67
2.2.8.4 Hydrophobic Interaction Chromatography (HIC) 67
2.2.8.5 Stand-up Monolayer Adsorption Chromatography (SMAC) 68
2.2.8.6 Differential Scanning Fluorimetry (DSF) 68
2.2.8.7 Flow-Induced Taylor Dispersion Analysis (FIDA) 68
2.2.8.8 *In Silico* Analysis 69
2.2.9 Statistical Analysis of Results 70
2.2.9.1 Ranked Multivariate Analysis 70
2.2.9.2 Analysis for Holistic Developability Parameter Ranking 70
3 Design and Characterisation of RNase H* Cysteine Substitution Mutants to Investigate Flow-Induced Protein Unfolding 73
3.1 Design and Production of RNase H* Cysteine Variants 73
3.1.1 RNase H* Cysteine Variant Design 74
3.1.2 RNase H* Cysteine Variant Production 75
3.1.3 RNase H* Cysteine Variant Purification 76
3.1.4 Cysteine Variants Structural Characterisation 79

3.1.5 Thermodynamic Stability of RNase H* Variants 83
3.2 RNase H* Aggregation Under Hydrodynamic Flow 86
3.2.1 Exposing Proteins to Flow with the Extensional Flow Device 86
3.2.2 Characterisation of EFD-Induced Aggregation Propensity in RNase H* Variants 88
3.2.3 Modelling RNase H* Aggregation Kinetics 91
3.2.4 Conformational Change in Bulk Most Affects Soluble Protein Loss 94
3.3 Investigating the Causes for Different Aggregation Behaviours 95
3.3.1 Aggregation-Prone Regions do not Explain Variations in Soluble Protein Loss 96
3.3.2 Thermodynamic Stability Changes do not Explain Variations in Soluble Protein Loss 98
3.4 Discussion 99
4 Development of an Assay to Quantify Hydrodynamic Flow-Induced Protein Unfolding 101
4.1 Aims and Challenges 101
4.2 Measuring Unfolding Using a Label-Free Assay 101
4.2.1 Proteolysis-Based Unfolding Assay 101
4.2.2 Enzyme Activity-Based Assay 104
4.3 Measuring Unfolding Using a Label-based Assay 107
4.3.1 IAEDANS Reactivity 108
4.3.2 Inhibition of Aggregation with L-arginine 109
4.4 Development of an IAEDANS-Based Labelling Assay 112
4.4.1 Densitometry for IAEDANS Labelling Quantification 112
4.4.2 HPLC-Based IAEDANS Labelling Assay 115
4.4.3 Absolute Labelling Measurements by HPLC 117
4.4.4 Absorbance is Not Sufficient for Accurate Detection of IAEDANS Labelling 118
4.4.5 Measuring Unfolding Using Fluorescence 119
4.4.6 Interpolation of Absorbance Values from a Standard Curve 120
4.4.7 HPLC-Based Quantification of IAEDANS-Labelling Compared to a Densitometry-Based Method 121
4.5 Discussion 123
5 Characterisation of Flow-Induced Conformational Changes in RNase H* 124
5.1 Flow-Induced Unfolding in Conditions with Reduced Aggregation 124
5.1.1 Off-Target Labelling of Cysteine Variants 124
5.1.2 Removal of Contribution from Off-target Labelling 126
5.1.3 RNase H* Cysteine Variants Exhibit Heterogeneous IAEDANS Labelling Under Hydrodynamic Flow 127

5.1.4 Structural Conclusions from IAEDANS Labelling Under Flow 130
5.1.5 Folding Sequence Does Not Determine Flow-Induced Unfolding 132
5.1.6 Flow-Induced Unfolding Corresponds to Aggregation 133
5.1.7 Thermodynamic Stability is Not Predictive of Flow-Induced Protein Unfolding 136
5.2 Flow-Induced Unfolding in Aggregation-Permissive Conditions 137
5.2.1 IAEDANS Labelling in Phosphate Buffer Experimental Design 137
5.2.2 Flow-Induced Labelling in Different RNase H* Components 138
5.2.3 Labelling in Aggregate Correlates with Aggregation Propensity in RNase H* Cysteine Substitution Variants 143
5.2.4 Structural Considerations from Labelling in Aggregation-Permissive Conditions 145
5.2.5 RNase H* Unfolding in Arginine Versus Phosphate Buffer 147
5.2.6 Flow-Induced Labelling Increase in Phosphate Buffer is Geometrically Determined 149
5.3 Discussion 150
5.3.1 Conclusions 150
5.3.2 Considerations and Future Work 152
6 Predicting Instability and Aggregation in Biopharmaceuticals 154
6.1 Introduction 154
6.1.1 Applying the EFD to Measure Flow-induced Aggregation in Different Biopharmaceutical Scaffolds 154
6.1.2 Establishing a Holistic Developability Parameter 154
6.1.3 Considering the Role of Extensional Flow in Developability 155
6.2 mAb Panel Characterisation 156
6.2.1 Study Design and Overview 156
6.2.2 HP-SEC Characterisation 160
6.2.3 Hydrodynamic Flow 163
6.2.4 Accelerated Stability 165
6.2.5 Differential Scanning Fluorimetry 168
6.2.6 HIC and SMAC 170
6.2.7 Flow-Induced Dispersion Analysis 173
6.2.8 *In Silico* Analysis 175
6.3 Statistical Analysis of Developability Assays 179
6.3.1 Multivariate Analysis 179
6.3.2 Flow-induced Aggregation and AS 184
6.3.3 Predicting Developability Using a Holistic Parameter 185
6.3.4 Predicting Accelerated Stability with a Holistic Developability Parameter 187

6.3.5 Refinement of HDP Scores 191
6.3.6 HDP Score Calculated with no Accelerated Stability Parameters 192
6.4 Discussion 195
6.4.1 A Role for the EFD in Biopharmaceutical Developability 195
6.4.2 Predicting Accelerated Stability with a Holistic Developability Parameter 196
7 Concluding Remarks and Future Avenues of Research 198
7.1 Concluding Remarks 198
7.2 Future Avenues of Research 200
8 Appendix 202
9 References 203

List of Figures

Figure 1.1: The Hydrophobic Effect 3
Figure 1.2: Proposed Protein Folding Pathway 5
Figure 1.3: Schematic Illustration of Protein Aggregation 9
Figure 1.4: IgG Structure 13
Figure 1.5: Aggregation-Inducing Steps in Biopharmaceutical Production 16
Figure 1.6: Shear and Extensional Flow 22
Figure 1.7: Inverse Concentration Dependence of Flow-Induced Aggregation 24
Figure 1.8: A Model to Describe Flow-Induced Aggregation 25
Figure 1.9: Evolved Conformational Change Under Hydrodynamic Flow 30
Figure 1.10: Devices to Expose Proteins to Shear 33
Figure 1.11: Devices to Expose Proteins to Extensional Flow and Shear 33
Figure 1.12: RNase H* Structure and Folding Pathway 35
Figure 2.1: PCR Thermal Cycling Protocol Schematic 47
Figure 2.2: Quantification of Pixel Intensity from SDS-PAGE Gel Bands 53
Figure 2.3: Extensional Flow Device Assembly 57
Figure 2.4: Standard Acetonitrile Gradient for Reverse-Phase Chromatography 59
Figure 2.5: HP-SEC TSKgel G3000SWxl Calibration Curve 60
Figure 2.6: RNase H* Oligonucleotide Substrate 63
Figure 2.7: RNase H* Phosphate-Buffered Unfolding Assay 65
Figure 2.8: Pearce 660 nm Protein Concentration Measurement Standard Curve 66
Figure 3.1: RNase H* Cysteine Variants 74
Figure 3.2: RNase H* Cysteine Variants Solvent Exposure 75
Figure 3.3: Representative Purification Process for RNase H* 77
Figure 3.4: Representative Intact LC-TOF-MS Spectra 78
Figure 3.5: RNase H* Variants Intrinsic Fluorescence Spectra 80
Figure 3.6: RNase H* Circular Dichroism Spectra Overlaid 81
Figure 3.7: Individual RNase H* Variants Circular Dichroism Spectra 82
Figure 3.8: Thermodynamic Stabilities and Associated m Values for RNase H* Cysteine Variants 84
Figure 3.9: RNase H* Variants Equilibrium Urea Denaturation Curves 85
Figure 3.10: The Extensional Flow Device 87
Figure 3.11: RNase H* Soluble Protein Loss with Hydrodynamic Flow 89
Figure 3.12: RNase H* Variable Soluble Protein Loss Behaviour 90
Figure 3.13: RNase H* Surface Coverage 91
Figure 3.14: EFD-Induced Aggregation Described by a Kinetic Model 92

Figure 3.15: EFD-Induced Aggregation in RNase H* Cysteine Variants Described by a Kinetic Model 93
Figure 3.16: Sensitivity Analysis of RNase H* Aggregation Mechanism Under Flow 95
Figure 3.17: Predicted Aggregation-Prone Regions in WT RNase H* 96
Figure 3.18: Aggregation Prone Regions do not Determine Aggregation Proneness of RNase H* Cystine Variants 97
Figure 3.19: Flow-Induced Aggregation Does Not Correlate with Thermodynamic Stability 99
Figure 4.1: RNase H* Proteolysis Unfolding Assay 103
Figure 4.2: RNase H* Enzyme Activity Assay 106
Figure 4.3: IAEDANS Labelling Assay for Unfolding 108
Figure 4.4: Arginine Succinate Protects Against Flow-Induced Loss of Soluble Protein 110
Figure 4.5: Effects of Arginine Succinate Buffer on Thermodynamic Stability 111
Figure 4.6: A137C RNase H* IAEDANS Labelling Quantified by Densitometry 113
Figure 4.7: Calibration Curve for Labelling Quantification by Densitometry 115
Figure 4.8: Reverse-Phase Chromatography of RNase H* 116
Figure 4.9: RNase H* and IAEDANS Reverse-Phase Chromatography Elution Profiles 117
Figure 4.10: A24C RNase H* Labelling Measured by Absorbance 119
Figure 4.11: Low Labelling can be Detected Effectively with Fluorescence 120
Figure 4.12: Free IAEDANS vs RNase H*-AEDANS Standard Curves 121
Figure 4.13: A137C IAEDANS Labelling Analysis Method Comparison 122
Figure 5.1: IAEDANS Labelling in WT RNase H* 125
Figure 5.2: IAEDANS-Labelled A137C TOF-MS Analysis 126
Figure 5.3: Flow-induced IAEDANS Labelling in RNase H* Cysteine Variants 129
Figure 5.4: IAEDANS Labelling Correlated with Flow-Induced Aggregation 130
Figure 5.5: EFD-Induced VS Quiescent IAEDANS Labelling in RNase H* Cysteine Variants 131
Figure 5.6: Foldon Thermodynamic Stability Measurements do not Correlate With EFD-Induced or Quiescent Unfolding 132
Figure 5.7: Calculated Solvent Accessible Surface Area Increases with Removal of Foldon Structure 133
Figure 5.8: IAEDANS Labelling Correlated with Flow-Induced Aggregation 134
Figure 5.9: Rank Order Analysis of Flow-Induced Unfolding and Aggregation 136
Figure 5.10: Thermodynamic Stability Versus Flow-Induced Unfolding 137
Figure 5.11: IAEDANS Labelling and Aggregate Recovery Assay 138
Figure 5.12: Unfolding Measured in Aggregated RNase H* 139
Figure 5.13: Unfolding Measured in Aggregated RNase H* 141

Figure 5.14: Unfolding Measured in Quiescent RNase H* 142
Figure 5.15: Degree of IAEDANS Labelling in Aggregate Protein Correlates with Degree of Aggregation 144
Figure 5.16: Flow-Induced Aggregation with and without IAEDANS 145
Figure 5.17: Total IAEDANS Labelling of RNase H* in Aggregation-Permissive Conditions 146
Figure 5.18: IAEDANS Labelling in Phosphate Versus Arginine Buffer 148
Figure 5.19: IAEDANS Labelling in Phosphate Versus Arginine Buffer Proposed Mechanism 149
Figure 5.20: IAEDANS Labelling Change in Aggregation-Permissive Conditions 150
Figure 6.1: Developability Assay Selection 159
Figure 6.2: HP-SEC Retention Times of mAb Panel at t=0 161
Figure 6.3: mAb Panel HP-SEC Analysis at t=0 162
Figure 6.4: mAb Panel EFD-Induced Aggregation 164
Figure 6.5: mAb Panel Accelerated Stability (40 °C, 4 weeks) 166
Figure 6.6: mAb Panel Accelerated Stability (45 °C, 2 weeks) 167
Figure 6.7: mAb Panel Differential Scanning Fluorimetry 169
Figure 6.8: mAb Panel Hydrophobic Interaction Chromatography 171
Figure 6.9: mAb Panel Stand-up Monolayer Adsorption Chromatography 172
Figure 6.10: mAb Panel Flow-Induced Dispersion Analysis 174
Figure 6.11: mAb Panel CamSol Analysis 176
Figure 6.12: mAb Panel TAP CDR Length and Hydrophobicity Analysis 177
Figure 6.13: mAb Panel TAP Charge-Based Parameters 178
Figure 6.15: Correlation and Clustering of Assays Measuring Developability Parameters with FIDA Data 181
Figure 6.16: Correlation and Clustering of Assays Measuring Developability Parameters without FIDA Data 182
Figure 6.14: Flow-induced Aggregation Performance Does Not Correlate with F_{ab} T_m 185
Figure 6.17: mAb Panel Naïve HDP Scores 187
Figure 6.18: Clustering of Assays Measuring Developability Parameters without 40 °C Accelerated Stability Data 188
Figure 6.19: Naïve HDP Score Correlates with Accelerated Stability at 40 °C when 45 °C Data is Included 189
Figure 6.20: Clustering of Panel mAbs 190
Figure 6.21: Refined HDP Score Correlated with Accelerated Stability 191
Figure 6.22: Clustering of Assays Measuring Developability Parameters with No Accelerated Stability Data Included 194
Figure 6.23: Naïve and Refined HDP Scores Do Not Correlate with Accelerated Stability at 40 °C When No Accelerated Stability Data is Included 194

List of Tables

Table 1.1: Hydrophobicity Index of Amino Acids 4
Table 2.1: Site-Directed Mutagenesis Reagents 46
Table 2.2: SDS-PAGE Gel Recipe for Analysis of Higher-MW Species 51
Table 2.3: SDS-PAGE Gel Recipe for Analysis of Lower-MW Species 52
Table 2.4: FIDA Protocol 69
Table 3.1: Rate Constants Produced from Fitting WT and STT to an EFD-Induced Aggregation Model 92
Table 6.1: Initial Monoclonal Antibody Panel 157
Table 6.2: Assays Used in this Developability Study 183
Table 6.3: β-Coefficients from MLR of Naïve HDP (no AS 40°C parameters) 192
Table 6.4: β-Coefficients from MLR of Naïve HDP (no AS parameters) 193

List of Abbreviations

- ABD-F = 4-Fluoro-7-sulfobenzofurazan
- ACC = Accessibility
- AFM = Atomic Force Microscopy
- AMD = Adsorbate Mass Density
- APR = Aggregation-Prone Region
- APS = Ammonium Persulphate
- AS = Accelerated Stability
- BSA = Bovine Serum Albumin
- BVP = Baculovirus Particle
- CD = Circular Dichroism
- CDR = Complementarity-Determining Region
- CIC = Cross-Interaction Chromatography
- CV = Column Volume
- DABCYL = (Dimethylamino)phenyl)azo)benzoic acid
- DLS = Dynamic Light Scattering
- DSC = Differential Scanning Calorimetry
- DSF = Differential Scanning Fluorimetry
- DSSP = Dictionary of Secondary Structure Prediction
- DTT = dithiothreitol
- EDTA = Ethylenediamenetetraacetic acid
- EFD = Extensional Flow Device
- ELISA = Enzyme-Linked Immunosorbent Assay
- F_{ab} = Antigen-binding Fragment
- FAM = Fluoresceine
- F_C = Crystallisable Fragment
- FF = Fast Flow
- FIDA = Flow-Induced Dispersion Analysis
- FPOP = Fast Photochemical Oxidation of Proteins
- HDP = Holistic Developability Parameter
- HDX = Hydrogen-Deuterium Exchange
- HIC = Hydrophobic Interaction Chromatography
- HMW = Higher-Molecular Weight
- HP = High Performance
- HT = High Tension
- HX = Hydrogen Exchange
- IAEDANS =
- IAEDANS = ((Iodoacetyl)amino)ethyl)amino)naphthalene-1-sulfonic acid)
- IEX = Ion Exchange Chromatography
- INN = International Non-proprietary Name
- IPTG = Isopropyl β-D-1-thiogalactopyranoside
- ITC = Isothermal Titration Calorimetry
- LB = Lysogeny Broth
- LC-TOF-MS = Liquid Chromatography Time-Of-Flight Mass Spectrometry
- mAb = Monoclonal Antibody
- MLR = Multiple Linear Regression
- MRE = Mean Residue Ellipticity
- PASTA = Prediction of Amyloid Structural Aggregation
- PBS = Phosphate-Buffered Saline
- PDA = Photodiode Array

- PNC = Patches of Negative Charge
- PPC = Patches of Positive Change
- PSH = Patches of Surface Hydrophobicity
- RNH = RNase H*
- RP-HPLC = Reverse-Phase HPLC
- RSA = Relative Solvent Accessibility
- SAP = Spatial Aggregation Predictor
- SB_NB_A = Surface and Bulk Nucleation with Bulk Aggregation
- SDS = sodium dodecyl sulphate
- SEC = Size Exclusion Chromatography
- SF_vCSP = Structural F_v Charge Symmetry Parameter
- SIC = Self-Interaction Chromatography
- SMAC = Stand-up Monolayer Adsorption Chromatography
- TAP = Therapeutic Antibody Profiler
- TEMED = tetramethylethylenediamene
- TFA = trifluoroacetic acid
- vWF = von Willebrand Factor

1 Introduction

1.1 Protein Folding and Unfolding

1.1.1 Protein Folding

Proteins make up a diverse group of macromolecules, which are comprised of polypeptide chains; sequences of amino acids that are covalently linked to one another by peptide (amide) bonds. The selection of 20 naturally occurring amino acids (plus selenocysteine and pyrrolysine, found in archaea) determine the structure and function of all proteins found in life (Young and Schultz, 2010)(Rother and Krzycki, 2010).

Proteins are essential in a range of biological processes and functions, from translation to cellular structure (Alberts *et al.*, 2007). More recently, proteins have found a place in medicine, as well as in less obvious applications such as in the production of biomaterials (Leader, Baca and Golan, 2008)(Abascal and Regan, 2018).

Experiments by Anfinsen *et al*., (1961) revealed that the folded structure adopted by a denatured polypeptide chain upon removal of denaturant is pre-determined by the primary sequence; the order and type of amino acids in the polypeptide chain. This is now the commonly accepted paradigm for protein folding. The interactions that govern this transition do so through a mixture of enthalpic contributions balanced with entropic penalties.

Interactions between residue side chains include ionic interactions between charged residues and dipoles, disulfide bonds between cysteine residues, hydrogen bonds and van der Waals interactions (Schulz and Schirmer, 1979). Additionally, hydrogen bonds between amide hydrogen atoms and carbonyl group oxygen atoms of the polypeptide ‘backbone’ facilitate collapse of the primary sequence into the most common secondary structure features; α-helices and β-sheets (Dobson, 2003).

Entropically, hydrophobicity of the amino acids in a protein sequence is a major driver of tertiary structure formation; if hydrophobic residues were exposed to a solution of water, H_2O molecules would form a cage-like structure of hydrogen bonds around the non-polar residue, with a significant entropic penalty (Southall, Dill and Haymet, 2002). This is the hydrophobic effect (fig 1.1).

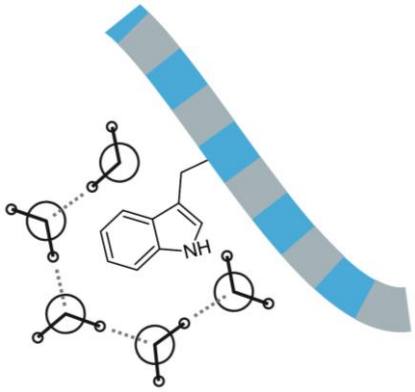


Figure 1.1: The Hydrophobic Effect

Schematic illustrating the entropically unfavourable cage-like hydrogen bond network that forms upon exposure of a hydrophobic residue to a polar solvent. Solvent molecules indicate H_2O, and dotted grey lines indicate hydrogen bonds.

The extent to which these residues promote folding is related to their hydrophobicity, with aromatic-containing residues such as tryptophan and saturated hydrocarbon-containing residues such as isoleucine inducing the greatest effect (table 1.1)

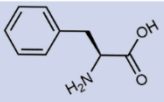
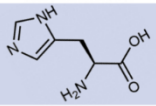
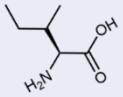
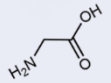
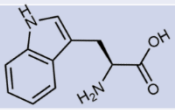
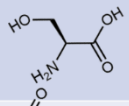
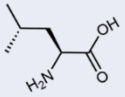
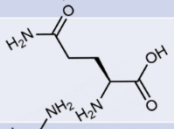
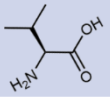
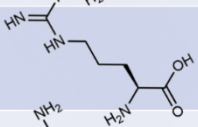
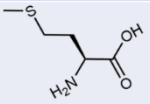
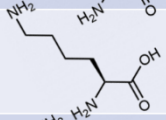
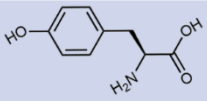
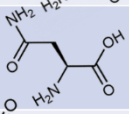
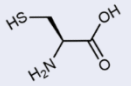
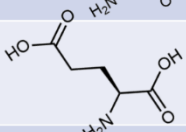
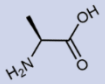
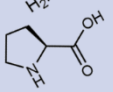
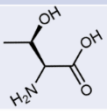
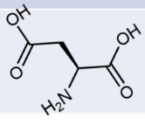
Residue	Normalised Hydrophobicity Index	Structure	Residue	Normalised Hydrophobicity Index	Structure
Phenylalanine (Phe)	100		Histidine (His)	8	
Isoleucine (Ile)	99		Glycine (Gly)	0	
Tryptophan (Trp)	97		Serine (Ser)	-5	
Leucine (Leu)	97		Glutamine (Gln)	-10	
Valine (Val)	76		Arginine (Arg)	-14	
Methionine (Met)	74		Lysine (Lys)	-23	
Tyrosine (Tyr)	63		Asparagine (Asn)	-28	
Cysteine (Cys)	49		Glutamic Acid (Glu)	-31	
Alanine (Ala)	41		Proline (Pro)	-46	
Threonine (Thr)	13		Aspartic Acid (Asp)	-55	

Table 1.1: Hydrophobicity Index of Amino Acids

Normalised hydrophobicity index given at pH 7, relative to glycine. Each residue was measured as part of an 18-residue peptide. Hydrophobicity was measured using reverse-phase HPLC, by Monera *et al.*, (1995). Taken from 'Amino Acids Reference Chart', Sigma-Aldrich.

Although some proteins reach the native state through sampling a series of discrete intermediates, some appear to transition from unfolded to native state, with no detectable intermediates (Fersht, 1997). It was thought that this gain of structure could be a random search through conformational space, but with the enormous number of possibilities, would not be possible within the timescales that had been observed at the time, even at the highest observed folding and unfolding rates; this is Levinthal's paradox (Levinthal, 1968). Although a range of proposals exist for folding pathways, it is now thought that proteins move towards the most thermodynamically favourable state, the native state, which can be through intermediates (Dill and Chan, 1997). Hydrophobic collapse

(described above) and accumulation of interactions between amino acids that favour the native state lead the molecule towards it, like a metaphorical object led to the bottom of a pit by gravity; formation of secondary structure can stabilise and facilitate formation of subsequent structure, and the amphipathic nature of α-helices and β-sheets facilitates formation of tertiary structure, with further stabilises the native conformation. Figure 1.2 illustrates the free energy of different states that may be present in an ensemble of folding protein, simulated for the monomeric protein acylphosphatase. Whilst specific non-covalent interactions are involved in formation of this structure, formation of covalent interactions can also stabilise the native-like interactions, with the magnitude of stabilising effects being varied. Additionally, in multi-domain proteins, gain of quaternary structure can stabilise the intermediates on-pathway to the native state.

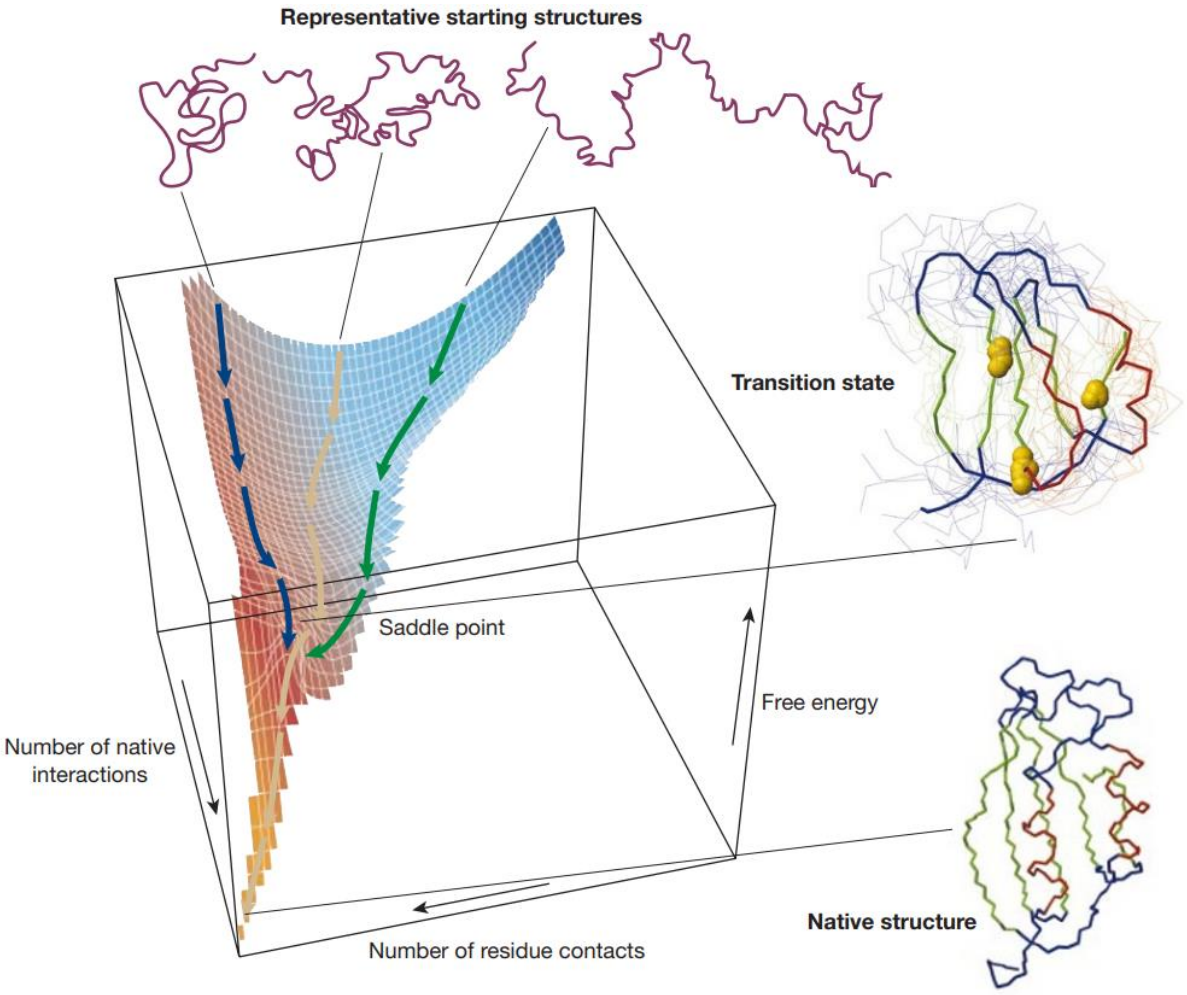


Figure 1.2: Proposed Protein Folding Pathway

Schematic produced using acylphosphatase as an example protein. Saddle point denotes a thermodynamic barrier at which a folding transition state can be present; a structure that is thermodynamically unfavourable and on-pathway to a more thermodynamically favourable state. Taken from Dobson., 2003.

1.1.2 Protein Unfolding

Protein unfolding is thought to result from loss of the interactions that maintain secondary and tertiary structure; hydrogen bonds, ionic interactions and van der Waals interactions, along with steric changes (Lumry and Eyring, 1954). A range of stimuli can lead to this transition.

Chemical denaturants, including urea and guanidine hydrochloride, are well-known and effective in unfolding proteins. Although the mechanism of action isn't fully understood, these two most frequently used molecules are chaotropic, with urea thought to interact with the amide backbone of the polypeptide chain through hydrogen bonds (Lim, Rösgen and Englander, 2009). Exposure to strong acids or bases also induce unfolding; extremes of pH can result in changes to protonation states of charged amino acids, resulting in repulsion between residues and loss of stabilising salt bridges (Yang and Honig, 1993).

Temperature induces protein unfolding; as temperature increases past the T_m of the protein, ΔG_{unf} decreases until the unfolded state becomes more thermodynamically favourable than the folded state. i.e., the unfolded state is at the bottom of the protein folding pathway in fig 1.2 (Lapidus, 2017). Proteins can also unfold under high pressure, where small cavities within the molecule's structure are thought to be destabilised, resulting in the cavity-free unfolded state being thermodynamically more favourable (Roche *et al.*, 2012).

Mechanical stimuli are thought to drive protein unfolding, by imparting direct force on the molecule to overcome the energy barrier to unfolding by breaking hydrogen bonds and other stabilising interactions (Javadi, Fernandez and Perez-Jimenez, 2013). The directionality of force application significantly affects how the molecule unfolds, with experimental techniques such as atomic force microscopy (AFM) frequently relying on tethers to impart force at two fixed positions on the protein's structure, and natural mechanical stresses such as shear thought to stress molecules in a more dispersed and unpredictable way (Rief *et al.*, 1997)(Sterpone, Derreumaux and Melchionna, 2018).

Interaction with interfaces may also be a contributor to protein unfolding; the presence of an air-liquid or solid-liquid interface may promote unfolding of a protein in solvent, by preferential interaction of the hydrophobic core with the interface (Nakanishi, Sakiyama and Imamura, 2001). In addition to interfaces, protein unfolding can be promoted by interaction with other molecules, protein or non-protein; Coan *et al.*, (2009) demonstrated unfolding of β-lactamase stimulated by interaction with rottlerin, a polyphenol.

Cellular processes *in vivo*, such as proteolysis and membrane translocation, can utilise protein unfolding. These often involve stabilisation of the unfolded state through hydrophobic interactions with another macromolecule, or unfolding driven by an energy input such as ATP (Matouschek, 2003).

Evidence indicates that unfolding can proceed through the same pathway as folding; Φ-value analysis of chymotrypsin inhibitor 2 revealed that folding and unfolding transition states are the same (Jackson *et al.,* 1993). Molecular dynamics-based studies appear to support this conclusion (McCully *et al.,* 2008). Although there is a lack of evidence for proteins with multiple intermediates, this work implies that if a folding pathway is known for a protein, the earliest intermediates may be the last to unfold.

1.1.3 Thermodynamic Stability

As described in section 1.1.1, protein fold stability is determined by the difference in thermodynamic stability between the native and unfolded state at equilibrium. Thermodynamic stability is measured as Gibb's free energy change requirement for unfolding, ΔG_{unf}. Gibbs free energy is a sum of enthalpic and entropic contributions:

$$\Delta G = \Delta H - T\Delta S$$

Equation 1.1

Where ΔG is change in Gibbs free energy, ΔH is change in enthalpy, T is temperature in Kelvin, ΔS is change in entropy.

Thermodynamic stability can be measured by chemical denaturation using a denaturant such as urea, in which the protein is exposed to a concentration range of the perturbant in to transition between folded and unfolded states (Pace, 1986). Once equilibrium is reached, folding status is often then detected using spectroscopic techniques. Alternatively, thermal methods such as differential scanning calorimetry (DSC), differential scanning fluorimetry (DSF) or isothermal titration calorimetry (ITC) can be used to measure thermodynamic stability. In the former example, changes to heat capacity (ΔC_p) are measured, which are indicative of unfolding; unfolded protein has a greater C_p than folded (Pace, 1990). T_m, the protein melt temperature, is also commonly employed as a parameter for protein stability (Johnson, 2013). T_m is the temperature at which 50% of a given species of interest is thermally denatured; at T_m, ΔG_{unf} = 0 kJ/mol; it is a parameter independent to ΔG_{unf} at equilibrium.

1.1.4 Mechanical Stability

Mechanical unfolding can be used to measure protein stability. Most often, atomic force microscopy is used to measure mechanical protein unfolding, involving tethering of the protein at two distinct positions, and pulling from these using a mechanical cantilever. Cantilever position and force input are captured to measure unfolding force (Forman and Clarke, 2007). For more sensitive force detection, optical tweezers can be used (Cecconi *et al.*, 2005). The force requirement for unfolding is usually the metric used to reflect stability.

However, the area of effect is not dispersed as with chemical or thermal unfolding; in the pulling experiments mentioned above, a finite number of angles are available to pull from, due to the limited tether points on a protein and experimental practicality. This is important as mechanical perturbations can be anisotropic; for example, (Brockwell *et al.*, 2003) showed that the stability of beta sheets in the protein E2lip3 varied by approximately 10-fold with different pulling geometries, suggesting stability measured by pulling experiments could vary wildly depending on choice of tether points.

1.1.5 Comparability of Protein Stability Measurements

Because there are a range of causes of protein denaturation, different stability measurements may not correlate with one another. For example, mechanical strength has been shown not to correspond to stability against chemical denaturation by urea. Φ-value analysis of four protein L mutations was used to probe whether the mutations were in positions that affect thermodynamic stability of the protein's transition state. Mechanical unfolding and chemical denaturation showed totally different values; the former indicating no change from the mutations, and the latter indicating significant change (Sadler *et al.*, 2009). This discrepancy was thought to be due to different modes of unfolding; local disruption of terminal β-strands with mechanical unfolding, and whole domain unfolding with chemical denaturation.

Similarly, the I27 domain of titin shows distinctly different unfolding intermediates when chemical denaturation is compared to mechanical (Best *et al.*, 2003), and more recently this was observed to be the case in a mutational analysis of SasG, a bacterial adhesion protein (Bruce, Paci and Brockwell, 2023). These observations are unsurprising, as thermodynamic stability by chemical denaturation is an equilibrium measure, whereas

unfolding under force is kinetic, directly influenced by the rate at which the molecule is exposed to force.

1.2 Protein Aggregation

1.2.1 Aggregation Mechanism

Even in thermodynamically stable examples, proteins do not exhibit 100% success in reaching the native state from an unfolded precursor: non-native species can form when the polypeptide accesses off-pathway free energy minima. Improperly folded proteins can be dysfunctional, and are the source of numerous diseases, including cystic fibrosis, Alzheimer's disease and types of cancer (Chaudhuri and Paul, 2006).

Alternatively, partially unfolded, misfolded, completely unfolded and even native state species can form aggregates. Protein aggregation is a process by which monomers or smaller oligomers interact to form higher-MW species. This can occur *in vitro* and *in vivo*, from small polypeptide chains to large, multi-domain proteins (Dobson, 2003)(Pang *et al.*, 2023). Aggregates can be amorphous or form structured fibrils, as illustrated in figure 1.3.

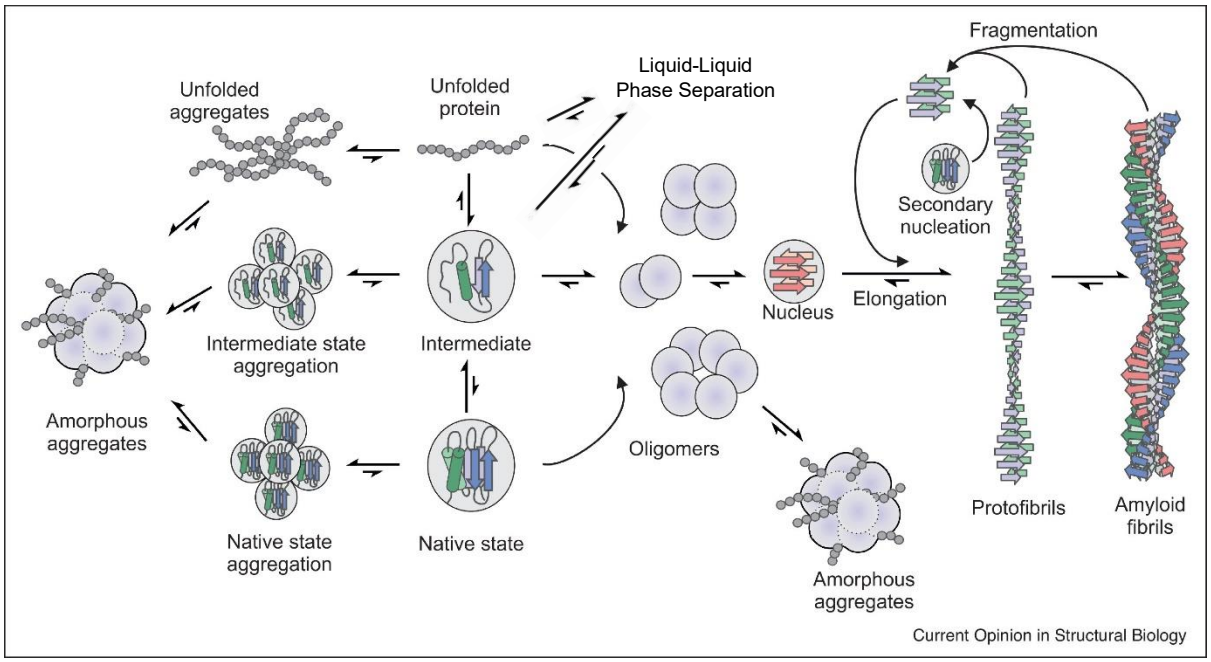


Figure 1.3: Schematic Illustration of Protein Aggregation

Schematic illustrating protein states leading to amorphous or fibrillar aggregates. Taken from (Ebo *et al.*, 2020). Adapted to include liquid/liquid phase separation (Ruff *et al.*, 2022).

Fibrillar aggregates are formed by addition of monomers to a polymer, which extends linearly and the rate of which is dependent on the available fibril ends; the presence of more short lengths of fibril increases aggregation rate compared to if fewer longer aggregates was present. This type of aggregation is more often observed in protein associated with amyloid disease, such as amyloid-β42 and α-synuclein (Borgia *et al.,* 2013). Alternatively, amorphous aggregate formation is thought to proceed by addition of monomers or oligomers in 3 dimensions, with surface area of existing aggregate a rate-limiting factor. Biopharmaceutical aggregation is associated with formation of amorphous aggregates (Borgia *et al.,* 2013). The structural character displayed by an aggregate is dependent on both the protein in question, and the stimuli to induce aggregation. For example, pH-induced aggregation may be reversible in some proteins, whereas temperature-induced aggregation may not (Kuroda, Y., 2022).

Assembly of monomers into an oligomer proceeds with a rate dependent on the monomers available in a system; this is primary nucleation. Alternatively, the surface of existing aggregates may catalyse aggregation, with rate dependent on both available monomer and existing aggregate; this is secondary nucleation. In fibrillar aggregates, release of fragments can provide these catalytic 'nuclei' (fig 1.3)(Cohen *et al.*, 2013).

With hydrophobic interactions being essential for protein folding, they can also contribute to the protein-protein interactions that lead to aggregation. Alongside these, van der Waals and ionic interactions are thought to contribute to aggregation (Mahler *et al.*, 2009). In addition, residues capable of forming disulphide or dityrosine bonds may lead to protein aggregation, particularly in oxidising environments (Cabra *et al.*, 2008)(Malencik and Anderson, 2003). These interactions can be reversible, and disaggregation may be facilitated with the unfolding stresses described in section 1.1.2, but are often not, especially under the conditions encountered *in vivo* (Booth *et al*., 1997)(Wang and Roberts, 2018).

Interestingly, the role of conformational stability in aggregation is not consistent for all proteins; the range of interactions responsible for aggregation dictate that unfolding could lead to protein-protein interactions, or repulsion, and a reduction or increase in colloidal stability. These effects are conditions and protein-dependent (Wang and Roberts, 2018).

1.2.2 Aggregation-Prone Regions

A prominent model for aggregation of natively folded proteins is through the exposure of aggregation-prone regions (APRs) by conformational changes (Booth *et al*., 1997)(Fink, 1998). These are short (5-15 residue) stretches of protein sequence that can interact inter- or intra-molecularly and facilitate oligomer formation. These oligomers can then act as nuclei for further aggregation, facilitating addition of further monomers to the aggregate (fig 1.3).

APRs have properties that are reflective of the amino acids of which they are comprised. These often contain amide side chains, or are hydrophobic (De Baets *et al*., 2014), and are rarely charged; exposure of charged residues can promote solubility. Serine and threonine can also be constituents of APRs, thought to be due to their high relative β-sheet propensity (Wang *et al*., 2009). Whilst aggregation of intrinsically disordered proteins containing APRs can be more easily inferred, APRs in structured proteins may be sequestered from solution under native conditions; making it more difficult to predict. In these proteins, the aggregation potential of an APR is therefore dependent on both its content and likelihood of exposure to protein-protein interactions (Van Durme *et al.*, 2016).

Additionally, the effects of 'gatekeeper' residues may dampen the aggregation potential of an APR. Charged residues including arginine, lysine, aspartic acid and glutamic acid can inhibit aggregate formation due to charge-charge repulsion between side groups. In addition, proline is thought to disrupt formation of β sheet-containing aggregate, due to its limited access to the psi and phi angles required compared to other amino acids, and inability to form H-bonds with the polypeptide backbone of other amino acids (Beerten *et al.*, 2012). The high conformational flexibility in glycine also disrupts formation of β-sheets, due to the entropic penalty of introducing a more sterically-constrained structure (Monsellier and Chiti, 2007). One of the above residues are frequently found at the end of aggregation-prone regions, and act as gatekeepers, reducing aggregation propensity (Rousseau, Serrano and Schymkowitz, 2006).

1.2.3 Consequences of Protein Aggregation

Problems with aggregates:

The transformation of monomer into aggregate has received considerable attention in recent years, with protein aggregation frequently considered problematic in the context of both disease and biopharmaceutical development. Aggregation of amyloid precursor

molecules is thought to be responsible for the formation of oligomers, to which the symptoms of neurodegenerative disorders are attributed, including Alzheimer's and Parkinson's diseases (Chiti and Dobson, 2017).

In addition, aggregation is thought to render proteins dysfunctional, as well as altering viscosity and solubility; factors of particular attention in medicinal proteins (Amin *et al.*, 2014). *In vivo*, immunogenicity is also thought to occur when part of the aggregate is recognised by T or B-cells as an epitope, against which antibodies are produced (Rosenberg, 2006)(Moussa *et al.*, 2016). Cytotoxicity has also been reported, with insulin aggregates showing toxicity to pancreatic β cells (Grudzielanek *et al.*, 2007). Toxicity from oligomers that form *en route* to fibrils (as illustrated in fig 1.3) has also been proposed, with α-synuclein oligomers thought to damage neuronal cells by insertion into cell lipid bilayers (Cascella *et al.*, 2022).

Potential Benefits of Aggregation

However, the potential benefits of aggregation have also been highlighted. Formation of reversible, colloidally stable aggregates show promise in storage and administration of therapeutics, potentially stabilising proteins against unfolding/loss of function, and promoting a more controlled 'sustained release' in patients (Dekel *et al.*, 2017).

In addition, usage of aggregate in materials science is developing; the structural integrity and consistent structure of β-sheet rich amyloid has given rise to suggestions of use as tissue scaffolds for injury repair (Cox *et al.*, 2005). Despite this optimism, uncontrolled protein aggregation is generally considered detrimental and a problem to be addressed, especially in the context of biopharmaceuticals.

1.3 Biopharmaceuticals

Biopharmaceuticals are a subset of therapeutics which are broadly defined as being "inherently biological in nature, and manufactured using biotechnology" (Rader, 2008). Although biopharmaceuticals can include enzymes, recombinant proteins and nucleic acid-based therapeutics, monoclonal antibodies (mAbs) make up the greatest proportion of biopharmaceuticals in use today (Elsayed *et al.*, 2023). Understanding the causes and prevalence of aggregation in biopharmaceuticals is of increasing importance, as this group of therapeutics is being adopted more and more rapidly to treat a huge range of disorders, reflected by the 10-fold increase in international non-proprietary names

(INNs)(meaning the molecule has at least entered clinical trials) in use over the past 15 years (Wilkinson and Hale, 2022).

Whilst a range of antibody formats exist, by far the most widely established in biopharmaceutics is that of IgG (immunoglobulin G). A standard IgG scaffold consists of two heavy and two light chains, which themselves are made up of 4 and 2 β-sheet rich immunoglobulin domains respectively, each stabilised by an intra-domain disulfide bond (fig 1.4)(Chiu *et al.*, 2019). The antibody binds to an antigen through the complementarity-determining regions (CDRs), binding sites found on the F_{ab} (antigen-binding fragment) 'arms'; the two regions that are connected to the F_c (crystallisable fragment) region by flexible linkers. This structural makeup offers a high degree of mobility in the molecule, while the β-sheet-rich structure conveys stability.

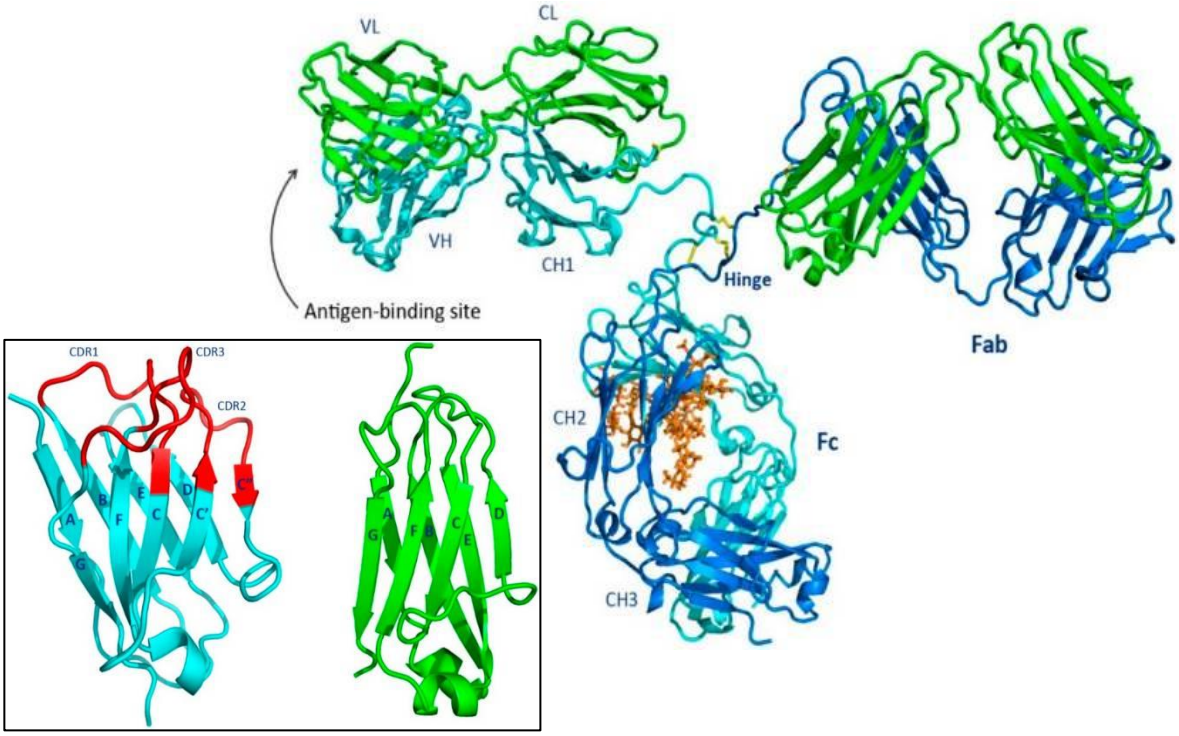


Figure 1.4: IgG Structure

Cartoon crystal structure showing a mouse IgG2a (PDB: 1IGT). Disulfide bonds are shown in yellow, stabilising glycan shown in orange. Inset: the β-sheet rich immunoglobulin fold of a F_{ab} region (cyan) and F_c region (green). The CDR regions are highlighted in red. Taken from Chiu et al., (2019).

1.3.1 Aggregation in Biopharmaceuticals

Aggregation has been shown to occur in a range of biopharmaceuticals; thought to be predominantly due to instability and unfolding of the native state leading to formation of amorphous aggregate (Pang *et al.*, 2023). Interestingly, it is thought that partial unfolding is a more significant contributor to aggregation than complete unfolding, due to the relatively high conformational stability of these molecules, and the fact that partially unfolded states are significantly more energetically favourable than the completely unfolded state (Wang, 2005). Additionally, the interaction of APRs in the F_v region of the antibody is thought to be the source of aggregation more often than in the rest of the protein; likely because of the requirement for high amino acid binding affinity often found in CDRs, which can be misappropriated to cause aggregation (Wang, Singh and Kumar, 2010). Hence, studies into biopharmaceutical aggregation are often focused on the F_v domains.

1.3.2 Causes of Aggregation in Biopharmaceuticals

A range of factors can initiate aggregation through conformational change, as outlined in section 1.1.2. However, the likelihood of protein aggregation can be increased by exposure to different conditions, many of which are experienced at different points throughout the lifecycle of a biopharmaceutical, as illustrated in figure 1.5.

High local protein concentrations have been found to correspond to protein aggregation, both *in vitro* and *in vivo*. Overexpression of β-lactamase was shown to increase aggregation in *E. coli* (Bowden and Georgiou, 1990), while aggregation in purification of biopharmaceuticals has been partially attributed to the high concentrations involved in overexpression (Paul *et al.*, 2015), filtration, chromatography (Arakawa, Ejima and Akuta, 2017) and formulation (Shire, Shahrokh and Liu, 2004) steps.

pH changes can be introduced during protein expression or formulation steps, but perhaps more obviously elution from a protein A column and subsequent pH hold for viral inactivation exposes proteins to low pH (Mazzer *et al.*, 2015). Aside from the effects on protein folding discussed in section 1.1.2, pH changes are thought to promote aggregation if a change in pH results in reduced charge repulsion between protein molecules; i.e. if pH changes to being near the protein's isoelectric point. A similar effect is observed with changes in ionic strength; charge shielding effects can reduce repulsion between proteins (Pandey, 2022).

Whilst excipients such as polysorbates are widely used to increase protein stability, fatty acid degradation products from polysorbates may also induce aggregation by providing a hydrophobic nucleus for aggregation (Larson *et al.*, 2020).

Hydrodynamic effects can be present throughout any stage that handles biopharmaceuticals as a liquid, from cell culture to administration in patients (Rathore and Rajan, 2008). This can include turbulence effects, extensional flow, shear, drag, positive and negative pressure effects, and momentum-based effects from bends in pipes. The implications of these forces are a subject of significant debate, and they are relatively understudied considering their ubiquitous nature in bioprocessing, and are frequently included under the umbrella term of 'shear' (Chalmers, 2015)(Thomas and Geer, 2011). Section 1.4.3 discusses the evidence for hydrodynamic flow-induced unfolding and aggregation in further detail.

Surfaces of varying materials are present throughout a biopharmaceutical's lifecycle, and have been shown to induce aggregation. The mechanism of this is subject to some contention, and depends on the surface in question; charged surfaces are thought to promote electrostatic interactions and aggregation, while more hydrophobic surfaces or air-liquid interfaces, such as those found in IV bags may cause interface-catalysed unfolding, similarly to as described in section 1.1.2 (Wu and Randolph, 2020)(Galush and Horst, 2015)(Kopp *et al.*, 2023). Shedding of particles from pumps and tubing can also provide larger surfaces areas to catalyse aggregation in this way (Adler and Allmendinger, 2023). Similarly, freeze/thaw cycles are thought to induce aggregation by producing ice crystals, which provide an ice-water interface which can catalyse protein denaturation. Freezing can also result in solute crystallisation, resulting in changes to pH and ionic strength of solutions (Cao *et al.*, 2003).

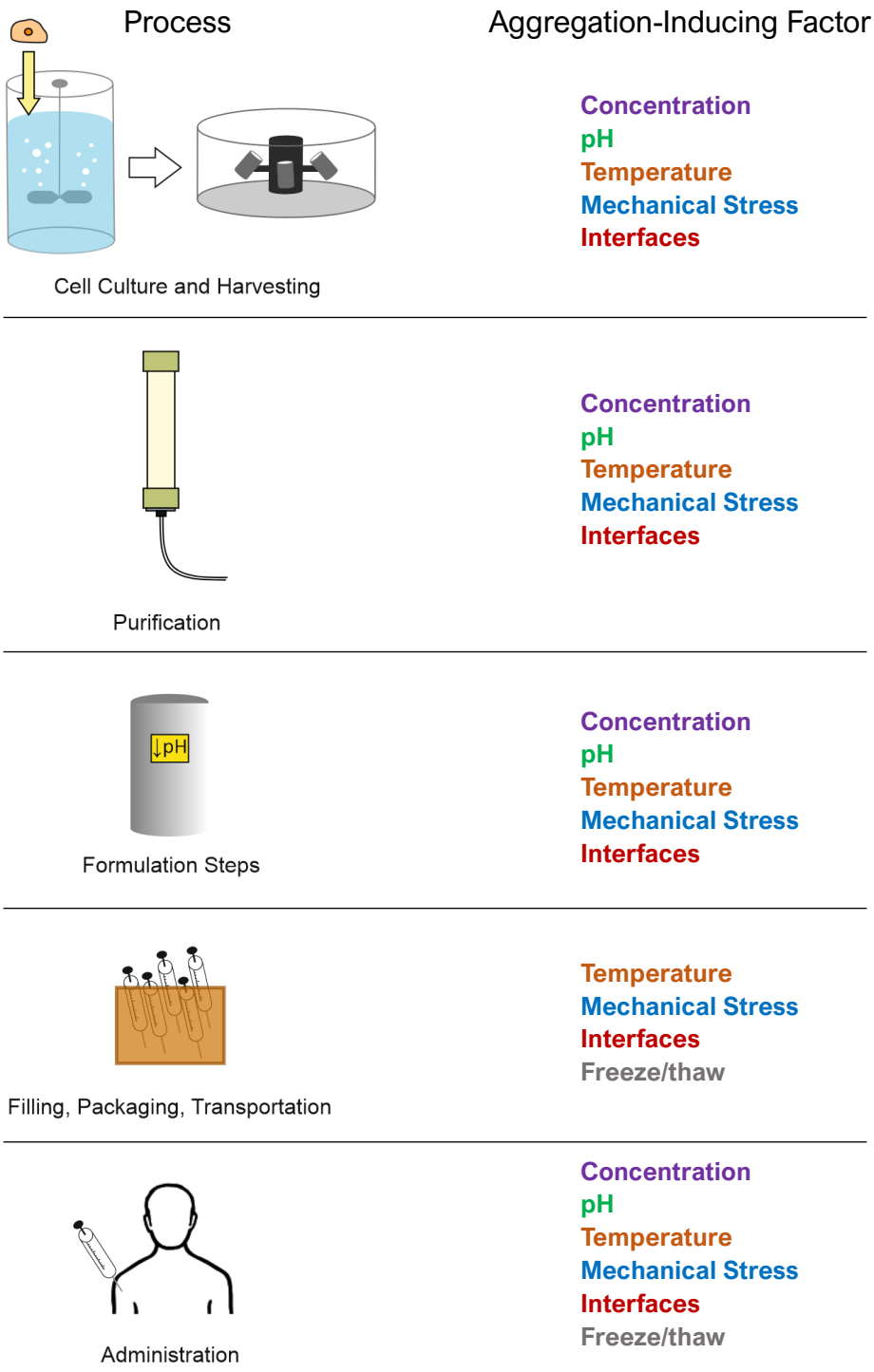


Figure 1.5: Aggregation-Inducing Steps in Biopharmaceutical Production

Schematic illustrating steps associated with biopharmaceutical production, and the aggregation-inducing factors that may be present at each step. Adapted from Elsayed *et al*., (2023).

1.3.3 Developability Studies in Biopharmaceuticals

The huge range of factors contributing to aggregation, and realities of the processes in a biopharmaceutical's lifecycle, makes identification and removal of causative factors very difficult. Therefore, focus is given to improvement of inherent aggregation-resistance; understanding the features that determine a molecule's propensity to aggregate is of vital worth, not only in developing aggregation-resistant biopharmaceuticals, but also in offering insight into disease-related aggregation.

Aggregation propensity and proclivity to retain an intact, functional format are features often summarised by description 'stability'. Measurements of a molecule's stability should be reflective of the broad stresses encountered during the biopharmaceutical's lifecycle, and aim to indicate how stable the molecule will be in practice. The process of analysis to predict how likely the molecule is to be stable and efficacious, or how 'developable' it is, should be done early in a candidate's development, use minimal material, and should consume as little time as possible (Zhang *et al.*, 2023). For these reasons, assays must be carefully selected, and a selection of popular developability assays have arisen as a result, discussed below.

1.3.4 Initial Developability Measurements

Initial characterisation of a candidate's 'stability' seeks to offer insight into a range of stability aspects, with minimal time investment. Popular measurements of structural stability are thermal assays, as they can measure a small volume of sample, in a multi-well plate format. Thermal shift assays can be used to measure unfolding to this end, and are often based on either binding of a dye to hydrophobic unfolded regions, in the case of DSF, or on calorimetry, in the case of DSC (both discussed in section 1.1.3)(Wolf Pérez *et al.*, 2022).

As promiscuous interactivity can be indicative of aggregation propensity, assays to probe for this are commonly used in developability. Popular methods include cross-interaction chromatography (CIC), which uses human antibodies as the stationary phase, and assays based on enzyme-linked immunosorbent assay (ELISA), which detect interactions with particles such as baculovirus, which contains a mixture of nucleic acid, carbohydrate, lipid and protein (Jacobs *et al.*, 2010)(Hötzel *et al.*, 2012). Retention times measured by various other chromatography-based techniques are also thought to be indicative of

aggregation propensity, for example standup monolayer adsorption chromatography (SMAC), which relies on elution through a proprietary Sepax Zenex SEC column (Kohli *et al.*, 2015). Additionally, techniques that analyse protein surface hydrophobicity, such as hydrophobic interaction chromatography (HIC), can be indicative of aggregation potential. Surface hydrophobicity is measured by changes in elution time as the protein interacts with the hydrophobic column stationary phase (Haverick *et al.*, 2014).

Assays to detect self-interaction propensity are also frequently used in developability. These can include affinity-capture self-interaction nanoparticle spectroscopy (AC-SINS) and self-interaction chromatography (SIC)(Liu *et al.*, 2014)(Tessier *et al.*, 2002). These techniques measure interactions between either nanoparticles or a chromatography resin, both coated with the protein of interest, and the free protein.

1.3.5 Accelerated Stability

Whilst more time-consuming than many initial measurements, accelerated stability (AS) is used as a developability parameter that is potentially more reflective of stresses a protein would experience, in which samples are held at varying temperatures, usually over the course of several weeks. This is more time-consuming than taking initial measurements, but only requires measurements of aggregate following incubation, usually by HPLC-SEC (Jain, Sun, *et al.*, 2017). AS is usually carried out around 25-40 °C and aims to be predictive of refrigerated (2-8 °C) storage conditions (Evers, Clénet and Pfeiffer-Marek, 2022).

However, despite the perceived simplicity of AS, studies are increasingly finding that mAb degradation proceeds by different pathways at different temperatures (Wälchli *et al.*, 2020). Bunc *et al*., (2022) suggested that at higher temperatures, aggregation proceeded by partial unfolding events, whereas at lower temperatures it was dominated by post-translational modifications. The authors also showed that predicting long-term stability (3 years) at 5 °C was not accurate using monomer loss data at 40 °C over 3 months, but demonstrated that using samples incubated at 25, 35 and 40 °C over 3 months, and integrating data into a model that described the two degradation methods allowed for accurate prediction of the long-term stability performance. Similarly, Chakroun *et al*., (2016) found that rank order of aggregation rates at 45 °C did not correlate with that of samples at 4 °C, for a mAb fragment ($A33F_{ab}$) in different solution conditions.

1.3.6 *In Silico* Predictors of Protein Aggregation

As described in section 1.1.1, the primary sequence of a protein is the fundamental determinant of its structure and stability. Increasingly, this is being utilised by predictive *in silico* tools to determine features of protein developability.

Stretches of sequence that contain aggregation-prone regions can be detected, using several open-source tools. These tools include Aggrescan (Conchillo-Solé *et al.*, 2007) and Zyggregator (Tartaglia and Vendruscolo, 2008), which both predict for aggregation based on the intrinsic aggregation propensity associated with amino acids, a value assigned from experimental studies. Alternatively, Waltz (Maurer-Stroh *et al.*, 2010), TANGO (Fernandez-Escamilla *et al*., 2004), PASTA (2.0)(Walsh *et al.*, 2014), and Spatial Aggregation Propensity (SAP)(Chennamsetty *et al.*, 2009) use more specific characteristics of amino acids such as charge and sequence, to predict aggregation propensity.

More bespoke algorithms for predicting whole-protein aggregation susceptibility also exist; SolubiS is an example where fold stability predictions by FoldX protein folding predictor are combined with APR predictions by TANGO, to give a score for overall aggregation-proneness (van der Kant *et al.*, 2019). FoldX predicts the change in thermodynamic stability when residues are mutated, and TANGO predicts the effect of mutations on β-sheet forming potential (Schymkowitz *et al.*, 2005)(Fernandez-Escamilla *et al.*, 2004).

The type of aggregation predicted for also varies between the algorithms; as TANGO and PASTA (2.0) predict for β-sheet forming potential, these algorithms may be more appropriate for predicting fibril-like aggregation formation, whereas Aggrescan and Zyggregator predict for aggregation including amorphous aggregation.

Therapeutic Anitibody Profiler (TAP) specifically addresses biopharmaceuticals, scoring molecules in 5 categories relative to a database of existing molecules (Raybould *et al.*, 2019). These 5 parameters are thought to be predictive of aggregation susceptibility, and are:

1) Total CDR length
2) Surface-exposed hydrophobic patches in the CDR
3) Regions of positive charge in the CDR
4) Regions of negative charge in the CDR
5) Patches of “charge symmetry” in the V_H and V_L domains; where one is negatively charged, and one is positively charged

1.3.7 Use of Developability Parameters for Protein Engineering

The idea of particular regions being principal causes of aggregation-propensity lends itself to the concept of protein engineering; designing sequences that are less susceptible to aggregation.

A range of strategies could be employed to prevent protein aggregation. In the examples where aggregation is a result of unfolding, enhancing the stability of the folded state is often a priority. Introduction of cross links such as disulfide bonds, amino acid substitutions and addition of ligands that stabilise the native state have all been evidenced to increase thermodynamic stability significantly (Pace, 1990). Of these options, amino acid substitution is the most broadly feasible; introduction of disulfide bonds can interfere with existing native interactions, and involvement of ligands can be impractical in molecules that bind to specific epitopes or substrates, such as biopharmaceuticals and enzymes.

Neutralising APRs with targeted mutagenesis may be logical, especially given the wealth of methods available to detect these stretches of sequence. Introduction of aggregation gatekeeper residues (discussed in section 1.2.2) has been shown effective in aggregation-prone biopharmaceuticals (van der Kant *et al.*, 2017).

However, given APRs are frequently found in CDRs, it is likely that retention of biopharmaceutical activity must be considered. As such, mutations must be selected carefully, with efficacy considered as well as stability. Alternative strategies to inhibit aggregation have also been investigated, with Courtois *et al*., (2016) generating AsnXThr and AsnXSer glycosylation sites (where X can be any amino acid except proline) in a mAb to shield APRs from interacting, without encroaching on native antigen-binding sites.

Mutations could also be focused on F_c regions, although this would only inhibit aggregation in a subset of cases. Chennamsetty *et al*., (2009) used SAP (section 1.3.6) to predict APRs in two antibodies, and showed increases in T_m, as well as reduction in aggregation, by mutating hydrophobic residues to charged residues, mostly in the F_c regions. However, this may only be applicable to a limited number of cases, as the less mobile nature of F_c regions means it is likely that structural perturbation is required to expose these APRs; neutralising F_c region APRs may affect stress-induced but not native-state aggregation.

Recently, protein engineering has become more high-throughput, with modalities developed to rapidly screen for specific substitutions that lower aggregation propensity

(Ebo *et al.*, 2020). Whilst these focus on the reduction in aggregation as the result, analysis of a specific variant from a mAb population screened in this way suggests that this is a result of increased conformational stability; indicating the beneficial effects of protein engineering may not always be a result of APR nullification (Lawrence, 2023).

1.3.8 Use of Developability Parameters for Protein Stability Predictions

As the different developability assays probe a range of features, recent studies have attempted to verify how these assays correlate with one another, with the goal of establishing an optimal 'toolkit' for developability analysis, which could most accurately predict a molecule's success. (Jain, Sun, *et al.*, 2017) applied a range of 12 assays to a panel of 137 antibodies, including many of those discussed in section 1.3.4; HIC, SMAC, CIC, AS, BVP ELISA analyses were among those included. In order to assess for redundancy between assays, these were grouped by hierarchical clustering. Although the assays mostly grouped as expected, AS was not grouped with any assay. This is interesting, because it indicated that although AS is the most time-consuming metric, the insight it provides is not offered by other assays in the dataset. With AS being used in biopharmaceutics to assess long-term stability over a condensed timeframe, other assays that correspond to this metric could be valuable.

The developability assays currently employed appear to be reasonably comprehensive in assaying for protein-protein interactions, as well as thermodynamic stability. However, mechanical stability is overlooked, with no regularly implemented assays reflecting how a molecule will react under the varied and significant mechanical forces encountered during biopharmaceutical production and handling. Recently, colleagues in the group have demonstrated that aggregation propensity following exposure to hydrodynamic flow may be predictive of a molecule's performance (Willis *et al.*, manuscript in preparation). Flow-induced aggregation was measured by HP-SEC in a panel of 9 different antibody/buffer formulations, and the panel was assayed using a range of techniques to produce 33 total developability parameters. Following hierarchical clustering of ranked data, flow-induced aggregation was found to group with long-term stability, at both 5 and 25 °C; two clinically relevant conditions. This highlights the possibility that a much more rapid analysis by flow-induced aggregation may be useful as a developability parameter that is predictive of protein stability over longer timeframes.

1.4 Hydrodynamic Flow

As described in section 1.3.2, hydrodynamic flow is near-ubiquitous throughout a biopharmaceutical's manufacture, processing and use lifecycle. Despite this, it has remained relatively understudied compared to other aspects, such as formulation.

While hydrodynamic flow encompasses a range of effects (section 1.3.2), shear and extensional flow are of particular interest, due to the large number of processes in biopharmaceutical manufacture that are expected to involve them. When fluid is flowing in a particular direction through a confined environment, such as a pipe, drag against the pipe walls can introduce a velocity gradient, whereby directional flow is greater in the centre of the pipe than at the edges (fig 1.6A). This velocity gradient is the source of shear effects; where molecules in the fluid are propelled anisotropically by flow (King, 2002). When fluid velocity increases in the direction of flow, for example in a pipe constriction, a gradient in velocity occurs (fig 1.6B). This can stress a molecule longitudinally, a phenomenon termed extensional (or elongational) flow (Keller and Odell, 1985).

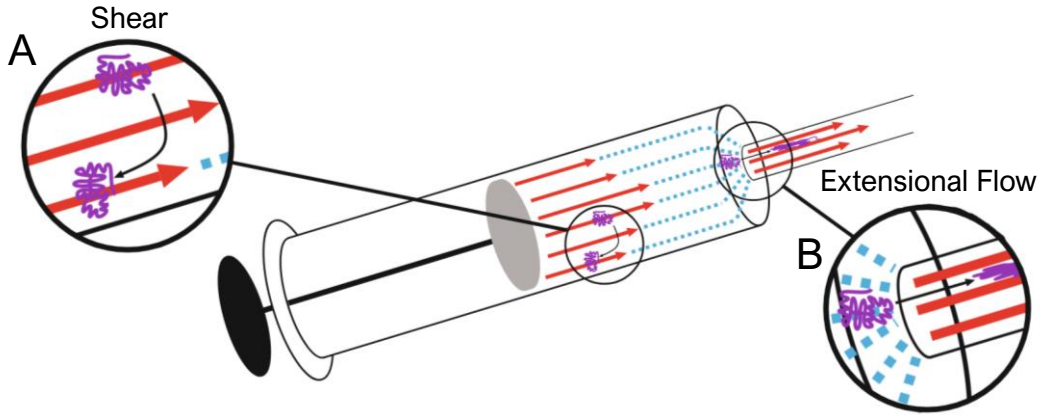


Figure 1.6: Shear and Extensional Flow

Schematic illustrating A) shear and B) extensional flow. A syringe is used as an example system in which these hydrodynamic forces may occur. These forces can be applied to a range of fluid flow environments. Red arrows indicate directional increase in fluid velocity.

The force imparted on a molecule in solution by shear is described as the shear rate (γ), and is velocity of the fluid in each layer (m/s)/distance between layers (m)(Bekard *et al.*, 2011). Therefore, shear rate units are s^{-1}. Strain rate resulting from extensional flow ($\dot{\varepsilon}$) is

described as dimensionless relative change in molecule length (ε)/time (s), giving units of s^{-1} for $\dot{\varepsilon}$.

1.4.1 Magnitudes of Hydrodynamic Forces

As hydrodynamic forces can occur at various stages in biopharmaceutical production and use, they are varied in presence and intensity, making a defined magnitude for experimental investigation difficult to prescribe. However, attempts have been made to model the shear forces in mixing devices such as rotators, orbital incubators, magnetic stirrers and vortex mixers, at varying RPM rates; revealing calculated maximum shear rates from 1711 to 8315 s^{-1} (Bai *et al.*, 2012). Additionally, modelled shear rates of over 15,000 s^{-1} have been calculated in homogenisers (Maa and Hsu, 1996). Shear rates in vial filling are also likely to be substantial; vial filling through a 10 cm long 20-gauge needle (0.6 mm internal diameter) was estimated to result in shear rates of 20,000 s^{-1} for 50 ms on a protein (Bee *et al.*, 2009).

Although offering a ballpark figure for shear rates to consider, this leaves the extent of hydrodynamic flow in other components such as filtration apparatus unaddressed, and does not consider extensional flow. The broad range of shear rates, and likely extensional flow strain rates, found in biomanufacture and processing equipment correspond to largely differing levels of aggregation induced. For example, comparison of orbital shaking over 4-24 hours with hydrodynamic flow-induced aggregation using a syringe and capillary-based flow device inducing high extensional strain rates, did not result in any correlation (Willis *et al.*, 2023). This highlights the importance of using instrumentation that can probe both extensional and shear-based flow fields, and indicates the use of relevant equipment in protein aggregation studies may only be applicable to the industrial process it is associated with.

Studies on shear have addressed a wide range of shear rates, perhaps for several reasons; the instrumentation available, lack of known magnitudes in processes of interest, and the fact that fundamental studies on these forces generally focus on a range of rates required to illicit an effect. Shear rates from at least 9.15 s^{-1} to 1 x 10^8 s^{-1} have been investigated (Bekard *et al.*, 2011)(Duerkop *et al.*, 2018). Some of the highest shear rates have not shown any effects on conformation and aggregation, while some of the lowest have, and vice versa. This varied response to flow has resulted in the field being divisive, although observed effects are strongly based on the molecule under consideration.

1.4.2 The Mechanism of Flow-Induced Protein Aggregation

The mechanism for flow-induced aggregation is not confirmed, however a consistent observation is that rather than proceeding solely in bulk, surface-mediated unfolding and/or aggregation plays a role (section 1.1.2). This is reflected by the concentration dependence of flow-induced aggregation; as protein concentration increases, rate of protein aggregation decreases as a percentage of total protein concentration. This inverse concentration dependence can be explained by the presence of two pathways through which a protein can aggregate; a 'bulk' in-solution pathway, and a surface-mediated pathway, which can become saturated at higher concentrations, resulting in a reduction in Δ% soluble protein loss (fig 1.7)(Treuheit, Kosky and Brems, 2002).

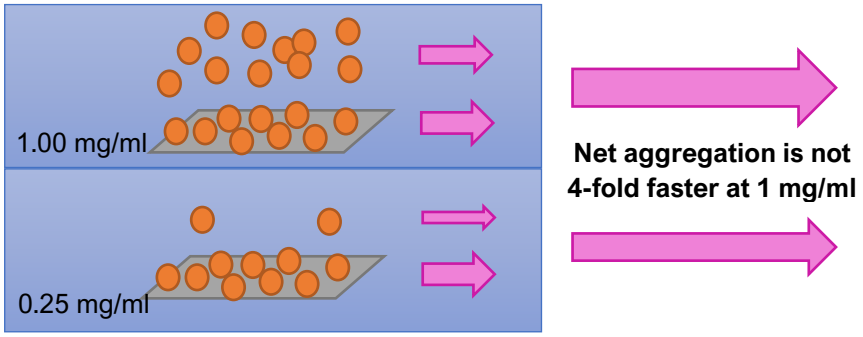


Figure 1.7: Inverse Concentration Dependence of Flow-Induced Aggregation

Surface-mediated aggregation stops increasing as the surface becomes saturated at higher concentrations, meaning a slower, bulk aggregation pathway is the only option to increase aggregation rate. Concentration denotes protein concentration, orange balls indicate protein molecules, pink arrows indicate aggregation proceeding, with thickness schematically proportional to rate.

Surfaces may catalyse aggregation similarly to as described in section 1.3.2, but with the additional consideration that flow and/or scraping effects from equipment may dislodge partially unfolded and/or aggregated proteins from surfaces, dispersing them and increasing sites for secondary nucleation (Kopp *et al.*, 2023).

Colleagues (Willis, Panagi, *et al.*, manuscript in preparation) have established a model which explains this aggregation behaviour in IgG mAbs, demonstrating that to describe the data, terms for both surface and bulk-mediated change of protein into an aggregation-prone state, and bulk-mediated aggregation, are necessary. In short, this is termed the surface and bulk nucleation with bulk aggregation (SB_NB_A) model. Figure 1.8A illustrates the model and includes the rate constants assigned to each step. Using a hydrodynamic flow device (described in section 1.4.4 below) to induce protein aggregation in an IgG, STT, the model fits experimental data well (fig 1.8B) Establishing a model to describe this process is a significant step in verifying which parts of the aggregation pathway are aggregation-determining for various molecules, allowing a greater understanding of the steps to be targeted in aggregation prevention.

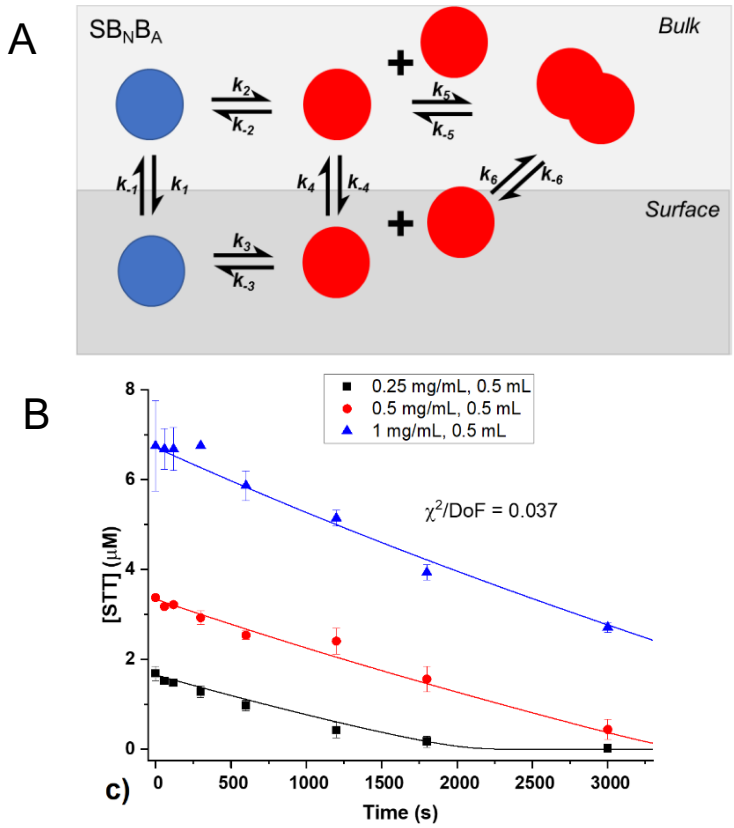


Figure 1.8: A Model to Describe Flow-Induced Aggregation

Surface and bulk nucleation with bulk aggregation (SB_NB_A) kinetic model schematic for flow-induced protein conformational change and aggregation, developed by Willis *et al*. (manuscript in preparation 2). Blue circles indicate molecules with native structure, red circles indicate molecules in an aggregation-competent state. B) Simulated soluble protein loss with flow-induced aggregation in STT, a monoclonal antibody. Datapoints produced experimentally shown as filled shapes. Solid lines indicate simulated fits to the data. Exposure to hydrodynamic flow shown increasing over time, with a calculated maximum $\dot{\varepsilon}$ of 11,750 s^{-1}. All data are N = 2, error bars = s.d. Taken from Willis *et al*. (manuscript in preparation 2).

Whilst the model contains 12 terms, 5 of these (k_1, k_5/k_{-5} and k_6/k_{-6}) remain fixed. Willis, Panagi, *et al.* analysed other models with the goal of fitting the data, whilst simulating extent of surface saturation values corresponding to those reported by surface ellipsometry.

The relatively simplistic Finke-Watzky (F-W) model of protein aggregation was used to fit the data (fig 1.9A)(Watzky and Finke, 1997). According to this model, protein aggregation proceeds entirely in bulk, and unfolding is promoted by the presence of aggregates. Additionally, surface-only (S_{NA}), and bulk-only (B_{NA}) nucleation and aggregation models were used to fit the data (figs 1.9B and 1.9C, respectively. None of these models were able to fit the data well, and although the bulk-only model showed the best fit, it did not fit the 0.25 mg/ml data well, in line with the observation of increased relative surface involvement in aggregation at lower concentrations, as discussed above.

More complex models including a surface nucleation only with only bulk aggregation (S_NB_A), and surface nucleation only with bulk and surface aggregation (S_NSB_A) model appear to fit data more closely than those previously tested (fig 1.10). However, a protein monolayer is not produced in the model, which contradicts experimental observations (section 3.2.2). Addition of further rate constants also worsened the fit, which led the conclusion that the model presented in figure 1.8 best describes the data, with terms for both surface and bulk nucleation, and bulk aggregation (SB_NB_A).

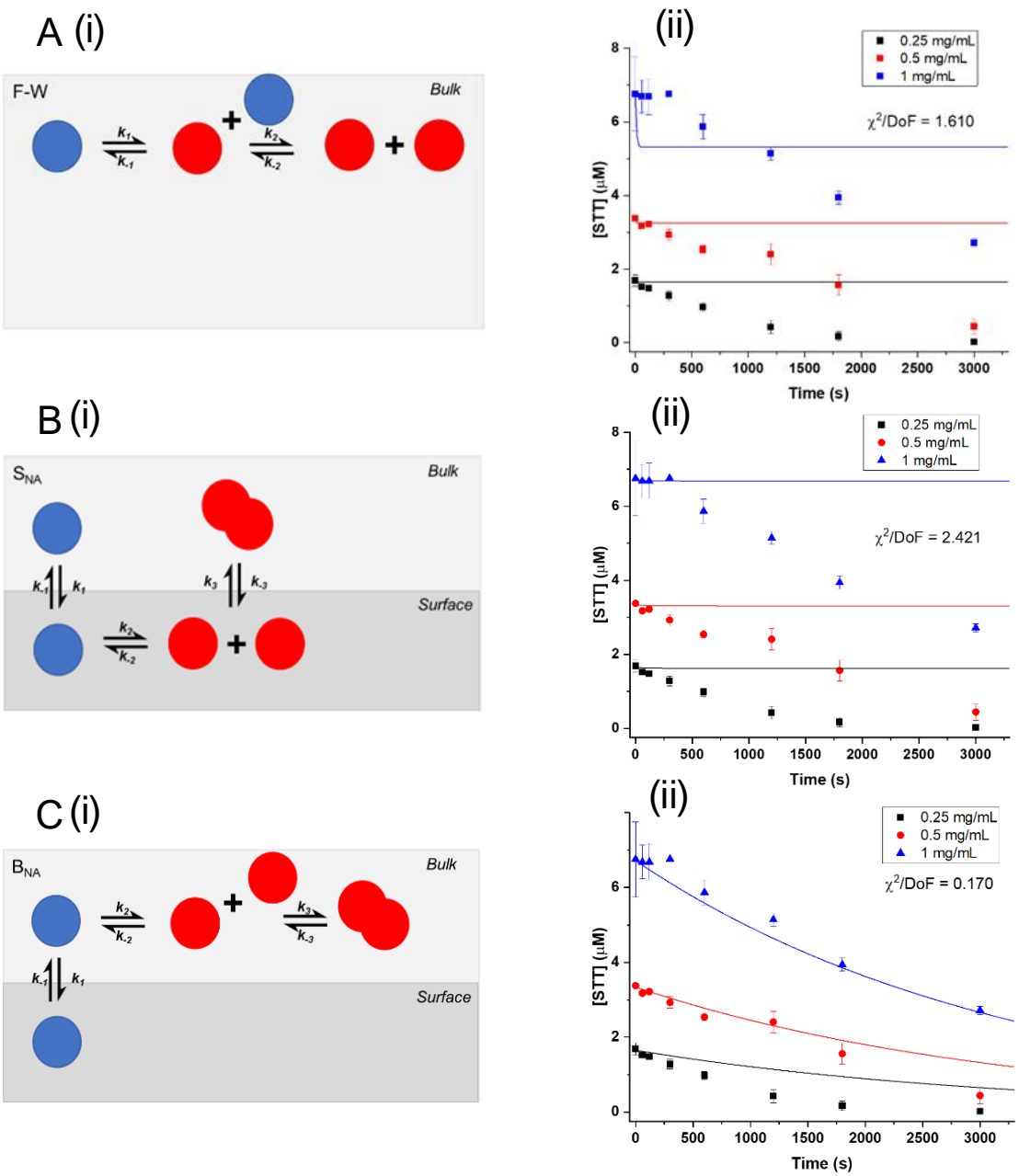


Figure 1.9: Simple Aggregation Models Do Not Describe EFD-Induced Aggregation Data

i) overview and ii) fits of data to a A) Finke-Watzky (F-W) two-state mechanism for aggregation, B) Surface-only nucleation and aggregation model (S_{NA}) or C) Bulk-only nucleation and aggregation model (B_{NA}). Blue circles indicate molecules with native structure, red circles indicate molecules in an aggregation-competent state. Datapoints produced experimentally shown as filled shapes. Solid lines indicate simulated fits to the data. Exposure to hydrodynamic flow shown increasing over time, with a calculated maximum $\dot{\varepsilon}$ of 11,750 s^{-1}. All data are N = 2, error bars = s.d. Taken from Willis *et al*. (manuscript in preparation 2).

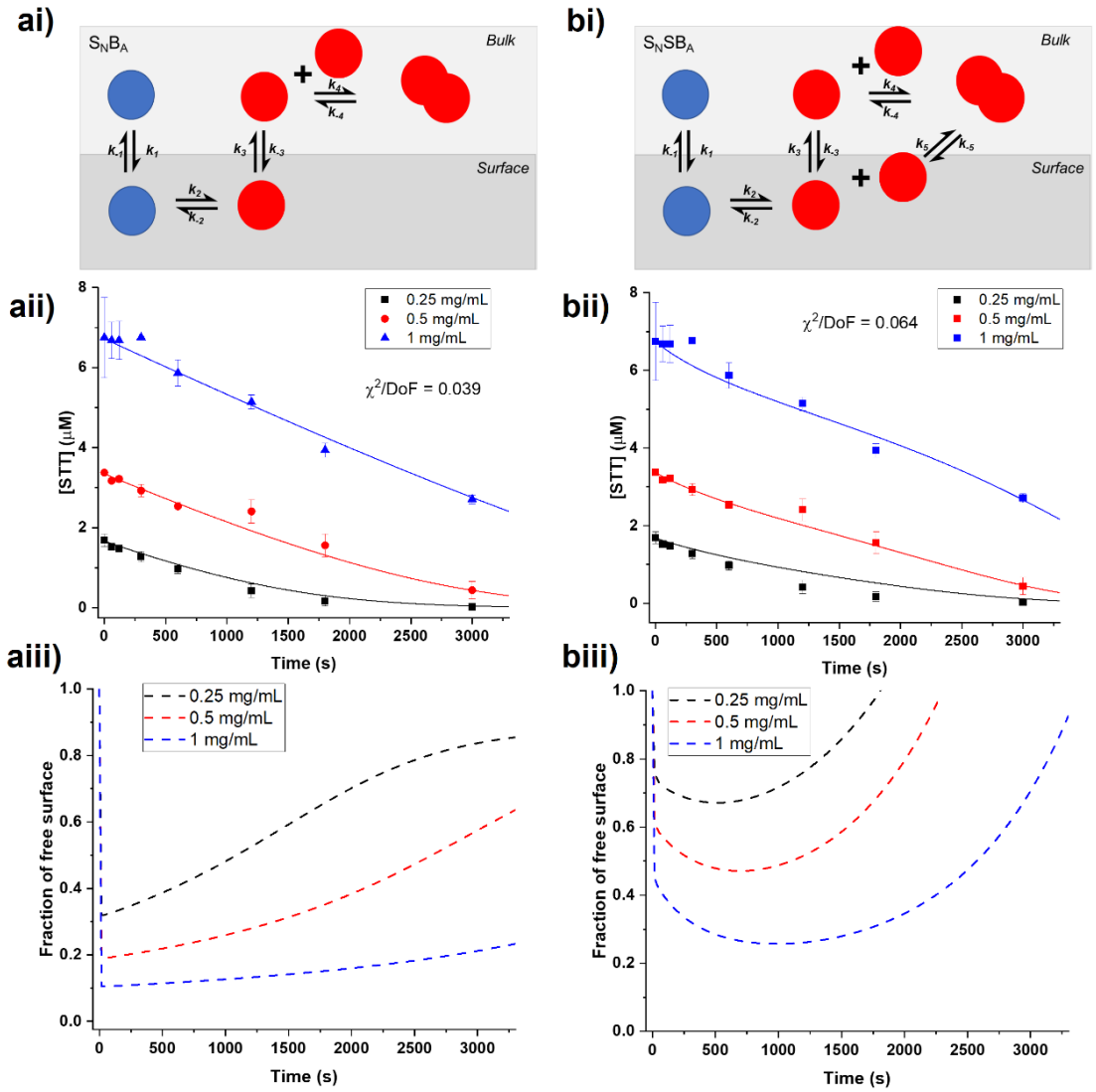


Figure 1.10: Surface Nucleation Alone Does Not Describe EFD-Induced Aggregation Data

i) overview and ii) fits of data to a A) Surface nucleation, bulk aggregation mechanism (S_NB_A) or a B) Surface nucleation, Surface and Bulk aggregation mechanism (S_NSB_A). Blue circles indicate molecules with native structure, red circles indicate molecules in an aggregation-competent state. Datapoints produced experimentally shown as filled shapes. Solid lines indicate simulated fits to the data. (iii) Fraction of free surface as a function of time, as output from each model simulation in KinTek Explorer. $[Surface]_0$ = 0.0338 μM in each case, corresponding to the concentration of antibody if a saturated monolayer was present. Exposure to hydrodynamic flow shown increasing over time, with a calculated maximum $\dot{\varepsilon}$ of 11,750 s^{-1}. All data are N = 2, error bars = s.d. Taken from Willis *et al*. (manuscript in preparation 2).

1.4.3 The Reported Effects of Hydrodynamic Flow on Protein Conformation

Despite the accepted paradigm that unfolding can lead to aggregation, whether flow-induced aggregation arises from unfolding is less understood. Efforts to understand the effects of hydrodynamic flow on proteins were initially focused on conformational change, with aggregation noticed quickly, but not a principal source of investigation until years later.

Shear was first considered, with Charm and Wong, (1970) showing that shear in a Teflon tube could reduce enzyme activity in rennet, catalase and carboxypeptidase, implying conformational change. Conversely, Thomas and Dunnill, (1979) found that shear rates of up to 10^6 s^{-1} in a capillary did not affect catalase activity, indicating that another component of flow may be the cause of unfolding effects previously observed.

Extensional flow was considered when Fuller and Leal (1980) compared shear and extensional flow, using a large polystyrene 'dumbbell' model of between 4 and 8 mDa, comprising two beads connected by a linker. When elongated by flow-induced strain, these produced a change in light scattering. Using a 4-roll mill instrument (fig 1.13A), the authors showed that at a lower strain rate, an extensional flow-dominated flow regime resulted in a 3-fold increase in scattering compared to a predominantly shear-stressed sample, indicating a higher degree of conformational change.

Later, Perkins *et al*., (1997) measured unfolding of individual fluorescently-labelled λ-bacteriophage DNA polymers, consistently 22 μm in starting length, and showed a high degree of heterogeneity in structural response to extensional flow. The authors noted that molecules with an extendable 'dumbbell-like' structure were more susceptible to flow-induced conformational change than the more compactly folded polymers, indicating structural determinants may be important in flow-induced unfolding. This indicates that morphology may be a key determinant in flow-induced unfolding and aggregation. Although amorphous aggregation has been more widely reported in mAbs than fibrillar aggregation (discussed in section 1.2.1), if a mAb formed fibrillar aggregates, its morphology may dictate that it is more susceptible to further unfolding and aggregation.

In support of the more significant potency of extensional flow versus shear, Sing and Alexander-Katz, (2010) found that shear rates of approximately 10,000 s^{-1} were required for conformational change in von Willebrand Factor (vWF), whereas unfolding occurred at extensional strain rates of only 300-600 s^{-1}. Similar results have been produced using

a range of molecules, protein and otherwise, with the general conclusion that hydrodynamic flow acts anisotropically on proteins. This is one of numerous studies involving vWF, which is evolved to experience flow-induced remodelling as part of the

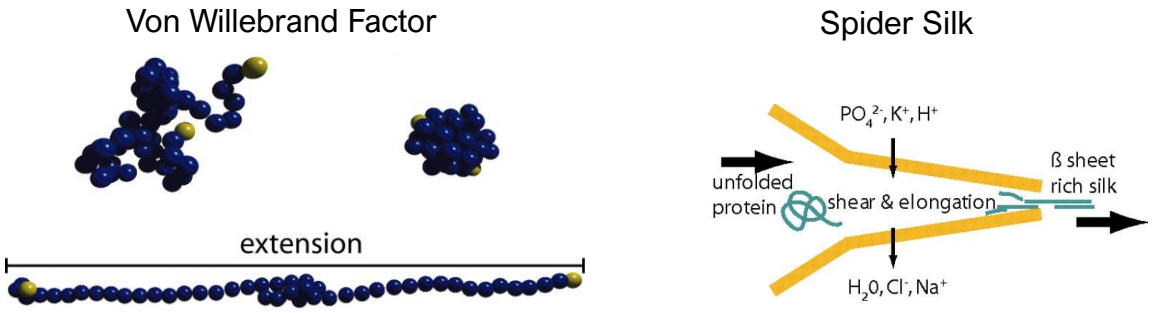


Figure 1.11: Evolved Conformational Change Under Hydrodynamic Flow

Examples of evolutionarily conserved flow-induced conformational change include von Willebrand Factor and Spider Silk. Taken from (Schneider *et al*., 2007) and (Rammensee *et al*., 2008).

blood clotting cascade. In this, flow exposes collagen binding sites and allows a protein network to form, facilitating platelet adhesion (fig 1.11) (Schneider *et al.*, 2007). In vWF, flow-induced conformational changes are predictable and well-understood. Another example is during formation of spider silk, which is dependent on proliferation of extended β sheets under flow (Rammensee *et al.*, 2008).

However, behaviour in globular proteins that have not evolved to access beneficial conformational transitions under flow is less predictable. Whether shear or extensional flow is causative in protein unfolding -or neither- is therefore still a subject of contention for the vast majority of proteins.

1.4.4 Studies with the Extensional Flow Device (EFD)

Through collaboration between the Radford, Brockwell and Kapur labs at the University of Leeds, an extensional flow device (EFD) designed to expose proteins to high extensional strain rates, was developed. Described in detail in section 2.2.5.1, the instrument can expose samples to high extensional strain (up to 2.3 $x10^4$ s^{-1}) and shear rates (up to 1 $x10^5$ s^{-1}), by forcing liquid protein sample from one syringe into another, through a capillary contraction.

Dobson *et al*., (2017) analysed flow-induced aggregation using this device, showing BSA, β-2-microglobulin, granulocyte colony stimulating factor and 3 mAbs all aggregated upon

exposure to flow. The authors also showed flow-induced unfolding of BSA, which was increased upon reduction of its stabilising disulphide bonds; indicating structural determinants that may affect flow-induced aggregation. A significant finding was that halving the length of the capillary connecting the two syringes, and therefore substantially reducing the exposure to shear without affecting extensional flow, did not affect protein aggregation; evidence that shear may not be the primary contributor to flow-induced aggregation.

Subsequently, over different 40 mAbs have been analysed and have shown a range of aggregation propensities under hydrodynamic flow (Willis *et al.*, 2020), with formulation components shown to significantly affect aggregation propensity (Willis *et al.*, 2023). When biopharmaceuticals were investigated in further detail, it was found that strain rate would more strongly influence aggregation in some than others, whereas increasing the number of exposure events always induced more protein aggregation (Willis *et al.*, 2018). This indicates that increasing magnitude of hydrodynamic flow may cause aggregation to a greater extent, perhaps due to more complete unfolding or a larger number of unfolding events, but in a protein-dependent manner. This is a facet of flow-induced aggregation that requires structural insight to be further understood.

These studies demonstrate varied responses to extensional flow and shear exhibited by different proteins, with shear in particular a source of contention, and extensional flow remaining understudied. Generally, it appears that extensional flow is a more potent contributor to protein aggregation than shear, but the intrinsic combination of the two in all studies and lack of insight into conformational changes under extensional flow make a mechanism of flow-induced aggregation difficult to establish.

1.5 Studying the Effects of Hydrodynamic Flow on Protein Unfolding

1.5.1 Methods to Study Hydrodynamic Flow

As significantly more attention has been given to shear, a wider range of devices has been employed to investigate it, compared to extensional flow. Couette chambers contain fluid between two cylinders, which create a unidirectional velocity gradient and expose suspended particles to shear when one is rotated (fig 1.12A)(Bekard *et al.*, 2011). Parallel-plate rheometers, in which liquid is held between two plates, and one is rotated,

have been widely used to study shear (Sharma and Pandey, 2021). A gradient in fluid velocity is created perpendicular to the directional motion of the rotating plate (fig 1.12B(i). Additionally, cone-and-plate devices have been explored, which operate on a similar principle of fluid contained between two rotating components (fig 1.12B(ii))(van der Veen *et al.*, 2004).

As extensional flow requires a directional velocity gradient, the necessity for solid container walls means shear is always present, to some degree, in studies on extensional flow. Therefore, while tuneable to varying degrees, the following techniques expose samples to both shear and extensional flow.

Four-roll mill apparatus modulate flow to be either shear-focused or extensional flow-focused by the directionality of 4 rollers, in a confined fluid system. Co-ordinating rollers so that flow enters the inter-roller space from two channels, and exits by two, increases extensional flow in the system (fig 1.13A(i), whereas unidirectional rotation by all rollers increases shear (fig 1.13A(ii)). Similarly to as described in figure 1.13A(i), cross-slot flow systems induce extensional flow in a system by entry of fluid through two converging channels, and acceleration out of two perpendicular exit channels (fig 1.13B). Cross-slot instruments can increase a molecule's residency time in extensional flow, to more clearly induce effects. Whilst valuable in fundamental studies, this type of prolonged exposure to flow is less reflective of the more transient effects that would be present industrially or clinically.

Capillary and tube-based systems are also popular, and while varying in delivery, generally involve unidirectional flow through a constriction (if extensional flow is of interest) or through a space of unform diameter (if shear is of interest). An example instrument is that of (Grigolato and Arosio, 2020), in which two syringes were connected by tubing and sample was transferred from one syringe to another through the tubing, by syringe compression (fig 1.13C). It is within this category of instruments that the similar EFD, described in sections 1.4.4 and 3.2.1, is found.

Microfluidic systems comprised of etched channels have also been used. While samples have been pumped through the geometry made of an inert material, more complex systems have been employed; (Schneider *et al.*, 2007) designed a device which relied on electrical stimulation of channels made of a piezoelectric material, where a pressure gradient in a liquid channel is formed and acts like a pump.

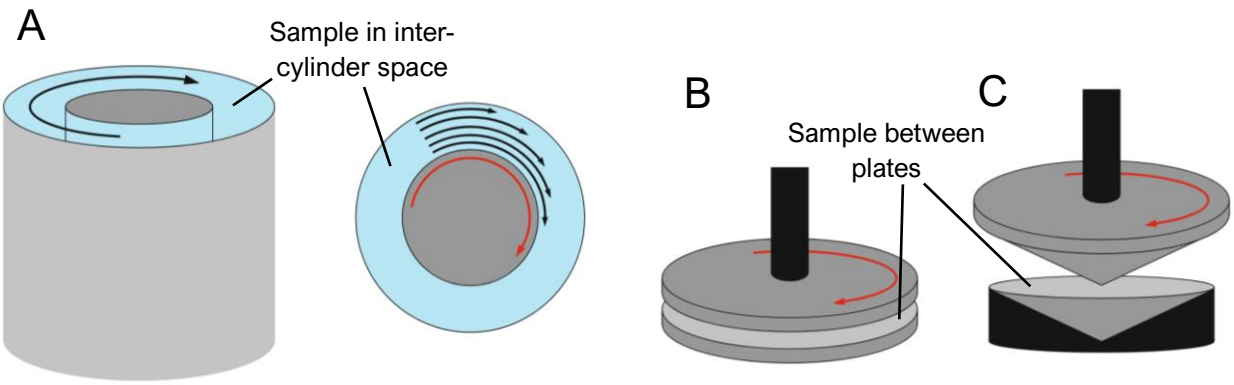


Figure 1.12: Devices to Expose Proteins to Shear

Schematics illustrate devices commonly employed to impart shear strain onto proteins. A) A Couette chamber, in which a velocity gradient is induced by rotation of a central cylinder. B) A parallel-plate rheometer, and C) a cone-and-plate device, which both expose proteins to shear by creation of a velocity gradient from rotation of a plate.

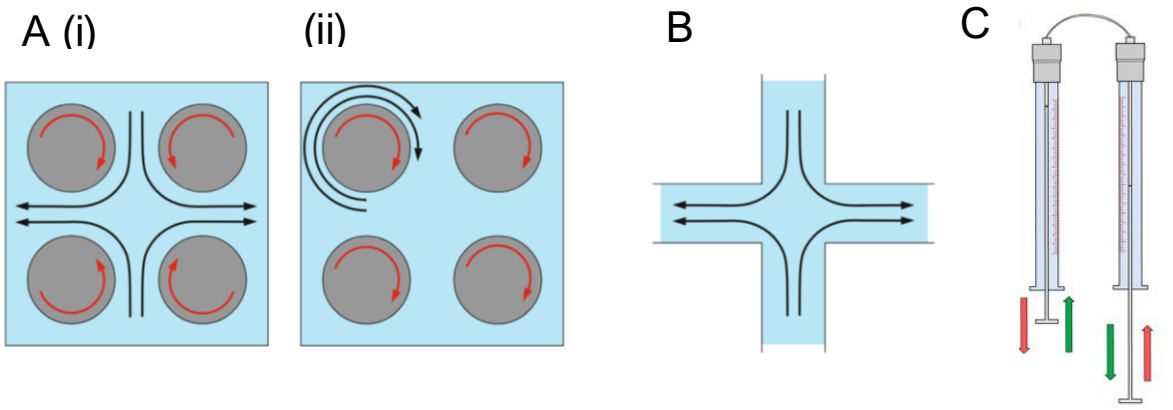


Figure 1.13: Devices to Expose Proteins to Extensional Flow and Shear

Schematics illustrate devices commonly employed to impart extensional flow and shear only proteins. Sample directionality indicated by black arrows. A) a four-roll mill apparatus can induce (i) extensional flow by converging flow fields (black arrows) or shear (ii) with circulating flow fields. B) Cross-slot channels can general extensional flow by converging flow fields. C) An example of a syringe and tubing-based device to expose proteins to extensional flow at constrictions, and shear along the tubing length. Taken from (Grigolato and Arosio, 2020).

1.6 Model Proteins for Flow-Induced Unfolding Studies

As discussed in sections 1.4.3 and 1.4.4, hydrodynamic flow is thought to impact proteins differently, depending on structural features and mechanical stability. Therefore, to fundamentally understand molecular detail of protein unfolding under flow in a way that can be applied to the widest possible range of molecules, a model protein that exhibits features of these different molecules would be an ideal mimetic.

1.6.1 RNase H

One of the most well-characterised proteins used in unfolding studies is *E. coli* RNase H. The 17.5 kDa monomer exhibits mixed α-helical/β-sheet secondary structure, and binds specifically to the RNA strand in DNA/RNA hybrids, before hydrolysing it (Katayanagi *et al.*, 1990)(Nakamura *et al.*, 1991).

Notably, a pseudo-WT variant of RNase H (RNase H*) has been developed, in which the 3 native cysteine residues, C13, C63 and C133 were replaced with alanine (Kanaya *et al.*, 1990). Interestingly, none of these cysteine residues are required for activity of the endonuclease, and the secondary/tertiary structure of the molecule remains the same. Additionally, the thermodynamic stability is not substantially different compared to the value reported in true WT at pH 6 (40.9 ± 1.1 kJ/mol for pseudo-WT, reported in fig. 4.5, and 38.3 ± 1.0 kJ/mol for true WT, reported by (Kimura *et al.,* 1992)).

This variant offers simplicity in folding and unfolding studies, removing potential effects from intermolecular disulphide bond formation. Folding and unfolding studies involving RNase H have, consequently, predominantly used RNase H*.

1.6.2 RNase H* Folding

RNase H* folding has been investigated with increasing sensitivity over the years. Initially, an intermediate state was detected by hydrogen exchange-NMR experiments, and was termed the I_{core} (Chamberlain, Handel and Marqusee, 1996). This corresponds roughly to helices A, B and D, as well as β-strand 4 (indicated in fig 1.14A). More recently, pulsed-labelling HX-MS experiments with a dead-time of 9 ms, have been used to detail the

folding sub-structures within the I_{core}, as well as including helix C in this grouping (fig 1.14B)(Hu *et al*., 2013). Using rapid continuous-flow mixing and intrinsic Trp fluorescence to detect conformational change, folding events as early as 60 μs after initiation were captured, and approximate rates of folding between steps were calculated (Rosen *et al*., 2015). Thermodynamic stabilities of the earliest detected intermediates were extracted from this data; in 20 mM sodium acetate (pH 5.5) and 50 mM potassium chloride, native state ΔG_{unf} = 39.7 ± 0.7 kJ/mol, I_{core} ΔG_{unf} = 18.0 ± 5.0 kJ/mol, and the intermediate containing only α-helix A and β-strand 4 ΔG_{unf} = 12.6 ± 2.9 kJ/mol. This illustrates the disproportionate contribution of different foldons to thermodynamic stability, as this foldon is geometrically relatively small (fig 1.14).

This categorisation of RNase H* structure into 'foldons' (defined regions containing secondary structure that are thought to comprise a step of a macromolecule's folding pathway) may offer a parameter that is indicative of structural stability; foldons are thought to stabilise subsequent folding, increasing the energy barrier to the unfolded state once formed (Maity *et al.*, 2005). Therefore, the folding order associated with RNase H* may indicate how susceptible different regions of the protein are to flow-induced unfolding.

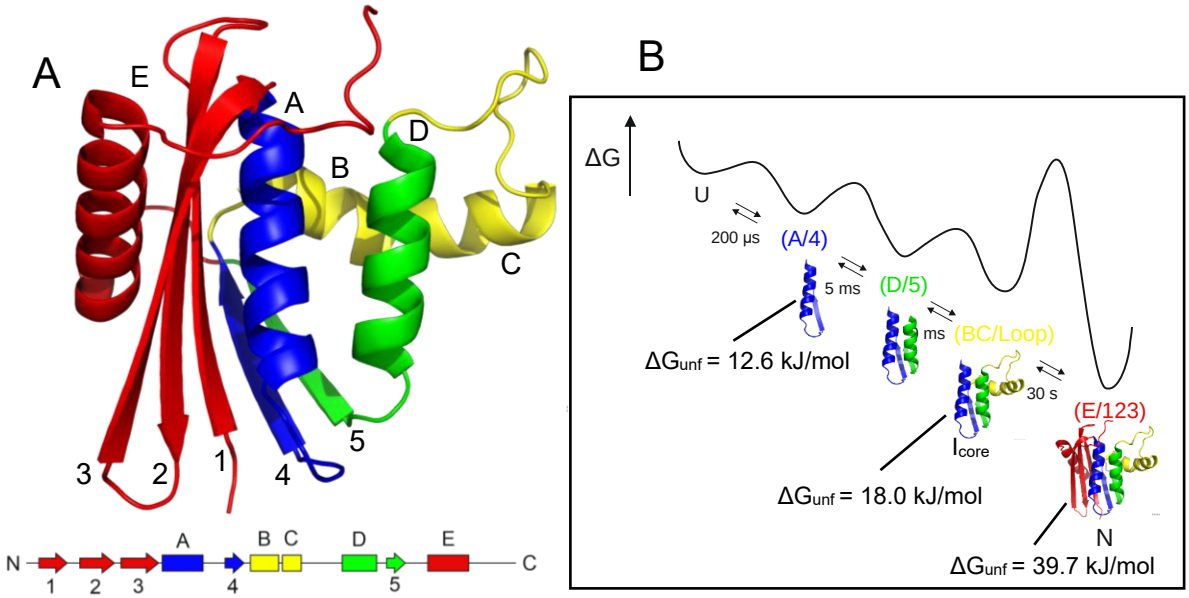


Figure 1.14: RNase H* Structure and Folding Pathway

Cartoon structure with topology map shown from N-terminus to C-terminus below. α-helices indicated by lettering, β-strands indicated by numbering. Foldons established by Hu *et al*., (2013) indicated in colour, with folding order indicated in (B), along with estimated timescales and thermodynamic stabilities associated with each foldon. PDB ID: 1F21. Adapted from (Hu *et al*., 2013)(Rosen *et al*., 2015).

1.6.3 RNase H* Unfolding

If mechanical unfolding was to proceed in the opposite direction to folding, it may be that positions requiring more complete unfolding would be less labile, due to protective effects from surrounding foldons. Literature on protein unfolding in general, and that specifically regarding RNase H*, suggests this may be the case; some proteins have been shown to unfold through intermediate states (Schlierf, Li and Fernandez, 2004)

Although RNase H* unfolding has been measured using chemical denaturation in numerous studies (Jensen *et al.*, 2020), mechanical unfolding has only been reported by (Cecconi *et al*., 2005), using optical tweezers. When tethered by residues 4 and 155, native RNase H* unfolds completely at ~19 pN. If relaxed and re-pulled without sufficient time to fully refold, RNase H* unfolds at ~5.5 pN, which was found to be a region corresponding to the I_{core} identified by Chamberlain *et al*., (1996). This indicates that following mechanical stress, the I_{core} forms within a shorter timeframe, and is substantially less mechanically stable, than the periphery. This supports the idea that the periphery, indicated by the red foldon in figure 1.14, protects the I_{core} from unfolding in the native structure.

1.7 Aims

The work presented in this thesis aimed to investigate the effects of hydrodynamic flow on protein structure using a model protein system, and explore how they might modulate aggregation propensity. The implications of these results on the 'developability' of protein biopharmaceuticals were considered throughout, to give insight into the structural features that determine a molecule's stability to flow fields encountered in manufacture. In more detail, the aims of this thesis are:

- To establish a method to probe protein conformational change under hydrodynamic flow, and to design and characterise variants of a model protein (RNase H*) that behave differently under hydrodynamic flow fields.

- To use this method to assess whether hydrodynamic flow fields, including relatively high levels of extensional flow, can induce conformational change in RNase H*, and establish if thermodynamic stability contributes to unfolding propensity under flow.

- To identify specific regions that are particularly susceptible to flow-induced unfolding, to gain an understanding of the structural determinants for flow-induced unfolding.

- To investigate if flow-induced conformational change leads to protein aggregation, by analysing aggregation propensity of the RNase H* dataset, and by detecting unfolded material in protein aggregate. This would be valuable in linking two phenomena that have been considered in the literature but not conclusively associated.

- Lastly, to compare flow-induced aggregation in a panel of 27 biopharmaceuticals with performance in other assays commonly used in biopharmaceutical development, to establish how flow relates to these measures. Additionally, establishing whether a combination of assays, including flow-induced aggregation, can be predictive of a molecule's long-term performance without the need for long-term experiments.

1.8 Results Chapters Overview

For clarity, an overview of the structure and contents of the results chapters is given here. Results chapters 3-5 focus on gaining a fundamental understanding of protein unfolding and aggregation under hydrodynamic flow, and chapter 6 focuses on how experiments involving hydrodynamic flow may be applied to biopharmaceutical development.

As discussed in section 1.6, a model protein could be an effective system to investigate how proteins in general unfold and aggregate under flow. RNase H* was selected as this model, and it was decided that labelling of single cysteine residues in RNase H* was to be used to investigate unfolding under flow, as discussed in chapter 4.1.

Chapter 3 describes the device used to expose proteins to defined hydrodynamic flow fields, as well as the design and production of RNase H* variants, each only containing a single cysteine residue, for site-specific labelling. Full structural characterisation of these variants is discussed, with attention given to the varying thermodynamic stabilities exhibited by them. In addition, the aggregation-proneness of these cysteine variants is quantified.

Investigation of label-free assays for measuring flow-induced unfolding is discussed in chapter 4, however it was determined that a labelling-based assay would be more appropriate. Development of the labelling assay is described in detail, with several of the RNase H* single-cysteine variants used to investigate assay sensitivity and reproducibility.

Chapter 5 then describes characterisation of flow-induced unfolding in the panel of RNase H* cysteine variants using the labelling assay and exploits the site-specificity of the assay to draw structural conclusions on how flow-induced unfolding proceeds. The varied biophysical properties of the cysteine variants are compared to the diverse response to flow-induced unfolding, to better understand the features that determine the latter. Lastly, to more closely attribute flow-induced aggregation to unfolding, labelling of aggregated material is analysed in detail.

Finally, chapter 6 considers flow-induced aggregation of biopharmaceuticals, examining an initial panel of 27 distinct monoclonal antibodies. The varied aggregation propensities indicate potential use for flow-induced aggregation as a developability metric; an indicator of how likely a mAb is to reach successful clinical use. Additionally, the concept of using a single parameter to determine biopharmaceutical developability is discussed,

and whether flow-induced aggregation has a place as part of this amongst a suite of other developability assays.

2 Thesis Methods

2.1 Materials

2.1.1 Equipment

2.1.1.1 Molecular Biology

Q5® Site-Directed Mutagenesis Kit (New England BioLabs®)

T100™ Thermal Cycler (Bio-Rad Laboratories Inc., CA, USA)

QIAprep® Spin Miniprep Kit (QIAGEN Group, Venlo, Netherlands)

NanoDrop 2000™ spectrophotometer (Thermo Fisher Scientific Inc., MA, USA)

2.1.1.2 Protein Expression and Purification

New Brunswick™ Innova 44 (Eppendorf, Hamburg, Germany)

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

Vibra-cell™ Ultrasonicator and 6 mm tip (Sonics & Materials, Inc., Leicestershire, UK)

ÄKTAprime plus (Cytiva, MA, USA)

ÄKTA start (Cytiva, MA, USA)

ÄKTA Go™ (Cytiva, MA, USA)

XK 16/40 column packed in-house with 50 ml Heparin Sepharose 6 Fast Flow resin (Cytiva, MA, USA)

5 ml bed volume pre-packed HiTrap™ heparin HP column (Cytiva, MA, USA)

6 ml RESOURCEQ anion exchange column (Cytiva, MA, USA)

Superdex™ 75 HiLoad 26/600 (320ml) gel filtration column (Cytiva, MA, USA)

2.1.1.3 Protein Purity Quantification

SDS-PAGE gel components; glass plates, silicone dividers, magnetic clamps, electrophoresis tank and cables (ATTO Ltd., Tokyo, Japan)

Powerpac Basic gel electrophoresis power source (Bio-Rad Laboratories, CA, USA)

Q9 Alliance Transilluminator (UVITEC, Cambridge, UK)

120 µl quartz glass high-performance cuvettes, 105.201-QS (Hellma® Analytics UK, Southend-on-Sea, UK)

UV-1800 UV/Visible scanning spectrophotometer (Shimadzu UK Ltd., Milton Keynes, UK)

2.1.1.4 Protein Characterisation

Xevo G2-XS Q-ToF Mass Spectrometer (Waters Corporation, Wilmslow, UK)

350 µl quartz glass cuvettes (110-1-40-QS)(Hellma® Analytics UK, Essex, UK)

Chirascan™ CD Spectrophotometer (Applied Photophysics Ltd., Surrey, UK)

1400 µl quartz glass high-performance cuvettes (114-QS)(Hellma® Analytics UK, Essex, UK)

PTI QuantaMaster™ spectrofluorometer (HORIBA Scientific, Kyoto, Japan)

Ellipsometer (J.A. Woolam, NE, USA)

Optima™ MAX-XP ultracentrifuge (Beckman Coulter Life Sciences, CA, USA)

TLA-100 fixed-angle rotor (Beckman Coulter Life Sciences, CA, USA)

200 µl polycarbonate open ultracentrifuge tubes (Beckman Coulter Life Sciences, CA, USA)

Fida 1 instrument (Fida Biosystems Aps, Søborg, Denmark)

Rotor-Gene™ 6000 rotary qPCR machine (Corbett Research UK, Cambridge, UK)

2.1.1.5 Hydrodynamic Flow Experiments

Breadboard base (Thor Labs Inc., NJ, USA)

Syringe Clamps (Thor Labs Inc., NJ, USA)

Ferrule compression fittings (Hamilton Company®, NV, USA)

Gilson P10 O-ring (Gilson Inc., WI, USA)

Borosilicate glass capillaries (Sutter Instrument®, CA, USA)

Arduino Microcontroller (Arduino®, Ivrea, Italy)

Stepper motor (Haydon Switch & Instrument Inc., CT, USA)

Gas-tight 1 mL syringes 1001 RN model (Hamilton Company®, NV, USA)

2.1.1.6 Analytical Chromatography

Nexera LC-40 Series HPLC system (Shimadzu UK Ltd., Milton Keynes, UK)

SPD-M20A photodiode array (Shimadzu UK Ltd., Milton Keynes, UK)

RF-20A detector (Shimadzu UK Ltd., Milton Keynes, UK)

Nucleosil 300 C4 column (Chromex Scientific Ltd., Dronfield, UK)

1260 Infinity II LC System (Agilent, CA, USA)

TSKgel G3000SWxl size exclusion column (Tosoh Bioscience LLC, Tokyo, Japan)

Proteomix Butyl HIC column (Sepax Technologies, Inc., DE, USA)

Zenix SEC-300 column (Sepax Technologies, Inc., DE, USA)

2.1.1.7 Other Equipment

Jenway 3020 Bench pH Meter (Bibby Scientific, Staffordshire, UK)

0.5 ml 7 kDa MWCO Zeba spin desalting columns (Thermo Fisher Scientific, MA, USA)

0.5 ml Ultrafree® 0.22 µm pore diameter centrifugal filter (Millipore®, MA, USA)

2.1.2 Software

PyMOL2 (version 2.5.5)(Schrödinger Inc., NY, USA)

ExPASy ProtParam tool (Swiss Institute of Bioinformatics, Lausanne, Switzerland)

NEBaseChanger™ (version 1.3.3)(New England Biolabs® Inc., MA, USA)

Igor Pro 7.03 (WaveMetrics, Inc., OR, USA)

CompleteEASE Ellipsometry Software (version 6.7)(J.A. Woolam, NE, USA)

KinTek Explorer Professional (version 6.3)(Kintek Corporation, PA, USA)

SpectraGryph (version 1.2)(Dr Friedrich Menges, Oberstdorf, Germany)

Origin 2021 graphing software (OriginLab® Software)

XLSTAT 2023 statistics software (Lumivero, CO, USA)

AlphaFold2 (version 2.3.2)(Google DeepMind, CA, USA)

Therapeutic Antibody Profiler (Oxford Protein Informatics Group, Oxford, UK)

CamSol (version 6.2)(Vendruscolo Lab, University of Cambridge, Cambridge, UK)

ImageJ FIJI (version 2.9.0)(Schindelin *et al.*, 2012)

2.1.3 Chemicals

All chemicals used were reagent grade unless otherwise stated.

Chemical	Supplier
A	
30 % (v/v) Acrylamide/Bis Solution (37.5:1 ratio)	Bio-Rad Laboratories Inc., CA, USA
Agar	Fisher Scientific, Loughborough, UK
Gradient-grade Acetonitrile (CH_3CN), CHROMASOLV™	Fisher Scientific, Loughborough, UK
Agarose	Melford Laboratories, Suffolk, UK
Ammonium Bicarbonate $(NH_4)HCO_3$	Sigma Life Sciences, MO, USA

Ammonium Chloride NH_4Cl	Sigma Life Sciences, MO, USA
APS (Ammonium persulphate)	Sigma Life Sciences, MO, USA
L-Arginine	Sigma Life Sciences, MO, USA
B	
Bromophenol blue	Sigma Life Sciences, MO, USA
C	
Chloramphenicol	Sigma Life Sciences, MO, USA
Calcium chloride ($CaCl_2$)	Sigma Life Sciences, MO, USA
D	
DTT (Dithiothreitol)	Formedium, Norfolk, UK
E	
Ethanol	Fisher Scientific UK, Loughborough, UK
EDTA (Ethylenediaminetetraacetic acid)	Fisher Scientific, Loughborough, UK
G	
Glycerol	Fisher Scientific, Loughborough, UK
Guanidine HCl	Sigma Life Sciences, MO, USA
H	
Hellmanex™ III alkaline cleaning liquid	Hellma® Analytics UK, Essex, UK
HPLC-grade H_2O	Merck, MA, USA
37 % (w/w) Hydrochloric acid (HCl)	Fisher Scientific, Loughborough, UK
I	
IAEDANS (5-((2-((iodoacetyl)amino)ethyl)amino)naphthalene-1-sulfonic acid)	Fisher Scientific, Loughborough, UK
IPTG (Isopropyl β-D-1-thiogalactopyranoside)	Fisher Scientific, Loughborough, UK
M	
Magnesium sulphate ($MgSO_4$)	Sigma Life Sciences, MO, USA
P	
Phosphoric acid (H_3PO_4)	Fisher Scientific, Loughborough, UK
Potassium chloride (KCl)	Fisher Scientific, Loughborough, UK
Potassium phosphate dibasic (K_2HPO_4)	Fisher Scientific, Loughborough, UK
Potassium phosphate monobasic (KH_2PO_4)	Fisher Scientific, Loughborough, UK
S	
Sodium acetate (CH_3CO_2Na)	Fisher Scientific, Loughborough, UK
Sodium chloride (NaCl)	Fisher Scientific, Loughborough, UK
Sodium sulphate (Na_2SO_4)	Fisher Scientific, Loughborough, UK
10 % SDS (Sodium dodecyl sulphate)	Severn Biotech, Kidderminster, UK
Sodium hydroxide (NaOH)	Fisher Scientific, Loughborough, UK
Sodium phosphate dibasic (Na_2HPO_4)	Thermo Scientific, Surrey, UK
Sodium phosphate monobasic (NaH_2PO_4)	Sigma Life Sciences, MO, USA
Sodium succinate dibasic hexahydrate	Sigma Life Sciences, MO, USA
SYPRO™ Orange (5000x)	Fisher Scientific, Loughborough, UK
T	
TEMED (Tetramethylethylenediamine)	Sigma Life Sciences, MO, USA
TFA (Trifluoroacetic acid)	Fisher Scientific, Loughborough, UK
Tris	Fisher Scientific, Loughborough, UK
U	
Urea	Fisher Scientific, Loughborough, UK

2.1.4 Buffers

Before use, buffers were filtered through a 0.22 µm filter. All buffers for chromatography use were degassed prior to use. pH of all buffers was adjusted with HCl or NaOH if necessary, except for AstraZeneca HP-SEC running buffer, which was adjusted with H_3PO_4. All buffers were prepared in-house unless a source is stated. All growth media were sterilised before use, at 121 °C and 15 psi, for 20 minutes.

- **Ammonium bicarbonate buffer** – 50 mM ammonium bicarbonate (pH 6)
- **Arginine succinate buffer** – 125 mM L-arginine, 20 mM sodium succinate, (pH 8)
- **AstraZeneca HP-SEC running buffer** – 0.1 M Na_2HPO_4, 0.1 M Na_2SO_4 (pH 6.8)
- **Gdn/DTT buffer** – 6 M Guanidine HCl, 50 mM Tris-HCl, 5 mM DTT (pH 6)
- **Heparin column buffer A** – 50 mM Tris-HCl, 20 mM NaCl, 10 mM $MgCl_2$, 0.5 mM EDTA (pH 8)
- **Heparin column buffer B** – 50 mM Tris-HCl, 1 M NaCl, 10 mM $MgCl_2$, 0.5 mM EDTA (pH 8)
- **HIC running buffer A** – 100 mM sodium phosphate mix (66.8 mM NaH_2PO_4, 33.2 mM Na_2HPO_4), 1.8 M ammonium sulphate (pH 6.5)
- **HIC running buffer B** – 100 mM sodium phosphate mix (66.8 mM NaH_2PO_4, 33.2 mM Na_2HPO_4)(pH 6.5)
- **Ion exchange chromatography high-salt buffer** – 20 mM NaH_2PO_4, 1 mM EDTA, 1 M NaCl
- **Ion exchange chromatography low-salt buffer** – 20 mM NaH_2PO_4, 1 mM EDTA, 50 mM NaCl
- **Phosphate Buffer** – 20 mM K_2HPO_4, 50 mM KCl (pH 8)
- **2x RNase H* Activity Assay Reaction Buffer** – 100 mM Tris-HCl, 11.6 mM $MgCl_2$, 2 mM DTT, 160 mM KCl, *either* 50 or 500 nM DNA/RNA molecular beacon (pH 8.1)
- **SDS-PAGE anode buffer** – 200mM tris.HCl (pH 8.9)
- **SDS-PAGE cathode buffer** – 100 mM tris, 100 mM tricine, 0.1% (w/v) SDS (pH 8.25)(Severn Biotech, Worcestershire, UK)
- **SDS-PAGE loading buffer** – 50mM tris.HCl, 100mM DTT, 2% (w/v) SDS, 0.1% (w/v) bromophenol blue, 10% (w/v) glycerol (pH 6.8)
- **SMAC running buffer** – 150 mM sodium phosphate mix (52.8 mM NaH_2PO_4, 97.2 mM Na_2HPO_4), pH 7
- **Urea unfolding buffer** – 6.5 M urea (pH 8)

2.1.5 Other Materials

2.1.5.1 Pre-prepared Reagent Mixes and Miscellaneous

Precision Plus Protein™ Dual Xtra prestained protein standard protein ladder (Bio-Rad Laboratories Inc., CA, USA)

Lyophilised thermolysin from *Geobacillus stearothermophilus* (Sigma Life Sciences, MO, USA)

Pearce™ 660 nm protein reagent (Thermo Fisher Scientific Inc.)

Miller dry LB mix (Sigma Life Sciences, MO, USA)

Coomassie Brilliant Blue R-250 Staining Solution (Biorad-Rad Laboratories Inc., CA, USA)

2.2 Methods

2.2.1 Molecular Biology

2.2.1.1 Design of RNase H* Cysteine Variants

Single cysteine substitutions were introduced into otherwise cysteine-free RNase H (section 2.2.1.2), to provide sites for covalent labelling with thiol-reactive IAEDANS. To act as a reporter of unfolding, sites chosen were in positions that are sequestered from solution when the protein is natively folded.

To do this, relative solvent accessibility (RSA) of each site was calculated using the solvent accessibility values of each residue predicted by the Dictionary of Protein Secondary Structure (DSSP)(Kabsch and Sander, 1983), divided by the maximum theoretical accessibility of that residue type, to account for differences in size. Solvent accessible area is given (in $Å^2$) as number of water molecules in the first hydration shell in contact with the residue, multiplied by 9.65 (a conversion factor that relates volume in the first hydration shell to number of water molecules)(Kabsch & Sander, 1983). RSA was calculated as follows for each amino acid:

$$RSA(\%) = \left(\frac{ACC}{maxACC}\right) \times 100$$

Equation 2.1

Where ACC is solvent accessible surface area of an amino acid, (in $Å^2$), maxACC is the maximum possible ACC for the amino acid, assuming it is in a tripeptide flanked by two glycine residues (Rost and Sander, 1994).

Residues with RSA values predicted to be below 10% were selected for mutagenesis. NEBaseChanger™ online primer generation tool was used to design optimal primers and associated optimal annealing temperatures. Primers were ordered from either Eurofins Genomics or Source BioScience Ltd.

2.2.1.2 Site-Directed Mutagenesis

A Q5® Site-Directed Mutagenesis Kit (New England BioLabs®) was used to introduce mutations that encode the single residue substitutions into WT RNase H* template sequence. The plasmid containing this template gene was a pAED4 vector with an ampicillin resistance gene and the RNase H* sequence under the control of a T7 promoter. This was kindly donated by the Marqusee lab (University of California, Berkley). The following reagents were combined in a thin-walled PCR tube and diluted with nuclease-free water:

Reagent	Final Content
Q5 Hot Start High-Fidelity Master Mix (2x)	1x
Forward Primer	0.5 µM
Reverse Primer	0.5 µM
Template DNA	25 ng

Table 2.1: Site-Directed Mutagenesis Reagents

The denaturation and annealing cycles in figure 2.1 were performed in thin-walled PCR tubes, using a T100™ Thermal Cycler (Bio-Rad Laboratories Inc.).

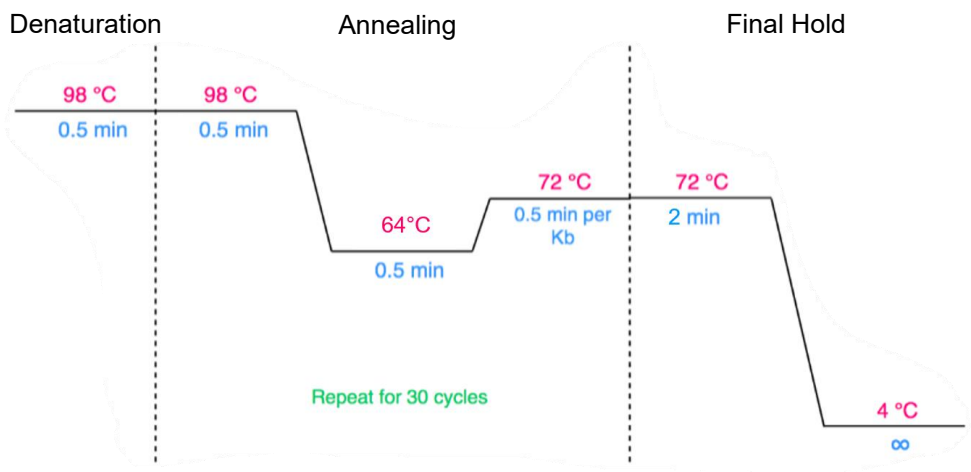


Figure 2.1: PCR Thermal Cycling Protocol Schematic

Times indicate hold times at each respective temperature, per step.

The resulting PCR product was treated with a Kinase, Ligase and DpnI (KLD) mix (New England BioLabs®) at a final concentration of 1x KLD Enzyme Mix and 1x KLD Reaction Buffer. This was left at 20°C for 5 minutes.

2.2.1.3 Plasmid DNA Purification

Following KLD treatment, plasmids were immediately transformed into DH5α competent *E. coli* cells, using the same heat-shock protocol outlined in section 2.2.2.1. Cells were plated onto LB agar petri dishes containing 100 µg/ml carbenicillin, and incubated overnight at 37 °C. All LB medium was made up to 2.5% (w/v) Miller LB mix/de-ionised H_2O and autoclaved to sterilise before use (section 2.1.4). Single colonies were isolated from the petri dishes and each of these were added to a separate conical flask containing 50 ml LB and incubated overnight at 37 °C, 200 rpm shaking. Plasmid DNA was extracted from 1.5 ml of the resulting cell culture, using a QIAprep® Spin Miniprep Kit (QIAGEN®) according to the protocol provided. Resulting pure DNA was quantified using a NanoDrop 2000™ spectrophotometer (Thermo Fisher Scientific Inc.) and stored at -20°C until further use. The collected plasmid DNA was delivered to Eurofins Genomics for Sanger sequencing to verify successful mutagenesis.

2.2.2 Protein Expression and Purification

2.2.2.1 Plasmid Transformation for Protein Expression in *E. coli*

To transform bacteria for protein expression, a heat-shock protocol was used. 50µl BL21(DE3)pLysS cells were defrosted on ice over 10 minutes, and 50-100 ng DNA was added. Cells were left for a further 10 minutes on ice before heat shock in a water bath at 42 °C, for 1 minute. Cells were returned to ice for 2 minutes and then 500 µl LB was added. Cells were incubated at 37°C, 200 rpm shaking for 1 hour, before centrifugation at 3500 rpm for 3.5 minutes. 450 µl supernatant was removed and the remaining 100 µl was suspended by pipette agitation before plating onto LB agar petri dishes containing chloramphenicol and carbenicillin.

Herein, all LB medium contains 100 µg/ml carbenicillin disodium, and 25 µg/ml chloramphenicol, added after sterilisation, when medium had cooled below approximately 50 °C. Single colonies were isolated from the petri dishes following overnight incubation at 37 °C. Each of these were added to a separate conical flask containing 50 ml LB and incubated overnight at 37 °C, 200 rpm shaking. 4 ml of the resulting culture was used to inoculate 1 L of LB containing chloramphenicol/carbenicillin, in 2.5 L conical flasks. 10x 1 L cell cultures were incubated at 37°C, 200 rpm shaking, until OD_{600} = 0.6 – 0.8. At which point, 1 M IPTG was added to each culture to reach a final concentration of 0.5 mM and induce expression via the inducible *lac* promoter. Cultures were then incubated overnight at 20 °C, 200 rpm shaking, and harvested the next morning. To isolate bacteria from the cultures, each was centrifuged at 4000 rpm for 30 minutes, at 10 °C. Supernatant was discarded and the pellet was collected and stored at -20 °C.

2.2.2.2 Cell Lysis by Sonication

Harvested pellets were defrosted at room temperature, and then homogenised in heparin buffer A (section 2.1.4)(50 ml buffer A per 10 g cell pellet). Cells were kept on ice for all succeeding steps. A Vibra-cell™ Ultrasonicator with 6 mm tip attached (Sonics & Materials, Inc.) was used at 75% amplitude, with 3 seconds sonication followed by 7

seconds rest, for a total of 1.5 min sonication time. 30 ml samples were sonicated at a time to lyse cells.

The lysate was centrifuged at 25,000 rpm for 30 minutes and 10°C, to separate insoluble lysate products from soluble. The lysate was analysed using an SDS-PAGE gel (section 2.2.3.1), and the supernatant containing soluble protein was purified.

2.2.2.3 Heparin Affinity Chromatography

The components associated with each purification step were analysed using SDS-PAGE (section 2.2.3.1). Lysate supernatant was loaded onto a XK 16/40 chromatography column (Cytiva), packed in-house with 50 ml heparin Sepharose Fast Flow 6 resin (Cytiva) and pre-equilibrated with ≥2 column volume (CV) equivalents heparin buffer A (section 2.1.4). The column was washed with buffer A until absorbance at 280 nm plateaued at a minimal value. At this point, any remaining protein was eluted by applying a gradient of heparin buffer B (section 2.1.4) up to 50% v/v, with fractions collected throughout the elution.

Consistently, it was noted that RNase H* variants would bind poorly to this column, meaning they would be eluted when the column was washed. This wash was further purified using 2x 5 ml pre-packed HiTrap™ heparin HP (high performance) columns (Cytiva), connected directly to one another to provide 10 ml of purifying resin volume. RNase H* was loaded and eluted as previously, however, with this resin eluted as the first major peak.

2.2.2.4 Ion Exchange Chromatography

To ensure nucleic acid is not present in purified RNase H*, ion-exchange chromatography was used. Fractions from heparin affinity chromatography containing RNase H* were applied to a 6 ml RESOURCEQ anion exchange column (Cytiva), pre-equilibrated with ≥2 column volume (CV) equivalents low-salt buffer (section 2.1.4). The column was washed with low-salt buffer until absorbance at 280 nm reached a plateau at a minimal value. Any remaining protein was eluted using a stepped gradient of increasing high-salt buffer (section 2.1.4), in 25 % (v/v) intervals from 0 to 100% high-salt buffer. Samples were collected as fractions throughout the gradient.

The positively-charged RNase H* did not bind to the column, and as such as eluted in the flow-through and wash steps, whilst nucleic acid bound to the column and was removed in the subsequent gradient elution. Nucleic acid content of the protein eluent was measured using a Shimadzu 1800 UV/Visible scanning spectrophotometer, as described in section 2.2.3.3 but with absorbance at 260 nm measured in addition to 280 nm.

2.2.2.5 Size-Exclusion Chromatography (SEC)

Nucleic acid-free RNase H* (section 2.2.2.4) was then injected on a Superdex 75 gel filtration column, pre-equilibrated with ≥2 CV equivalents ammonium bicarbonate buffer (section 2.1.4). 5 ml RNase H* sample was injected per run, with peaks reaching ≥ 20 mAU collected as fractions. To reduce the number of injections required, samples were often concentrated before injection using Amicon® Ultra-15 centrifugal filter units (3 kDa MWCO)(Merck KGaA), at 4000 xg, 10℃.

2.2.2.6 Lyophilisation

Purified RNase H* samples in ammonium bicarbonate buffer were snap-frozen in liquid nitrogen before lyophilisation in a Heto Powerdry PL3000 (Thermo Fisher Scientific Inc.). Protein was lyophilised for up to 1 week and stored at -20 ℃ until further use. Lyophilised protein dissolved in liquid was always filtered through a 0.22 µm syringe filter before use, to remove insoluble components.

2.2.3 Protein Purity Quantification

2.2.3.1 SDS-PAGE

For analysis of purification products, sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) gels were prepared to a formula intended for analysis of higher-MW species (table 2.2), to allow relative content of species with a range of MW values to be visualised. For gels containing fluorescent samples to be analysed

quantitatively, gels were prepared to a formulation intended for analysis of lower-MW species (table 2.3), as only analysis of RNase H* (17.5 kDa) was needed.

Per gel, an 8 cm x 10 cm glass gel casing was clamped together, with a 1.5 mm spacer and silicone divider between. The resolving components were combined, with APS and TEMED added last, and the mixture poured into the casing (ensuring space is left for a 12-lane comb to be inserted without touching the mixture). If using a gel mix for lower-MW species, a thin (~2 mm) layer of 100% ethanol was deposited on the surface, to level the resolving gel evenly. The resolving gel was left to set for approximately 1 hour, before the ethanol was removed. The stacking gel mix was then prepared, with APS and TEMED added last, and the comb added before leaving the gel to set for approximately 1 hour. If using a gel mix for analysis of higher-MW species, the step involving ethanol could be omitted and the stacking gel added immediately on top of the resolving gel. If not used immediately, gels were stored at 4°C for up to a week, wrapped in water-dampened tissue.

Component	Resolving Gel	Stacking Gel
Acrylamide (30% w/v)/bis-acrylamide (0.8% w/v)	3.75 ml	0.415 ml
De-ionised H_2O	1.7 ml	1.7 ml
Tris-HCl 1.5M, pH 8.8	1.9 ml	-
Tris-HCl 1M, pH 6.8	-	0.315 ml
SDS (10% w/v)	75 µl	25 µl
Ammonium Persulphate (10% w/v)	75 µl	25 µl
TEMED	5 µl	5 µl

Table 2.2: SDS-PAGE Gel Recipe for Analysis of Higher-MW Species

Volumes are given per gel.

Component	Resolving Gel	Stacking Gel
Acrylamide (30% w/v)/bis-acrylamide (0.8% w/v)	3.75 ml	0.415 ml
De-ionised H_2O	0.22 ml	1.86 ml
Tris-HCl 3M, 0.3% (w/v) SDS, pH 8.45	2.5 ml	0.775 ml
Glycerol	1 ml	-
Ammonium Persulphate (10% w/v)	25 µl	50 µl
TEMED	5 µl	5 µl

Table 2.3: SDS-PAGE Gel Recipe for Analysis of Lower-MW Species

Volumes are given per gel.

Samples were diluted 2-fold with 2x loading buffer (50mM tris.HCl pH 6.8, 100mM DTT, 2% (w/v) SDS, 0.1% (w/v) bromophenol blue, 10% (w/v) glycerol), and held in a boiling water bath for 5 minutes. If denaturation was not required, samples were diluted the same way into loading buffer, without the presence of DTT and boiling. Gels were inserted into gel running tanks, which were then filled with cathode buffer (section 2.1.4) in the cathode chamber, and anode buffer in the anode chamber (section 2.1.4). 10 µl of diluted sample was loaded into each well. Using a Bio-Rad Powerpac Basic to supply current, gels were run at 30 mA per gel, until the stacking gel was surpassed. After which, the amplitude was increased to 60 mA per gel. Gels were removed from glass casings stained for at least 1 hour in InstantBlue® Coomassie blue stain (Expedeon Ltd.), before rinsing thoroughly with distilled and deionised H_2O (ddH_2O) to remove free stain. Gels were all imaged using A Q9 Alliance Transilluminator (UVITEC). Stained protein content was imaged using the 'SDS-PAGE white light' setting, with an exposure time of 0.720 s.

2.2.3.2 Densitometry

Both bands resulting from imaging with visible light, and those resulting from fluorescent labelling of protein were quantified using the Q9 Alliance software associated with the instrument. Each lane was selected manually to include minimal overlap with neighbours. The baseline for each lane was determined as the lowest pixel intensity reached in the band, and was subtracted (fig 2.2A). This could vary across a gel, as illumination by the lamp was not always of even intensity. Each lane was displayed as a 2-D plot of pixel intensity versus position down the lane. Each band within a lane was selected; the edge

of a band being determined as the point at which pixel intensity slope increases beyond 45° from horizonal (fig 2.2B).

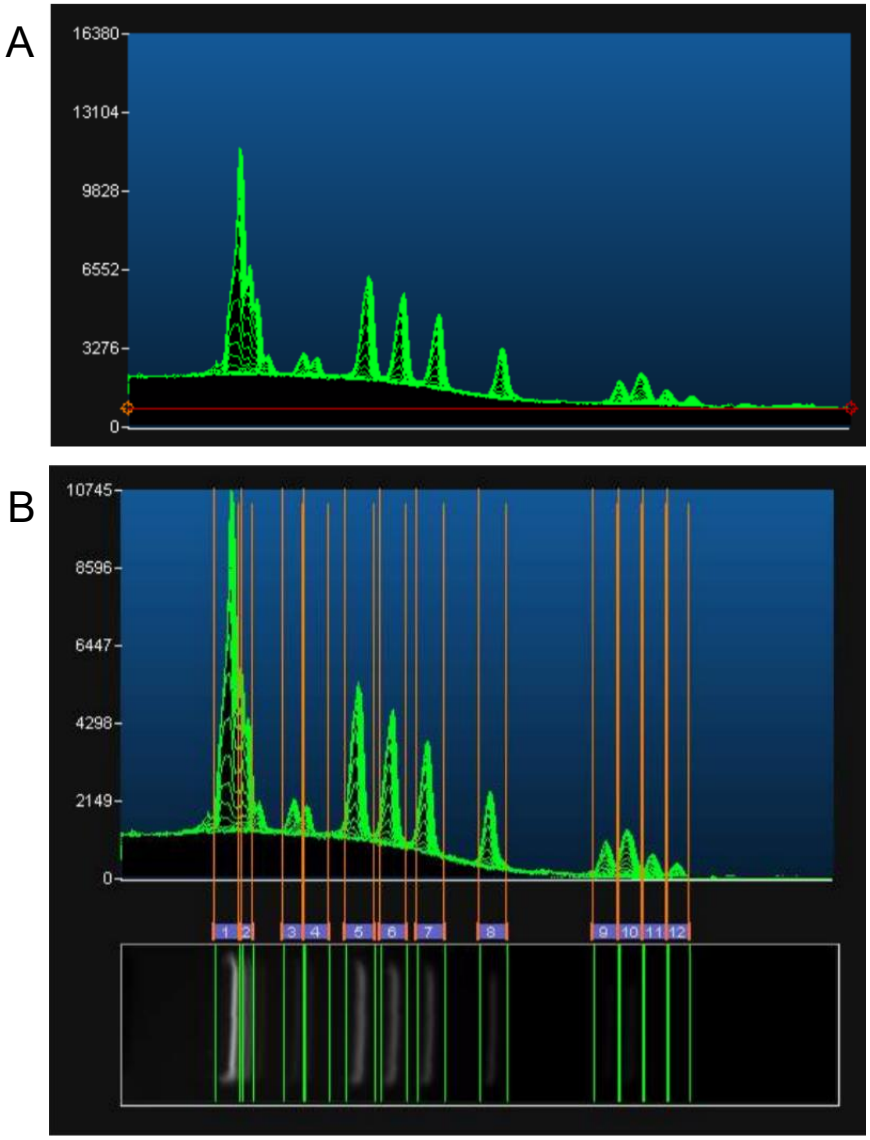


Figure 2.2: Quantification of Pixel Intensity from SDS-PAGE Gel Bands

A) Linear background pixel intensity removal by introduction of a baseline (red line). B) Pixel intensity peak selection by introduction of separations (orange lines).

2.2.3.3 Protein Quantification by Spectrophotometry

For routine protein quantification, samples were measured in 120 µl quartz glass high-performance cuvettes (105.201-QS, Hellma® Analytics) with pathlength of 1 cm, using a Shimadzu 1800 UV/Visible scanning spectrophotometer. 120 µl buffer was used as a

blank measurement, before the same volume of protein sample was measured. Absorbance at 280 nm was used to calculate protein concentration from the Beer-Lambert equation:

$$A = \varepsilon cl$$

Equation 2.2

Where *A* = absorption (AU), *l* = pathlength (cm), *c* = protein concentration (M), ε = molar extinction coefficient of the protein of interest ($M^{-1}cm^{-1}$), calculated using the online ExPASy ProtParam tool (Swiss Institute of Bioinformatics), and the primary sequence associated with each protein. For WT RNase H*, ε = 40450 $M^{-1}cm^{-1}$.

2.2.4 Protein Characterisation

2.2.4.1 Intact LC-TOF-MS

Once expressed and purified, RNase H* WT and each cysteine variant was analysed by liquid chromatography-TOF (time-of-flight) mass spectrometry (LC-MS) using a Xevo G2-XS Q-ToF (Waters Corporation) mass spectrometer to verify correct sample content, successful mutagenesis, and purity. Samples were prepared at 20 μM in ammonium bicarbonate buffer (section 2.1.4) and analysed. The University of Leeds Mass Spectrometry Facility instrumentation used in this work is supported by the BBSRC (BB/M012573/1) and analysis was carried out by Rachel George.

2.2.4.2 Circular Dichroism Spectroscopy (CD)

RNase H* was analysed by circular dichroism spectroscopy (CD) at 0.1 mg/ml in either potassium phosphate or urea buffer (section 2.1.4). 250 μl was measured per sample using a Chirascan™ CD Spectrophotometer (Applied Photophysics Ltd.). λ_{ex} = 280 nm, with λ_{em} detected from 180-260nm, at 20℃. Cuvette path length = 0.1 cm. For each wavelength point, mean residue ellipticity (MRE) was calculated:

$$Mean\ Residue\ Ellipticity = \frac{\left(\frac{\theta}{10 \times c \times d}\right)}{154}$$

Equation 2.3

Where θ = ellipticity (mdeg), c = concentration of protein (M), d = pathlength (cm).

Data points with high tension (HT) voltage values above 600 V were deemed unacceptably noisy and were disregarded.

2.2.4.3 Intrinsic Fluorescence Spectroscopy

The fluorescence emission of 0.1 mg/ml RNase H* solutions were measured using a PTI QuantaMaster™ spectrofluorometer (HORIBA Scientific), with excitation slit widths set to 1 nm and emission slit widths set to 2 nm. 1400 µl quartz glass high-performance cuvettes (114-QS, Hellma® Analytics) with a path length of 1 cm were used with 600 µl sample volume. To characterise the folding status of each RNase H* variant, λ_{em} = 280 nm and an emission wavelength scan of 320 – 400 nm was used.

2.2.4.4 Equilibrium Unfolding for Thermodynamic Stability Measurements

RNase H* in phosphate buffer (section 2.1.4) with a range of urea concentrations was prepared, by diluting a stock of 1 mg/ml RNase H* into the correct buffers, to reach a final concentration of 0.1 mg/ml RNase H* with urea from 0 – 8 M, at 0.2 M intervals. These samples were left at 25 ℃ overnight in a water bath to equilibrate.

Intrinsic fluorescence was used to quantify unfolding, as described in section 2.2.4.3 but with a single emission wavelength measured and averaged over 30 s. The cuvette chamber was held at 25 ℃ and 1 minute 30 s was allowed for thermal equilibration before taking measurements. λ_{ex} = 280 nm and λ_{em} = 320 nm were used to allow maximal intensity difference between folded and unfolded states.

Igor Pro 7.03 (WaveMetrics, Inc.) was used to fit the raw fluorescence intensity data to the following equation:

$$f_{[D]} = \frac{(a[D] + b)\, exp\left(\frac{\Delta G_{unf} - (m[D])}{RT}\right)(c[D] + d)}{1 + exp\left(\frac{\Delta G_{unf} - (m[D])}{RT}\right)}$$

Equation 2.4

Where $f_{[D]}$ = fluorescence emission intensity (320 nm) at a given urea concentration, *a* and *c* = pre- and post-unfolding transition baseline gradients, *b* and *d* = pre- and post-unfolding transition y intercept values, [*D*] = urea concentration (M), *R* = gas constant at standard conditions (8.314 J K^{-1} mol^{-1}), *T* = temperature (298.15 K), ΔG_{unf} = free energy requirement for urea-mediated unfolding (J mol^{-1}), m = rate of $\Delta\Delta G_{unf}$ with varying urea concentration (J mol^{-1} M^{-1}).

ΔG_{unf} and m values were obtained from this fitting, with error values given as standard deviation, calculated from residuals of data points to the fit.

Extracted fits to the data and original datapoints could then be normalised, to reflect the proportion of the sample in the data that is folded (P_N):

$$P_N = \frac{f_{[D]} - (c[D] + d)}{(a[D] + b) - (c[D] + d)}$$

Equation 2.5

$\Delta\Delta G_{unf}$ values were calculated, and error propagated using the min-max method.

2.2.5 Hydrodynamic Flow Experiments

2.2.5.1 Extensional Flow Device Assembly

A device designed and assembled in-house by colleagues was employed to expose proteins to hydrodynamic flow; termed the extensional flow device (EFD)(Dobson *et al.*, 2017). Two 1 ml capacity borosilicate syringes with an internal diameter of 4.61 mm (1001 RN SYR, Hamilton Company®), were clamped facing one another, with a 75 mm long borosilicate capillary connecting the internal volume of each. The internal diameter of 0.3 mm in the capillary results in a constriction of over 15x, corresponding to a 238-fold increase in fluid velocity when the syringe plunger is compressed.

The capillary was held in place and the connection made watertight by a combination of a P10 pipette O-Ring (Gilson Inc.), which itself was held tight to the outlet of the syringe by a PFA conical ferrule and PEEK cup ferrule, all tightened in place by a screw-on nut (Hamilton Company®)(fig 2.3B). Before each experiment, the syringes were rinsed in the buffer of interest, leaving approximately 0.3 ml in one of the syringes. The capillary was attached to the syringe, and the pre-loaded buffer was expelled from the unattached end, to remove air. 0.5 ml protein solution was taken up for each experiment, and the open end of the capillary was immediately inserted into the second syringe by a O-Ring/ferrule/nut connection. Once this syringe pair was secured to the apparatus (shown fully assembled in fig 2.3A), a stepper motor with end plates attached to a central screw

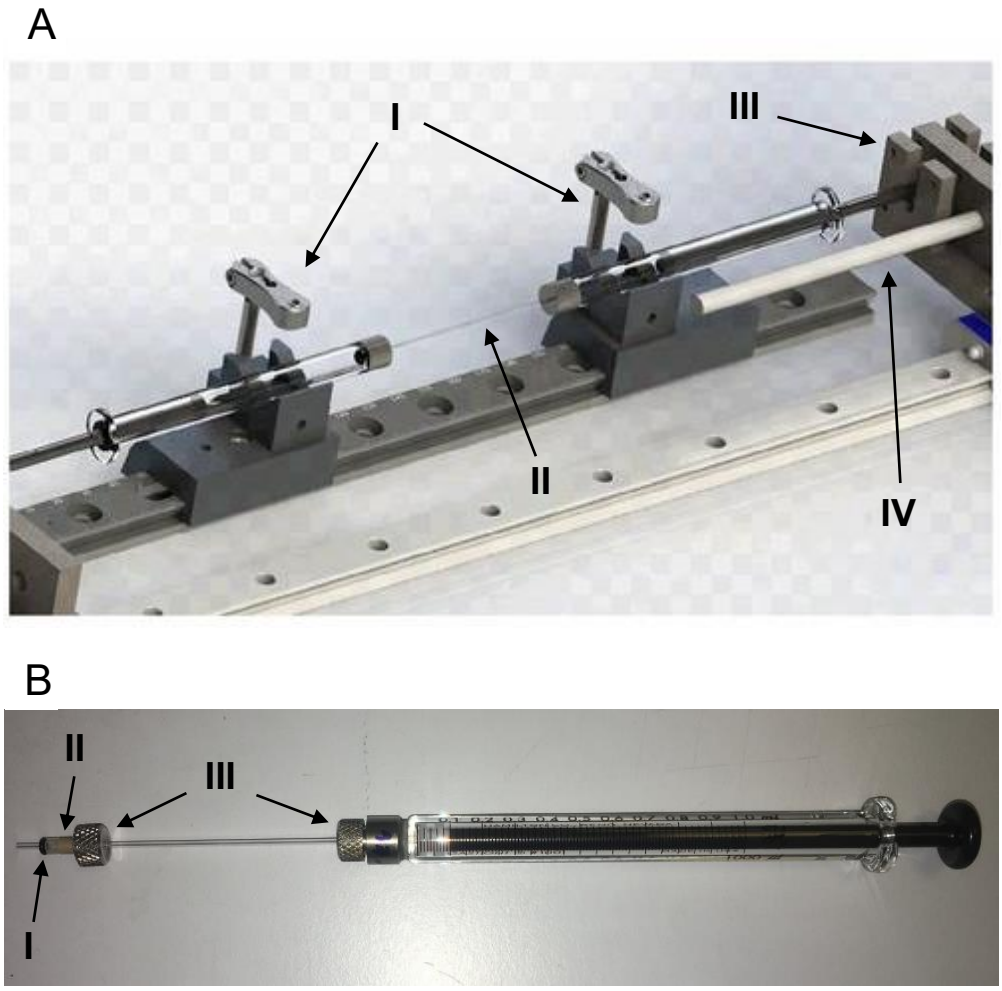


Figure 2.3: Extensional Flow Device Assembly

A) Extensional flow device apparatus comprises two 1ml syringes held in place by clamps shown loose here (I), a glass capillary connecting the two syringes (II), and a metal end plate used to compress the plunger of each syringe (III). The metal end plate moves along the length of a screw (IV), made to rotate by a stepper motor (not shown) Taken from Dobson *et al* (2017). B) One syringe with a capillary and O-rings (I), ferrules (II) and nut connections (III) attached.

was used to compress the syringes turn wise, for a set number of passes; each pass denoting a full compression of one syringe plunger.

The motor was operated at either 8 mm/s or 16 mm/s plunger velocity. These correspond to extensional strain rates of 11750 s^{-1} and 23421 s^{-1} respectively, and capillary wall shear rates of 50375 s^{-1} and 100751 s^{-1}, respectively. After the desired pass counts were completed, the samples were removed from the syringes and the instrument disassembled. The disposable capillaries were replaced with each new sample, and all other components were cleaned thoroughly using 2% (v/v) Hellmanex™ III alkaline cleaning liquid (Hellma® Analytics) followed by ddH_2O. The EFD was always operated at room temperature.

2.2.5.2 Analytical Reverse-Phase HPLC (RP-HPLC)

A Nucleosil 300 C4 column (Chromex Scientific Ltd) was used with a Nexera LC-40 Series HPLC system (Shimadzu UK Ltd.) to resolve individual components of a sample by hydrophobicity and quantify absorbance/fluorescence, respectively. Samples were prepared for injection as described in section 2.2.6.1 if analysed for solely aggregation, and were prepared as described in section 2.2.7.3 or 2.2.7.4 if unfolding was also to be measured. The Column was pre-equilibrated before use with ≥40 ml (10 times the column volume) 0.1 % TFA, 5% ACN (v/v) in HPLC-grade H_2O. Samples for analysis were transferred to plastic 300 µl HPLC vials (VWR™ International), and stored in the instrument sample tray at 4℃ while awaiting injection. Sample injection was frequently 50 µl, however this was reduced for a sample if fluorescence intensity exceeded 2,000,000 counts (arbitrary intensity units). Following sample injection, a gradient from 5% to 80% (v/v) ACN was established, as illustrated in fig 2.4. Wash and method steps were run at 1 ml/min.

An SPD-M20A photodiode array (Shimadzu UK Ltd.) was used to measure absorbance of each sample. For fluorescence measurements, a RF-20A detector (Shimadzu UK Ltd.) was turned on ≥30 minutes ahead of use. For detection of IAEDANS fluorescence (λ_{ex} = 336 nm, λ_{em} = 490 nm), sensitivity and gain were set to medium and 1x, respectively.

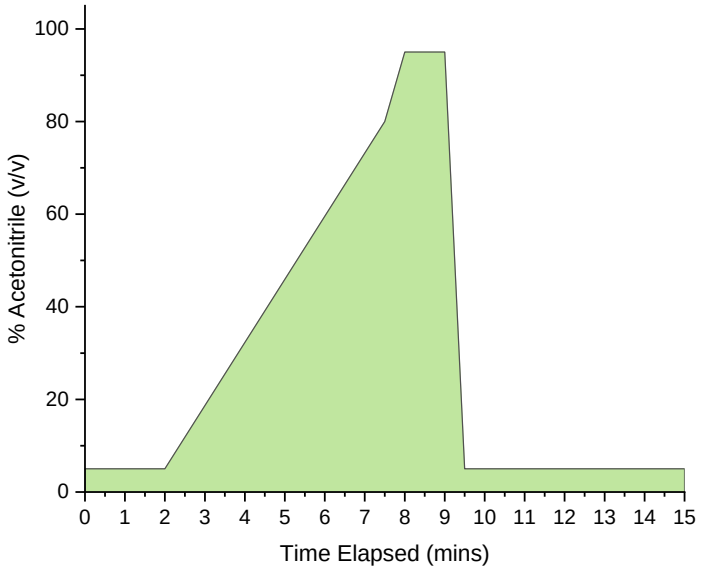


Figure 2.4: Standard Acetonitrile Gradient for Reverse-Phase Chromatography

2.2.5.3 Analytical Size Exclusion Chromatography-HPLC (HP-SEC)

HP-SEC was predominantly used to quantify aggregation in biopharmaceuticals, but RNase H* has also been analysed using this method. A TSKgel G3000SWxl size exclusion column (mean pore size 25 nm, mean particle size 5 µm, internal diameter 7.8 mm, length 30 cm)(Tosoh Bioscience LLC) was used to analyse the macromolecular status of protein samples. As with RP-HPLC, a Nexera LC-40 Series HPLC system (Shimadzu UK Ltd.) was used, but with only absorption measured, using the SPD-M20A photodiode array (Shimadzu UK Ltd.). A standard SEC running buffer (section 2.1.4) as employed by AstraZeneca was used for wash and elution steps. Before injecting samples, 2 CV equivalents of running buffer were used to wash the system. Sample injection volume was 50 µl, eluted at 0.5 ml/min over 30 minutes total measurement time. Absorbance peaks at 280 nm were integrated by area using the LabSolutions software provided with the HPLC system. A calibration curve was used to calculate predicted elution times of approximately 21.2 minutes for RNase H* and 17.9 minutes for a 150 kDa mAb (fig 2.5).

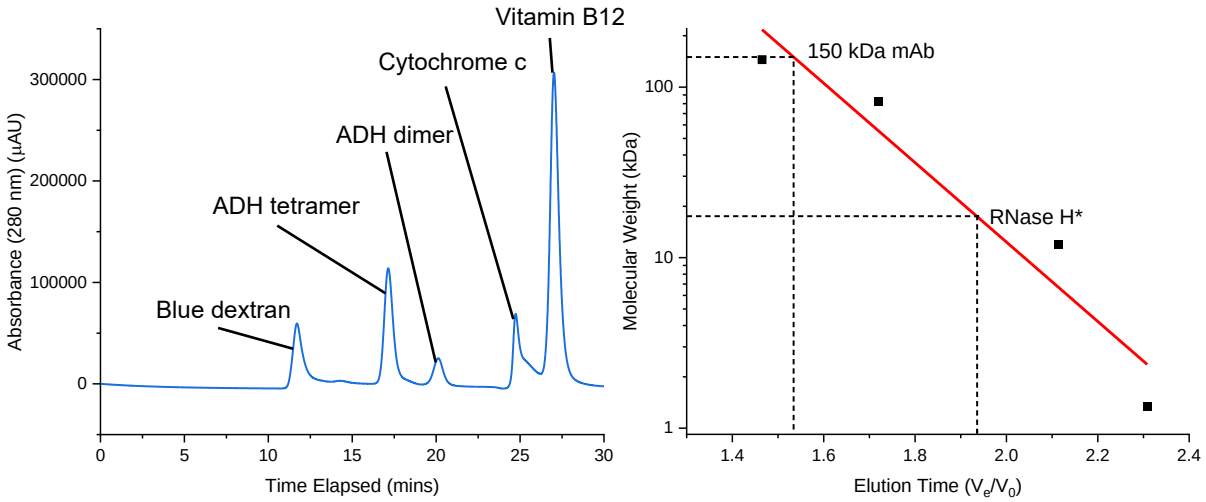


Figure 2.5: HP-SEC TSKgel G3000SWxl Calibration Curve

Source chromatogram (left) shown with peak contents indicated. Calibration curve (right) shown alongside, with calculated positions for a 150 kDa mAb and RNase H* indicated. ADH is alcohol dehydrogenase.

2.2.5.4 Analysis of Protein Surface Coverage by Spectroscopic Ellipsometry

Spectroscopic ellipsometry was carried out by Dr Leon F. Willis, as described in Willis *et al.*, (manuscript in preparation). A J.A. Woolam ellipsometer was used with λ_{ex} = 380 – 1000 nm, with an incidence angle of 70°. Polished 3.5 x 20 mm^2 silica wafers were used as the surface substrate and lowered into a 100 µl custom PEEK cuvette flow chamber, pre-irrigated with phosphate buffer (section 2.1.4). Baseline measurements were collected for 5 minutes at 5 ml/min flow rate, before replacement of the buffer with WT RNase H* in phosphate buffer at the desired concentration. Data was then collected for 20 minutes at 5 ml/min; this was deemed sufficient for surface coverage to plateau for all concentrations examined. Buffer was then pumped through the chamber for 5 minutes at 5 ml/min. Finally, the pump was turned off and measurements were collected for a further 15 minutes. Analysis was carried out by Dr Willis using CompleteEase Software (Version 6.7).

2.2.6 Quantification of Protein Aggregation

2.2.6.1 Analysis of RNase H* Aggregation

A modified version of the EFD was employed for this work, in which 3 pairs of syringes are used simultaneously. RNase H* was prepared at 0.25, 0.5 and 1.0 mg/ml, and each were assigned to the same syringe and fittings pair for each experiment, to minimise contamination between concentration regimes.

Samples were exposed to flow at 16 mm/s for the desired pass count (section 2.2.5.1). Immediately after stressing, samples were quenched by placing on ice. Control samples were left at room temperature for the same duration as the longest EFD-stressed sample experiment. Duplicate samples were clarified by centrifuging 150 µl at 30,000 RPM for 30 minutes, using an Optima™ MAX-XP ultracentrifuge with a TLA-100 fixed-angle rotor and polycarbonate open tubes (Beckman Coulter Life Sciences). 100 µl supernatant was collected, taking care not to disturb the pellet. Duplicate samples were pooled and analysed using RP-HPLC, measuring absorbance only (section 2.1.4). Peaks pertaining to RNase H* monomer were integrated to give monomer content, and percentage monomer remaining was calculated relative to the quiescent control sample.

2.2.6.2 Kinetic Modelling of RNase H* Aggregation

All modelling was performed using KinTek Explorer Professional version 6.3 chemical modelling software (Johnson, Simpson and Blom, 2009), and a kinetic model established by colleagues Willis *et al*., (manuscript in preparation). Each component of the 5-state flow-induced aggregation model was included, with 6 pairs of rate constants total, as in figure 3.14A. Data were manipulated as soluble protein concentration change with time. In RNase H* variants, 1, 0.5 and 0.25 mg/ml correspond to 57, 26.5 and 14.26 µM, respectively. All soluble protein loss experiments were conducted at 16 mm/s plunger velocity, with each pass corresponding to 3 seconds, meaning the greatest experimental time was 3000 seconds, with a total simulation time of 3300 seconds. Surface was included as a reagent with a concentration of 0.29 µM, and the on-rate to the surface (k_1) was fixed at 2.26/s. These values were calculated from measurements taken using spectroscopic ellipsometry (section 2.2.5.4) in Willis *et al.* (manuscript in preparation) that

indicate maximal average surface coverage of 82.9%, and using the calculated internal surface area of the EFD (1290 mm^2) with the R_H of WT RNase H* (2.2 nm). Additionally, aggregation rate constants (k_5/k_{-5}, and k_6/k_{-6}) were fixed at 0.453 and 0.00223, for the on and off-rates, respectively. This left 7 variable rate constants.

2.2.7 Flow-Induced Unfolding

2.2.7.1 Thermolysin-mediated Proteolysis Assay

Thermolysin from *Geobacillus stearothermophilus* (Sigma Life Sciences T7902-100MG) was dissolved into buffer containing 2.5 M NaCl and 10 mM $CaCl_2$, to produce a stock at 50x the desired concentration (all assay concentrations can be found in section 4.2.1) A stock RNase H* solution of 0.5102 mg/ml was prepared in 102 mM sodium acetate and 10 mM $CaCl_2$, then adjusted to pH 5.5 with HCl. The RNase H* and sodium acetate concentrations were set so that addition of 11 µl thermolysin stock to 539 µl RNase H* stock would result in 550 µl at final concentrations of 0.5 mg/ml and 100 mM, respectively.

Thermolysin stock was either added to the RNase H* stock immediately before exposure to flow, or immediately after. To quench thermolysin activity, 15 µl sample was immediately diluted with 5 µl 50 mM EDTA solution (pH 8). The resulting 20 µl samples were analysed using SDS-PAGE (section 2.2.3.1).

2.2.7.2 RNase H* Enzyme Activity Assay

A custom oligonucleotide substrate for RNase H* was produced, containing a fluorescein molecule at the 5' end and a DABCYL quencher at the 3' end (fig 2.6)(Corona and Tramontano, 2015). Oligonucleotides were produced by Eurofins Genomics. All steps involving water used nuclease-free water. The oligonucleotides were first suspended to a known concentration ≥100 µM in 100 mM Tris-HCl (pH 8.1) buffer, containing 0.5 mM EDTA. The oligonucleotides were combined with one another, as well as NaCl stock, to give a final concentration of 50 µM, and 50 mM NaCl in a thin-walled PCR tube. Using a T100™ thermal cycler (Bio-Rad Laboratories Inc.), the mixture was heated to 95 °C and held for 2 mins, before cooling to 25 °C over a period of 45 min, then holding at 4 °C. The

product was stored at -20°C if not used immediately. The molecular beacon was added as a component of the rection buffer (section 2.1.4) when ready for use.

Figure 2.6: RNase H* Oligonucleotide Substrate

A) Sequence is shown 5'to 3'. FAM denotes fluoresceine tethered to the end nucleotide, DAB denotes DABCYL tethered to the end nucleotide. B) Schematic showing the assembled oligonucleotide.

RNase H* was dissolved in 2x arginine succinate buffer (section 2.1.4) at pH 6, then serially diluted in the same buffer to reach the 2x the desired protein concentration. RNase H* was then diluted 2-fold into reaction buffer (section 2.1.4), either immediately before or after exposure to hydrodynamic flow. As soon as RNase H* had been exposed to the molecular beacon for the desired length of time, the fluorescence intensity of the sample was measured with λ_{ex} = 495 nm, either measuring an emission wavelength scan, or λ_{em} = 515 nm over a set time course.

2.2.7.3 IAEDANS Labelling in Urea

Throughout all work involving IAEDANS, the fluorophore was protected from light using foil covers, where possible. The evening before a planned experiment, a concentrated stock of the RNase H* variant of interest was prepared in urea unfolding buffer (section 2.1.4), alongside a stock of concentrated 5-((2-((iodoacetyl)amino)ethyl)amino)naphthalene-1-sulfonic acid (IAEDANS) in urea unfolding buffer. A volume of the two stocks was mixed gently, along with additional buffer, so that the final concentration of RNase H* was 28.5 μM (0.5 mg/ml), and IAEDANS was 1.43 mM (50 x that of RNase H*). This was left overnight to ensure unfolding of RNase H* and optimal IAEDANS labelling, so an internal standard for maximal labelling was available. This sample was often serially diluted with urea buffer, producing 6 reference samples in total, to visually depict the degree of adherence to a linear (fluorescence/[labelled protein]) fit, and guide the eye when comparing samples.

Each sample was diluted 2-fold into DTT to quench, for a final concentration of 500 mM DTT. These were then desalted using 0.5 ml 7 kDa MWCO Zeba spin desalting columns (Thermo Fisher Scientific), according to the manufacturer's instructions. Samples produced in this way were either analysed by SDS-PAGE (section 2.2.3.1) or RP-HPLC (section 2.2.5.2).

2.2.7.4. IAEDANS Labelling in Arginine

Stocks of RNase H* and IAEDANS were prepared in arginine succinate buffer, similarly to as described above (section 2.2.7.3), and mixed to give the same final concentrations. Immediately after mixing, samples were loaded into the EFD and stressed at 16 mm/s plunger velocity for the desired pass count (section 2.2.5.1). Control samples were prepared in the same way, but rather than being loaded into the EFD, were left at room temperature for the same duration as the highest-pass-count EFD-stressed sample. In some cases, 50 µl was taken from the control at 1, 10, 20, 30, 40 and 50 minutes from mixing, to capture quiescent labelling rate.

Samples were diluted with DTT to quench, and desalted using Zeba spin columns, as described above (section 2.2.7.3). Samples produced in this way were either analysed by SDS-PAGE (section 2.2.3.1) or RP-HPLC (section 2.2.5.2).

2.2.7.5 IAEDANS Labelling in Phosphate Buffer

RNase H* was prepared and stressed as described in section 2.2.7.4, using phosphate buffer (section 2.1.4) rather than arginine succinate for all buffer steps. The following experiment is depicted schematically in figure 2.7. After stressing of each sample and collection from the syringes, 0.75 ml Gdn/DTT buffer (section 2.1.4) was taken up into one syringe, and 100 passes at 16 mm/s were delivered, to remove as much insoluble surface-deposited protein as possible. The resulting solution was collected and quenched with 500 mM DTT. All Samples except these Gdn/DTT solutions were processed by ultracentrifugation as described in section 2.2.6.1.

To solubilise components, 50 µl of each supernatant sample, and each remaining pellet sample were diluted with 200 µl Gdn/DTT buffer and mixed thoroughly, before overnight incubation. All samples were then buffer-exchanged into Gdn/DTT buffer using 0.5 ml 7 kDa MWCO Zeba spin desalting columns (Thermo Fisher Scientific) according to the manufacturer's instructions. Following buffer exchange, samples were analysed using RP-HPLC, as in section 2.2.5.2 but with an additional 3 column volume equivalents at 5% acetonitrile following each sample injection. This was to ensure removal of any Gdn/DTT buffer from the sample before elution.

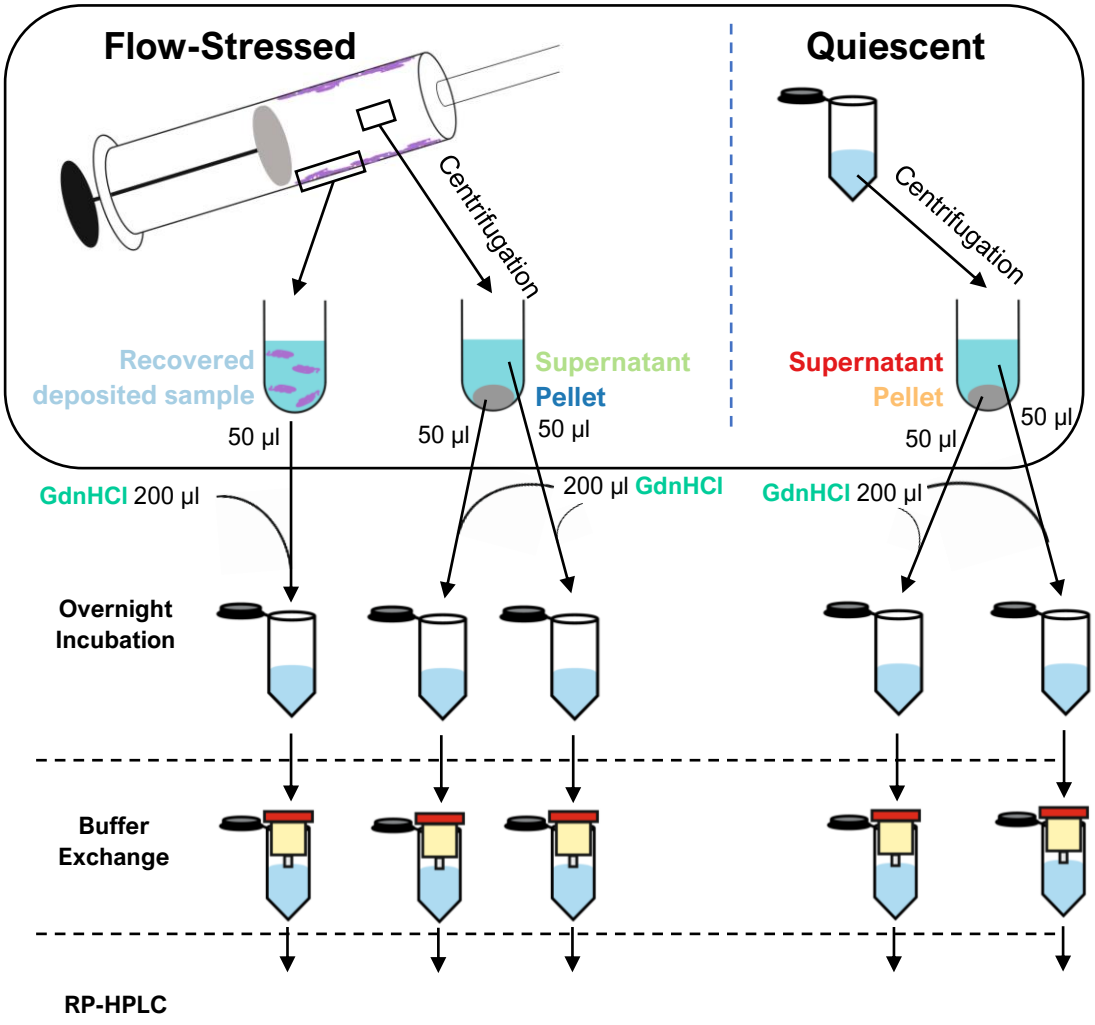


Figure 2.7: RNase H* Phosphate-Buffered Unfolding Assay

Schematic illustrating processing of RNase H* samples prior to analysis, following IAEDANS labelling in the presence of hydrodynamic flow.

2.2.8 Analysis of Biopharmaceutical Developability

2.2.8.1 Preparation of Biopharmaceutical Samples

Some of the biopharmaceutical panel selected for analysis were not in PBS. Before analysis, these biopharmaceuticals were buffer-exchanged into PBS using 0.5 ml 7 kDa MWCO Zeba spin desalting columns (Thermo Fisher Scientific) according to the manufacturer's instructions.

As the some of the sequences for the therapeutics analysed were proprietary, the primary sequence and details on molar extinction coefficient were unavailable to the author. Therefore, to calculate protein concentration, Pearce™ 660 nm protein reagent (Thermo Fisher Scientific Inc.) was used. 50 µl protein sample was mixed thoroughly but gently with 700 µl reagent, and left for 5 minutes before measurement against a protein-free sample, using a Shimadzu 1800 UV/Visible scanning spectrophotometer and 1 mm pathlength plastic cuvettes. A standard curve was produced, using MEDI 1912 STT at a known concentration, and this was used to interpolate protein concentration from absorbance values in all molecules analysed (fig 2.8).

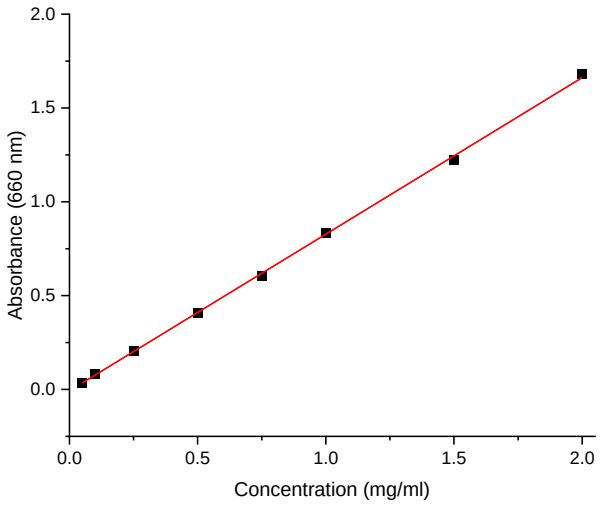


Figure 2.8: Pearce 660 nm Protein Concentration Measurement Standard Curve

Absorbance of 0.05 – 2 mg/ml MEDI1912 STT samples measured at 660 nm.

2.2.8.2 Analysis of Flow-Induced Aggregation

Samples were prepared at 0.5 mg/ml in PBS, before being stressed in the EFD as described in section 2.2.5.1, for 0 to 150 passes, at 8 mm/s. Samples were collected and clarified by removal of large aggregates, using a 0.22 µm pore diameter centrifugal filter (Millipore), at 16,700x g for 1 minute. 50 µl of filtrate was analysed using as TOSOH G3000swxl column, as described in section 2.2.5.3, but with an Agilent HPLC system. Peaks were integrated using the OpenLab ChemStation software (Agilent) associated with the instrument.

2.2.8.3 Accelerated Stability Analysis

Samples were prepared at 0.5 mg/ml in PBS and incubated in sealed plastic 1.5 ml Eppendorf tubes for 2 weeks at 45 °C, or 4 weeks at 40 °C. Following incubation, samples were collected and clarified, before analysis by HP-SEC, as described above (section 2.2.5.3). Multimer, monomer and fragment peaks were identified for integration by comparing elution times and peak areas to those of the same protein type at t=0.

2.2.8.4 Hydrophobic Interaction Chromatography (HIC)

Samples were prepared at 1 mg/ml in PBS, before being diluted 1:1 in hydrophobic interaction chromatography (HIC) running buffer A (section 2.1.4). Samples were analysed by absorbance (280 nm) using a Nexera LC-40 HPLC system (similarly to as described in section 2.2.5.3), but by injection onto a Butyl HIC column (Sepax), initially using pure HIC running buffer A. Samples were eluted from the column using a gradient from 0-100% HIC running buffer B (section 2.1.4), at a flow rate of 1 ml/min and a run time of 25 minutes per sample. Retention times for each peak were collected using the LabSolutions software provided with the HPLC system.

2.2.8.5 Stand-up Monolayer Adsorption Chromatography (SMAC)

Samples were prepared at 1 mg/ml in PBS, before being analysed by absorbance (280 nm) using the Nexera LC-40 HPLC system, similarly to as described in section 2.2.5.3. Samples were injected onto a Zenix SEC-300 column (Sepax), and eluted at 0.35 mg/ml over 20 minutes in SMAC running buffer (section 2.1.4). Retention times for each peak were collected using the LabSolutions software provided with the HPLC system.

2.2.8.6 Differential Scanning Fluorimetry (DSF)

A Rotor-Gene™ 6000 rotary qPCR machine (Corbett Research) was used to measure protein unfolding, by fluorescent detection of SYPRO™ Orange (Invitrogen™) dye binding. SYPRO Orange (5000x) was diluted to 200x with PBS, and each sample for analysis was prepared at 1 mg/ml in PBS. 20 µl of each protein was added to 10 µl SYPRO orange at these concentrations, and mixed thoroughly. These mixtures were transferred to thin-walled qPCR tubes on a 72-position rotor and held at 40 °C for 5 minutes in the qPCR instrument. These were then heated from 40 – 95 °C, with 60 s given between each 1 °C change in temperature, and changes in fluorescence (λ_{ex} = 300 nm, λ_{em} = 570 nm) were continuously measured.

Fluorescence versus temperature was plotted in SpectraGryph free open-source software (Menges), and first derivatives of data were captured, before being fitted to multiple gaussian peaks using Origin 2021 graphing software (OriginLab® Software). The peak maxima were taken to correspond to t_m values for each molecule.

2.2.8.7 Flow-Induced Taylor Dispersion Analysis (FIDA)

Analysis of protein-protein interactions by analysis of Taylor Dispersion was conducted using a Fida 1 instrument (Fida Biosystems Aps)(Otzen, Buell and Jensen, 2021). Samples were prepared in PBS at 0.5, 1, 2, 3 and 4 mg/ml, and made up to 50 µl in 96-well plates. Per sample, between 10 and 50 nl was injected into a fused silica capillary, with an internal diameter of 75 µm. PBS was used as both the wash and running buffer. The capillary was washed and loaded according to the following method, with samples measured at 25°C:

Step	Pressure (mbar)	Time (s)
Buffer Rinse and Equilibration	3500	140
Sample Injection	50	10
Sample Mobilisation and Measurement	400	300

Table 2.4: FIDA Protocol

The instrument produced ensemble values for hydrodynamic radius (R_H) in each sample, according to the following equation:

$$R_H = \frac{K_B T}{6\pi\eta D}$$

Equation 2.6

Where K_B = Boltzmann Constant, T = temperature (Kelvin), η = dynamic viscosity of PBS (1.02 cP at 25 °C (Ahmed *et al.*, 2021)) and D = diffusion coefficient, calculated as follows:

$$D = \frac{{R_c}^2 t_d}{24\sigma^2}$$

Equation 2.7

Where R_c = channel radius, t_d = average elution time and σ = peak variance.

To calculate k_D with R_H given by the instrument, the diffusion interaction parameter for each molecule, D was first calculated by rearranging equation 2.6. D was plotted against protein concentration, and k_D was extracted from a linear fit to this data.

2.2.8.8 *In Silico* Analysis

To protect intellectual property, sequences of all molecules in the mAb panel were undisclosed throughout the study. A collaborator, Isabelle Sermadiras (AstraZeneca) produced homology models of each molecule using AlphaFold2 (Google DeepMind), and analysed these using the Therapeutic Antibody Profiler (TAP) and structurally-corrected CamSol algorithms (Raybould *et al.*, 2019)(Sormanni, Aprile and Vendruscolo, 2015). Analysis using both techniques was performed on 05/12/2023 and 06/12/2023.

2.2.9 Statistical Analysis of Results

All analysis of correlations and clustering was carried out using Origin 2021 graphing software (OriginLab® Software).

2.2.9.1 Ranked Multivariate Analysis

Hierarchical clustering of RNase H* variants was carried out, first by ranking data manually, and then by using the 'hierarchical cluster analysis' function, with a Euclidean type calculation for distance between observables:

$$d_{ik} = \left(\sum_{j=1}^{p} (x_{ij} - x_{kj})^2 \right)^{\frac{1}{2}}$$

Equation 2.8

Where d_{ik} = distance between observables, i.e. between RNase H* variants, x_{ij} and x_{kj} = rank of an observable, p = the number of variables, i.e. 2; flow-induced soluble protein loss and labelling.

Ranked data was also analysed by pairwise Spearman's rank correlation and Pearson's correlation using Origin functions.

2.2.9.2 Analysis for Holistic Developability Parameter Ranking

With the aim of producing a holistic developability parameter with which to rank therapeutic molecules, data had to be processed. As the assays included produced results on different scales, all data was scaled, to give a value for each datapoint relating to position within the distribution:

$$V_{ij} = \frac{y - Y_{50\%}}{Y_{80\%} - Y_{20\%}}$$

Equation 2.9

Where V_{ij} = scaled value, y = unscaled value, $Y_{50\%}$ = median, $Y_{80\%}$ = 80% percentile value, and $Y_{20\%}$ = 20th percentile value for a particular dataset. This analysis method was taken from Jain, Boland, *et al.*, (2017).

Scaled values were then normalised on a scale of 0 to 1, with 0 being desirable and 1 being undesirable, in the context of biopharmaceutical development:

$$NV_{ij} = \frac{(V_{ij} - minV_{ij})}{(maxV_{ij} - minV_{ij})}$$

Equation 2.10

Where NV_{ij} = normalised value, $minV_{ij}$ = minimum value and $maxV_{ij}$ = maximum value for a particular scaled dataset. If a dataset was of the nature that a higher value (i.e. value closer to 1) was beneficial, each datapoint was subtracted from 1, to invert the sequence, so 0 was always most desirable.

The average NV_{ij} of a molecule in each of the assay groups established in section 6.3.1 was calculated:

$$Grp\overline{x} = \frac{(\sum NV_{ij})_x}{n}$$

Equation 2.11

Where $Grp\overline{x}$ = mean NV_{ij} within each group, $(\sum NV_{ij})_x$ = sum of NV_{ij} scores associated with assay group x, and n = number of assays associated with group x.

Lastly, the holistic developability parameter (HDP) was calculated by dividing $Grp\overline{x}$ by the number of groups established in section 6.3.1, i.e. 4:

$$HDP = \frac{\sum Grp\overline{x}}{4}$$

Equation 2.12

This produces the initial or 'naïve' HDP.

As different assay groups define the HDP to differing extents, this can be measured using multiple linear regression (MLR). Origin's MLR function was applied to the 4 assay groups (the independent variables), with the dependant variable being the developability parameter HDP was to be tested against (e.g. 40°C, 4 weeks accelerated stability). In this analysis, naïve HDP must be calculated without inclusion of the assay used to produce the developability parameter of interest. The β-coefficients produced from this analysis

were multiplied against the $Grp\bar{x}$ values for each assay group, and these were averaged as in equation 2.12 to give the ‘refined’ HDP.

3 Design and Characterisation of RNase H* Cysteine Substitution Mutants to Investigate Flow-Induced Protein Unfolding

As discussed in section 1.4.3, the contribution of hydrodynamic flow to protein aggregation is contentious, particularly in the case of extensional flow. This is largely due to the absence of a mechanistic understanding of flow-induced conformational change. As a fundamental understanding is lacking, it is logical to use a model protein to investigate this, so that results may give an understanding as to how a wider spectrum of molecules behave.

3.1 Design and Production of RNase H* Cysteine Variants

E. coli RNase H* was selected for use in this study, due to its mixed structural features and well-characterised folding pathway (sections 1.6.1 to 1.6.3). To probe conformational changes to RNase H* in a site-specific manner, it was eventually decided that a rapid and effectively permanent labelling-based assay would be most suitable for analysis of unfolding (described in detail in sections 4.3 and 4.4). In this assay, a thiol-reactive fluorescent molecule reacts rapidly with cysteine residues when they become exposed to solution following flow-induced conformational change.

The absence of cysteine residues in RNase H* (section 1.6.1) allowed site-specificity of labelling, as independent RNase H* variants each containing only 1 cysteine residue could be designed, without further mutagenesis being required.

In this chapter, a range of RNase H* cysteine variants were produced, before being characterised and analysed for flow-induced aggregation propensity. These cysteine variants showed diverse thermodynamic stabilities when analysed in chemical denaturation studies, and varied aggregation propensities, which were not easily explained by the available information, emphasising the need for structural insight.

3.1.1 RNase H* Cysteine Variant Design

To provide a range of positions for detection of unfolding across the RNase H* structure, cysteine substitution mutations were introduced (fig 3.1). Each of the resulting variants only contained 1 mutation at a time, so labelling could be attributed to a particular site. These were positioned across all of the foldons established by (Hu *et al.*, 2013), a foldon being a defined region containing secondary structure that is thought to comprise a step of a macromolecule's folding pathway. This consideration was made so that differences in unfolding propensity due to folding order would be detectable. Mechanical unfolding by optical tweezers revealed that a region corresponding to the red foldon (residues 1 – 42 and 121-155, last in the folding sequence) unfolds with a higher force input than the remaining structure; foldons that lie earlier in the folding pathway may be protected by the stability of this region (Cecconi *et al.*, 2005).

To facilitate use of cysteines as indicators of unfolding, prospective substitution sites were first analysed for predicted solvent accessibility, using the structure of WT RNase H* (PDB ID: 1F21) with DSSP (section 2.2.1.1). The selected cysteine variants showed low levels of predicted solvent exposure using the most sterically favourable predicted conformations (fig 3.2), with only I82C and H114C showing a solvent-exposed thiol group. This indicates minimal solvent exposure of these residues is likely to occur under native conditions, resulting in lower levels of quiescent labelling.

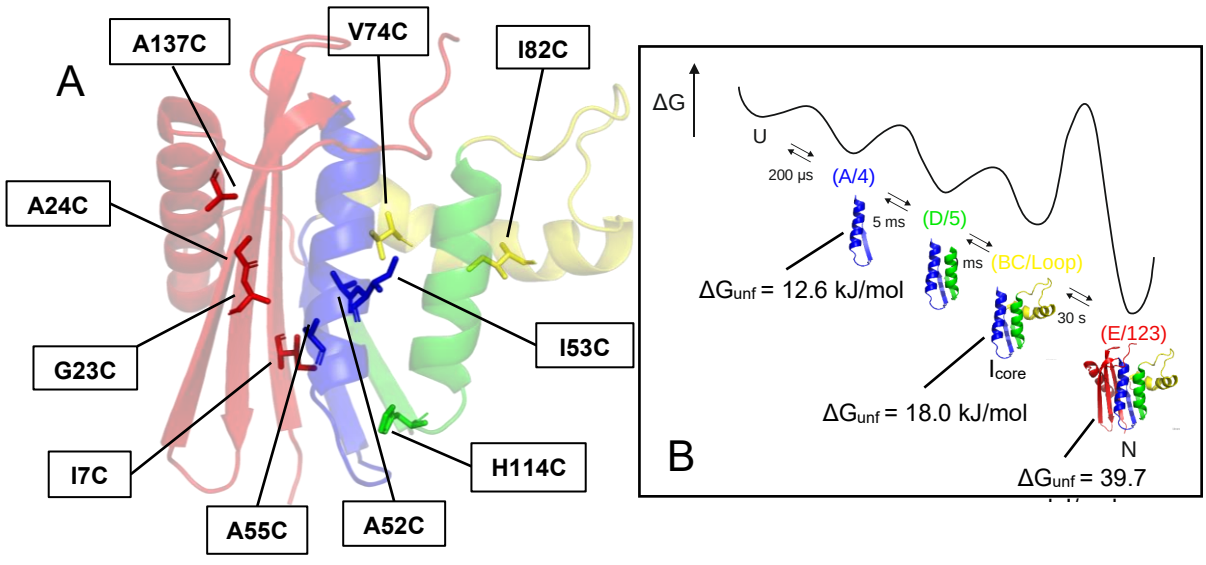


Figure 3.1: RNase H* Cysteine Variants

A) Cartoon structure showing sites of cysteine subsitutions on WT RNase H* (PDB: 1F21). Coloured according to foldon order in the free energy diagram (B) showing predicted foldons, approximate calculated folding time and ΔG_{unf} contributions associated with each. Adapted from Hu *et al*. (2013) and Rosen *et al*. (2015).

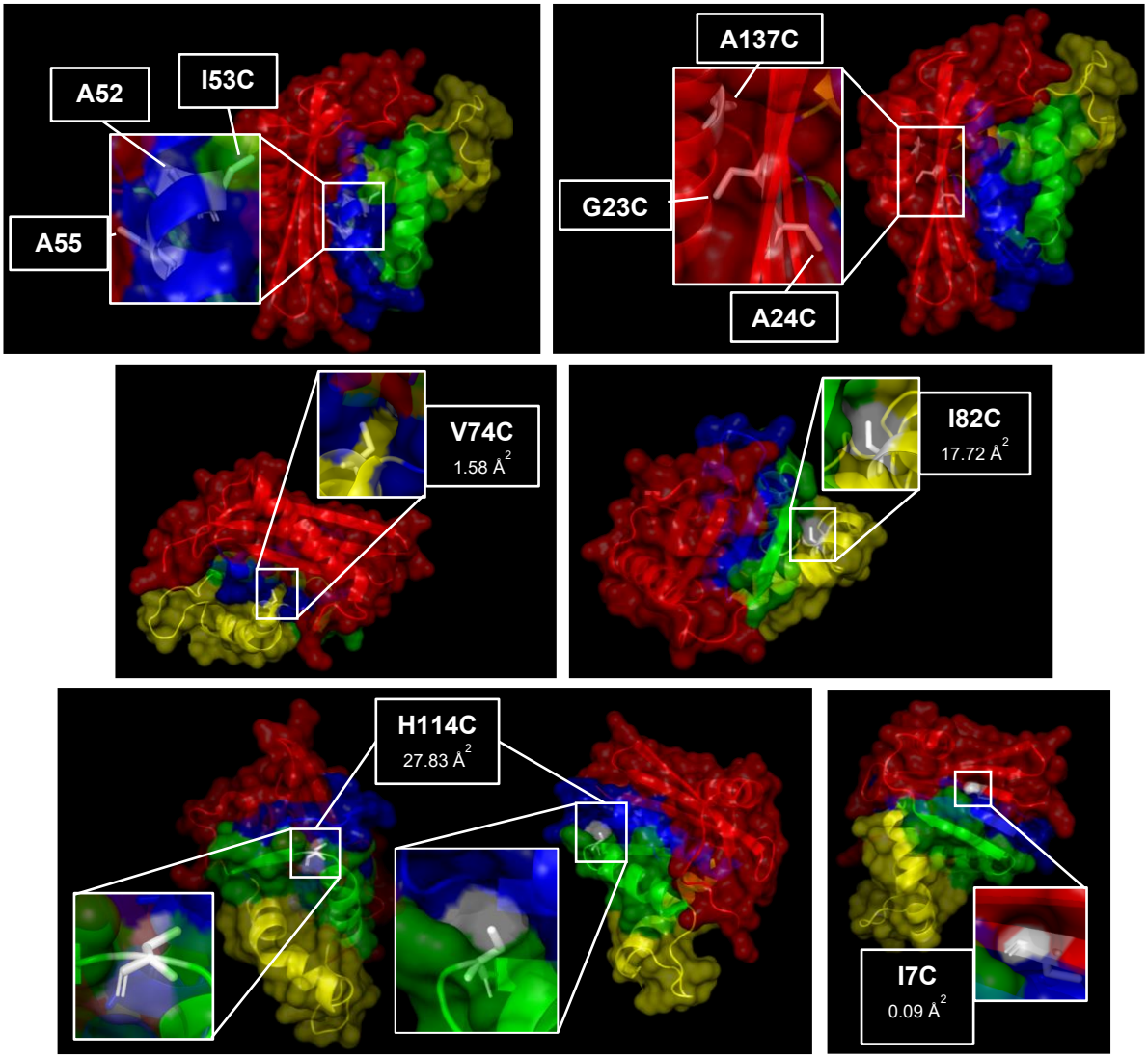


Figure 3.2: RNase H* Cysteine Variants Solvent Exposure

Surface exposure of cysteine residues (white) in each variant, with solvent accessible surface area values calculated by pyMOL. Images generated using pyMOL. Foldons shown with corresponding colours throughout. Labelled positions without Å values indicate no predicted solvent exposure. Maximum possible solvent accessibility for a cysteine residue in RNase H* is 135 $Å^2$.

3.1.2 RNase H* Cysteine Variant Production

The desired cysteine substitutions were first introduced into WT RNase H* as described in section 2.2.1.2. The gene containing RNase H* was in the form of a pAED4 vector, provided by the Marqusee Lab (UC, Berkeley). As well as the RNase H* gene, this plasmid contained an ampicillin resistance gene under the control of a T7 promoter, allowing controllable expression by addition of IPTG. Once each RNase H* variant had been produced, they were cultured and harvested, before being purified (section 2.2.2).

3.1.3 RNase H* Cysteine Variant Purification

In each instance, RNase H* was purified according to a 3-step protocol; heparin affinity chromatography, then ion exchange chromatography, then gel filtration chromatography. Following each purification step, samples were analysed using SDS-PAGE (section 2.2.3.1). An example chromatogram and SDS-PAGE gel from each step is given in figure 3.3.

As RNase H* is a nucleic acid binding protein that is positively charged at neutral pH (pI = 8.83), it could be purified using heparin affinity chromatography (section 2.2.2.3). A heparin-based column resin contains negatively charged sulphate groups which can interact with RNase H*, and these interactions can be overcome by addition of NaCl to elute bound protein. A higher volume fast-flow (FF) resin was used as an initial clarification step, in which RNase H* did not strongly bind but was released in the wash step, with more positively-charged impurities thought to be removed by the column. A high-performance (HP) resin with greater binding affinity was then used to bind and purify the clarified RNase H* (fig 3.3A and B).

Ion exchange chromatography (IEX) was then implemented as it was deemed effective in removing contaminating nucleic acid from the sample; the positively charged column interacts strongly with nucleic acid in the sample, but not with RNase H*, allowing the protein to be eluted rapidly and distinct from this contaminant (fig 3.3C)(section 2.2.2.4). As nucleic acid absorbs at 280 nm, removal was essential for accurate determination of protein concentration.

Lastly, size exclusion chromatography (SEC) was used to further purify the sample and remove non-monomer sample (fig 3.3D)(section 2.2.2.5). The ratio of absorbance between 260 nm and 280 nm was measured spectrophotometrically to determine content of nucleic acid contamination, with values below 0.6 deemed acceptable.

Purified samples were then analysed by TOF-MS to determine correct content and purity, before being lyophilised ahead of further use (sections 2.2.3.3 and 2.2.2.6, respectively). A representative mass spectrum for WT RNase H* is shown in figure 3.4, with a high level of purity indicated by the presence of only one major peak at 17500 kDa.

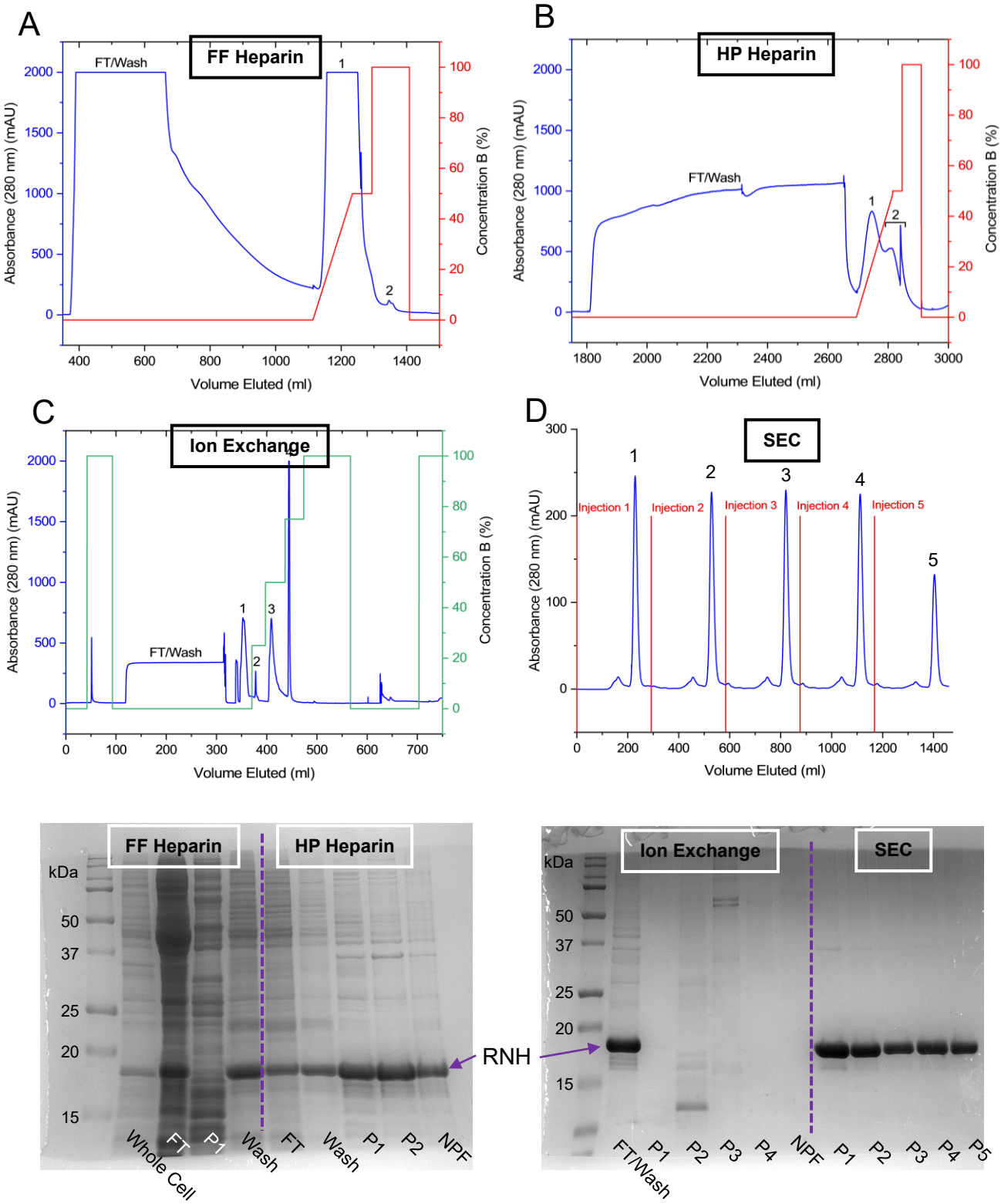


Figure 3.3: Representative Purification Process for RNase H*

I7C RNase H* purified by A) heparin fast flow (FF), B) heparin high-performance (HP), C) ion exchange and D) size exclusion chromatography (SEC). Associated SDS-PAGE gel analyses of peaks identified by each method are shown below. Numbers denote peak number, RNH denotes RNase H*. High-MW gels were used (table 2.2).

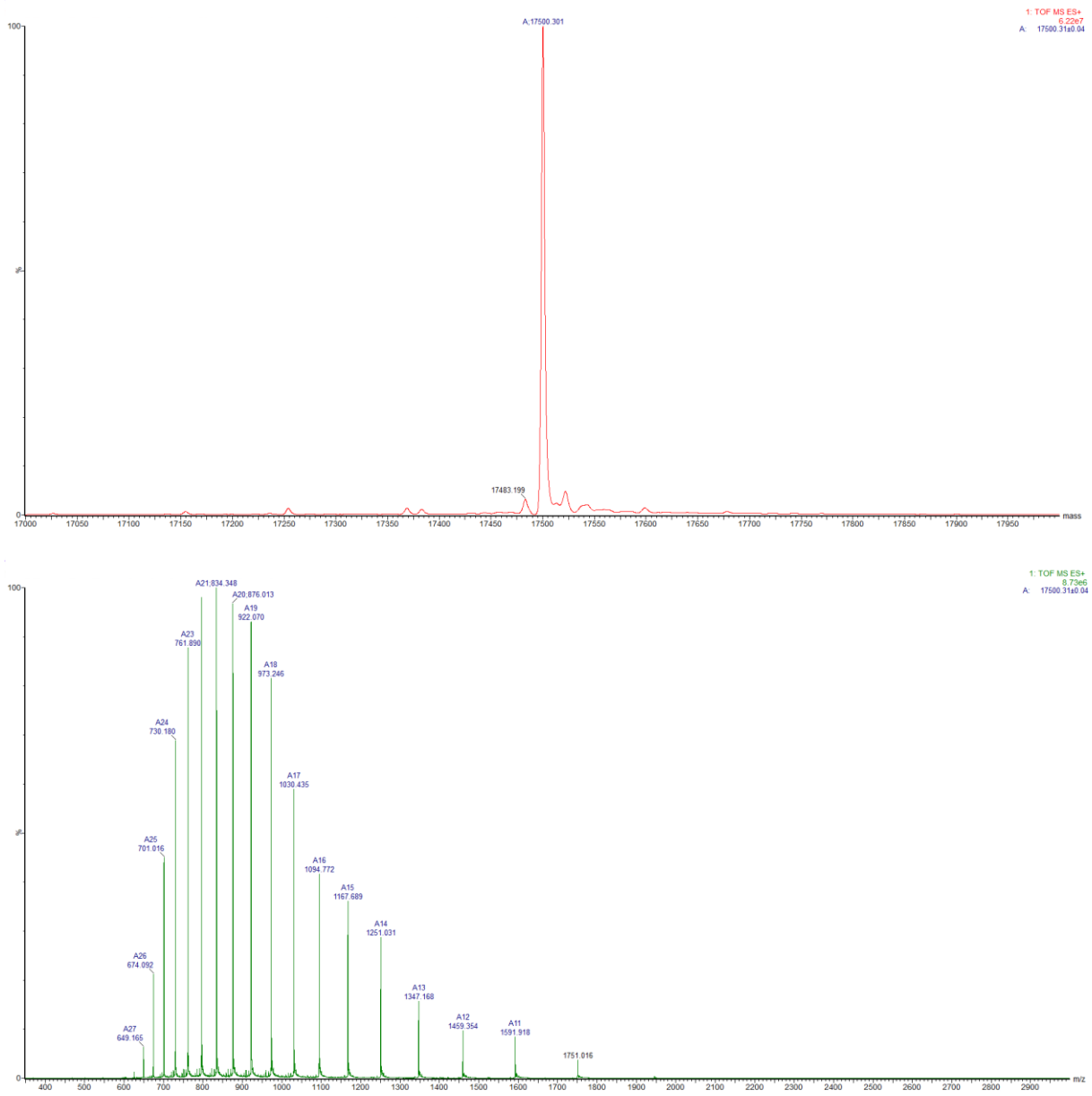


Figure 3.4: Representative Intact LC-TOF-MS Spectra

20 μM WT RNase H* analysed. Expected intact MW = 17500 kDa. A) Deconvoluted and B) raw spectra.

3.1.4 Cysteine Variants Structural Characterisation

To draw conclusions about the structural features that stabilise a protein against flow-induced unfolding and aggregation, the native secondary structure of RNase H* should be unperturbed under native conditions. To measure structural changes resulting from the cysteine substitutions, far-UV circular dichroism (CD) and intrinsic fluorescence emission spectra were collected. CD is able to report on changes to secondary structure content, whereas intrinsic fluorescence, mainly from Trp residues, gives information on the presence of tertiary structure.

As the 6 Trp residues in RNase H* are concentrated between positions 81 and 120, as well as being geometrically close, intrinsic Trp fluorescence may not be able to report on folding of the whole protein (fig 3.5A). However, the normalised spectra and CD spectroscopy data all appear to overlay with those of WT RNase H*; indicating the cysteine variants contain the secondary and tertiary structure associated with RNase H*. The peak in intrinsic fluorescence around 330 nm indicates that the Trp residues that are recessed from solution due to the presence of tertiary structure (fig 3.5B). If all Trp residues were solvent exposed, the emission maximum would be expected at 348 nm.

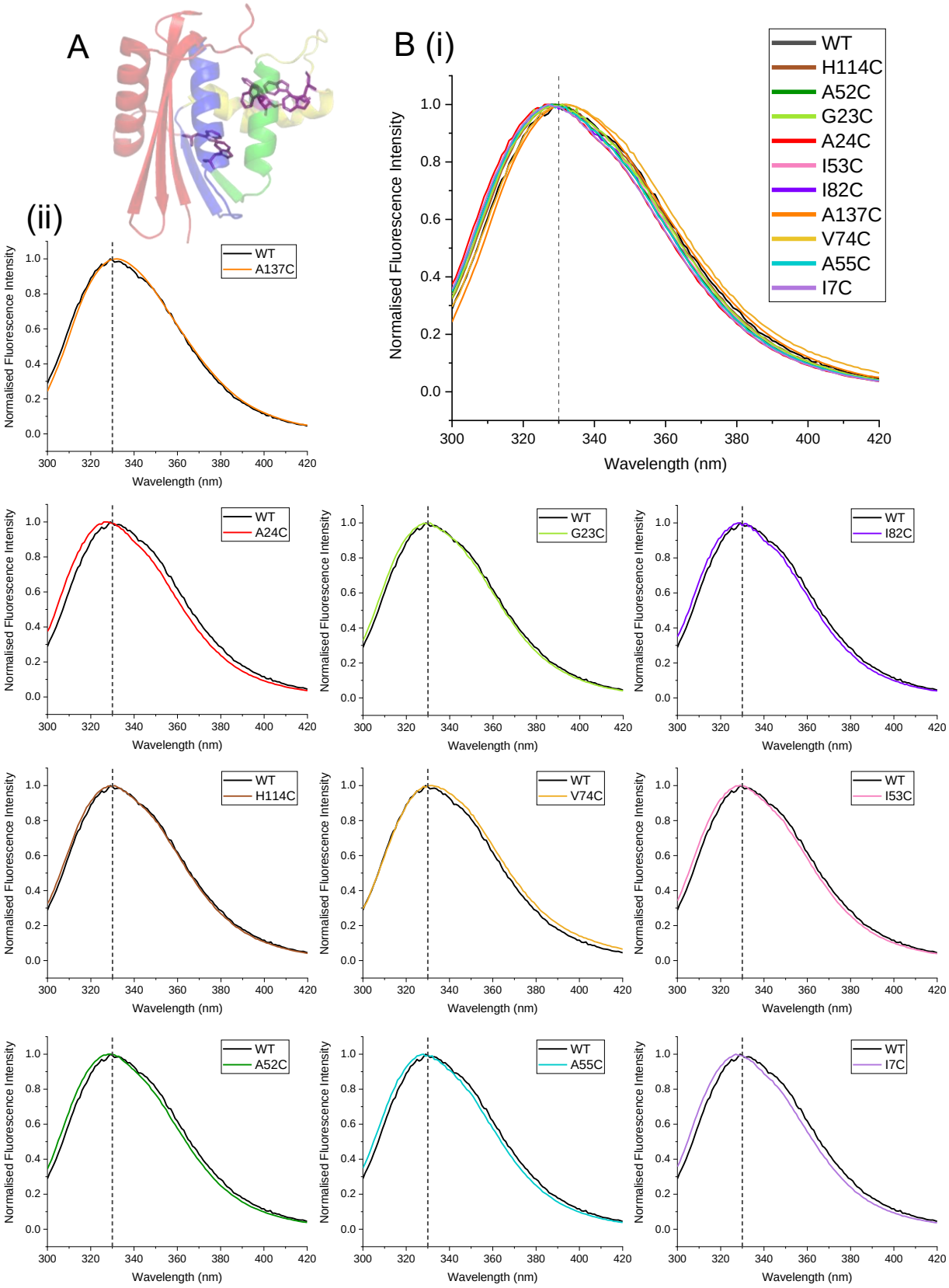


Figure 3.5: RNase H* Variants Intrinsic Fluorescence Spectra

A) Cartoon representation of WT RNase H* showing the positions of Trp residues (purple). B) Normalised fluorescence emission spectra for each cysteine variant shown (i) overlaid or (ii) with only WT overlaid, dotted line indicates 330 nm. λ_{ex} = 280 nm.

The broad ‘trough’ pattern in the CD spectra indicates a combination of the ‘double dip’ shape indicative of α-helices (with minima at 208 and 222 nm), and the ‘single dip’ at 217 nm, indicative of β-sheet structure (fig 3.6). This suggests that mixed α-helix/β-sheet structure is present. This characterisation allows the assumption that broadly the same secondary structure features are present in each variant.

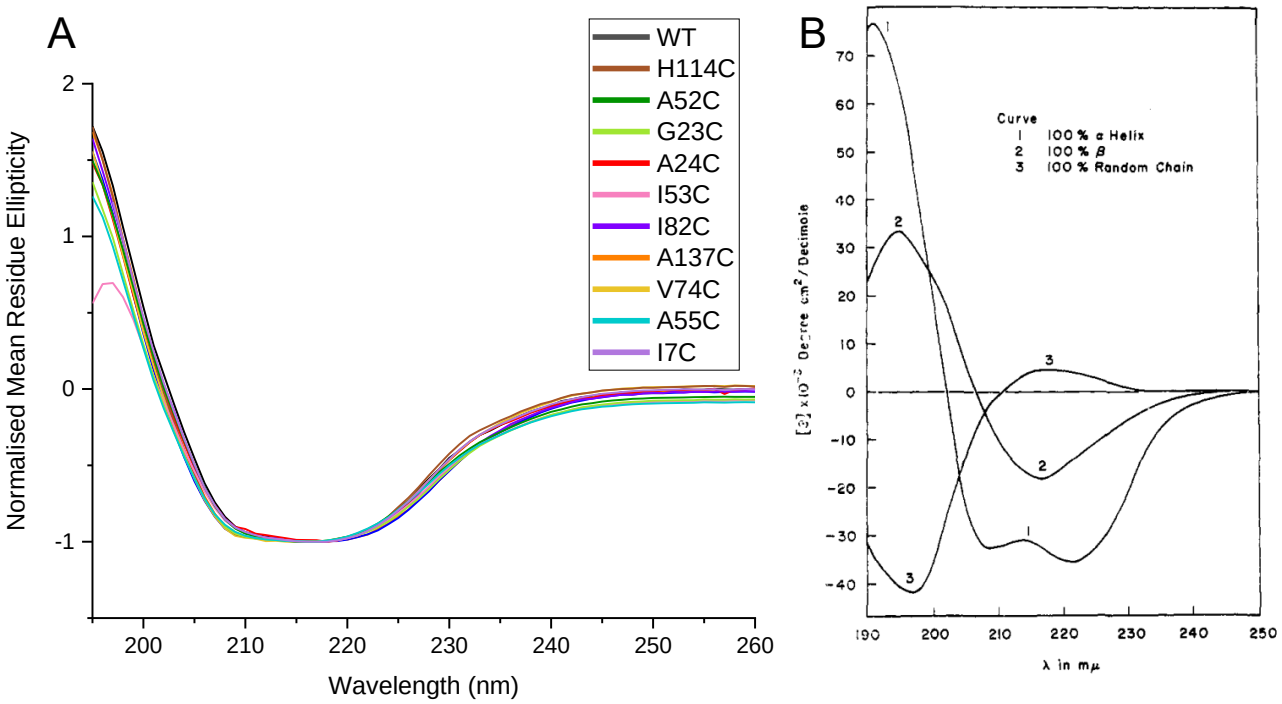


Figure 3.6: RNase H* Circular Dichroism Spectra Overlaid

A) CD spectra for all cysteine variants and WT RNase H* in phosphate buffer shown overlaid, normalised to the minimum mean residue ellipticity values reached. B) Example spectra for common secondary structure features shown for use as a reference, taken from Greenfield and Fasman, (1969).

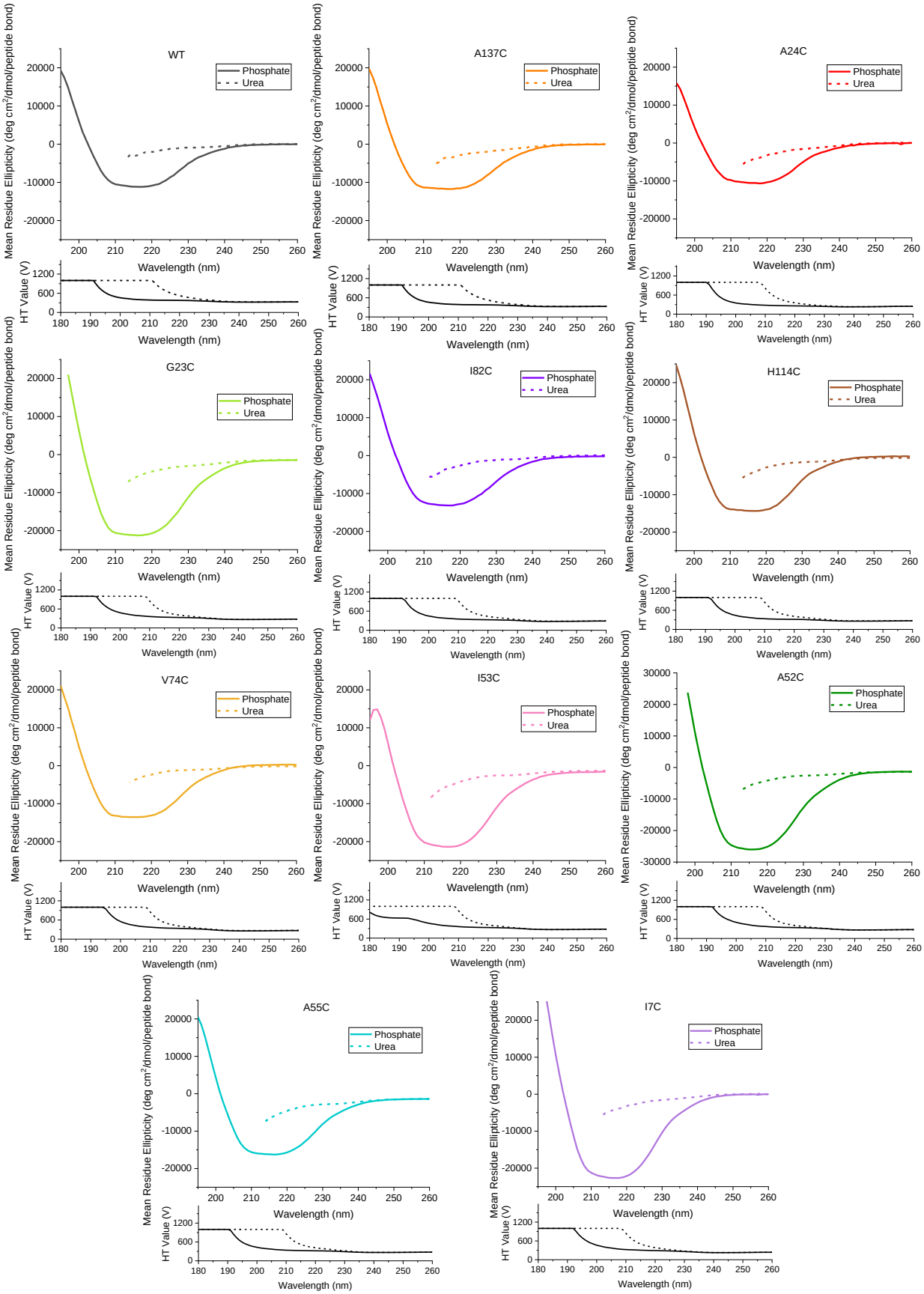


Figure 3.7: Individual RNase H* Variants Circular Dichroism Spectra

Circular dichroism spectra shown with mean residue ellipticity displayed for each variant, in either phosphate (solid line) or urea buffer (dashed line). High tension voltage values shown below. RNase H* concentrations between 0.1 and 0.2 mg/ml were used.

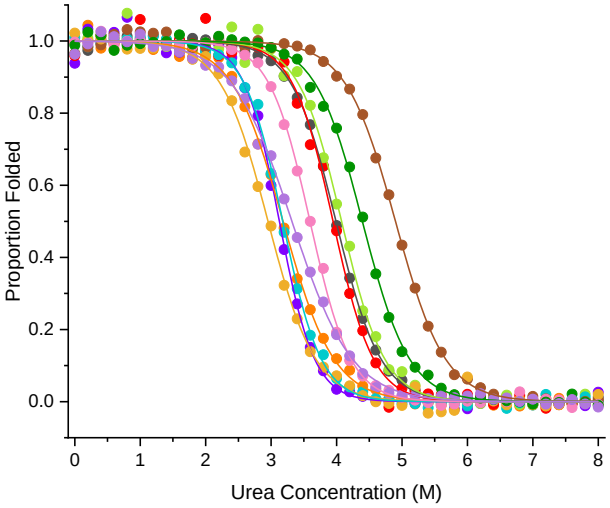
3.1.5 Thermodynamic Stability of RNase H* Variants

Thermodynamic stability is a frequently cited metric in determining a protein's resistance to aggregation (De Simone *et al.*, 2011)(Schön *et al.*, 2015)(Bunc *et al.*, 2022). Therefore, understanding how this parameter correlates with flow-induced unfolding could offer valuable insight. In each cysteine variant, thermodynamic stability was measured by chemical denaturation with urea at equilibrium, monitored by intrinsic fluorescence emission intensity (section 2.2.4.4). Data were fit to a two-state transition model as applied previously to RNase H* (Spudich, Miller and Marqusee, 2004). A range of thermodynamic stabilities were observed (figs 3.7 and 3.8), with some variants showing higher ΔG_{unf} values than WT; increases of up to 2.9 kJ/mol were observed in A24C and H114C. Mutations resulting in destabilisation showed more substantial effects, with reductions in ΔG_{unf} of up to 11 kJ/mol.

It was thought that these variations in effect might be because the cysteine substitution mutations were found in different foldons, which vary in their contribution to global protein stability (discussed in section 1.6.2). Interestingly, the position of substitutions in relation to the I_{core} (the largest and most stable folding intermediate) did not correlate with magnitude of changes in thermodynamic stability. However, the varied residue types selected for mutation may mean that if topology was playing a role in determining thermodynamic stability changes, it is obscured by the fact that different amino acid types were substituted, leading to differences from changes in entropies, non-covalent interactions, hydrophobicity and the hydrophobic effect. (Sturtevant, 1994).

Each variant appears to unfold according to a two-state mechanism; although intermediates may be present, they are lowly populated at equilibrium to the point where they are not detectable with this method. Despite the observation of similar secondary and tertiary structure character by CD and intrinsic fluorescence (section 3.1.4), a range of m-values is observed, suggesting differences in native or unfolded states between different variants and WT (Greene and Pace, 1974). This could be due to packing differences between secondary structure features, and is important to take into account when making considerations on unfolding propensity of the variants.

A



B

Variant	ΔG_{unf}(kJ/mol)	m(kJ/mol/M)
WT	28.9 ± 0.7	7.2 ± 0.1
A137C	20.3 ± 0.9	6.4 ± 0.3
I82C	28.9 ± 1.8	9.3 ± 0.6
A24C	31.8 ± 2.3	8.0 ± 0.6
G23C	30.8 ± 2.1	7.6 ± 0.5
A55C	21.6 ± 1.0	8.5 ± 0.3
H114C	31.8 ± 1.6	6.5 ± 0.3
V74C	20.2 ± 1.3	6.8 ± 0.4
I53C	28.6 ± 0.9	8.0 ± 0.3
A52C	30.3 ± 1.1	6.9 ± 0.3
I7C	17.9 ± 0.9	5.4 ± 0.2

Figure 3.8: Thermodynamic Stabilities and Associated m Values for RNase H* Cysteine Variants

Thermodynamic stabilities characterised by equilibrium denaturation curves, measured by intrinsic fluorescence spectroscopy. A) Data for each RNase H* cysteine substitution variant shown overlaid, with fits to a two-state model shown as solid lines. B) Corresponding thermodynamic stabilities and m values for each RNase H* variants. All samples analysed at pH 8.0, 25°C in 20 mM K_2HPO_4, 50 mM KCl buffer. Error values shown are standard deviations estimated for each coefficient using Igor Pro, from residuals of datapoints to the fit.

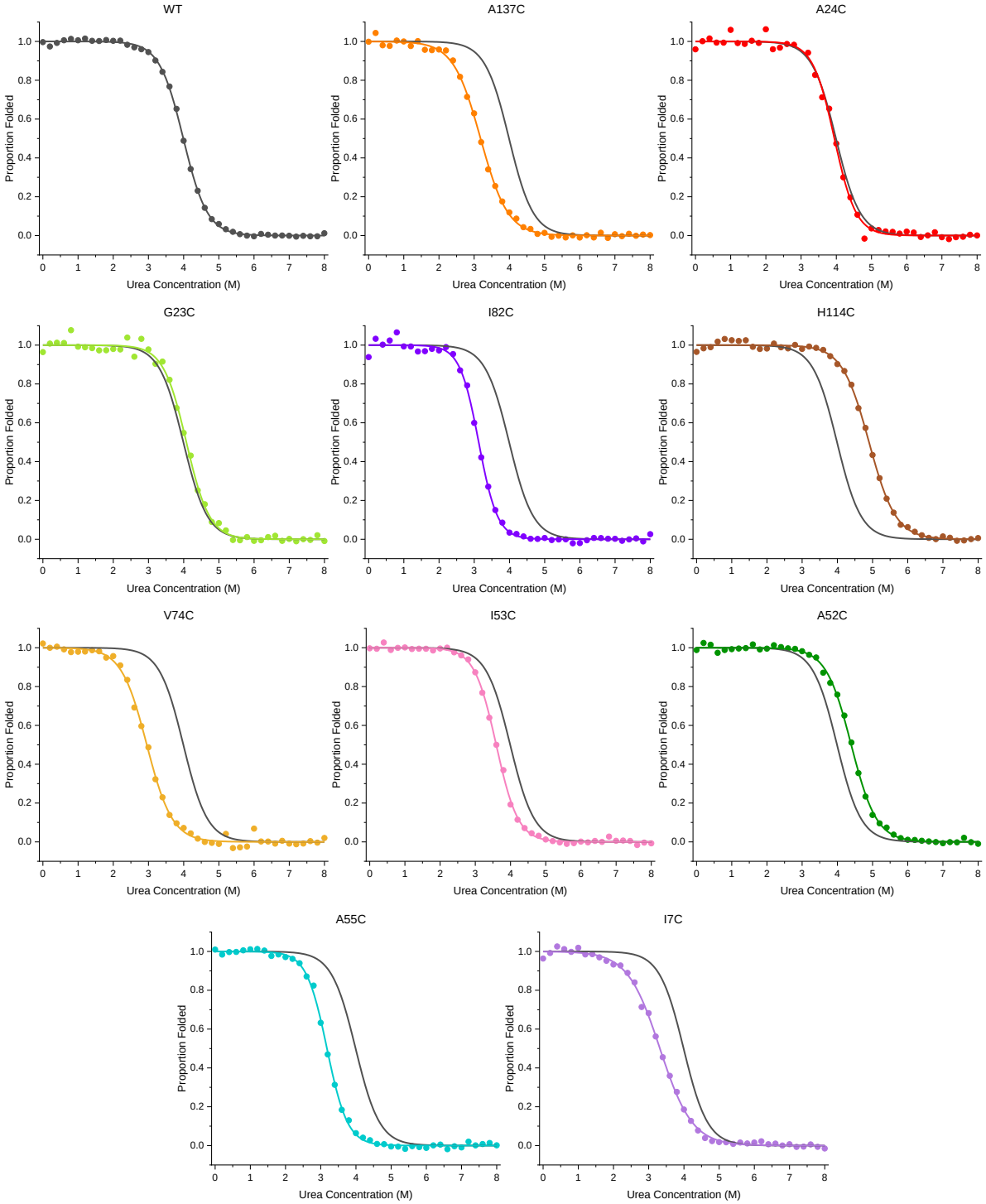


Figure 3.9: RNase H* Variants Equilibrium Urea Denaturation Curves

Individual RNase H* cysteine variant equilibrium unfolding curves, with fits to a two-state model shown as solid lines. WT RNase H* curve fit shown overlaid as a solid grey line.

3.2 RNase H* Aggregation Under Hydrodynamic Flow

3.2.1 Exposing Proteins to Flow with the Extensional Flow Device

To investigate the contribution of hydrodynamic flow to protein unfolding and aggregation, a bespoke device was employed, termed the extensional flow device (EFD). Unlike the Couette chambers commonly used in shear-centric studies (as discussed in section 1.5.1), the EFD comprises two syringes facing each other, connected by a narrow capillary. When a liquid protein solution is loaded into one of the syringes, it can be transferred into the other by compression of the syringe plunger, through the capillary. The decrease in internal diameter of over 15-fold, (4.61 mm to 0.3 mm) at the junction between the syringe barrel and capillary causes an increase in fluid velocity, exposing the molecules in solution to extensional strain in the direction of flow (section 1.4)(fig 3.10).

The solution is also exposed to significant shear stress along the length of the capillary, as well as the syringe barrel in the EFD(section 1.4). As discussed in sections 1.4.3 and 1.4.4, the effects of shear in other instruments is a subject of contention, but within the EFD shear does not appear to be the main contributing factor to aggregation.

Computational fluid dynamics simulations carried out by Dobson *et al*. (2017) illustrate that at 8 mm/s plunger velocity, fluid velocity increases by approximately 238-fold upon passing into the capillary. This is accompanied by a maximum strain rate of 11,750 s^{-1}, at the point of volume contraction in the device (fig 3.10B). A range of different plunger velocities has been investigated, which each produce specific maximum strain and shear rates. Below 25 mm/s plunger velocity, flow is expected to remain laminar. This is essential when studying the effects of extensional flow, to ensure that denaturation and aggregation are not a result of turbulence effects. Previously, 8 mm/s has been predominantly employed, as at this speed the resulting strain rates were able to distinguish a range of stabilities in antibodies (Willis *et al.*, 2018). The relatively high stability of the model protein employed in this work however, required use of increased plunger velocities; 16 mm/s was used predominantly in chapters 3 and 4. As chapter 5 examines the effects of flow on therapeutics, 8 mm/s was used, allowing comparisons to previous work.

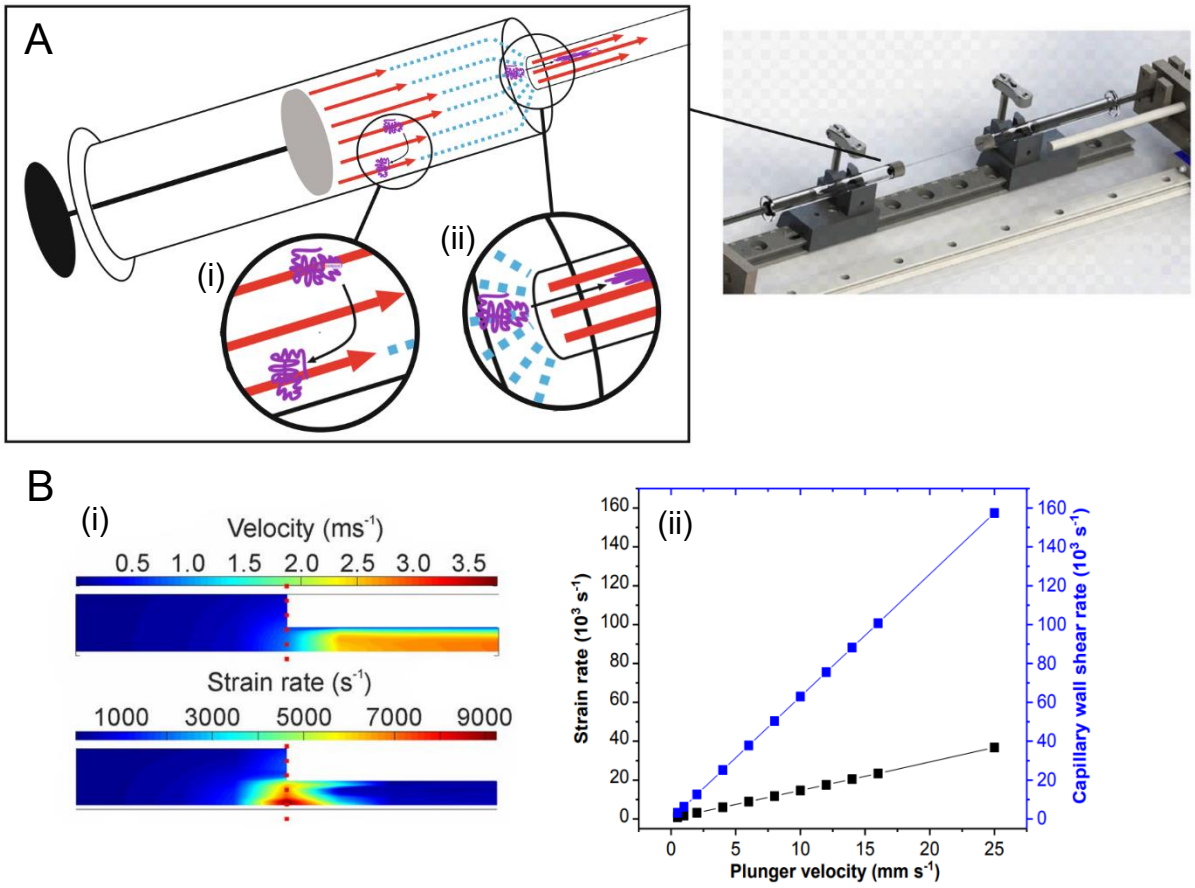


Figure 3.10: The Extensional Flow Device

A) Schematic illustrating i) shear and ii) extensional flow in the EFD. Red arrows indicate directional increase in fluid velocity. B) i) Computational fluid dynamics modelling of a 2D cross-section of the EFD, showing fluid velocity and corresponding strain/shear rates at 8 mm/s plunger velocity. ii) Extensional flow (strain rate) and shear (capillary wall shear rate) dependence on plunger velocity. Taken from (Willis, 2018).

As increasing plunger velocity has been shown to increase soluble protiein loss in biopharmaceuticals (Willis, 2018)(Willis *et al.*, 2020), it was often modulated in experiments where a shorter experimental time was desirable, as an alternative to changing time exposed to flow.

Due to the defined nature of these forces, and EFD experiments being measured by quantifiable ‘passes’ (where one pass is defined as complete emptying of one syringe into the other by EFD-driven plunger compression), proteins can be exposed to known magnitudes and durations of hydrodynamic flow.

3.2.2 Characterisation of EFD-Induced Aggregation Propensity in RNase H* Variants

While protein aggregation can be promoted by a number of factors (section 1.3.2), primary sequence is thought to play a pivotal role in directing a molecule's fate, with single substitutions rescuing molecules from or consigning them to aggregation (Wang, 2005)(Ebo *et al.*, 2020). The effect on aggregation of introducing cysteine substitutions in different positions throughout RNase H* was assessed, by measuring soluble protein content following exposure to flow (section 2.2.6.1). Loss of soluble protein was attributed to formation of insoluble higher-order aggregates.

It has previously been shown in that EFD-induced aggregation of monoclonal antibodies displays an inverse concentration dependence between 0.25 and 1 mg/ml protein concentration (Willis *et al.*, 2023), whereby as protein concentration increases, loss rate decreases as a percentage of total protein concentration. This inverse concentration dependence can be explained by the presence of two aggregation pathways; 'bulk' in-solution pathway, and a surface-mediated pathway, which can become saturated at higher concentrations, resulting in a slowing of aggregation rate increase (Treuheit, Kosky and Brems, 2002). This behaviour is described in detail in section 1.4.2.

To assess whether RNase H* displayed an inverse concentration dependence, WT RNase H* and each cysteine substitution variant was subjected to between 100 and 1000 passes in the EFD at 16 mm/s, using either 0.25, 0.5 or 1 mg/ml starting protein concentration (section 2.2.6.1). Samples were then clarified to remove large aggregates, and loss of soluble protein was then quantified relative to a quiescent control sample using RP-HPLC, as described in section 2.2.5.2.

After 1000 passes, or 50 minutes in the EFD at 16 mm/s (assay endpoint), soluble protein loss in the cysteine variants varied substantially across 3 concentration regimes: from 75% to 94% of the initial protein content at 0.25 mg/ml, 54-90% at 0.5 mg/ml and 31-68% at 1 mg/ml. (fig 3.11). A clear inverse concentration dependence was observed; mean soluble protein loss at the endpoint was 85.4%, 69.0% and 47.8% at 0.25, 0.5 and 1 mg/ml respectively (3.10A).

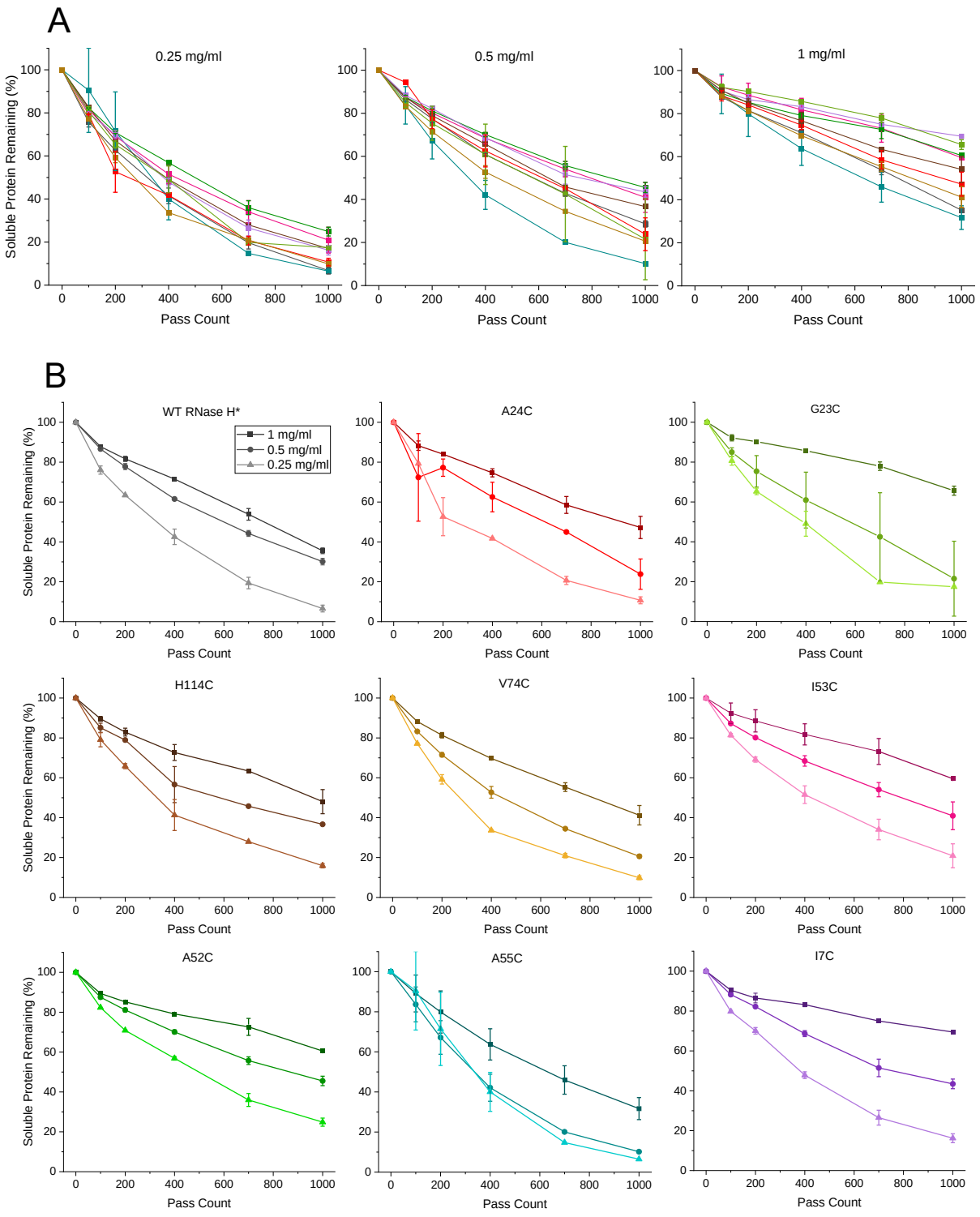


Figure 3.11: RNase H* Soluble Protein Loss with Hydrodynamic Flow

Concentrations indicated refer to starting RNase H* concentrations. Samples were stressed for the pass counts indicated, at 16 mm/s plunger velocity. Error bars denote standard deviation of the mean, n=2. A) Overlaid soluble protein loss data shown for comparison between RNase H* variants. B) Individual variants shown to allow comparison in soluble protein loss between concentration regimes.

Interestingly, the RNase H* variants (including WT) displayed different types of aggregation behaviour across the concentration range tested, with some (for example V74C) showing a transition to more significant monomer loss at higher concentrations (0.25 and 0.5 mg/ml), and others (for example WT) showing more significant monomer loss only at 0.25 mg/ml (indicated in fig 3.12).

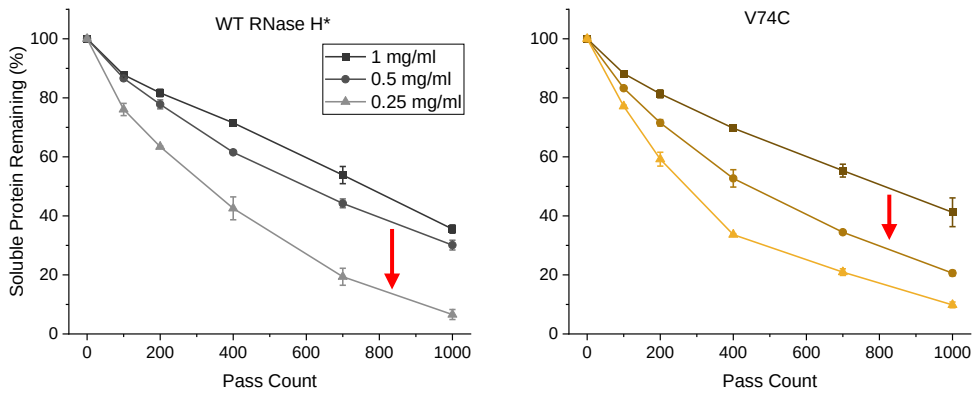


Figure 3.12: RNase H* Variable Soluble Protein Loss Behaviour

WT RNase H* shows different EFD-induced aggregation behaviour to V74C. Red arrows indicate a transition from less to more aggregation that is different to in the other RNase H* variant. Data taken from fig 3.11.

Such behaviour could be explained by groups of variants having different affinities for a surface, resulting in different availability of surface area for aggregation. While time constraints precluded analysis of the variants, the extent of monolayer formation was examined for WT RNase H* at 0.25, 0.5 and 1 mg/ml to examine if significant changes occurred, suggestive of the concentration regime being around K_D for the RNase H*/surface interaction.

To achieve this, surface ellipsometry was performed by a colleague, Dr Leon F. Willis (details can be found in section 2.2.5.4. When carried out in the same phosphate buffer as used in aggregation and unfolding studies (section 2.1.4), RNase H* quickly reached an adsorbate mass density (AMD) corresponding to 75-90% coverage (fig 3.13A). Notably, AMD rapidly increased and then reached a plateau within 10 minutes in each concentration regime. This, the fact that coverage does not increase significantly with higher protein concentrations, and that coverage is close to the maximum predicted value, suggests that RNase H* has saturated the available surface at all concentrations investigated.

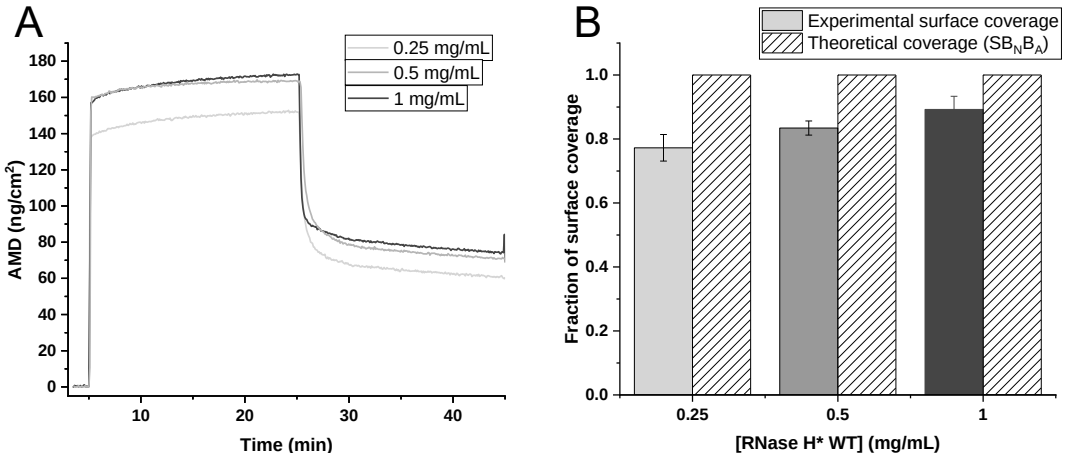


Figure 3.13: RNase H* Surface Coverage

A) Adsorbate mass density (AMD) change over time in the presence of WT RNase H* at the concentrations indicated. Lines correspond to lines of best fit. B) Fraction of surface coverage is given as measured maximal AMD reached, relative to the calculated theoretical AMD of a WT RNase H* monolayer (191 ng/cm^2). Error is standard deviation of the mean, n=2. Taken from Willis *et al*., (manuscript in preparation).

3.2.3 Modelling RNase H* Aggregation Kinetics

Due to the involvement of multiple aggregation pathways, it is important to decipher which steps are most important in driving aggregation, and how this hierarchy might vary between different proteins. A model established by colleagues (Willis *et al.*, manuscript in preparation), and based on experimental data, shows that the model that best fits the data is one where molecules partially unfold and nucleate in bulk and on surface, and aggregate in bulk. This is termed the SB_NB_A model (depicted in fig 3.14A).

This model was applied to experimental data showing loss of soluble protein concentration in RNase H*, using the mean maximum surface coverage value of 82.7% from surface ellipsometry measurements above (fig 3.13B) to estimate total surface area in the syringe/capillary assembly, available as a discrete concentration (0.29 μM)(section 2.2.6.2). Kintek chemical kinetics software was then used to generate fits to raw data using this model, with rate constants associated with each step produced.

When initially fitted to EFD-induced aggregation of the IgG mAb STT, the fitting describes the data well at all 3 concentrations (χ^2/DoF = 0.037)(fig 3.14B). The model was applied to WT RNase H* aggregation data and fitted well, demonstrating it is applicable to substantially different proteins (χ^2/DoF = 1.13). Several rate constants were markedly different between RNase H* and STT (table 3.1); both rate constants involving

dissociation from the surface, k_{-1} and k_4, were slower in RNase H*, indicating higher surface affinity. Additionally, conformational change in bulk, described by k_2 and k_{-2}, was more rapid and less reversible in RNase H*. This offers insight into the mechanisms that govern the reduced aggregation propensity in RNase H* compared to STT; suggesting it is a result of slower dissociation from the surface.

Protein	K_1	k_{-1}	k_2	k_{-2}	k_3	k_{-3}	k_4	k_{-4}	k_5	k_{-5}	k_6	k_{-6}
WT RNase H*	2.26	7.25E-06 ±0.453	2.08E-04 ±0.000416	2.02E-07 ±0	0.0241 ±0.256	0.0328 ±1.14	0.72 ±0.345	167 ±110	0.453	0.00223	0.453	0.00223
STT	0.337	1.23E-4	1.08E-4	2.11E-6	0.024	0.107	21.3	33.9	0.453	0.00250	0.453	0.00250

Table 3.1: Rate Constants Produced from Fitting WT and STT to an EFD-Induced Aggregation Model

Rate constants (s^{-1}) produced from fits shown in fig 3.13. Error is standard error, estimated by KinTek.

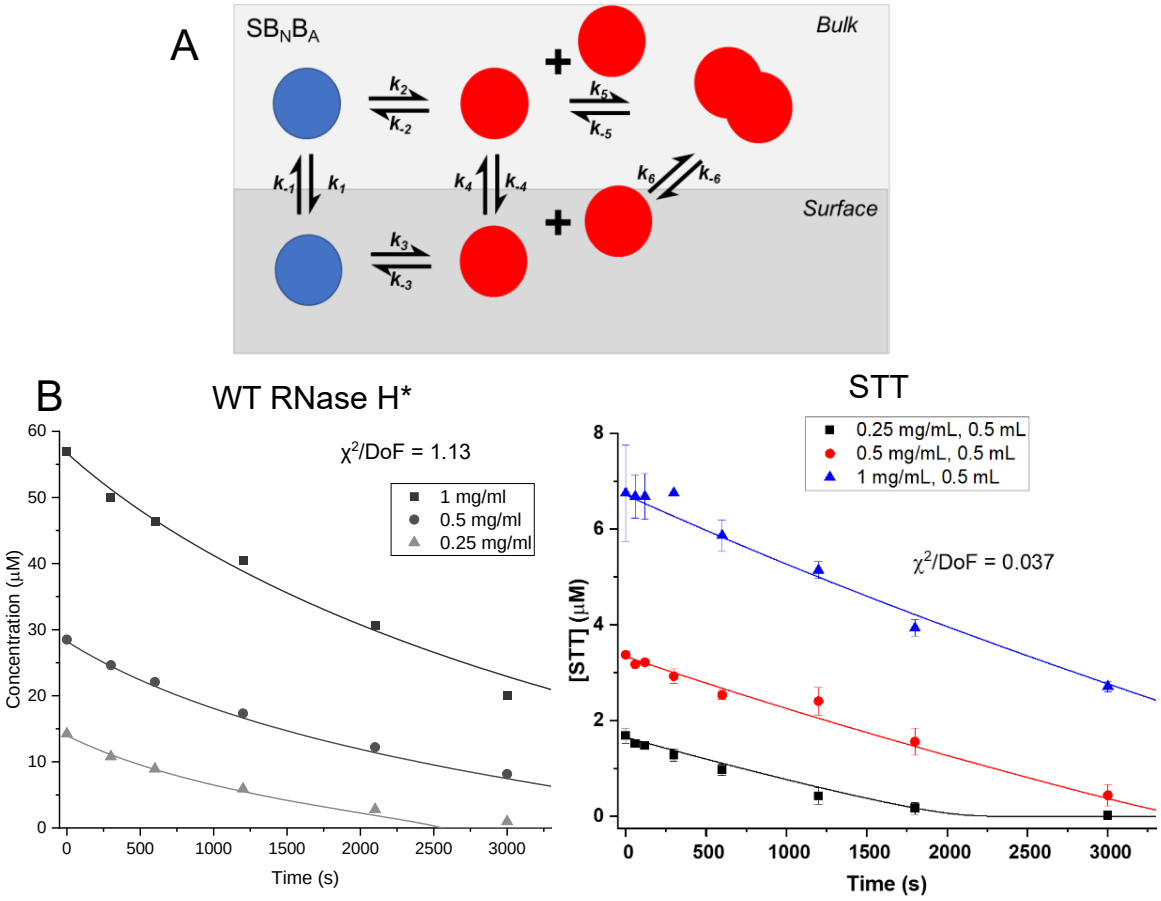


Figure 3.14: EFD-Induced Aggregation Described by a Kinetic Model

A) Surface and bulk nucleation with bulk aggregation (SB_NB_A) kinetic model schematic for flow-induced protein conformational change and aggregation, developed by Willis *et al*. (manuscript in preparation). B) Simulated soluble protein loss with EFD-induced aggregation in WT RNase H* and STT, a monoclonal antibody. Datapoints produced experimentally shown as filled shapes. Solid lines indicate simulated fits to the data.

To rationalise the differences in aggregation behaviour in the RNase H* cysteine variants, the model described above was applied to the datasets acquired in section 3.2.2. Surprisingly, the fits to data were worse, with χ^2/DoF values ranging from 10.13 to 34.13 (fig 3.15).

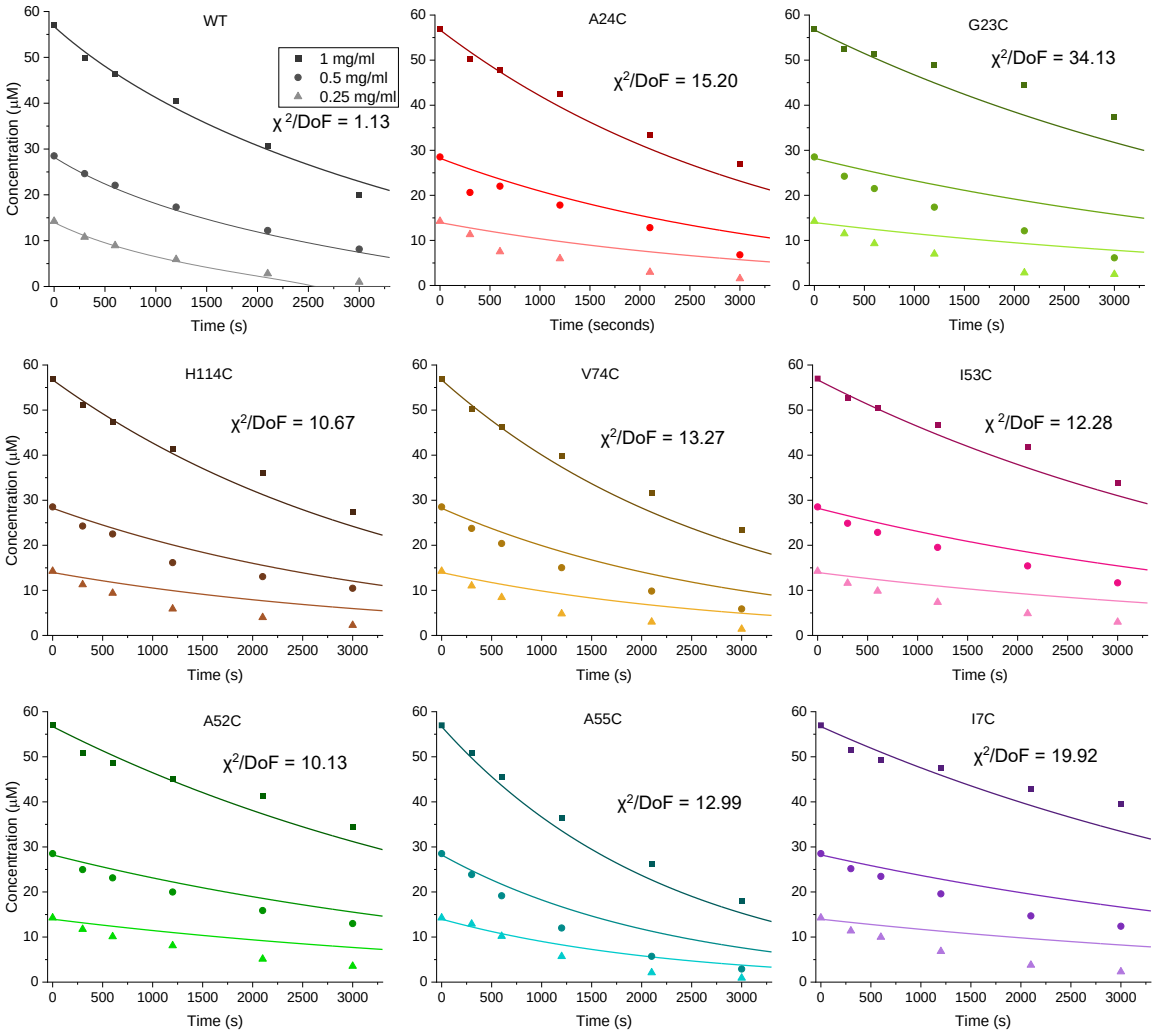


Figure 3.15: EFD-Induced Aggregation in RNase H* Cysteine Variants Described by a Kinetic Model

Simulated soluble protein loss with EFD-induced aggregation, using the SB_NB_A model. Datapoints produced experimentally shown as filled shapes. Solid lines indicate simulated fits to the data.

3.2.4 Conformational Change in Bulk Most Affects Soluble Protein Loss

As fits were poor, conclusions were not drawn from the rate constants produced (included in table A2, appendix). Instead, a sensitivity analysis was conducted to investigate how changing different rate constants affects simulated monomer loss according to the model, and identify which rate constants are more prominent in progressing EFD-induced aggregation. I7C RNase H* was used as the representative cysteine variant during this analysis, due to the degree of monomer and calculated rate constants being close to the middle of the cysteine variant range.

As discussed in section 1.4.2, a simpler model does not describe monomer loss data under hydrodynamic flow. However, the majority of rate constants could be divided or multiplied by 10^6 with no noticeable change in soluble protein loss; indicating the steps they are associated with are necessary but not rate-determining. Those that did result in a change are included in fig 3.16. k_2 and k_{-2} had by far the greatest effect on soluble protein loss; a 10-fold increase in k_2 results in almost complete loss of soluble protein by the endpoint of the assay, whereas a 10-fold decrease results in complete protection against soluble protein loss. This indicates that under these conditions, aggregation-competent species in bulk solution are pivotal in aggregation, a conclusion further supported by the observation that the only other rate constant causing a major effect is k_4, the constant for release of aggregation-prone species from the surface into the bulk.

The significance of these specific processes indicates that understanding the root causes for these differences in bulk unfolding and surface affinity is therefore fundamentally important in characterising the mechanisms that drive flow-induced aggregation.

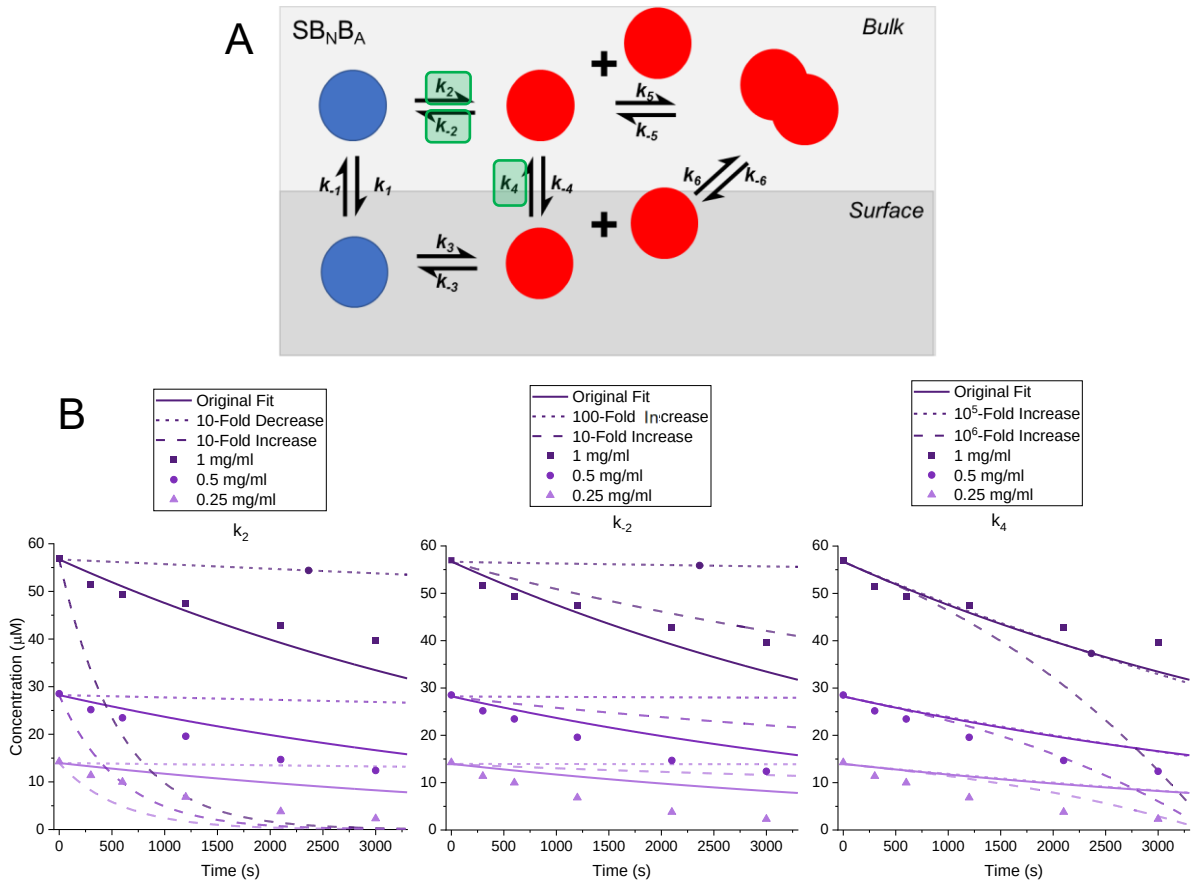


Figure 3.16: Sensitivity Analysis of RNase H* Aggregation Mechanism Under Flow

A) SB_NB_A kinetic model schematic with rate constants of significance highlighted in green. B) Fitted I7C RNase H* soluble protein loss data used for analysis (fig 3.14). Datapoints produced experimentally shown as filled shapes. Solid lines indicate simulated fits to the data, dotted lines indicate simulated fits with the increase in rate constant indicated in the figure legends, dashed lines indicate simulated fits with indicated decrease in rate constant.

3.3 Investigating the Causes for Different Aggregation Behaviours

This diversity in aggregation behaviour and apparent surface affinities from individual substitutions implies that certain regions within the protein are more important in determining aggregation mechanism and propensity. Understanding the root causes of these differences is a key part of biopharmaceutical design and could be used to direct protein engineering of future therapeutics more effectively, while also having implications in research of disease processes.

3.3.1 Aggregation-Prone Regions do not Explain Variations in Soluble Protein Loss

It was considered that the differences in aggregation propensity could simply be explained by the alteration of the intrinsic aggregation propensity of RNase H* upon substitution with cysteine. For example, introduction of a cysteine residue into an aggregation-prone region (APR)(discussed in section 1.2.2) could interfere with its hydrophobicity, reducing its aggregation potential (Wang *et al.*, 2009). Conversely, the thiol group is prone to oxidation which could promote aggregation through intermolecular disulfide bonds (Lévy *et al.*, 2019).

APRs were predicted using Aggrescan 3D (version 2.0) and TANGO. Aggrescan 3D was selected because it can integrate 3D structure into calculations, as well as intrinsic predicted aggregation propensity of the primary sequence from residue scores (Zambrano *et al.*, 2015). It also considers solvent accessibility of aggregation-prone stretches, to ensure the hydrophobic core is not automatically assigned as an APR. TANGO was included as a partially orthogonal algorithm that predicts β-sheet aggregate forming-propensity of a protein from primary sequence data (Fernandez-Escamilla *et al.*, 2004). The WT RNase H* primary sequence was analysed using both of these methods,

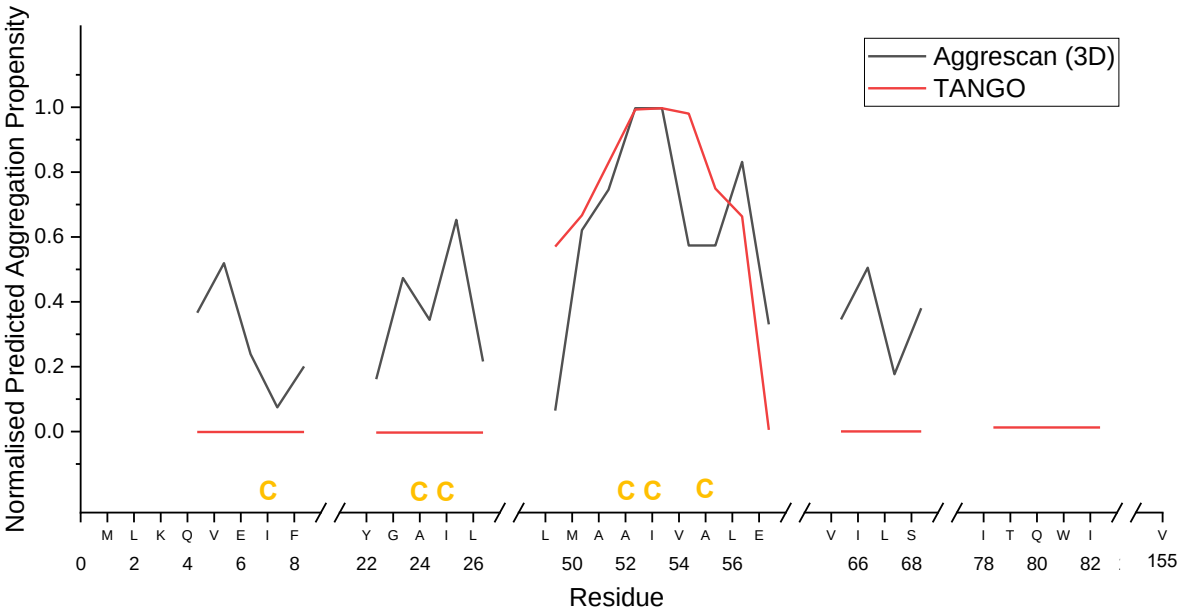


Figure 3.17: Predicted Aggregation-Prone Regions in WT RNase H*

Normalised aggregation-proneness scores for flagged aggregation-prone regions detected by both Aggrescan 3D and TANGO. Regions flagged as aggregation-prone are indicated by the algorithm used to detect them. Sites of cysteine substitution mutations used in this study are indicated with gold Cs. TANGO aggregation potential predictions outside of residue 49-57 are non-zero, but miniscule in comparison.

and a high degree of overlap in detected length and aggregation potential of the APRs (from residues 49-57) was measured (fig 3.17).

Interestingly, there was no significant difference in aggregation propensity in sites within APRs versus those outside APRs (fig 3.18A). As the APR-predictive algorithms assigned aggregation propensity scores to detected APRs, it was thought that discrepancies in aggregation propensity may be due to APRs contributing unequally to aggregation. However, no correlation was detected between normalised aggregation propensity score and aggregation propensity of a RNase H* cysteine variant found in each position at 0.25, 0.5 or 1 mg/ml protein concentration (fig 3.18B).

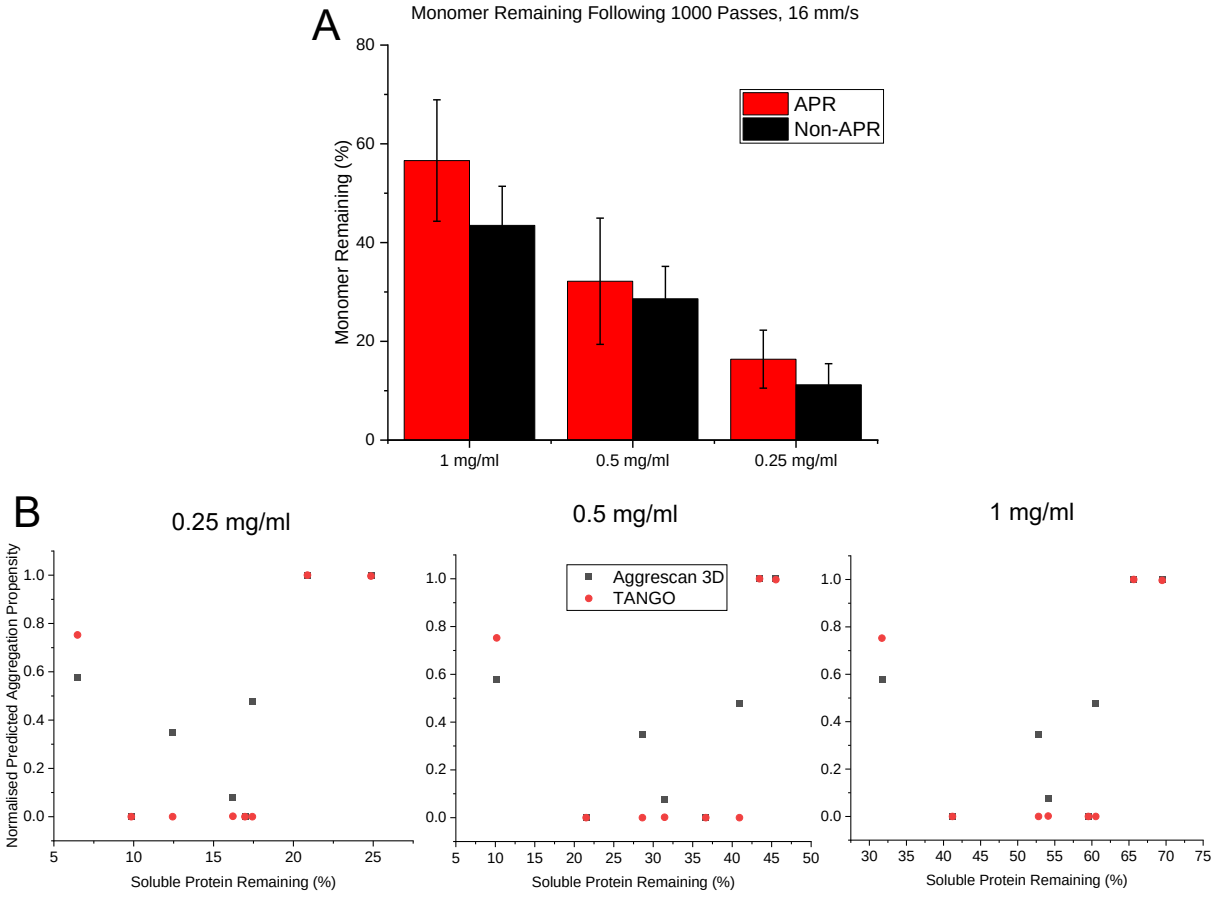


Figure 3.18: Aggregation Prone Regions do not Determine Aggregation Proneness of RNase H* Cystine Variants

A) Average aggregation propensities of variants found within and outside of predicted APRs are similar. Error bars denote standard deviation of the mean. B) Normalised aggregation propensity scores of RNase H* cysteine variant substitution sites do not correlate with aggregation propensity of those variants. Analysis by Aggrescan 3D and TANGO shown separately.

These data indicate that the varied aggregation behaviour and surface affinities of RNase H* variants cannot be solely explained by the positions of cysteine substitutions in relation to predicted APRs. As k_2/k_{-2} (the terms for bulk unfolding/folding) were shown by sensitivity analysis to be prominent in determining aggregation propensity, the effects of these cysteine mutations on folding was investigated.

3.3.2 Thermodynamic Stability Changes do not Explain Variations in Soluble Protein Loss

Although an important parameter in biopharmaceutical development, thermodynamic stability doesn't correspond to resistance to mechanical unfolding, a parameter often measured by molecular tweezer or cantilever-based experiments such as atomic force microscopy (AFM)(discussed in section 1.1.5). This is illustrated by mechanically resilient molecules such as Ubiquitin, which do not exhibit distinctly high thermodynamic stabilities (Kundu, Saha and Gangopadhyay, 2020).

Whilst it is difficult to draw a correlation based on the dataset available, it appears that flow-induced aggregation correlates best with global thermodynamic stability at 0.5 mg/ml and 1 mg/ml protein concentrations, where bulk-mediated unfolding is thought to be more prevalent than at 0.25 mg/ml (fig 3.19). This indicates that if thermodynamic stability does reduce aggregation propensity, it is because of resistance to flow-induced unfolding rather than surface-mediated infolding. However, the lack of clear correlation suggests that flow-induced aggregation should be assayed for independently, in the context of measuring a protein's resistance to flow-induced aggregation.

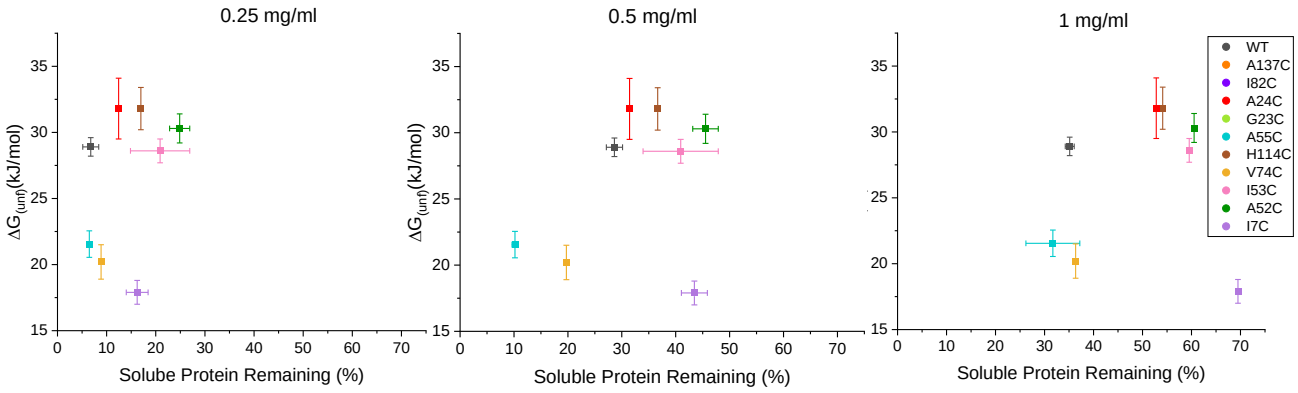


Figure 3.19: Flow-Induced Aggregation Versus Thermodynamic Stability

Each concentration value indicates the RNase H* concentration used for measuring soluble protein loss. Error values shown are standard deviations estimated for each coefficient for equilibrium unfolding data using Igor Pro, from residuals of fit (fig 3.8). Error values for soluble protein remaining are standard deviations of the mean for soluble protein remaining values (n=2).

To establish the causes for differences in k_2/k_{-2}, the rate constants governing bulk conformational change and therefore EFD-induced aggregation, a method to gather detailed structural insight into EFD-induced conformational changes is needed. Chapter 4 addresses development of this unfolding assay.

3.4 Discussion

A range of RNase H* variants each with single cysteine substitutions have been produced to provide sites for a labelling in the event of conformational change. Whilst containing similar secondary and tertiary structure content, these variants showed a variety of thermodynamic stabilities. When stressed with defined hydrodynamic forces using a bespoke flow device, these variants display a range of aggregation propensities under flow as well as inverse concentration dependencies for aggregation – indicating the presence of both surface and bulk-mediated aggregation.

Although a model has been previously developed (Willis *et al*., manuscript in preparation), and describes aggregation data well in the mAb STT as well as WT RNase H*, the model did not fit data for cysteine substitution variants as well. However, through changing individual rate constants, it was found that aggregation rate is most sensitive to change

in k_2, the constant for protein conformational change in bulk; this further validates the need for characterisation of unfolding under flow.

The differences observed in aggregation propensity between variants cannot be explained solely by changes to APRs. This may be due to the requirement of conformational change for flow-induced protein aggregation, if an APR is sequestered from solution when under a protein's native conditions. Finally, there was no clear correlation between thermodynamic stability and aggregation propensity in the RNase H* variants analysed. This is a useful conclusion, as it indicates the uniqueness of flow and the disparity between stability measurements from chemical denaturation at equilibrium and mechanical stability of a molecule under flow; measurements of thermodynamic stability usually taken in the biopharmaceutical industry may be insufficient to determine a protein's propensity for flow-induced aggregation.

Wang *et al.*, (2017) showed that global increases in thermodynamic stability of endostatin, a 20 kDa model therapeutic protein, did not result in higher colloidal stability or reduced aggregation. Small structural perturbations have been shown to contribute to aggregation more significantly than completely unfolded species, and this may be reflected in flow-induced aggregation (section 1.3.1)(Wang, 2005). This emphasises the need for an assay that is sensitive to these relatively localised changes; the next chapter of this thesis will address the development of this, and the insight it has produced.

4 Development of an Assay to Quantify Hydrodynamic Flow-Induced Protein Unfolding

4.1 Aims and Challenges

As discussed in section 1.4.3, determining whether proteins unfold under flow is of vital importance in establishing the mechanism of flow-induced aggregation; unfolding is known to frequently precede aggregation (Booth *et al.*, 1997)(Pang *et al.*, 2023). Whilst the end point of this process can be measured relatively simply through quantification of soluble protein loss (section 3.2.2), partial unfolding is often transient in nature, and therefore requires a different strategy for quantification (Ashton *et al.*, 2010).

Notably, the EFD employed in our group's work creates a highly dynamic environment which is not conducive to methods requiring molecules to be fixed, for example, crystallography-based methods and electron microscopy. This reflects the different environments encountered throughout the lifetime of a biopharmaceutical, from production to administration *in vivo*. In addition, folding timescales can be rapid; <60 μs for α-helix A and β-strand IV in WT RNase H* (Rosen *et al.*, 2015). As such, methods involving rapid detection/capture of unfolded states are required, to ensure these events are not missed.

This chapter describes how a sensitive method for assaying this process was established and optimised, eventually resulting in a final label-based assay being employed, which shows high sensitivity and detects measurable conformational change in RNase H* following exposure to either denaturant or hydrodynamic flow.

4.2 Measuring Unfolding Using a Label-Free Assay

4.2.1 Proteolysis-Based Unfolding Assay

Label-free assays were first investigated, as they avoid any alterations to structure or stability that arise from a labelling-based method. (Park and Marqusee, 2005) reported the use of *Geobacillus stearothermophilus* thermolysin as an enzyme that degrades partially or completely unfolded conformations of RNase H*. This specificity arises from

the enzyme's propensity to hydrolyse peptide bonds at hydrophobic residues, which are often sequestered from solution in a natively folded protein. Degradation products can be separated from any remaining intact protein using SDS-PAGE, and quantified using densitometry of the resulting bands.

A137C RNase H* was used to assess the viability of the assay (fig 3.1). 0.5 mg/ml (28.5 mM) RNase H* was incubated for varying times in sodium acetate buffer with varying concentrations of thermolysin (indicated in fig 4.1)(section 2.2.7.1). 50 mM NaCl and 10 mM $CaCl_2$ were used to enhance the stability and thereby facilitate activity of thermolysin; it has 4 calcium-binding sites, to which Ca^{2+} binding increases stability through electrostatic interactions and hydrogen bonds (Tajima *et al.*, 1976). The mechanism by which NaCl stabilises the protein is unknown (Inouye, Kuzuya and Tonomura, 1998). 12.5 mM EDTA was used to chelate Ca^{2+} and quench thermolysin activity.

Under quiescent conditions, longer exposure to, and higher concentrations of, thermolysin appears to produce cleavage products adding up to approximately 17.5 kDa, the molecular weight of RNase H* (fig 4.1A). Unfortunately, degradation products which add up to approximately the mass of intact thermolysin were also visible, suggesting self-degradation (fig 4.1B). This autolysis has been reported in the literature, particularly under stress conditions such as high temperature and where Ca^{2+} ions are unavailable (Vriend and Eijsink, 1993). More importantly, an increase thermolysin-mediated degradation of RNase H* was not detectable with hydrodynamic flow.

As flow-induced unfolding may be responsible for thermolysin self-degradation, a 'pulsed' approach was considered, in which thermolysin was added as rapidly as possible to RNase H* following exposure to flow, for 1 minute. When compared to samples exposed *in situ* to thermolysin for the same length of time under flow, there was no noticeable increase in RNase H* degradation products, despite increasing the pass count to 200 to maximise any detectable differences (fig 4.1C). Altering molar concentration ratio of thermolysin:RNase H* was then investigated, so that regardless of self-degradation, a sufficient concentration of thermolysin would be available to cleave any unfolded RNase H*. Despite showing what appeared to be new degradation products with a 1:1 ratio (fig 4.1D), the results proved irreproducible.

The lack of visible differences suggests that the number of molecules unfolding is too small for this relatively insensitive assay to detect. This indicates both that aggregation may arise from a relatively small number of unfolding events, and that to detect differences between these, a more sensitive assay is required.

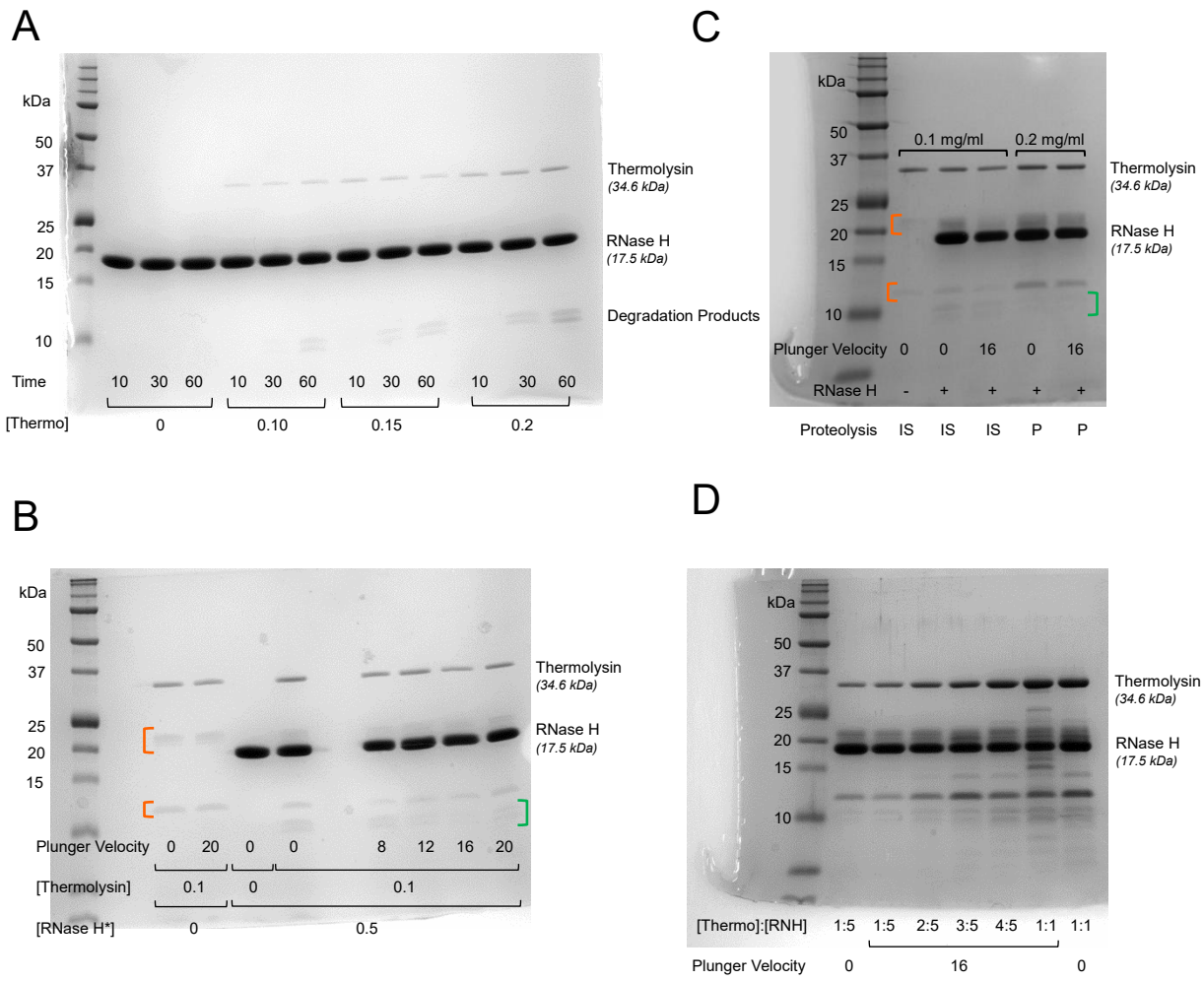


Figure 4.1: RNase H* Proteolysis Unfolding Assay

A) 0.5 mg/ml RNase H* incubated with varying thermolysin concentrations. Incubation time displayed in minutes. B) Exposure to flow at varying plunger velocities (indicated) for 100 passes each. 0 mm/s implies control, left quiescent for 10 minutes before quenching. Orange brackets indicate suggested thermolysin degradation products, green brackets indicate suggested RNase H* degradation products. C) 0.5 mg/ml RNase H* either exposed to thermolysin *in situ* (IS) in the EFD or exposed to thermolysin through a pulse (P) immediately after the EFD experiment. Samples were stressed for 200 passes or incubated at room temperature quiescent for 10 minutes D) Thermolysin pulsed into RNase H* at varying [thermolysin]:[RNase H*] ratios as indicated ([Thermo]:[RNH]), following 200 passes in the EFD or quiescent incubation at room temperature for 10 minutes. A137C RNase H* used throughout. Concentrations indicated are all in mg/ml. Plunger velocity values are in mm/s. Low-MW gels were used (table 2.3)

4.2.2 Enzyme Activity-Based Assay

WT RNase H* specifically degrades RNA in DNA/RNA hybrid structures. As enzyme structure is closely tied to function, perturbations resulting from hydrodynamic flow would likely affect RNase H* enzyme activity (Fischer, 1894). This would be a more direct and sensitive method to detect conformational change, as it removes the dependence on activity of a second macromolecule.

Enzyme activity was measured quantitatively using a short DNA/RNA substrate, in which fluorescein is tethered to the 3′ end of the RNA, and a 4-((4-(dimethylamino)phenyl)azo)benzoic acid succinimidyl ester (DABCYL) quencher is tethered to the 5′ end of the DNA strand (termed “molecular beacon” hereon)(section 2.2.7.2). When the RNA is cleaved by RNase H* activity, the fluorescein molecule is liberated from the quenching effect and an increase in fluorescence emission intensity at 515 nm should be measurable (Rizzo *et al.*, 2002). The experiment was conducted in arginine succinate buffer, to minimise the amount of protein loss due to aggregation (discussed in section 4.3.2).

Continuous fluorescence measurements of WT RNase H* with 25 nM molecular beacon in reaction buffer (section 2.1.4) at room temperature show increasing intensity with time, demonstrating the ability of the assay to measure RNase H* enzyme activity (fig 4.2A). 5 nM RNase H* was used in future experiments, as it retains linear activity over 20 minutes, implying the molecular beacon remains in excess over this timeframe. Under a number of strain rates and pass counts in the EFD, the molecular beacon does not show a peak at 515 nm in the absence of RNase H*, indicating cleavage is a result of enzyme activity, not flow (fig 4.2B). The peak at 495 nm is likely elastic Rayleigh scattering from the excitation beam ($\lambda_{ex} = 495$ nm).

When RNase H* was exposed to flow with the molecular beacon *in situ*, the change in fluorescence emission intensity (515 nm)(Δfluorescence intensity) appeared considerably different between samples that had been exposed to flow or left quiescent (fig 4.2Ci). Unfortunately, the results were difficult to reproduce (fig 4.2Cii). Finally, RNase H* was incubated with the molecular beacon following exposure to flow, with the aim of capturing the effects of structural perturbations that remain present over the order of seconds. This method was also not reproducible, and differences in Δfluorescence intensity were less often detected.

This may be in part due to difficulties in keeping the dead-time equal between samples; if RNase H* structure is only very briefly perturbed, effects on enzyme activity may be missed in the dead-time. Use of a rapidly reactive covalent labelling molecule *in situ* that is specific to the unfolded state may be an alternative that circumvents this issue.

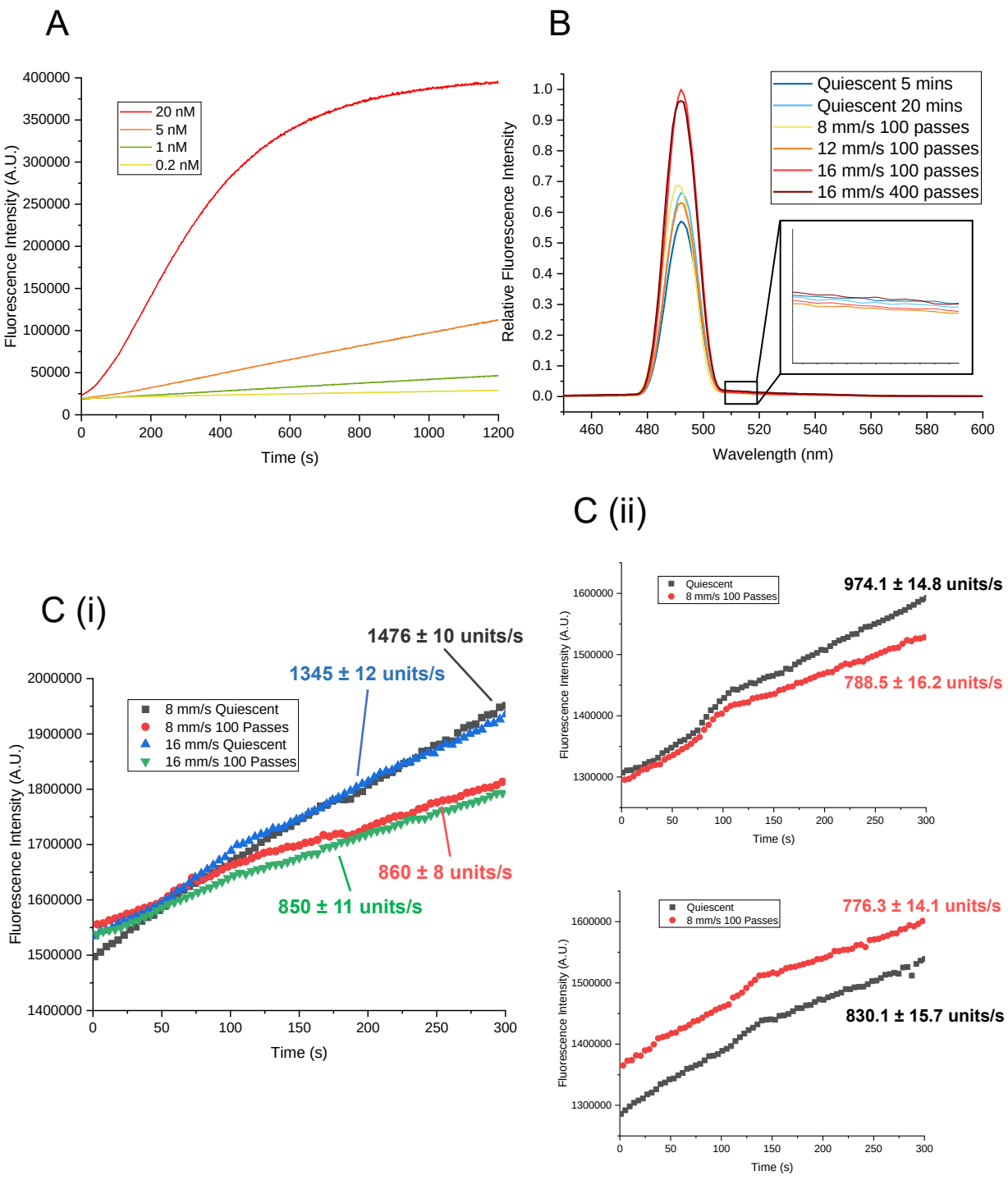


Figure 4.2: RNase H* Enzyme Activity Assay

A) RNase H* enzyme activity as a function of enzyme concentration. 25 nM molecular beacon used. RNase H* concentrations used as indicated in figure legend. Measurements taken every second. B) Hydrodynamic flow does not affect fluorescence intensity in the absence of RNase H*. 25 nM molecular beacon analysed with and without hydrodynamic flow. λ_{ex} = 495 nm. C) Enzyme activity (indicated by change in fluorescence intensity units) modulated by hydrodynamic flow. 250 nM molecular beacon used. Units refer to arbitrary fluorescence intensity units. i) Initial results. ii) Separate biological repeats. For all time-based measurements, λ_{ex} = 495 nm, λ_{em} = 515 nm. mm/s values indicate EFD plunger velocity.

4.3 Measuring Unfolding Using a Label-based Assay

Suitable methods could include use of a detectable fluorescent probe, which, with sufficient reactivity and in a molecular excess to RNase H*, could rapidly capture unfolded states. The panel of 10 cysteine-containing RNase H* variants provide the unique opportunity to probe unfolding in specific regions of the protein using a thiol-reactive label, and the varied stabilities and aggregation propensities may be informative in determining which features correspond to flow-induced unfolding.

Previously, Dobson *et al.*, (2017) investigated protein unfolding using a covalent fluorescent label. Bovine serum albumin (BSA) was labelled using 5-((2-((iodoacetyl)amino)ethyl)amino)naphthalene-1-sulfonic acid (IAEDANS), a thiol-reactive label for surface-exposed cysteine residues that may be indicative of unfolding. The high reactivity of cysteine with a range of groups, as well as being the least naturally solvent-exposed residue, makes it a good labelling target (Marino and Gladyshev, 2010). At pH 8, IAEDANS reacts with thiol-containing groups via nucleophilic substitution (fig 4.3A), to covalently AEDANS-label the species containing the thiol. Benefits of using IAEDANS as the probe include its distinct absorption and fluorescence profile (minimally overlapping with that of intrinsic protein fluorescence)(fig 4.3B), storage lifetime and reactivity/fluorescence at a practical pH range (Hudson and Weber, 1973).

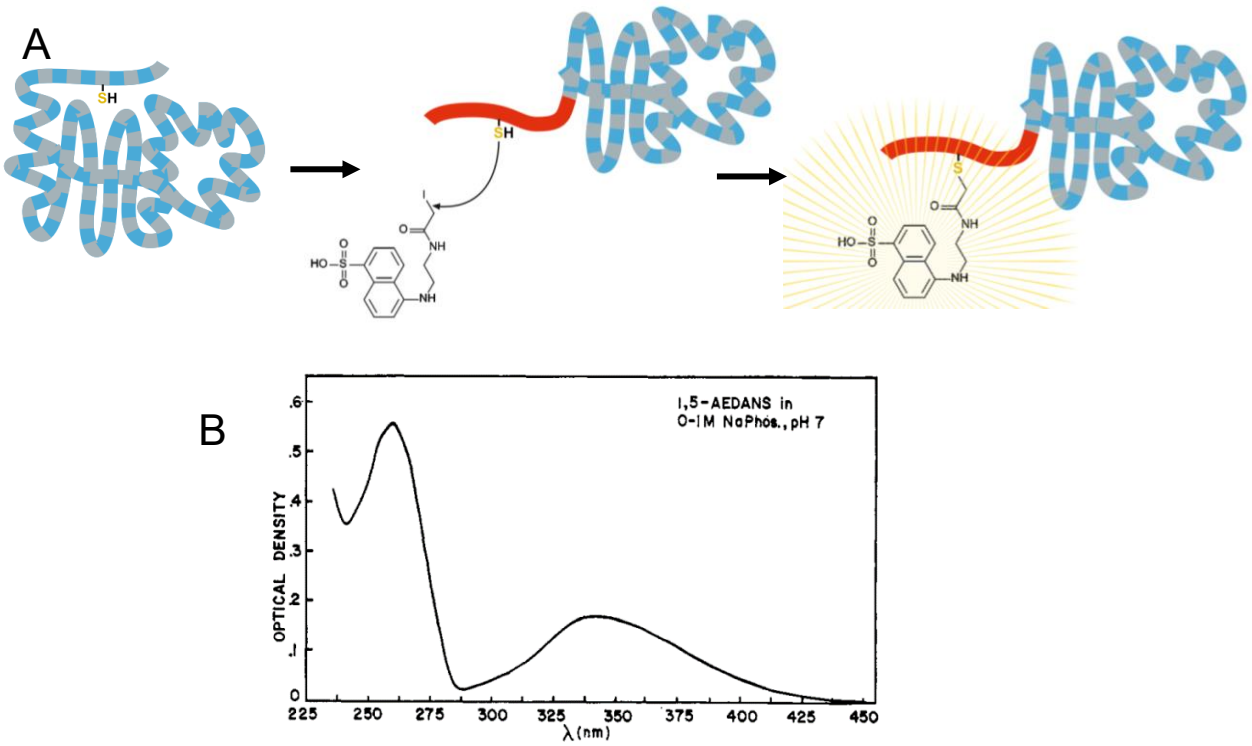


Figure 4.3: IAEDANS Labelling Assay for Unfolding

A) Schematic illustrating IAEDANS labelling for detection of unfolding. B) IAEDANS Absorption spectrum, in 0.1 M sodium phosphate, pH 7. Taken from (Hudson and Weber, 1973).

4.3.1 IAEDANS Reactivity

It is necessary to conduct the assay under EX1 kinetics, whereby the rate of labelling is limited by unfolding of RNase H*, rather than reactivity of the label. This ensures unfolded states will be maximally detected.

Reaction of the thiol group in a cysteine from thioredoxin with iodoacetamide at pH 8 is approximately k_{app} = 107.8 M^{-1} s^{-1} (in 100 mM Tris-HCl, 2 mM EDTA, 25°C)(Kallis and Holmgren, 1980). Whilst this remains an estimate due to differences in reactivity from protein to protein, initial maximum rate of reaction between IAEDANS and RNase H* can be calculated as 4.38 μM/s according to the equation:

$$r = k_2[A][B]$$

Equation 4.1

Where r = initial rate constant, k_2 is the apparent bimolecular rate constant of iodoacetamide with a sulfhydryl group at pH 8, [A] and [B] are the initial concentrations of RNase H* and IAEDANS (0.0285 mM and 1.425 mM, respectively).

This initial rate of reaction should be sufficient to capture unfolded states, provided these have a lifetime in the order of milliseconds or seconds. If all RNase H* molecules were unfolded completely, the reaction would be over within 6 seconds. However, the maximum unfolding per second detected in an RNase H* variant thus far is 19 nM/s. As IAEDANS labelling of cysteines proceeds optimally between around pH 8.5, all IAEDANS labelling experiments were conducted at pH 8, which was also deemed sufficiently far from the RNase H* pI of 8.83 (computed from primary sequence using ExPASy, Swiss Institute of Biophysics)(Hudson and Weber, 1973)(Poole, 2015).

4.3.2 Inhibition of Aggregation with L-arginine

The hypothesis of this work is that flow induces partial unfolding. if this unfolding exposes regions with relatively high aggregation propensity, aggregation through protein-protein interactions will occur. In order to measure EFD-induced conformational changes most effectively, loss of soluble sample due to aggregation must be limited, to maximise the number of molecules that can unfold from the native state in solution, and ensure transiently unfolded states can be labelled.

Although the molecular mechanism is a subject of contention, L-arginine is broadly considered a potent inhibitor of protein aggregation. Suggestions exist that arginine sterically hinders protein-protein interactions through ‘molecular crowding’ effects (Baynes, Wang and Trout, 2005). Additionally, weak (approx. 2.9 ± 1.4 kcal/mol) π-cation interactions between the guanidinium group in arginine and aromatic amino acids (which are frequently cited as components of APRs) have been proposed as protective against aggregation through APR contact and/or facilitative of disaggregation (Gallivan and Dougherty, 1999)(Dougherty, 1996)(Tsumoto *et al.*, 2004).

The RNase H* cysteine substitution variants produced and characterised in chapter 3 were exposed to hydrodynamic flow in the EFD for between 100 and 1000 passes at 16 mm/s, as described in section 3.2.2, but in buffer containing 125 mM L-arginine rather than phosphate. Additionally, RNase H* was only analysed at 0.5 mg/ml, as it was thought to aggregate through both bulk and surface pathways at this concentration (discussed in section 3.2.2).

When compared to samples stressed in the same way in phosphate buffer, all arginine-buffered RNase H* variants aggregated to a lesser extent (fig 4.4B). The degree of protection is variable between variants, although all show increased soluble protein remaining compared to samples in phosphate buffer (64.9% mean soluble protein remaining versus 16.6% in phosphate buffer). As such, a buffer containing 125 mM L-arginine (section 2.1.4) has been the choice solvent for unfolding work, as it allows loss of soluble protein due to aggregation to be reduced in said experiments.

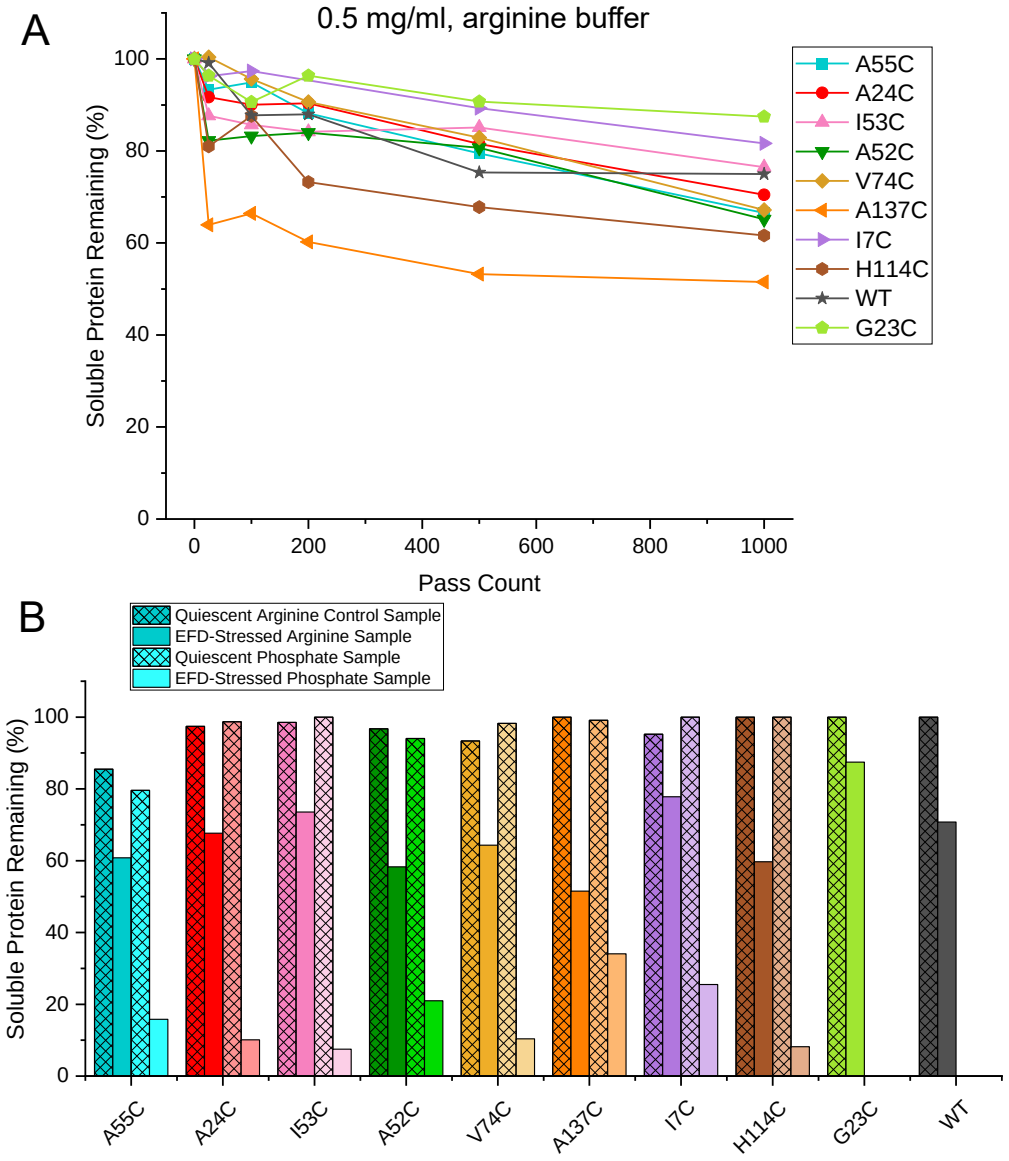


Figure 4.4: Arginine Succinate Protects Against Flow-Induced Loss of Soluble Protein

Soluble protein remaining in IAEDANS labelling experiments. 0.5 mg/ml RNase H* samples stressed at 16 mm/s plunger velocity in the EFD. A) Soluble protein remaining relative to the mean quiescent control samples, in arginine succinate buffer (section 2.1.4). B) RNase H* variants stressed in arginine (dark colours) or phosphate (light colours) buffer. Soluble protein remaining shown relative to quiescent control samples measured at t = 1 minute. Control samples are measured at t = 50 minutes. EFD-stressed samples were measured after 1000 passes.

Thermodynamic stability of WT RNase H* remains largely unchanged in the presence of arginine buffer compared to phosphate buffer at the same pH (fig 4.5A), whilst a slight decrease in thermodynamic stability was reported in a test protein with different structure to RNase H*, I27 (provided by a colleague, Sophie Cussons)(fig 4.5B). This is supported by findings in the literature, which has notably also reported a decrease in T_m values with other proteins, and arginine concentrations ranging from 100-750 mM (Taneja and Ahmad, 1994)(Arakawa and Tsumoto, 2003). Additionally, (Arakawa *et al.*, 2007) revealed that unlike guanidine, an alternative aggregation inhibitor, arginine did not preferentially interact with the surface of native state lysozyme over water. This indicates the protective effects of arginine may not be due to stabilisation of a folded, native state; supporting the aforementioned theories that arginine modulates aggregation rather than folding.

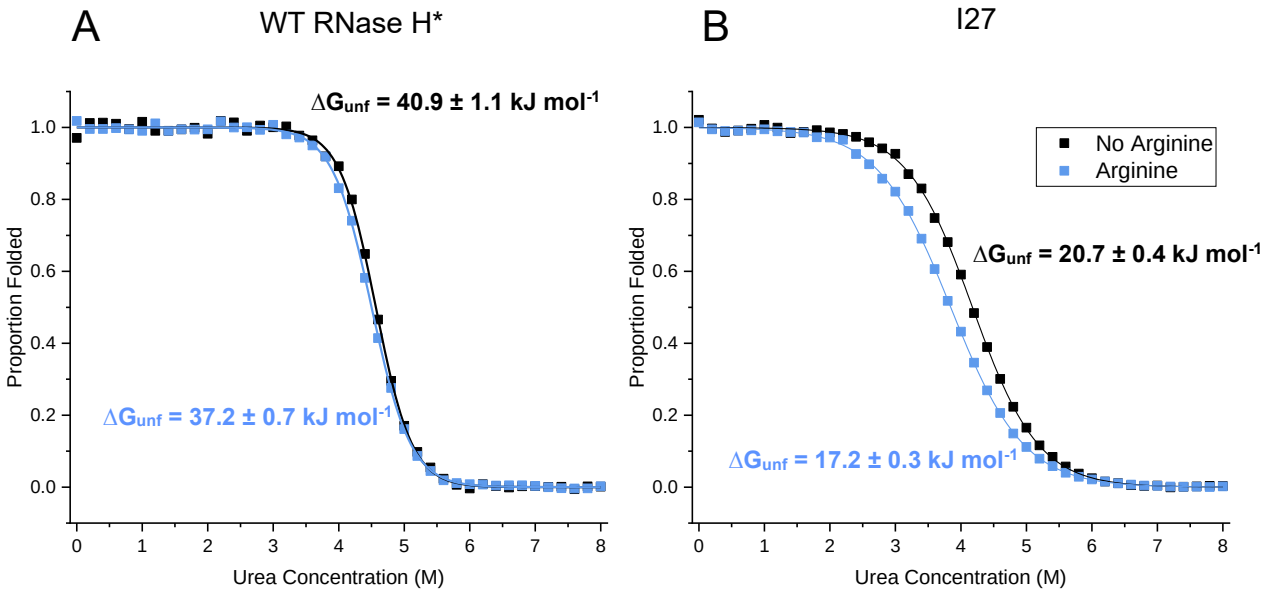


Figure 4.5: Effects of Arginine Succinate Buffer on Thermodynamic Stability

Equilibrium unfolding of 0.1 mg/ml samples measured by intrinsic Trp fluorescence. A) WT RNase H*. B) I27 monomer, provided by a colleague, Sophie Cawood. λ_{ex} = 280 nm, λ_{em} = 320 nm. All samples at pH 6.0, 25°C and buffered in 20 mM K_2HPO_4, 50 mM KCl. 125 mM L-arginine, 20 mM Na succinate included for arginine-containing samples. Error values shown are standard deviations estimated for each coefficient using Igor Pro, from residuals of datapoints to the fit.

4.4 Development of an IAEDANS-Based Labelling Assay

A137C RNase H* in arginine succinate buffer was used to optimise the assay. This is because residue 137 is in the region that folds last according to the Hu *et al*. (2013) schematic, and is relatively peripheral in the structure, suggesting it may unfold more readily under flow than one of the more central regions (section 3.1.1, fig 3.1). In addition, early characterisation of A137C showed a reduced ΔG_{unf} compared to WT ($\Delta\Delta G_{unf}$ = 8.6 ±1.6 kJ/mol, section 3.1.5, fig 3.8)(Zhang *et al.*, 2023).

To facilitate rapid capture of unfolded states, it was important to ensure that IAEDANS concentration was not a limiting factor. A 50x final molar excess of IAEDANS in arginine succinate buffer was added to each sample before stressing in the EFD (section 2.2.7.4). Following exposure to flow, the IAEDANS labelling reaction was quenched by addition of an excess of dithiothreitol (DTT), before the sample was analysed for changes in fluorescence intensity and protein concentration.

4.4.1 Densitometry for IAEDANS Labelling Quantification

Initially, SDS-PAGE followed by densitometry of bands was adopted as the method to quantify protein unfolding, as it enabled quantification of both protein concentration and fluorescence, using a transilluminator with white light/UV capabilities and quantification of band pixel intensities from images (sections 2.2.3.1 and 2.2.3.2). This involved electrophoresing samples and imaging the resulting gels for quantification of bands corresponding to RNase H* molecular weight.

This densitometry method was able to yield results quickly but had several limitations. To prevent fluorescence intensity from being an arbitrary number that is only relative to non-stressed sample, a urea-unfolded ‘100%’ labelled sample was required as a benchmark for relative unfolding. RNase H* of the variant under investigation was prepared to the same final concentration and conditions as the experimental samples, but using 6.5 M urea at the same pH as the buffer used, rather than the original buffer. This was incubated overnight at room temperature, and was deemed sufficient to unfold all RNase H* variants featured in this work. Although these samples appear unfolded, evidence (presented in section 5.1.1) is required to determine if these urea-unfolded controls are 100% labelled in absolute terms.

As in previous work investigating aggregation propensity (section 3.2.2), 16 mm/s plunger velocity was used in unfolding experiments featuring RNase H*, to ensure a sufficient, detectable response. Upon stressing A137C RNase H* in the EFD for various pass counts at this speed, fluorescent labelling was analysed by densitometry (fig 4.6). Labelling is increased when compared to a quiescent control and a trend towards increased labelling with higher strain rates can be distinguished. This indicates both that hydrodynamic flow may induce unfolding, and that quantification of this through densitometry may be a viable option.

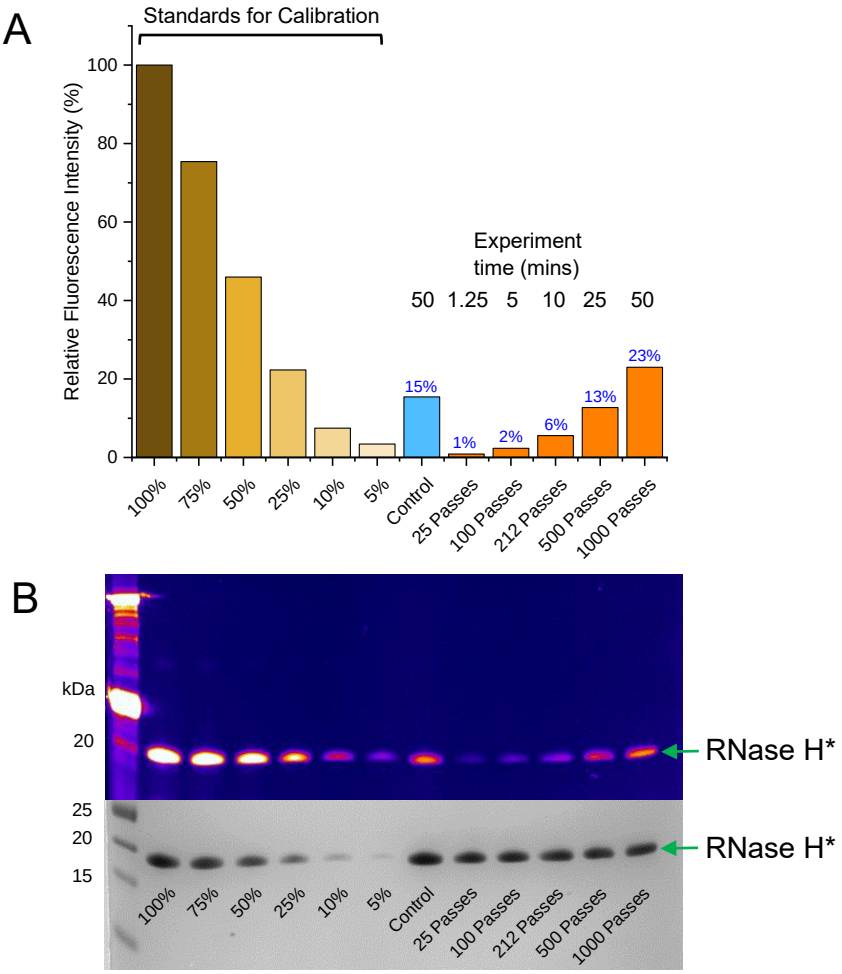


Figure 4.6: A137C RNase H* IAEDANS Labelling Quantified by Densitometry

0.5 mg/ml A137C RNase H* stressed in the EFD at 16 mm/s, for varying pass counts, as indicated, in arginine succinate buffer. A) Quantification of labelling from densitometry of imaged SDS-PAGE gel shown below. Flow-stressed sample fluorescence intensity values are shown above bars relative to the undiluted '100%' urea-unfolded sample. B) SDS-PAGE gel showing bands UV-imaged before Coomassie blue stain (above) and after stain (below). Image enhanced for visibility using FIJI image processing package.

Despite optimisations such as varying gel composition to contain higher concentrations of acrylamide/bis acrylamide, and removal of unreacted IAEDANS using Zeba™ centrifuge filtration columns (section 2.2.7.3), it was noted that fluorescent bands are often difficult to see with the naked eye, and gel-based images frequently require adjustment to increase clarity for presentation. This suggests that the fluorescence imaging system may be operating near its lower limit for fluorescence detection, and that human error may be introduced during quantification steps.

To understand the error associated with quantification, a standard curve was prepared. A dilution series from 0 – 100x was prepared from a stock of 0.5 mg/ml A137C RNase H*, incubated overnight in urea in the presence of IAEDANS, as described above. Each sample was electrophoresed and analysed by densitometry, as described above. Expected fluorescence intensity was plotted based on the dilution factor of each sample, and fluorescence intensity measured by densitometry was plotted against this. This allows the point(s) at which the results deviate from the expected value to be identified (fig 4.7A).

Below 7.5% measured fluorescence intensity (relative to the undiluted '100%' standard), the deviation from a straight line was deemed too severe for points below this value to be sufficiently accurate. Removal of the points below 7.5% improved the R^2 from 0.937 to 0.993 (fig 4.7B). This trend suggests that the less fluorescent bands may be quantified with lower accuracy, because they are at the lower detection limit for the chemiluminescence gel imaging system, and/or the contribution from background fluorescence of the gel stifles the fluorescence detected from the bands. Consequently, the assay requires IAEDANS labelling of a sample to give fluorescence intensity >7.5% of the undiluted urea-unfolded standard, to be deemed sufficiently accurate.

These observations steered experiments towards conditions that promote labelling of ≥7.5% relative to the internal standard. This was deemed a considerable limitation, as the heterogeneity observed in aggregation propensities between cysteine variants (section 3.2.2) indicated flow-induced unfolding would also vary, if the latter was the cause of the former. Thus, an alternative capable of accurately measuring lower extents of labelling was investigated.

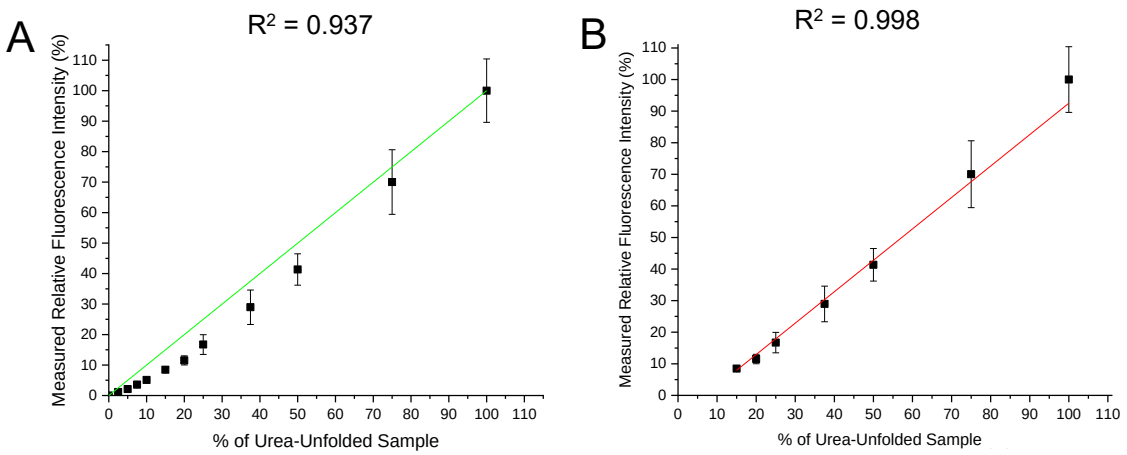


Figure 4.7: Calibration Curve for Labelling Quantification by Densitometry

A) IAEDANS labelling measured by UV fluorescence. Green guideline for comparison of results against a linear dependence fit, assuming no lower limit of detection. Error is standard deviation of mean, n = 2. B) Exclusion of less visibly fluorescent points improves linearity of measured fluorescence intensity with increasing sample concentration. Red line indicates a linear fit.

4.4.2 HPLC-Based IAEDANS Labelling Assay

The ability to accurately measure low levels of labelling is important for this work, as it not only expands the breadth of possible experiments, but also allows investigation of RNase H* variants that are less susceptible to IAEDANS labelling.

Quantification by high-pressure liquid chromatography (HPLC) allows more accurate analysis of IAEDANS-labelled samples; automation of processes such as sample injection and measurement in a microfluidic system, rather than on a gel, reduces error and increases assay throughput. The Shimadzu HPLC system used in this work operates both an absorbance detector and photodiode array, as well as a fluorescence detector and lamp. This means both absorbance at a range of wavelengths and fluorescence can be measured in the same sample simultaneously.

To examine the efficacy of this quantification method, different RNase H* cysteine variants were exposed to hydrodynamic flow in the presence of IAEDANS as described in section 4.4.1, before quenching with 0.5 M DTT. Following quenching and buffer exchange into 0.1% trifluoroacetic acid (TFA), samples were analysed using Reverse-Phase-HPLC (RP-HPLC)(section 2.2.5.2). 0.1% TFA is pH 2.1 and ensures a denaturing environment for RNase H*, allowing detection of fluorescent labels with minimal interference from nearby side chains.

In RP-HPLC, RNase H* interacts with the column through hydrophobic interactions which can be disrupted when a gradient of strongly hydrophobic acetonitrile is introduced, allowing RNase H* to be detected at a predictable position in the resulting chromatogram (fig 4.8).

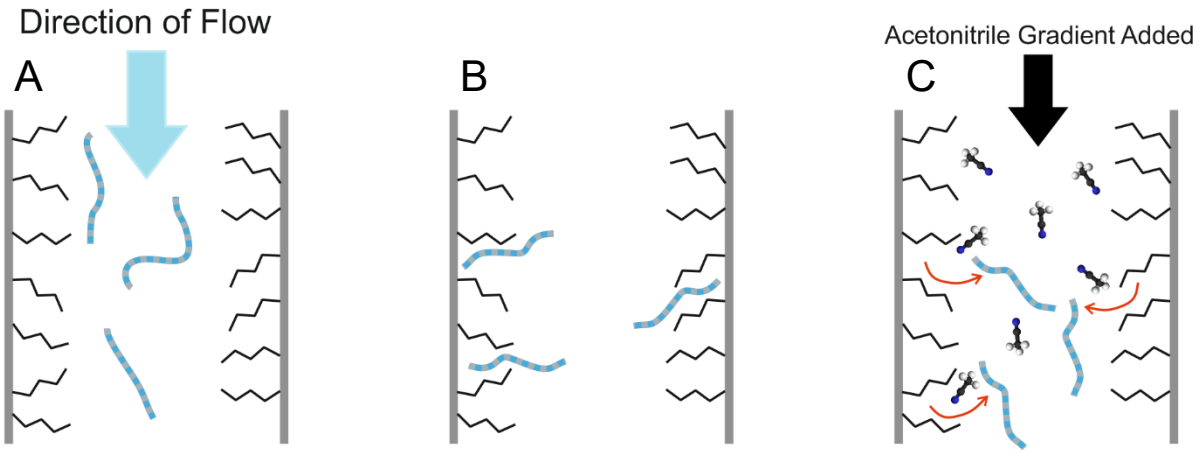


Figure 4.8: Reverse-Phase Chromatography of RNase H*

Schematic illustrates A) injection of RNase H* (dashed blue and grey) onto the column, B) Interaction with hydrophobic butyl groups in the stationary phase, C) Elution of RNase H* through addition of an acetonitrile gradient.

As retention time is not linearly associated with hydrophobicity, a calibration curve is not practical for identifying RNase H*. Therefore, known controls containing only RNase H* and only IAEDANS were analysed. Elution of RNase H* alone shows that RNase H* elutes at approximately 9.2 minutes (fig 4.9A). None of the cysteine variants since analysed have appeared to deviate significantly from this. Elution of IAEDANS alone gives a peak at approximately 3.2 minutes, as expected for a smaller molecule with far fewer hydrophobic interactions. However, IAEDANS that has been quenched with DTT (as occurs in the presence of RNase H*) elutes later (clearly visible in fig 4.11). This may be due to the quenching with DTT in the RNase H* sample, causing removal of the iodine and an increase in hydrophobicity.

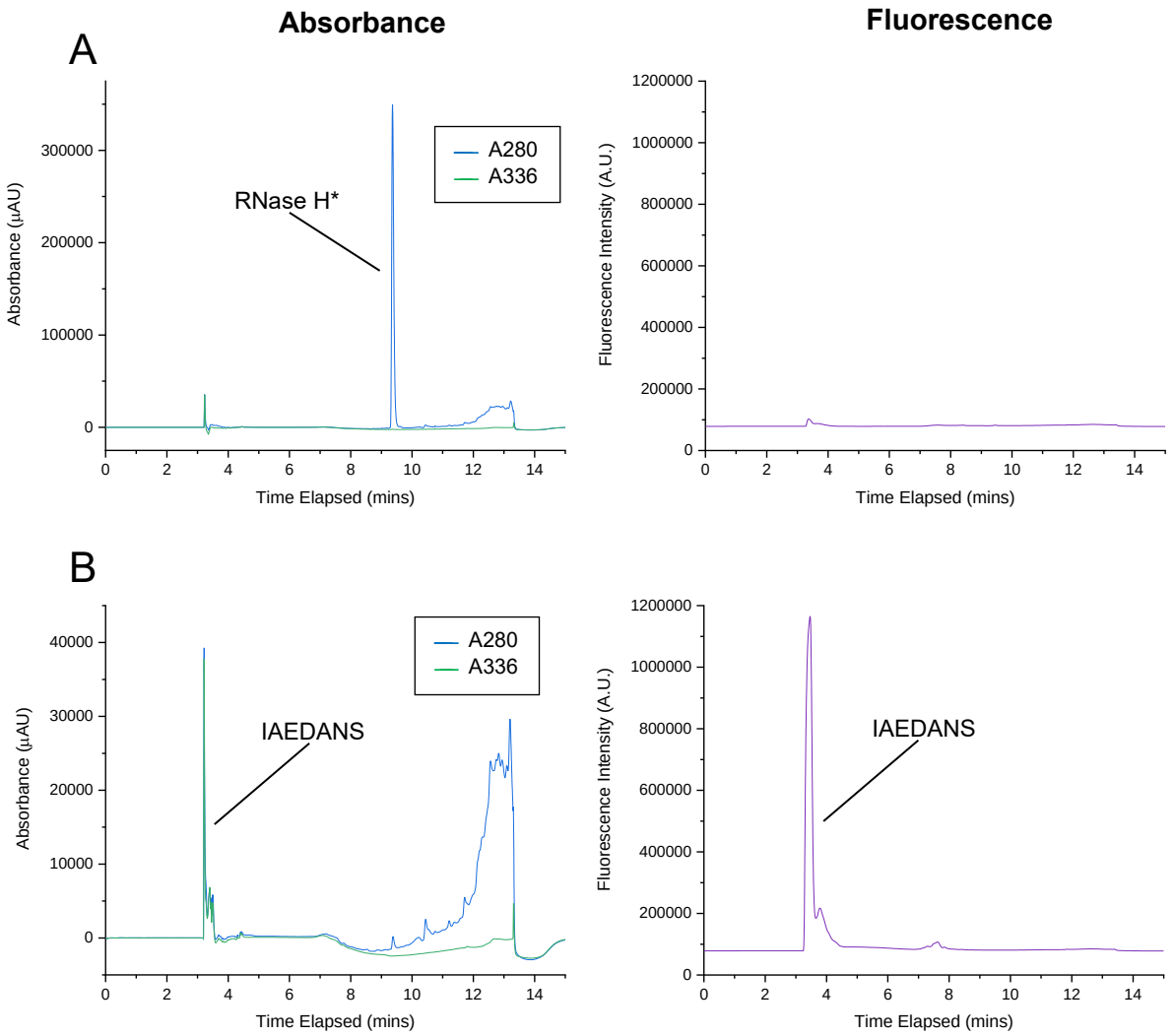


Figure 4.9: RNase H* and IAEDANS Reverse-Phase Chromatography Elution Profiles

HPLC-produced chromatograms showing absorbance and fluorescence elution profiles of A) 28.52 μM (0.5 mg/ml) A137C RNase H* alone and B) 1.43 mM IAEDANS alone. Both samples were in arginine succinate buffer. 50 μl RNase H*, 0.5 μl IAEDANS solution injected per sample. For fluorescence measurements, λ_{ex} = 336 nm, λ_{em} = 490 nm.

4.4.3 Absolute Labelling Measurements by HPLC

The capacity for the HPLC photodiode array used to detect absorbance can measure multiple wavelengths simultaneously (section 2.2.5.2). In principle, this allows quantification of the contents in a peak by analysis of the absorbance properties resulting from the reacted AEDANS and RNase H*, as the molar extinction coefficients of both are known. This was applied to labelling of RNase H* cysteine variants with AEDANS.

AEDANS absorbs significantly at 280 nm, as well its maximum at 336 nm, meaning the peak corresponding to labelled and unlabelled RNase H* contains an absorbance contribution at 280 nm, from the AEDANS label. To calculate the concentration of RNase H* present in the peak, this absorbance contribution must be accounted for, according to the following equation:

$$[RNase\ H^*] = \frac{A_{280}^{RNase\ H^*} - (A_{336}^{IAEDANS} \times 0.57)}{\varepsilon_{280}^{RNase\ H^*}}$$

Equation 4.2

Where A is absorbance at a given wavelength, ε is molar extinction coefficient at a given wavelength and 0.57 corresponds to a correction factor, A_{280}/A_{max} of free IAEDANS (Kim *et al.*, 2008).

The proportion of RNase H* molecules that are labelled can then be calculated using the following equation:

$$Proportion\ RNase\ H\ Labelled = \frac{A_{336}^{IAEDANS}}{\varepsilon_{336}^{IAEDANS} \times [RNase\ H^*]}$$

Equation 4.3

4.4.4 Absorbance is Not Sufficient for Accurate Detection of IAEDANS Labelling

To measure the sensitivity of the assay in measuring AEDANS-labelled RNase H*, a cysteine-containing RNase H* variant (A24C in this case) was unfolded with urea, in the presence of IAEDANS (as described in section 4.2.1) and analysed for absorbance using the reverse-phase HPLC method. The elution profile shows a peak at the expected retention time of ~9.2 mins for RNase H*, which contains both unlabelled and AEDANS-labelled RNase H*; the hydrophobicity increase upon AEDANS labelling should not be significant enough to produce a separate peak (fig 4.10).

In all samples, only minor additional peaks were detected (visible in fig 4.10A), indicating the 0.1% TFA running buffer and pre-injection protocol was sufficient to solubilise the majority of sample for detection (Burra and Thakur, 2016).

As some of the cysteine variants were expected to be lowly labelled, it was necessary to assess the sensitivity of absorbance measurements. Consequently, it was found that IAEDANS absorbance at 336 nm was below the limit for accurate measurement using the Shimadzu SPD-M40 photodiode array (PDA), when less than 50% of the sample was labelled. This is reflected by the lack of linearity when a range of different volumes of urea-unfolded A24C RNase H* with IAEDANS (prepared as described in section 4.4.1) were measured and absorbance was plotted against sample labelling % equivalent (fig 4.10B). Volume loaded was modulated by instructing the auto-sampler injection volume.

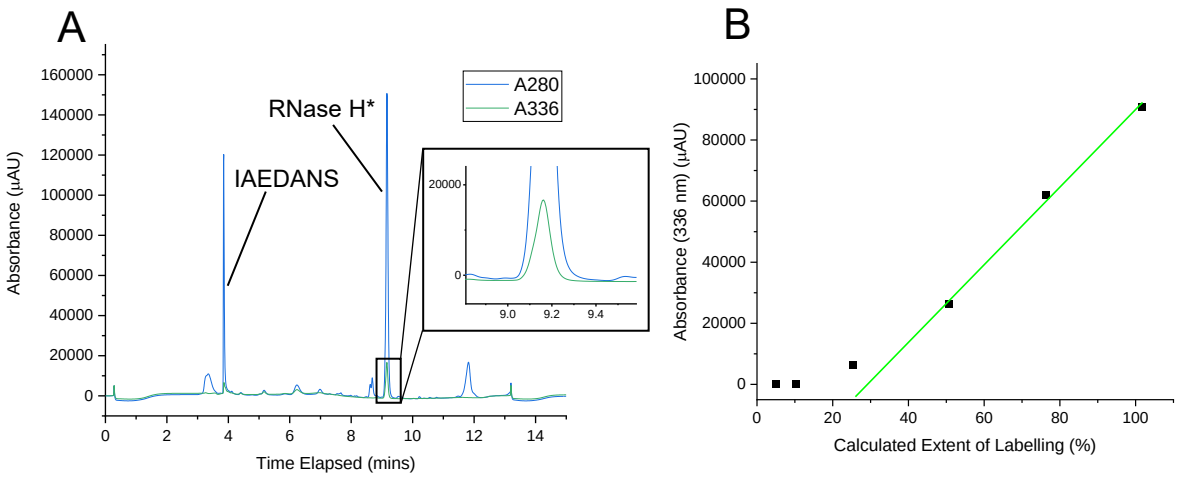


Figure 4.10: A24C RNase H* Labelling Measured by Absorbance

14.26 μM (0.25 mg/ml) A24C RNase H* prepared in 6.5 M urea (pH 8). A) 100 μl A24C RNase H* loaded. B) standard curve with 5-100 μl A24C RNase H* loaded, extent of labelling calculated using the most absorbent sample (100 μl injection) and given relative to that sample.

4.4.5 Measuring Unfolding Using Fluorescence

As fluorescence is generally more sensitive than absorbance, it was decided that despite the appeal of directly quantifying extent of labelling through absorbance, measuring labelling through fluorescence was more appropriate for detecting labelling in samples that were labelled to a low extent. The peak area around 9.2 mins elution time gives a measure for relative fluorescent labelling.

Fig 4.11 shows that the fluorescence intensity of unfolded and AEDANS-labelled A24C RNase H* (prepared and loaded as described above) is directly proportional to the volume injected, from 100 µl down to 0.1 µl (0.1% - 100% labelled sample equivalent). This shows the drastic improvement in sensitivity using fluorescence compared to absorbance to measure labelling.

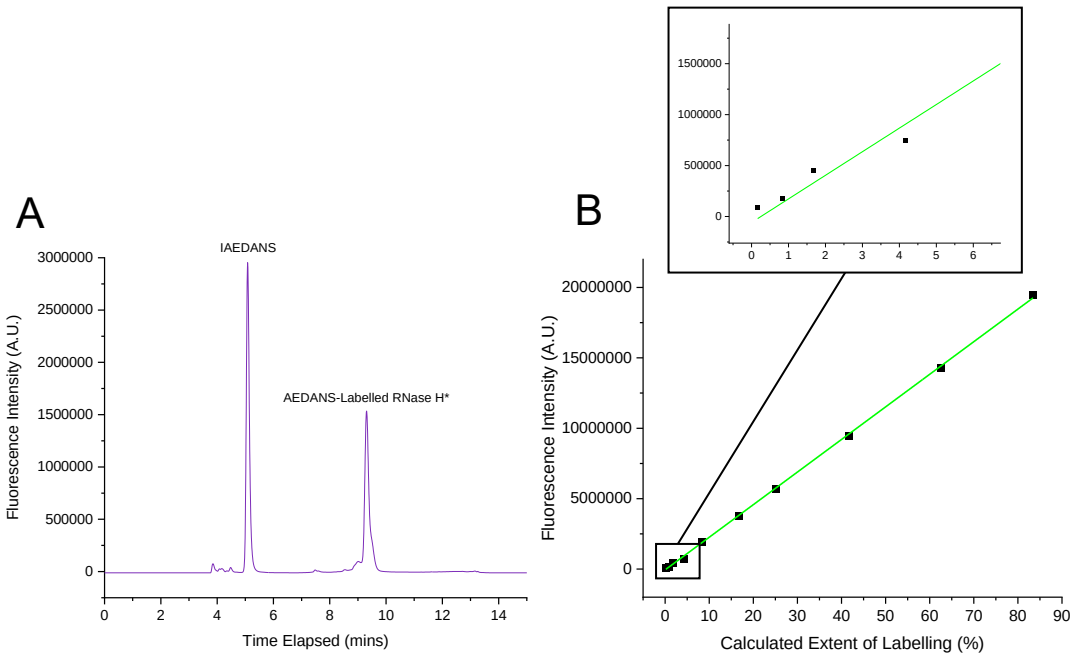


Figure 4.11: Low Labelling can be Detected Effectively with Fluorescence

Labelling in a urea-unfolded RNase H* cysteine variant (A24C) is readily detectable at low concentrations using fluorescence. 0.25 mg/ml sample loaded (14.25 mM). A) Chromatogram resulting from 50 µl A24C RNase H* loaded. B) Standard curve including 0.1 – 50 µl RNase H* loaded, extent of labelling calculated using the most fluorescent sample (100 µl injection) and given relative to that sample. λ_{ex} = 336 nm, λ_{em} = 490 nm.

4.4.6 Interpolation of Absorbance Values from a Standard Curve

In order to quantify IAEDANS labelling from fluorescence intensity values using the equations presented in section 4.4.3, it was necessary to interpolate A336 from a standard curve of fluorescence vs A336. IAEDANS was measured alone in water, however figure 4.12A shows that fluorescence intensity relative to A336 is much greater in RNase H*-AEDANS. Therefore, for interpolating A336 values from fluorescence intensity, urea-unfolded G23C RNase H* was used to produce a standard curve for interpolation (fig 4.12B).

Subsequently, A336 values interpolated from this standard curve could be used with A280 values (measured directly, due to protein content remaining sufficiently high for accurate detection) to calculate extent of sample labelling, by applying equations 4.2 and 4.3. This use of the relative high sensitivity of fluorescence allowed quantification of labelling in a broader range of labelled samples.

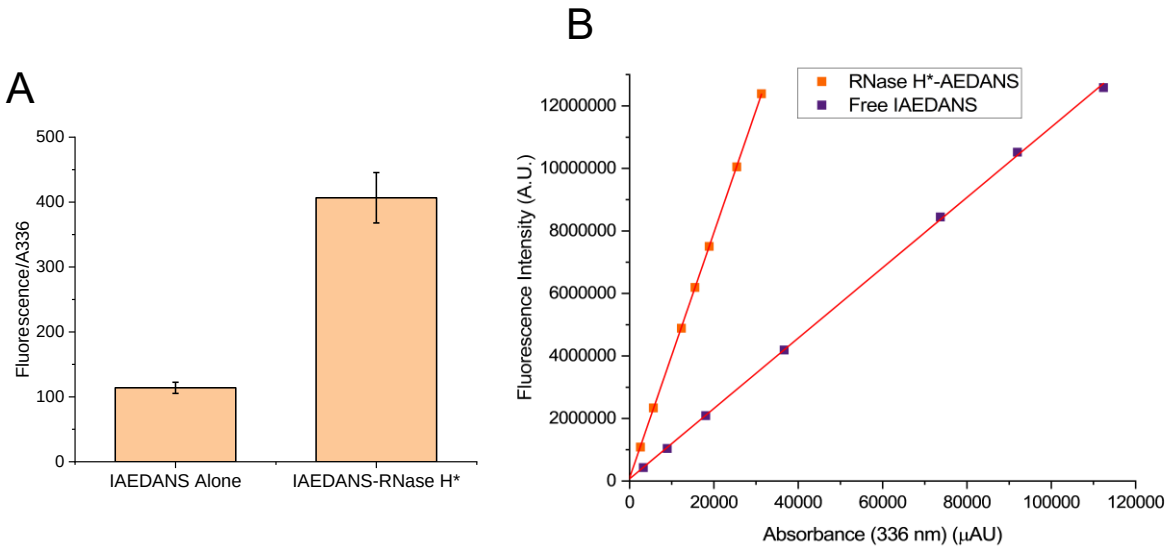


Figure 4.12: Free IAEDANS vs RNase H*-AEDANS Standard Curves

Values denote integrated peak areas from HPLC chromatograms. A) Average ratio of fluorescence (A.U.) to A336 (µAU) ratio across all sample injections analysed in (B) is significantly greater in RNase H*-AEDANS than free IAEDANS. Error bars denote standard deviation of the mean of all sample injections. B) Standard curves of fluorescence/A336 used for interpolation of A336 values from fluorescence intensity values. Samples volumes of 1-50 µl were loaded. All samples were at 0.5 mg/ml (28.5 µM) in urea unfolding buffer.

4.4.7 HPLC-Based Quantification of IAEDANS-Labelling Compared to a Densitometry-Based Method

To assess the comparability of HPLC-based detection of labelling to the original densitometry-based method, 0.5 mg/ml A137C RNase H* was exposed to flow in arginine succinate buffer (as described in section 4.4.1) at 16 mm/s for varying pass counts. Labelling of the sample was analysed using densitometry and HPLC, with calculated extent of labelling shown in figure 4.13. Both methods produce a similar trend towards greater labelling with higher pass count, as well as showing greater fluorescence in the 1000 passes sample compared to the control. The quiescent control can only be fairly compared to the 1000 passes sample, as these are the only two that are exposed to

IAEDANS for the same amount of time (50 minutes, plus dead-time associated with apparatus assembly). The relatively small number of labelled molecules detected, particularly at the lower pass counts, supports the observations in section 4.2.1; a small number of unfolding events may be responsible for flow-induced aggregation.

While both methods reflect similar trends in flow-induced labelling, the lower limit for accurate detection is substantially lower with HPLC-based fluorescence detection, with standard curves indicating accurate detection below 1% labelling equivalent fluorescence intensity, versus loss of linearity below approximately 15% in the case of samples quantified by densitometry (shown in figs 4.11 and 4.7, respectively). Thus, HPLC-based quantification was used for all analysis of IAEDANS labelling herein.

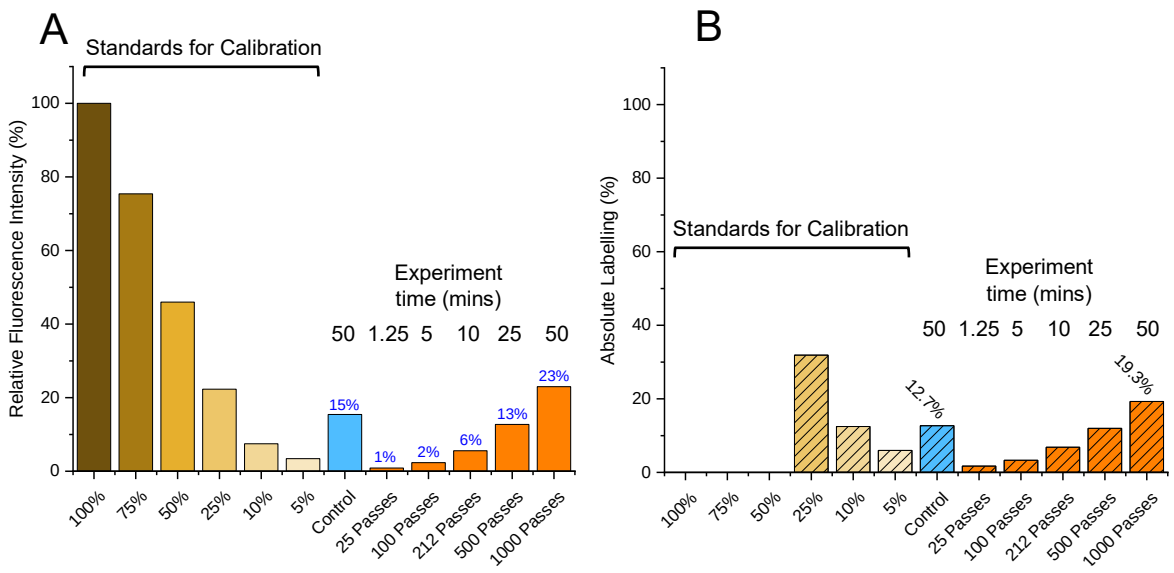


Figure 4.13: A137C IAEDANS Labelling Analysis Method Comparison

Analysis of IAEDANS labelling by densitometry and HPLC show similar trends. 0.5 mg/ml (28.5 mM) A137C RNase H* stressed in the EFD at 16 mm/s for varying pass counts as indicated, in arginine succinate buffer. A) Analysis by densitometry with SDS-PAGE gel, taken from fig 4.6B. B) Analysis with HPLC. Values above 25% of the '100% labelled' urea-unfolded sample were not included due to being above the maximum intensity threshold for the HPLC RF-20 fluorescence detector.

4.5 Discussion

Analysing the endpoint of the aggregation cascade is insufficient to understand the steps that precede it, and therefore structural insight into the conformational changes that occur is needed to understand what makes a protein susceptible to aggregation. This chapter describes development of a sensitive, reproducible method to measure protein unfolding under hydrodynamic flow, using the panel of RNase H* cysteine variants designed and produced in chapter 3.

To simulate flow fields that would be imparted on molecules during a biopharmaceutical's lifecycle, a bespoke extensional flow device (EFD) has been adopted, and allows protein solutions to be stressed in a controllable manner. A number of label-free assays were investigated for detecting conformational changes in RNase H* under flow, but were insufficiently sensitive and reproducible.

Use of the thiol-reactive label IAEDANS was considered, due to its high reactivity and irreversible covalent labelling mechanism. This also offers the additional benefit of being site-specific when used to label RNase H* variants, as each contains a single cysteine residue in known positions. To minimise loss of soluble sample to aggregation, labelling was analysed in arginine buffer during all optimisation assays, shown here to inhibit protein aggregation whilst inducing a slight decrease in protein thermodynamic stability.

The labelling assay initially appeared effective, and the method of quantification was optimised throughout the chapter; showing that HPLC-based quantification was significantly more reproduceable and sensitive than separation of samples by SDS-PAGE and quantification by densitometry. Absolute % labelling of the RNase H* ensemble could be measured using absorbance at 280 nm and 336 nm to indicate protein and IAEDANS concentration, respectively. The low sensitivity of absorbance measurements has meant that fluorescence was used to calculate absorbance at 336 nm, which has proved reproducible and sensitive down to below 1% sample labelling. Once this optimised assay for unfolding was established, the panel of RNase H* cysteine variants could be analysed in detail, the results and implications of which are discussed in chapter 5.

5 Characterisation of Flow-Induced Conformational Changes in RNase H*

5.1 Flow-Induced Unfolding in Conditions with Reduced Aggregation

Using the optimised IAEDANS-based labelling assay, conformational changes under hydrodynamic flow could be characterised for the 10 diverse RNase H* cysteine variants. with this panel, a wealth of structural insight into protein unfolding could be obtained and this applied to future development of biopharmaceuticals.

5.1.1 Off-Target Labelling of Cysteine Variants

Whilst IAEDANS is most readily reactive with cysteine (Mendoza and Vachet, 2009), reaction with other residues has been reported; namely histidine, lysine, methionine, and in some cases tyrosine (Heinrikson *et al.*, 1965)(Korman and Clarke, 1955)(Gundlach, Stein and Moore, 1959)(Whitehurst *et al.*, 2007). Due to imidazole and lysyl groups being predominantly reduced (and therefore nucleophilic) at pH 8, additional labels may result from reaction with these rather than other, less pH-sensitive, groups (Gurd, 1967).

To test specificity of IAEDANS labelling for cysteine in this assay, WT RNase H* (containing no cysteine residues) was stressed in arginine buffer for up to 1000 passes at 16 mm/s, equal to the longest exposure any cysteine variant would receive, and extent of labelling was quantified (described in section 2.2.7.4). After 1000 passes, the sample was approximately 1.9% labelled (fig 5.1), the majority being due to quiescent labelling. This suggests the non-cysteine labelling sites are either surface exposed or in a labile position under quiescent conditions.

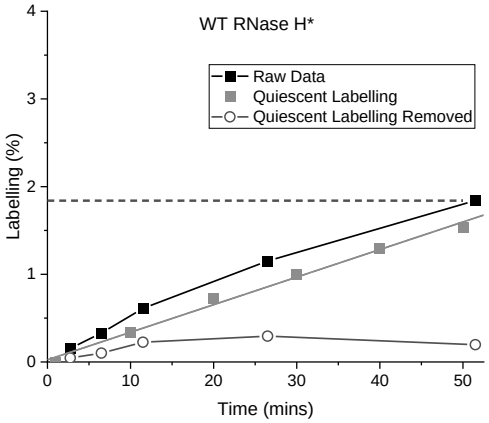


Figure 5.1: IAEDANS Labelling in WT RNase H*

WT RNase H* exhibits IAEDANS labelling despite the absence of cysteine residues. 0.5 mg/ml RNase H* exposed to 16 mm/s plunger velocity for 25-1000 passes (1.5-50 mins) in arginine succinate buffer. Dotted line indicates maximum total labelling measured.

To investigate this further, A137C RNase H* unfolded in urea overnight and labelled with IAEDANS (section 2.2.7.3) was examined using intact mass measurements from mass spectrometry (conducted by a colleague, Dr Samantha Lawrence).

No peaks corresponding to the molecular weight of unlabelled A137C RNase H* were detectable, indicating the overnight incubation with urea was sufficient to facilitate near-complete IAEDANS labelling of this RNase H* variant. It appears that aside from a single-labelled population, there also exists a doubly and triply AEDANS-labelled population (fig 5.2) in this variant; a mass of 17,532 Da is expected for unlabelled protein, with addition of 306 Da per AEDANS label. Additions of ~16 Da likely correspond to addition of oxygen.

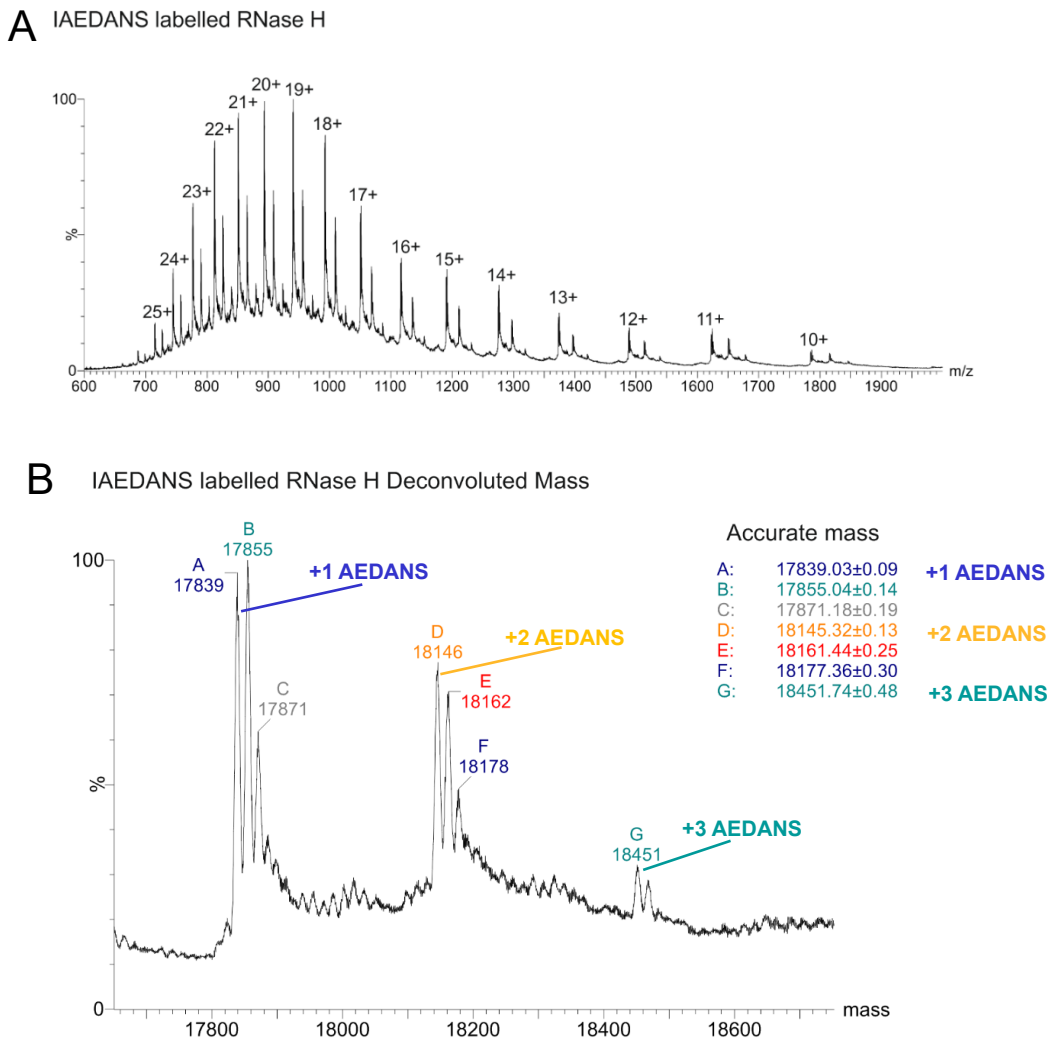


Figure 5.2: IAEDANS-Labelled A137C TOF-MS Analysis

20 μM A137C RNase H* in ammonium bicarbonate buffer (section 2.1.4) used. A) Raw spectra B) De-convoluted mass spectrum.

5.1.2 Removal of Contribution from Off-target Labelling

De-convoluting data to remove the effects of off-target labelling is challenging, as the extent of it is likely variable between cysteine variants. This is indicated by IAEDANS labelling in A52C, which is presented in the next section (fig 5.3) and labels to a lesser extent than WT RNase H* (up to 1.1% labelled compared to 1.8%), perhaps due to changes in stability resulting from the cysteine substitution.

In low-intensity measurements, the greater proportional effect of noise may also contribute to variability in off-target labelling between variants. As such, rather than

subtracting the labelling measured in WT directly from all other results, this estimated off-target labelling is displayed as a threshold for guidance, below which labelling may be predominantly due to off-target reactions (fig 5.1, dashed line).

A method to limit side reactions is to reduce the concentration of IAEDANS in relation to the protein, however a significant excess is necessary to ensure the highest chance of labelling occurring upon unfolding and cysteine exposure. As the presence of secondary and tertiary IAEDANS labels reduces the specificity of the assay, identifying the sites of these additional labels would be beneficial. To do this, mass spectrometry following digestion of the RNase H* sequence could be employed.

5.1.3 RNase H* Cysteine Variants Exhibit Heterogeneous IAEDANS Labelling Under Hydrodynamic Flow

Once this threshold for labelling was established, each variant was exposed to hydrodynamic flow for between 25-1000 passes (1.25 – 50 minutes total time under flow), in the presence of IAEDANS to capture unfolding events (section 2.2.7.4) and in arginine succinate buffer, to inhibit aggregation (section 4.3.2).

Increasing time exposed to flow resulted in greater labelling in all variants (fig 5.3). The fact that this varies between variants with cysteine residues in different positions suggests this increase is a result of an accumulation of events in which thiol groups become accessible to IAEDANS, rather than a reaction with an already fully surface-exposed thiol in each. The exception to this is in I82C, in which labelling rapidly increases before reaching a plateau, with flow-induced labelling increasing at a similar rate to quiescent labelling. This behaviour indicates that the 17.72 Å exposure of the cysteine residue calculated for I82C in section 3.1.1 is substantial enough to allow almost complete IAEDANS labelling of the sample ensemble within 10 minutes. As such, this variant was discarded from analysis of EFD-induced unfolding, as effects from flow could not be detached from quiescent labelling.

The remaining cysteine variants showed a range of labelling propensities, from 1% to 34% total sample labelling. Also, all variants exhibited some degree of labelling in quiescent samples, which increases linearly with time, suggesting that it is due to either exposure of thiol groups in the native state, or structural fluctuations that occur under quiescent conditions.

As discussed above (section 5.1.2), the contribution from off-target labelling was thought not to be constant between variants, so controls had to be prepared for each variant, to capture this variability. For each variant, the linear fit of quiescent labelling was interpolated from, and this was subtracted from total measured labelling at each point measured, to quantify labelling attributable to the effects of the EFD.

Importantly, each cysteine variant labels more under flow than quiescently, indicating the EFD causes conformational changes that expose cysteine residues. The consequences of these conformational changes to aggregation are explored in section 5.2.3; the purpose of this section is to investigate regional unfolding under flow through the site-specificity of this labelling assay.

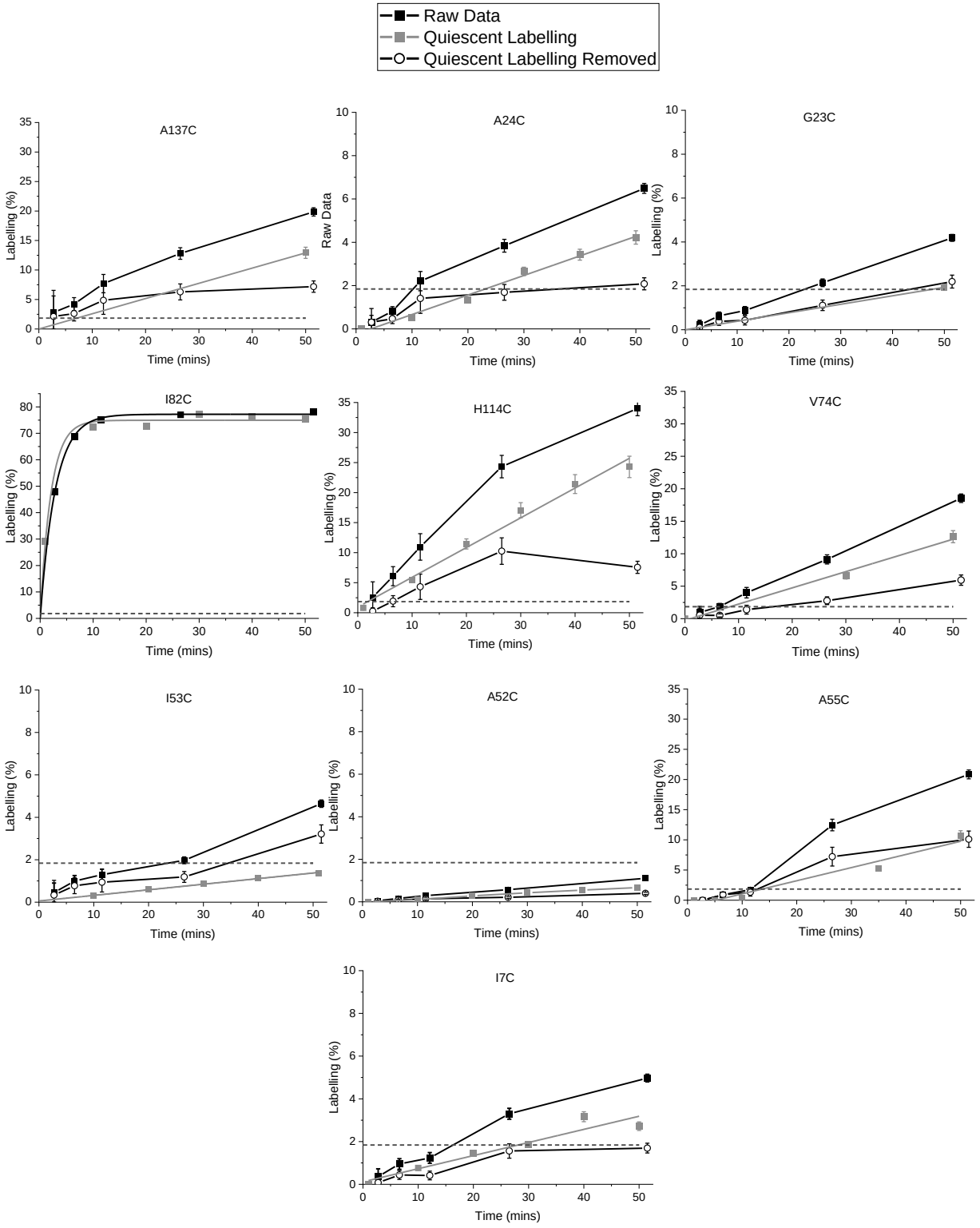


Figure 5.3: Flow-induced IAEDANS Labelling in RNase H* Cysteine Variants

Cysteine substitution position determines quiescent and flow-induced IAEDANS labelling. 0.5 mg/ml of each RNase H* variant exposed to 16 mm/s plunger velocity for 25-1000 passes (1.5-50 mins) in arginine succinate buffer. Dotted line indicates maximum labelling measured in WT RNase H*. Error bars shown are estimate standard deviation of the mean, based on the value for G23C RNase H*, which was calculated from repeat measurements, n=2.

The extent of flow-induced IAEDANS labelling in each variant can be more readily compared in figure 5.4A, which displays labelling after 1000 passes in the EFD, and quiescent labelling at the equivalent timepoint, interpolated from the quiescent standard curve produced for each variant, as described above.

Interestingly, quiescent and flow-induced labelling correlate, suggesting flow may coax already labile regions of the protein into further unfolding. This is illustrated in figure 5.4B, showing that most variants associate with a linear trend in this regard. Those that lie furthest away from this trend, namely A55C and I53C, may indicate sites that are relatively more affected by hydrodynamic flow, and as such show the greatest proportional increase in labelling when under flow.

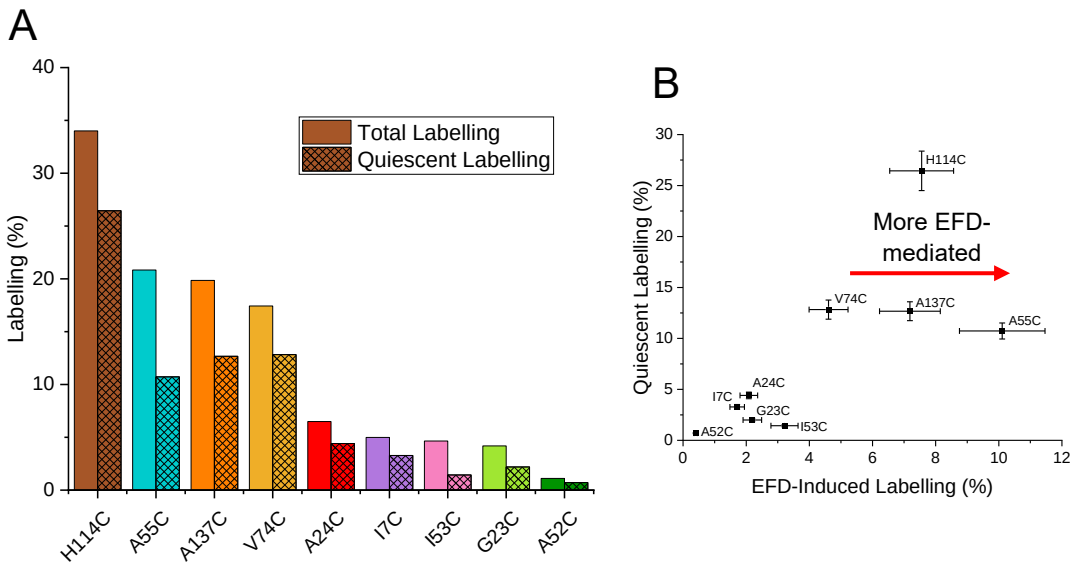


Figure 5.4: IAEDANS Labelling Correlated with Flow-Induced Aggregation

A) Labelling in 0.5 mg/ml RNase H* in arginine succinate buffer shown after 1000 passes at 16 mm/s plunger velocity in EFD-stressed samples (total labelling), or 50 minutes quiescent in control samples. Error bars shown are estimates based on the standard deviation associated with G23C RNase H*, which was calculated from repeat measurements, n=2.

5.1.4 Structural Conclusions from IAEDANS Labelling Under Flow

To verify this concept, and characterise which regions in RNase H* are more susceptible to EFD-induced perturbation, labelling intensity can be mapped onto the RNase H* structure. As expected, quiescent labelling appears to be related to the structural features that govern how protected the residues are from solution, and by extension the degree of motion that would be required to expose these. For example, A52C and I53C both label

minimally, and are found in the hydrophobic core (fig 5.5A). H114C displays the opposite behaviour, and is located on a peripheral loop, connecting two α-helices. Other variants such as A137C and V74C show moderate levels of quiescent labelling, perhaps because while shielded by whole secondary structure features, are not within any known regions of high structural integrity.

Once the contribution from quiescent labelling has been subtracted, EFD-induced labelling can be mapped to the RNase H* structure (fig 5.5B). As with quiescent labelling, structural insight can be gained from these labelling patterns. The high relative labelling in A55C RNase H* (10.1% EFD-induced labelling) suggests that under flow, the structure protecting this residue from labelling is perturbed, likely being strand 3 of the β-sheet (fig 1.14). Alternatively, α-helix A may be perturbed, although the low degree of overall labelling in A52C and I53C suggests the hydrophobic core would remain intact in this instance. This is a region of 3 defined triads of hydrophobic contacts consisting of 5 leucine, 2 alanine, 1 tryptophan and 1 isoleucine residues between helices A and D (fig 1.14)(Katayanagi *et al.*, 1992).

The fact that the core remains protected in this way suggests that a set of hydrophobic interactions in this orientation offers significant mechanical stability against flow-induced unfolding, even when one of the contacts is replaced with a polar cysteine residue.

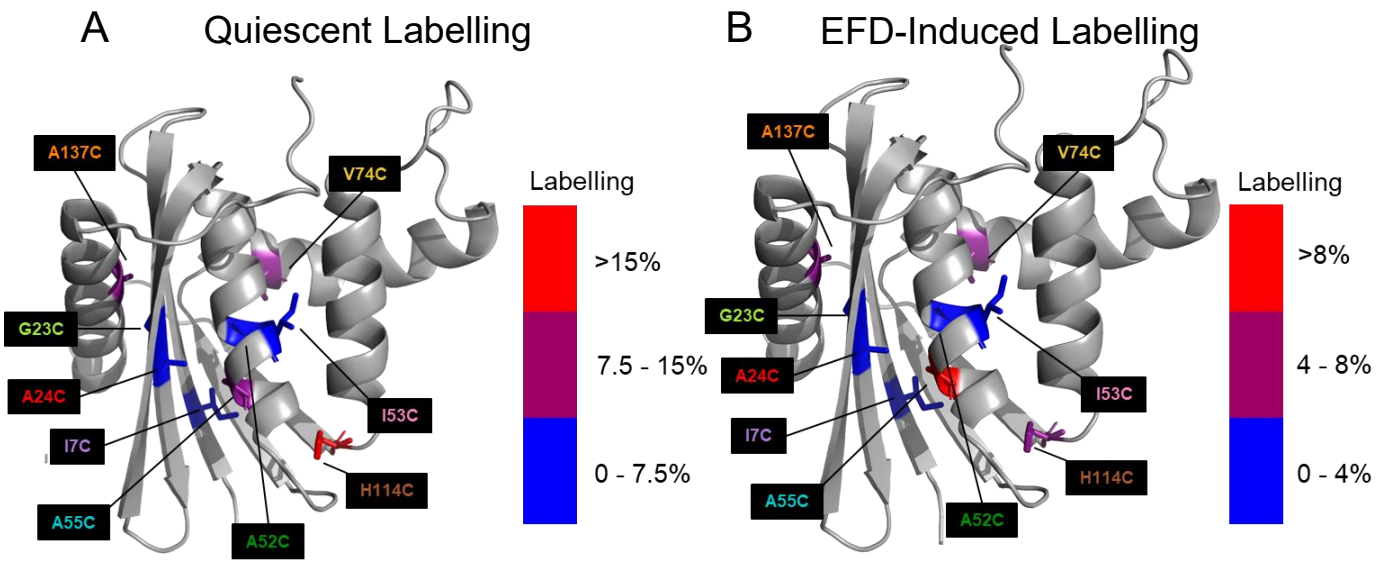


Figure 5.5: EFD-Induced VS Quiescent IAEDANS Labelling in RNase H* Cysteine Variants

A) Quiescent labelling and B) EFD-induced labelling with quiescent labelling subtracted, mapped onto RNase H* cartoon structure (PyMOL). Labelling values taken from fig 5.4A, with EFD-induced labelling being total labelling with quiescent labelling subtracted.

In section 5.1.3 above, it was concluded that flow may further expose regions that are already susceptible to transient exposure under quiescent conditions. The observation that labelling is greater in more peripheral sites in RNase H* structure, for example, A137C, V74C and H114C, appears to support this conclusion (fig 5.5B).

5.1.5 Folding Sequence Does Not Determine Flow-Induced Unfolding

One of the appeals of utilising RNase H* as a model protein is the well-characterised folding pathway (section 1.6.2), and the fact that the stabilities of all but one of the defined foldons has been established (Rosen *et al.*, 2015). The thermodynamic stability of each of these foldons was considered as a parameter which could be indicative of unfolding propensity, however it did not correlate with either quiescent or flow-induced unfolding (fig 5.6). Using thermodynamic stability of foldons may be too coarse-grained to draw clear correlations, with the limited dataset available. Foldon thermodynamic stability values were also taken from the literature, without consideration for the effects of the cysteine substitutions.

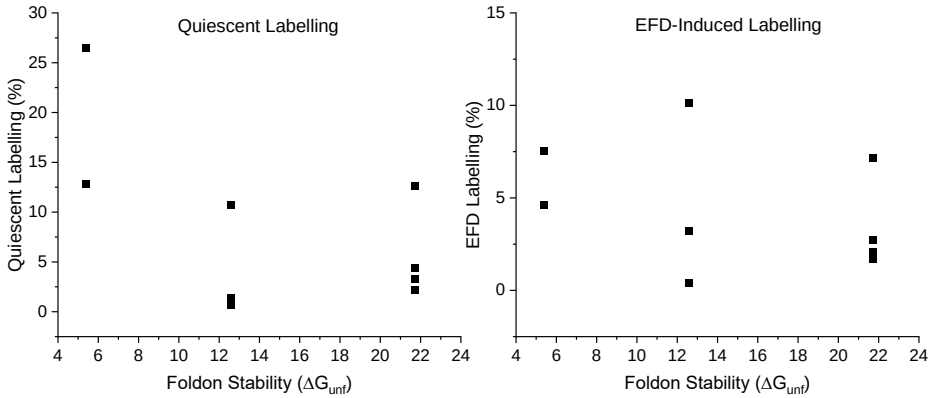


Figure 5.6: Foldon Thermodynamic Stability Measurements do not Correlate With EFD-Induced or Quiescent Unfolding

Unfolding measured by IAEDANS labelling does not correlate with foldon thermodynamic stabilities taken from Rosen *et al*. (2015). Labelling values taken from fig 5.4A, with EFD-induced labelling being total labelling with quiescent labelling subtracted.

As discussed in section 1.1.5, mechanical unfolding may proceed by perturbations that are geometrically localised, relative to the chemical denaturation used to establish the thermodynamic stability values discussed above. To investigate whether foldon order relates to the unfolding behaviour of RNase H* under hydrodynamic flow, solvent accessible surface area (SASA) was calculated for each RNase H* variant, when the

foldon-associated structure is removed in reverse order to folding (fig 5.7). The calculated values suggest that unfolding proceeds in the reverse foldon order, with the majority of unfolding in the ensemble being in the red and yellow (ultimate and penultimate, respectively) foldons; A55C is expected to experience maximal exposure upon loss of the red foldon, and is the most labile of the RNase H* variants (fig 5.4), whereas A52C and I53C require unfolding of all foldons aside from the blue I_{early} region, before they exhibit maximal exposure, and are both the least susceptible to flow-induced unfolding (fig 5.4).

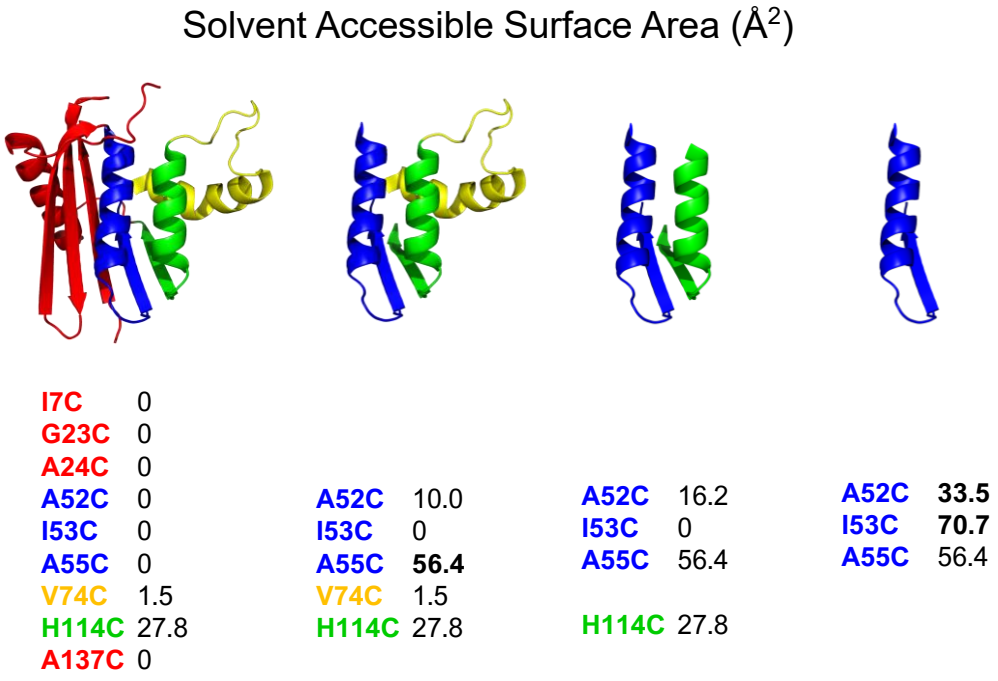


Figure 5.7: Calculated Solvent Accessible Surface Area Increases with Removal of Foldon Structure

Solvent accessible surface area (SASA) values calculated using WT RNase H* structure (1F21) in PyMOL. Maximal calculated SASA values are indicated in **bold**.

5.1.6 Flow-Induced Unfolding Corresponds to Aggregation

Whilst it is known that protein unfolding often precedes aggregation, flow-induced aggregation is not widely attributed to flow-induced unfolding (Booth *et al.*, 1997)(Pang *et al.*, 2023). To investigate this, flow-induced IAEDANS labelling (section 5.1.3) was

plotted against flow-induced soluble protein loss (section 3.2.2)(fig 5.8). These two metrics appear to correlate, in both quiescent and EFD-stressed samples.

Flow-induced unfolding and labelling with IAEDANS may thus either result in greater aggregation, or aggregation may facilitate further unfolding. This is discussed throughout the chapter and broader thesis (section 5.2.5). As all RNase H* variants contain the expected secondary and tertiary structure associated with the native state under quiescent conditions (section 3.1.4), it is assumed that the early stages of aggregation proceed via unfolding from the native state. However, it may be that aggregation is further propagated by protein unfolding catalysed by interaction with existing aggregates (Cohen *et al.*, 2013)(Meisl *et al.*, 2022). Whilst this has been described in models, detailed experimental validation of this process is lacking (Andrews and Roberts, 2007).

H114C appears to be an outlier regarding this trend, but only when quiescent labelling is included (i.e. in fig 5.8A and C); this is likely a result of the cysteine residue in H114C being partially exposed under native conditions (27.8 $Å^2$ predicted solvent accessible surface area), meaning “background” quiescent labelling in this variant is much higher than others (fig 3.1).

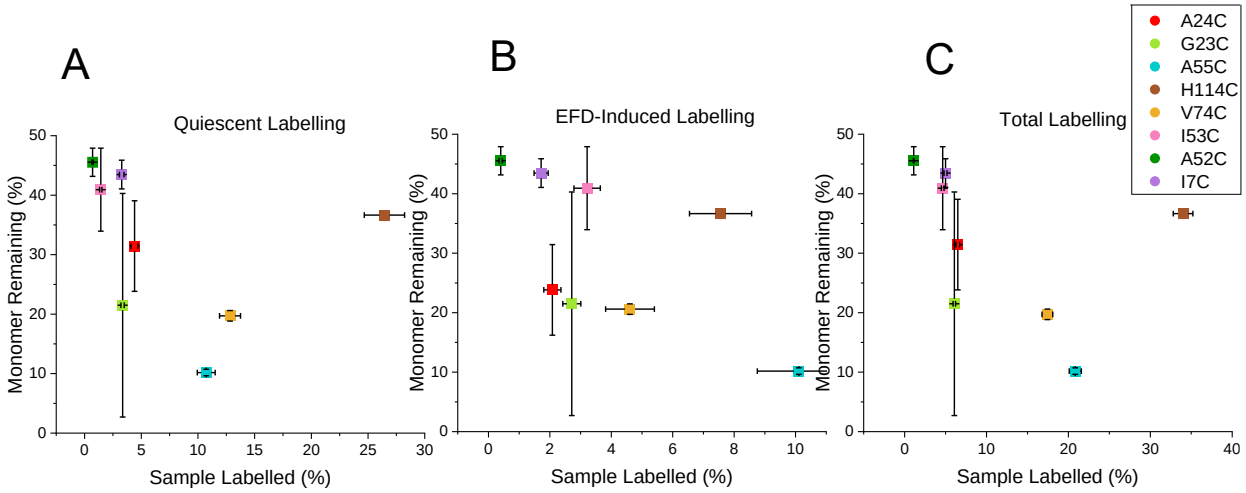


Figure 5.8: IAEDANS Labelling Correlated with Flow-Induced Aggregation

For A) quiescent labelling, B) EFD-induced labelling (total labelling with quiescent removed), and C) total labelling: P = -0.813, -0.594 and -0.873. R^2 = 0.662, 0.353, 0.762, respectively. Quiescent and Total labelling fits were produced excluding H114C RNase H* (indicated in brown). Labelling values taken from fig 5.4A. Error bars shown for labelling are estimates based on the standard deviation associated with G23C RNase H*, which was calculated from repeat measurements, n=2. Error bars for monomer remaining are standard deviation of the mean, n=2.

The fact that when H114C is disregarded, both quiescent and flow-induced labelling correlate with flow-induced aggregation, would suggest that quiescent fluctuations and flow-induced fluctuations are both important. This is compounded by the fact that the best fit to a linear co rrel ation occurs when both are included; $R^2 = 0.762$.

Variants were ranked according to labelling and soluble protein remaining, to allow further analysis. Correlation of the rank order is statistically significant (fig 5.9A), and the data was hierarchically clustered, to group variants into those with the highest resistance to flow-induced unfolding and aggregation (section 2.2.9.1)(fig 5.9B). Interestingly A52C and I7C were grouped, despite being found on two different foldons (fig 3.1); again, disproving the hypothesis that foldon order may determine stability. Geometrically, both variants contain substitution sites relatively central in RNase H*, suggesting protection from unfolding by local structure is a key determinant of resistance to flow-induced aggregation.

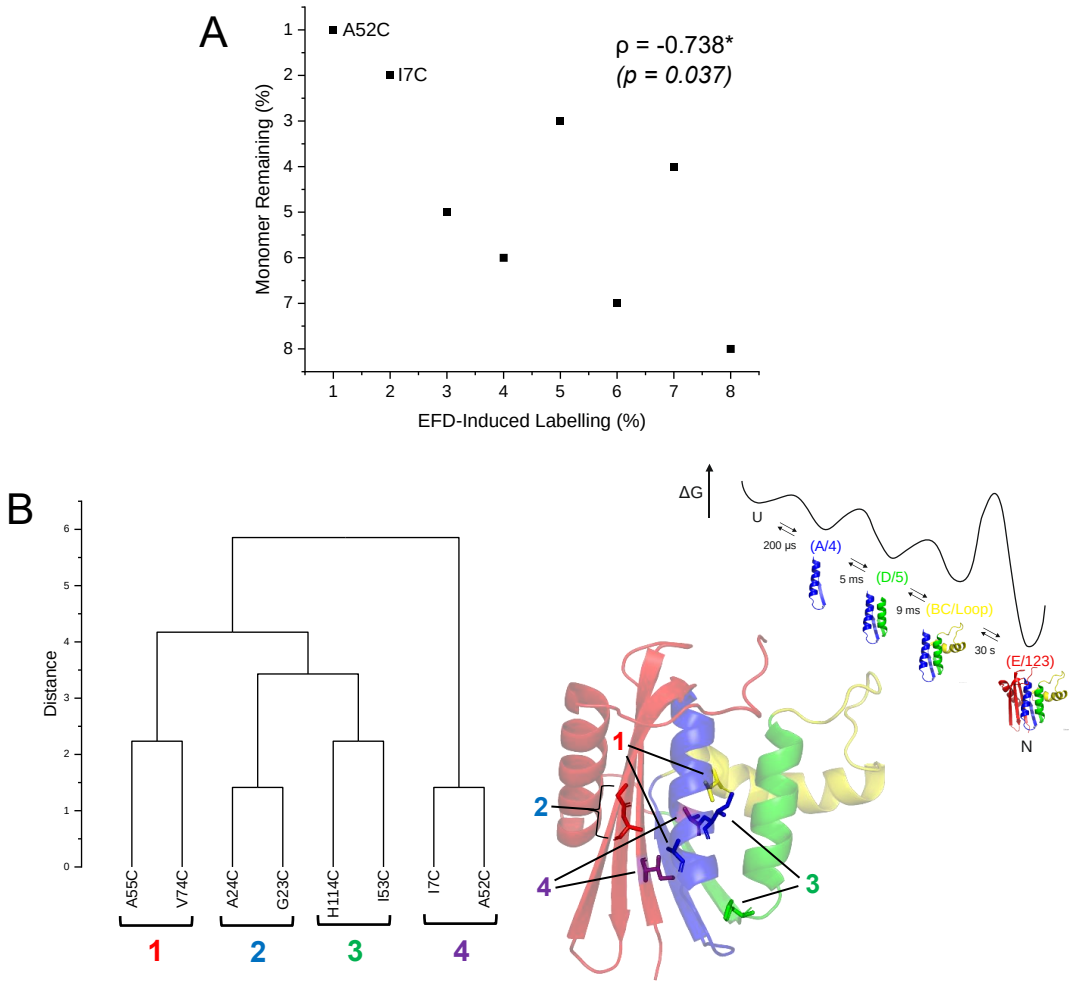


Figure 5.9: Rank Order Analysis of Flow-Induced Unfolding and Aggregation

A) Rank order of soluble protein remaining versus EFD-induced sample labelling for each RNase H* variant analysed. Raw data presented in figure 5.7B. Person's correlation coefficient indicated with statistical significance shown. B) Hierarchical clustering analysis of the rank order data set in (A)(left), with pairings shown mapped onto the RNase H* cartoon structure (right). Hierarchical folding pathway established by Hu *et al*. (2013) shown to indicate foldon order. All analysis carried out using Origin 2021 graphing software.

5.1.7 Thermodynamic Stability is Not Predictive of Flow-Induced Protein Unfolding

As discussed in section 3.3.2, thermodynamic stability is a commonly used developability parameter, but does not appear to be predictive of flow-induced aggregation susceptibility. In the RNase H* variants analysed, flow-induced unfolding does not correlate with thermodynamic stability, indicating the flow-induced labelling observed is not a result of destabilisation from introduction of cysteine substitutions (fig 5.10). This also aligns with the previous conclusions that chemical denaturation at equilibrium is a

poor predictor for mechanical stability under flow, and therefore both should be considered separately in the context of developability.

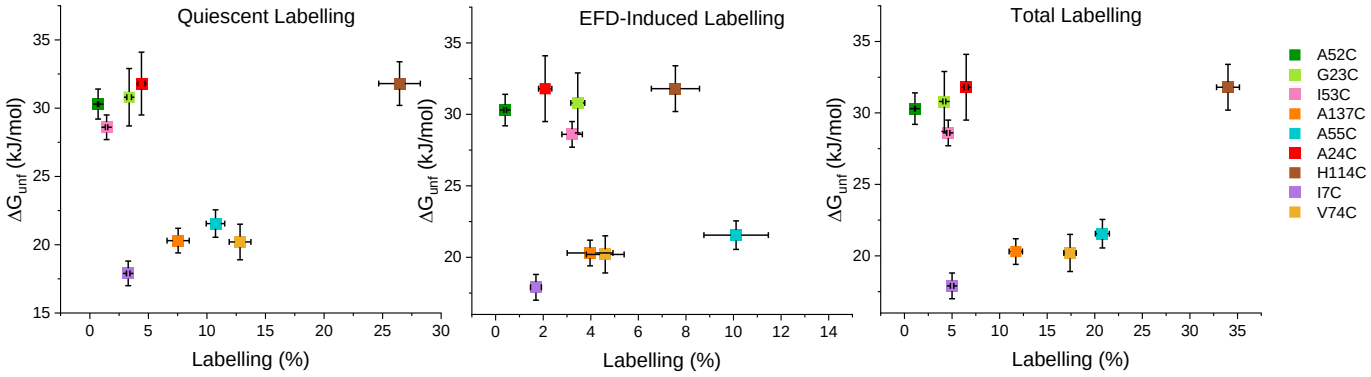


Figure 5.10: Thermodynamic Stability Versus Flow-Induced Unfolding

Equilibrium chemical denaturation and IAEDANS labelling data taken from Figs 3.8 and 5.4. Error bars are standard deviation of the equilibrium denaturation fit for each variant. Error bars shown for labelling are estimates based on the standard deviation associated with G23C RNase H*, which was calculated from repeat measurements, n=2. ΔG_{unf} error values for each variant are standard deviations estimated using Igor Pro, from residuals of datapoints to the equilibrium unfolding curve fit.

5.2 Flow-Induced Unfolding in Aggregation-Permissive Conditions

5.2.1 IAEDANS Labelling in Phosphate Buffer Experimental Design

As previously discussed (section 1.4.3), flow-induced unfolding has not been widely assigned as a key cause of aggregation in proteins. Whilst unfolding can be measured effectively in arginine buffer, and correlates with aggregation in RNase H* variants, further study is needed to directly attribute the source of aggregate material with unfolded protein.

To achieve this, RNase H* variants were stressed in the EFD using the same IAEDANS labelling assay as in section 5.1.3, but with phosphate buffer (section 2.2.7.5) at the same pH, rather than arginine. This was to permit aggregation, whilst measuring unfolding with labelling. Following exposure to flow, samples were collected and clarified using

ultracentrifugation to isolate aggregated insoluble, and soluble non-aggregated material (fig 5.11A). Additionally, it was realised that insoluble protein sample was deposited on the borosilicate syringe surface, accumulating over the course of each experiment (fig 5.11B). To recover this material, the emptied syringe barrel was filled and irrigated with 6 M GdnHCl containing DTT (section 2.2.7.5), before collection of the resulting mixture. This was to solubilise the mixture and ensure reduction of any disulfide bond-forming cysteine residues.

All components were solubilised in the same GdnHCl mixture overnight, and measured using RP-HPLC (section 2.2.5.2). Protein content and labelling percentage were quantified in each variant, to establish the fate of flow-unfolded protein.

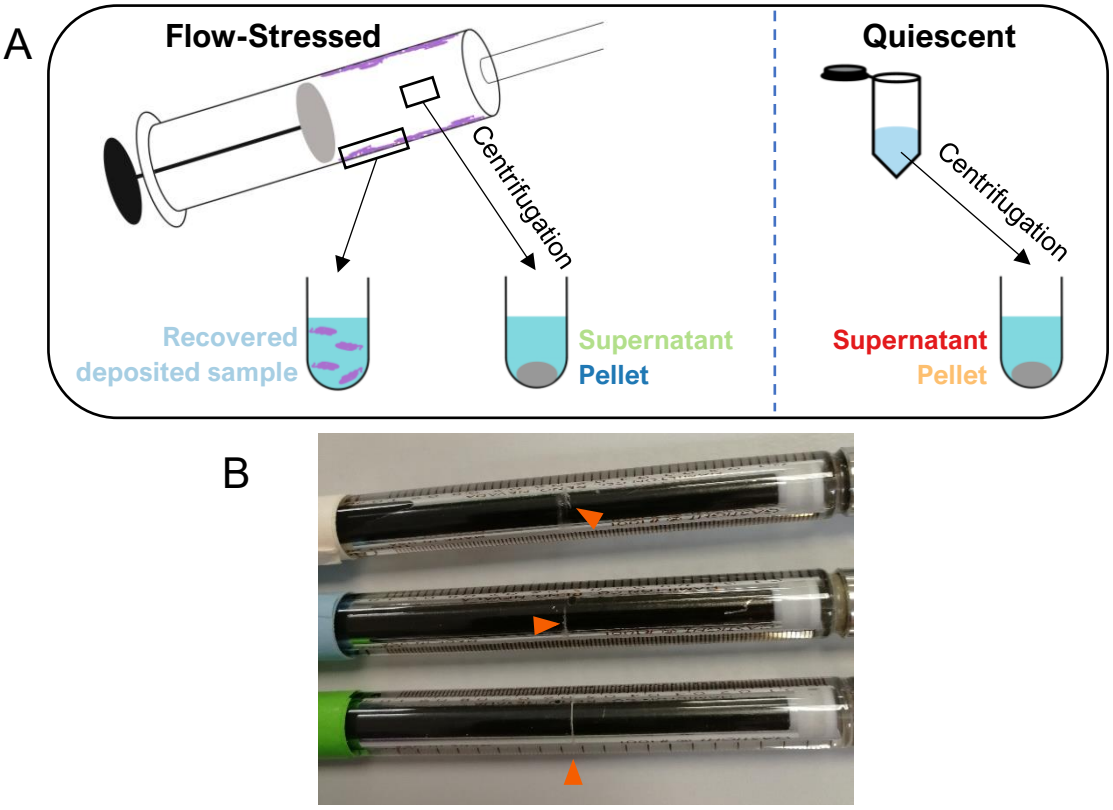


Figure 5.11: IAEDANS Labelling and Aggregate Recovery Assay

A) Schematic illustrating collection of aggregated and soluble material, including clarification steps by ultracentrifugation. B) Surface-deposited insoluble protein indicated by orange triangles on each syringe.

5.2.2 Flow-Induced Labelling in Different RNase H* Components

IAEDANS concentration was used to reflect total unfolded protein in each component. In the representative example of I53C RNase H*, the sample found in the insoluble pellet

following exposure to flow in the EFD contained the most IAEDANS label; after 1000 passes this was 90-fold greater than the soluble portion of the same sample (fig 5.12A). This is further evidence associating flow-induced unfolding and aggregation. Interestingly, while the deposited aggregate recovered from the syringe contributed to a lesser extent to overall labelling in the sample, it was found that this material was labelled to the same extent as the insoluble pellet collected from the liquid contents of the syringe (fig 5.12B). This suggests that it has a similar character (i.e. aggregate containing a relatively high degree of labelling), but there is just less of it.

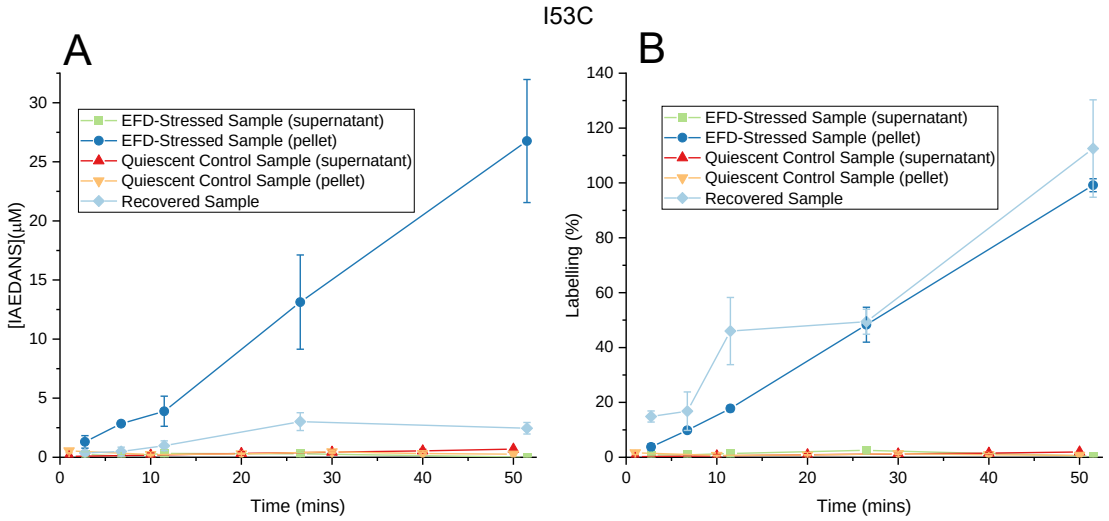


Figure 5.12: Unfolding Measured in Aggregated RNase H*

I53C RNase H* shown as a representative example. 0.5 mg/ml I53C RNase H* stressed at 16 mm/s for varying pass counts as indicated, in phosphate buffer. A) IAEDANS content of each component and B) percentage labelling of each component. Error is standard deviation of the mean, n=2.

The finding that a significant proportion of labelling is in surface-deposited aggregate is an important one, which highlights several possibilities.

One is that unfolding and/or aggregation may proceed predominantly on the surface and be released into bulk solution, perhaps by the sloughing effect of the syringe plunger against the barrel surface with each pass. Alternatively, unfolding and/or aggregation may occur in bulk and unfolded material could reach the surface and provide a nucleus for aggregation, or bulk-aggregated material may be deposited on the surface following formation in bulk. The analysis of RNase H* aggregation data in phosphate buffer using the SB_NB_A model indicates that the predominant step which influences insoluble aggregate formation is the bulk unfolding step, as described in section 3.2.4.

Regardless of where the unfolding and aggregation occurs, hydrodynamic flow is clearly causative in this; in I53C, after 1000 passes the insoluble portion of flow-stressed samples are proportionally 50-fold more IAEDANS-labelled than the equivalent quiescent sample (fig 5.12B). In addition, labelling increases with increasing pass count, indicating a dose-dependency of hydrodynamic flow and labelling that is concurrent with data in aggregation-inhibiting arginine buffer (section 5.1.3).

This method was applied to the RNase H* cysteine substitution variants, and a similar trend was observed with the exception of I82C; which contained substantially more IAEDANS labelling in the supernatant than other variants, further suggesting the residue is highly accessibly to IAEDANS under native conditions (fig 5.13).

Considering the percentage of total labelling from each component (recovered insoluble sample, insoluble pellet from liquid sample, and soluble supernatant from liquid sample) illustrates the aggregation occurring over time in the EFD (fig 5.13). Most cysteine variants show that as the experiment proceeds, soluble material unfolds and is labelled in solution, but goes on to aggregate, meaning the overall percentage content of soluble labelled protein does not increase. An interesting observation is that surface-deposited labelled aggregate reaches a plateau in terms of absolute concentration, illustrating the finite nature of the surface. To probe this in further detail, the characteristics of these surface deposits including morphology could be investigated, perhaps by design of a spectrophotometrically competent flow device that would allow observation of samples *in situ*. Although precluded by time constraints, this concept is discussed in further detail in section 1.7.2.

Under quiescent conditions labelling was substantially lower (between 10 and 70-fold less, after 50 minutes/1000 passes) compared to when stressed by the EFD, in all variants aside from I82C, discussed above. This further supports a role for hydrodynamic flow in protein unfolding. Interestingly, unfolded/labelled RNase H* remains in solution to a much greater extent in the absence of flow; approximately 50% versus an average of <5%. This suggests that flow is also a key driver of the transition from unfolded protein to aggregation.

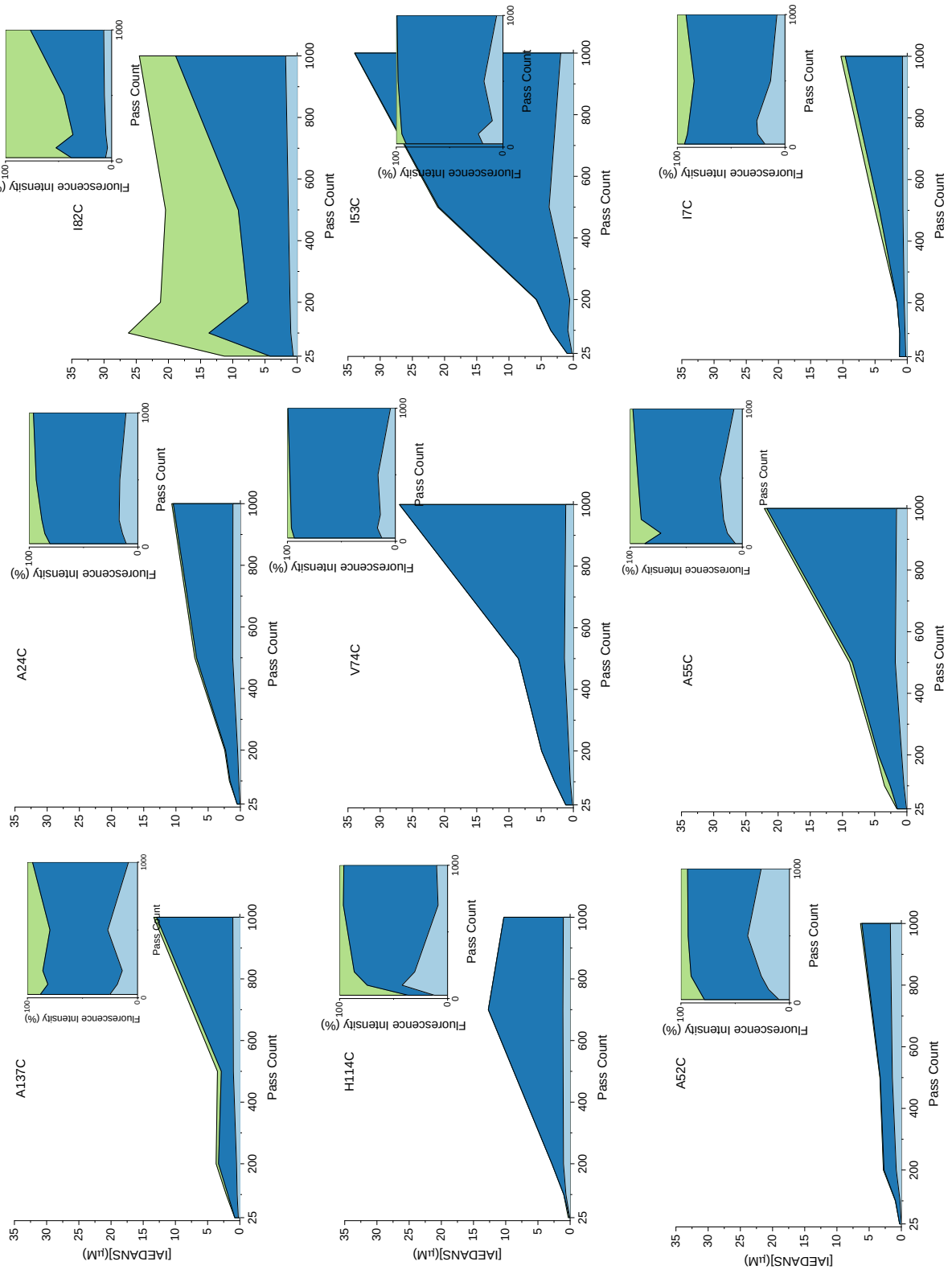


Figure 5.13: Unfolding Measured in Aggregated RNase H*

IAEDANS content in RNase H* components following exposure to hydrodynamic flow, indicating absolute extent of labelling in each sample. Percentage of labelling arising from each component shown in insets. 0.5 mg/ml RNase H* variants stressed at 16 mm/s for varying pass counts as indicated, in phosphate buffer.

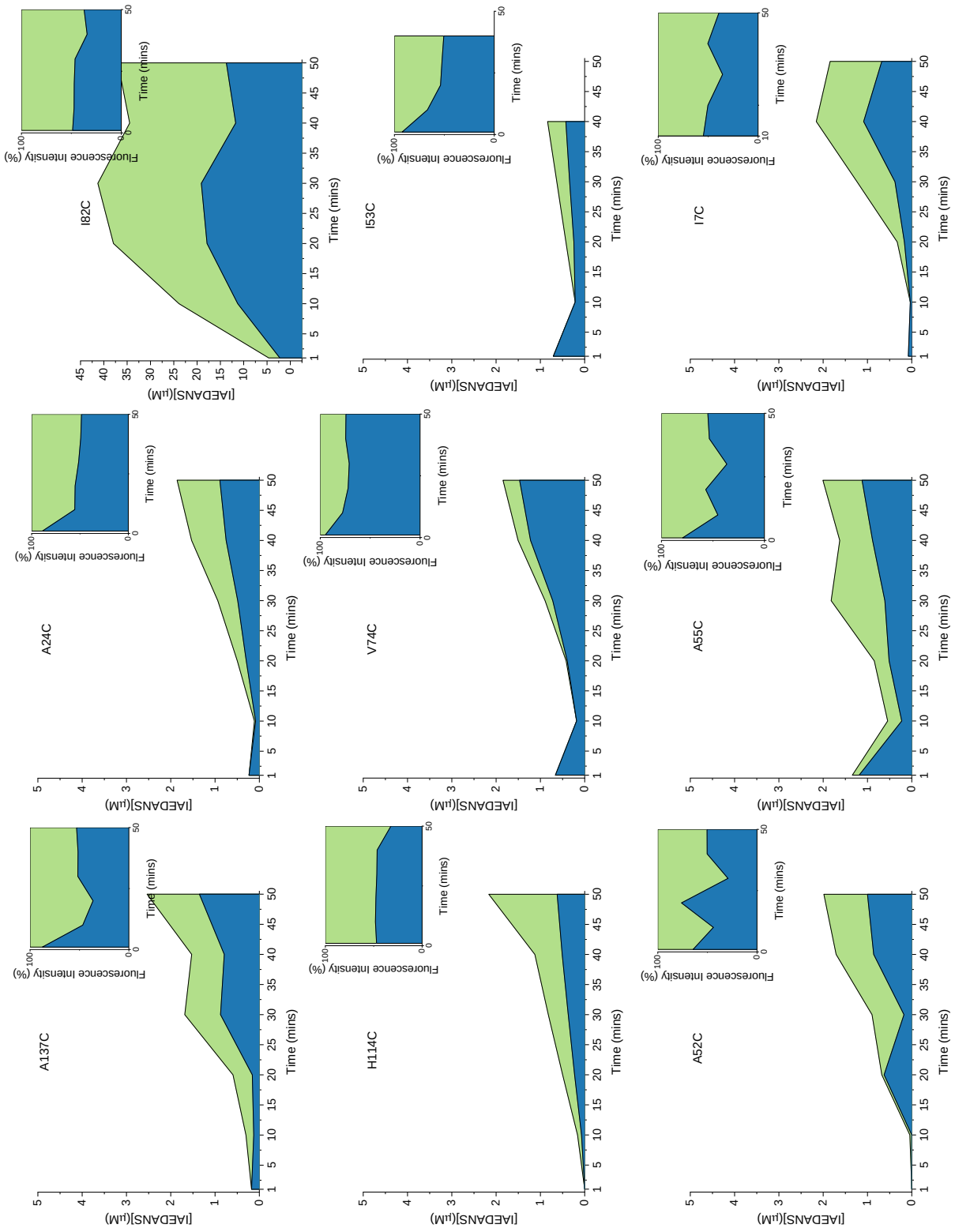


Figure 5.14: Unfolding Measured in Quiescent RNase H*

IAEDANS content in RNase H* components following various timepoints quiescent at room temperature, indicating absolute extent of labelling in each sample. Percentage of labelling arising from each component shown in insets. 0.5 mg/ml RNase H* variants left quiescent at room temperature and sampled at varying timepoints as indicated, in phosphate buffer.

5.2.3 Labelling in Aggregate Correlates with Aggregation Propensity in RNase H* Cysteine Substitution Variants

The previous sections (5.2.1 and 5.2.2) have demonstrated that in RNase H* cysteine variants, exposure to hydrodynamic flow in the EFD increases labelling with IAEDANS, a reporter for unfolding, and that aggregated material containing label also increases under these conditions. However, whether unfolded material proceeds to form this aggregate, while suggested in section 5.1.6, is still unclear. To further associate the two observations, percentage labelling in the pellet was shown to correlate with soluble protein loss (fig 5.15A). This relationship varies between approximately 0.66 to 1.07 IAEDANS label equivalents per aggregated RNase H* molecule (calculated for each datapoint in fig 5.15A), indicating that each molecule experiencing a significant enough unfolding event to become IAEDANS labelled, is likely to go on to aggregate.

The range in this relationship may be due to two factors: 1) unfolding events that do not expose labelling sites or are not captured by the IAEDANS reaction (the latter likely to make up a very small portion of cases due to the high reactivity and concentration of IAEDEANS (section 4.3.1)). 2) Off-target labelling, a possibility in accordance with previous observations in the scale of this contribution (section 5.1.1). Also, the fact that combined components in the EFD-stressed sample, as illustrated in fig 5.15B(ii), add up to approximately 60% of the initial protein concentration suggests a significant part of the sample was not recovered.

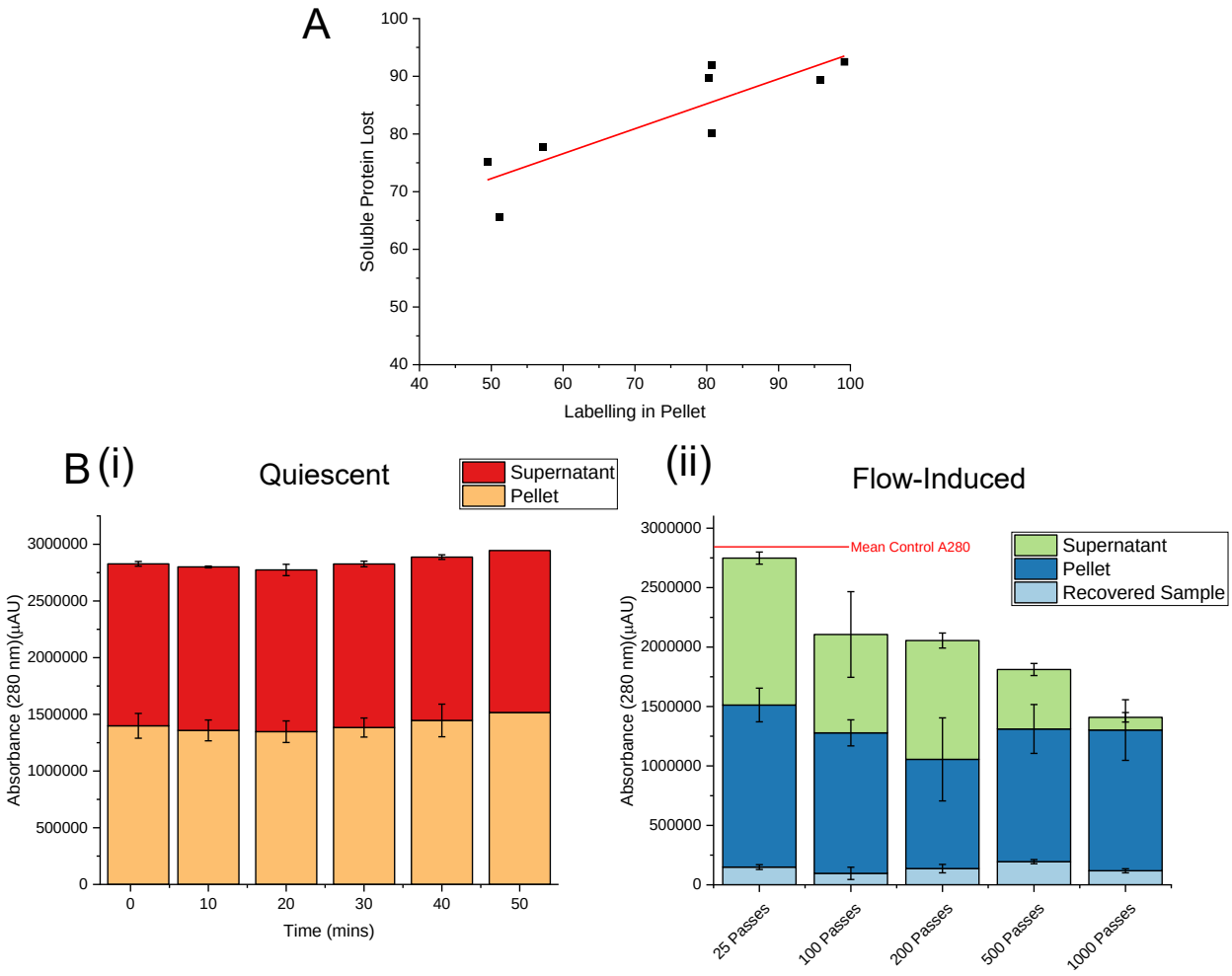


Figure 5.15: Degree of IAEDANS Labelling in Aggregate Protein Correlates with Degree of Aggregation

A) Soluble protein loss correlated with labelling in the pellet, in EFD-stressed samples. ρ = 0.867, R^2 = 0.752. B) Example soluble protein loss data for I53C RNase H* (i) quiescent and (ii) following exposure to hydrodynamic flow sampled at various times and stressed for various pass counts, as indicated on x-axes. Samples were in phosphate buffer. Error bars are standard deviation of the mean, n=2.

It was considered that the addition of IAEDANS could modulate aggregation, due to its size and hydrophobic aromatic group/polar sulfonic acid group. It appeared that in phosphate buffer and following 1000 passes at 16 mm/s plunger velocity, EFD-induced aggregation was on average greater in samples containing IAEDANS, than those without; approximately 15% soluble protein remained compared to a quiescent control, versus 32%, respectively (fig 5.16). Although not fully accounting for aggregation, the presence of an IAEDANS label likely amplifies the increase in aggregation observed with EFD-induced unfolding.

This highlights a weakness in the design of this experiment; aggregation observed is not reflective of that which would proceed in natural RNase H*; a consideration that should

be made if applying this type of unfolding assay to molecules of clinical interest. To circumvent this issue, label-free assays, or assays based on non-fluorescent quantification that do not require addition of large chemical groups could be used to characterise flow-induced unfolding. These could include hydrogen/deuterium exchange (HDX) methods, or more emergent techniques, such as fast photochemical oxidation of proteins (FPOP), which have already been applied to monitor conformational change of biopharmaceutics in the EFD (Lawrence, 2023).

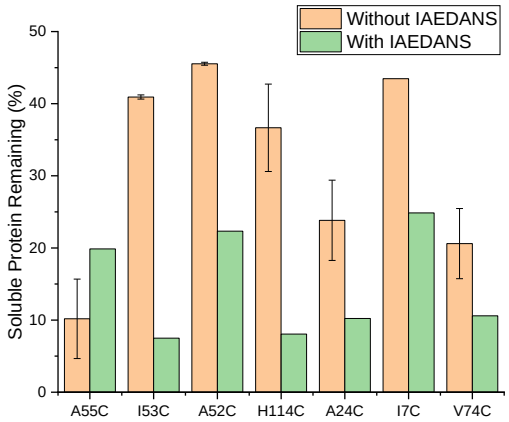


Figure 5.16: Flow-Induced Aggregation with and without IAEDANS

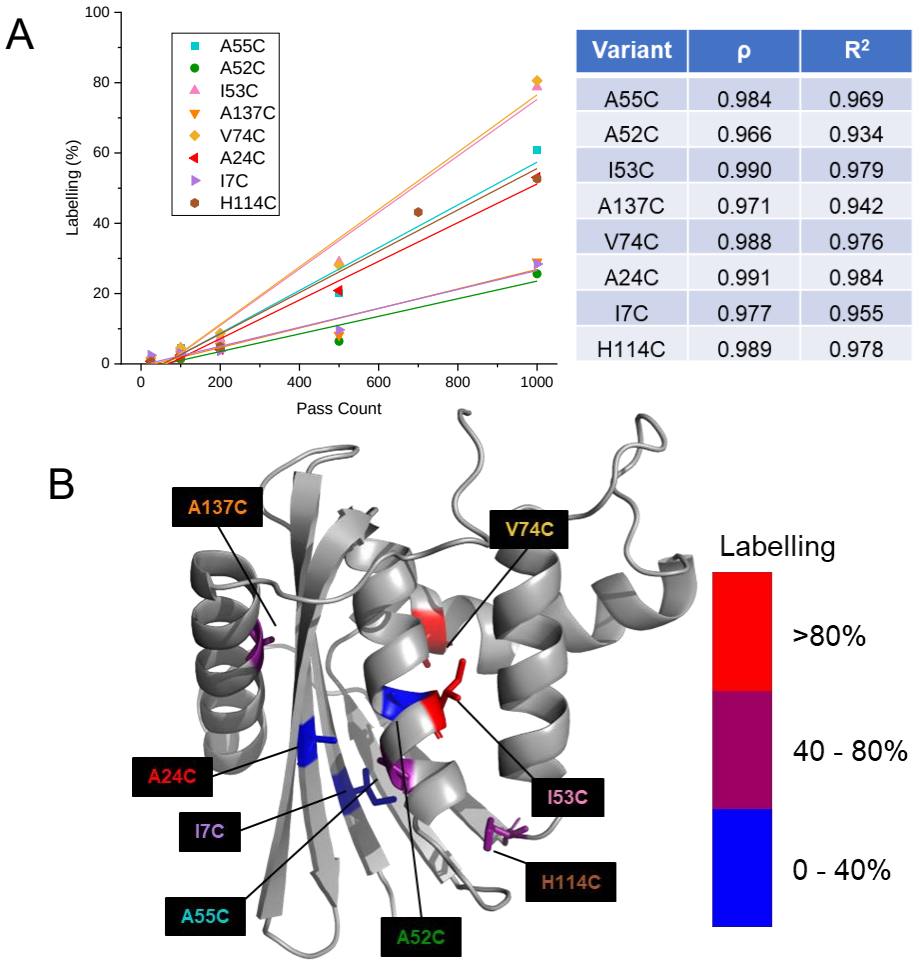
0.5 mg/ml of each RNase H* variant stressed at 16 mm/s in phosphate buffer. Soluble protein remaining calculated relative to a control sample, left quiescent for 50 minutes.

5.2.4 Structural Considerations from Labelling in Aggregation-Permissive Conditions

When fluorescence contributions from all components from EFD-stressed samples (i.e. the recovered insoluble sample, insoluble pellet from liquid sample, and soluble supernatant from liquid sample), are combined, labelling appears close to linear (fig 5.17A). This suggests IAEDANS labelling in this system is first order, limited by RNase H* unfolding rather than concentration and reactivity of IAEDANS.

To gain structural insight into unfolding in aggregation-permissive conditions, IAEDANS labelling of the insoluble aggregate and soluble sample, following 1000 passes at 16 mm/s, was mapped onto RNase H* structure (fig 5.17B). Interestingly, the pattern in observed in phosphate buffer (aggregation-permissive) is different to when aggregation is supressed by arginine (section 5.1.4); labelling is still relatively high in sites that have a greater degree of structure surrounding them. For example. whilst relative protection of

the significantly buried A52C and I7C (grouped together in fig 5.9B, by their low extent of labelling and aggregation) appears low compared to other residues, total IAEDANS labelling is still significantly higher than in arginine.



Variant	ρ	R^2
A55C	0.984	0.969
A52C	0.966	0.934
I53C	0.990	0.979
A137C	0.971	0.942
V74C	0.988	0.976
A24C	0.991	0.984
I7C	0.977	0.955
H114C	0.989	0.978

Figure 5.17: Total IAEDANS Labelling of RNase H* in Aggregation-Permissive Conditions

0.5 mg/ml RNase H* exposed to flow at 16 mm/s in phosphate buffer. A) Percentage labelling of total RNase H* sample with exposure to flow. Linear fits to the data shown as solid lines, with pearson's R-values and R^2 values shown for each variant. B) Total EFD-induced labelling with quiescent labelling subtracted, mapped onto RNase H* cartoon structure (PyMOL).

5.2.5 RNase H* Unfolding in Arginine Versus Phosphate Buffer

Previously, it was considered that flow-induced unfolding might be further propagated by aggregation (section 5.1.6), rather than aggregation being solely caused by flow-induced unfolding. To investigate this, flow-induced unfolding in arginine and phosphate buffers was compared, to evaluate the difference between unfolding in aggregation-inhibitory and permissive conditions, respectively.

For this analysis, protein depositions that had been recovered from washing of syringes with GdnHCl were not included. This was to ensure a fair comparison of labelling between phosphate and arginine buffered samples, as samples in arginine buffer did not show any obvious deposition so were not exposed to the same treatment.

Across all RNase H* variants, after 1000 passes flow-induced unfolding was significantly greater in phosphate buffer than arginine (fig 5.18A). At lower pass counts, it appears that some variants show greater flow-induced labelling in arginine, however this is due to a higher level of quiescent labelling, rather than flow-mediated effects (fig 5.18B)i). Increased quiescent labelling in arginine could be due to the slight thermodynamic destabilisation that this buffer induces (section 4.3.2). Although thermodynamic stability measured in phosphate buffer does not correlate with labelling in arginine (fig 5.10), further work is needed to define if this would be different, was thermodynamic stability measured in arginine buffer. As arginine's stabilising effects are thought to be due to steric effects and stabilisation of unfolded states, inhibiting aggregation (discussed in section 4.3.2), it was deemed more likely that arginine was not stabilising the folded state.

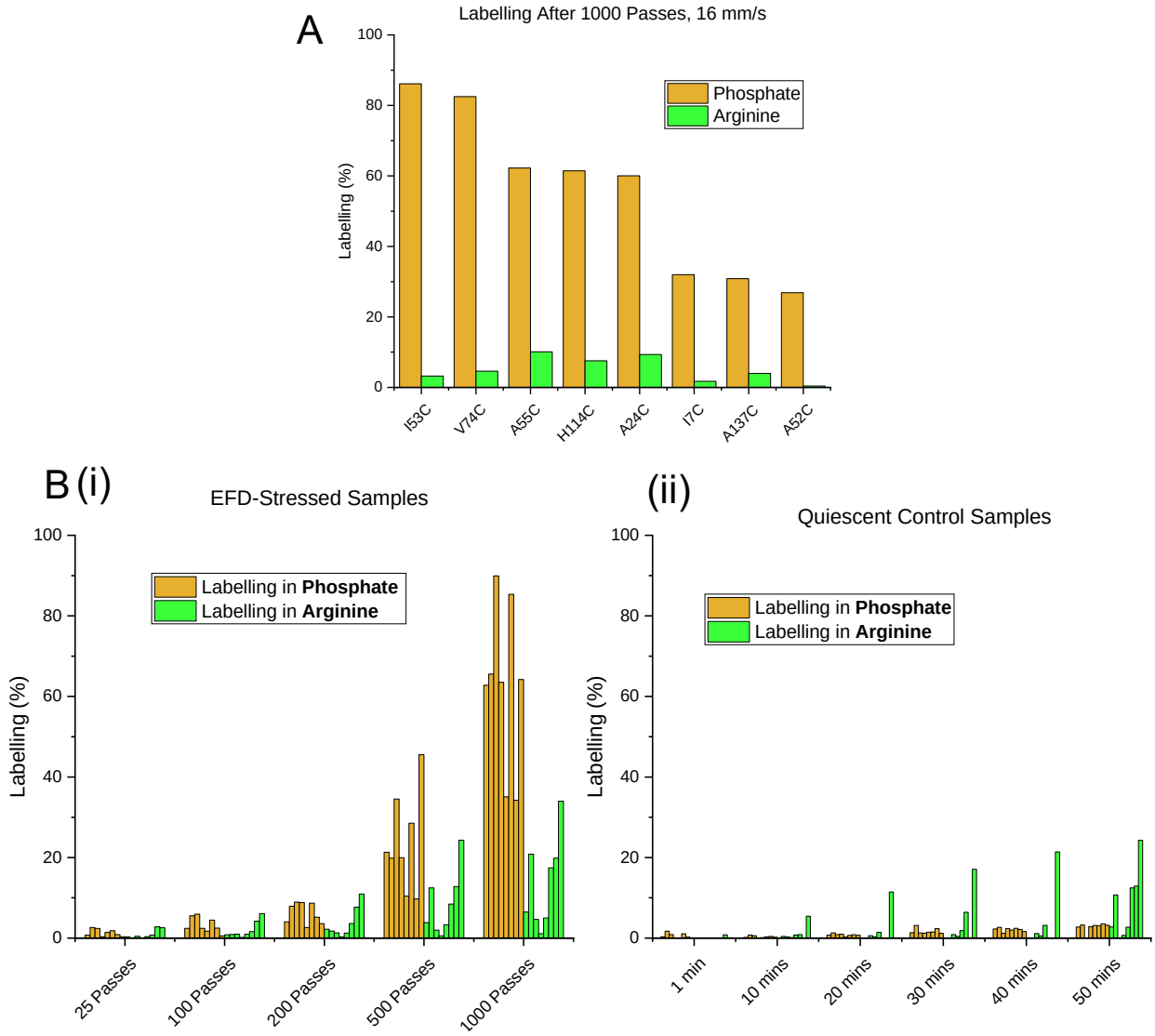


Figure 5.18: IAEDANS Labelling in Phosphate Versus Arginine Buffer

Comparison of percentage labelling between combined supernatant/pellet in phosphate samples and unclarified samples in arginine buffer. Results from 0.5 mg/ml RNase H*, stressed at 16 mm/s for the pass counts indicated, were used. A) Comparison of labelling following 1000 passes between variants, with corresponding quiescent labelling for each variant subtracted. B) Labelling with increasing i) pass number and ii) quiescent control time before sampling.

However, a prominent difference between phosphate and arginine-buffered systems is that aggregation is much greater in the former. Without arginine present, progression of aggregation could result in further unfolding, resulting in exposure of usually well-protected sites, such as residue 52, and further labelling (if these sites contain cysteine substitutions)(illustrated in fig 5.19). This type of occurrence has been described but is poorly characterised experimentally, and could describe a mechanism of flow-induced unfolding and aggregation (Andrews and Roberts, 2007). However, as the proportion of

labelled sample increases, the number of labelling-competent molecules decreases, which could obscure detectable increases in labelling.

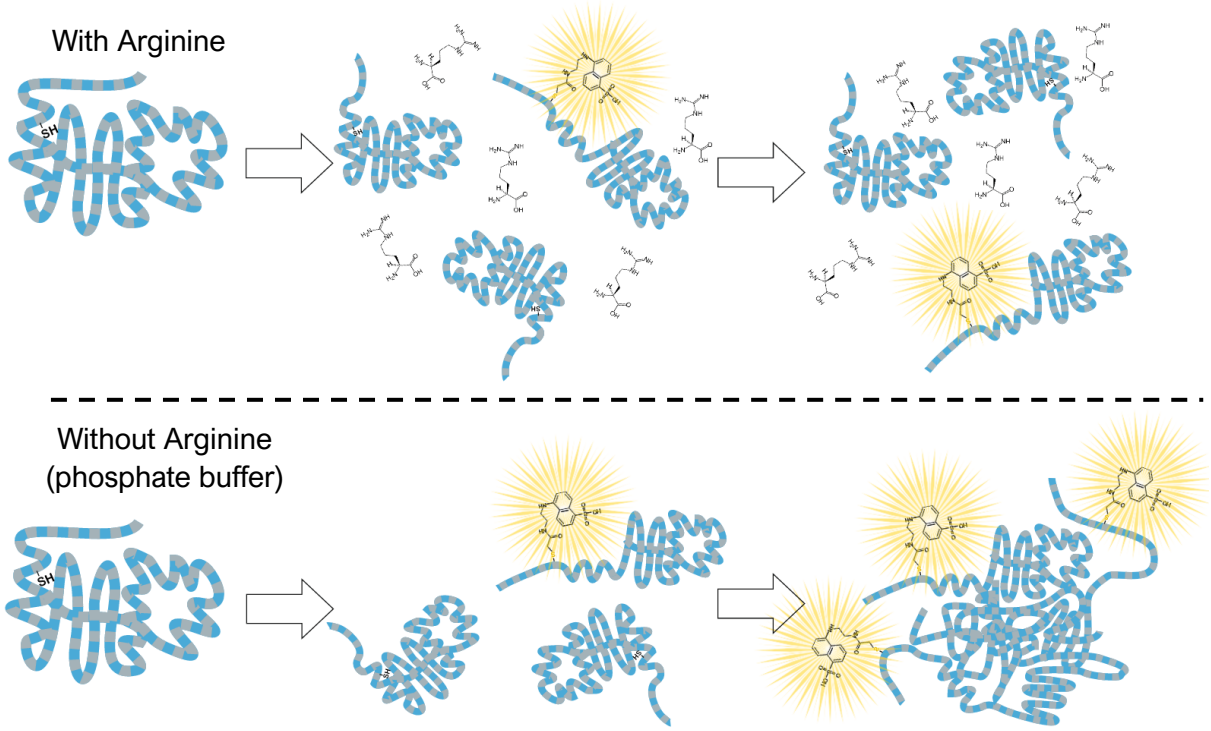


Figure 5.19: IAEDANS Labelling in Phosphate Versus Arginine Buffer Proposed Mechanism

RNase H* unfolding and IAEDANS labelling may be further progressed by aggregation, which can be inhibited by arginine, shown schematically.

5.2.6 Flow-Induced Labelling Increase in Phosphate Buffer is Geometrically Determined

To further understand this possible mechanism of increased labelling, fold-increase in labelling from arginine to phosphate buffer was calculated for each RNase H* variant. RNase H* variants appear to fall into three groups; those that experience a minimal increase in labelling in phosphate (≤8-fold), those that experience a moderate increase (18-27-fold), and A52C, which experiences the greatest increase (67-fold)(fig 5.20A). Interestingly, these groups closely correspond to the site of the cysteine variants relative to RNase H* structure; more peripheral variants do not experience as much increase as those that are located in centrally (fig 5.20B). This may be due to the greater degree of structure protecting from unfolding in more central positions, and suggests that in

phosphate buffer, RNase H* unfolds more completely, exposing sites in the core that are better protected from unfolding in arginine.

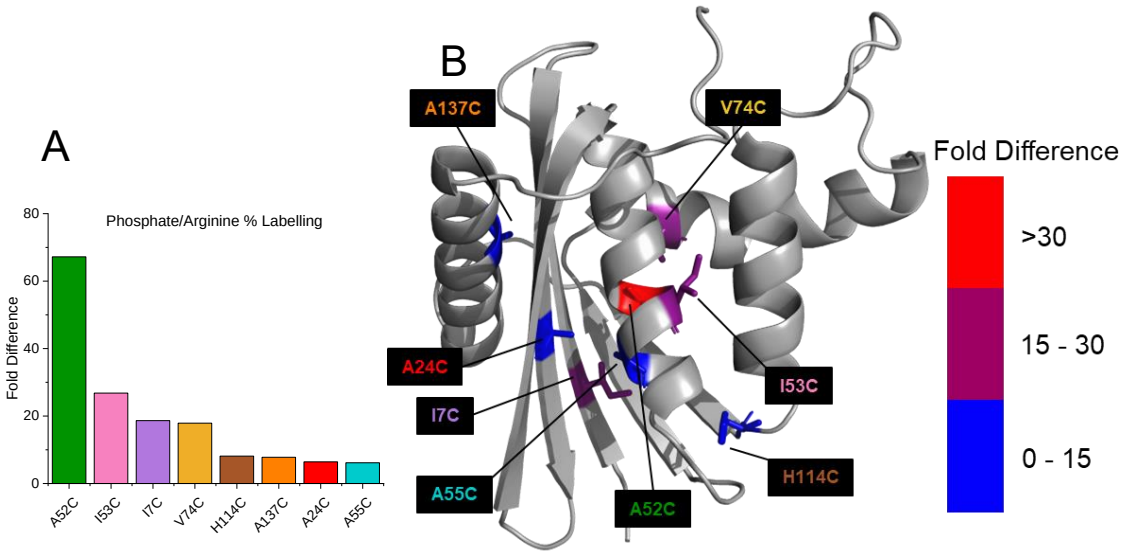


Figure 5.20: IAEDANS Labelling Change in Aggregation-Permissive Conditions

Fold difference in IAEDANS labelling from arginine to phosphate buffer. 0.5 mg/ml RNase H* variants compared, following 1000 passes at 16 mm/s. Combined supernatant/pellet in phosphate samples compared with unclarified samples in arginine buffer. A) Shown as a bar chart for comparison. B) Mapped onto RNase H* cartoon structure (PyMOL).

5.3 Discussion

5.3.1 Conclusions

The work presented in this chapter applies the methodology developed in chapter 4, with the panel of RNase H* variants produced and characterised in chapter 3, to investigate the effects of hydrodynamic flow on protein unfolding, with the aim of identifying whether this unfolding leads to protein aggregation. Whilst the use of IAEDANS labelling to measure unfolding was shown to be limited in specificity by off-target labelling, IAEDANS has proved sufficiently reactive to show a significant difference in labelling of flow-stressed RNase H* compared to a quiescent control, despite the challenges of a dynamic environment and potentially transient unfolding events.

Analysis of unfolding in arginine succinate buffer has shown repression of aggregation (fig 4.4), allowing analysis of unfolding in a mainly soluble sample. Analysis of cysteine

variants with a range of labelling sites offers detailed structural insight into the regions in RNase H* that are differentially susceptible to flow. This has revealed that the hydrophobic core is the most well-protected region of the protein, and that peripheral regions are not only more susceptible to flow-induced perturbations, but are also more labile under quiescent conditions.

It was thought that flow-induced unfolding propensity might be attributed to the stabilities of 'foldons' established by (Hu *et al.*, 2013), as RNase H* substitution variants were introduced across all 4 of the established foldons. However, the order in which RNase H* folds doesn't appear to explain variability in labelling, either in quiescent or EFD-induced labelling. For example, the most readily labelled variant (A55C) and the most lowly-labelled variant (A52C) contain cysteine substitutions within the same foldon, the region highlighted in blue that is thought to fold first under native conditions (section 3.1). Previously, mechanical unfolding from the N- and C- termini using optical tweezers (from tethers attached to residues 4 and 155) showed that the first region to unfold under force corresponded to the final foldon to form (Cecconi *et al.*, 2005); a more coarse-grained finding that is not explained by the heterogeneity in unfolding of different sites within foldons presented here. This indicates the complexity of flow-induced unfolding and suggests that force is not exerted in one plane, a valuable conclusion but one that may be limited to RNase H* due to its globular structure; larger macromolecules including antibodies have been shown to adopt specific alignments under flow (Engstler *et al.*, 2007).

The outcome of this flow-induced unfolding was then investigated, finding that it correlated with aggregation. Furthermore, when the IAEDANS labelling assay was used to measure flow-induced unfolding in the presence of aggregation, it was found that aggregated material was highly labelled in comparison to soluble protein. This directly associates the two processes, indicating flow-induced unfolding leads to aggregation, or vice versa. The observation of highly labelled, non-water soluble protein, deposited on syringe surfaces, also suggests a role for surface-mediated aggregation, a process that has been observed in the literature, and is predicted to occur by modelling of RNase H* aggregation data (sections 3.2.3 and 3.2.4)(Kopp *et al.*, 2023).

The combination of two separate buffer-system (arginine as a repressor of aggregation, phosphate as a permitter) approaches reveals that under aggregation-permissive conditions RNase H* unfolds significantly more when aggregation is not supressed by arginine. This not likely to be due to arginine suppressing unfolding; quiescent samples in arginine label to the same or greater extent than those in phosphate. It could be that

this is a result of aggregation stimulating further unfolding, following flow-induced unfolding. It may be in this way that a small number of initial flow-induced unfolding events are responsible for significant loss of sample, and – extrapolated to apply to biopharmaceuticals – failure of therapeutics.

This information is valuable for a fundamental understanding of flow-induced aggregation that has not been achieved at this detail before. In addition, the implications on biopharmaceutical design and stability analysis are significant, and could aid in directing development of future therapeutics, with greater stability and therefore reduced aggregation propensity.

5.3.2 Considerations and Future Work

Whilst all samples were analysed using RP-HPLC, with the expectation that the 0.1% TFA (~pH 2) running buffer was sufficient to disaggregate RNase H* and allow detection of the total sample, samples labelled in phosphate and arginine were not treated the same, with samples in phosphate experiencing an additional solubilisation step in GdnHCl/DTT (section 5.2.1). Therefore, IAEDANS labelling experiments in arginine with GdnHCl/DTT solubilisation steps should be considered, to allow the most accurate comparison. However, as discussed in section 4.4.4, significant multimer peaks were not detected in any chromatograms, and 500 mM DTT was used to quench all labelling reactions, which would be sufficient to reduce disulfide bonds.

Alternative residues may also be valuable in interrogating different regions of RNase H*, without requiring introduction of new mutations; labelling of lysine or tyrosine residues for example could be combined with a peptide digest, and analysed using MS, to determine the positions of labels. In addition, other labels such as ABD-F (4-Fluoro-7-sulfobenzofurazan) could be investigated, although the low cost and high reactivity, coupled with convenience of thiol chemistry, make IAEDANS an effective probe (Toyooka and Imai, 1984).

Use of a label-free method to detect protein unfolding would still be desirable; to determine the effects of introducing single cysteine mutations into RNase H*, each variant requires comprehensive structural characterisation. Hydrogen-deuterium exchange (HDX) techniques such as HDX mass spectrometry (HDX-MS) and HDX nuclear magnetic resonance spectrometry (HDX-NMR) could give detailed residue-by-residue

information on protein unfolding; an improvement on the lower-resolution local information afforded by measuring exposure of single labelling sites.

6 Predicting Instability and Aggregation in Biopharmaceuticals

6.1 Introduction

6.1.1 Applying the EFD to Measure Flow-induced Aggregation in Different Biopharmaceutical Scaffolds

Throughout this thesis, results and their implications have been considered in the context of biopharmaceutical development. RNase H* was used as a model protein system with expectation that the fundamental mechanistic insight could be applied to more complex biopharmaceuticals. Previous chapters were focused on flow-induced aggregation of this model protein.

For biopharmaceutical development, it is essential to establish if the diversity of behaviour under flow exhibited between variants of RNase H*, is also seen in biopharmaceuticals. This is necessary for defining the features that determine aggregation propensity in these molecules, and later, development of aggregation-resistant molecules. Several biopharmaceuticals have already been studied in detail using the EFD, as well as larger panels in less detail (Willis *et al.*, 2018)(Willis *et al.*, 2020). While F_c fusion protein have been investigated alongside traditional IgG molecules (Willis *et al.*, 2023), the increasing breadth of modalities in use and popularity of bispecific mAbs in particular indicates that analysis of biopharmaceuticals in this format is highly important; as of 2022, 79 multi specifics have reached clinical trials (Wilkinson and Hale, 2022).

6.1.2 Establishing a Holistic Developability Parameter

The other aspect of this chapter is based on developability: the concept of predicting a potential therapeutic molecule's likelihood of translation into a manufacturable and stable medicine, by integrating a range of assays designed to probe features of the candidate molecule (Xu *et al.*, 2019). Each developability parameter aims to assay for a different potential pitfall, and sensitively distinguish molecules that will experience stability and efficacy problems.

Using a large number of assays can be both time-consuming and expensive, with potential for significant consumption of material. While this may be necessary to ensure therapeutics that are appropriate for use, only recently have studies investigated whether there may be redundancy in the selection of assays used. Jain *et al*., (2017) examined a panel of 137 antibodies with 12 different assays, and found assays grouped significantly when testing similar attributes. This raises the question: could developability be predicted using a single, holistic parameter which combines the most effective combination of assays without a loss in predictive ability?

Recently, colleagues have established a method to produce a holistic developability parameter (HDP) from 28 outputs of 12 assays, and shown that HDP score correlates with stability over 6 months at 25 °C in a selection of 9 mAb/formulation combinations (section 1.3.8)(Willis *et al.,* manuscript in preparation). The work in this chapter will seek to apply this principal to a panel of 27 mAbs with different targets and morphologies, to assess the widespread application of this analysis.

6.1.3 Considering the Role of Extensional Flow in Developability

An additional consideration made throughout this chapter is whether flow-induced aggregation studies could extend more broadly from their original use as experiments used to gain fundamental insight into hydrodynamic flow. Could results from the EFD be applied as a developability parameter?

When compared to *in silico* predictors of insolubility and aggregation (CamSol, TAP and Solubis, (Sormanni, Aprile and Vendruscolo, 2015)(Raybould *et al.*, 2019)(van der Kant *et al.*, 2019)) using a panel of 33 mAbs, flow-induced aggregation showed no correlation (Willis *et al.*, 2020). This perhaps indicates the complexity of flow-induced aggregation and mechanical unfolding that is purported to precede it; as discussed in section 1.1.2, considerations such as pulling geometry and buffer conditions play major roles and are not accounted for by sequence or even structure-based computational models.

Establishing if commonly used experimental characterisation techniques correspond to flow-induced aggregation would be beneficial in understanding whether this technique could be predictive of those, or if it is able to provide entirely novel insight.

6.2 mAb Panel Characterisation

6.2.1 Study Design and Overview

Whilst the usage of bi-specifics and other antibody modalities has been increasing year-on-year, the IgG antibody format still makes up approximately 75% of therapeutic mAbs (Wilkinson and Hale, 2022). The majority of the dataset is therefore of that format, although bispecific molecules were also analysed, to investigate if differences were detectable in this emerging and increasingly popular modality. The bispecifics investigated were all different combinations of two F_V regions (Bis 1-5), aside from Bis 6, which was independent. All 27 mAbs were provided by AstraZeneca Cambridge, and analysed in PBS throughout, with those not provided in PBS being buffer exchanged into this before use (similarly to as described in section 2.2.7.5). These molecules have been anonymised, but were assigned identifiers which remain consistent throughout, and are listed in table 6.1 along with basic details.

ID	Structural Character	Approximate MW (kDa)
mAb 1	mAb	150
mAb 2	mAb	150
mAb 3	mAb	150
mAb 4	mAb	150
mAb 5	mAb	150
mAb 6	mAb	150
mAb 7	mAb	150
mAb 8	mAb	150
mAb 9	mAb + engineered Cys	150
mAb 10	mAb	150
mAb 11	mAb	150
mAb 12	mAb	150
mAb 13	mAb	150
mAb 14	mAb	150
mAb 15	mAb	150
mAb 16	mAb	150
mAb 17	mAb	150
mAb 18	mAb + engineered Cys	200
Bis 1	Bispecific (mAb + ScF_V)	200
Bis 2	Bispecific (mAb + ScF_V)	200
Bis 3	Bispecific (mAb + ScF_V)	200
Bis 4	Bispecific (mAb + ScF_V)	200
Bis 5	Bispecific (mAb + ScF_V)	200
Bis 6	Bispecific (mAb + ScF_V)	200
mAb 19	mAb	150
mAb 20	mAb	150
mAb 21	mAb	150

Table 6.1: Initial Monoclonal Antibody Panel

mAb refers to monospecific mAb, and Bis refers to bispecific mAb. ScF_v refers to single-chain F_v.

The mAb panel was first analysed by a series of techniques, intended to probe different developability features and cover a range of assays employed in industry. The developability assays were selected to span the different groups established by hierarchical clustering in the work by both Jain *et al*., (2017) and Willis *et al.* (manuscript in preparation), but using an adjusted roster of assays (as indicated in fig 6.1). Briefly, this included chromatography assays (HIC, SMAC, HP-SEC) which quantify retention time, and sample morphology in the case of HP-SEC, DSF measuring T_m values, a diffusivity assay (FIDA) measuring sample hydrodynamic radius and a diffusion interaction parameter (k_D), *in silico* tools (CamSol, TAP), which give scores informative of

aggregation propensity and solubility, and accelerated stability (at either 4 weeks at 40 °C or 2 weeks at 45 °C), measured by HP-SEC and quantified by change in sample morphology over time. The details of each assay are discussed further in the context of the results.

Baculovirus particle adsorption (BVP) and viscosity measured by rheology were omitted, both techniques being part of groups that were well-addressed by other assays (fig 6.1B(i). As the correlations observed at 1 mg/ml protein concentration were similar to those at 0.5 mg/ml, the latter concentration was investigated; low sample consumption is one of the benefits of considering EFD-induced aggregation as a potential developability parameter.

Of the total panel of 27 molecules, analysis by every assay was possible for only 16 (fig 6.1B(ii)). Removal of molecules was predominantly due to inability to obtain a quantifiable output parameter for a particular assay, but also resulted from lack of available material. Only molecules that could be analysed by all assays were used for the concluding analysis presented in this chapter.

When assay results were processed into a holistic developability parameter (HDP), only data from panel mAbs that had been analysed by every assay were used. As the HDP was produced with the aim of predicting molecule stability, a single metric was selected to indicate long-term stability. Accelerated stability (AS) is widely used to predict for long-term stability in a shorter timeframe, with lower material requirements. Although the timeframes vary, AS studies are frequently conducted over the course of weeks or months (Kuzman *et al.*, 2021). Change in sample monomer content was selected as the parameter indicative of molecule stability, after 4 weeks incubation at 40 °C; a routinely used AS protocol (Evers, Clénet and Pfeiffer-Marek, 2022). In this study, it is also the assay with the highest time and material requirements, so would be highly beneficial to predict for, provided acquiring the HDP score is less consumptive.

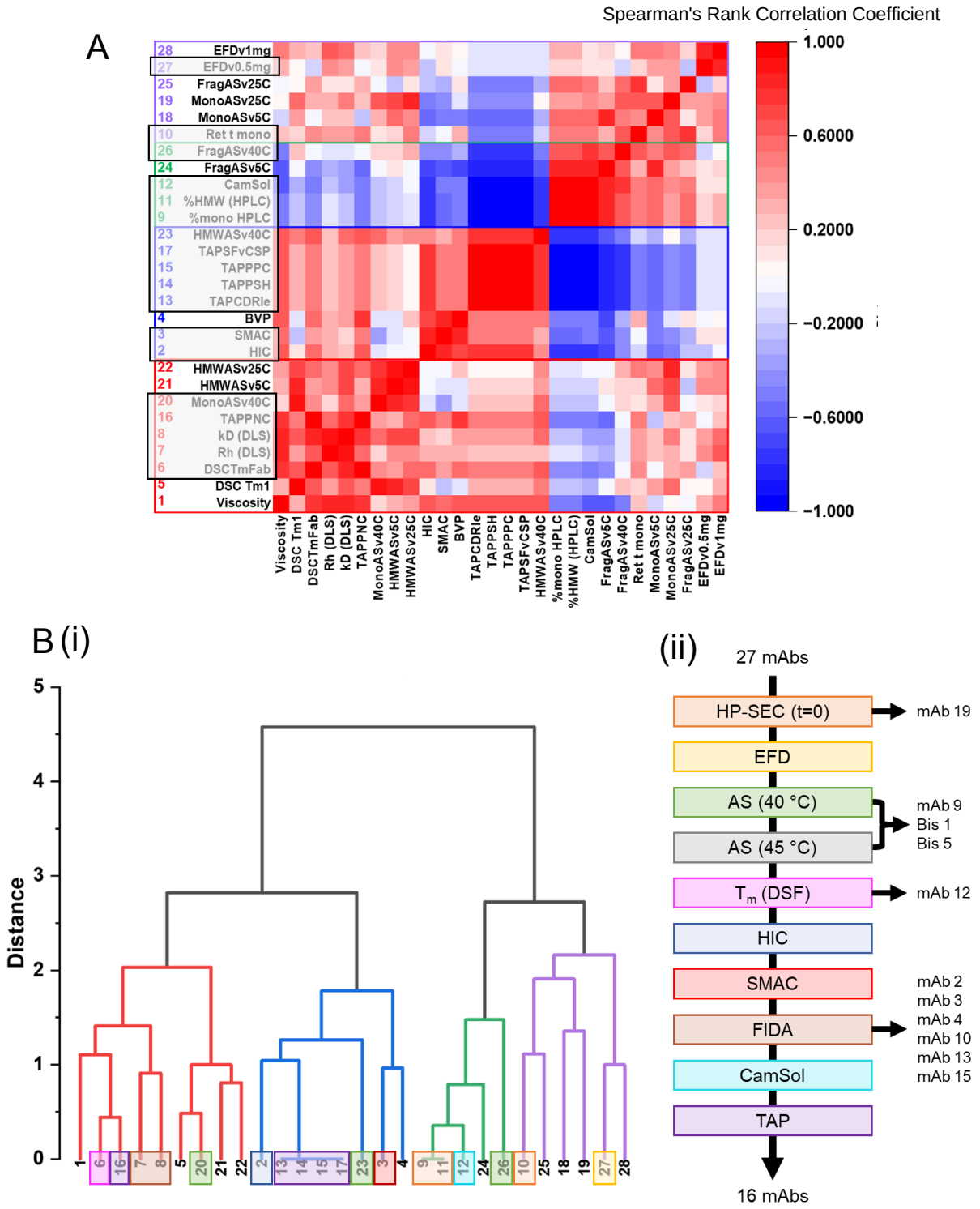


Figure 6.1: Developability Assay Selection

A) Spearman's rank correlation coefficient heatmap for assay comparisons taken from Willis *et al.* (manuscript in preparation), with assays selected for analysis of the mAb panel in this work highlighted (grey) B)(i) Assay groupings from (A). Assays selected for analysis are highlighted (coloured). (ii) Selected assays shown in order of completion, with removed mAbs indicated next to the step at which they were removed. The mAbs analysed by all methods were used for hierarchical clustering, to determine which assays are related to one another, and for generation of the holistic developability parameter (HDP), which was correlated with accelerated stability (AS) scores, to determine if the latter can be predicted for with the assay panel used.

6.2.2 HP-SEC Characterisation

Initial characterisation of the mAb panel was carried out using high-performance size-exclusion chromatography (HP-SEC), a technique frequently employed in developability studies and industrial application (section 1.3.4). Monomer retention time was measured, as well as monomer content (mono), higher molecular weight species (HMW) content, and fragment (frag) content as a percentage of total detectable Absorbance (280 nm) area.

Samples were analysed following clarification to remove impurities and any aggregate (section 2.2.8.2). All 27 mAbs were analysed by this assay, however mAb 19 was undetectable by HP-SEC, indicating the molecule may have aggregated to such an extent it was entirely removed by the clarification step. As such, it was excluded from further analysis.

Although the increased MW of the bispecifics compared to monospecifics (approximately 200 kDa vs 150 kDa) is a contributor to reduced retention time, this was not adjusted to allow data to remain unadulterated for analysis; the increased MW is an intrinsic characteristic of the bispecific molecules, and due to it not being the only factor that determines retention time, the effects of it would be difficult to isolate.

As a result, the bispecific mAbs exhibited the smallest retention times (fig 6.2A), and the majority of monospecifics associated with a roughly gaussian dispersion (fig 6.2B). However, mAbs 2, 4, 13 and 10 demonstrated longer retention times, indicating possible interactions with the column and heterogeneity in the mAb panel. Regarding monomer, HMW, and fragment content, data was more long-tailed with the majority population showing favourable characteristics, indicating a subset of the panel performed significantly worse than others (fig 6.3B). Remarkably, chromatograms for all bispecific mAbs contained solely monomer (fig 6.3A).

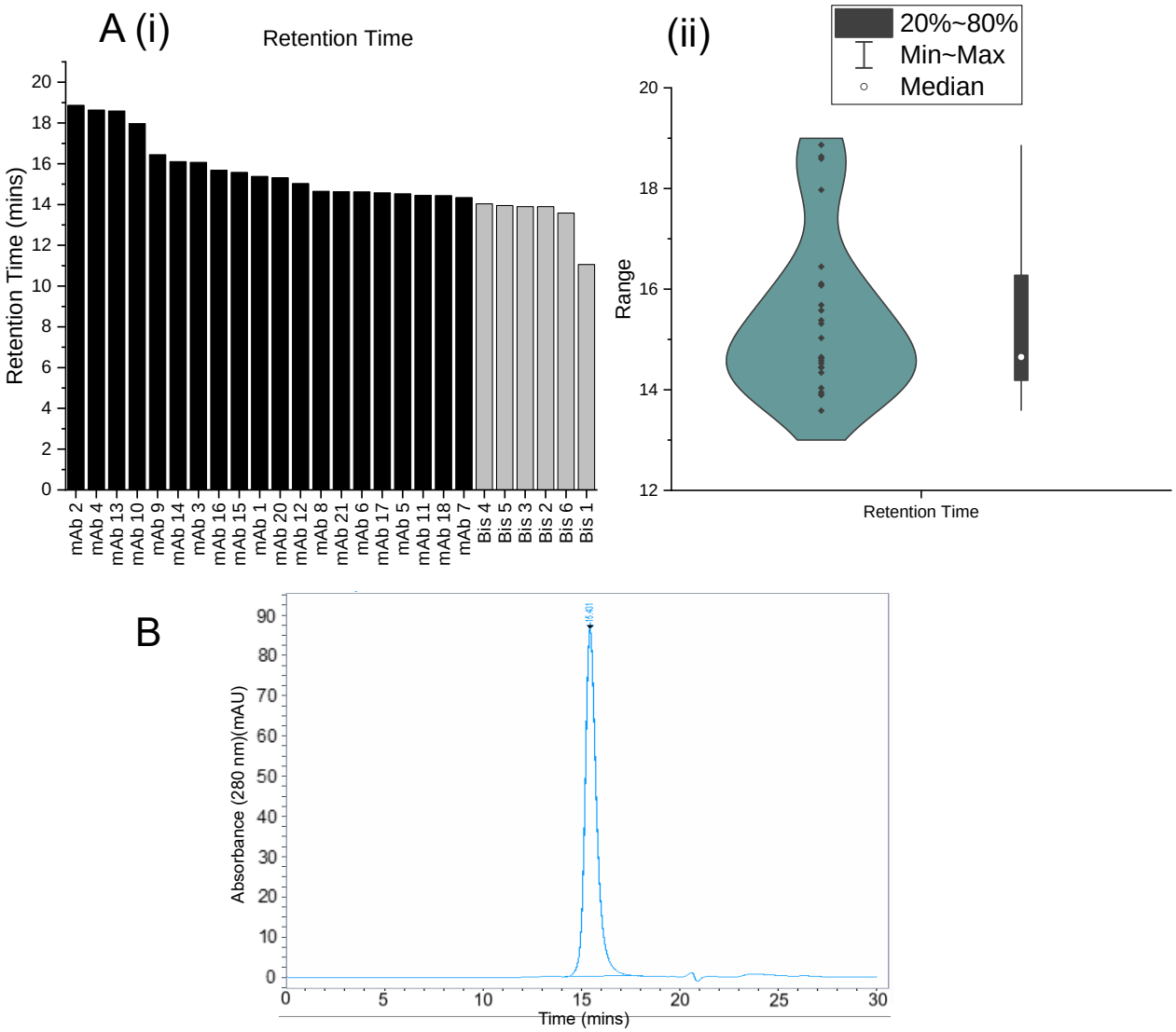


Figure 6.2: HP-SEC Retention Times of mAb Panel at t=0

0.5 mg/ml samples analysed by HP-SEC. A(i) Retention times given for monospecific (black) and bispecific (grey) mAbs measured at t=0. (ii) Violin and box distribution plots of the dataset. B) Representative HP-SEC chromatogram showing elution of mAb 1 and integration boundary selected for the main peak.

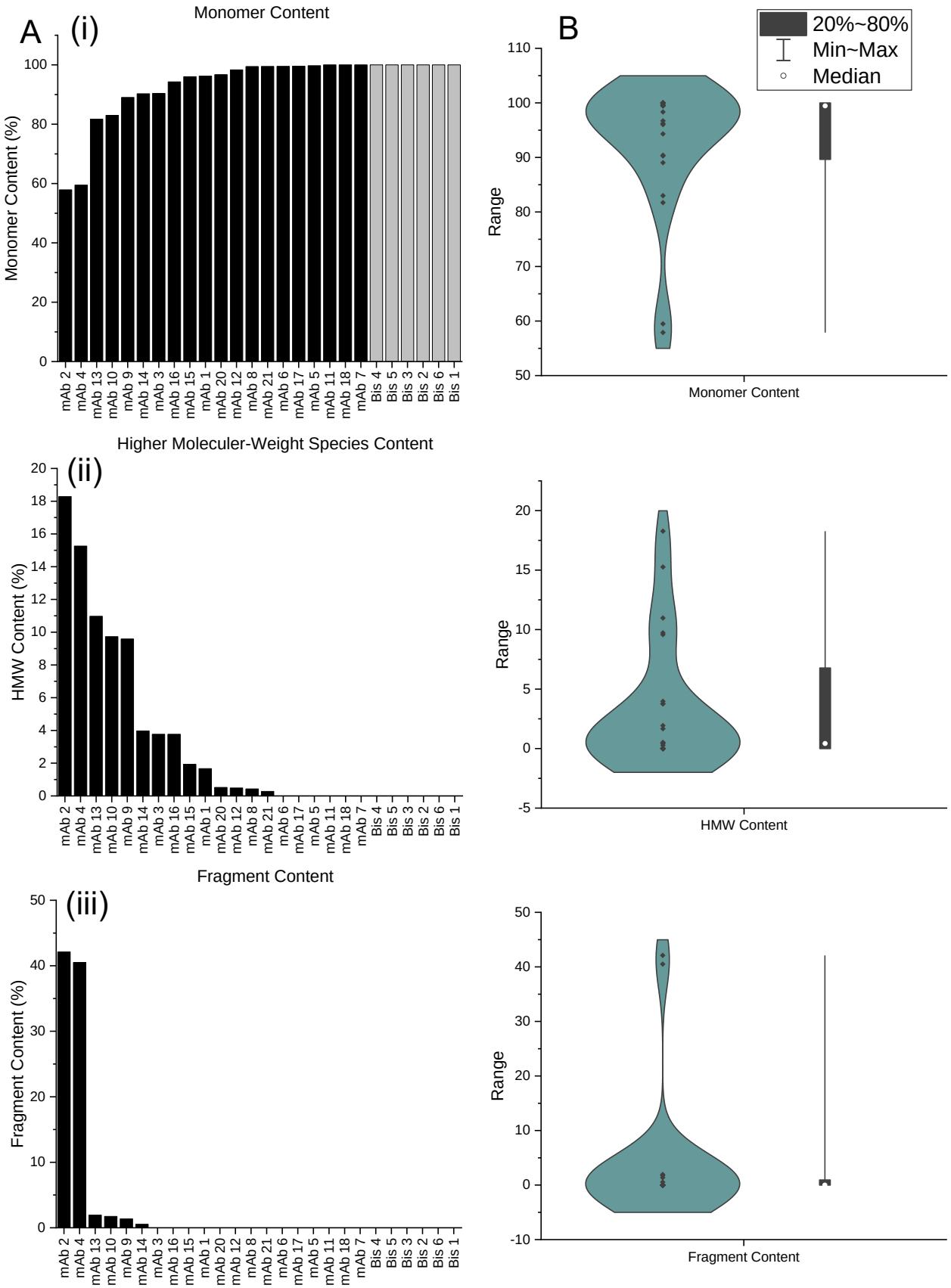


Figure 6.3: mAb Panel HP-SEC Analysis at t=0

0.5 mg/ml samples analysed by HP-SEC. (i) % monomer, (ii) % HMW and (iii) fragment content in monospecific (black) and bispecific (grey) mAbs measured at t=0. B) Violin and box distribution plots associated with each component.

6.2.3 Hydrodynamic Flow

The EFD was used first to assay for protein stability; initial rate of aggregation under flow was chosen as the developability parameter. This involved stressing 0.5 mg/ml sample at 8 mm/s plunger velocity for 10, 20, 50 and 150 passes. This relatively low level of hydrodynamic stress was chosen because it minimised experimental time, whilst still allowing detection of protein aggregation in a highly reproducible manner (fig 6.4B (inset)). A control sample was left quiescent at room temperature for the same duration as the 150 passes sample would be exposed to (16.5 minutes). 0.5 mg/ml was deemed sufficient for accurate detection by HP-SEC analysis, and minimised sample consumption; one iteration of this experiment consumed 1.25 mg of a given mAb. Before analysis by HP-SEC, samples were clarified to remove insoluble aggregate, to more clearly visualise monomer peaks (section 2.2.8.2).

Interestingly, EFD-induced monomer loss shows a long-tailed distribution, clearly illustrated by fig 6.4A(ii). This pattern indicates several mAbs behave in a much less desirable way than others, according to this assay, and is the type of trend observed in several of the assays selected by Jain *et al*., (2017). Whether or not this ranking provided by EFD-induced monomer loss is useful depends on its redundancy with other assays, which is discussed in section 6.3.2 and then 6.4.1. Based on the information available for each mAb, there doesn't appear to be a clear correlation between EFD-induced monomer loss and mono/bispecificity status, with the bispecific molecules being dispersed throughout the dataset.

Notably, bis 2, bis 5 and mAb 4 showed an apparent increase in monomer content at 20 and 50 passes (fig 6.4B). This was thought to be due to interactions with the column and low sample stability (clearly reflected in figs 6.2A and 6.3A, in the case of mAb 4), which convoluted integration of peaks. Using denaturing conditions and reverse-phase HPLC, such as that applied in the quantification of RNase H* would likely be a more suitable option for future analysis, as this would remove convolution by HMW species and fragments, with soluble protein content used as the metric for aggregation rather than monomer content (section 2.2.5.2). For the analysis in this chapter, the 3 mAbs mentioned above were the only molecules that exhibited these difficulties, and as the net slope of monomer change was negative in each, they were included in analysis with the expectation that they would cause a minor increase in error.

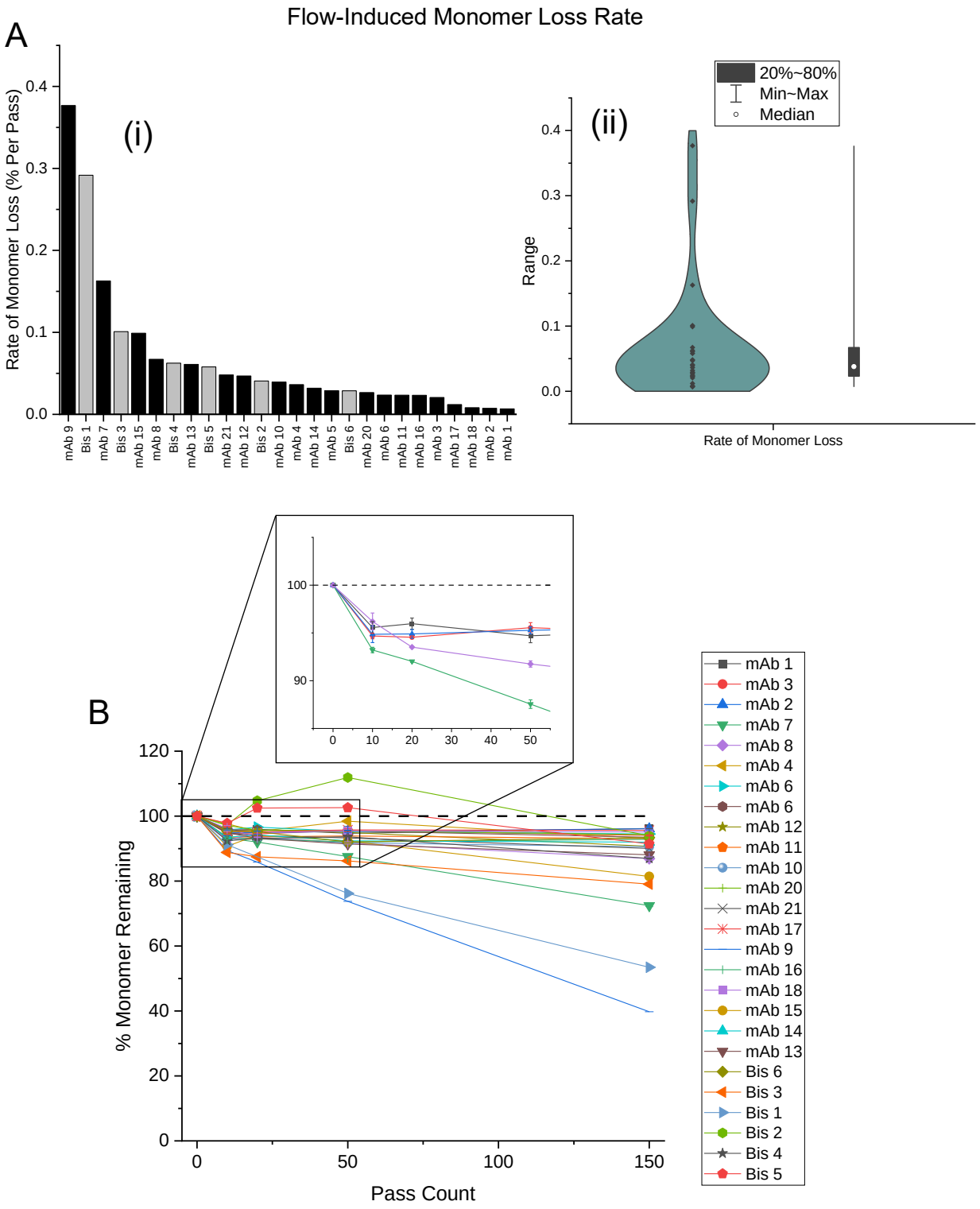


Figure 6.4: mAb Panel EFD-Induced Aggregation

A)i) Rate of monomer loss by monospecific (black) and bispecific (grey) mAbs. ii) Violin and box distribution plots of the dataset. B) Percentage soluble monomer remaining following exposure to flow at 8 mm/s for pass counts indicated. Inset: 5 mAbs shown with error bars for clarity. Error is standard deviation of the mean, n=2. Dotted guideline indicates 100% monomer remaining.

6.2.4 Accelerated Stability

Accelerated stability (AS) is frequently used in industry to predict long-term protein stability, and does not group closely with other assays when AS scores are hierarchically clustered with those resulting from other assays (figs 6.16 and 6.17). Samples were incubated at 0.5 mg/ml at 40 °C or 45 °C for 4 weeks or 2 weeks, respectively, before being clarified and quantified by HP-SEC (section 2.2.5.3). The temperatures and timeframes used were deemed to be industrially relevant, as suggested by the project's industrial collaborator and referenced in literature (Evers, Clénet and Pfeiffer-Marek, 2022).

As evaporation was thought to occur, a small concentrating effect was noticed in samples that had been incubated. Therefore, rather than measuring monomer content change from t=0 to t=assay completion, the proportions of total sample made up by mono, HMW, and frag (the sample components) were measured as in section 6.2.2. The change in relative content in each of these species was used as a separate metric of soluble monomer loss, aggregate formation, and mAb fragmentation, respectively. Bis 6 showed a small (8.4%) increase in apparent monomer content, which was attributed to difficulties integrating data, and so monomer change was fixed at 0 for this molecule. Bis 1 and mAb 9 produced chromatograms that could not be accurately integrated, indicating these molecules are highly unstable and possibly interact with the column. Both were excluded from further analysis.

Again, in the remaining samples no correlation was observable between percentage occupancy of the different sample components and mono/bispecificity. However, data appear long-tailed, albeit less dramatically than with hydrodynamic flow (section 6.2.3), perhaps indicating this technique offers greater potential to separate mAbs by developability.

Interestingly, it also appears that after both 4 weeks at 40 °C and 2 weeks at 45 °C, HMW content could both increase and decrease, depending on the mAb in question, whereas fragment content only increased with incubation. This is expected as once fragmented, mAbs would not reassociate under the same conditions, whereas HMW content may decrease through either disaggregation or by reaching a critical size where aggregates become excluded by the pre-HPLC clarification step (Vlasak and Ionescu, 2011).

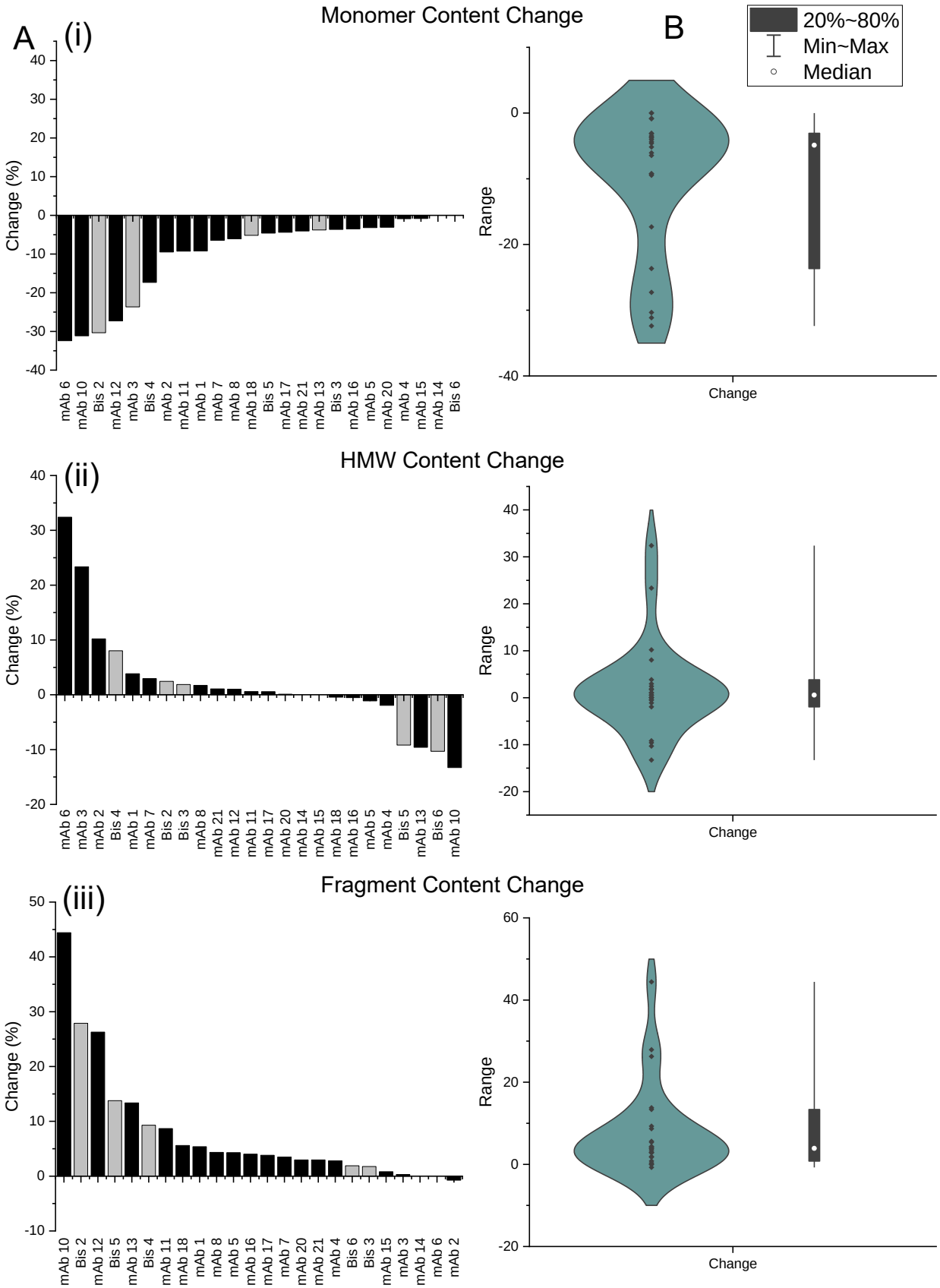


Figure 6.5: mAb Panel Accelerated Stability (40 °C, 4 weeks)

A) 0.5 mg/ml samples analysed by HP-SEC. Change in (i) % monomer, (ii) % HMW and (iii) fragment content in monospecific (black) and bispecific (grey) mAbs. B) Violin and box distribution plots associated with each component.

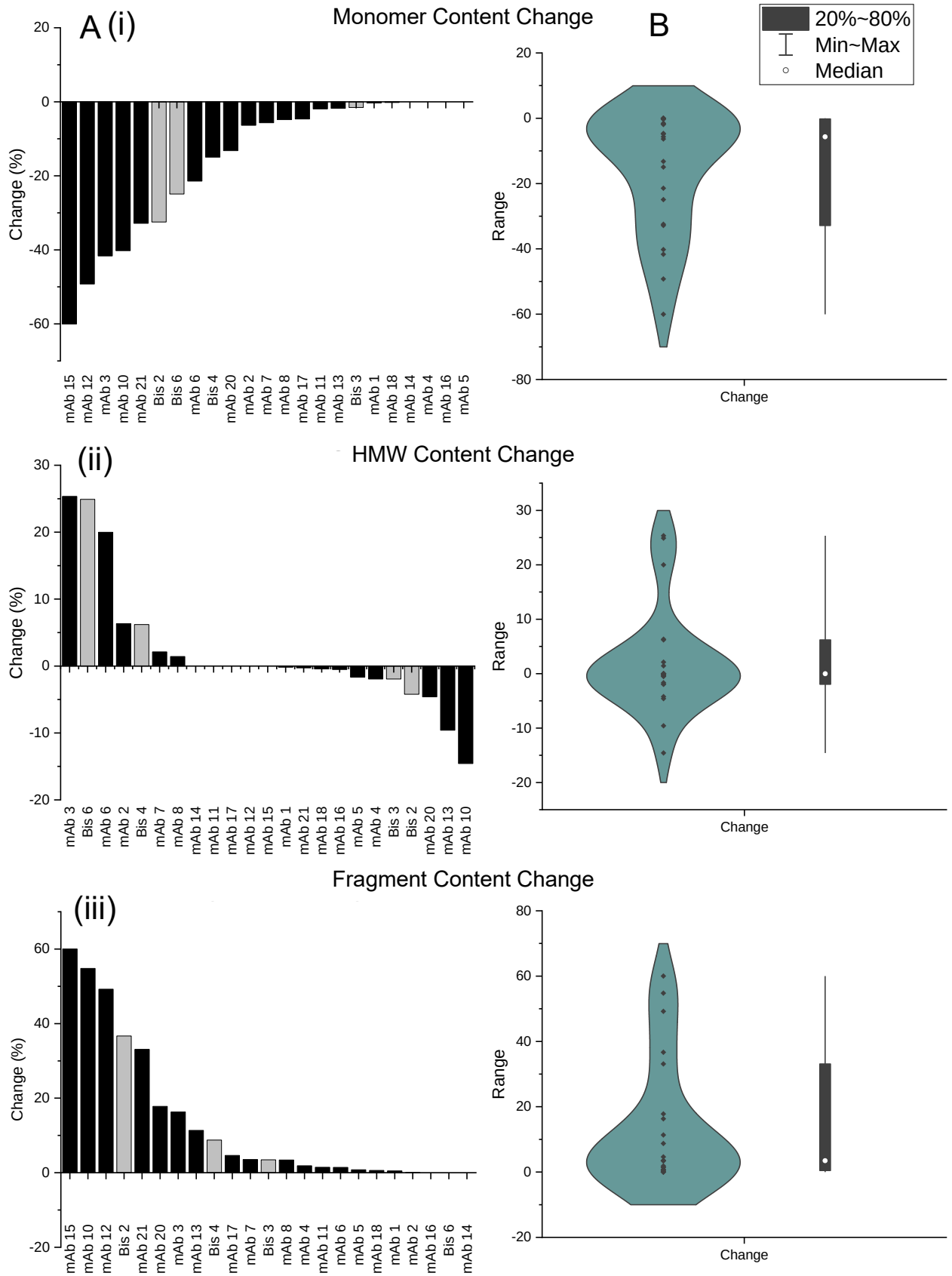


Figure 6.6: mAb Panel Accelerated Stability (45 °C, 2 weeks)

0.5 mg/ml samples analysed by HP-SEC. Change in (i) % monomer, (ii) % HMW and (iii) fragment content in monospecific (black) and bispecific (grey) mAbs. B) Violin and box distribution plots associated with each component.

6.2.5 Differential Scanning Fluorimetry

DSF was used to extract melting temperature (T_m) values for each mAb, with the major T_m transition taken as a developability parameter (section 2.2.8.6). This transition was expected to be denaturation of the F_{ab}, which generally results in a larger and better-defined transition than the F_C region, with a higher associated T_m (Tischenko *et al.*, 1982)(Niedziela-Majka *et al.*, 2015).

The majority of molecules showed T_m values within a symmetrical gaussian distribution of around 64-72 °C, but several were either side of this range, indicating that this popular developability parameter is able to identify a range of thermodynamic stabilities in the panel (fig 6.7B). Interestingly, the bispecifics all exhibited moderate thermal stabilities in relation to the rest of the dataset (fig 6.7A). This indicates that the 50 kDa addition of the single-chain F_v region to an IgG does not result in a remarkably low T_m, a promising observation for this increasingly popular modality. Unfortunately, mAb 12 was removed at this step due to lack of available sample material.

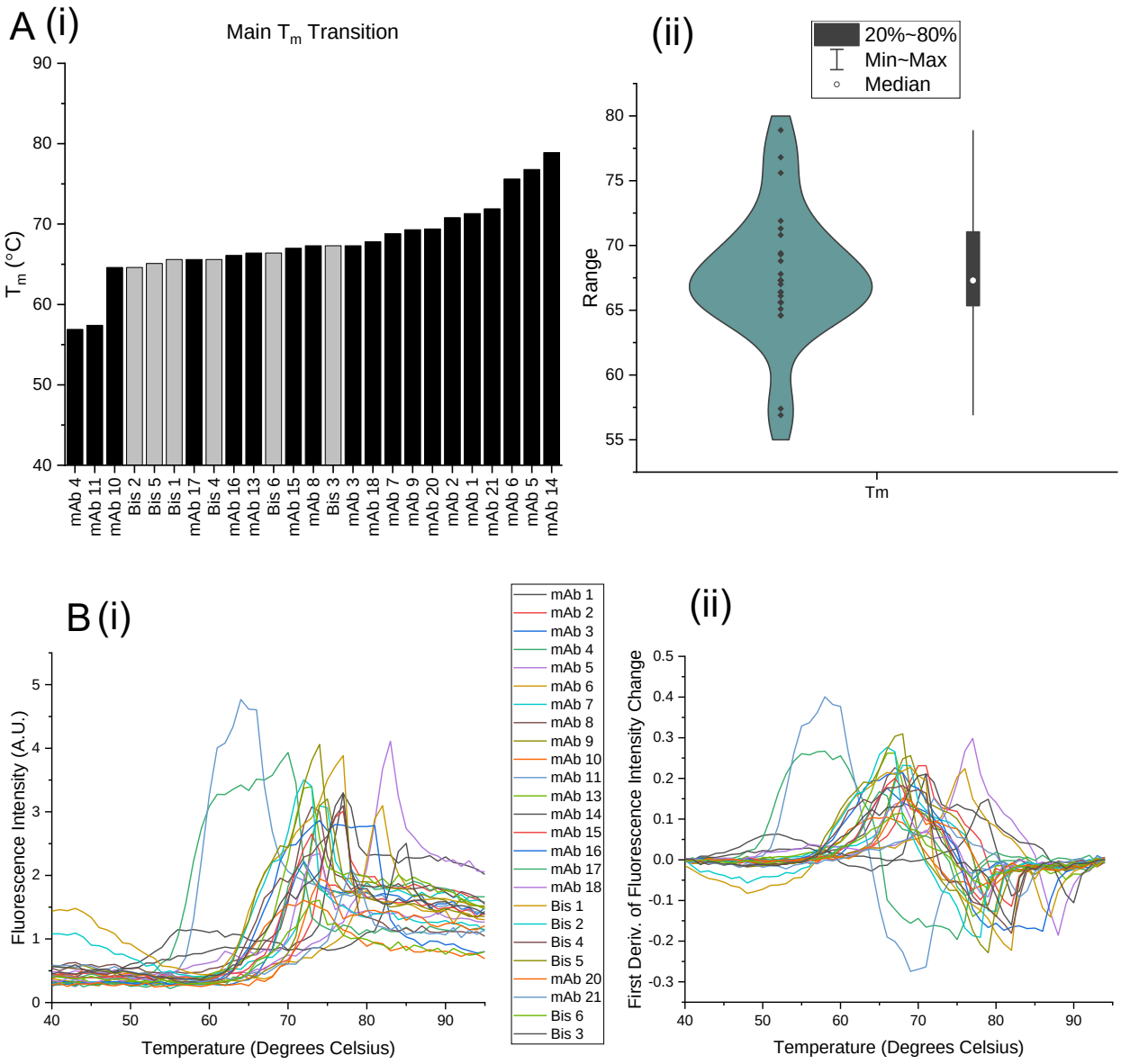


Figure 6.7: mAb Panel Differential Scanning Fluorimetry

A) Major T_m transition temperatures measured by DSF. (i) Data for monospecific (black) and bispecific (grey) mAbs. (ii) Violin and box distribution plots of the dataset. B)(i) Overlaid thermal melt curves showing absorbance (280 nm) change with temperature and (ii) first derivative of fluorescence intensity change with temperature.

6.2.6 HIC and SMAC

Both hydrophobic interaction chromatography (HIC) and stand-up monolayer adsorption chromatography (SMAC) were found in the same group when clustered in previous studies (Jain, Sun, *et al.*, 2017)(Willis *et al*., manuscript in preparation). Both techniques reveal the extent to which molecules in the mAb panel interact with a column, HIC through the protein's surface-exposed hydrophobicity and SMAC through interactions with a proprietary SEC resin. As such, retention time was collected for each molecule when analysed at t=0, with those that did not elute being assigned the full elution cycle time as the retention time (i.e. the least desirable result).

The spread of data is similar with HIC and SMAC both appears to be slightly two-tiered; the majority of mAbs in the panel showed a symmetrical gaussian distribution of 'good' scores relative to the dataset, whereas a smaller subset showed significantly 'poorer' relative scores (figs 6.8A(ii) and 6.9A(ii)). However, as several samples did not elute from the columns, these had to be assigned equal retention times, causing a reduction in detail of the dataset and meaning this subset of worse-performing molecules may not be similar in reality.

Interestingly, while 3/5 of the bispecifics did not elute from the HIC column, all bispecifics showed relatively low SMAC retention times. While the lack of information on the Zenix column used for SMAC makes detailed biochemical conclusions difficult, these results suggest the addition of the ScF_v in bispecifics can increase hydrophobic interactions and does not alter the type of interactions involved in a Zenix SEC elution. These results suggest that particularly in the case of bispecifics, these two techniques are non-redundant with each other, despite being closely related.

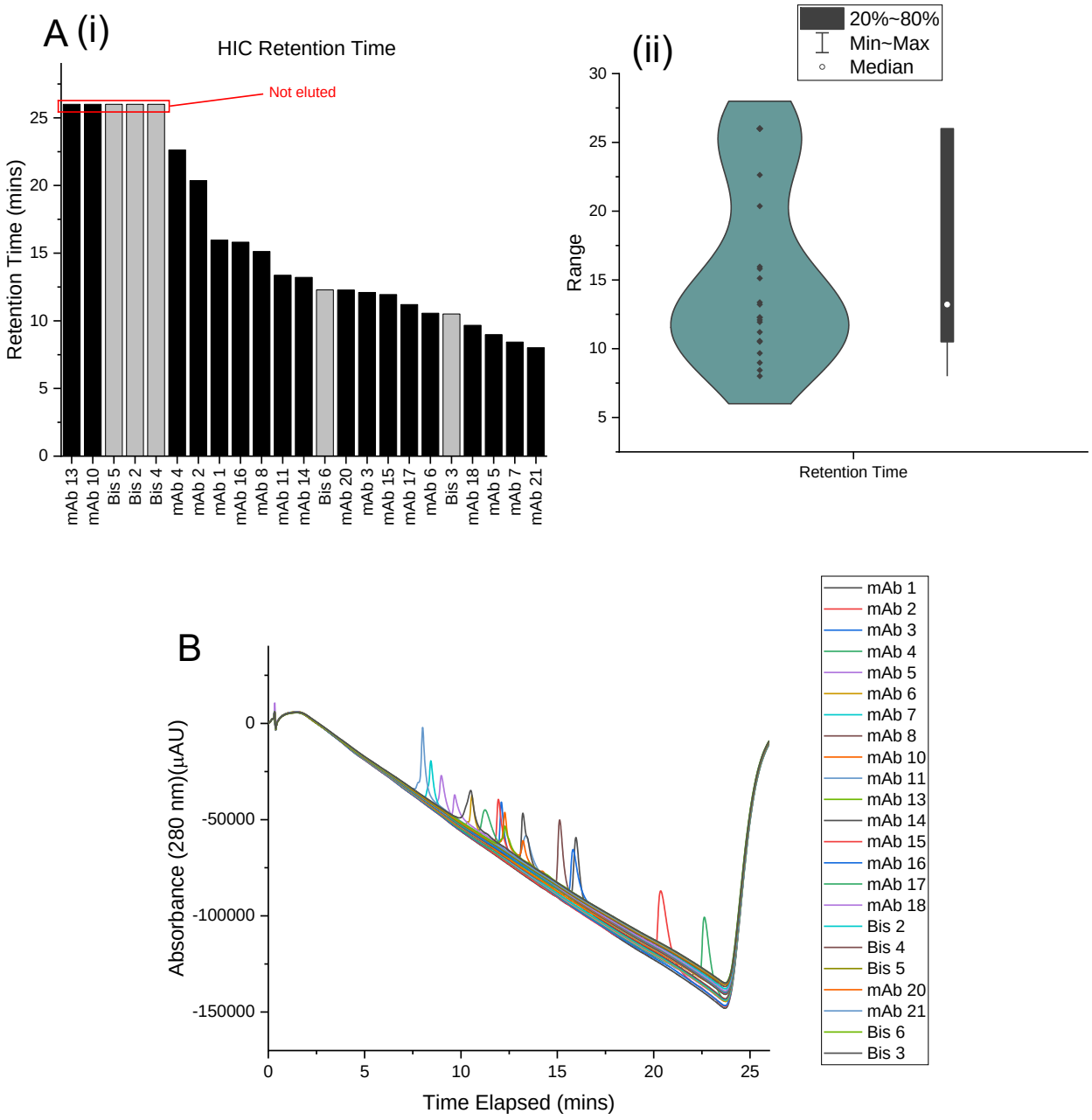


Figure 6.8: mAb Panel Hydrophobic Interaction Chromatography

A) HIC elution times of each mAb. (i) Data for monospecific (black) and bispecific (grey) mAbs. (ii) Violin and box distribution plots of the dataset. B) Overlaid HIC chromatograms.

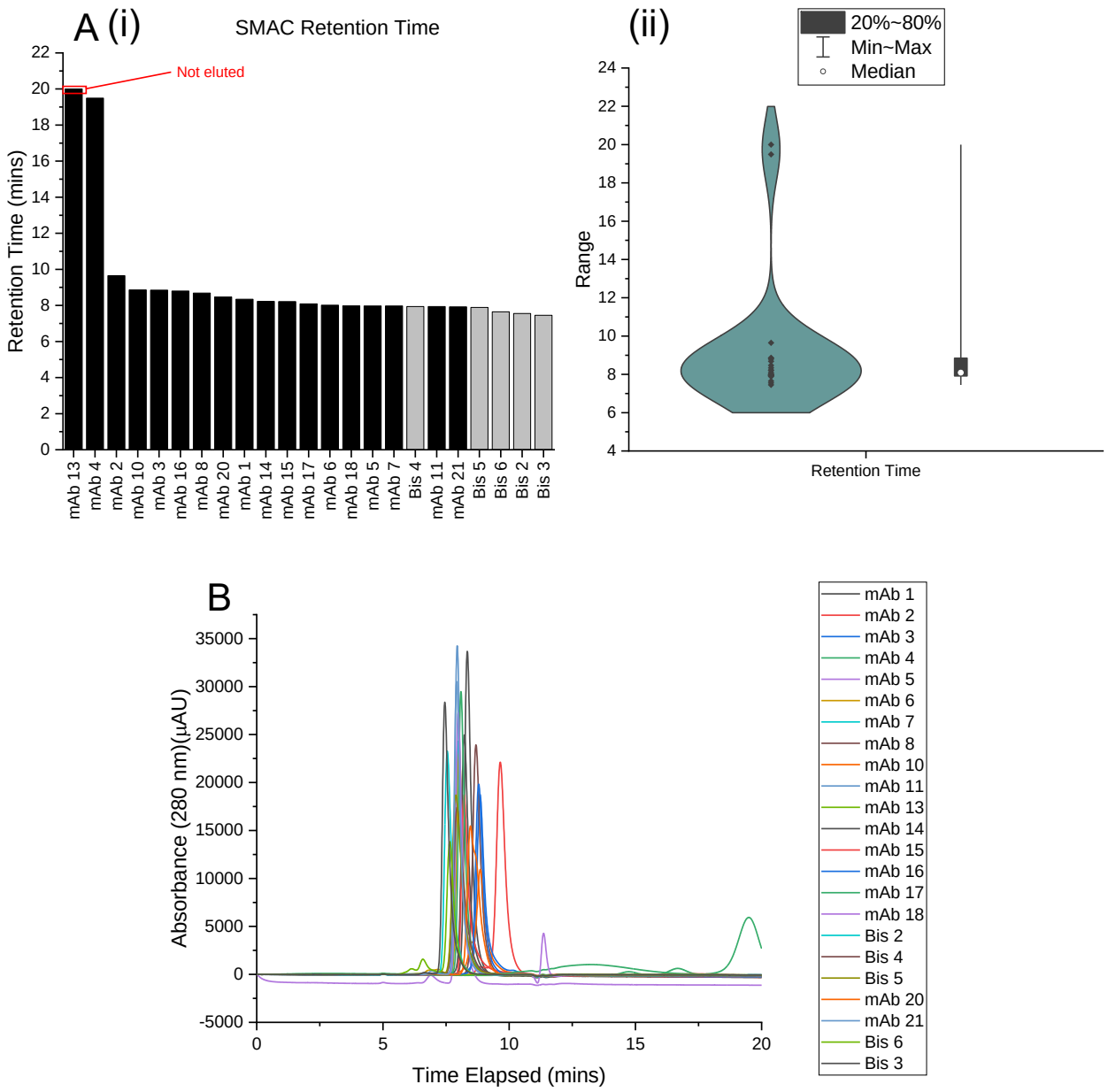


Figure 6.9: mAb Panel Stand-up Monolayer Adsorption Chromatography

A) SMAC elution times of each mAb. (i) Data for monospecific (black) and bispecific (grey) mAbs. (ii) Violin and box distribution plots of the dataset. B) Overlaid SMAC chromatograms.

6.2.7 Flow-Induced Dispersion Analysis

The diffusion interaction parameter (k_D) and hydrodynamic radius (R_H) of the mAb panel at t=0 were both used as developability parameters, as they are indicative of aggregation propensity and viscosity (Zhang *et al.*, 2023); k_D indicates how diffusion of a protein changes with concentration, thereby indicates the presence of protein-protein interactions, and hydrodynamic radius can change with propensity to form multimers and aggregates (Kingsbury *et al.*, 2020). Although normally measured using dynamic light scattering (DLS) to give diffusion coefficients from which the parameters can be calculated, flow-induced dispersion analysis (FIDA) was used as it is significantly more high-throughput and can inform on both these parameters while using a small amount of sample (5-200 ng)(section 2.2.8.7).

Due to either sample preparation issues or aggregation in the samples, several mAbs were not successfully analysed using this method. Additionally, only one concentration point for mAb 15 was available, meaning that whilst R_H was measurable, k_D could not be calculated. This reduced the total number of mAbs analysed by all methods to 16. The mAbs removed from the study at this step were mAbs 2, 3, 4, 10, 13 and 15.

Upon analysis by FIDA both k_D and R_H were dispersed across the dataset (fig 6.10B). Perhaps this is indicative of the broad range of attributes that contribute to protein-protein interactions, and the diversity of structural features in IgG mAbs. The bispecifics were reasonably dispersed throughout both datasets, with a slight tendency towards higher R_H, which is likely reflective of the additional ~50 kDa ScF_V unit.

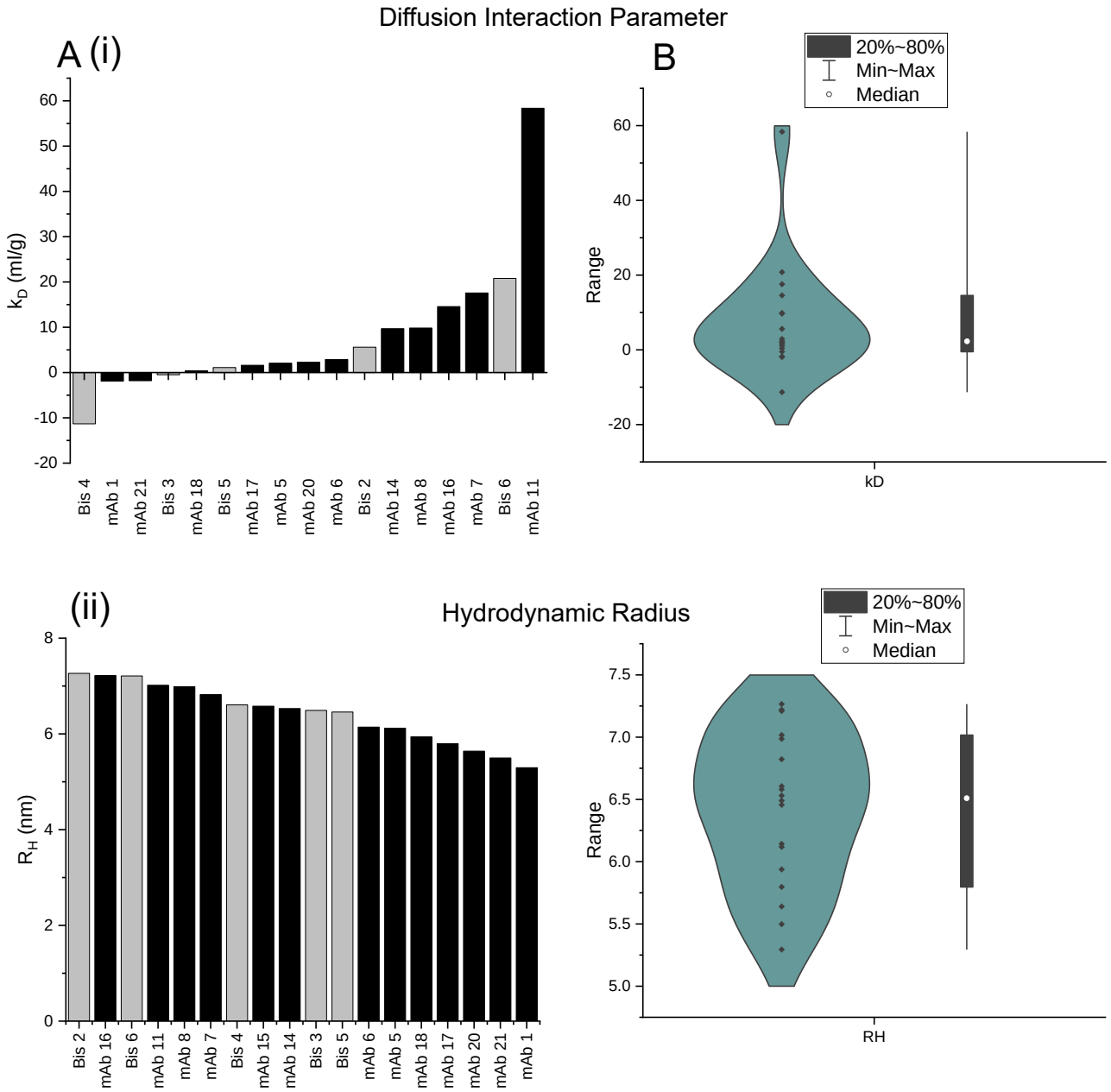


Figure 6.10: mAb Panel Flow-Induced Dispersion Analysis

FIDA analysis of mAb Panel. A) Calculated values for (i) diffusion interaction parameter and (ii) hydrodynamic radii of monospecific (black) and bispecific (grey) mAbs. B) Violin and box distribution plots of the datasets.

6.2.8 *In Silico* Analysis

In silico methods to probe developability are increasingly popular, consuming no material and often only requiring primary sequence information. These methods can compute structures and developability metrics using algorithms which are frequently open source (section 1.3.6).

For this study, CamSol and therapeutic antibody profiler (TAP) were employed (section 2.2.8.8). CamSol offers information on whole-protein solubility; a key consideration for developability. Using the primary sequence and a protein structure PDB file, CamSol assigns a score for solubility of each amino acid residue, which it 'smooths' based on the effects of 3 residues either side, to account for effects of neighbouring residues (Sormanni, Aprile and Vendruscolo, 2015). From this, it calculates an intrinsic solubility score based on the amino acid sequence of the protein of interest. This intrinsic solubility score is then adjusted based on the proximity of different residues in the 3D structure, producing a structurally corrected CamSol score, which was used as a developability parameter.

Using sequence data for mAb F_V regions, TAP produces information on 5 metrics associated with mAb F_v regions; total CDR length, patches of CDR surface hydrophobicity (PSH), patches of CDR positive charge (PPC), patches of CDR negative charge (PNC) and a structural F_v region charge symmetry parameter (SF_vCSP). Each of these measurements are given as a value that is relative to a reference database of mAbs which is continuously updated, and as of 30th January 2023 contained 644 entries. A high score in each metric aside from SF_vCSP is cited as being indicative of aggregation propensity, with a lower SF_vCSP indicating aggregation (Raybould *et al.*, 2019).

As discussed in section 6.2.1, the bispecific molecules are each comprised of different combinations of the same mAb and ScF_v, aside from Bis 6. As no differences were present in within the F_v regions, both computational methods were unable to distinguish between those of the same mAb/ScF_v origin. As these have been analysed independently throughout the chapter, they were left as independent datapoints, however the 'tightening' effects on the spread of data in these datasets should be noted.

The structure corrected CamSol scores were slightly long-tailed towards less desirable solubility, indicating several mAbs would be less soluble than others, but few would be noticeably more soluble (fig 6.11B). Interestingly, CDR length analysis by TAP showed a

very similar trend, with long-tailing towards less-desirable longer CDR regions; perhaps indicating the role of CDR length in solubility (fig 6.12B(i)).

CDR hydrophobicity scores showed a different trend, with two groups apparent (fig 6.12 B(ii)). However mAb 13, the variant with greatest CDR length and poorest CamSol score, also showed by far the greatest content of CDR hydrophobicity, indicating lack of solubility may be due to exposed hydrophobic patches. Charge-related TAP scores were far more divisive; with a minority subset performing significantly worse than the other molecules (fig 6.13B). The varied dispersion of data from TAP analysis is promising, as it indicates that five parameters analysed are able to differentiate molecules, rather than grouping them all together.

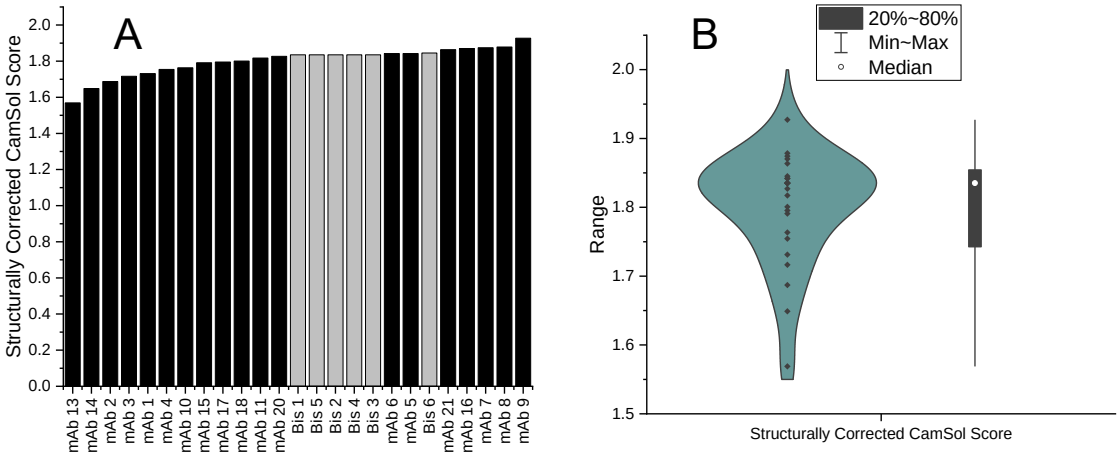


Figure 6.11: mAb Panel CamSol Analysis

A) Structurally corrected CamSol values calculated for monospecific (black) and bispecific (grey) mAbs. B) Violin and box distribution plots of the datasets.

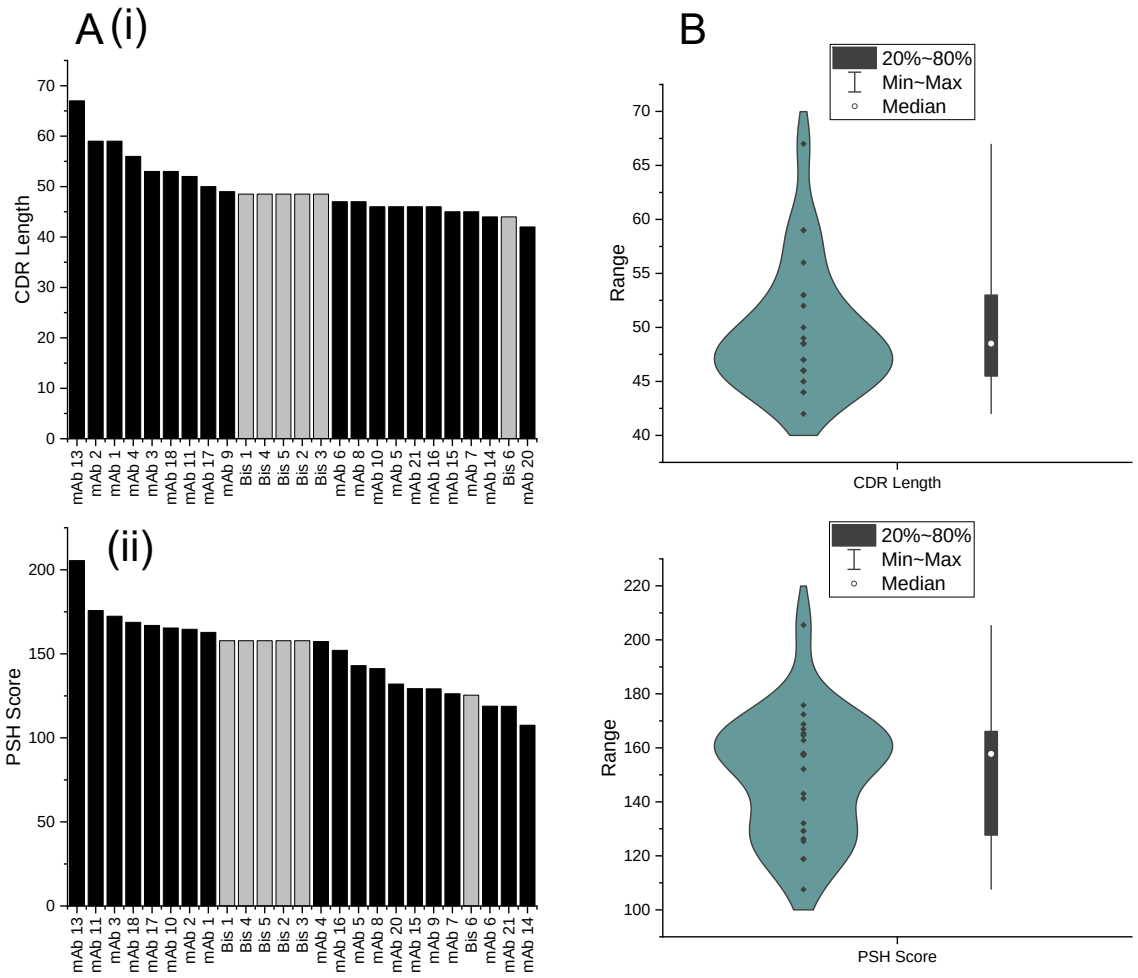


Figure 6.12: mAb Panel TAP CDR Length and Hydrophobicity Analysis

A) Therapeutic antibody profiler scores for (i) CDR length and (ii) PSH, calculated for monospecific (black) and bispecific (grey) mAbs. B) Violin and box distribution plots of the datasets.

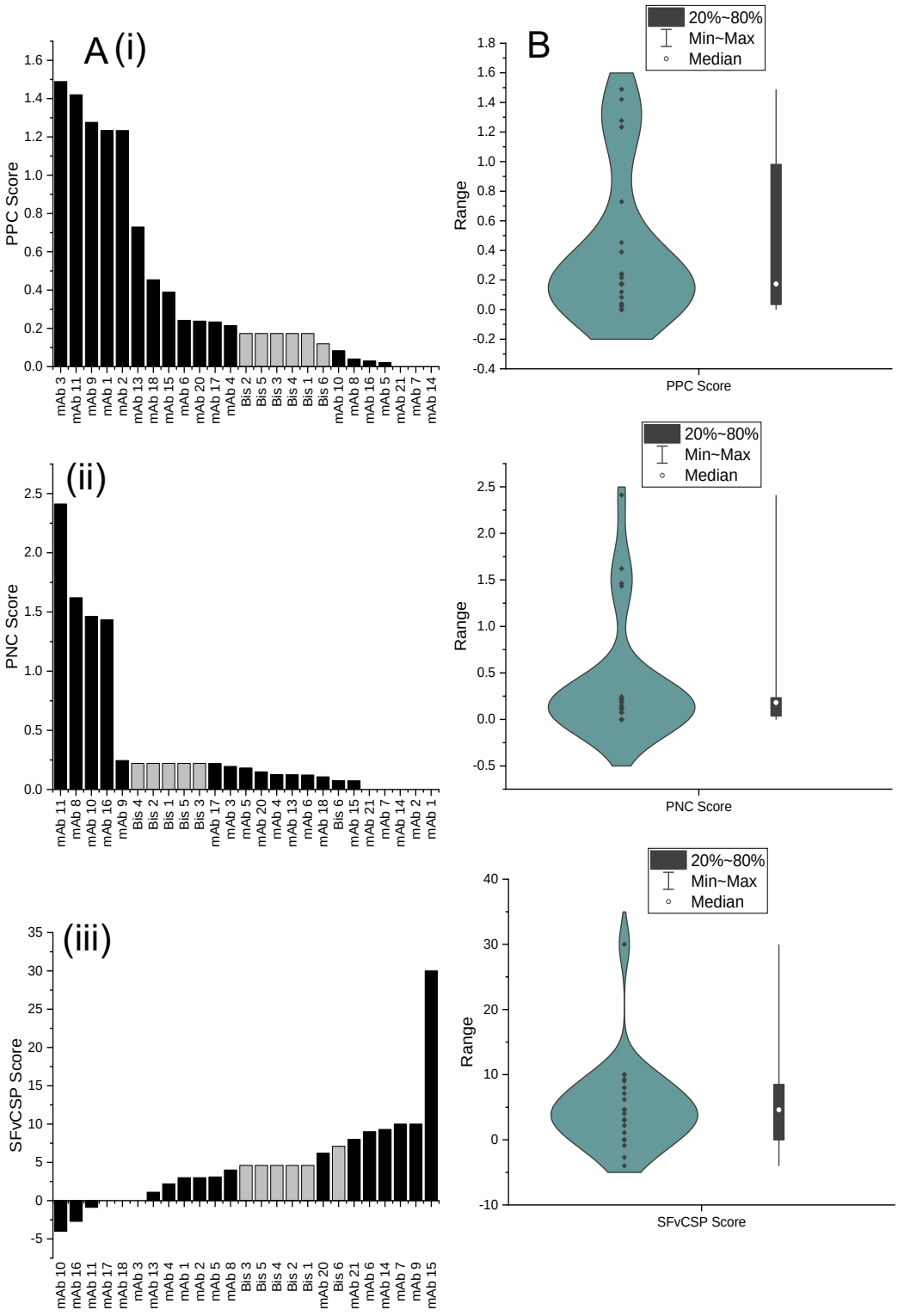


Figure 6.13: mAb Panel TAP Charge-Based Parameters

A) Therapeutic antibody profiler scores for (i) PPC, (ii) PNC and (iii) SFvCSP, calculated for monospecific (black) and bispecific (grey) mAbs. B) Violin and box distribution plots of the datasets.

6.3 Statistical Analysis of Developability Assays

The assays investigated were selected to probe a range of characteristics in the mAb panel, with the intention of establishing firstly how hydrodynamic flow correlates with other developability parameters, and secondly whether the ensemble of assays investigated can combine to give a single, holistic developability parameter.

6.3.1 Multivariate Analysis

To investigate how assays relate to one another, mAb performance in every assay was ranked, and ranked performance was compared between every possible assay pair, by Spearman's rank correlation (section 2.2.9.1), to produce a heatmap of assay correlation (figs 6.15A and 6.16A).

As discussed in section 6.2.7, mAbs 2, 3, 4, 10, 13 and 15 were not successfully analysed by FIDA. As this substantially reduced the size of the dataset, it was thought that further analysis should involve both the dataset with FIDA data without these mAbs (16 mAbs included), as well as the dataset without FIDA data but including these mAbs (22 mAbs included), with analysis done in tandem but kept separate. It was thought that inclusion of this assay could bias data because an unexpectedly high number of events occurred in which data had to be discarded, for example suspected inclusion of air bubbles in samples. For clarity, each assay conducted and the associated developability parameters, groupings and numbers are provided in table 6.2.

Correlation coefficients resulting from this analysis were used to hierarchically cluster assays, to identify those which rank mAbs similarly (figs 6.15B and 6.16B). Similarly to work by colleagues on a slightly larger assay selection (Willis *et al*., manuscript in preparation), 4 assay groups were distinguishable.

As expected, developability parameters from the same assay family correlate well with one another; parameters resulting from TAP correspond together, with the same trend observed for AS and t=0 measurements by HP-SEC. Additionally, other logical relationships are noticeable; Retention time of monomer at t=0 correlates well with SMAC retention time, both SEC-based techniques measuring interactions of the protein of interest with column matrices at t=0. Interestingly, EFD-induced aggregation does not

group with assays that reflect intrinsic biophysical properties of proteins, an observation also made by Willis *et al*., (2020). This is perhaps due to the high complexity of flow-induced aggregation, which is thought to be dependent on a mixture of excursions from the native state and protein-protein interactions, which are modulated by factors such as mechanical stability, which are not considered in these assays. Instead, EFD-induced aggregation was found to correlate slightly with several accelerated stability parameters, which indicate more complex degradation.

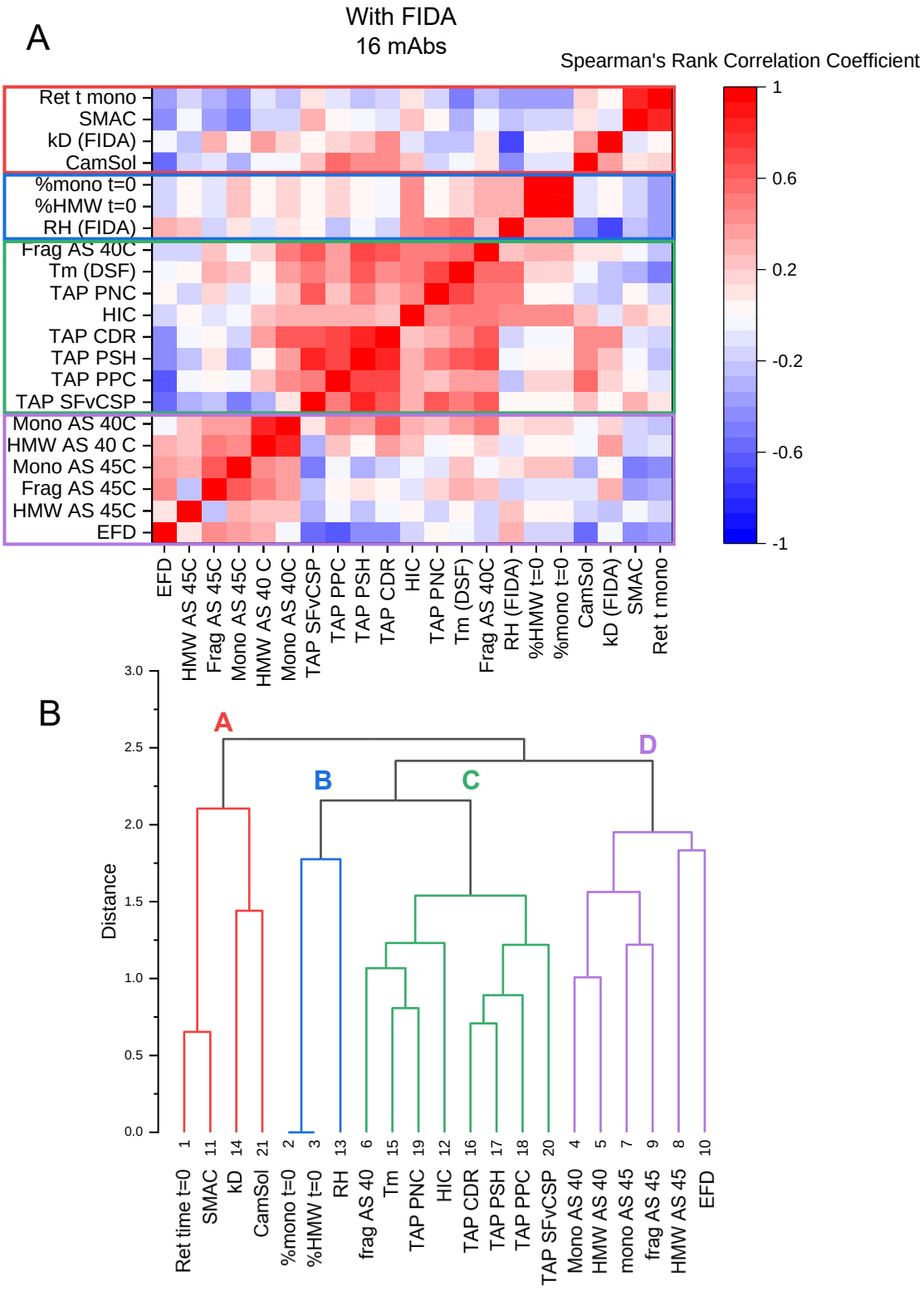


Figure 6.15: Correlation and Clustering of Assays Measuring Developability Parameters with FIDA Data

A) Heatmap showing pairwise Spearman's Rank Correlation Coefficient of mAb panel rankings between assay pairs. B) Hierarchical clustering of correlation coefficients for each assay identifies four clusters (A-D). Different assay groups indicated by different branch colours and boxes highlighted in (A). 16 panel mAbs analysed. Abbreviations are defined in table 6.2.

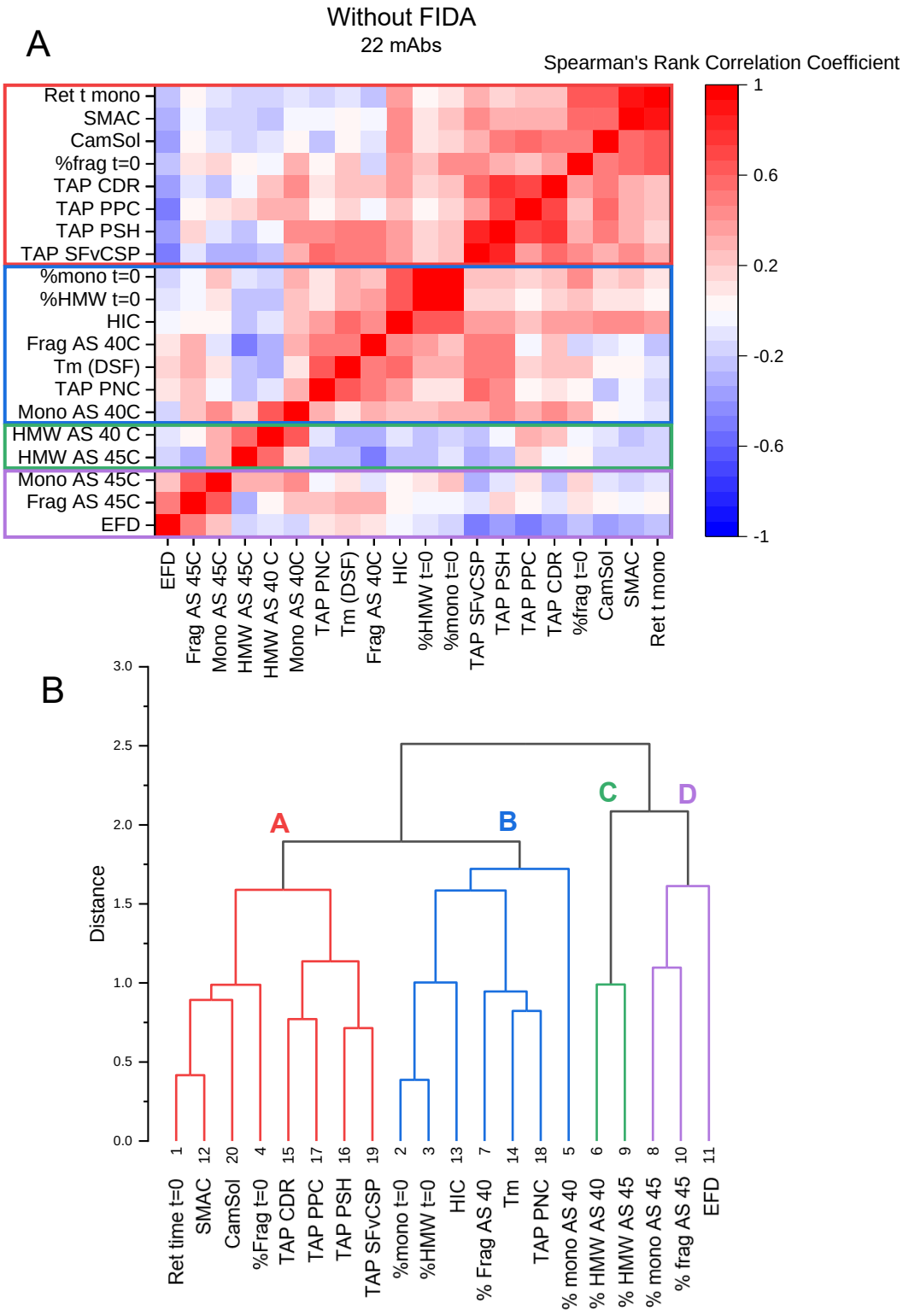


Figure 6.16: Correlation and Clustering of Assays Measuring Developability Parameters without FIDA Data

A) Heatmap showing pairwise Spearman's Rank Correlation Coefficient of mAb panel rankings between assay pairs. B) Hierarchical clustering of correlation coefficients for each developability parameter identifies four clusters (A-D). Different assay groups indicated by different branch colours and boxes highlighted in (A). 22 panel mAbs analysed. Abbreviations are defined in table 6.2.

Assay	Developability Parameter	Shortened Name	Parameter ID		Assay group	
			With FIDA	Without FIDA	With FIDA	Without FIDA
HP-SEC	Retention time of monomer at t=0 (mins)	Ret t mono	1	1	A	A
	% sample occupied by monomer at t=0	%mono t=0	2	2	B	B
	% sample occupied by high-molecular weight species at t=0	%HMW t=0	3	3	B	B
	% sample occupied by fragment at t=0	%frag t=0		4		A
Accelerated stability measured by HP-SEC	% change in monomer content from t=0 after 4 weeks at 40 °C	Mono AS 40	4	5	D	B
	% change in higher molecular weight species content from t=0 after 4 weeks at 40°C	HMW AS 40	5	6	D	C
	% change in fragment content from t=0 after 4 weeks 40 °C	Frag AS 40	6	7	C	B
	% change in monomer content from t=0 after 2 weeks at 45 °C	Mono AS 45	7	8	D	D
	% change in higher molecular weight species content from t=0 after 2 weeks at 45°C	HMW AS 45	8	9	D	C
	% change in fragment content from t=0 after 2 weeks 45 °C	Frag AS 45	9	10	D	D
EFD-induced aggregation measured by HP-SEC	Monomer loss rate (slope)	EFD	10	11	D	D
SMAC	Retention time (mins)	SMAC	11	12	A	A
HIC	Retention time (mins)	HIC	12	13	C	B
FIDA	R_H (nm)	RH (FIDA)	13		B	
	k_D (ml/g)	KD (FIDA)	14		A	
DSF	T_m (F_{ab})	Tm	15	14	C	B
TAP	CDR length score	TAP CDR	16	15	C	A
	Patches of CDR surface hydrophobicity score	TAP PSH	17	16	C	A
	Patches of CDR positive charge score	TAP PPC	18	17	C	A
	Patches of CDR negative charge score	TAP PNC	19	18	C	B
	Structural F_v charge symmetry parameter	TAP SFvCSP	20	19	C	A
CamSol	Structurally corrected CamSol Score	CamSol	21	20	A	A

Table 6.2: Assays Used in this Developability Study

Assays used are listed, with corresponding developability parameters produced, shortened names and associated numbers/clustered groups.

6.3.2 Flow-induced Aggregation and AS

As the EFD is a unique instrument not previously used in commercial developability, the positive correlation coefficient with monomer loss and fragment formation at 45 °C (Spearman's ρ = 0.35/0.22 and 0.44/0.50, respectively, with/without FIDA data) is noteworthy. Experimental techniques that are indicative of AS performance, while taking a fraction of the time, would be highly beneficial.

By clustering analysis, this relationship is confirmed by grouping of EFD with AS developability parameters in both datasets (i.e. with and without FIDA data; figs 6.15B and 6.16B). However, correlation between these assays is generally weak, denoted by the clustering distance parameter branch length.

Previous work by colleagues Willis *et al*. (manuscript in preparation) using a selection of 9 mAb/formulations, showed that parameters associated with EFD-induced flow correlated more with longer-term stability at 5 and 25 °C (up to 6 and 18 months respectively) than AS at 40 °C for a shorter timeframe (up to 3 months)(fig 6.1A). For the Willis dataset, Spearman's rank correlation coefficients were 0.40 and 0.35 when EFD was compared to monomer loss from AS at 5 and 25 °C respectively, versus -0.27 at 40 °C.

This suggests that while EFD performance may be indicative of AS performance at lower temperatures for longer timeframes, compared to the assays in this chapter it is more effectively used as an independent assay for characterising stability against hydrodynamic flow, a unique stress that is not currently assayed for in developability studies. The fact that it can distinguish between panel mAbs that perform well and those that perform poorly (fig 6.4) supports its application as a developability parameter. The poor correlations observed between EFD-induced aggregation and measures of thermodynamic stability in RNase H* (section 3.3.2) support the idea that this method probes a completely different facet of stability, and therefore could be considered as part of the developability toolkit. This is reinforced by results from the mAb panel when T_m is compared to EFD performance (fig 6.14).

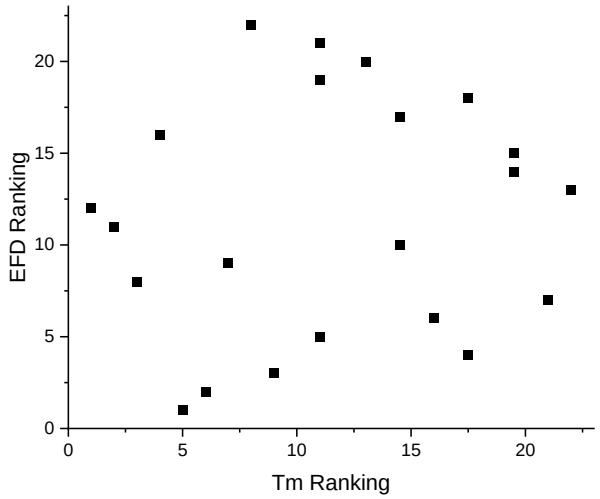


Figure 6.14: Flow-induced Aggregation Performance Does Not Correlate with F_{ab} T_m

mAb panel ranking of EFD-induced monomer loss shown versus ranking of F_{ab} T_m measured by DSF.

6.3.3 Predicting Developability Using a Holistic Parameter

As shown by Jain *et al*., (2017), Willis *et al*. (manuscript in preparation), and throughout this chapter, most of the investigated developability assays contain a subset of molecules which perform significantly more poorly than the mean; often exhibited by a long-tailed data distribution. In the case of Jain *et al*., the 137 molecules investigated had all surpassed at least phase II clinical trials, indicating that poor performance in individual assays does not wholly determine overall likelihood of clinical success. The authors noted that a larger number of successive poor scores in assays, so called 'red flags' reduced the likelihood of a biopharmaceutical to be approved. Therefore, a holistic developability parameter (HDP) which encapsulates a range of assays into one score may be more informative.

Additionally, if the HDP correlated with accelerated stability (AS), it could be used to de-risk this developability parameter, which can be time and material-consuming (section 1.3.5). Consequently, the predictive power of the HDP was investigated. The intended use for calculated HDP scores is to accurately predict a biopharmaceutical's developability. As the panel of mAbs used must remain anonymised, the HDP scores cannot be correlated with known developability issues, however as described in section

(6.2.1), monomer loss after 4 weeks at 40 °C (%mono AS 40C) was chosen as an AS metric that is routinely used in industry.

This score should comprise assays that are non-redundant; the assays chosen for this chapter were selected based on the groups established by hierarchical clustering carried out by Willis *et al*. (manuscript in preparation). Assays present in each of the 4 reported groups were selected.

To calculate HDP, a normalised score was calculated for every developability parameter measured in each panel mAb, with all calculations explained in detail in section 2.2.9.2. The score ranges from 0-1, with 0 being closest to ideal and 1 being furthest away from ideal. The average scores for each assay group were established, and then these were averaged over the 4 groups, to give a single 'naïve HDP score', where the score is calculated by weighting each assay equally.

As discussed above, analysis was calculated both with FIDA developability parameters R_H and k_D, and without. For both examples with and without FIDA, percentage of sample occupied by fragment at t=0, measured by HP-SEC (%frag t=0) could not be included, as the number of mAbs in the panel containing fragment at t=0 was too low to create a ranked series suitable for analysis.

Interestingly, panel mAbs with higher naïve HDP scores are those that could not be measured by FIDA; perhaps indicating that good molecule stability is prerequisite for analysis by FIDA. A technique with high rank separating power is beneficial, but only if it can analyse a broad array of candidates. Interestingly, all 4 of the mAbs that had shown any fragment content at t=0 by HP-SEC (fig 6.3A(iii)) and reached the FIDA stage of the study (mAbs 2, 4, 10 and 13), could not be analysed by FIDA. This indicates that whilst stringent, the selectivity shown by this assay may be informative in terms of excluding the most unstable candidate molecules.

Without FIDA, a clear range of HDP scores is present in the mAb panel, with a broader distribution than many of the developability assays (fig 6.17). This indicates the HDP could be used to effectively rank molecules, removing the precedence of markedly 'bad' or 'good' scores in individual assays. With FIDA included, a more monodisperse trend is observed, but with a smaller range (with the exception of bis 2).

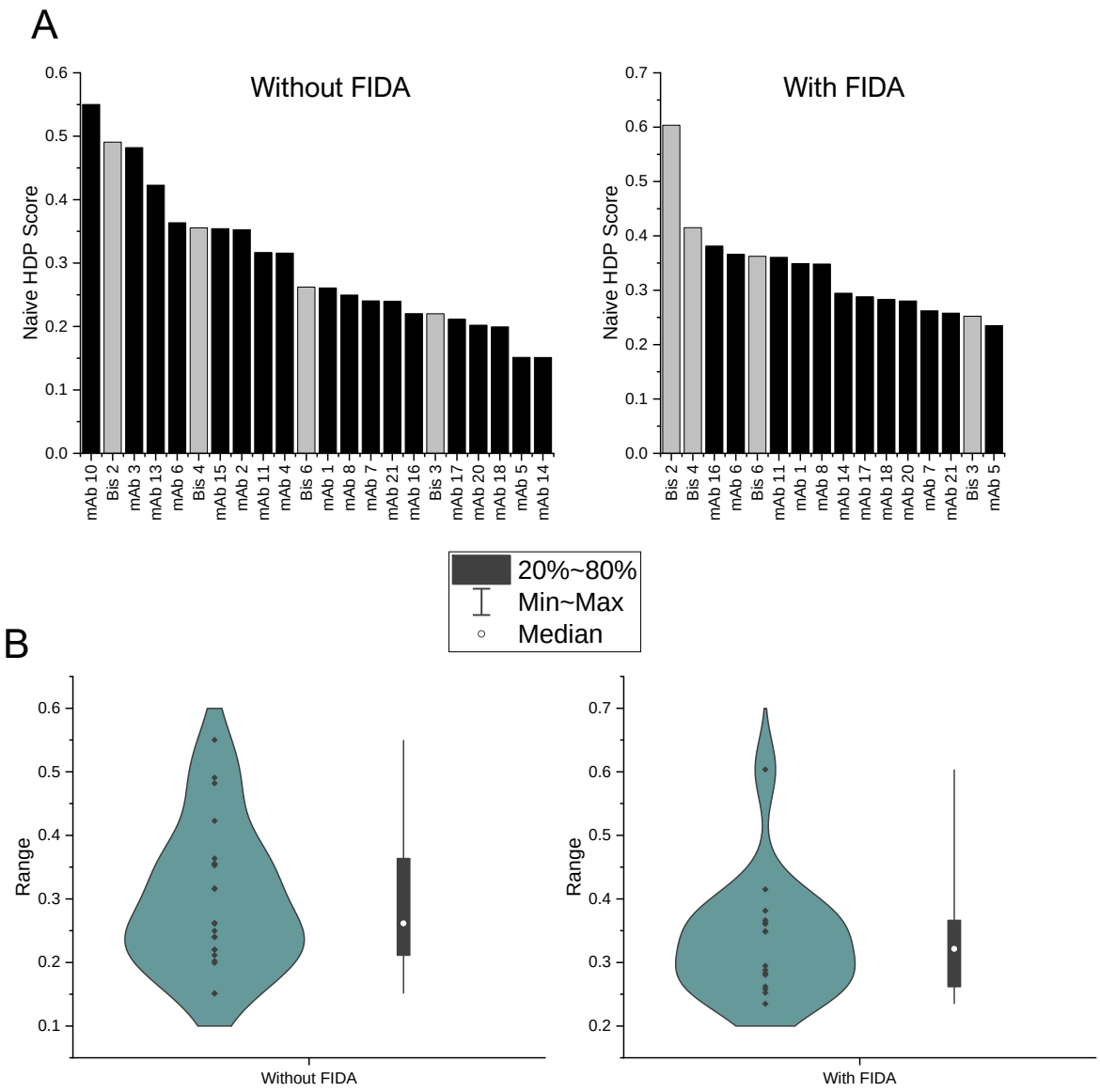


Figure 6.17: mAb Panel Naïve HDP Scores

A) Naïve holistic developability parameter scores are given for mono (dark) and bispecific (light) antibodies with and without FIDA scores included in the analysis. B) Violin and box distribution plots associated with each dataset.

6.3.4 Predicting Accelerated Stability with a Holistic Developability Parameter

To ascertain whether a HDP calculated from developability parameters used throughout this chapter would correlate with %mono AS 40C, HDP scores were first calculated for each panel mAb: A) without any 40 °C AS data, and B) without any 40 or 45 °C AS data. The latter being a truer reflection of whether the selected parameters could predict mAb stability from rapid (<1 day) developability assays.

To calculate HDP as described above (section 6.3.3), after removal of AS data developability parameters must retain the same groupings established by hierarchical clustering (figs 6.15/6.16). Recalculation of the Spearman's rank correlation coefficients and groupings by hierarchical clustering showed that with FIDA included, the developability parameters grouped differently, whereas without FIDA, they remained the same. This was true of both datasets excluding 40 °C AS data (fig 6.18) and datasets excluding both AS datasets (fig 6.22). HDP was therefore calculated without inclusion of FIDA data, in both instances.

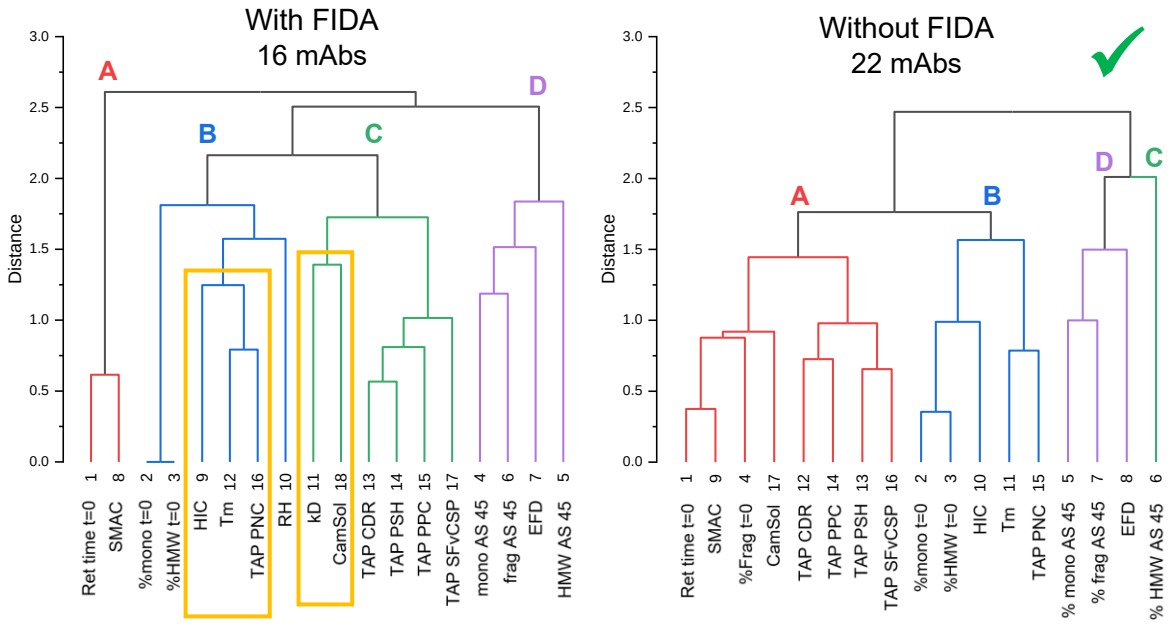


Figure 6.18: Clustering of Assays Measuring Developability Parameters without 40 °C Accelerated Stability Data

Hierarchical clustering of correlation coefficients for each developability parameter, with and without developability parameters from FIDA. Different assay groups indicated by different branch colours. With FIDA included, groupings of assays indicated by yellow boxes change compared with figure 6.15B.

Calculated HDP values were ranked and correlated with ranked %mono AS 40C (fig 6.19), to test whether the score could be predictive of mAb stability. Whilst a clear positive correlation is observed in most of the dataset, mAb 4, 13, 15 and bis 6 appear to belong to a distinct population, with high relative stability at 40 °C. To determine if this subset formed a grouping of mAbs with similar properties, antibodies were clustered according to developability attributes (fig 6.20A), with measures of intrinsic properties (CamSol scores, T_m values and all TAP scores) also considered (fig 6.20B). Clustering of mAbs did not reveal similarities, either by analysis of all developability parameters measured, or by developability parameters related to intrinsic characteristics of the molecules. This suggests that whilst the majority of molecules in a dataset may conform to the relationship of HDP vs AS score, outliers are not predictable or explicable with the selection of assays used in this analysis.

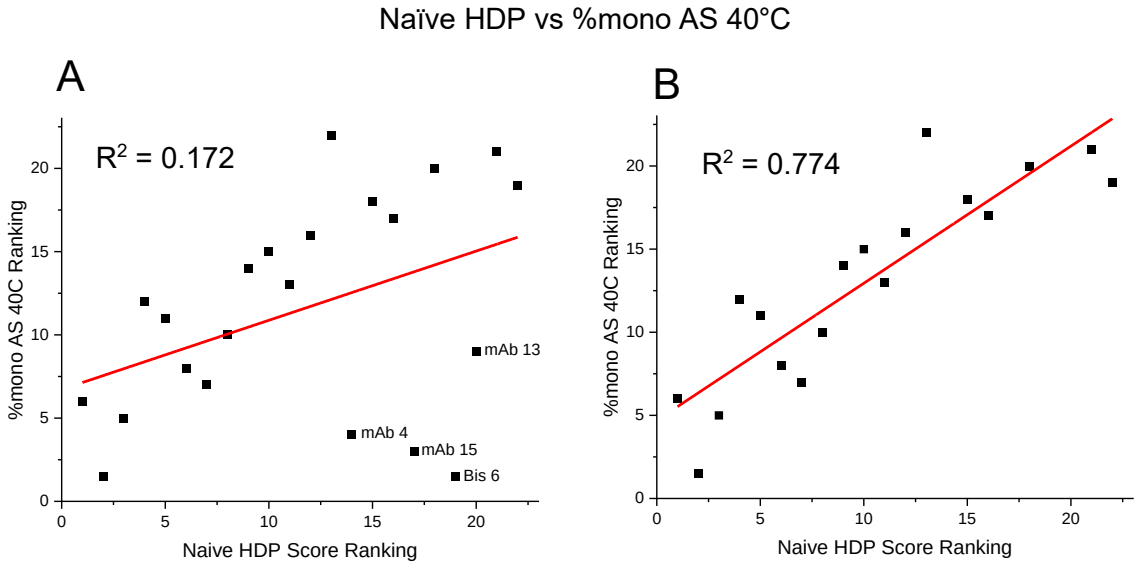


Figure 6.19: Naïve HDP Score Correlates with Accelerated Stability at 40 °C when 45 °C Data is Included

Native holistic developability parameter ranking correlated with monomer loss after 40 °C AS rankings. 45 °C data included. Analysis shown with (A) and without (B) mAbs that are considered as part of a distinct population (shown labelled).

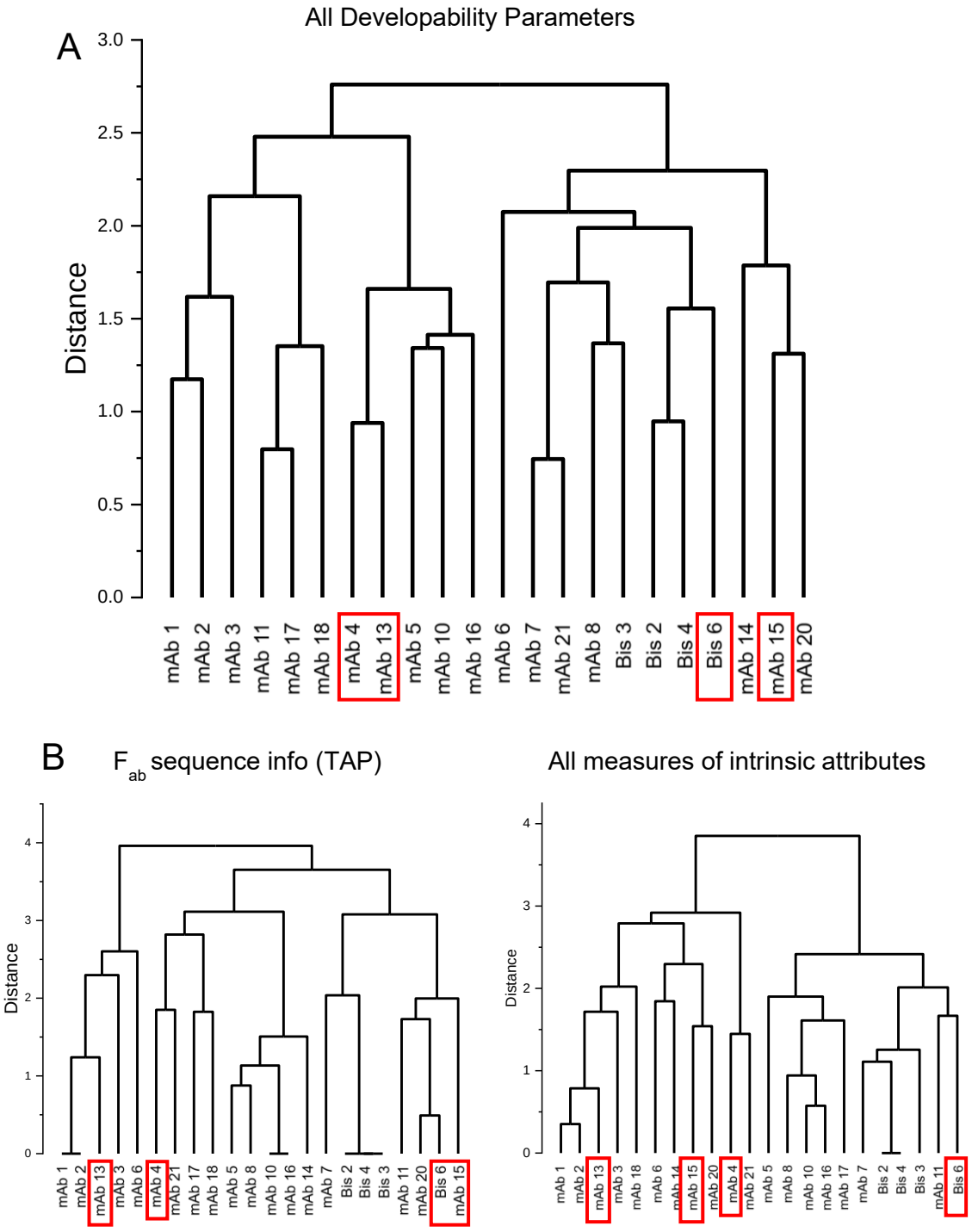


Figure 6.20: Clustering of Panel mAbs

Hierarchical clustering of panel mAbs by A) all developability parameters (excluding FIDA) and B) parameters resulting from intrinsic features (CamSol scores, T_m values and all TAP scores). mAbs thought to be potential outliers from the main population (identified in fig 6.19A) are indicated by red boxes.

6.3.5 Refinement of HDP Scores

As assumed in the naïve HDP score, each assay group is unlikely to contribute equally to a prediction of protein stability. To account for this and improve the closeness of fit in the HDP/AS ranking relationship, multiple linear regression (MLR) was used to calculate the contribution from each group (in fig 6.18) in predicting % monomer remaining after 4 weeks at 40 °C (%mono AS 40C) ranking, and refine the fit (fig 6.21)I. The contribution from each assay group was given as a β-coefficient value, which was used to weight the naïve HDP scores associated with each group (section 2.2.9.2).

Upon weighting of each assay group naïve HDP score, a whole 'refined HDP score' could be calculated from these. Predicting %mono AS 40C with the refined HDP yields an improved fit, matching the majority of the dataset well, but being insufficient to describe the data when the 'subpopulation' mAbs are included (fig 6.21). This indicates that while a HDP calculated with this assay set is promising, it is unable to describe AS risk in a population of 22 mAbs.

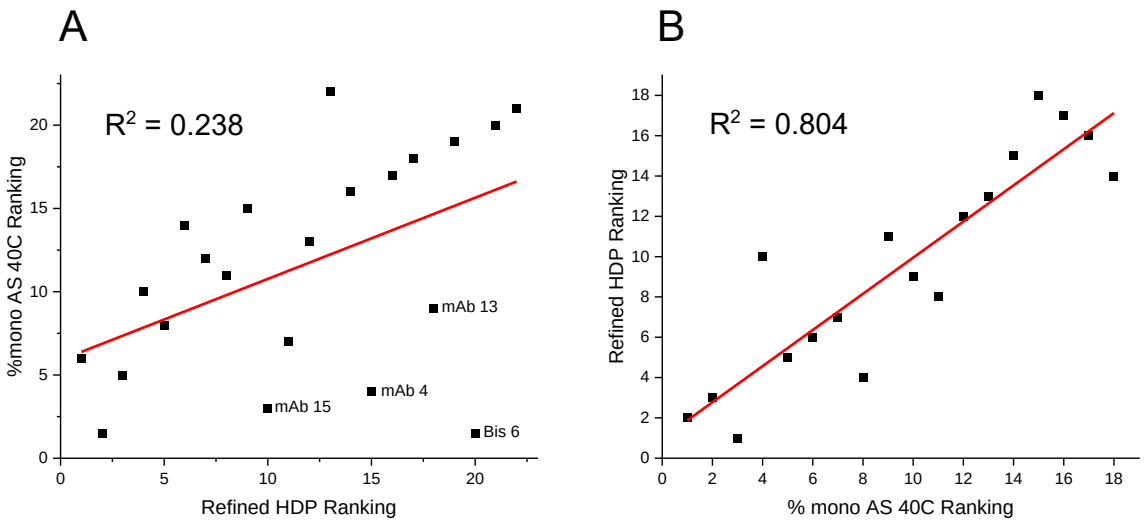


Figure 6.21: Refined HDP Score Correlated with Accelerated Stability

MLR-refined holistic developability parameter ranking correlated with 40 °C AS monomer loss rankings. Analysis shown with (A) and without (B) mAbs that are considered part of a distinct population (shown labelled).

The β-coefficients produced from MLR, given in table 6.3, are indicative of which assay groups are most predictive of accelerated stability. These values suggest that using naïve HDP ranking, group B contributes most to the prediction, regardless of whether the more isolated population of mAbs identified in fig 6.19 are included. This may be due to the contribution from T_m, as conformational stability at the 40 °C AS temperature chosen is likely a factor in the extent of observed monomer loss.

A

Group	β-Coefficient
A	4.01
B	11.67
C	6.68
D	2.55

B

Group	β-Coefficient
A	6.05
B	18.02
C	15.97
D	7.44

Table 6.3: β-Coefficients from MLR of Naïve HDP (no AS 40°C parameters)

β-coefficients calculated A) with and B) without residues identified in figure 6.19.

Interestingly, group C, containing only percentage change in higher-molecular weight species after 2 weeks at 45 °C (%HMW AS 45C), contributes more than group D, which contains two AS conditions and EFD. This result was unexpected, as AS at 40 and 45 °C were considered close enough together they could be predictive of one another. This could be a result of the limited number of assays, meaning contribution from only one assay in group C could more easily offer a good fit, because the likelihood of having goodness of fit reduced by a poor performing assay is greatly reduced.

Importantly, group A contributes relatively little to prediction of %mono AS 40C. The parameters in this group are all based on intrinsic properties of molecules at t=0 and are unlikely to accurately predict the complex process of protein conformational change and aggregation over time.

6.3.6 HDP Score Calculated with no Accelerated Stability Parameters

As AS measurements at 45 °C are similar in design to those conducted at 40 °C, the analysis conducted in sections 6.3.4 and 6.3.5 above were repeated, with all AS data removed. Using this dataset, the HDP would be constituted of experiments which take at most a day of experimental time, and each use below 1.25 mg of protein per sample.

Again, only the dataset excluding FIDA retains the same assay groupings upon hierarchical clustering, however as 6 developability parameters are removed at this point, group C is removed, and number of assays in group D is diminished to only include EFD (fig 6.22).

A naïve HDP value was calculated for each panel mAb from this dataset, and this was refined against %mono AS 40C ranking using MLR, as described above (section 6.3.5). The β-coefficients calculated for the refined HDP score support the conclusion that group B is dominant in predicting AS at 40 °C (table 6.4). The minor contribution from EFD, the only remaining assay in group D, suggests that this is not highly indicative of AS at 40 °C and may have been the cause for group D's lacklustre contribution, even when AS data at 45 °C were included.

Group	β-Coefficient
A	1.50
B	10.61
D	1.50

Table 6.4: β-Coefficients from MLR of Naïve HDP (no AS parameters)

β-coefficients calculated with no AS data included in HDP calculations.

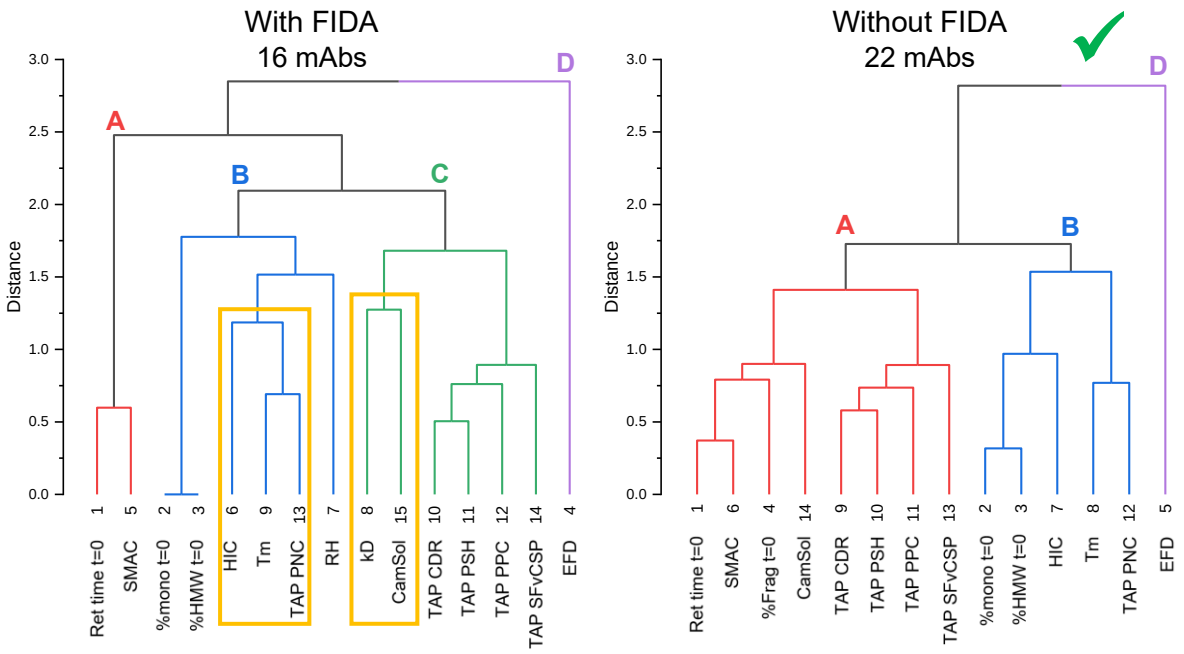


Figure 6.22: Clustering of Assays Measuring Developability Parameters with No Accelerated Stability Data Included

Hierarchical clustering of correlation coefficients for each developability parameter, with and without developability parameters from FIDA. Different assay groups indicated by different branch colours. With FIDA included, groupings of assays indicated by yellow boxes change.

Neither the naïve nor refined HDP rankings correlate with %mono AS 40C rankings, indicating that without a partial contribution from AS experiments, the selected dataset is unable to predict for AS performance at 40 °C (fig 6.23).

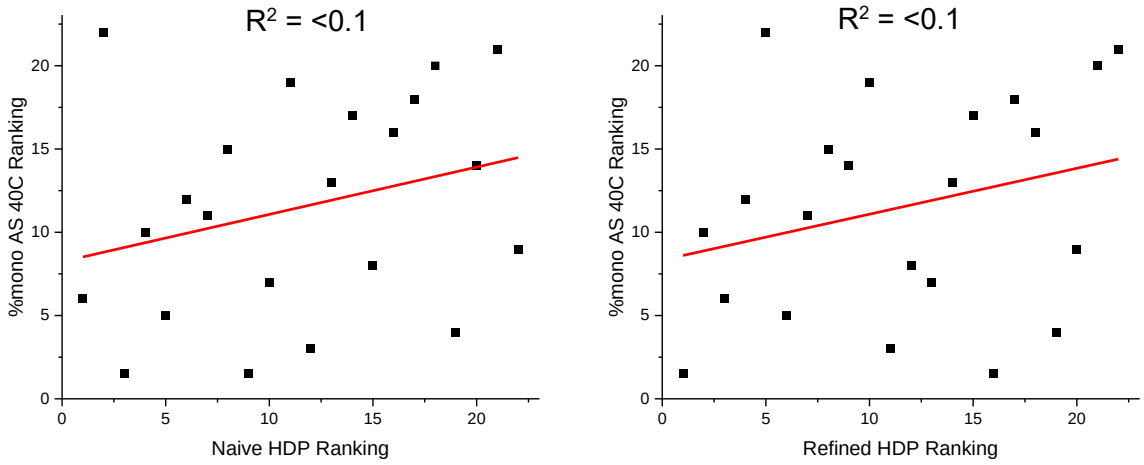


Figure 6.23: Naïve and Refined HDP Scores Do Not Correlate with Accelerated Stability at 40 °C When No Accelerated Stability Data is Included

A) Naïve and B) Refined holistic developability parameter rankings correlated with monomer loss after 40 °C AS ranking, without any AS data included in HDP calculation.

6.4 Discussion

The previous three chapters (3-5) have demonstrated both hydrodynamic flow-induced unfolding and aggregation in a model protein, RNase H*, using a bespoke extensional flow device (EFD). However, the broader application of this instrument may not have been fully realised; the unique challenge to developability that flow poses suggests that an assay which can quantify flow-induced aggregation propensity is lacking in the current developability toolbox.

In this chapter, a selection of developability parameters including hydrodynamic flow have been measured in a panel of 27 total mAbs, including 21 monospecific and 6 bispecific molecules. The aims were two-fold; Firstly, identify how hydrodynamic flow impacts a range of antibodies with attention given to emergent bispecific modalities, and investigate how EFD-induced aggregation relates to outputs from a range of existing developability assays. Secondly, combine the established assays to produce a holistic developability parameter (HDP), and determine if the selected assay set is able to establish a HDP that could be predictive of AS in a much shorter experimental timeframe, and with less material consumption.

6.4.1 A Role for the EFD in Biopharmaceutical Developability

EFD-induced aggregation was able to distinguish mAbs that were particularly susceptible to flow-induced aggregation, with the dataset showing a long-tailed distribution characteristic of many developability assays (Jain, Sun, *et al.*, 2017). Analysis of the mAb panel by other developability assays revealed a range of scores in all other assays, indicating the dataset were subject to other developability pitfalls. Remarkably, when clustered with other assays, EFD did not associate closely with any others, indicating a low-level of redundancy. This is in line with observations by Willis *et al*., (2020), in which the novel insight offered by the EFD was highlighted in the context of 33 mAbs previously examined in detail with a similar set of assays.

This information, coupled with observations in section 5.1.7, suggest a protein's mechanical stability against flow-induced unfolding and aggregation is completely different to that measured by thermal or chemical means, and as discussed in section 1.1.4, is thought to be anisotropic, meaning likely not measurable using overly simplistic

unidirectional pulling experiments. Thus, devices such as the EFD which are able to impart hydrodynamic flow on proteins in a defined manner may be viable options as part of a developability toolkit.

6.4.2 Predicting Accelerated Stability with a Holistic Developability Parameter

Production of a HDP using the selected developability assays appeared to predict monomer content changes with 4-week AS at 40 °C (%mono AS 40C), provided 2-week AS data at 45 °C was included. Using multiple linear regression (MLR) to identify the assay groups contributing most to this prediction revealed that assays reporting on intrinsic biophysical properties were less able to predict AS, perhaps indicative that stability under thermal stress over time is defined by unfolding, which is at least partially captured by T_m measurements (found in group B, which contributed most to AS prediction).

However, without contribution from AS data, HDP is not predictive of %mono AS 40C. This indicates that the selected dataset is insufficient to predict AS. Notably, removal of AS parameters results in assay group D only containing 1 developability parameter, and group C being removed completely. This means assays within less-populated groups contribute disproportionally to the HDP score, as HDP is calculated from averaged assay group scores. This variation in assay number per group is an inherent weakness of the HDP; due to the nature of the assays addressing a broad range of features, the selection was unlikely to lead to an equal number of assays in each group. A potential method to address this could be to select from the groupings established in the larger dataset by Willis *et al*. (manuscript in preparation), so that there are an equal number of assays in each group. From this, HDP could be calculated, and then MLR conducted, to weight assay groups depending on their contribution, but with an equal selection of input data.

Alternatively, AS at 40 °C may not be a suitable parameter for long-term stability predictions. Recommended storage temperatures for biopharmaceuticals are usually between 2 and 8 °C (Evers, Clénet and Pfeiffer-Marek, 2022). Whilst groups have had success in predicting long term stability at these temperatures using a range of AS conditions (usually between 5 and 40 °C)(Evers, Clénet and Pfeiffer-Marek, 2022), using only 40 and 45 °C may be inappropriate as AS at higher temperatures is thought to follow a different mechanism as in long-term low-temperature storage (Brummitt, Nesta and Roberts, 2011). Wälchli *et al*., (2020) showed that at 5 °C, aggregation is limited to small

multimer formation, whereas at 40 °C, larger aggregates form and agglomerate. Additionally, rate of monomer loss was shown to be non-Arrhenius. While the mechanism behind these differences is not understood, the authors proposed that more excursions from the native state are possible at higher temperatures, perhaps facilitating addition of further multimers that would not be possible at lower temperatures, where a smaller extent of unfolding is present.

Additionally, when a larger dataset of assays was clustered, including 5 and 25 °C long-term stability data at up to 18 and 6 months respectively, 40 °C stability up to 3 months did not group with these, and only correlated with T_m (Willis *et al*., manuscript in preparation). This study also revealed that long-term stability parameters group with EFD-induced aggregation, indicating a possible application of flow-induced aggregation for prediction of long-term molecule stability.

7 Concluding Remarks and Future Avenues of Research

7.1 Concluding Remarks

The overarching aim of this thesis was to understand whether hydrodynamic flow, with a particular focus on extensional flow fields, induced partial unfolding of proteins and if so, the role of these meta-stable states in flow-induced aggregation.

To achieve this, the model protein RNase H* was employed, and the absence of cysteine residues in its sequence was exploited to design a series of cysteine substitution variants, which each contained a single cysteine residue at a different site in the protein. Provided these sites remained sequestered from solution under native folding conditions, they could act as reporters for protein unfolding, detectable by reaction with the thiol-reactive fluorophore IAEDANS.

This panel of variants was produced and structurally characterised throughout chapter 3, revealing that despite sharing secondary and tertiary structure similarities, the different cysteine substitutions resulted in a range of thermodynamic stabilities, an observation of interest as one of the aims of this study was to investigate the relationship between thermodynamic stability and resistance to flow-induced unfolding.

RNase H* variants were exposed to defined flow fields, using a bespoke ‘extensional flow device’ (EFD) previously introduced by Dobson *et al*. (2017). *In situ* IAEDANS labelling of RNase H* variants indicated flow-induced conformational changes occurred, however the original method of quantifying fluorescent SDS-PAGE bands through densitometry was not sufficiently sensitive. To address this, quantification of samples with an HPLC fluorescence detector was adopted, with assay development discussed in chapter 4.

Prior to this, label-free assays were also investigated, as the prospect of detecting RNase H* unfolding without the need for covalent modification was appealing. However, both use of an oligonucleotide substrate that was designed to fluoresce upon cleavage by folded, enzymatically active RNase H*, and use of a proteolytic enzyme thermolysin, with the purpose of cleaving unfolded RNase H*, were not deemed viable. Both assays were not sufficiently reproducible, due to the dead-time between initiation and the first measurement being taken, as well as the self-degradation observed in thermolysin.

Subsequently, the refined IAEDANS labelling assay was used in chapter 5 to characterise the effects of hydrodynamic flow on the panel of RNase H* cysteine variants in

aggregation-inhibiting arginine buffer; the assay was able to probe for conformational change in a site-specific manner, revealing a high degree of heterogeneity between sites. This offers insight into the mechanism of flow-induced protein unfolding that has not been previously established, indicating firstly that a protein sample unfolds to a greater extent under flow than when quiescent, secondly that the extent of unfolding is not always complete – indicated by the heterogeneity between samples, and thirdly that sites found within a previously identified region of stability – the RNase H* hydrophobic core – are likely to display greater mechanical stability under flow.

This latter conclusion has significant implications for biopharmaceutical design; as discussed in 1.3.2, flow fields of varying intensities are widely present during biopharmaceutical manufacture. As such, understanding the structural features needed to design flow-resistant proteins could allow development of increasingly stable and manufacturable biopharmaceutics.

As discussed in section 1.2.1, protein unfolding is thought to be a significant cause of aggregation. As the aggregation propensities of the RNase H* variants were diverse, these values were compared to the extents of unfolding observed, and found to correlate. To further verify this relationship, flow induced unfolding was measured in aggregation-permissive phosphate buffer, and insoluble aggregate sample was isolated and solubilised, before being measured for fluorescent IAEDANS labelling (indicative of unfolding). It was found aggregated protein was labelled to a much greater extent than soluble protein in flow-stressed samples, but this was not mirrored in quiescent samples. This suggests that while unfolding may occur under quiescent conditions, it is more common under hydrodynamic flow, and this unfolded material is much more liable to aggregate.

Numerous metrics have been investigated to predict a protein's aggregation propensity, and two of the most readily considered are thermodynamic stability and content of aggregation-prone regions (Sormanni, Aprile and Vendruscolo, 2015). It was found that neither of these correlated with flow-induced unfolding or aggregation in the RNase H* variants, alluding to the complicated nature of flow and the fact that mechanical and thermodynamic stability are distinct. This suggests that hydrodynamic flow may need to be considered independently when establishing how stable a protein is, especially in the context of developability.

This was further investigated in chapter 6, whereby a panel of 27 mAbs, including mono and bispecific IgGs, were analysed for flow-induced aggregation using the EFD. To

ascertain how EFD-induced aggregation compares to other developability assays commonly used in industry, the panel were interrogated using 9 other assays, chosen to represent different 'families' of assays identified by Willis *et al*. (manuscript in preparation) that each probe different aspects of a candidate molecule's developability. It was found that, as observed in Willis *et al*. (2020), EFD-induced aggregation in the mAb panel did not correlate well with performance in the other assays included, further supporting the role for hydrodynamic flow-induced aggregation as an independent and unique parameter, but one that should not be ignored due to its ubiquitous nature in biopharmaceutical manufacturing.

Willis *et al*. (manuscript in preparation) also found that EFD-induced aggregation could be included as part of a single, holistic, developability parameter (HDP) which condenses performance scores in all assays analysed, and that this score was predictive of long-term stability when stored at 25 °C for up to 6 months. The potential of the panel presented in chapter 6 to predict shorter-term accelerated stability results, which usually take at least 4 weeks to obtain and are routinely employed for stability studies in industry (Evers, Clénet and Pfeiffer-Marek, 2022), was investigated. However, the analysis found that assembly of the HDP from the parameters chosen was not predictive of accelerated stability when panel mAbs were stored for 4 weeks at 40 °C. This is thought to be due to the different mechanism of degradation that occurs in samples at elevated temperature compared to at 25 °C (Wälchli *et al.*, 2020), and perhaps indicates that like hydrodynamic flow, accelerated stability is a unique parameter that cannot be predicted by other simplistic assays.

7.2 Future Avenues of Research

Whilst the IAEDANS labelling assay can report region-specific information on flow-induced conformational change, design and characterisation of cysteine substitution variants is time-consuming and requires selectivity in the choice of substitution positions.

An alternative may be to use a more promiscuous labelling strategy, to target naturally present residues in RNase H*. For example, WT RNase H* contains 5 tyrosine residues, which, due to their hydrophobic nature are likely to be sequestered from solution. A reactive molecule such as N-acetylimidazole could be used to label exposed tyrosine side chains, and the sites of labels could then be identified by peptide mapping and mass spectrometry (Mendoza and Vachet, 2009).

For a more detailed insight into the structural changes that occur under flow, EFD-induced perturbations have been measured in antibodies using fast photochemical oxidation of proteins (FPOP)(Lawrence, 2023), a method involving modification of exposed amino acid side chains with laser-generated hydroxyl radicals, which can be detected by mass-spectrometry. However, this method does not generate hydroxyl radicals *in situ*, meaning an assay dead-time exists between stressing under flow and exposure to the radicals, in which proteins may refold. An alternative to circumvent this issue could be to develop a flow device with a photometrically-accessible viewing window to allow generation of hydroxyl radicals, perhaps at the region of maximal extensional strain or shear rate.

This concept could also be applied to microscopy techniques, for example confocal FRET could be used to detect distance change between two engineered fluorophores in a protein of interest. If the constriction in the flow apparatus, where extensional flow is thought to dominate the flow regime, could be viewed and compared to a region of no constriction, where shear is thought to dominate, this could help distinguish between the effects of the two forces, which have historically been entangled in the literature.

For higher-resolution structural insight, assays based on hydrogen exchange with deuterium atoms (HX) upon exposure of backbone amides could also be viable; both methods could offer residue-specific detail and rapidly capture transient excursions from the native state.

Whilst this work suggests regions that are more susceptible to unfolding and therefore label more also contribute more to aggregation, effects from introduction of cysteine residue substitutions may impact other regions in the protein, causing the observed changes in aggregation. To confirm the importance of particular regions, a disulphide bond or other relatively strong single chemical bond could be introduced into the region thought to unfold, ‘stapling’ it to the rest of the structure (i.e. strand 3 of the β-sheet, thought to be perturbed as discussed in section 5.1.4). If introduction of this bond resulted in a reduction in aggregation, it could be inferred that it was the region responsible for aggregation in that cysteine variant.

8 Appendix

RNase H* Variant	Expected MW (Da)	Measured MW (Da)
WT	17500.81	17500.30 ± 0.04
A137C	17,532.88	17531.82 ± 0.08
A24C	17,532.88	17532.59 ± 0.06
G23C	17,546.90	17546.56 ± 0.06
I82C	17,490.80	17490.47 ± 0.04
V74C	17,504.82	17505.4
H114C	17,466.82	17467.4
A55C	17,532.88	17533.34 ± 0.03
I53C	17,490.80	17491.52 ± 0.04

Table A1: Expected and Measured Intact Mass Values for RNase H* and Cysteine Variants

Variant	K_1	k_{-1}	K	k_2	k_{-2}	K	k_3	k_{-3}	K	k_4	k_{-4}	K	k_5	k_{-5}	K	k_6	k_{-6}	K
WT	2.26	7.25E-06	311,724	2.08E-04	2.02E-07	1029	0.0241	0.0328	0.735	0.72	167	0.0043	0.453	0.00223	203.139	0.453	0.00223	203.139
G23C	2.26	1.90E-07	1.19E-07	1.94E-04	3.68E-06	52.7	0.0455	13	0.0035	2.02E-07	0.000159	0.00127	0.453	0.00223	203.139	0.453	0.00223	203.139
A55C	2.26	3.21E-07	7.04E-06	4.37E-04	1.29E-04	3.39	0.0252	23.6	0.00107	1.20E-07	0.000268	4.48E-04	0.453	0.00223	203.139	0.453	0.00223	203.139
I53C	2.26	8.35E-08	2.71E-07	2.02E-04	6.33E-08	3190	0.103	5.76	0.0179	4.62E-07	6.99E-05	0.00661	0.453	0.00223	203.139	0.453	0.00223	203.139
A52C	2.26	3.68E-07	6.14E-06	2.00E-04	1.17E-04	1.71	0.022	27	8.15E-04	1.05E-07	3.08E-04	3.41E-04	0.453	0.00223	203.139	0.453	0.00223	203.139
V74C	2.26	1.73E-07	1.31E-07	3.48E-04	4.05E-06	85.9	0.0501	11.8	0.00425	2.23E-07	0.000145	0.00154	0.453	0.00223	203.139	0.453	0.00223	203.139
H114C	2.26	1.80E-07	1.26E-07	2.84E-04	3.89E-06	73.0	0.049	12.1	0.00405	2.15E-07	0.00015	0.00143	0.453	0.00223	203.139	0.453	0.00223	203.139
A24C	2.26	3.42E-07	6.61E-06	2.99E-04	1.23E-04	2.43	0.0236	25.2	9.37E-04	1.13E-07	0.000287	3.94E-04	0.453	0.00223	203.139	0.453	0.00223	203.139
I7C	2.26	1.93E-07	1.17E-07	1.76E-04	3.62E-06	48.6	0.0448	13.2	0.00340	1.99E-07	0.000162	0.00123	0.453	0.00223	203.139	0.453	0.00223	203.139

Table A2: Modelled Rate Constants for RNase H* Cysteine Variants

Rate constants produced from simulated fits (fig 3.15). Greater flux values are indicated in darker red.

9 References

Abascal, N.C. and Regan, L. (2018) ‘The past, present and future of protein-based materials’, *Open Biology*, 8(10). Available at: https://doi.org/10.1098/rsob.180113.

Adler, M. and Allmendinger, A. (2023) ‘Filling unit operation for biological drug products: Challenges and considerations’, *Journal of Pharmaceutical Sciences*, 3549(23). Available at: https://doi.org/10.1016/j.xphs.2023.11.017.

Ahmed, S.H., Espinoza-Sánchez, N.A., El-Damen, A., Fahim, S.A., Badawy, M.A., Greve, B., El-Shinawi, M., Götte, M. and Ibrahim, S.A. (2021) ‘Small extracellular vesicle-encapsulated miR-181b-5p, miR-222-3p and let-7a-5p: Next generation plasma biopsy-based diagnostic biomarkers for inflammatory breast cancer’, *PLOS ONE*, 16(4), p. e0250642. Available at: https://doi.org/10.1371/journal.pone.0250642.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2007) *Molecular Biology of the Cell*. W.W. Norton & Company. Available at: https://doi.org/10.1201/9780203833445.

Amin, S., Barnett, G. V., Pathak, J.A., Roberts, C.J. and Sarangapani, P.S. (2014) ‘Protein aggregation, particle formation, characterization & rheology’, *Current Opinion in Colloid & Interface Science*, 19(5), pp. 438–449. Available at: https://doi.org/10.1016/j.cocis.2014.10.002.

Andrews, J.M. and Roberts, C.J. (2007) ‘A Lumry−Eyring Nucleated Polymerization Model of Protein Aggregation Kinetics: 1. Aggregation with Pre-Equilibrated Unfolding’, *The Journal of Physical Chemistry B*, 111(27), pp. 7897–7913. Available at: https://doi.org/10.1021/jp070212j.

Anfinsen, C.B., Haber, E., Sela, M. and White, F.H. (1961) ‘The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain’, *Proceedings of the National Academy of Sciences*, 47(9), pp. 1309–1314. Available at: https://doi.org/10.1073/pnas.47.9.1309.

Arakawa, T., Ejima, D., Tsumoto, K., Obeyama, N., Tanaka, Y., Kita, Y. and Timasheff, S.N. (2007) ‘Suppression of protein interactions by arginine: A proposed mechanism of the arginine effects’, *Biophysical Chemistry*, 127(1–2), pp. 1–8. Available at: https://doi.org/10.1016/j.bpc.2006.12.007.

Arakawa, T., Ejima, D. and Akuta, T. (2017) ‘Protein aggregation under high concentration/density state during chromatographic and ultrafiltration processes’, *International Journal of Biological Macromolecules*, 95, pp. 1153–1158. Available at: https://doi.org/10.1016/j.ijbiomac.2016.11.005.

Arakawa, T. and Tsumoto, K. (2003) ‘The effects of arginine on refolding of aggregated proteins: Not facilitate refolding, but suppress aggregation’, *Biochemical and Biophysical Research Communications*, 304(1), pp. 148–152. Available at: https://doi.org/10.1016/S0006-291X(03)00578-3.

Ashton, L., Dusting, J., Imomoh, E., Balabani, S. and Blanch, E.W. (2010) ‘Susceptibility of Different Proteins to Flow-Induced Conformational Changes Monitored with Raman Spectroscopy’, *Biophysical Journal*, 98(4), pp. 707–714. Available at: https://doi.org/10.1016/j.bpj.2009.10.010.

Bai, G., Bee, J.S., Biddlecombe, J.G., Chen, Q. and Leach, W.T. (2012) ‘Computational fluid dynamics (CFD) insights into agitation stress methods in biopharmaceutical development’, *International Journal of Pharmaceutics*, 423(2), pp. 264–280. Available

at: https://doi.org/10.1016/j.ijpharm.2011.11.044.

Baynes, B.M., Wang, D.I.C. and Trout, B.L. (2005) 'Role of arginine in the stabilization of proteins against aggregation', *Biochemistry*, 44(12), pp. 4919-4925. Available at: https://doi.org/10.1021/bi047528r.

Bee, J.S., Stevenson, J.L., Mehta, B., Svitel, J., Pollastrini, J., Platz, R., Freund, E., Carpenter, J.F. and Randolph, T.W. (2009) 'Response of a concentrated monoclonal antibody formulation to high shear', *Biotechnology and Bioengineering*, 103(5), pp. 936–943. Available at: https://doi.org/10.1002/bit.22336.

Beerten, J., Jonckheere, W., Rudyak, S., Xu, J., Wilkinson, H., De Smet, F., Schymkowitz, J. and Rousseau, F. (2012) 'Aggregation gatekeepers modulate protein homeostasis of aggregating sequences and affect bacterial fitness', *Protein Engineering Design and Selection*, 25(7), pp. 357–366. Available at: https://doi.org/10.1093/protein/gzs031.

Bekard, I.B., Asimakis, P., Bertolini, J. and Dunstan, D.E. (2011) 'The effects of shear flow on protein structure and function', *Biopolymers*, 95(11), pp. 733–745. Available at: https://doi.org/10.1002/bip.21646.

Best, R.B., Fowler, S.B., Toca Herrera, J.L., Steward, A., Paci, E. and Clarke, J. (2003) 'Mechanical Unfolding of a Titin Ig Domain: Structure of Transition State Revealed by Combining Atomic Force Microscopy, Protein Engineering and Molecular Dynamics Simulations', *Journal of Molecular Biology*, 330(4), pp. 867–877. Available at: https://doi.org/10.1016/S0022-2836(03)00618-1.

Booth, D.R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W.L., Fraser, P.E., Hawkins, P.N., Dobson, C.M., Radford, S.E., Blake, C.C.F. and Pepys, M.B. (1997) 'Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis', *Nature*, 385(6619), pp. 787–793. Available at: https://doi.org/10.1038/385787a0.

Borgia, M. B., Nickson, A. A., Clarke, J., Hounslow, M. J. (2013) 'A Mechanistic Model for Amorphous Protein Aggregation of Immunoglobulin-like Domains' *Journal of the American Chemical Society*, 135(17), 6456-6464. Available at: https://doi.org/10.1021/ja308852b

Bowden, G.A. and Georgiou, G. (1990) 'Folding and aggregation of beta-lactamase in the periplasmic space of Escherichia coli.', *Journal of Biological Chemistry*, 265(28), pp. 16760–16766. Available at: https://doi.org/10.1016/S0021-9258(17)44825-3.

Brockwell, D.J., Paci, E., Zinober, R.C., Beddard, G.S., Olmsted, P.D., Smith, D.A., Perham, R.N. and Radford, S.E. (2003) 'Pulling geometry defines the mechanical resistance of a β-sheet protein', *Nature Structural Biology*, 10, pp. 731–737. Available at: https://doi.org/10.1038/nsb968.

Bruce, A.J.E., Paci, E. and Brockwell, D.J. (2023) 'Collagen-like Motifs of SasG: A Novel Fold for Protein Mechanical Strength', *Journal of Molecular Biology*, 435(6), p. 167980. Available at: https://doi.org/10.1016/j.jmb.2023.167980.

Brummitt, R.K., Nesta, D.P. and Roberts, C.J. (2011) 'Predicting accelerated aggregation rates for monoclonal antibody formulations, and challenges for low-temperature predictions', *Journal of Pharmaceutical Sciences*, 100(10), pp. 4234–4243. Available at: https://doi.org/10.1002/jps.22633.

Bunc, M., Hadži, S., Graf, C., Bončina, M. and Lah, J. (2022) 'Aggregation Time Machine: A Platform for the Prediction and Optimization of Long-Term Antibody Stability Using Short-Term Kinetic Analysis', *Journal of Medicinal Chemistry*, 65(3), pp.

2623–2632. Available at: https://doi.org/10.1021/acs.jmedchem.1c02010.

Burra, G. and Thakur, A.K. (2016) ‘Unaided trifluoroacetic acid pretreatment solubilizes polyglutamine peptides and retains their biophysical properties of aggregation’, *Analytical Biochemistry*, 494, pp. 23–30. Available at: https://doi.org/10.1016/j.ab.2015.10.006.

Cabra, V., Vázquez-Contreras, E., Moreno, A. and Arreguin-Espinosa, R. (2008) ‘The effect of sulfhydryl groups and disulphide linkage in the thermal aggregation of Z19 α-zein’, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1784(7–8), pp. 1028–1036. Available at: https://doi.org/10.1016/j.bbapap.2008.04.002.

Cao, E., Chen, Y., Cui, Z. and Foster, P.R. (2003) ‘Effect of freezing and thawing rates on denaturation of proteins in aqueous solutions’, *Biotechnology and Bioengineering*, 82(6), pp. 684–690. Available at: https://doi.org/10.1002/bit.10612.

Cascella, R., Bigi, A., Cremades, N. and Cecchi, C. (2022) ‘Effects of oligomer toxicity, fibril toxicity and fibril spreading in synucleinopathies’, *Cellular and Molecular Life Sciences*, 79(3), p. 174. Available at: https://doi.org/10.1007/s00018-022-04166-9.

Cecconi, C., Shank, E.A., Bustamante, C. and Marqusee, S. (2005) ‘Direct Observation of the Three-State Folding of a Single Protein Molecule’, *Science*, 309(5743), pp. 2057–2060. Available at: https://doi.org/10.1126/science.1116702.

Cecconi, G., Shank, E.A., Bustamante, C. and Marqusee, S. (2005) ‘Direct observation of the three-state folding of a single protein molecule’, *Science*, 309(5743), pp. 2057–2060. Available at: https://doi.org/10.1126/science.1116702.

Chakroun, N., Hilton, D., Ahmad, S.S., Platt, G.W. and Dalby, P.A. (2016) ‘Mapping the Aggregation Kinetics of a Therapeutic Antibody Fragment’, *Molecular Pharmaceutics*, 13(2), pp. 307–319. Available at: https://doi.org/10.1021/acs.molpharmaceut.5b00387.

Chalmers, J.J. (2015) ‘Mixing, aeration and cell damage, 30+ years later: what we learned, how it affected the cell culture industry and what we would like to know more about’, *Current Opinion in Chemical Engineering*, 10, pp. 94–102. Available at: https://doi.org/10.1016/j.coche.2015.09.005.

Chamberlain, A.K., Handel, T.M. and Marqusee, S. (1996) ‘Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH’, *Nature Structural Biology*, 3(9), pp. 782–787. Available at: https://doi.org/10.1038/nsb0996-782.

Charm, S.E. and Wong, B.L. (1970) ‘Enzyme inactivation with shearing’, *Biotechnology and Bioengineering*, 12(6), pp. 1103–1109. Available at: https://doi.org/10.1002/bit.260120615.

Chaudhuri, T.K. and Paul, S. (2006) ‘Protein-misfolding diseases and chaperone-based therapeutic approaches’, *The FEBS Journal*, 273(7), pp. 1331–1349. Available at: https://doi.org/10.1111/j.1742-4658.2006.05181.x.

Chennamsetty, N., Voynov, V., Kayser, V., Helk, B. and Trout, B.L. (2009) ‘Design of therapeutic proteins with enhanced stability’, *Proceedings of the National Academy of Sciences*, 106(29), pp. 11937–11942. Available at: https://doi.org/10.1073/pnas.0904191106.

Chiti, F. and Dobson, C.M. (2017) ‘Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade’, *Annual Review of Biochemistry*, 86(1), pp. 27–68. Available at: https://doi.org/10.1146/annurev-biochem-061516-045115.

Chiu, M.L., Goulet, D.R., Teplyakov, A. and Gilliland, G.L. (2019) 'Antibody Structure and Function: The Basis for Engineering Therapeutics', *Antibodies*, 8(4), p. 55. Available at: https://doi.org/10.3390/antib8040055.

Coan, K.E.D., Maltby, D.A., Burlingame, A.L. and Shoichet, B.K. (2009) 'Promiscuous Aggregate-Based Inhibitors Promote Enzyme Unfolding', *Journal of Medicinal Chemistry*, 52(7), pp. 2067–2075. Available at: https://doi.org/10.1021/jm801605r.

Cohen, S.I.A., Linse, S., Luheshi, L.M., Hellstrand, E., White, D.A., Rajah, L., Otzen, D.E., Vendruscolo, M., Dobson, C.M. and Knowles, T.P.J. (2013) 'Proliferation of amyloid-β42 aggregates occurs through a secondary nucleation mechanism', *Proceedings of the National Academy of Sciences*, 110(24), pp. 9758–9763. Available at: https://doi.org/10.1073/pnas.1218402110.

Conchillo-Solé, O., de Groot, N.S., Avilés, F.X., Vendrell, J., Daura, X. and Ventura, S. (2007) 'AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides', *BMC Bioinformatics*, 8(1), p. 65. Available at: https://doi.org/10.1186/1471-2105-8-65.

Corona, A. and Tramontano, E. (2015) 'RNase H Polymerase-independent Cleavage Assay for Evaluation of RNase H Activity of Reverse Transcriptase Enzymes', *Bio-protocol*, 5(16). Available at: https://doi.org/10.21769/BioProtoc.1561.

Courtois, F., Agrawal, N.J., Lauer, T.M. and Trout, B.L. (2016) 'Rational design of therapeutic mAbs against aggregation through protein engineering and incorporation of glycosylation motifs applied to bevacizumab', *mAbs*, 8(1), pp. 99–112. Available at: https://doi.org/10.1080/19420862.2015.1112477.

Cox, D.L., Lashuel, H., Lee, K.Y.C. and Singh, R.R.P. (2005) 'The Materials Science of Protein Aggregation', *MRS Bulletin*, 30(6), pp. 452–457. Available at: https://doi.org/10.1557/mrs2005.123.

Dekel, Y., Machluf, Y., Gefen, T., Eidelshtein, G., Kotlyar, A., Bram, Y., Shahar, E., Reslane, F., Aizenshtein, E. and Pitcovski, J. (2017) 'Formation of multimeric antibodies for self-delivery of active monomers', *Drug Delivery*, 24(1), pp. 199–208. Available at: https://doi.org/10.1080/10717544.2016.1242179.

Dill, K.A. and Chan, H.S. (1997) 'From Levinthal to pathways to funnels', *Nature Structural & Molecular Biology*, 4(1), pp. 10–19. Available at: https://doi.org/10.1038/nsb0197-10.

Dobson, C.M. (2003) 'Protein folding and misfolding', *Nature*, 426(6968), pp. 884–890. Available at: https://doi.org/10.1038/nature02261.

Dobson, J., Kumar, A., Willis, L.F., Tuma, R., Higazi, D.R., Turner, R., Lowe, D.C., Ashcroft, A.E., Radford, S.E., Kapur, N. and Brockwell, D.J. (2017) 'Inducing protein aggregation by extensional flow', *Proceedings of the National Academy of Sciences of the United States of America*, 114(18), pp. 4673–4678. Available at: https://doi.org/10.1073/pnas.1702724114.

Dougherty, D.A. (1996) 'Cation-π Interactions in Chemistry and Biology: A New View of Benzene, Phe, Tyr, and Trp', *Science*, 271(5246), pp. 163–168. Available at: https://doi.org/10.1126/science.271.5246.163.

Duerkop, M., Berger, E., Dürauer, A. and Jungbauer, A. (2018) 'Impact of Cavitation, High Shear Stress and Air/Liquid Interfaces on Protein Aggregation', *Biotechnology Journal*, 13(7). Available at: https://doi.org/10.1002/biot.201800062.

Van Durme, J., De Baets, G., Van Der Kant, R., Ramakers, M., Ganesan, A., Wilkinson,

H., Gallardo, R., Rousseau, F. and Schymkowitz, J. (2016) ‘Solubis: a webserver to reduce protein aggregation through mutation’, *Protein Engineering Design and Selection*, 29(8), pp. 285–289. Available at: https://doi.org/10.1093/protein/gzw019.

Ebo, J. S., Saunders, J., Devine, P., Gordon, A. M., Warwick, A., S., Schiffrin, B., Chin, S. E., England, E., Button, J. D., Lloyd, C., Bond, N. J., Ashcroft, A. E., Radford, S. E., Lowe, D. C., Brockwell, D. J. (2020) ‘An in vivo platform to select and evolve aggregation-resistant proteins.’, *Nature communications*, 11(1), p. 1816. Available at: https://doi.org/10.1038/s41467-020-15667-1.

Ebo, J. S., Guthertz, N., Radford, S.E. and Brockwell, D.J. (2020) ‘Using Protein Engineering to Understand and Modulate Aggregation’, *Current Opinion in Structural Biology*, 60, pp. 157–166.

Elsayed, A., Jaber, N., Al-Remawi, M. and Abu-Salah, K. (2023) ‘From cell factories to patients: Stability challenges in biopharmaceuticals manufacturing and administration with mitigation strategies’, *International Journal of Pharmaceutics*, 645, p. 123360. Available at: https://doi.org/10.1016/j.ijpharm.2023.123360.

Engstler, M., Pfohl, T., Herminghaus, S., Boshart, M., Wiegertjes, G., Heddergott, N. and Overath, P. (2007) ‘Hydrodynamic Flow-Mediated Protein Sorting on the Cell Surface of Trypanosomes’, *Cell*, 131(3), pp. 505–515. Available at: https://doi.org/10.1016/j.cell.2007.08.046.

Evers, A., Clénet, D. and Pfeiffer-Marek, S. (2022) ‘Long-Term Stability Prediction for Developability Assessment of Biopharmaceutics Using Advanced Kinetic Modeling’, *Pharmaceutics*, 14(2), p. 375. Available at: https://doi.org/10.3390/pharmaceutics14020375.

Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J. and Serrano, L. (2004) ‘Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins’, *Nature Biotechnology*, 22(10), pp. 1302–1306. Available at: https://doi.org/10.1038/nbt1012.

Fersht, A.R. (1997) ‘Nucleation mechanisms in protein folding’, *Current Opinion in Structural Biology*, 7(1), pp. 3–9. Available at: https://doi.org/10.1016/S0959-440X(97)80002-4.

Fink, A.L. (1998) ‘Protein aggregation: Folding aggregates, inclusion bodies and amyloid’, *Folding and Design*, 3(1), pp. R9–R23. Available at: https://doi.org/10.1016/S1359-0278(98)00002-9.

Fischer, E. (1894) ‘Einfluss der Configuration auf die Wirkung der Enzyme’, *Berichte der deutschen chemischen Gesellschaft*, 27(3), pp. 2985–2993. Available at: https://doi.org/10.1002/cber.18940270364.

Forman, J.R. and Clarke, J. (2007) ‘Mechanical unfolding of proteins: insights into biology, structure and folding’, *Current Opinion in Structural Biology*, 17(1), pp. 58–66. Available at: https://doi.org/10.1016/j.sbi.2007.01.006.

Fuller, G.G. and Leal, L.G. (1980) ‘Flow birefringence of dilute polymer solutions in two-dimensional flows’, *Rheologica Acta*, 19(5), pp. 580–600. Available at: https://doi.org/10.1007/BF01517512.

Gallivan, J.P. and Dougherty, D.A. (1999) ‘Cation-π interactions in structural biology’, *Proceedings of the National Academy of Sciences*, 96(17), pp. 9459–9464. Available at: https://doi.org/10.1073/pnas.96.17.9459.

Galush, W.J. and Horst, T.A. (2015) ‘Vented Spikes Improve Delivery from Intravenous

Bags with No Air Headspace', *Journal of Pharmaceutical Sciences*, 104(7), pp. 2397–2400. Available at: https://doi.org/10.1002/jps.24466.

Greene, R.F. and Pace, C.N. (1974) 'Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin.', *The Journal of biological chemistry*, 249(17), pp. 5388–93. Available at: http://www.ncbi.nlm.nih.gov/pubmed/4416801.

Greenfield, N.J. and Fasman, G.D. (1969) 'Computed circular dichroism spectra for the evaluation of protein conformation', *Biochemistry*, 8(10), pp. 4108–4116. Available at: https://doi.org/10.1021/bi00838a031.

Grigolato, F. and Arosio, P. (2020) 'Synergistic effects of flow and interfaces on antibody aggregation', *Biotechnology and Bioengineering*, 117(2), pp. 417–428. Available at: https://doi.org/10.1002/bit.27212.

Grudzielanek, S., Velkova, A., Shukla, A., Smirnovas, V., Tatarek-Nossol, M., Rehage, H., Kapurniotu, A. and Winter, R. (2007) 'Cytotoxicity of Insulin within its Self-assembly and Amyloidogenic Pathways', *Journal of Molecular Biology*, 370(2), pp. 372–384. Available at: https://doi.org/10.1016/j.jmb.2007.04.053.

Gundlach, H.G., Stein, W.H. and Moore, S. (1959) 'The nature of the amino acid residues involved in the inactivation of ribonuclease by iodoacetate', *The Journal of biological chemistry*, 234(7), pp. 1754–1760. Available at: https://doi.org/10.1016/s0021-9258(18)69920-x.

Gurd, F.R.N. (1967) 'Carboxymethylation', *Methods in Enzymology*, 11, pp. 532–541. Available at: https://doi.org/10.1016/S0076-6879(67)11064-1.

Haverick, M., Mengisen, S., Shameem, M. and Ambrogelly, A. (2014) 'Separation of mAbs molecular variants by analytical hydrophobic interaction chromatography HPLC: overview and applications', *mAbs*, 6(4), pp. 852–858. Available at: https://doi.org/10.4161/mabs.28693.

Heinrikson, R.L., Stein, W.H., Crestfield, A.M. and Moore, S. (1965) 'The Reactivities of the Histidine Residues at the Active Site of Ribonuclease Toward Halo Acids of Different Structures.', *The Journal of biological chemistry*, 240, pp. 2921–34. Available at: https://doi.org/14342316.

Hötzel, I., Theil, F.-P., Bernstein, L.J., Prabhu, S., Deng, R., Quintana, L., Lutman, J., Sibia, R., Chan, P., Bumbaca, D., Fielder, P., Carter, P.J. and Kelley, R.F. (2012) 'A strategy for risk mitigation of antibodies with fast clearance', *mAbs*, 4(6), pp. 753–760. Available at: https://doi.org/10.4161/mabs.22189.

Hu, W., Walters, B.T., Kan, Z.Y., Mayne, L., Rosen, L.E., Marqusee, S. and Englander, S.W. (2013) 'Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry', *Proceedings of the National Academy of Sciences of the United States of America*, 110(19), pp. 7684–7689. Available at: https://doi.org/10.1073/pnas.1305887110.

Hudson, E.N. and Weber, G. (1973) 'Synthesis and characterization of two fluorescent sulfhydryl reagents', *Biochemistry*, 12(21), pp. 4154–4161. Available at: https://doi.org/10.1021/bi00745a019.

Inouye, K., Kuzuya, K. and Tonomura, B. (1998) 'Sodium chloride enhances markedly the thermal stability of thermolysin as well as its catalytic activity', *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, 1388(1), pp. 209–214. Available at: https://doi.org/10.1016/S0167-4838(98)00189-7.

Jackson, S.E., elMasry, N., Fersht, A.R., (1993) 'Structure of the hydrophobic core in the transition state for folding of chymotrypsin inhibitor 2: A critical test of the protein engineering method of analysis', *Biochemistry*, 32(42), pp. 11270-11278. Available at: https://doi.org/10.1021/bi00093a002

Jacobs, S.A., Wu, S.-J., Feng, Y., Bethea, D. and O'Neil, K.T. (2010) 'Cross-Interaction Chromatography: A Rapid Method to Identify Highly Soluble Monoclonal Antibody Candidates', *Pharmaceutical Research*, 27(1), pp. 65–71. Available at: https://doi.org/10.1007/s11095-009-0007-z.

Jain, T., Sun, T., Durand, S., Hall, A., Houston, N. R., Nett, J. H., Sharkley, B., Bobrowicz, B., Caffrey, I., Yu, Y., Cao, Y., Lynaugh, H., Brown, M., Baruah, H., Gray, L. T., Krauland, E. M., Xu, Y., Vasquez, M., Wittrup, K. D. (2017) 'Biophysical properties of the clinical-stage antibody landscape', *Proceedings of the National Academy of Sciences*, 114(5), pp. 944–949. Available at: https://doi.org/10.1073/pnas.1616408114.

Jain, T., Boland, T., Lilov, A., Burnina, I., Brown, M., Xu, Y. and Vásquez, M. (2017) 'Prediction of delayed retention of antibodies in hydrophobic interaction chromatography from sequence using machine learning', *Bioinformatics*. Edited by A. Valencia, 33(23), pp. 3758–3766. Available at: https://doi.org/10.1093/bioinformatics/btx519.

Javadi, Y., Fernandez, J.M. and Perez-Jimenez, R. (2013) 'Protein Folding Under Mechanical Forces: A Physiological View', *Physiology*, 28(1), pp. 9–17. Available at: https://doi.org/10.1152/physiol.00017.2012.

Jensen, M.K., Samelson, A.J., Steward, A., Clarke, J. and Marqusee, S. (2020) 'The folding and unfolding behavior of ribonuclease H on the ribosome', *Journal of Biological Chemistry*, 295(33), pp. 11410–11417. Available at: https://doi.org/10.1074/jbc.RA120.013909.

Johnson, C.M. (2013) 'Differential scanning calorimetry as a tool for protein folding and stability', *Archives of Biochemistry and Biophysics*, 531(1–2), pp. 100–109. Available at: https://doi.org/10.1016/j.abb.2012.09.008.

Johnson, K.A., Simpson, Z.B. and Blom, T. (2009) 'Global Kinetic Explorer: A new computer program for dynamic simulation and fitting of kinetic data', *Analytical Biochemistry*, 387(1), pp. 20–29. Available at: https://doi.org/10.1016/j.ab.2008.12.024.

Kabsch, W. and Sander, C. (1983) 'Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features', *Biopolymers*, 22(12), pp. 2577–2637. Available at: https://doi.org/10.1002/bip.360221211.

Kallis, G.B. and Holmgren, A. (1980) 'Differential reactivity of the functional sulfhydryl groups of cysteine-32 and cysteine-35 present in the reduced form of thioredoxin from Escherichia coli.', *The Journal of biological chemistry*, 255(21), pp. 10261–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7000775.

Kanaya, S., Kimura, S., Katsuda, C. and Ikehara, M. (1990) 'Role of cysteine residues in ribonuclease H from Escherichia coli. Site-directed mutagenesis and chemical modification', *Biochemical Journal*, 271(1), pp. 59–66. Available at: https://doi.org/10.1042/bj2710059.

van der Kant, R., Karow-Zwick, A. R., van Durme, J., Blech, M., Gallardo, R., Seeliger, D., Abfalg, K., Baatsen, P., Compernolle, G., Gils, A., Studts, J. M., Schulz, P., Garidel, P., Schymkowitz, J., Rousseau, F. (2017) 'Prediction and Reduction of the Aggregation of Monoclonal Antibodies', *Journal of Molecular Biology*, 429(8), pp. 1244–1261. Available at: https://doi.org/10.1016/j.jmb.2017.03.014.

van der Kant, R., van Durme, J., Rousseau, F. and Schymkowitz, J. (2019) ‘SolubiS: Optimizing Protein Solubility by Minimal Point Mutations’, in, pp. 317–333. Available at: https://doi.org/10.1007/978-1-4939-8820-4_21.

Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T. and Morikawa, K. (1990) ‘Three-dimensional structure of ribonuclease H from E. coli’, *Nature*, 347(6290), pp. 306–309. Available at: https://doi.org/10.1038/347306a0.

Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Nakamura, H., Ikehara, M., Matsuzaki, T. and Morikawa, K. (1992) ‘Structural details of ribonuclease H from Escherichia coli as refined to an atomic resolution’, *Journal of Molecular Biology*, 223(4), pp. 1029–1052. Available at: https://doi.org/10.1016/0022-2836(92)90260-Q.

Keller, A. and Odell, J.A. (1985) ‘The extensibility of macromolecules in solution; A new focus for macromolecular science’, *Colloid & Polymer Science*, 263(3), pp. 181–201. Available at: https://doi.org/10.1007/BF01415506.

Kim, Y., Ho, S.O., Gassman, N.R., Korlann, Y., Landorf, E. V., Collart, F.R. and Weiss, S. (2008) ‘Efficient Site-Specific Labeling of Proteins via Cysteines’, *Bioconjugate Chemistry*, 19(3), pp. 786–791. Available at: https://doi.org/10.1021/bc7002499.

Kimura, S., Nakamura, H., Hashimoto, T., Oobatake, M., Kanaya, S. (1992) 'Stabilization of Escherichia coli Ribonuclease HI by Strategic Replacement of Amino Acid Residues with Those from the Thermophilic Counterpart' *The Journal of Biological Chemistry*, 267(30), 21535-21542.

King, R.P. (2002) ‘Flow of fluids in piping systems’, in *Introduction to Practical Fluid Flow*. Elsevier, pp. 9–54. Available at: https://doi.org/10.1016/B978-075064885-1/50002-8.

Kingsbury, J.S., Saini, A., Auclair, S.M., Fu, L., Lantz, M.M., Halloran, K.T., Calero-Rubio, C., Schwenger, W., Airiau, C.Y., Zhang, J. and Gokarn, Y.R. (2020) ‘A single molecular descriptor to predict solution behavior of therapeutic antibodies’, *Science Advances*, 6(32). Available at: https://doi.org/10.1126/sciadv.abb0372.

Kohli, N., Jain, N., Geddie, M.L., Razlog, M., Xu, L. and Lugovskoy, A.A. (2015) ‘A novel screening method to assess developability of antibody-like molecules’, *mAbs*, 7(4), pp. 752–758. Available at: https://doi.org/10.1080/19420862.2015.1048410.

Kopp, M.R.G., Grigolato, F., Zürcher, D., Das, T.K., Chou, D., Wuchner, K. and Arosio, P. (2023) ‘Surface-Induced Protein Aggregation and Particle Formation in Biologics: Current Understanding of Mechanisms, Detection and Mitigation Strategies’, *Journal of Pharmaceutical Sciences*, 112(2), pp. 377–385. Available at: https://doi.org/10.1016/j.xphs.2022.10.009.

Korman, S. and Clarke, H.T. (1955) ‘Carboxymethyl Proteins’, *Journal of Biological Chemistry*, 221(1), pp. 133–141. Available at: https://doi.org/10.1016/S0021-9258(18)65235-4.

Kundu, P., Saha, S. and Gangopadhyay, G. (2020) ‘Mechanical Unfolding of Single Polyubiquitin Molecules Reveals Evidence of Dynamic Disorder’, *ACS Omega*, 5(16), pp. 9104–9113. Available at: https://doi.org/10.1021/acsomega.9b03701.

Kuroda, Y., (2022) 'Biophysical studies of amorphous protein aggregation and in vivo immunogenicity', *Biophysical Reviews*, 14(6), 1495-1501, Available at: https://doi.org/10.1007/s12551-022-01011-y.

Kuzman, D., Bunc, M., Ravnik, M., Reiter, F., Žagar, L. and Bončina, M. (2021) ‘Long-

term stability predictions of therapeutic monoclonal antibodies in solution using Arrhenius-based kinetics', *Scientific Reports*, 11(1), p. 20534. Available at: https://doi.org/10.1038/s41598-021-99875-9.

Lapidus, L.J. (2017) 'Protein unfolding mechanisms and their effects on folding experiments', *F1000Research*, 6, p. 1723. Available at: https://doi.org/10.12688/f1000research.12070.1.

Larson, N.R., Wei, Y., Prajapati, I., Chakraborty, A., Peters, B., Kalonia, C., Hudak, S., Choudhary, S., Esfandiary, R., Dhar, P., Schöneich, C. and Middaugh, C.R. (2020) 'Comparison of Polysorbate 80 Hydrolysis and Oxidation on the Aggregation of a Monoclonal Antibody', *Journal of Pharmaceutical Sciences*, 109(1), pp. 633–639. Available at: https://doi.org/10.1016/j.xphs.2019.10.069.

Lawrence, S.M. (2023) 'Detection of flow-induced perturbations of antibody structure using mass spectrometry (Doctoral dissertation, University of Leeds)', *White Rose eTheses Online*.

Leader, B., Baca, Q.J. and Golan, D.E. (2008) 'Protein therapeutics: a summary and pharmacological classification', *Nature Reviews Drug Discovery*, 7(1), pp. 21–39. Available at: https://doi.org/10.1038/nrd2399.

Levinthal, C. (1968) 'Are there pathways for protein folding?', *Journal de Chimie Physique*, 65, pp. 44–45. Available at: https://doi.org/10.1051/jcp/1968650044.

Lévy, E., El Banna, N., Baïlle, D., Heneman-Masurel, A., Truchet, S., Rezaei, H., Huang, M.-E., Béringue, V., Martin, D. and Vernis, L. (2019) 'Causative Links between Protein Aggregation and Oxidative Stress: A Review', *International Journal of Molecular Sciences*, 20(16), p. 3896. Available at: https://doi.org/10.3390/ijms20163896.

Lim, W.K., Rösgen, J. and Englander, S.W. (2009) 'Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group', *Proceedings of the National Academy of Sciences of the United States of America*, 106(8), pp. 2595–600. Available at: https://doi.org/10.1073/pnas.0812588106.

Liu, Y., Caffry, I., Wu, J., Geng, S.B., Jain, T., Sun, T., Reid, F., Cao, Y., Estep, P., Yu, Y., Vásquez, M., Tessier, P.M. and Xu, Y. (2014) 'High-throughput screening for developability during early-stage antibody discovery using self-interaction nanoparticle spectroscopy', *mAbs*, 6(2), pp. 483–492. Available at: https://doi.org/10.4161/mabs.27431.

Lumry, R. and Eyring, H. (1954) 'Conformation Changes of Proteins', *The Journal of Physical Chemistry*, 58(2), pp. 110–120. Available at: https://doi.org/10.1021/j150512a005.

Maa, Y.F. and Hsu, C.C. (1996) 'Effect of high shear on proteins.', *Biotechnology and bioengineering*, 51(4), pp. 458–65. Available at: https://doi.org/10.1002/(SICI)1097-0290(19960820)51:4<458::AID-BIT9>3.0.CO;2-H.

Mahler, H.-C., Friess, W., Grauschopf, U. and Kiese, S. (2009) 'Protein aggregation: Pathways, induction factors and analysis', *Journal of Pharmaceutical Sciences*, 98(9), pp. 2909–2934. Available at: https://doi.org/10.1002/jps.21566.

Maity, H., Maity, M., Krishna, M.M.G., Mayne, L. and Englander, S.W. (2005) 'Protein folding: The stepwise assembly of foldon units', *Proceedings of the National Academy of Sciences*, 102(13), pp. 4741–4746. Available at: https://doi.org/10.1073/pnas.0501043102.

Malencik, D.A. and Anderson, S.R. (2003) 'Dityrosine as a product of oxidative stress

and fluorescent probe', *Amino Acids*, 25(3–4), pp. 233–247. Available at: https://doi.org/10.1007/s00726-003-0014-z.

Marino, S.M. and Gladyshev, V.N. (2010) 'Cysteine Function Governs Its Conservation and Degeneration and Restricts Its Utilization on Protein Surfaces', *Journal of Molecular Biology*, 404(5), pp. 902–916. Available at: https://doi.org/10.1016/j.jmb.2010.09.027.

Matouschek, A. (2003) 'Protein unfolding — an important process in vivo?', *Current Opinion in Structural Biology*, 13(1), pp. 98–109. Available at: https://doi.org/10.1016/S0959-440X(03)00010-1.

Maurer-Stroh, S., Debulpaep, M., Kuemmerer, N., de la Paz, M.L., Martins, I.C., Reumers, J., Morris, K.L., Copland, A., Serpell, L., Serrano, L., Schymkowitz, J.W.H. and Rousseau, F. (2010) 'Exploring the sequence determinants of amyloid structure using position-specific scoring matrices', *Nature Methods*, 7(3), pp. 237–242. Available at: https://doi.org/10.1038/nmeth.1432.

Mazzer, A.R., Perraud, X., Halley, J., O'Hara, J. and Bracewell, D.G. (2015) 'Protein A chromatography increases monoclonal antibody aggregation rate during subsequent low pH virus inactivation hold', *Journal of Chromatography A*, 1415, pp. 83–90. Available at: https://doi.org/10.1016/j.chroma.2015.08.068.

McCully, M.E., Beck, D.A.C., Dagget, V. (2008) 'Microscopic reversibility of protein folding in molecular dynamics simulations of the engrailed homeodomain', *Biochemistry*, 47(27), 7079-7089. Available at: https://doi.org/10.1021/bi800118b

Meisl, G., Xu, C.K., Taylor, J.D., Michaels, T.C.T., Levin, A., Otzen, D., Klenerman, D., Matthews, S., Linse, S., Andreasen, M. and Knowles, T.P.J. (2022) 'Uncovering the universality of self-replication in protein aggregation and its link to disease', *Science Advances*, 8(32). Available at: https://doi.org/10.1126/sciadv.abn6831.

Mendoza, V.L. and Vachet, R.W. (2009) 'Probing protein structure by amino acid-specific covalent labeling and mass spectrometry', *Mass Spectrometry Reviews*, 28(5), pp. 785–815. Available at: https://doi.org/10.1002/mas.20203.

Menges, F. (no date) 'Spectragryph - optical spectroscopy software Version 1.2.16.1 http://www.effemm2.de/spectragryph/'.

Monera, O.D., Sereda, T.J., Zhou, N.E., Kay, C.M. and Hodges, R.S. (1995) 'Relationship of sidechain hydrophobicity and α-helical propensity on the stability of the single-stranded amphipathic α-helix', *Journal of Peptide Science*, 1(5), pp. 319–329. Available at: https://doi.org/10.1002/psc.310010507.

Monsellier, E. and Chiti, F. (2007) 'Prevention of amyloid-like aggregation as a driving force of protein evolution', *EMBO reports*, 8(8), pp. 737–742. Available at: https://doi.org/10.1038/sj.embor.7401034.

Moussa, E.M., Panchal, J.P., Moorthy, B.S., Blum, J.S., Joubert, M.K., Narhi, L.O. and Topp, E.M. (2016) 'Immunogenicity of Therapeutic Protein Aggregates', *Journal of Pharmaceutical Sciences*, 105(2), pp. 417–430. Available at: https://doi.org/10.1016/j.xphs.2015.11.002.

Nakamura, H., Oda, Y., Iwai, S., Inoue, H., Ohtsuka, E., Kanaya, S., Kimura, S., Katsuda, C., Katayanagi, K. and Morikawa, K. (1991) 'How does RNase H recognize a DNA.RNA hybrid?', *Proceedings of the National Academy of Sciences*, 88(24), pp. 11535–11539. Available at: https://doi.org/10.1073/pnas.88.24.11535.

Nakanishi, K., Sakiyama, T. and Imamura, K. (2001) 'On the adsorption of proteins on solid surfaces, a common but very complicated phenomenon', *Journal of Bioscience*

and Bioengineering, 91(3), pp. 233–244. Available at: https://doi.org/10.1016/S1389-1723(01)80127-4.

Niedziela-Majka, A., Kan, E., Weissburg, P., Mehra, U., Sellers, S. and Sakowicz, R. (2015) 'High-Throughput Screening of Formulations to Optimize the Thermal Stability of a Therapeutic Monoclonal Antibody', *SLAS Discovery*, 20(4), pp. 552–559. Available at: https://doi.org/10.1177/1087057114557781.

Otzen, D.E., Buell, A.K. and Jensen, H. (2021) 'Microfluidics and the quantification of biomolecular interactions', *Current Opinion in Structural Biology*, 70, pp. 8–15. Available at: https://doi.org/10.1016/j.sbi.2021.02.006.

PACE, C. (1990) 'Measuring and increasing protein stability', *Trends in Biotechnology*, 8, pp. 93–98. Available at: https://doi.org/10.1016/0167-7799(90)90146-O.

Pace, C.N. (1986) 'Determination and Analysis of Urea and Guanidine Hydrochloride Denaturation Curves', *Methods in Enzymology*, 131, pp. 266–280.

Pandey, L.M. (2022) 'Physicochemical factors of bioprocessing impact the stability of therapeutic proteins', *Biotechnology Advances*, 55, p. 107909. Available at: https://doi.org/10.1016/j.biotechadv.2022.107909.

Pang, K.T., Yang, Y.S., Zhang, W., Ho, Y.S., Sormanni, P., Michaels, T.C.T., Walsh, I. and Chia, S. (2023) 'Understanding and controlling the molecular mechanisms of protein aggregation in mAb therapeutics', *Biotechnology Advances*, 67, p. 108192. Available at: https://doi.org/10.1016/j.biotechadv.2023.108192.

Park, C. and Marqusee, S. (2005) 'Pulse proteolysis: A simple method for quantitative determination of protein stability and ligand binding', *Nature Methods*, 2, pp. 207–212. Available at: https://doi.org/10.1038/nmeth740.

Paul, A.J., Schwab, K., Prokoph, N., Haas, E., Handrick, R. and Hesse, F. (2015) 'Fluorescence dye-based detection of mAb aggregates in CHO culture supernatants', *Analytical and Bioanalytical Chemistry*, 407(16), pp. 4849–4856. Available at: https://doi.org/10.1007/s00216-015-8672-8.

Perkins, T.T., Smith, D.E. and Chu, S. (1997) 'Single polymer dynamics in an elongational flow', *Science*, 276(5321), pp. 2016–2021. Available at: https://doi.org/10.1126/science.276.5321.2016.

Poole, L.B. (2015) 'The basics of thiols and cysteines in redox biology and chemistry', *Free Radical Biology and Medicine*, 80, pp. 148–157. Available at: https://doi.org/10.1016/j.freeradbiomed.2014.11.013.

Rader, R.A. (2008) '(Re)defining biopharmaceutical', *Nature Biotechnology*, 26, pp. 743–751. Available at: https://doi.org/10.1038/nbt0708-743.

Rammensee, S., Slotta, U., Scheibel, T. and Bausch, A.R. (2008) 'Assembly mechanism of recombinant spider silk proteins', *Proceedings of the National Academy of Sciences*, 105(18), pp. 6590–6595. Available at: https://doi.org/10.1073/pnas.0709246105.

Rathore, N. and Rajan, R.S. (2008) 'Current Perspectives on Stability of Protein Drug Products during Formulation, Fill and Finish Operations', *Biotechnology Progress*, 24(3), pp. 504–514. Available at: https://doi.org/10.1021/bp070462h.

Raybould, M.I.J., Marks, C., Krawczyk, K., Taddese, B., Nowak, J., Lewis, A.P., Bujotzek, A., Shi, J. and Deane, C.M. (2019) 'Five computational developability guidelines for therapeutic antibody profiling', *Proceedings of the National Academy of Sciences*, 116(10), pp. 4025–4030. Available at: https://doi.org/10.1073/pnas.1810576116.

Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J.M. and Gaub, H.E. (1997) 'Reversible Unfolding of Individual Titin Immunoglobulin Domains by AFM', *Science*, 276(5315), pp. 1109–1112. Available at: https://doi.org/10.1126/science.276.5315.1109.

Rizzo, J., Gifford, L.K., Zhang, X., Gewirtz, A.M. and Lu, P. (2002) 'Chimeric RNA-DNA molecular beacon assay for ribonuclease H activity', *Molecular and Cellular Probes*, 16(4), pp. 277–283. Available at: https://doi.org/10.1006/mcpr.2002.0423.

Roche, J., Caro, J.A., Norberto, D.R., Barthe, P., Roumestand, C., Schlessman, J.L., Garcia, A.E., García-Moreno E., B. and Royer, C.A. (2012) 'Cavities determine the pressure unfolding of proteins', *Proceedings of the National Academy of Sciences*, 109(18), pp. 6945–6950. Available at: https://doi.org/10.1073/pnas.1200915109.

Rosen, L.E., Kathuria, S. V., Matthews, C.R., Bilsel, O. and Marqusee, S. (2015) 'Non-native structure appears in microseconds during the folding of E. coli RNase H', *Journal of Molecular Biology*, 427(2), pp. 443–453. Available at: https://doi.org/10.1016/j.jmb.2014.10.003.

Rosenberg, A.S. (2006) 'Effects of protein aggregates: An Immunologic perspective', *AAPS Journal*, 8(3), pp. E501–E507. Available at: https://doi.org/10.1208/aapsj080359.

Rost, B. and Sander, C. (1994) 'Conservation and prediction of solvent accessibility in protein families', *Proteins: Structure, Function, and Genetics*, 20(3), pp. 216–226. Available at: https://doi.org/10.1002/prot.340200303.

Rother, M. and Krzycki, J.A. (2010) 'Selenocysteine, Pyrrolysine, and the Unique Energy Metabolism of Methanogenic Archaea', *Archaea*, 2010, pp. 1–14. Available at: https://doi.org/10.1155/2010/453642.

Rousseau, F., Serrano, L. and Schymkowitz, J.W.H. (2006) 'How Evolutionary Pressure Against Protein Aggregation Shaped Chaperone Specificity', *Journal of Molecular Biology*, 355(5), pp. 1037–1047. Available at: https://doi.org/10.1016/j.jmb.2005.11.035.

Ruff, K.M., Choi, Y.H., Cox, D., Ormsby, A.R., Myung, Y., Ascher, D.B., Radford, S.E., Pappu, R. V. and Hatters, D.M. (2022) 'Sequence grammar underlying the unfolding and phase separation of globular proteins', *Molecular Cell*, 82(17), pp. 3193-3208.e8. Available at: https://doi.org/10.1016/j.molcel.2022.06.024.

Sadler, D.P., Petrik, E., Taniguchi, Y., Pullen, J.R., Kawakami, M., Radford, S.E. and Brockwell, D.J. (2009) 'Identification of a Mechanical Rheostat in the Hydrophobic Core of Protein L', *Journal of Molecular Biology*, 393(1), pp. 237–248. Available at: https://doi.org/10.1016/j.jmb.2009.08.015.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J., White, D. J., Hartenstain, V., Eliceiri, K., Tomancak, P., Cardona, A. (2012) 'Fiji: an open-source platform for biological-image analysis', *Nature Methods*, 9(7), pp. 676–682. Available at: https://doi.org/10.1038/nmeth.2019.

Schlierf, M., Li, H. and Fernandez, J.M. (2004) 'The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques', *Proceedings of the National Academy of Sciences of the United States of America*, 101(19), pp. 7299–7304. Available at: https://doi.org/10.1073/pnas.0400033101.

Schneider, S.W., Nuschele, S., Wixforth, A., Gorzelanny, C., Alexander-Katz, A., Netz, R.R. and Schneider, M.F. (2007) 'Shear-induced unfolding triggers adhesion of von Willebrand factor fibers', *Proceedings of the National Academy of Sciences of the United States of America*, 104(19), pp. 7899–7903. Available at: https://doi.org/10.1073/pnas.0608422104.

Schön, A., Clarkson, B.R., Siles, R., Ross, P., Brown, R.K. and Freire, E. (2015) 'Denatured state aggregation parameters derived from concentration dependence of protein stability', *Analytical Biochemistry*, 488, pp. 45–50. Available at: https://doi.org/10.1016/j.ab.2015.07.013.

Schulz, G.E. and Schirmer, R.H. (1979) *Principles of Protein Structure*. New York, NY: Springer New York (Springer Advanced Texts in Chemistry). Available at: https://doi.org/10.1007/978-1-4612-6137-7.

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. and Serrano, L. (2005) 'The FoldX web server: an online force field', *Nucleic Acids Research*, 33(Web Server), pp. W382–W388. Available at: https://doi.org/10.1093/nar/gki387.

Sharma, L.G. and Pandey, L.M. (2021) 'Shear-induced aggregation of amyloid β (1–40) in a parallel plate geometry', *Journal of Biomolecular Structure and Dynamics*, 39(17), pp. 6415–6423. Available at: https://doi.org/10.1080/07391102.2020.1798814.

Shire, S.J., Shahrokh, Z. and Liu, J. (2004) 'Challenges in the development of high protein concentration formulations', *Journal of Pharmaceutical Sciences*, 93(6), pp. 1390–1402. Available at: https://doi.org/10.1002/jps.20079.

De Simone, A., Dhulesia, A., Soldi, G., Vendruscolo, M., Hsu, S.-T.D., Chiti, F. and Dobson, C.M. (2011) 'Experimental free energy surfaces reveal the mechanisms of maintenance of protein solubility', *Proceedings of the National Academy of Sciences*, 108(52), pp. 21057–21062. Available at: https://doi.org/10.1073/pnas.1112197108.

Sing, C.E. and Alexander-Katz, A. (2010) 'Elongational Flow Induces the Unfolding of von Willebrand Factor at Physiological Flow Rates', *Biophysical Journal*, 98(9), pp. L35–L37. Available at: https://doi.org/10.1016/j.bpj.2010.01.032.

Sormanni, P., Aprile, F.A. and Vendruscolo, M. (2015) 'The CamSol Method of Rational Design of Protein Mutants with Enhanced Solubility', *Journal of Molecular Biology*, 427(2), pp. 478–490. Available at: https://doi.org/10.1016/j.jmb.2014.09.026.

Southall, N.T., Dill, K.A. and Haymet, A.D.J. (2002) 'A View of the Hydrophobic Effect', *The Journal of Physical Chemistry B*, 106(3), pp. 521–533. Available at: https://doi.org/10.1021/jp015514e.

Spudich, G.M., Miller, E.J. and Marqusee, S. (2004) 'Destabilization of the Escherichia coli RNase H Kinetic Intermediate: Switching between a Two-state and Three-state Folding Mechanism', *Journal of Molecular Biology*, 335(2), pp. 609–618. Available at: https://doi.org/10.1016/j.jmb.2003.10.052.

Sterpone, F., Derreumaux, P. and Melchionna, S. (2018) 'Molecular Mechanism of Protein Unfolding under Shear: A Lattice Boltzmann Molecular Dynamics Study', *Journal of Physical Chemistry B*, 122(5), pp. 1573–1579. Available at: https://doi.org/10.1021/acs.jpcb.7b10796.

Sturtevant, J.M. (1994) 'The thermodynamic effects of protein mutations', *Current Opinion in Structural Biology*, 4(1), pp. 69–78. Available at: https://doi.org/10.1016/S0959-440X(94)90062-0.

Tajima, M., Urabe, I., Yutani, K. and Okada, H. (1976) 'Role of calcium ions in the thermostability of thermolysin and Bacillus subtilis var. amylosacchariticus neutral protease.', *European journal of biochemistry*, 64(1), pp. 243–7. Available at: https://doi.org/10.1111/j.1432-1033.1976.tb10293.x.

Taneja, S. and Ahmad, F. (1994) 'Increased thermal stability of proteins in the presence of amino acids', *Biochemical Journal*, 303(1), pp. 147–153. Available at:

https://doi.org/10.1042/bj3030147.

Tartaglia, G.G. and Vendruscolo, M. (2008) 'The Zyggregator method for predicting protein aggregation propensities', *Chemical Society Reviews*, 37(7), p. 1395. Available at: https://doi.org/10.1039/b706784b.

Tessier, P.M., Vandrey, S.D., Berger, B.W., Pazhianur, R., Sandler, S.I. and Lenhoff, A.M. (2002) 'Self-interaction chromatography: a novel screening method for rational protein crystallization', *Acta Crystallographica Section D Biological Crystallography*, 58(10), pp. 1531–1535. Available at: https://doi.org/10.1107/S0907444902012775.

Thomas, C.R. and Dunnill, P. (1979) 'Action of shear on enzymes: Studies with catalase and urease', *Biotechnology and Bioengineering*, 21(12), pp. 2279–2302. Available at: https://doi.org/10.1002/bit.260211209.

Thomas, C.R. and Geer, D. (2011) 'Effects of shear on proteins in solution', *Biotechnology Letters*, 33(3), pp. 443–456. Available at: https://doi.org/10.1007/s10529-010-0469-4.

Tischenko, V.M., Zav'yalov, V.P., Medgyesi, G.A., Potekhin, S.A. and Privalov, P.L. (1982) 'A Thermodynamic Study of Cooperative Structures in Rabbit Immunoglobulin G', *European Journal of Biochemistry*, 126(3), pp. 517–521. Available at: https://doi.org/10.1111/j.1432-1033.1982.tb06811.x.

Toyooka, T. and Imai, K. (1984) 'New fluorogenic reagent having halogenobenzofurazan structure for thiols: 4-(aminosulfonyl)-7-fluoro-2,1,3-benzoxadiazole', *Analytical Chemistry*, 56(13), pp. 2461–2464. Available at: https://doi.org/10.1021/ac00277a044.

Treuheit, M.J., Kosky, A.A. and Brems, D.N. (2002) 'Inverse relationship of protein concentration and aggregation.', *Pharmaceutical research*, 19(4), pp. 511–6. Available at: https://doi.org/10.1023/a:1015108115452.

Tsumoto, K., Umetsu, M., Kumagai, I., Ejima, D., Philo, J.S. and Arakawa, T. (2004) 'Role of Arginine in Protein Refolding, Solubilization, and Purification', *Biotechnology Progress*, 20(5), pp. 1301–1308. Available at: https://doi.org/10.1021/bp0498793.

van der Veen, M.E., Van Iersel, D.G., Van der Goot, A.J. and Boom, R.M. (2004) 'Shear-Induced Inactivation of Alpha-Amylase in a Plain Shear Field', *Biotechnology Progress*, 20(4), pp. 1140–1145. Available at: https://doi.org/10.1021/bp049976w.

Vlasak, J. and Ionescu, R. (2011) 'Fragmentation of monoclonal antibodies.', *mAbs*, 3(3), pp. 253–63. Available at: https://doi.org/10.4161/mabs.3.3.15608.

Vriend, G. and Eijsink, V. (1993) 'Prediction and analysis of structure, stability and unfolding of thermolysin-like proteases', *Journal of Computer-Aided Molecular Design*, 7(4), pp. 367–396. Available at: https://doi.org/10.1007/BF02337558.

Wälchli, R., Vermeire, P.-J., Massant, J. and Arosio, P. (2020) 'Accelerated Aggregation Studies of Monoclonal Antibodies: Considerations for Storage Stability', *Journal of Pharmaceutical Sciences*, 109(1), pp. 595–602. Available at: https://doi.org/10.1016/j.xphs.2019.10.048.

Walsh, I., Seno, F., Tosatto, S.C.E. and Trovato, A. (2014) 'PASTA 2.0: an improved server for protein aggregation prediction', *Nucleic Acids Research*, 42(W1), pp. W301–W307. Available at: https://doi.org/10.1093/nar/gku399.

Wang, S., Zhang, X., Wu, G., Tian, Z. and Qian, F. (2017) 'Optimization of high-concentration endostatin formulation: Harmonization of excipients' contributions on colloidal and conformational stabilities', *International Journal of Pharmaceutics*, 530(1–

2), pp. 173–186. Available at: https://doi.org/10.1016/j.ijpharm.2017.07.057.

Wang, W. (2005) ‘Protein aggregation and its inhibition in biopharmaceutics’, *International Journal of Pharmaceutics*, 289(1–2), pp. 1–30. Available at: https://doi.org/10.1016/j.ijpharm.2004.11.014.

Wang, W. and Roberts, C.J. (2018) ‘Protein aggregation – Mechanisms, detection, and control’, *International Journal of Pharmaceutics*, 550(1–2), pp. 251–268. Available at: https://doi.org/10.1016/j.ijpharm.2018.08.043.

Wang, X., Das, T.K., Singh, S.K. and Kumar, S. (2009) ‘Potential aggregation prone regions in biotherapeutics’, *mAbs*, 1(3), pp. 254–267. Available at: https://doi.org/10.4161/mabs.1.3.8035.

Wang, X., Singh, S.K. and Kumar, S. (2010) ‘Potential Aggregation-Prone Regions in Complementarity-Determining Regions of Antibodies and Their Contribution Towards Antigen Recognition: A Computational Analysis’, *Pharmaceutical Research*, 27(8), pp. 1512–1529. Available at: https://doi.org/10.1007/s11095-010-0143-5.

Watzky, M. A.; Finke, R. G. (1997) 'Transition Metal Nanocluster Formation Kinetic and Mechanistic Studies. A New Mechanism When Hydrogen Is the Reductant: Slow, Continuous Nucleation and Fast Autocatalytic Surface Growth', *Journal of the American Chemical Society*, 119 (43), 10382–10400. Available at: https://doi.org/10.1021/ja9705102.

Whitehurst, C.B., Soderblom, E.J., West, M.L., Hernandez, R., Goshe, M.B. and Brown, D.T. (2007) ‘Location and Role of Free Cysteinyl Residues in the Sindbis Virus E1 and E2 Glycoproteins’, *Journal of Virology*, 81(12), pp. 6231–6240. Available at: https://doi.org/10.1128/JVI.02859-06.

Wilkinson, I. and Hale, G. (2022) ‘Systematic analysis of the varied designs of 819 therapeutic antibodies and Fc fusion proteins assigned international nonproprietary names’, *mAbs*, 14(1). Available at: https://doi.org/10.1080/19420862.2022.2123299.

Willis, L.F. (2018) ‘The effects of flow on therapeutic protein aggregation’, *White Rose eTheses Online*.

Willis, L.F., Kumar, A., Dobson, J., Bond, N.J., Lowe, D., Turner, R., Radford, S.E., Kapur, N. and Brockwell, D.J. (2018) ‘Using extensional flow to reveal diverse aggregation landscapes for three IgG1 molecules’, *Biotechnology and Bioengineering*, 115(5), pp. 1216–1225. Available at: https://doi.org/10.1002/bit.26543.

Willis, L.F., Kumar, A., Jain, T., Caffry, I., Xu, Y., Radford, S.E., Kapur, N., Vásquez, M. and Brockwell, D.J. (2020) ‘The uniqueness of flow in probing the aggregation behavior of clinically relevant antibodies’, *Engineering Reports*, 2(5). Available at: https://doi.org/10.1002/eng2.12147.

Willis, L.F., Toprani, V., Wijetunge, S., Sievers, A., Lin, L., Williams, J., Crowley, T.J., Radford, S.E., Kapur, N. and Brockwell, D.J. (2023) ‘Exploring a Role for Flow-Induced Aggregation Assays in Platform Formulation Optimisation for Antibody-Based Proteins’, *Journal of Pharmaceutical Sciences*, 00, pp. 1–12. Available at: https://doi.org/10.1016/j.xphs.2023.10.031.

Willis, L.F., Panagi, I., Page, A.D., Nathathan, T.M., Lawrence, S.M., Richer, R., Darton, N., Bond, N.J., Turner5, R.C., Kapur, N., Radford, S.E. and Brockwell, D.J. (no date) ‘A unified mechanism of flow-induced protein aggregation is governed by both surface and bulk-mediated processes’.

Willis, L.F., Trayton, I., Saunders, J.C., Brùque, M.G., Birch, W.D., Westhead, D.R., Day,

K., Devine, P.W.A., Lloyd, C., Kapur, N., Radford, S.E., Darton, N.J. and Brockwell, D.J. (no date) ‘Rationalising mAb candidate screening using a single holistic developability parameter’.

Wolf Pérez, A.-M., Lorenzen, N., Vendruscolo, M. and Sormanni, P. (2022) ‘Assessment of Therapeutic Antibody Developability by Combinations of In Vitro and In Silico Methods’, *Methods in Molecular Biology*, 2313, pp. 57–113. Available at: https://doi.org/10.1007/978-1-0716-1450-1_4.

Wu, H. and Randolph, T.W. (2020) ‘Aggregation and Particle Formation During Pumping of an Antibody Formulation Are Controlled by Electrostatic Interactions Between Pump Surfaces and Protein Molecules’, *Journal of Pharmaceutical Sciences*, 109(4), pp. 1473–1482. Available at: https://doi.org/10.1016/j.xphs.2020.01.023.

Xu, Y., Wang, D., Mason, B., Rossomando, T., Li, N., Liu, D., Cheung, J. K., Xu, W., Raghava, S., Katiyar, A., Nowak, C., Xiang, T., Dong, D. D., Sun, J., Beck, A., Liu, H. (2019) ‘Structure, heterogeneity and developability assessment of therapeutic antibodies’, *mAbs*, 11(2), pp. 239–264. Available at: https://doi.org/10.1080/19420862.2018.1553476.

Yang, A.-S. and Honig, B. (1993) ‘On the pH Dependence of Protein Stability’, *Journal of Molecular Biology*, 231(2), pp. 459–474. Available at: https://doi.org/10.1006/jmbi.1993.1294.

Young, T.S. and Schultz, P.G. (2010) ‘Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon’, *Journal of Biological Chemistry*, 285(15), pp. 11039–11044. Available at: https://doi.org/10.1074/jbc.R109.091306.

Zambrano, R., Jamroz, M., Szczasiuk, A., Pujols, J., Kmiecik, S. and Ventura, S. (2015) ‘AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures’, *Nucleic Acids Research*, 43(W1), pp. W306–W313. Available at: https://doi.org/10.1093/nar/gkv359.

Zhang, W., Wang, H., Feng, N., Li, Y., Gu, J. and Wang, Z. (2023) ‘Developability assessment at early-stage discovery to enable development of antibody-derived therapeutics’, *Antibody Therapeutics*, 6(1), pp. 13–29. Available at: https://doi.org/10.1093/abt/tbac029.