

The Effects of Flow on Therapeutic Protein Aggregation



UNIVERSITY OF LEEDS



Leon Fitzroy Willis

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

The University of Leeds

Astbury Centre for Structural Molecular Biology

School of Molecular and Cellular Biology

September 2018

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

© Leon Fitzroy Willis and the University of Leeds

Statement regarding inclusion of published work in this thesis

The candidate confirms that the work submitted is his/her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others:

Chapters 3 and 4 "Characterising protein aggregation induced by extensional flow" include work published in the following publication: 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*, 114 (18), 4673-4678. Chapter 4 "Mapping the aggregation landscapes of proteins under defined fluid fields" includes work from the following publication: Willis, L.F., Kumar, A., Dobson, J., Bond, N., 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, 1216–1225.

In Dobson et al. 2017 *Proc. Natl. Acad. Sci.*, The author contribution reads: " J.D., A.K., L.F.W., A.E.A., S.E.R., N.K., and D.J.B. designed research; J.D., A.K., and L.F.W. performed research; R. Tuma, D.R.H., R. Turner, and D.C.L. contributed new reagents/analytic tools.; J.D., A.K., L.F.W., R. Tuma, D.R.H., R. Turner, D.C.L., A.E.A., S.E.R., N.K., and D.J.B. analyzed data; and J.D., A.K., L.F.W., A.E.A., S.E.R., N.K., and D.J.B. wrote the paper." In Willis et al. 2018, *Biotechnol. Bioeng.* I prepared the vast majority of the protein samples ahead of stress in the flow device. I co-performed the vast majority of the flow stress experiments with A Kumar (joint-first author). I performed the vast majority of the pelleting assays, with A Kumar measuring the A280 of the vast majority of samples following overnight incubation. I performed the vast majority of the final analysis of the data, in addition to the writing of the manuscript (submitted as part of Year 3 Draft Paper and Thesis Plan in FBS). J Dobson helped conceive the study. All authors contributed to the writing of the final manuscript.

Acknowledgements

Firstly, I would like to thank my supervisory team, Dr David Brockwell, Prof Nikil Kapur and Prof Alison Ashcroft, for their guidance and encouragement of me over the past four years and hopefully, the years to come. I also wish to thank Professor Sheena Radford for her help and tutelage despite not formally being part of my supervisory team.

Secondly, I wish to thank the EPSRC Centre for Doctoral Training in Emergent Macromolecular Therapies for funding my studentship. In particular, I wish to thank Professor Paul Dalby and the other academics in the Department of Biochemical Engineering at UCL, who have introduced me to a field I did not know existed before my PhD, of which I now feel proud to be part of. I wish to thank the friends I made during my training events for making this experience as enjoyable as possible. I also wish to thank MedImmune LLC and Adimab LLC for providing additional funding, materials and advice over the last four years. Without this, most of the work in this thesis would not exist. I particularly wish to thank Drs David Lowe, Richard Turner, Nick Bond and Paul Varley (MedImmune) and Drs Tushar Jain, Yingda Xu and Max Vasquez (Adimab).

Thirdly, I wish to thank Drs John Dobson and Amit Kumar for their endless help, advice and friendship. As the original 'squishing team,' it's been great to get this project off the ground with you. I wish to thank all members of the Radford and Brockwell groups, past and present, for not just being great colleagues, but also great friends. I must also thank the lab manager, Mr G Nasir Khan, who always makes time for you despite his busyness! I wish to thank Drs Nuria Benseny-Cases, Sophie Goodchild, Esther Martin, James Ault and Iain Manfield for offering their time and advice to teach me many practical techniques throughout my PhD.

Finally, I wish to thank my past teachers who always encouraged me to aim higher. Thanks also to all my friends and family for their timeless support and encouragement- without you I would not have written this thesis! I particularly wish to thank my parents, Debbie and Fitzroy, their respective partners Chris and Tricia, and my grandparents Janet and Roy Wilson, as they have all made me the man I am today. I dedicate this thesis to my late grandparents, Iris and Fred Willis, who I hope would be proud.

At this point people usually thank their other half... As my luck has not stretched that far, I think I will end the sentiment here and talk instead about science!

Abstract

To date, over 70 monoclonal antibody (mAb) biopharmaceuticals have been approved, allowing effective treatment of serious diseases such as cancer. In addition to improving human health, these powerful medicines are very valuable, generating billions of dollars in sales annually. Like all proteins, environmental changes can cause mAbs to unfold, misfold and aggregate. Aggregation can block the progress of mAbs to market, as aggregates have been linked to adverse effects in patients. The hydrodynamic forces mAbs encounter during their manufacturing process have long been thought to be one of the causes of aggregation. This link remains tenuous, however, partly due to a lack of knowledge surrounding how specific flow fields (e.g. shear and extensional flows) perturb protein structure.

To assess the effects of flow on therapeutic protein aggregation, a recently developed, bespoke Extensional Flow Device (EFD) was characterised, which mimics the hydrodynamic forces mAbs encounter at manufacturing scale. In this thesis, the model proteins BSA and three mAbs (WFL, mAb1 and STT) were subjected to the defined fluid fields present in the EFD, with the resulting aggregates characterised using an array of biophysical techniques. The data show that protein aggregation can be induced by extensional flow. The extent of aggregation depends on a protein's sequence and topology, in addition to the flow conditions and buffer composition. For example, the mAbs WFL and STT show disparate aggregation behaviour following hydrodynamic stress, despite having >99 % sequence identity, with the generic mAb1 somewhere in between the two. Reinforced by data from a screen of 33 clinically relevant mAbs, the data in this thesis support future use of the EFD to: explore flow-induced protein aggregation mechanisms; improve mAb bioprocessing and; screen mAb candidates to select sequences and/or formulations which are resistant to potentially deleterious hydrodynamic forces, facilitating the development of next-generation mAb therapeutics.

Table of Contents

Contents

Statement regarding inclusion of published work in this thesis.....	ii
Acknowledgements.....	iii
Abstract	iv
Table of Contents	v
List of Tables	x
List of Figures	xi
List of Abbreviations	xvi
1. Introduction.....	1
1.1 Protein folding, misfolding and aggregation	1
1.2 Introduction to biopharmaceuticals	4
1.2.1 . Biological molecules as therapies: A revolution in modern medicine. 4	
1.2.2 Monoclonal antibodies (mAbs): The dominant products in the biopharmaceutical arena.....	5
1.2.3 Biochemical engineering: Selection, expression, purification and formulation of therapeutic antibodies	12
1.2.4 ‘The aggregation problem’ in the biopharmaceutical industry.....	22
1.2. Techniques used to study biopharmaceutical aggregation.....	26
1.3.1 <i>In silico</i> predictors of aggregation-prone regions.....	26
1.3.2. Spectroscopic and spectrometric techniques to probe higher-order structure	29
1.2.3. Separation-based methods to probe biopharmaceutical aggregation. 37	
1.2.4. Light scattering and related methods to detect, size and count aggregates in solution.....	40
1.4 Fluid flows and their effects on molecular and cellular structure	44
1.4.1 Defined fluid fields: Shear flow vs extensional flow	44
1.4.2 Effects of flow and hydrodynamic forces on cells	49

1.4.3 Effects of flow on polymers and DNA	50
1.4.4 Effects of flow on proteins	51
1.5 Basis of the study	54
1.6 Aim and Objectives	55
2. Materials and Methods.....	57
2.1. Materials.....	57
2.1.1 Equipment.....	57
2.1.2 Chemicals (all reagent grade unless stated).....	59
2.2. Methods.....	60
2.2.1. Proteins used in the study and their preparation.....	60
2.2.2 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)	62
2.2.3 Extensional flow apparatus and experimental setup.....	64
2.2.4. Dynamic light scattering (DLS).....	66
2.2.5 Nanoparticle Tracking Analysis	68
2.2.6 Transmission Electron Microscopy (TEM)	69
2.2.7 Insoluble protein pelleting assay	69
2.2.8 Native mass spectrometry of BSA.....	70
2.2.9 Fluorescence correlation spectroscopy (FCS) of BSA.....	71
2.2.10 IAEDANS labelling of BSA in the presence and absence of TCEP.....	72
2.2.11 Nile Red and 8-Anilinonaphthalene-1-sulphonic acid (ANS) fluorescence spectroscopy.....	73
2.2.12 Chemical cross-linking of WFL with BS ³	74
2.2.13 High-performance liquid chromatography (HPLC) monomer loss assay	75
2.2.14 Semi-denaturing sodium dodecyl sulphate polyacrylamide gel electrophoresis (SD-SDS-PAGE).....	75
2.2.15 Batch mode static light scattering	76
2.2.16 Seeding and cross-seeding assays.....	78
2.2.17 Adimab protein dataset experiments and analyses	78

2.2.18 Error propagation method	81
3. Characterising protein aggregation induced by extensional flow	83
3.1 Overview of the extensional flow device (EFD).....	83
3.1.1. Computational fluid dynamics under standard flow conditions	83
3.1.2. Computational fluid dynamics under variable flow conditions.....	86
3.2 Characterisation of BSA aggregation under defined fluid fields.....	87
3.2.1. Characterising BSA aggregate quantity and morphology	87
3.2.2 Measuring aggregate size and dispersity using biophysical techniques .	91
3.3 Assessing the aggregation of therapeutic antibodies under flow.....	99
3.4 Determining the key drivers of aggregation in the extensional flow device	104
3.4.1. Dissecting the role of shear in BSA's aggregation.....	105
3.4.2. Looking for evidence of particle shedding from the device	106
3.4.3 Conformation re-modelling of BSA under extensional flow.....	107
3.4.4 Assessing the role of energy in the EFD and its effects on BSA's	
aggregation.....	109
3.5 Discussion	112
4. Mapping the aggregation landscapes of proteins under defined fluid fields...	117
4.1. Determining the aggregation landscapes of BSA, WFL, mAb1 and STT ...	117
4.2 The effect of concentration on the extent of flow-induced aggregation	122
4.3 Understanding the role of shear in monoclonal antibody aggregation.....	124
4.4 The role of the buffer environment in flow-induced mAb aggregation	125
4.5 Manipulating the aggregation pathways of WFL and STT.....	127
4.6 Discussion.....	130
5. Towards a mechanism of flow-induced monoclonal antibody aggregation	133
5.1. Preface.....	133
5.2 Using chemical tools to probe the N* state	135
5.3 Following the pathway of mAbs from monomers to aggregates.....	137
5.3.1 Using HPLC to determine the loss of monomer as a function of pass	
number	137

5.3.2	Detecting oligomer formation using SD-SDS-PAGE	141
5.3.3	Using light scattering techniques to track the size and mass of aggregates over time	144
5.4	Characterising the size distribution, stability and nucleating ability of flow- induced mAb aggregates.....	148
5.5	Discussion	156
5.5.1	N*-state characterisation	156
5.5.2	Tracking aggregation as a function of pass number	157
5.5.3	Characterising nucleation in the EFD	160
5.5.4	The EFD mechanism data in the context of established aggregation models.....	160
6.	Screening the aggregation propensity of therapeutic mAbs using the extensional flow device	163
6.1	Overview of the Jain, Sun et al. <i>Proc. Natl. Acad. Sci.</i> (2017) Adimab paper 163	
6.2	Selection of a new mAb family from the Adimab dataset	169
6.3	Screening the Adimab IgGs using the extensional flow device	171
6.3.1	Generating the EFD pelleting data	171
6.3.2	Statistical comparison of the EFD data to those from twelve alternative biophysical assays.....	175
6.4	Using bioinformatics to attempt to rationalise the EFD data obtained for the 33 mAbs.....	181
6.4.1	Are some families of mAbs more aggregation-prone than others?	181
6.4.2	Do the simple, intrinsic properties of the mAb sequences correlate with the amount of aggregate formed by the EFD?	183
6.4.3	Using Solubis to identify APRs in the variable domains of the Adimab proteins.....	186
6.5	Discussion.....	194
7.	Concluding remarks and future directions	196
8.	Appendices	202
Appendix 1.	Ancillary Tables	202

Appendix 2. Ancillary Figures.....	207
9. References	213

List of Tables

- Table 1.1 A selection of some of the advantages and disadvantages of biopharmaceuticals compared to small-molecule drugs.....5
- Table 1.2. Summary of the aggregation algorithms.....29
- Table 1.3. Summary of the spectroscopic methods mentioned above.....33
- Table 1.4. Summary of the traditional separation-based methods for aggregate analysis.....39
- Table 2.1. Gel components for 12% (w/v) SDS-PAGE gels.....62
- Table 2.2. Initial cross-linking gel recipe.....63
- Table 2.3. Lower % (8% w/v) SDS-PAGE gel for later cross-linking experiments.....63
- Table 2.4. 6% (w/v) SD-SDS-PAGE gel recipe.....63
- Table 6.1 Summary of the twelve techniques used to analyse the 137 mAbs in the Jain et al. paper.....164&165

List of Figures

• Figure 1.1 Representation of a protein folding pathway (left) and ‘folding funnel’ (right).....	2
• Figure 1.2. Schematic of a protein aggregation pathway.....	3
• Figure 1.3. Topology diagram for a constant Ig domain.....	6
• Figure 1.4. The structure of IgG1 molecules.....	7
• Figure 1.5. Antibody fragments and a ‘knobs-into-holes’ bispecific antibody (bsAb).....	10
• Figure 1.6. Schematic of the antibody-drug conjugate (ADC) Kadcyla® (Ado-Trastuzumab emtansine).....	11
• Figure 1.7. An overview of the drug discovery pipeline.....	13
• Figure 1.8. Overview of mouse hybridoma technology.....	14
• Figure 1.9. Overview of phage display technology.....	15
• Figure 1.10. Overview of yeast display technology.....	16
• Figure 1.11. Overview of ribosome display.....	17
• Figure 1.12. Schematic of a typical mAb manufacturing process.....	19
• Figure 1.13. Potential inducers of mAb aggregation.....	25
• Figure 1.14. Mass spectrometry-based methods to study HOS in proteins including mAbs.....	36
• Figure 1.15. Schematic of AF4.....	39
• Figure 1.16. Schematic of Rayleigh scattering and detection.....	40
• Figure 1.17. Schematic of the NTA technology.....	43
• Figure 1.18. Schematic examples of shear flow (a) and extensional flow (b).....	45
• Figure 1.19. Schematics of commonly employed flow devices from the literature.....	48
• Figure 2.1a. Syringe and capillary used for extensional flow stress experiments.....	65
• Figure 2.1b. Mechanical driver of the extensional flow device.....	65
• Figure 2.1c. The connected syringes in the extensional flow device.....	66
• Figure 3.1. CFD results for the device, overlaid on a schematic of the EFD.....	84
• Figure 3.2. CFD result for the central streamline at a plunger velocity of 8 mm s ⁻¹	85
• Figure 3.3. CFD results for the EFD at variable plunger velocity.....	86

- Figure 3.4. Further purification of BSA using gel-filtration chromatography (see Methods Section 2.2.1).....88
- Figure 3.5. Determination of the amount of BSA lost from solution following hydrodynamic stress in the EFD.....89
- Figure 3.6. TEM images of BSA after subjection to hydrodynamic stress in the EFD at a plunger velocity of 8 mm s⁻¹ for the number of passes and concentration stated.....90
- Figure 3.7. Representative correlation functions obtained by DLS for Quiescent BSA (left) and BSA after 2000 passes in the EFD (right).....91
- Figure 3.8. Plot of the z-average radii of BSA stressed for 0 (quiescent sample) to 2000 passes at a plunger velocity of 8 mm s⁻¹, obtained by Cumulants analysis of the DLS data.....92
- Figure 3.9. Plot of polydispersity index (PDI) values, calculated from the z-average radii measurements (and their widths) obtained from Cumulants DLS analysis of the BSA samples in Figure 3.8, against the number of passes.....93
- Figure 3.10. DLS Regularisation plot for quiescent BSA (black) and BSA after 2000 passes of stress at a plunger velocity of 8 mm s⁻¹.....94
- Figure 3.11. Particle number concentration vs size plots for stressed BSA samples analysed using NTA.....95
- Figure 3.12. Total particle count for the stressed BSA samples detected by NTA in Figures 3.10 a–d.....96
- Figure 3.13. Plot of the average R_H values obtained from fitting FCS data for quiescent (0 passes) and stressed (10 – 2000 passes) BSA solutions.....97
- Figures 3.14. Native mass spectra for a) quiescent BSA (5 mg mL⁻¹) b) after 20 passes and c) after 100 passes.....98
- Figure 3.15. Biophysical characterisation of WFL following stress in the EFD at a concentration of 0.5 mg mL⁻¹ at a plunger velocity of 8 mm s⁻¹.....100
- Figure 3.16. Biophysical characterisation of STT following stress in the EFD at a concentration of 0.5 mg mL⁻¹ at a plunger velocity of 8 mm s⁻¹.....101
- Figure 3.17. Biophysical characterisation of mAb1 following stress in the EFD at a concentration of 0.5 mg mL⁻¹ at a plunger velocity of 8 mm s⁻¹.....102
- Figure 3.18 Plot of the amount of protein left in solution following 0 (quiescent), 20 or 100 passes of stress in the flow device, then the samples analysed with the insoluble protein pelleting assay.....103

- Figure 3.19. Plot of the amount of excess protein in the pellet fraction for 5 mg mL⁻¹ BSA following 1000 passes of stress at 8 mm s⁻¹.....105
- Figure 3.20. Correlation function plot obtained by DLS analysis of 500 µL of 25 mM ammonium acetate, pH 5.1 following 1000 passes of stress in the EFD (plunger velocity = 8 mm s⁻¹).....106
- Figure 3.21. SDS-PAGE analyses of 5 mg mL⁻¹ BSA, incubated with 5 mM IAEDANS and/or 0.5 mM TCEP, in the presence and absence of flow.....108
- Figure 3.22. Jaspe and Hagen elementary model for proteins unfolding due to flow.....109
- Figure 3.23. Distributions and dissipation of energy in the EFD.....110
- Figure 3.24. Quantification of the amount of insoluble BSA formed when stressed at different strain rates111
- Figure 4.1. Overview of the parameter space that could potentially be explored with the flow device.....118
- Figure 4.2. Aggregation landscapes for a) BSA, b) WFL c) mAb1 and d) STT.....119
- Figure 4.3. Plot of percentage aggregated protein for WFL (green) and STT (blue) following 20 and 50 passes of stress in the EFD (respectively) at a concentration of 0.5 mg mL⁻¹ at a plunger velocity of 0.5 mm s⁻¹.....121
- Figure 4.4 Titration plot of % protein in pellet for WFL (green), STT (blue) and mAb1 (magenta) following stress for 100 passes in the EFD at different concentrations.....123
- Figure 4.5 Shear-length dependence of the three IgG1s in the EFD. 124
- Figure 4.6. Buffer screen of the mAbs with the EFD. WFL (green), STT (blue) and mAb1 (magenta) were dialysed into (left to right).....126
- Figure 4.7 The shear-dependence of WFL stressed for 100 passes in arginine+succinate buffer.....128
- Figure 4.8 Shear-dependence of STT at a plunger velocity of 16 mms⁻¹ ($\dot{\epsilon} = 23,421 \text{ s}^{-1}$, $\dot{\gamma} = 104,000 \text{ s}^{-1}$).....129
- Figure 5.1. Outline of the proposed mechanism of aggregation of mAb aggregation induced by extensional flow.....134
- Figure 5.2. Chemical cross-linking of WFL to identify its dimer interface..136
- Figure 5.3. HPLC chromatograms for a) WFL, b) mAb1 and c) STT after stress for the given number of passes at a plunger velocity of 8 mm s⁻¹.....138

- Figure 5.4. Plot of percentage monomer lost as determined by HPLC (left-hand y- axis and black lines) vs loss of protein from the supernatant (right-hand y-axis and blue lines) as a function of the number of passes through the device (x-axis).....140
- Figure 5.5. 6% (w/v) Semi-denaturing SDS-PAGE of a) WFL (green), b) mAb1 (magenta) and c) STT (blue) samples following stress in the EFD.....142
- Figure 5.6. Semi-denaturing SDS-PAGE of a) WFL, b) mAb1 and c) STT samples following stress in the EFD.....143
- Figure 5.7. Cumulants analysis of DLS data for WFL (green), mAb1 (pink) and STT (blue) obtained following no stress (0 passes) or 1-100 passes in the EFD.....145
- Figure 5.8. Polydispersity index (PDI) calculation of the samples shown in Figure 5.7.....146
- Figure 5.9. Plot of uncertainty averaged molecular weight (M_{avg}) for WFL, mAb1 and STT samples determined using static light scattering.....146
- Figure 5.10. Regularisation analysis of DLS data for mAb1 a) after 50 passes, b) supernatant following 50 passes and centrifugation and c) quiescent mAb1.....149
- Figure 5.11. Seeding experiment and supernatant characterisation with a) WFL (green) and b) STT (blue).....152
- Figure 5.12. Cross-seeding assay of WFL (green) with STT (blue) in 150 mM ammonium acetate buffer, pH 6.0.....154
- Figure 5.13. Cross-seeding assay of WFL (green) with STT (blue) in 125 mM L-arginine + 20 mM sodium succinate buffer, pH 6.0.....155
- Figure 5.14. Proposed aggregation models in the EFD.....161
- Figure 6.1. Hierarchical clustering tree of biophysical assays in the Jain et al. paper.....166
- Figure 6.2. Cluster diagram of the 137 mAbs according to their relatedness in the biophysical assays.....168
- Figure 6.3. Plot of 'distance from ideal' red flag rank against protein number from Figure 6.2.....170
- Figure 6.4. 'Ranging shot' for Daratumumab.....171
- Figure 6.5. Plot of the percentage insoluble protein following stress in the EFD.....173
- Figure 6.6. Histogram and cumulative frequency diagram for the pelleting data in Figure 6.5.....174

- Figure 6.7. Box and whisker diagram showing the common statistical features in the pelleting data from Figure 6.5.....174
- Figure 6.8. Hierarchical clustering ‘Family trees’ of assays used to assay the biophysical properties of the 33 Adimab IgG1s.....176
- Figure 6.9. Cluster diagrams for the 33 mAbs according to their relatedness in the biophysical assays, both without (left) and with (right) flow data.....178
- Figure 6.10. Plot of % protein in pellet following stress in the EFD against the cluster number for the Adimab mAbs when a) flow is excluded and b) included in the analysis.....179
- Figure 6.11. Scatter plot of the rank of each mAb including and excluding data from the extensional flow device.....180
- Figure 6.12. Phylogenetic tree for the 33 mAbs. The mAbs at the bottom of the tree are least related to those at the top.....182
- Figure 6.13. Plot of the % protein in pellet (from Figure 6.5) against a) pI and b) Intrinsic CamSol score at pH 7 for the scFv of the 33 mAbs stressed in the study.....183
- Figure 6.14. Plot of the % protein in pellet (from Figure 6.5) against a) Net CDR net charge at pH 7 for all 33 mAbs and b) K_D for 26/33 mAbs stressed in the study against their targets.....185
- Figure 6.15. scFv homology models (V_H and V_L , left and right respectively) for a) Alemtuzumab, b) Etralizumab, c) Guselkumab and d) Bevacizumab....187
- Figure 6.16. scFv homology models (V_H and V_L , left and right respectively) for a) Galiximab, b) Ipilimumab, c) Fezakinumab and d) Golimumab.....188
- Figure 6.17. scFv homology models (V_H and V_L , left and right respectively) for a) Vedolizumab, b) Foralumab, c) Daclizumab and d) Cixutumumab.....189
- Figure 6.18. ‘Stretch plot’ of the APRs identified in the 33 mAbs using Solubis.....190
- Figure 6.19. Plot of the Solubis score ($\Delta G^{\text{contrib}} \times \text{TANGO score}$ for the total number of APRs) vs the % protein in pellet formed in the pelleting assay for each mAb after stress.....191
- Figure 6.20. Plot of the summed TANGO score against the change in interfacial APR solvent-accessible surface area for the 33 Adimab scFvs..193

List of Abbreviations

- Ab = Antibody
- AC-SINS = Affinity Capture- Self-Interaction Nanoparticle Spectroscopy
- ADC = Antibody-Drug Conjugate
- ADCC = Antibody-Dependent Cellular Cytotoxicity
- ADH = Alcohol Dehydrogenase
- AF4 = Asymmetrical Flow Field-Flow Fractionation
- AFM = Atomic Force Microscopy
- ANS = 8-Anilino-1-Naphthalenesulfonic Acid
- APR = Aggregation-Prone Region
- APS = Ammonium Persulphate
- Arg = Arginine
- AS = Accelerated Stress
- Asn = Asparagine
- Asn = Asparagine
- Asp = Aspartic Acid
- AUC = Analytical Ultra Centrifugation
- BS³ = Bis(Sulfosuccinimidyl)Suberate
- BSA = Bovine Serum Albumin
- bsAb = Bi-Specific Antibody
- BVP = Baculovirus Particle
- CD = Circular Dichroism
- CDR = Complementarity-Determining Region
- CFD = Computational Fluid Dynamics
- C_H = Constant Heavy Domain
- CHO = Chinese Hamster Ovary
- CIC = Cross-Interaction Chromatography
- C_L = Constant Light Domain
- CPL = Circularly Polarised Light
- CSI-BLI = Clone Self-Interaction by Biolayer Interferometry
- DAR = Drug:Antibody Ratio
- DHFR = Dihydrofolate Reductase
- DLS = Dynamic Light Scattering

- DO = Dissolved Oxygen
- DSF = Differential Scanning Fluorimetry
- DSP = Downstream Processing
- DTT = Dithiothreitol
- EFD = Extensional Flow Device
- ELISA = Enzyme-Linked Immunosorbent Assay
- EMA = European Medicines Agency
- ESI = Electrospray Ionisation
- Fab = Antigen-Binding Fragment
- FACS = Fluorescence-Assisted Cell Sorting
- Fc = Crystallisable Fragment
- FCS = Fluorescence Correlation Spectroscopy
- FDA = Food and Drug Administration
- FTIR = Fourier Transform Infra-Red
- Glu = Glutamic Acid
- HCP = Host Cell Proteins
- HDX = Hydrogen-Deuterium Exchange
- HEK = Human Embryonic Kidney
- HEPES = 4-(2-Hydroxyethyl)Piperazine-1-Ethanesulphonic Acid
- HIC = Hydrophobic Interaction Chromatography
- HOS = Higher-Order Structure
- HPLC = High-Performance Liquid Chromatography
- IAEDANS = 5-{2-(Iodoacetamido)Ethylamino}Naphthalene-1-Sulphonic Acid
- ICH = International Council for Harmonisation
- IEX = Ion-Exchange Chromatography
- Ig = Immunoglobulin
- IM-MS = Ion- Mobility-Mass Spectrometry
- LENP = Lumry-Eyring Nucleation Polymerisation
- m/z = Mass-To-Charge Ratio
- mAb = Monoclonal Antibody
- MACS = Magnetism-Assisted Cell Sorting
- MALLS = Multi-Angle Laser Light Scattering
- MD = Molecular Dynamics
- MoA = Mechanism of Action
- MS = Mass Spectrometry

- MS/MS = Tandem Mass Spectrometry
- Nb = Nanobody
- NGF = Nerve Growth Factor
- NMR = Nuclear Magnetic Resonance
- NTA = Nanoparticle Tracking Analysis
- PASTA = Prediction of Amyloid Structure Aggregation
- PCR = Polymerase Chain Reaction
- PDI = Polydispersity Index
- Pro = Proline
- PSR = Polyspecificity Reagent Binding Assay
- QbD = Quality by Design
- R_H = Hydrodynamic Radius
- SAP = Spatial Aggregation Propensity
- SASA = Solvent-Accessible Surface Area
- scFv = Single-Chain Variable Fragment
- SD-SDS-PAGE = Semi-Denaturing Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
- SEC = Size-Exclusion Chromatography
- SG = Salt-Gradient AC-SINS
- SLS = Static Light Scattering
- SMAC = Standup Monolayer Adsorption Chromatography
- SV = Sedimentation Velocity
- TCEP = Tris(2-Carboxyethyl)Phosphine Hydrochloride
- TDA = Taylor-Dispersion Analysis
- TEM = Transmission Electron Microscopy
- TEMED = Tetramethylethylenediamine
- TNF = Tumour Necrosis Factor
- TPP = Target Product Profile
- Tris = Tris(hydroxymethyl) aminomethane
- Trp = Tryptophan
- UF/DF = Ultrafiltration/Diafiltration
- USP = Upstream Processing
- V_H = Variable Heavy Domain
- V_L = Variable Light Domain
- vWf = Von Willebrand Factor

1. Introduction

1.1 Protein folding, misfolding and aggregation

Proteins are one of Nature's most important classes of macromolecule, carrying out many vital life processes. Proteins are biological polymers, made up of amino acid monomers. There are 20 such canonical α -amino acids which constitute the molecular building blocks of a protein (Voet and Voet, 2011). These amino acids are linked together by peptide bonds to form a polypeptide chain. The primary sequence contains all of the necessary chemical information to allow the protein to fold into its correct three-dimensional structure (Anfinsen et al., 1961; Anfinsen, 1973). Structural collapse and burial of hydrophobic amino acid residues into the core of proteins, *via* the hydrophobic effect, makes this process thermodynamically favourable (Cheung et al., 2002; Fersht, 1999).

It was originally thought that proteins fold along a linear pathway. But this is simplistic, especially considering the vast number of conformations available to a nascent polypeptide chain other than the correctly folded, native three-dimensional structure. Levinthal proposed that no protein would be able to randomly sample these conformations to get to the desired structure on a biologically relevant timescale and that it must therefore be a guided search (Levinthal, 1968). As the protein folding field came into fruition, the 'pathway' idea (from unfolded to folded protein) evolved to that of an energy landscape (Anfinsen, 1973; Englander and Mayne, 2014). Proteins fold energetically downhill, decreasing their conformational freedom in the process (Dill and Chan, 1997; Englander and Mayne, 2014). These two ideas are summarised in Figure 1.1.

There is not necessarily one exclusive route through which a protein can fold to the native state; this can occur through various paths (Radford et al., 1992; Radford and Dobson, 1995). Any intermediate states on the pathway can have some secondary structure, but often lack the packed hydrophobic core of a properly folded protein. These partially structured intermediates have thus been described as 'molten globules' (Roder and Colón, 1997).

These intermediate states present proteins with an additional problem; they are susceptible to being trapped in 'off-pathway' kinetic wells. These misfolded species are capable of self-associating in a process called aggregation (Dobson, 2004; Jahn and Radford, 2008). Proteins are described as metastable entities, in

that there is a fine energetic balance between their native and structurally perturbed intermediate states. ΔG_{unfold} is of the order of 20-60 kJ mol⁻¹ for the unfolding of a protein from the native state (Fersht, 1999).

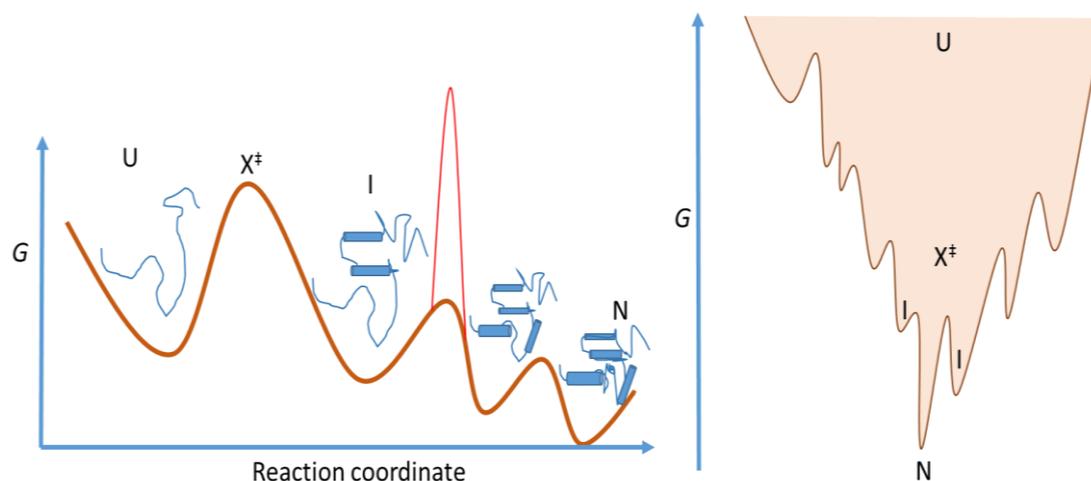


Figure 1.1 Representation of a protein folding pathway (left) and ‘folding funnel’ (right). On the folding pathway, the unfolded protein (U) must go through transition states (X^{\ddagger}). Any intermediates (I) may be trapped by kinetic barriers (red). Once these have been traversed, the protein can fold into the native structure (N), which has the lowest free energy (G). The folding funnel illustrates how some protein folding pathways have many more intermediate states and kinetic traps than others. This means one route may be much smoother than another going from the unfolded to the correctly folded native state (N). Figure adapted from Englander and Mayne, 2014.

The number of favourable *intramolecular* contacts within a protein increases as it folds from the unfolded to the native ground state (Fersht, 1999). But favourable *intermolecular* interactions can also form between proteins, allowing them to form complexes (Garcia-Seisdedos et al., 2017). Whilst protein oligomerisation can be functionally desirable with natively folded proteins as the substrates, it is generally deleterious to function if this occurs with unfolded or misfolded proteins (Jahn and Radford, 2008; Roberts, 2014; Williamson, 2012). An aggregate can thus be defined as any proteinaceous entity composed of two or more non-native protein monomers (Ratanji et al., 2014). The increased number of intermolecular, contacts (typically hydrophobic and van der Waals) which form between the proteins in an aggregate can make aggregation enthalpically favourable. This is especially true of ordered, fibrillar aggregates- amyloid fibrils (Jahn and Radford,

2008; Roberts, 2014; Voet and Voet, 2011). Amyloids are primarily characterised by their cross- β architecture, making them very stable (Chiti and Dobson, 2009; Roberts, 2014; Wei and Roberts, 2010).

Aggregates can form either reversibly or irreversibly, depending on the protein in question and its environment (Mahler et al., 2009; Wang et al., 2010; Wang and Roberts, 2010). Aggregation pathways are generally described by long-standing models such as the Lumry-Eyring Nucleation Polymerisation (LENP) model (Andrews and Roberts, 2007; Li and Roberts, 2009; Lumry and Eyring, 1954). The exact pathway a protein traverses will once again depend on its physico-chemical environment and the nature of the protein itself (Figure 1.2) (Roberts, 2014).

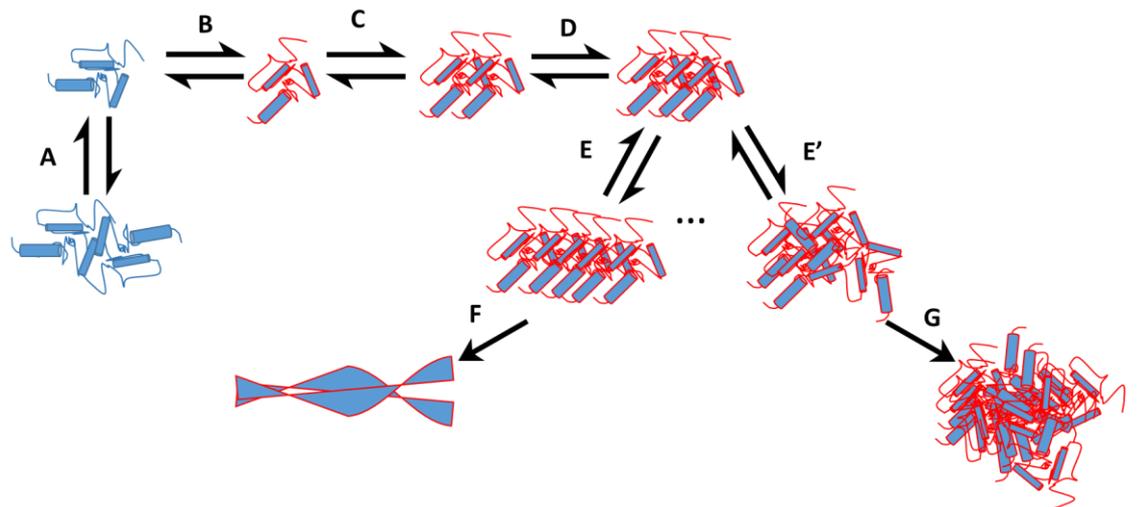


Figure 1.2. Schematic of a protein aggregation pathway. The association of natively folded monomers to form complexes can often be functionally desirable (A). Environmental perturbation (See Section 1.2.4) can lead to structural perturbation of the native protein, causing it to partially unfold (B). These partially unfolded species can then aggregate to form dimers (C) and trimers (D). These reversible association steps can continue in an ordered (E) or disordered fashion (E'). Ordered aggregates can eventually form (amyloid fibrils, F) or insoluble amorphous aggregates (G). These final species are formed irreversibly, representing the thermodynamic end-point of the aggregation pathway. Adapted from Roberts 2014.

Aggregation is thus a key problem faced by proteins. *In vivo*, cells use a variety of strategies to prevent this from happening to nascent proteins synthesised on the ribosome. A network of molecular chaperones (such as Hsp70 and Hsp90) bind to the exposed hydrophobic residues of folded and misfolded proteins to prevent their aggregation (Morán Luengo et al., 2018; Young et al., 2004). Chaperonins, such as GroEL, use the power of ATP to sequester unfolded proteins *in vivo* and assist their folding (Yan et al., 2018). Sometimes these mechanisms break down, leading to aggregation diseases (Dobson, 2004; Jahn and Radford, 2008) or bacterial cell death (Khodaparast et al., 2018). However, how environmental stresses perturb proteins, as well as how proteins evade aggregation *in vitro*, remain key questions in the protein aggregation field. These questions will be posed time and again in this thesis.

1.2 Introduction to biopharmaceuticals

1.2.1. Biological molecules as therapies: A revolution in modern medicine

Small-molecule therapies, such as aspirin, have been used to treat diseases for in excess of 100 years (Sneader, 2000). However, as time and biochemical knowledge have progressed, larger biological molecules have been developed as therapeutic entities (Walsh, 2003). These therapies, commonly referred to as biologics or biopharmaceuticals, have been in existence in a modern guise for much less time than small molecules. The different modes of action afforded to biologics compared to their small-molecule counterparts make them therapeutically advantageous. Vaccines such as those against tuberculosis and tetanus are amongst the most common and powerful early biologics (Walsh, 2003).

Protein-based therapies have emerged as the dominant player in the biologics sphere (Aggarwal, 2014). Early examples include blood factors, e.g. Factor VIII (Schramm, 2014) and insulin (Zhou et al., 2016), initially derived from mammalian sources such as human blood donations or cattle pancreases, respectively. Animal-derived therapeutics pose obvious moral and ethical questions. The advent of recombinant DNA technology led to the first recombinant human protein therapeutic, Humulin® (human insulin), being developed by Eli Lilly and

approved by the Food and Drug Administration (FDA) in 1982 (Sanlioglu et al., 2013). Since this time, many other recombinant enzymes and proteins have been approved for therapeutic use in the 1980's and beyond (Appendix 1 in (Walsh, 2003)). The advantages and disadvantages of small-molecule drugs and biopharmaceuticals are summarised in Table 1.1.

Table 1.1 A selection of some of the advantages and disadvantages of biopharmaceuticals compared to small-molecule drugs. Advantages are underlined in green, with disadvantages underlined in red.

Small-molecule drug	Biopharmaceutical
<u>Organic chemistry is an older and well-established field</u>	<u>Younger field with less expertise</u>
<u>Wide-range of delivery routes</u>	<u>Delivery options limited</u>
<u>Small size (<500 Da) makes characterisation and analysis easier</u>	<u>Large size (10's-100's kDa) makes analysis complex</u>
<u>Increased toxicity risk due to off-target interactions</u>	<u>Off-target interactions less likely, leading to low toxicity</u>
<u>Potential for weaker interaction strength with target</u>	<u>Potent, highly specific binding to the target</u>

1.2.2 Monoclonal antibodies (mAbs): The dominant products in the biopharmaceutical arena

Very recently, cell and gene therapies are emerging as the latest classes of biopharmaceutical being developed to treat individual patients with great success (Bajgain et al., 2018; Dunbar et al., 2018). However, the monoclonal antibody (mAb) has emerged as the most profitable and most important class of biologic over the last thirty years (Aggarwal, 2014), with over 74 approved to date and several hundred currently in clinical trials (Carter and Lazar, 2018; Strohl, 2018). Additionally, mAbs are also widely used as diagnostic reagents in biochemical assays (Clark, 2006; Vallée et al., 2016).

Antibody structure and function

Antibodies (Abs) are the immunoglobulin proteins used by the humoral immune system to bind to foreign macromolecular entities. These proteins feature immunoglobulin domains of ~110 residues, comprising a β -sheet rich Immunoglobulin (Ig) fold (Voet and Voet, 2011; Williamson, 2012). A conserved disulphide bridge 'staples' the two sheets of the Ig fold together (Huber et al., 1976). The structure and topology of the Ig domain is shown in Figure 1.3.

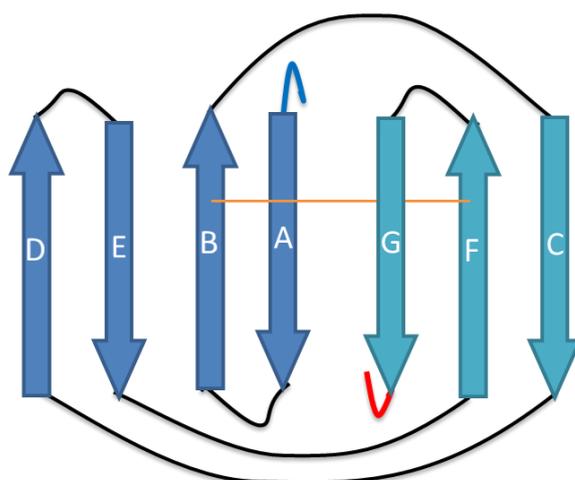


Figure 1.3. Topology diagram for a constant Ig domain. The four beta strands in dark blue comprise the inner sheet of the Ig domain, whilst those in light blue comprise the outer sheet. Black lines denote the loops connecting the strands. The blue and red lines at the end of strands A and G denote the N- and C-termini respectively. The orange line denotes the disulphide bond linkage. Adapted from (Spada et al., 1998).

Abs are heterodimeric proteins comprising two heavy chains and two light chains (Schroeder et al., 2010). The light chains possess two Ig domains, whereas the heavy chains possess either four or five Ig domains (Schroeder et al., 2010; Williamson, 2012).

Humans possess five different isotypes of antibody, which differ in terms of their type of heavy chain: IgM (950 kDa), IgA (360–720 kDa), IgE (190 kDa), IgD (160 kDa) and IgG (150 kDa) (Schroeder et al., 2010; Voet and Voet, 2011). The IgG isotype dominates the biopharmaceutical sector, due to them possessing long half-lives *in vivo* (Schroeder et al., 2010). As such, IgGs will be the focus of the thesis. A schematic of an IgG is shown in Figure 1.4.

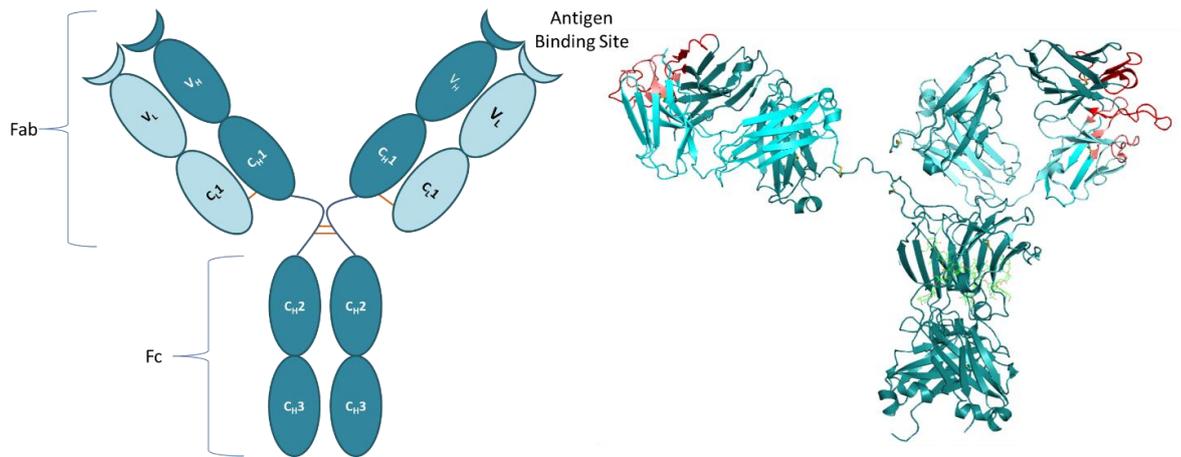


Figure 1.4. The structure of IgG1 molecules. Schematic (left): Each ellipse represents one β -sheet rich Ig domain. The heavy chains (dark turquoise) and light chains (light turquoise) are connected with disulphide bridges (orange). The heavy chains have three constant domains (C_{H1-3}) whilst the light chains only possess one (C_{L1}). Two disulphide bridges connect the heavy chains together near the flexible hinge region (blue lines) of the molecule. The bottom half of the antibody represents the crystallisable fragment (Fc), whereas the top half of the molecule is the antigen binding fragment (Fab). Three hypervariable loops are responsible for antigen binding in each of the two variable domains (V_H and V_L for variable heavy and light respectively) at the top of the molecule. **Crystal structure (right):** The crystal structure of an IgG1, colour-coded as in the schematic. CDRs in the variable domains are coloured in red and pink (V_H and V_L) respectively. N-linked glycans are coloured in green. Schematic adapted from (Buss et al., 2012). Crystal structure from (Saphire et al., 2001).

The variable regions of an antibody bind to their target antigens which reside at the top of the antigen binding Fab region (Figure 1.2). The motif recognised by the antibody is called the epitope, whilst the binding interface on the antibody is called the paratope (Feige and Buchner, 2014; Schroeder et al., 2010). Computational analysis of antibody paratopes have shown they are enriched in aromatic amino acids (Tyr, Phe and Trp), in addition to short-chain hydrophilic amino acids (Ser, Thr, Asn, Asp and Gly) (Peng et al., 2014). These residues help drive the specificity and high affinity of mAbs to their targets. The loops which afford antibodies their binding ability are known as complementarity-determining regions (CDRs) (Tiller and Tessier, 2015).

The hinge and Fc regions can also play important roles in the antibody's function. The hinge region affords an IgG conformational flexibility. IgGs can be further categorised into different subclasses (Vidarsson et al., 2014). These subclasses are IgG1, IgG2, IgG3 and IgG4. They have various structural differences which affect their function *in vivo*: their hinge regions can be of varying length (IgG3's have hinges four or five times the length of the other classes'); the number of disulphides connecting the heavy chains can differ (IgG1 and IgG4 = 2, IgG2 = 4 and IgG3s = 11) and their so-called effector functions, mediated through the Fc region, can differ widely (Buss et al., 2012; Vidarsson et al., 2014).

The Fc region contains a conserved Asn297 residue in the C_H2 domain, which is glycosylated in all IgGs (Vidarsson et al., 2014). The glycan greatly impacts the conformation of an IgG (Krapp et al., 2003; Zhang, 2015), affecting not only Fc receptor binding (Lin et al., 2015) but also the thermal stability of the mAb as a whole (Zheng et al., 2011).

Cells possess several different receptors capable of binding to the Fc region (Wang et al., 2008). FcRn binding is important in maintaining the long half-lives of IgGs *in vivo*, as FcRn binding 'recycles' mAbs inside cells (Wang et al., 2008). Other receptors give IgGs 'effector function'. Interaction with the Fcγ receptor can trigger antibody-dependent cellular cytotoxicity (ADCC) (Buss et al., 2012). mAbs can also activate the complement system through binding to the C1q receptor (Buss et al., 2012; Williamson, 2012). This ultimately results in death of the target cell (Carter and Lazar, 2018). IgG1 and IgG3 mAbs are far better at activating these latter mechanisms than IgG2 and IgG4s, thus one can tailor the isotype of the mAb to suit the desired mechanism of action (MoA) (Buss et al., 2012; Carter and Lazar, 2018). This wide diversity also explains why mAbs are so common in the biopharmaceutical arena.

[New mAb modalities](#)

A mAb in its conventional form is not the only type of antibody therapeutic in development or even on the market (Aggarwal, 2014; Carter and Lazar, 2018; Elgundi et al., 2017; Rodgers and Chou, 2016). The discovery that camelids (camels, llamas, alpacas etc.) produce antibodies which lack a light chain in 1993 (Hamers-Casterman et al., 1993) has led to their increasing use as research tools (Hu et al., 2017) and potential therapeutics (Peyvandi et al., 2016). These 'nanobodies' (Nbs) are much smaller than their mAb counterparts, consisting of

just one variable domain, giving them high levels of tissue penetration (Hu et al., 2017). Other antibody fragment molecules have been constructed, including single-chain variable fragments (scFvs) (Weisser and Hall, 2009), as well as just the Fab fragments (if Fc function is unnecessary) (Kennedy et al., 2017). scFvs consist of one V_H and V_L domain connected together with a short GS-linker (Weisser and Hall, 2009). Fab fragments can exist as either monomers or as a Fab' dimer (i.e. a full-length mAb minus the Fc region) (Kennedy et al., 2017). Fc fusion proteins have also been generated, whereby proteins or peptides are fused to an antibody Fc to extend the *in vivo* half-life of the former molecule. Etanercept is the most prominent example of an Fc-fusion, with this Tumour Necrosis Factor (TNF)- α inhibitor having 2016 sales exceeding \$8.9 bn (Strohl, 2018).

Advances in protein engineering have allowed biotechnologists to construct various novel modalities from antibody fragments to increase the functionality of the final molecule. For example, two different binding specificities can be exhibited by bi-specific antibodies (bsAbs). This can involve swapping the chains of two individual mAbs to create a hybrid, or the conjugation of antibody fragments onto a full-length mAb (Chiu and Gilliland, 2016; Kontermann and Brinkmann, 2015; Tiller and Tessier, 2015). These multi-functional molecules are finding promise in treating various cancers (Ayyar et al., 2016). A detailed description of the numerous methods employed to engineer these chimeric molecules is beyond the scope of this introduction. One method of note though is the 'knobs-into-holes' methodology. Here, mutations are introduced into the C_H3 domains of each heavy chain, yielding a 'knob' in one chain's C_H3 domain and a 'hole' in the other. Additional disulphide engineering makes the heterodimerisation step efficient (Merchant et al., 1998). The Ab fragments and bsAbs discussed above are summarised in Figure 1.5.

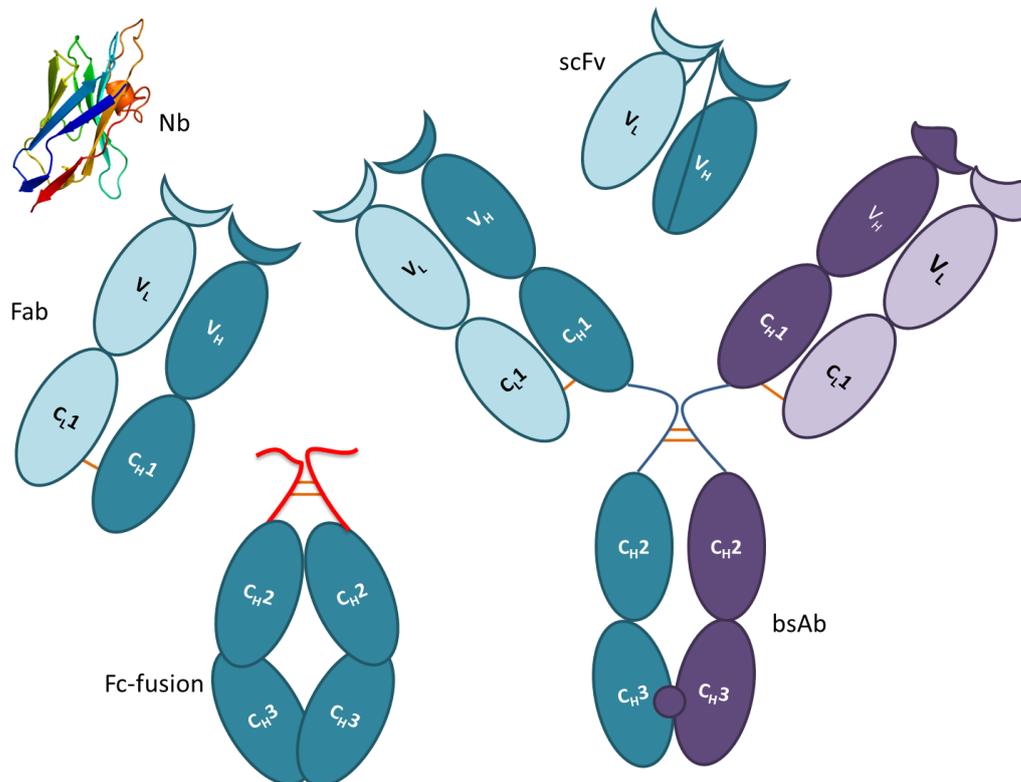


Figure 1.5. Antibody fragments and a ‘knobs-into-holes’ bispecific antibody (bsAb). The nanobody (Nb) is the smallest antibody-based fragment, consisting of just one Ig domain derived from camelids. Structure = PDB 1MEL (Desmyter et al., 1996). Single-chain variable fragments (scFvs) typically consist of a variable heavy and light chain tethered together with a flexible glycine-serine linker. Sometimes, Fabs possess enough desired functionality to be used in isolation. Functional peptides (red curves) or proteins can be conjugated to Fc domains as an Fc-fusion. Finally, bi-specific antibodies generated using the ‘knobs-into-holes’ method have two completely different halves, thus these molecules can bind to two different antigens. Adapted from Strohl, 2018.

Antibody-drug conjugates (ADCs) are another major family of antibody-based biopharmaceuticals. These molecules exploit the high specificity of the antibody to its target to deliver a drug ‘payload’ to the required site (Ducry, 2012; Kennedy et al., 2017; Tiller and Tessier, 2015). Often, the conjugated drugs are highly potent cytotoxic agents. In isolation, off-target interactions with the drug would lead to side-effects being observed in patients, which are greatly reduced when administered as an ADC (de Goeij and Lambert, 2016). Three ADCs have been approved to date, with many more in development (Kennedy et al., 2017; Mullard, 2018). ADCs have three key components: the cytotoxic drug, a conjugating linker and the antibody itself, as summarised in Figure 1.6.

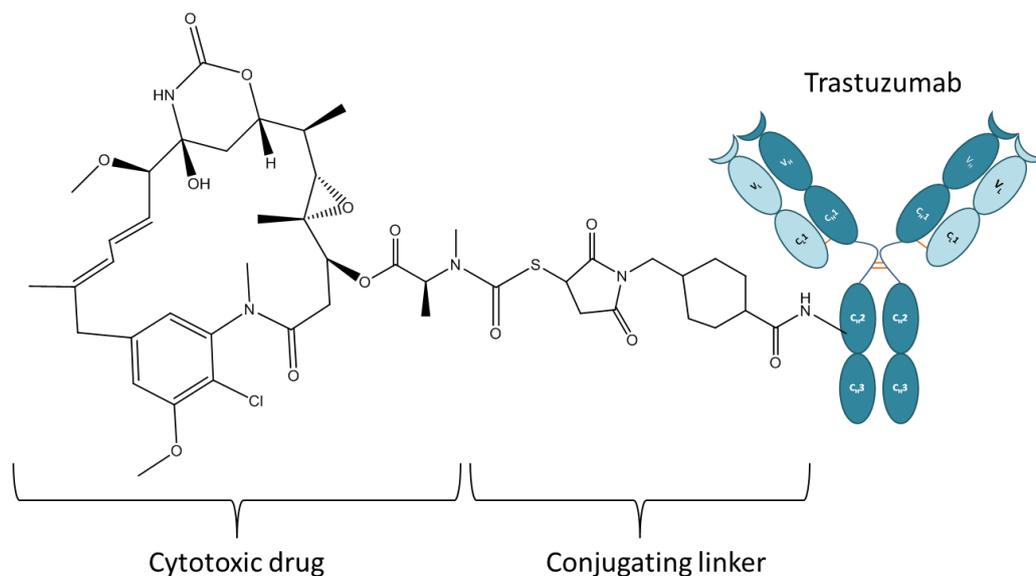


Figure 1.6. Schematic of the antibody-drug conjugate (ADC) Kadcyła® (Ado-Trastuzumab emtansine). Trastuzumab (Herceptin®) targets HER2 positive breast cancer cells. The conjugation of the microtubule-inhibiting drug emtansine via a stable maleimide thioester linkage. Multiple copies of the drug conjugate to the mAb. The mAb and drug are not drawn to scale. Structure adapted from Elgundi, 2016.

The large number of lysine residues adorning the surface of antibodies allows one to conjugate drugs to these sites, as has been done for Kadcyła® (Akkapeddi et al., 2016). However, optimising the conjugation conditions to obtain the correct drug:antibody ratio (DAR) is a key consideration in the development of ADCs (Ducry, 2012). Cysteine residues are often used as well, either by reducing the intact Ab (Doronina et al., 2003) or by introducing them through engineering (Junutula et al., 2008). There is however a risk of forming disulphide-linked aggregates during the conjugation step, or scrambling the disulphides in the Ab (Chudasama et al., 2016) More site-specific conjugation strategies are being developed to obviate these issues (Akkapeddi et al., 2016). The chemistry of the linker is critical, as chemically labile linkers would be cleaved prematurely in the blood stream. This could lead to premature release of the drug, causing toxic side-effects (Kennedy et al., 2017). Some linkers are designed to be cleaved *in vivo* following targeting to the lysosome (Erickson et al., 2006). Kadcyła®, for example, features a non-cleavable linker, which increases the stability of the molecule (Chen et al., 2016).

In summary, the affinity of antibody-based molecules to their targets, as well as the wide number of functionalities that can be incorporated into them accounts for their dominance in the biopharmaceutical market.

1.2.3 Biochemical engineering: Selection, expression, purification and formulation of therapeutic antibodies

In order for a candidate therapeutic protein to become a finished, marketed product, various steps must be traversed. This can be summarised by a drug development pipeline (DiMasi et al., 2016; Walsh, 2003), as shown in Figure 1.7. In the early stages of the pipeline, drug targets are identified and lead compounds are discovered (Walsh, 2003). Following optimisation of these compounds, pre-clinical studies are performed on animal models, before licenses are sought to begin clinical trials in humans (Plenge, 2016; Walsh, 2003).

Clinical trials are generally broken down into four stages (Piantadosi, 2017; Walsh, 2003): Phase I (20–100 patients), where doses are adjusted to find the safest levels; Phase II (100–300 patients), where the safety, activity and efficacy at a set dose is assessed; Phase III (1000–3000 patients), where the new treatment is compared to alternatives or placebos to really assess the impact the medicine will have compared to those already on the market and Phase IV (Post-marketing), where patients taking the approved medicine are monitored. Phase IV can open up new therapeutic avenues to extend the patent lifetime of the drug (Piantadosi, 2017).



Figure 1.7. An overview of the drug discovery pipeline. Basic research identifies the culprits of disease, which are to be targeted by drugs. Lead compounds are identified which interact with the target in a desirable fashion. After optimisation, toxicity studies are performed in animal models, before the material is tested in humans. During clinical trials, dosing regimens, toxicity and efficacy are monitored in carefully chosen patient groups. Trials often start with small groups (tens of patients) before growing to involve hundreds to thousands of patients. If efficacious and safe, a licensing application is made to a regulatory body, such as the US Food and Drug Administration (FDA) or the European Medicines Agency (EMA). The whole process can take 10–15 years and cost billions of dollars. Adapted from (DiMasi et al., 2016; Walsh, 2003)

As this thesis focusses on therapeutic antibodies, the discovery, manufacture and delivery of this class of proteins will be discussed in the remainder of this section.

Antibody discovery

In vivo, B-cells produce antibodies against antigens detected by the humoral immune system. However, billions of different B-cells are necessary to produce the many different antibodies (estimated at 10^{10} – 10^{12}) found in humans (Elgundi et al., 2017; Williamson, 2012). These polyclonal antibodies would not generally be desirable as drug products. The revolutionary work of Köhler and Milstein in 1975 developed monoclonal antibodies using mouse hybridoma technology (Köhler and Milstein, 1975). This technique is summarised in Figure 1.8. Briefly, it involves the fusion of mutant mouse myeloma cells with spleen cells from a mouse immunised against an antigen X. Selection for ‘hybridoma’ cells leads to proliferation of cells producing the antibody against the antigen X (Brekke and Sandlie, 2003). This swiftly led to the approval of the first monoclonal antibody therapy, Orthoclone OKT3 in 1986, a murine mAb which prevents kidney organ transplant rejection (Rodgers and Chou, 2016). However, the murine lineage of the mAbs derived in this fashion meant immunogenic effects were sometimes observed in patients (Kuypers and Vanrenterghem, 2004). Consequently, gene technologies were developed to ‘humanise’ mAbs by adding mouse variable

domains into a human mAb scaffold to make a chimeric mAb (Morrison et al., 1984) or humanise all of the scaffold except the CDRs by grafting in the mouse CDRs (Jones et al., 1986). Alternatively, genetically engineered mice can be used to produce fully human IgGs (Lonberg et al., 1994).

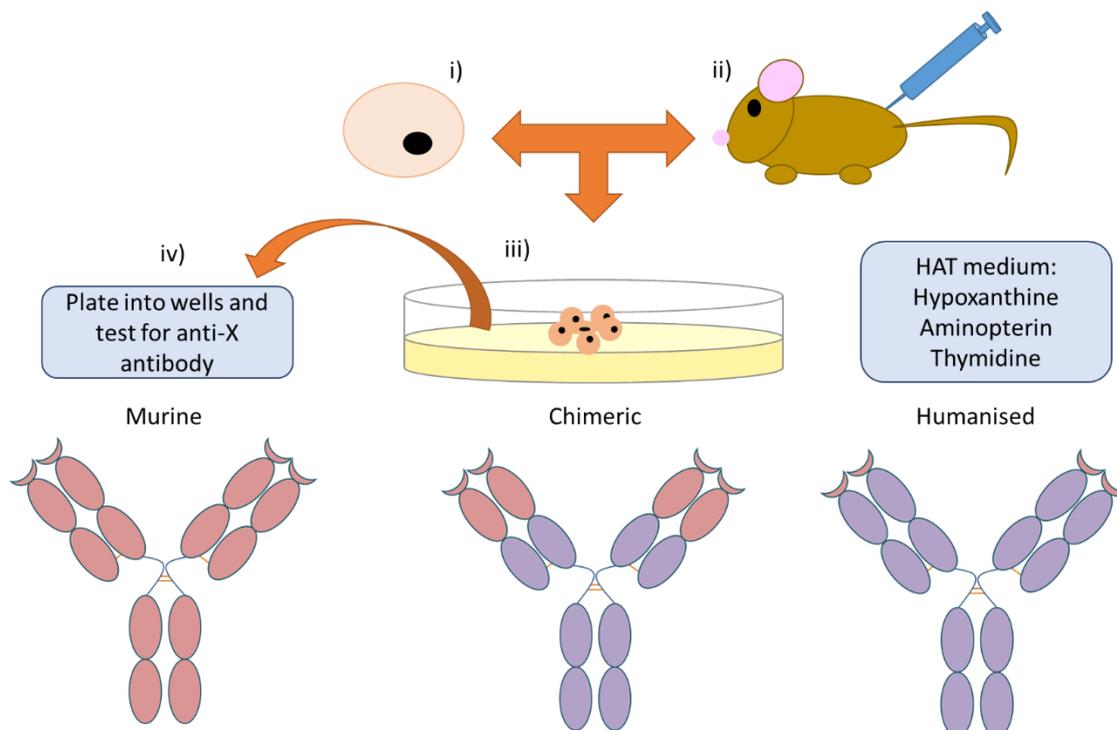


Figure 1.8. Overview of mouse hybridoma technology. *i) A murine myeloma cell line is generated. ii) These cells are fused with mouse spleen lymphocytes (B-cells) following immunisation against a desired antigen (target). iii) The myeloma cells lack the ability to grow on HAT medium, thus only those which successfully fuse with the murine spleen cells will survive to form a hybridoma. iv) B-cells usually do not survive for long when cultured in vivo unlike the hybridomas, which readily produce antibodies against the desired antigen. Screening the clones to identify those producing the murine mAb allows one to grow these cells at scale to make the product. Murine mAbs (pink) can cause immunogenic effects in humans, hence engineering strategies have been developed to make chimeric or humanised mAbs (purple = human sequence). Adapted from Clark, 2006; Rodgers and Chiu, 2016; Walsh, 2003; Voet and Voet, 2011.*

The other key breakthrough that enabled the development of fully human mAbs was phage display technology (Rodgers and Chou, 2016; Winter et al., 1994). Briefly, this involves the expression of human V_H and V_L genes on bacteriophage

coat proteins, e.g. pIII (Smith, 1985). Such fusions do not affect the infectivity of the phage, which readily displays the protein of interest on its surface. Importantly, this technology avoids the use of immunisation or hybridoma technology, reducing the ethical concerns associated with antibody technologies (Marks et al., 1991). The great 'success story' to emerge from phage display was the approval of adalimumab (Humira®) in 2002. This phage display-derived mAb is used in the treatment of chronic inflammation and had 2017 sales exceeding \$18bn, making it the most valuable drug in terms of sales in the world (Urquhart, 2018). How these filamentous phages are used to select for paratopes with high affinity for their epitopes is summarised in Figure 1.9.

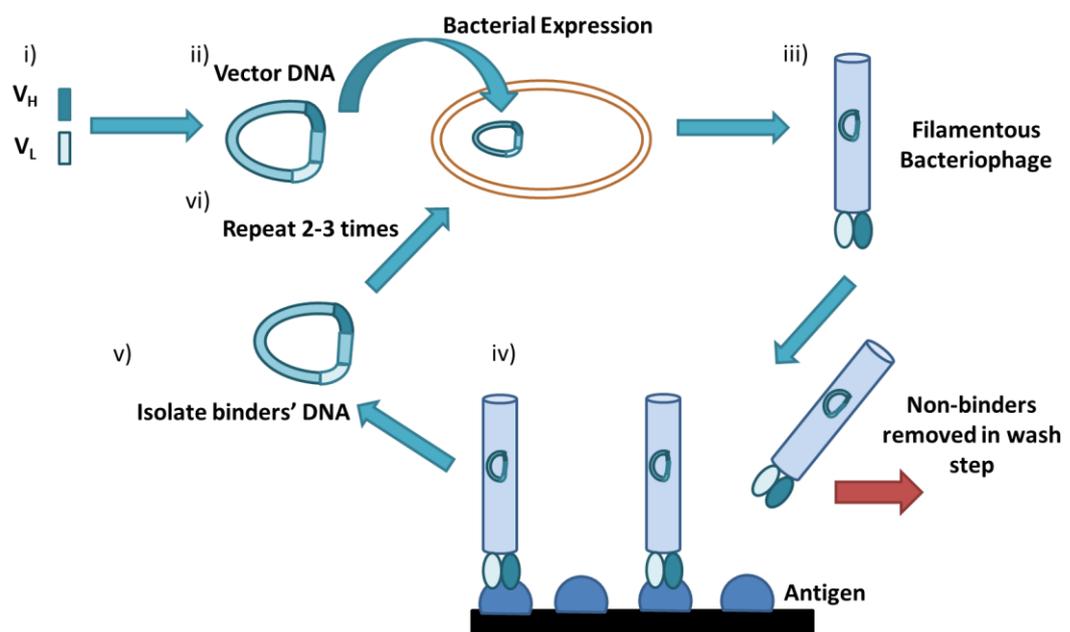


Figure 1.9. Overview of phage display technology. *i) Desired variable heavy and light chain sequences are selected. ii) These sequences are cloned into plasmids, transformed into bacteria and expressed. iii) Infection with filamentous phage results in expression of the scFv of interest on the phage's coat protein. iv) Immobilisation of the antigen of interest and panning with the phages allows one to rapidly screen for binding, eliminating non-binders. v) E. coli are re-infected to allow the binding phage DNA to be isolated and identified. Repeating the cycle increases the affinity for the target. Adapted from Brekke and Sandlie, 2003; Clark, 2006.*

Another development to increase the throughput and identification of potent human mAbs was yeast display technology (Feldhaus et al., 2003). Briefly, this involves the generation of a scFv library, which is incorporated into yeast, which express the scFv on their surface. One can then use flow cytometry-based

methods to sort and enrich binders from non-binders (Chao et al., 2006). Yeast display offers a further advantage in that *Saccharomyces cerevisiae* is a eukaryote. This means human scFv fragments will go through the same ER-trafficking pathways as occurs in the body, thus the folding and processing of these proteins is highly suited to this system (Elgundi et al., 2017). This technology is summarised in Figure 1.10 and Appendix Figure 1.

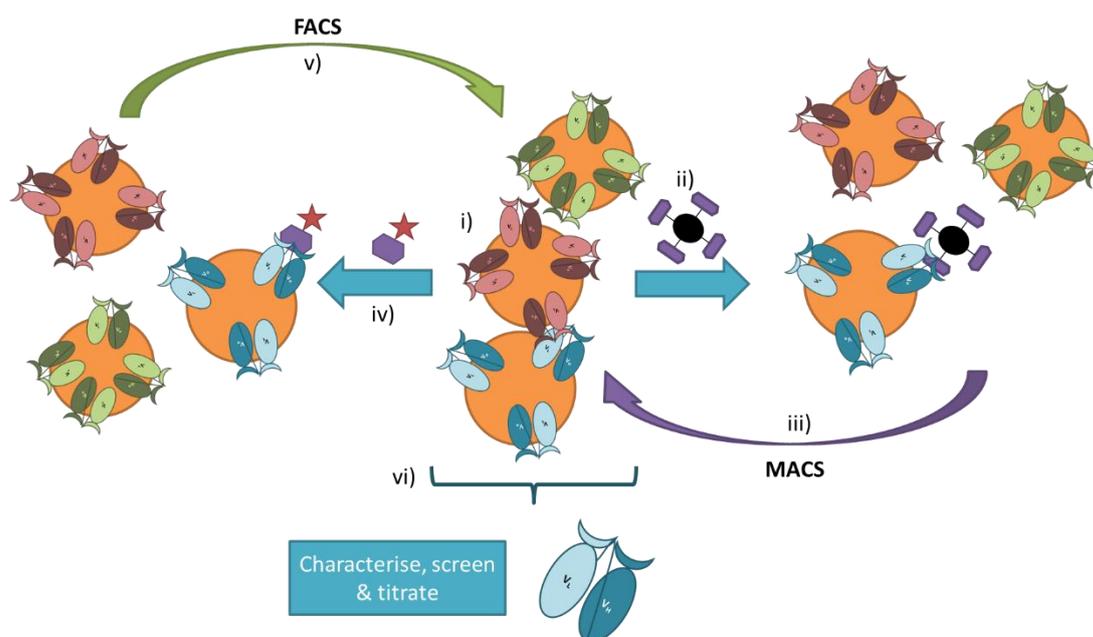


Figure 1.10. Overview of yeast display technology. i) The scFv sequence of interest (from a library) is expressed on the surface of yeast as a fusion protein (Appendix Figure 1). ii) The antigen of interest is conjugated to magnetic beads. iii) Incubation with the yeast allows one to pull out and expand binding cells using magnetism-assisted cell sorting (MACS). iv) The antigen of interest is conjugated to a fluorophore. v) Cells which bind the antigen are sorted and expanded using fluorescence-assisted cell sorting (FACS). This can be repeated with a mutagenized library to increase the affinity for the antigen. vi) Individual clones are sequenced, with the affinity of the scFv: antigen complex determined.

Adapted from Boder et al, 2012; Chao et al, 2006.

Finally, *in vitro* ribosome display has also emerged as a novel antibody discovery strategy (Hanes et al., 2000; Zahnd et al., 2007). This technique can screen very large (10^{13} clones $>$) libraries of scFvs in an entirely *in vitro* fashion, negating the need to use microbes or animals (Elgundi et al., 2017). This technique is summarised in Figure 1.11.

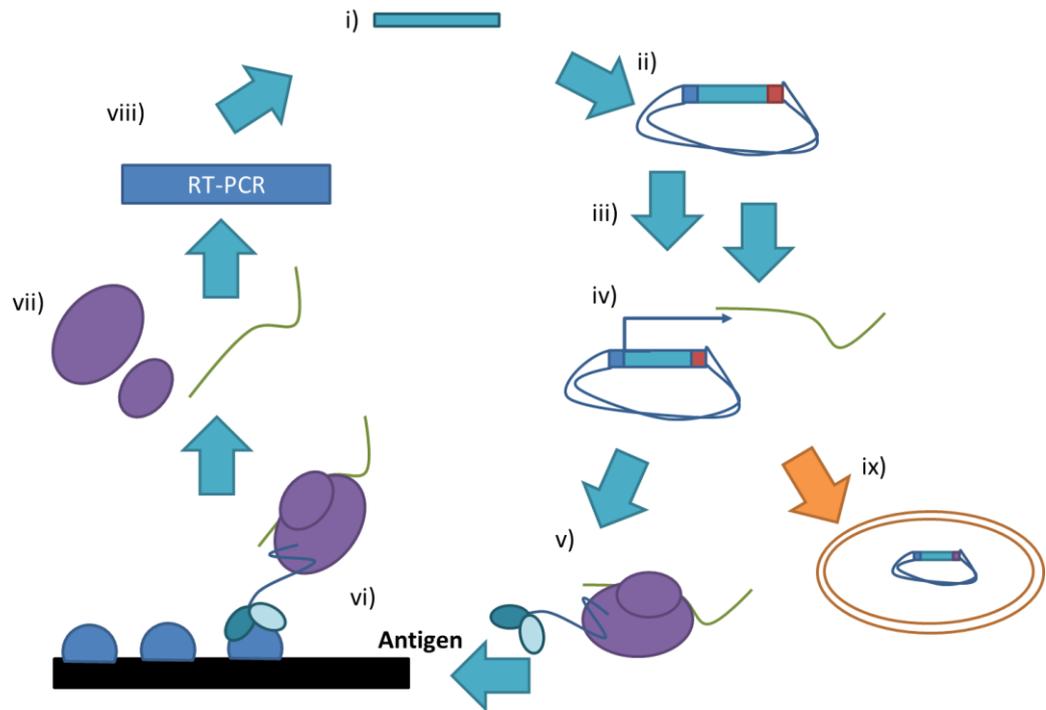


Figure 1.11. Overview of ribosome display. *i and ii) The scFv library is prepared and cloned into the ribosome display plasmid. iii) The Polymerase Chain Reaction (PCR) amplifies the amount of plasmid DNA. iv) The insert is transcribed in vitro to yield the mRNA (green line). v) The mRNA is translated in vitro, but stalls on the ribosome due to a lack of stop codon. vi) The lack of stop codon also means the mRNA-ribosome-protein complex remains intact. This complex is screened against immobilised antigen as in phage display (Figure 1.8). vii) Binding complexes are dissociated, then reverse transcription PCR performed on the mRNA. viii) This generates successful binding sequences which can be modified to increase target affinity in another round of display. ix) Desirable sequences can then be clones into *E. coli* vectors for expression and characterisation. Adapted from Hanes et al. 2000; Zahnd et al, 2007.*

Once candidate sequences are identified that bind to the target using any of the technologies above, one or more rounds of affinity maturation are then performed. This essentially exploits the error-prone PCR used to make the highly diverse sequence library to try and increase the binding affinity of the scFvs, selected by e.g. phage/ribosome display, for their targets (Chiu and Gilliland, 2016; Groves and Nickson, 2012). Typically, the long-loop V_H CDR3 is mutated, as this often has the biggest impact on a mAb's affinity for its antigen, in addition to limiting the number of mutated sequences one needs to generate (Yang et al., 1995). Once potent, human or humanised mAbs have been identified, the drug development process can really begin, allowing companies to focus on producing ever larger

quantities of the molecule of interest as the pipeline progresses. This will be discussed in the following sections.

Mammalian cells for mAb expression

Being multi-domain proteins with numerous post-translational modifications (such as glycosylation), full-length monoclonal antibodies are generally produced in mammalian expression systems (Elgundi et al., 2017; Spadiut et al., 2014). Of all of the cell systems available, the Human Embryonic Kidney 293 (HEK-293) cell line is typically used to transiently express mAbs at the research scale (Russell et al., 1977; Vink et al., 2014). Conversely, the Chinese Hamster Ovary (CHO) cell has become the 'workhorse' of the biopharmaceutical industry in terms of being a production cell line (Birch and Racher, 2006; Jayapal et al., 2007; Walsh, 2003). The plasmids used to code for the protein of interest, together with their selectable markers and promoter sequences, is usually transfected into the cells using electroporation or using cationic lipids to facilitate DNA entry into the target cells (Li et al., 2010). Ideally the cell line of interest will: be transfected efficiently; be easy to screen for transfection (e.g. using methotrexate if the cell line is usually deficient in dihydrofolate reductase (DHFR), with the gene for the latter enzyme present in the plasmid bearing the protein of interest) and produce high titres of quality recombinant protein (Birch and Racher, 2006).

Once the cell line has been selected, it must be made into a cell bank, which is usually thawed, expanded in small-scale cell culture and the inoculum expanded before the cells enter a large-scale bioreactor (Fieder et al., 2005; Shukla and Thömmes, 2010).

Bioprocessing: From bioreactor to purified mAb

Monoclonal antibody production is described as a 'platform process'- largely all products of this type are produced in a very similar fashion (the 'platform'). A typical mAb production process is shown in Figure 1.12, with each of the individual unit operations discussed below.

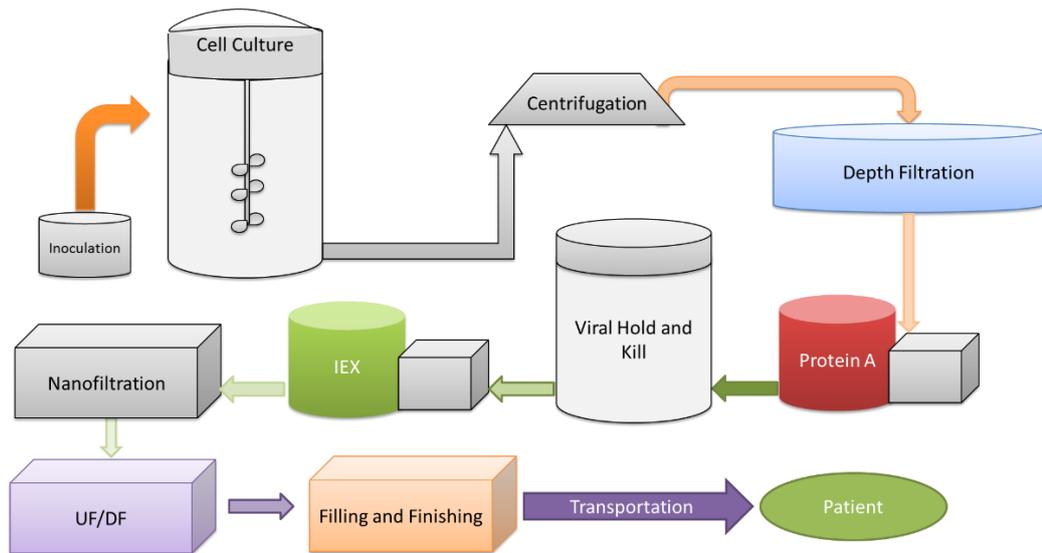


Figure 1.12. Schematic of a typical mAb manufacturing process. Many inoculum steps are necessary to have the desired cell density to begin large scale fermentation and over-expression of the product of interest. Cells are usually separated from product during centrifugation and depth filtration (silver and light orange arrows). The clarified feed stock is then loaded onto a Protein A column to capture the mAb product. Elution of the product at low pH can be beneficial for virus inactivation (dark green arrows). Polishing steps using ion-exchange chromatography (IEX) allow any residual impurities to be removed prior to nanofiltration/virus filtration (light green arrows). Ultra-/dia-filtration concentrates the product and buffer-exchanges it into the desired formulation buffer. This may be frozen for use at a later stage, or the product put into vials (light purple arrow). The product can then be transported (dark purple arrow) to hospitals/pharmacies ahead of administration to the patient. Adapted from Shukla and Thömmes, 2010.

Following the inoculum steps, cells are grown in a production-scale bioreactor. For some mAb products, this can take place at 20,000 L scale (Birch and Racher, 2006). The bioreactor design is critical to the survival of the cells (Nienow, 2006). Probes are often built in to monitor pH, temperature and dissolved oxygen (DO) content. Various feed lines can maintain these parameters should they change during production (Nienow, 2006). The turbines/impellers used to keep the cells in suspension and the medium mixed are often of a ‘marine’ design to minimise potential damage to mammalian cells (see Section 1.4.2) (Varley and Birch, 1999). These systems grow cells in disposable bags, with advantages being

reduced infrastructure costs and cleaning times compared with stainless steel bioreactors (Rogge et al., 2015). In modern production, components of the cell medium are usually chemically defined, as the traceability and uniformity of the medium exceeds that of serum-derived media (Birch and Racher, 2006). Furthermore, there are fewer ethical implications for the product as compared to when animal-derived media is used (Li et al., 2010). An emerging trend in upstream processing (USP) is the use of single-use bioreactors for smaller scale (e.g. 2000 L) cultures (Challener, 2017). A typical cell culture at scale takes ~15 days, with the batch often being 'fed' during the stationary phase to increase the titre of product (Li et al., 2010). Modern production cell lines can produce several g L⁻¹ of mAb product (Azevedo et al., 2009).

Following these USP steps, the product must be separated away from any contaminants and cell components during downstream processing (DSP). DSP is often thought to be the bottleneck of bioprocessing, especially given the ever larger titres of mAb produced from USP (Capela et al., 2017; Kelley, 2009). The initial clarification steps are usually referred to as primary recovery steps (Shukla and Thömmes, 2010). Typically, centrifugation is used to remove the cells away from the product. The centrifuge is typically of a disk-stack design, which minimises the drying of the product, as well as the subjection of cells and product to overly harsh flow conditions (see Section 1.4.2) (Hutchinson et al., 2006). A depth filter of reasonable porosity (> 0.2 µm) is used to provide yet more clarification of the feed stream (van Reis and Zydney, 2007). Care must be taken during processing to avoid the formation of a so-called 'filter cake'; fouling and clogging of the membrane which slows the transition of the feed stream through the bioprocessing pipeline (Goldrick et al., 2017).

The defined structures of antibodies can be exploited to purify them from solution. Protein A chromatography has emerged as the key capture step in mAb purification (Gagnon, 2012). Protein A is a ~45 kDa virulence factor produced by *Staphylococcus aureus* which predominantly binds to the Fcs of IgGs *in vivo* (Palmqvist et al., 2002). This protein is usually produced recombinantly in high quantity, chemically bound to chromatographic resins and packed into a column. At high production scale, Protein A columns can be over a metre in diameter (Wang and Mann, 2009). The true price of Protein A resin is estimated to cost ~€10,000 L⁻¹ (Franzreb et al., 2014). The ubiquitous nature of Protein A chromatography, in addition to many technological improvements in Protein A resin, mean it is likely to continue to dominate mAb capture in the future (Dransart

et al., 2018). mAbs are usually eluted from Protein A resin by decreasing the pH. Whilst this may adversely impact product quality (see Section 1.2.4), this does aid the inactivation of any viruses present in the feed. A viral hold tank may therefore appear at this stage of the bioprocess (Mazzer et al., 2015; Shukla and Thömmes, 2010).

Following Protein A chromatography, polishing steps are necessary to remove any residual contaminants, such as DNA, viruses, medium components and host cell proteins (HCPs) from the process feed (Shukla and Thömmes, 2010). Typically, anion/cation exchange or hydrophobic interaction (HIC) chromatography are used for this purpose (Capela et al., 2017). Membrane chromatography has been developed for this purpose, as the faster flow-through compared to traditional column chromatography can help accelerate the purification of the mAb product (van Reis and Zydney, 2007). A virus filtration step (pore size of ~ 20 nm) will remove any viruses that have not been cleared by the previous unit operations (van Reis and Zydney, 2007; Shukla and Thömmes, 2010).

Once the product is purified, it can be buffer exchanged into its fill/finish or formulation buffer. This is usually done using an ultrafiltration/diafiltration (UF/DF) step, with the flow set-up in a tangential flow arrangement as opposed to dead-end (Figure 1.12). This step is crucial, as many mAb-based products have to be delivered into patients at high concentrations (50–150 mg mL⁻¹) (Baek and Zydney, 2018; Elgundi et al., 2017; Eschbach and Vermant, 2008). The product then undergoes filling and finishing. This could be into vials/ pre-filled syringes for clinical/home administration (Mitragotri et al., 2014; Wang, 1999) or lyophilisation (freeze-drying), often at a dedicated facility, to be reconstituted before administration to the patient (Vollrath et al., 2018). The transportation of the finished drug product can be very important, as will be discussed in Section 1.2.4

Formulation and delivery of therapeutic monoclonal antibodies

The components of a biopharmaceutical formulation are very important for both the colloidal and kinetic stability of the product, as well as its pharmacokinetics (Wang, 2015). The need for an aqueous environment to maintain the correct three-dimensional structure of a protein means that therapeutic mAbs are formulated as liquids, except where lyophilised. Oral delivery of a mAb would lead to proteolysis of the drug in the stomach, hence mAb-based products are

generally delivered intravenously or through sub-cutaneous injection into the patient (Wang et al., 2008). Intravenous delivery requires patient hospitalisation, which increases the cost of treatment and inconveniences the patient (Snelling, 2008). Biopharmaceutical companies have thus sought to move towards subcutaneous administration of mAbs in recent years (Eisenstein, 2011; Wang, 2015). Whilst this means that the product can be administered by the patients themselves (or with the assistance of a district healthcare worker) in their own home (Snelling, 2008), the product has to be kept stable in solution at 4°C for up to two years (Randolph and Carpenter, 2007). Furthermore, pre-filled syringe devices have had to be developed to facilitate injection of the drug into the patient (minimising both the length of time needed for delivery and pain) (Mitragotri et al., 2014).

This formulation challenge has been a big issue for the biopharmaceutical industry. However, common buffer components and excipients have been identified to allow mAb products to be administered in this fashion. These typically include amino acids such as histidine and arginine, polyols such as trehalose and mannose and surfactants such as polysorbate (Tween®) 20 and 80 (Daugherty and Mrsny, 2006; Wang, 1999; Zbacnik et al., 2017).

In summary, the manufacture of therapeutic antibodies, from identification of mAbs which bind tightly to the desired target *in-vitro*, to the final formulated product is a complex process. Factors which adversely affect the molecule (efficacy and safety) form the subject of the next section.

1.2.4 'The aggregation problem' in the biopharmaceutical industry

Once administered into a patient, therapeutic antibodies will seek their target, typically an extracellular membrane protein or a secreted protein present in the blood (Rodgers and Chou, 2016). Any contaminants present in the product will also be injected into the patient. Whilst these may be benign, it is well documented that adverse reactions can be a consequence in patients, with severity ranging from intolerance to the medicine to anaphylaxis and death (Büttel et al., 2011; Moussa et al., 2016; Rosenberg, 2006; Wang et al., 2012). Gradual intolerance

and/or resistance of these expensive medicines, whilst not life-threatening, increases the gap between the cost of the drug and its benefit in patients. For example, Humira® patients can form anti-drug antibodies to adalimumab which neutralise the latter's therapeutic effect (Van Schouwenburg et al., 2014). The risk aggregates pose to patient safety and drug efficacy means their levels are monitored closely by the drug regulators such as the US FDA, e.g. US Pharmacopeia (788) for injectable drugs (Pharmacopeia). Typically, aggregate levels are kept below 5% of the total protein content, though no 'defined' level is set by the regulator (Mahler et al., 2009; van Reis and Zydney, 2007).

Aggregation can occur at any point in a biopharmaceutical's lifetime (Cromwell et al., 2006). Some antibodies can begin to show signs of aggregation when expressed at low level (1–10 mg) in HEK cells early on in development (see Figure 1.6) (Dobson et al., 2016). This could be due to inherently aggregation-prone sequences within the protein of interest (Meric et al., 2017) (see Section 1.3.1). It is of great industrial importance to identify rapidly mAb candidates which display aggregation propensity early on in development, as this decreases the likelihood of encountering problems later down the development pipeline (Jain et al., 2017).

For a mAb to enter clinical trials, larger quantities of material are needed. This is when bioprocessing itself can induce aggregation in certain mAbs. Aggregation can take place during cell culture for a variety of reasons, including: over-expression of large quantities (5–10 g L⁻¹) of protein, which will stress the cells' protein quality control machinery (chaperones, chaperonins etc.) (Powers and Balch, 2013); changes in pH and temperature in the bioreactor which could induce aggregation in certain proteins (Jing et al., 2012), extractables and leachables from the plastics found in single-use bioreactors (BPSA Extractables and Leachables Subcommittee, 2007) and stirring in the bioreactor (see Section 1.4). Aggregation at this early stage decreases the amount of material that can be purified downstream.

Aggregation could also occur during downstream processing, due to: pH and ionic strength changes encountered during chromatography (Cromwell et al., 2006; Mazzer et al., 2015; Wang, 1999); interactions between the protein and stainless steel surfaces inducing aggregation (Bee et al., 2010); adsorption of hydrophobic mAbs and aggregates at air-water interfaces (Bee et al., 2012; Koepf et al., 2018) and adverse interactions/fouling of membranes (van Reis and Zydney, 2007).

Aggregates which form during DSP could render a drug 'unmanufacturable', halting its development (Kola and Landis, 2004).

At the formulation stage, many strategies are employed to mitigate aggregation of the product and to afford the drug a long shelf-life (see Section 1.2.3). Arginine is a commonly employed excipient in biopharmaceuticals, as this amino acid is thought to block hydrophobic interactions between mAbs with its aliphatic side chain and interact electrostatically with side-chains on the mAb *via* its charged termini and guanidinium R-group (Baynes et al., 2005; Kim et al., 2016). Surfactants such as Polysorbate 20 and 80 are often used to compete with therapeutic proteins for the potentially damaging air-water interfaces that are encountered throughout manufacture, storage and administration (Agarkhed et al., 2017; Bee et al., 2011; Koepf et al., 2018; Zheng et al., 2017). For example, these damaging interfaces can also be formed when the product undergoes freeze-thaw (Hawe et al., 2009) or upon reconstitution from a lyophilised powder (Mahler et al., 2009). These protective excipients can breakdown in the presence of UV light, with the resulting products causing damaging effects such as oxidation and cleavage of the mAb (Kerwin, 2008; Kim et al., 2016; Mahler et al., 2009; Tamizi and Jouyban, 2016). As well as oxidation, deamidation is a particularly challenging problem for formulated mAbs, potentially leading to reduced activity and increased levels of aggregation (Daugherty and Mrsny, 2006; Phillips et al., 2017; Tamizi and Jouyban, 2016).

The formulated mAb is usually filled into vials or pre-filled syringes. In this fill-finish operation, the harsh hydrodynamic stress imparted onto the drug product may influence its aggregation (Section 1.4) – this is particularly concerning as the biopharmaceutical does not undergo any further purification after this stage (Cromwell et al., 2006). The drug product must then be transported before being administered/dispensed to a patient. The storage conditions and hydrodynamic forces generated during transport may all influence the aggregation of the product. This is very difficult to control from outside the manufacturing environment (Fleischman et al., 2017).

One of the final hurdles to overcome for the development of a commercially successful antibody-based biopharmaceutical is the injection into patients at high concentration (>50 mg mL⁻¹). If delivered subcutaneously, the product is usually supplied as a pre-filled syringe (Eisenstein, 2011), with the barrel lubricated with silicone oil (Bee et al., 2011). There is a growing body of literature to suggest that

silicone oil may influence the aggregation, and subsequent immunogenicity of biopharmaceuticals, led by Carpenter, Randolph and colleagues (Chisholm et al., 2015; Gerhardt et al., 2015; Teska et al., 2016). The high concentration of the product may increase its viscosity, making injection into the patient slow and potentially painful (Mitragotri et al., 2014). Methods of improving and predicting the 'syringeability' of biopharmaceuticals in on-going, but complex (see Sections 1.3 and 1.4) (Baek and Zydney, 2018). All of these potential stresses, which could induce the aggregation of a mAb-based biopharmaceutical, are summarised in Figure 1.13.

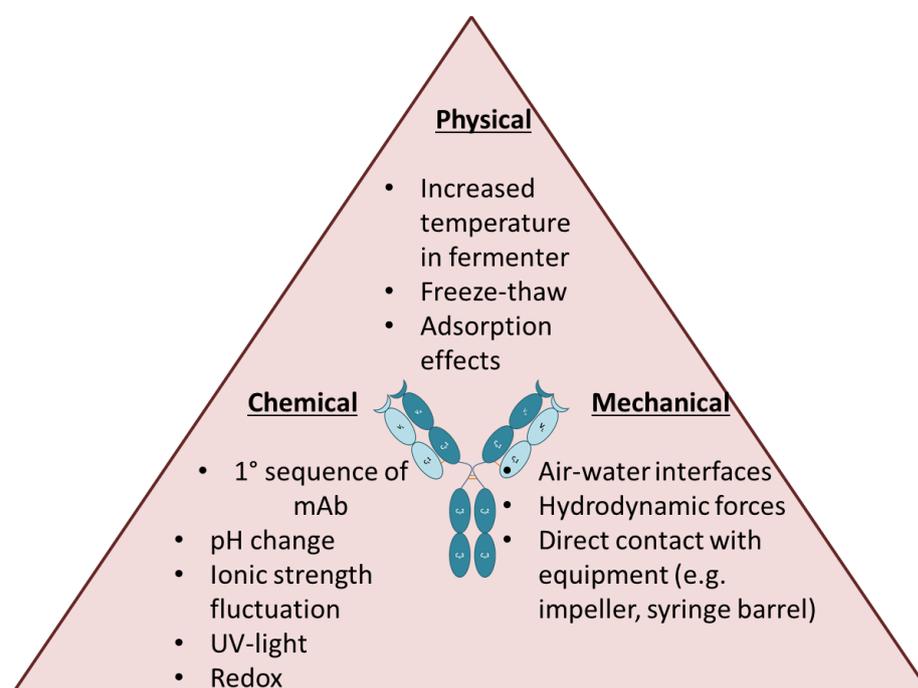


Figure 1.13. Potential inducers of mAb aggregation. The stresses mAbs encounter are physical, chemical and mechanical in nature. How mAbs respond to these stresses is sequence dependent.

With aggregation being a concern throughout a biologic's lifetime, regulatory bodies have sought to force biopharmaceutical companies to minimise such issues through the 'Quality by Design' (QbD) framework. This was implemented by the International Council for Harmonisation (ICH), comprising a group of the world's largest drug regulators in an attempt to standardise their expectations regarding drug safety. QbD can be defined as "a systematic approach to development that begins with predefined objectives and emphasises product and process understanding and process control, based on sound science and quality risk management" (ICH, 2009). This essentially means engineering 'quality' into

the process of making a drug, to achieve the desired Target Product Profile (TPP). The TPP includes criteria such as maximal levels of aggregate in the final drug product, glycosylation pattern, potency etc. Deviations from the TPP may hinder the 'developability' of a mAb. Developability can be broken down into 'manufacturability' (the ease of making the product in the manner discussed in Section 1.2.3), safety (the specificity for the target in a patient and a lack of side-effects) and activity (the pharmacodynamics and kinetics of the drug) (Zurdo et al., 2015). Aggregation can clearly impact the developability of a biopharmaceutical. Computational and experimental methods are thus employed to respectively predict and characterise aggregation. These methods form the subject of the next section.

1.2. Techniques used to study biopharmaceutical aggregation

1.3.1 *In silico* predictors of aggregation-prone regions

Hydrophobic amino acids are usually buried in the core of a protein when it folds (Williamson, 2012). Partial unfolding of a protein, instigated by the various stresses biopharmaceuticals encounter (see Figure 1.12) can expose these usually buried hydrophobic patches to the solvent. These regions can then self-associate, i.e. the protein can aggregate (Roberts, 2014). Regions within proteins that have been predicted and experimentally verified to self-associate are called aggregation-prone regions (APRs) (Fernandez-Escamilla et al., 2004; Meric et al., 2017). An APR can be defined as a region of a protein 5–15 amino acids in length, typically hydrophobic in character, that often forms aggregates rich in β -content (Ganesan et al., 2016; Khodaparast et al., 2018; Meric et al., 2017).

Various APR prediction algorithms exist, all with distinct caveats and assumptions. A comprehensive review of all of them is beyond the scope of this thesis. The algorithms can generally be divided into those which analyse just the primary sequence of the protein (termed 'intrinsic') and those which map any APRs onto a structure of the protein in question (termed 'structure-corrected').

Many APR predictors were developed using proteins/peptides which either form amyloid or β -rich aggregates (Meric et al., 2017). One of the best characterised is

the TANGO algorithm (Fernandez-Escamilla et al., 2004). This statistical mechanics algorithm uses the physicochemical properties of an amino acid stretch (five residues), its hydrophobicity, charge at a particular pH etc. and computationally determines what conformations this residue could occupy in an aggregate. Those with a β -sheet propensity 'TANGO' score above 5% were found experimentally to be more amyloidogenic, i.e. more aggregation-prone than those below the threshold (Fernandez-Escamilla et al., 2004).

TANGO has recently been paired with the FoldX force field (Schymkowitz et al., 2005) to create Solubis (van der Kant et al., 2017). This programme identifies the APR 'hot-spots' using TANGO, whilst identifying whether or not the regions are stabilising to the protein structure using FoldX (which calculates the free energy of a protein from its structure (Schymkowitz et al., 2005)). This script virtually mutates the amino acids in a TANGO zone to 'gatekeeper' residues (Arg, Glu, Asp, Asn and Pro (Rousseau et al., 2006)) to evaluate the contribution of the TANGO zone to the stability of the protein. Mutation of a residue buried in the hydrophobic core of a protein to a gatekeeper would be destabilising. Conversely, mutation of exposed hydrophobic residues is stabilising to the overall protein. A pre-requisite of this algorithm is the availability of a 3D structure or homology model.

Whilst TANGO and Solubis predict *generally* aggregation-prone regions, the Waltz algorithm (also developed by Serrano, Schymkowitz, Rousseau and colleagues) specifically predicts amyloidogenic regions within proteins. This is based on a bioinformatics and experimental analysis of hexapeptides linked to amyloid. Of particular note, the algorithm scores the position of an amino acid in terms of its likely contribution to form an amyloid fibril, as the conformation of such residues is crucial to the formation of the very regular amyloid fibril (Maurer-Stroh et al., 2010).

A similar algorithm is the prediction of amyloid structure aggregation (PASTA) (Trovato et al., 2007). As well as predicting the intrinsic APRs within an input sequence, it calculates the energy of forming a parallel or anti-parallel β -sheet. The cross- β core gives amyloid fibrils their high stability, thus understanding the likelihood of forming such species is desirable. A recent improvement to the algorithm (PASTA 2.0) incorporates a larger reference dataset, in addition to understanding the individual effect of certain mutations on the stability of the protein being analysed (Walsh et al., 2014).

AGGRESCAN is another web-based algorithm which detects APRs or ‘hot spots’ within protein sequences. This uses a dataset based on the aggregation of amyloidogenic proteins, to which sliding windows (5–11 residues) of the input sequence are compared. The a^4v represents the aggregation propensity of an amino acid based on its neighbouring residues, whilst the HST is the “hot spot” threshold- essentially the average aggregation propensity of the twenty canonical amino acids based on their frequency in proteins (de Groot et al., 2005). A hotspot is defined as “five or more sequentially continuous residues with an a^4v larger than the HST and none of them is a proline (aggregation breaker)” (Conchillo-Solé et al., 2007). Recently, a structure-corrected version of AGGRESCAN has been developed (AGGRESCAN-3D) which incorporates the specific structural environment of a protein’s amino acids, improving the algorithm’s ability to predict the aggregation of partially unfolded proteins (Zambrano et al., 2015).

The Spatial Aggregation Propensity (SAP) algorithm is one of very few which focusses on therapeutic, rather than amyloidogenic, proteins. This algorithm performs molecular dynamics (MD) simulations to compute the dynamically exposed solvent-accessible surface area of a mAb, then corrects this value to represent the time-averaged hydrophobicity of these patches on the surface of the protein. A homology model of the mAb can be input into SAP and mutations made *in silico*, to engineer mAbs with reduced hydrophobic patches and therefore reduced aggregation propensity (Chennamsetty et al., 2009). A later paper from the same group combined the SAP score with the *in silico* calculated net charge of the mAb in a given formulation to generate the ‘developability index’. mAbs predicted to be aggregation-prone tended to correlate with long-term shelf life data (Lauer et al., 2012).

The prediction of the aggregation propensity of proteins has not been the sole focus of computational biologists over the last decade. The CamSol algorithm can compute the intrinsic *solubility* of a protein from its primary sequence, as well as a corrected CamSol score that accounts for the protein’s structure (Sormanni et al., 2015). A CamSol score below -1 identifies ‘poorly-soluble’ regions, whereas a CamSol score exceeding +1 suggests the region is highly soluble. A sliding seven-residue window is used for the calculation. The algorithm, in a similar fashion to SAP, has been used to guide protein engineering, to make mutations which improve the solubility of therapeutic proteins (Sormanni et al., 2015; Sormanni et al., 2017). All of the algorithms discussed above are summarised in Table 1.2.

Table 1.2. Summary of the aggregation algorithms.

Algorithm	Amyloid-focussed?	Structural Information?
TANGO	Yes, but also β -rich aggregates.	No
Solubis	Same as TANGO	Yes
PASTA	Yes	Informed the algorithm design, but structure not fed into the programme
AGGREGSCAN	Yes	In 3D version
SAP	No	Yes- crucial to its function
CamSol	No- solubility is the focus	If desired, yes.

The ability to predict the of aggregation propensity and solubility of proteins, including those of therapeutic interest facilitates the screening and re-design of proteins with desirable biophysical properties (Deng et al., 2016). In isolation however, each *in silico* method in isolation cannot predict the ‘developability’/‘manufacturability’ of therapeutic proteins due to the diversity of aggregation mechanisms. Various *in vitro* methods, which probe different length scales and sample sizes, must be employed in order to gain a holistic understanding of a therapeutic protein’s propensity to aggregate. These methods will be discussed in the rest of this section.

1.3.2. Spectroscopic and spectrometric techniques to probe higher-order structure

Within the QbD paradigm, a detailed understanding of the product is expected by the regulator (see Section 1.2.4). The TPP of a biopharmaceutical often includes features such as the extent of glycosylation, post-translational modifications, and presence of correct disulphides etc. Spectroscopic techniques (Mahler et al., 2009) and more recently, mass spectrometry are employed by the biopharmaceutical industry to assess the conformational stability and integrity of biologics (Chen et al., 2011; Tian and Ruotolo, 2018). These techniques are used to assess the Higher-order structure (HOS) of biopharmaceuticals.

Spectroscopic methods

Spectroscopy can be defined as the interaction of electromagnetic radiation with matter (Pavia et al., 2010). The chemical environment of a biomolecule, from its local conformation to its solution conditions, affects how it will respond to electromagnetic radiation. The wavelength of the radiation crucially affects which parts of a protein are probed in a spectroscopy experiment (Williamson, 2012).

Fourier Transform Infra-Red (FTIR) spectroscopy works on the basis of protein amide backbones stretching asymmetrically, which causes changes in dipole moments. Amide bonds, and by extension proteins, can therefore absorb infra-red radiation (Pavia et al., 2010). FTIR probes the secondary structure of a protein, with the local conformation of each amide bond affecting the wavenumber (the reciprocal of wavelength, with units of cm^{-1}) absorbed (Yang et al., 2015; Zölls et al., 2012). The amide-I region between 1600–1700 wavenumbers is the main region of interest in an FTIR measurement. A broad band at $\sim 1630 \text{ cm}^{-1}$ is usually indicative of an intramolecular, antiparallel β -sheet, (e.g. in mAb solutions), whilst α -helices give a band at 1652 cm^{-1} (Campbell, 2012; Harn et al., 2007; Yang et al., 2015). The spectrum is usually processed by taking its second derivative (to separate overlapping peaks), as well as subtracting the water peak from the amide-I region at 1640 cm^{-1} which would ordinarily mask the protein's signal (Yang et al., 2015). The HOS of the biopharmaceutical sample can thus be determined, generating a 'spectral fingerprint'. Any deviation away from the native state spectrum could be indicative of aggregation. FTIR can be performed on both liquid samples and lyophilised (solid) samples, making it quite versatile (Yang et al., 2015; Zölls et al., 2012). However, these data are fairly low resolution in isolation (Jiskoot et al., 2012).

Circular dichroism (CD) spectroscopy is another commonly employed technique to probe the secondary structure of biological molecules, complementary to FTIR (Zölls et al., 2012). It works on the differential absorption of left- or right-handed circularly polarised light (CPL) by a chiral molecule (Zaccai et al., 2017). CPL is formed from the sum of two electromagnetic waves which are perpendicular and phase-shifted with respect to one another (Campbell, 2012). Proteins are chiral molecules, as they are made up of L- amino acids. A protein's secondary structure can be probed using far-UV radiation (190-250 nm), whilst tertiary structure can be probed in the near UV region, based on the position of aromatic residues in the protein of interest (Zaccai et al., 2017). α -helices give a characteristic double

minimum at 209 and 222 nm, whilst β -sheets have one maximum at 198 nm and one minimum at 218 nm (Kelly et al., 2005). Loss of ellipticity in these secondary structural regions could be indicative of unfavourable conformational changes in the biopharmaceutical. Whilst CD is a very sensitive technique, the need for low protein concentrations ($\sim 0.2 \text{ mg mL}^{-1}$) and UV-silent buffers (Kelly et al., 2005) could impede the use of CD for biopharmaceuticals in their final, formulated state (Harn et al., 2007; Sharma et al., 2014).

Raman spectroscopy can also be used to gain a spectral fingerprint of a protein sample (Zölls et al., 2012). It works on the basis of an incident photon interacting with a molecule and being scattered at a longer or shorter wavelength to that of the incident photon. Molecules are Raman-active if rotation or vibration, brought about by the incident photon, is accompanied by a change of polarisability. Photons which lose energy following collision undergo Stokes scattering, whilst those which gain energy undergo anti-Stokes scattering (Atkins and de Paula, 2010). Raman spectral bands are weaker than IR signals as only 1 in 10^7 photons undergo Stokes scattering. Water barely scatters in the Raman region, thus solvent type is not an issue (Campbell, 2012). Raman has been developed recently to distinguish silicone oil droplets from protein in biopharmaceutical formulations which are administered by pre-filled syringes, a much discussed issue in biopharmaceutical analysis (Lankers et al., 2008) (see Sections 1.2.3 and 1.2.4), in addition to monitoring protein unfolding under flow (Ashton et al., 2009; Ashton et al., 2010; Carpenter et al., 1999) (see also Section 1.4.4).

Both intrinsic and extrinsic fluorescence techniques are used to analyse the extent of aggregation in protein therapeutic products. Fluorescence involves the excitation of a molecule, followed by emission of a photon upon relaxation. This is always of a longer wavelength than that of the incident photon. Tryptophan (Trp) residues are the most prominent fluorophores in proteins (Chen and Barkley, 1998).

Joubert *et al.* monitored the intrinsic fluorescence of a mAb solution by exciting the Trp at 280 nm and monitoring emission at 285–450 nm (Joubert et al., 2011). Unfolding causes Trp residues to emit photons of longer wavelengths than when they are buried in the hydrophobic core of a protein, based on their exposure to a more polar environment. This can help determine whether or not a protein is folded (Mahler et al., 2009), in addition to determining if aggregation has occurred depending on the change in emission wavelength observed (Joubert et al., 2011).

Mis-folded or aggregated proteins tend to have exposed hydrophobic surfaces. Extrinsic fluorophores such as 8-Anilino-1-naphthalenesulfonic acid (ANS) (Joubert et al., 2011), SYPRO Orange (He et al., 2010) and Thioflavin T (ThT) (Chakroun et al., 2016) have all been used to monitor the unfolding aggregation of proteins. These dyes bind to hydrophobic regions of the protein, which increases their fluorescence signal based on their chemical environment. Fluorescence is very sensitive, so small (μM) amounts of sample are needed when extrinsic dyes are used in the analysis. The greater the extent of aggregation in the sample, the stronger the fluorescence emission (Sharma and Kalonia, 2010).

The formulation of a biopharmaceutical, e.g. the presence of surfactants, may alter the fluorescence characteristics of the chosen dye (Carpenter et al., 1999; Mahler et al., 2009; Wang, 1999). The fluorophore itself may not bind to partially-unfolded native species amongst the conformational ensemble (Sharma and Kalonia, 2010) , or prefer to bind to one conformer of aggregate, e.g. ThT's preference to bind to amyloid fibrils (Thompson et al., 2015).

Nuclear Magnetic Resonance (NMR) is increasingly being used as a means to interrogate aggregation in biopharmaceuticals (Kiss et al., 2018; Wishart, 2013). A comprehensive overview of NMR theory is beyond the scope of this thesis. Briefly, nuclei which possess spin (such as ^1H , ^{13}C , ^{15}N , ^{31}P in biomolecules exist in two degenerate spin states: spin up and spin down (Pavia et al., 2010). These two states have a small energy difference in the presence of a magnetic field (Williamson, 2012). Nuclei either align with the magnetic field (lower energy state) or oppose it (higher energy state). The net sum of all of nuclear spins is an energy difference called the magnetisation (Williamson, 2012). A radio frequency pulse is therefore used to excite the nuclei further, causing them to precess. The energy of the radio wave matches the magnetisation. Removing the radio pulse causes the nuclei to relax, which can be detected. The chemical environment of a spinning nucleus dictates the frequency at which it resonates, known as the chemical shift. The resolution of NMR can thus be at the atomic level (Williamson, 2012). In terms of biopharmaceutical analysis, NMR spectra can be difficult to interpret due to the large size of the protein samples and the naturally low abundance of ^{13}C and ^{15}N (Wishart, 2013). NMR is increasingly being used in the biopharmaceutical industry thanks to technological advancements in spectrometers decreasing sample concentrations (Berkowitz et al., 2012). NMR is especially being used to analyse the extent of glycosylation/ HOS of mAb-based

biopharmaceuticals (Berkowitz et al., 2012; Kiss et al., 2018; Poppe et al., 2015; Wishart, 2013). Recently, NMR methods have been developed to study protein aggregation based solely on the water relaxation signal; something which is generally regarded as undesirable in biological NMR (Feng et al., 2015). The spectroscopies discussed in this section are summarised in Table 1.3.

Table 1.3. Summary of the spectroscopic methods mentioned above.

Technique	Advantage	Disadvantage
FTIR	Solid or liquid samples	H ₂ O can mask signal
CD	Sensitive methods	Buffer composition is important
Raman	Very non-invasive to the sample	Low resolution data obtained, i.e. not atomic-level
Intrinsic Fluorescence	Sensitive method	Ensemble technique
Extrinsic Fluorescence	Wide array of different fluorophores available	Buffer environment can affect the performance of the dye
NMR	Powerful method, capable of atomic level resolution	Resolution limited with large proteins, e.g. mAbs

Mass spectrometry-based techniques

Mass spectrometry (MS) is used extensively in the analysis of biopharmaceuticals, as it requires very small quantities of protein. The technique involves the ionisation of the sample, with the ions then sorted/analysed and then detected. Mass spectrometers measure the mass to charge ratio (m/z) of ions (Campbell, 2012; Glish and Vachet, 2003). Amino acid modifications and incorrect/missing post-translational modifications can all be detected by subjecting a biopharmaceutical to analysis with MS or tandem mass spectrometry (for sequencing from which residue level information is gleaned) (Bobst et al., 2008; Phillips et al., 2017; Rosati et al., 2014). Electrospray ionisation (ESI) was developed as an ionisation method for proteins in the 1980's (Fenn et al., 1990). For ESI, the biopharmaceutical must be transferred from the formulation buffer into a volatile buffer such as ammonium acetate (a volatile solution commonly used at pH 6.8) or ammonium bicarbonate, which could lead to the observation of artefacts in the gas-phase (Konermann, 2017). Native proteins in the gas-phase generally occupy fewer (and lower) charge states than when unfolded, thus this technique can assess whether a protein is unfolded or not (Tian and Ruotolo, 2018) (see Figure 1.14a).

An extension of this method is ion- mobility-mass spectrometry (IM-MS). Here, after ESI, the ions are extracted into the vacuum of the mass spectrometer. They are then allowed to drift through a region filled with an inert gas, which impedes their passage through the mass spectrometer. More expanded ions, with a larger surface area, collide more frequently with the gas atoms, thus arrive at the detector much later than their compact counterparts of the same mass and m/z values (Chen et al., 2011). While IM-MS is being used more in the analysis of mAbs (Ferguson and Gucinski-Ruth, 2016), its use is limited by expensive instrumentation and the tendency for mAbs to collapse in the gas-phase; i.e. have a collisional cross-section measured by IM-MS much smaller than that expected from orthogonal methods (Devine et al., 2017; Ferguson and Gucinski-Ruth, 2016; Pacholarz et al., 2014).

Chemical cross-linking is becoming an attractive option for mapping protein-protein interactions using mass spectrometry (Figure 1.14b). Various linker types and conjugating group chemistries are employed, but those that react with common nucleophiles in proteins (lysines and cysteines) are mainly used (Leitner, 2016). Protein complexes can be 'trapped' by the covalent cross-linker, digested and then sequenced using mass-spectrometry, with the additional advantage that the covalent modification between peptide and cross-linker is not labile, allowing one ample time to analyse the sample (Leitner, 2016). Once again, this has been used for mapping antibody self-interactions (Devine, 2016) However, cross-linking can generally be difficult to optimise and perform (Iacobucci and Sinz, 2017), meaning it is not yet routinely used for epitope mapping (Opuni et al., 2018).

Finally, various covalent labelling strategies can be employed to capture molecular snapshots of reactive and solvent-accessible residues using specific chemical labels (Borotto et al., 2015; Mendoza and Vachet, 2009) and hydroxyl radicals (Watson and Sharp, 2012; Zhang et al., 2015b). These 'footprinting' approaches (Figure 1.14c) can be powerful for the reasons discussed above, but can suffer from generating large amounts of data to analyse (Rinas et al., 2016) and, in the case of hydroxyl radical footprinting, the requirement for specialist equipment (Xu and Chance, 2007; Zhang et al., 2014; Zhang et al., 2015b).

Hydrogen-Deuterium Exchange (HDX) MS is becoming gradually more widespread to assess the HOS of mAbs (Tian and Ruotolo, 2018), having been used to measure the dynamics and folding of smaller proteins for several

decades, e.g. (Clarke and Itzhakit, 1998; Malhotra and Udgaonkar, 2016). Briefly, the method works on the fact that when incubated in a deuterated solvent, the protons along a protein's backbone will exchange with deuterons, depending on the solvent-accessibility of these residues, pH, temperature and the chemical nature of the residues involved (Clarke and Itzhakit, 1998). The deuterium incorporated into a protein increases the latter's mass. Following quench and digest (usually with Pepsin at ~ pH 2), the peptides which exchanged (i.e. incorporated D) can be mapped back onto the full-length protein (Wei et al., 2014) (Figure 1.14d). This has been used to map mAb association interfaces (Dobson et al., 2016) and epitopes (Li et al., 2017), as well as conformations of biologics (Nazari et al., 2016). However, the method has many condition-dependent steps which can affect its reproducibility, such as the fast back-exchange of D to H. These mean analyses need to be ultrafast, with temperatures kept as low as possible and pH maintained at pH 3 (hence pepsin is used as the protease. These issues can be better controlled by the development of commercially available automation (Wales et al., 2008).

In addition to detecting the modifications above, amino acid modifications and incorrect/missing post-translational modifications can all be detected by subjecting a biopharmaceutical to analysis with tandem mass spectrometry (MS/MS) (Bobst et al., 2008; Phillips et al., 2017; Rosati et al., 2014) (see Figure 1.14e). This typically involves the enzymatic digest of a protein of interest, followed by the on-line separation of peptides by liquid chromatography *en route* to analysis by mass spectrometry. This allows one to re-assemble the peptides to determine the overall sequence of the protein, so-called 'bottom-up' proteomics (Tian and Ruotolo, 2018). These mass spectrometry-based methods are summarised in Figure 1.14.

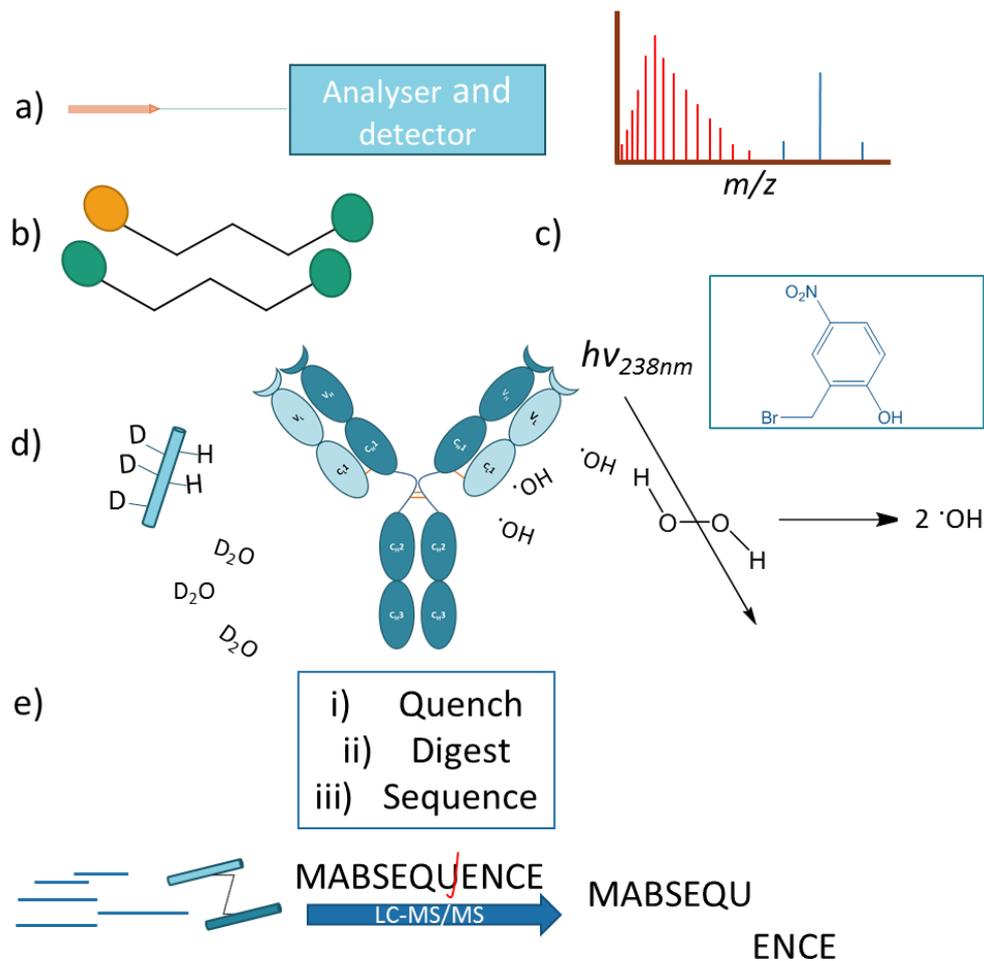


Figure 1.14. Mass spectrometry-based methods to study HOS in proteins including mAbs. a) ESI is a soft ionisation method, preserving the native structure of proteins in the gas-phase (blue peaks). Unfolded proteins have more protonation sites exposed, thus they occupy more charge states (red peaks). b) Chemical cross-linkers typically feature electrophilic groups with good leaving group ability, stapling neighbouring reactive proteins together. These can be homofunctional (green circles) or heterofunctional as desired. c) Hydroxyl radical footprinting can be performed using a KrF laser to fragment H_2O_2 into hydroxyl radicals which covalently label solvent-accessible residues. Specific labelling reagents, such as Koshland's reagent (for Trp residues, blue box) can also be used. d) HDX-MS is useful for mapping the solvent-accessible surface area of proteins. The degree of deuterium uptake can be quantified after analysing the protein via e) Proteolytic digestion, followed by LC/MS allows one to identify a protein from its peptide fragment. Residue-level information can be gained if tandem mass spectrometry (MS/MS) is employed, mapping modifications etc. as above. Adapted from Tian and Ruotolo, 2018.

1.2.3. Separation-based methods to probe biopharmaceutical aggregation.

As suggested in Figure 1.2, aggregates have different physicochemical characteristics to native, monomeric proteins. Size-exclusion chromatography (SEC) separates proteins on the basis of size (volume and molecular weight) and shape. The technique works by using a column packed with a porous gel matrix, e.g. Sepharose. The pores in the gel beads only allow proteins of a certain size (i.e. hydrodynamic radius) to permeate through them, whilst larger proteins/aggregates are excluded. These proteins will thus be eluted from the column much faster than the smaller, natively folded protein. UV absorbance at 280 nm is used to detect protein elution, based on tryptophan being the strongest absorbing chromophore in proteins. This is plotted with respect to elution volume to produce a chromatogram. The area under the curve of the chromatogram is proportional to the total mass of protein eluted. Protein standards of known molecular weight are used to calibrate the column; i.e. to relate the known protein's molecular weight to its elution volume. Proteins between 0.1–600 kDa can be effectively separated, depending on the choice of SEC column and conditions (Price and Nairn, 2009). Advantageously, SEC is relatively easy to perform thus it has been the standard separation-based technique for aggregate studies of biopharmaceuticals for many years, often coupled with High-Performance Liquid Chromatography (HPLC) to offer a short analysis time of ~15 minutes (Chaudhuri et al., 2014; Jain et al., 2017; Joubert et al., 2011; Mahler et al., 2009; Price and Nairn, 2009; Wang, 1999).

However, the method is not perfect, as sample-matrix interactions, they can occur, affecting the sample's elution (Dobson et al., 2016; Nicoud et al., 2014). These effects can be minimised by controlling the ionic strength and pH of the running buffer (Nicoud et al., 2014). Hydrophobic interactions can create false elution volumes and thus inaccurate determinations of molecular weight (Philo, 2006). An additional disadvantage of SEC is that the method assumes all proteins are spherical. Also, SEC is *not* suitable for the analysis of insoluble aggregates, as particulate matter would foul the column. Therefore, samples and buffers are typically passed through a 0.22 µm filter before use. Furthermore, soluble aggregates could dissociate reversibly in the running buffer, as the running buffer can greatly dilute samples. Consequently, the dispersity of the eluted protein may not accurately reflect the aggregated state of a biopharmaceutical formulation

prior to injection onto the column (Mahler et al., 2009). Therefore, orthogonal techniques are necessary to validate the results of SEC analysis.

Analytical Ultra Centrifugation (AUC) is one such orthogonal method that has been used to separate monomeric proteins from their oligomers/aggregates and to quantify these species in a variety of studies (e.g. (Scott et al., 2002; Yang et al., 2016). For aggregate analysis, AUC is usually operated in sedimentation velocity (SV) mode, as the increasing force (several thousand xg) applied to the molecules separates larger species from smaller ones. The sedimentation coefficient and diffusion coefficient (see Section 1.3.5) can be used to then calculate the molecular weight of the sample, from the work of Svedberg (Arakawa et al., 2006; Price and Nairn, 2009; Zaccari et al., 2017). Advantageously, SV-AUC can be performed in the original sample buffer, provided it is not too viscous. However, sample analysis can be difficult, the instrumentation is expensive and the technique is destructive to the sample (Arakawa et al., 2006).

In addition to centrifugation, asymmetrical flow field-flow fractionation (AF4) is being used as an orthogonal method to SEC (Bria et al., 2016; Mahler et al., 2009; Ripple and Dimitrova, 2012; Zölls et al., 2012). How the method works is shown in Figure 1.15. AF4 can separate and characterise particles of 1 nm–100 μm in size, without the need for a stationary phase (Zölls et al., 2012). One potential issue with AF4 is the tendency of protein aggregates to adhere to membranes, thus fouling it (Ripple and Dimitrova, 2012). As discussed in Section 1.2.3, this issue is observed throughout the manufacturing process of biopharmaceuticals (Goldrick et al., 2017). This could potentially then promote further aggregation in the instrument, as surfaces are known to influence the aggregation of biopharmaceuticals (see Section 1.2.4).

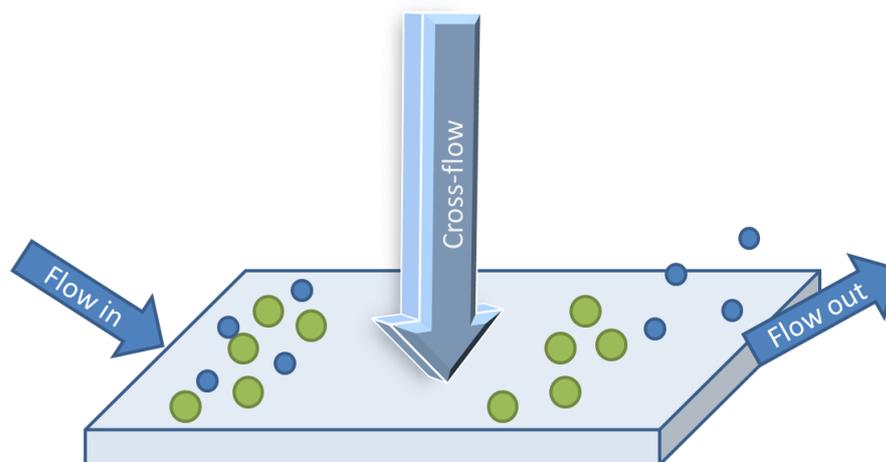


Figure 1.15. Schematic of AF4. The sample flows into the system, where it then encounters a cross-flow, forcing the solutes towards the membrane. Smaller species (e.g. monomeric proteins, blue circles) can diffuse away from the membrane faster than their larger counterparts (e.g. aggregates, green circles), thus they are detected earlier. Adapted from (Postnova, 2018).

A plethora of other, novel technologies have emerged over the last decade from academic and research laboratories to separate and size aggregates, including: adaptations of traditional electrophoretic methods (Halfmann and Lindquist, 2008); microfluidic technologies to separate monomers from aggregates such as Taylor-Dispersion Analysis (TDA) (Hawe et al., 2011) and other emerging microfluidic methods (Arosio et al., 2016; Kopp and Arosio, 2018); as well as novel chromatographic methods such as cross-interaction chromatography (Jacobs et al., 2010). These assays are too numerous to discuss in detail but are mentioned as they are allowing scientists to quantify protein aggregation in ways that way not possible using traditional methodologies. These traditional methods are summarised in Table 1.4.

Table 1.4. Summary of the traditional separation-based methods for aggregate analysis.

Technique	Advantage	Disadvantage
SEC	The industry 'gold-standard' method, well-established	Column unable to resolve oligomers over broad mass range
AUC	Minimal sample preparation	Complex analysis and destructive
AF4	Can separate species over large size range	Potential for aggregate-membrane interactions

1.2.4. Light scattering and related methods to detect, size and count aggregates in solution

Sub-visible particles are the utmost concern to regulators and biopharmaceutical manufacturers (Wang et al., 2012). These are too large to be separated from monomeric proteins and oligomers by SEC and too small to be analysed through visual inspection. These particles thus fall in the range of 0.1 μm –100 μm (24). Light-scattering methods are principally used to determine the size of particles in the 1–100 nm size range but can be used to analyse species up to $\sim 5 \mu\text{m}$ in size. Static Light Scattering (SLS) is a commonly used technique in which laser light is elastically scattered by a sample. Elastic scattering occurs when the sample's electrons scatter the incident radiation in all directions at the same frequency, i.e. no energy is lost (Mahler et al., 2009; Wyatt, 1993; Zölls et al., 2012). This is also called Rayleigh Scattering (Campbell, 2012). The light scattered by a macromolecule can have different phases; thus, the scattered light waves can interfere constructively and destructively with one another. The intensity of this scattered light is proportional to the molar mass of the scatterer, as well as its shape (Wyatt, 1993). This technique is called Multi-Angle Laser Light Scattering (MALLS). A schematic of this process is shown in Figure 1.16.

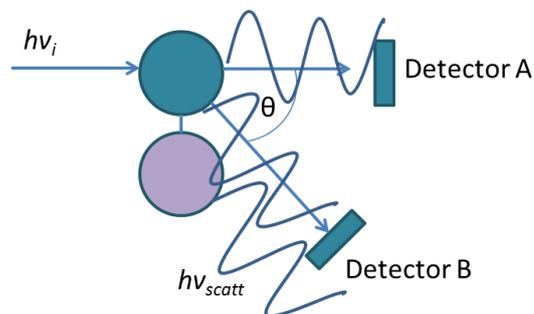


Figure 1.16. Schematic of Rayleigh scattering and detection. Incident photons, $h\nu_i$, are scattered by macromolecules. The blue and purple regions of the molecule (e.g. two subunits of a protein) can scatter the photon differently if the molecule is greater than $\sim 10 \text{ nm}$ in radius. Interference between these scattered photons then has an angular dependence, represented by θ . This in turn is related to the root-mean square radius of the macromolecule. Adapted from (Minton, 2016).

MALLS is frequently combined with SEC, in order to produce a rapid technique that can not only separate soluble aggregates, but determine their absolute molecular weight as well, based on the hydrodynamic radius of the protein

(Campbell, 2012; Wyatt, 1993). This has been done for model proteins (Oliva and Farin, 2001) as well as those of biopharmaceutical interest (Chaudhuri et al., 2014; Philo, 2006).

SEC-MALLS is a powerful technique, as any discrepancies in elution volume caused by aggregate-matrix interactions (SEC only) can usually be clarified with the MALLS detection (Philo, 2006). This is because the molecular weight determined by MALLS is not affected by elution volume, provided the eluted fraction's concentration and 'differential refractive index increment' or dn/dc is known (24, 37). For a non-glycosylated protein, this value is constant = 0.186 ml/g (Oliva and Farin, 2001). It is important to note however, that mAbs are usually heavily glycosylated (Section 1.2.2), so calibration may be necessary when SEC-MALLS is used in the analysis of these products. The latest developments in SEC-MALLS allow for the analysis of mAbs with very small sample volumes of 1–6 μL (Hsieh and Wyatt, 2017).

Dynamic Light Scattering (DLS) is the other common scattering technique. It is an ensemble technique which monitors the change in scattered light intensity over time due to Brownian motion of the particles (Arakawa et al., 2006; Campbell, 2012; Hassan et al., 2014; Minton, 2016; Zaccai et al., 2017). Buffer molecules collide with the solute molecules, causing the phases of their scattered photons to fluctuate. A particle's hydrodynamic radius can be determined from measurement of its diffusion coefficient, based on Stokes-Einstein equation (Hassan et al., 2014; Minton, 2016) (Equation 1.1).

$$D = \frac{k_B T}{6\pi\eta R_h}$$

Equation 1.1 Stokes-Einstein equation for determining the hydrodynamic radius (R_h) of a particle using DLS (SI unit = m). k_B = Boltzmann constant, T = absolute temperature (K), η = dynamic viscosity ($\text{kg m}^{-1}\text{s}^{-1}$), D = Diffusion coefficient ($\text{m}^2 \text{s}^{-1}$).

Protein particles in biopharmaceuticals between 1 nm and $\sim 5 \mu\text{m}$ in size can be detected and sized by DLS, provided the sample is not too concentrated (Fischer and Schmidt, 2016; Minton, 2016). Larger particles scatter much more intensely than smaller ones, i.e. Intensity, $I \propto r^6$ (Filipe et al., 2010; Minton, 2016; Philo, 2006). This means a large bias is placed on larger oligomers and particles than smaller ones. Filipe et al. established that DLS cannot differentiate between two

sets of particles in a polydisperse system unless they differ in size by at least three fold (Filipe et al., 2010).

The size ratio between aggregates in a biopharmaceutical formulation may be smaller than 1:3, meaning potentially immunogenic aggregates may not be detected. The technique should therefore be used *semi-quantitatively* in conjunction with other techniques; a widely-held view within the biopharmaceutical community (Fischer and Schmidt, 2016; Mahler et al., 2009).

A related method to DLS is Fluorescence Correlation Spectroscopy (FCS). Rather than measuring the fluctuations of elastic scattering with time, FCS measures fluctuations in fluorescence emission intensity with time (Haustein and Schwille, 2007; Zaccai et al., 2017). The sample of interest is typically fluorescently labelled, thus the increased sensitivity of the dye allows one to detect single-molecules at a time (Zaccai et al., 2017). The confocal volume of the apparatus used to detect the molecules is of the order of 1 fL (Haustein and Schwille, 2007). It has been used to detect oligomers in amyloid aggregation mechanisms (Guan et al., 2015; Tipping et al., 2015), as well as now being developed to track antibody aggregation in academic research groups (Imamura et al., 2017). The specialised equipment and the need for fluorescently labelled protein has probably limited the utilisation of FCS routinely in industry despite the advantages mentioned above.

Many new techniques have been developed in recent years to aid the counting and characterisation of nanoparticles. Nanoparticle Tracking Analysis (NTA) utilises a camera to visualise and track the motion of particles which dynamically scatter laser light, within the confines of a defined field of view (Carr and Wright, 2015; Filipe et al., 2010). The lower limit of resolution for the technique for proteins/soft matter is around 50 nm (Zhou et al., 2016). Although this appears disadvantageous in comparison to DLS, as NTA cannot size and detect monomeric proteins, NTA has the ability to distinguish between particles in polydisperse samples with greater sensitivity. DLS can only distinguish between particles that have a three-fold size difference (Filipe et al., 2010). NTA can distinguish between particles (including therapeutic protein aggregates) which have a ~0.5 to 2-fold size difference (Filipe et al., 2010; Vasudev et al., 2015).

Furthermore, the instrument can determine the concentration of each pool of particles, based on the defined amount of sample illuminated by its laser (Zhou et al., 2016). NTA tracks particles individually, whereas DLS is an ensemble technique; hence the latter may not fully represent the aggregated nature of the

sample (Filipe et al., 2010; Vasudev et al., 2015). A schematic of the apparatus is shown in Figure 1.17.

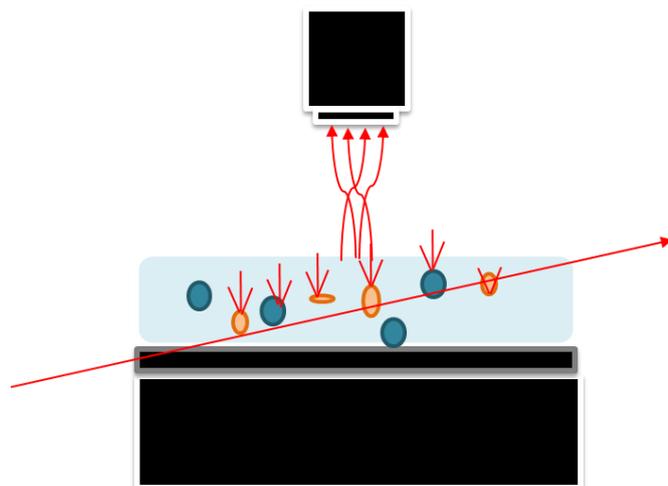


Figure 1.17. Schematic of the NTA technology. Incident laser light (642 nm) is passed through a glass prism inside the sample chamber. The angle of incidence and the functionalization of the surface (black rectangle, grey outline) together create an intense, thin laser beam. This means the light scattered from the particles can be seen at 20x magnification using a microscope (top). Coupling a charge-coupled device camera to the microscope allows one to capture videos of the moving particles. The videos are then processed by the Nanosight® software and the sizes and concentrations of particles in the sample are generated. Figure adapted from Carr and Wright, 2015.

Traditional methods for the identification and quantification of sub-visible and visible particles in biopharmaceutical samples include light obscuration (Hawe et al., 2009), turbidity (Wolfrum et al., 2017) and visual inspection (Mahler et al., 2009). Limitations of these methods include a lack of resolution between different scattering species using turbidity (Barnard et al., 2011) and human error (visual inspection (Melchore, 2011)). However, newer technologies are rapidly being developed to size and count particles in this space, including Micro Flow Imaging (MFI) (Hamrang et al., 2015; Zölls et al., 2013) and Resonant Mass Measurement (RMM) (Christie et al., 2014; Zheng et al., 2017; Zölls et al., 2013). This expansion of the analytical repertoire is providing researchers with many new ways to characterise particles and aggregates in mAb products that weren't possible ten years ago. Many of the methods mentioned discussed in Section 1.3 have been used to study flow-induced unfolding and aggregation of proteins, which is discussed in the next section.

1.4 Fluid flows and their effects on molecular and cellular structure

Proteins are exposed to a variety of different stresses in their lifetime, as discussed in Sections 1.1–1.3. However, one of the most under-studied areas is how hydrodynamic forces influence the unfolding and aggregation of proteins. This remains a very controversial topic in the field. This section of the introduction will lead into the aims and objectives of this PhD thesis.

1.4.1 Defined fluid fields: Shear flow vs extensional flow

Fluid mechanics is the field devoted to the study of the physical properties of fluids over space and time. In order to understand how flows influence the molecules being transported within them, it is crucial to define their underlying fluid mechanics. Of the various types of fluid flow, two of the most common are shear flows and extensional (elongational) flows (Bekard et al., 2011). Shear flows exist when layers of fluid (lamellae) travel over one another at differing velocities in the direction of the flow. This generates a velocity gradient perpendicular to the direction of the flow (Figure 1.18a) (Thomas and Geer, 2011). The word ‘shear’ has often been misused as an umbrella term to describe any hydrodynamic force (Thomas and Geer, 2011). Conversely, an extensional flow is generated where fluid is forced to accelerate, with a linear velocity gradient generated in the direction of flow (Figure 1.18b) (Thomas and Geer, 2011).

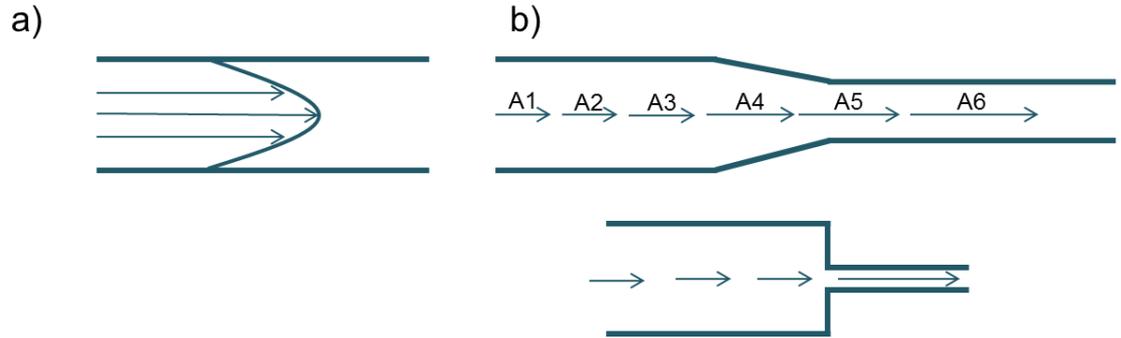


Figure 1.18. Schematic examples of shear flow (a) and extensional flow (b).
 a) Friction at the pipe wall leads to the pressure-driven flow adopting a parabolic profile, with the fluid moving fastest in the centre and slowest at the wall. The shear rate is thus maximal at the wall. b) A contraction in the pipe forces the fluid to accelerate to overcome the contraction. Each arrow represents a velocity vector, with length proportional to the velocity. An abrupt contraction (lower figure) leads to a rapid increase in the fluid velocity in the direction of the flow.
 Adapted from Bekard et al. 2011; Thomas and Geer, 2011.

Key equations for shear and extensional flows

Shear flows are usually described in terms of the shear rate, γ , which describes the velocity gradient perpendicular to the direction of flow. This gradient arises from thin lamellae of fluid travelling over one another at different velocities, separated by infinitesimally small distances (Equation 1.2).

$$\gamma = \frac{dv}{dy}$$

Equation 1.2. Equation for shear rate, γ (s^{-1}). v = velocity of each layer in the fluid ($m s^{-1}$) and y = distance between the layers (m).

The shear rate is maximal at the pipe wall (Figure 1.18a). Furthermore, the force a shear flow imparts on the pipe wall (expressed per unit area) is described as the shear stress, τ (Equation 1.3).

$$\tau = \eta \frac{dv}{dy}$$

$$= \eta\gamma$$

Equation 1.3. Equation for shear stress at a pipe wall, τ ($N m^{-2}$). η is the dynamic viscosity of the fluid ($N s m^{-2}$) and γ is the shear rate (s^{-1}).

Extensional flows are described in terms of the strain rate, $\dot{\epsilon}$ (Equation 1.4).

$$\dot{\epsilon} = \frac{d\epsilon}{dt}$$

where $\epsilon = \frac{l_f - l_0}{l_0}$

Equation 1.4. Equation for strain rate $\dot{\epsilon}$ (s^{-1}), which is the change in strain ϵ , with time, t . Strain can be described as the relative change in the length (in m) of a fluid element, or object within the flow from its original length, l_0 , to a final length, l_f . As strain is dimensionless, the strain rate has units of s^{-1} .

It is often important to discuss the flow in terms of being laminar (the fluid moving in defined layers as discussed above, like a gentle stream) or turbulent (merged streamlines and vortexes, like white-water rapids). The flow regime can be established using the Reynolds number, Re (Equation 1.5).

$$Re = \frac{\rho v L}{\eta}$$

Equation 1.5. Equation for Reynolds number, where ρ = density of the fluid ($kg m^{-3}$), v = velocity of the fluid ($m s^{-1}$), L = characteristic length (m), η = dynamic viscosity of the fluid ($N s m^{-2}$ or $kg m s^{-1}$). In a pipe flow, L can be substituted with the diameter of the pipe (in m).

The Reynolds number essentially describes the ratio of inertial to viscous forces in the fluid. A Reynold's number below 2000 is indicative of a laminar flow, whilst one exceeding 4000 is indicative of a turbulent flow (King, 2002), with an ill-defined transition region between 2000–4000.

Devices to generate shear and extensional flows

As well as the simple pipe designs shown above, researchers have developed many ways to generate shear and extensional flows. Viscometers are often used to generate these flows, especially those with a Couette flow design (Figure 1.19a). These devices can subject samples to defined shear rates for prolonged periods of time (Bekard et al., 2011; Fardin et al., 2014). The four-roll mill design has been employed to subject proteins to both shear and extensional flows, depending on the direction each of its rollers are rotated (Figure 1.19b) (Simon et al., 2011) –see Section 1.4.4. Another common way to generate extensional flow is to use a so-called cross-slot device, where fluids are made to collide with one another, generating an extensional flow (Figure 1.19c) (Perkins et al., 1997; Renner and Doyle, 2015).

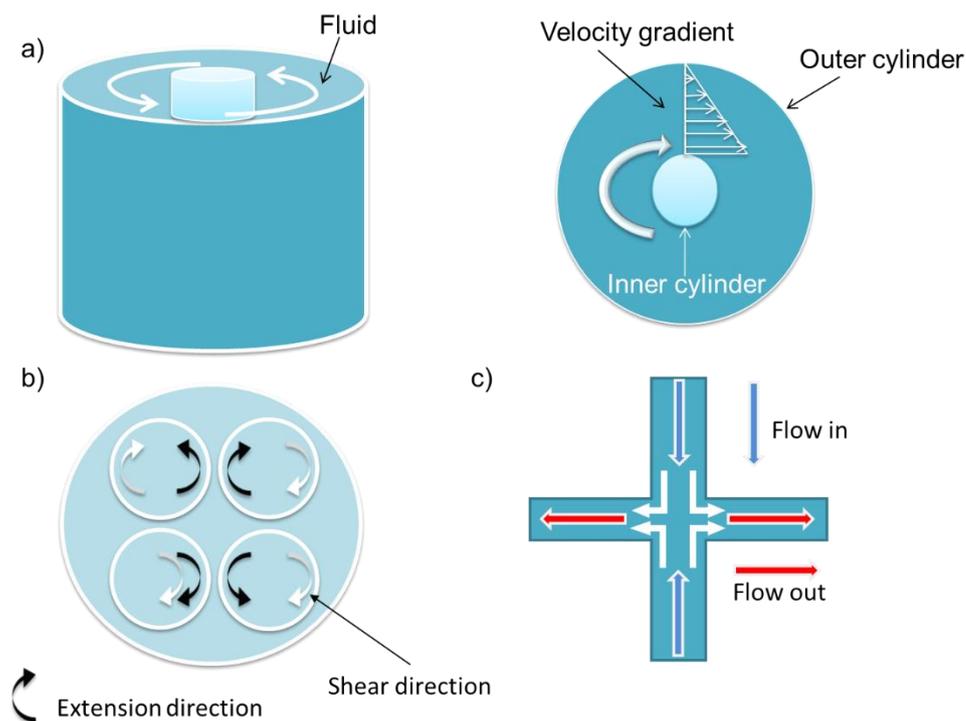


Figure 1.19. Schematics of commonly employed flow devices from the literature. a) A Couette flow device. The sample is placed in the device and the inner cylinder is rotated at a defined rate, inducing shear in the fluid. The plan view (right-hand image) shows the velocity gradient in the flow device, with the flow being slowest at the stationary, outer cylinder wall. b) Plan view of a four-roll mill apparatus. Protein solution is input into the device (similar to a, but with four Teflon rollers inside) and sealed. The rollers can be rotated independently in the directions indicated to generate extensional (black arrows) or shear flow (white arrows). c) A cross-slot device to generate extensional flow. Fluid is input into the device from opposing directions (blue arrows). The streams collide, generating extensional flow (white arrows). The fluid then exits the channels (red arrows). Adapted from (Bae et al., 2016; Bekard et al., 2011; Simon et al., 2011).

The exact design of flow device used depends on the sample volume, whether or not optics and other *in situ* measurements are being performed, as well as whether or not the device is custom-made or commercially available. How these different devices, and the flow they generate, affect molecules and cells will be discussed next.

1.4.2 Effects of flow and hydrodynamic forces on cells

Many researchers have investigated how flow affects mammalian cells for both medical (e.g. Jain et al., 2016) and biotechnology purposes (e.g. Sinha et al., 2015). It is thought that shear flows in the body affect the positions of red blood cells in the body's vasculature (Carboni et al., 2018), as well as their morphology (Lanotte et al., 2016). Blood can be described as a non-Newtonian fluid, as its viscosity changes when it is subjected to shear stress (Davies, 2009). Shear stress in the body can vary, from 1×10^{-3} Pa (N m^{-2}) in regions of the body with poor circulation to a few Pa (Abaci et al., 2015).

It is thought that shear and extensional flows can deform red blood cells, particularly in cases of atheroma where blockages generate extensional flows (Balaguru et al., 2016). Various microfluidic devices have been built to investigate these haematological phenomena, with geometries akin to those shown in Figure 1.18 (Carboni et al., 2016; Carboni et al., 2018; Fitzgibbon et al., 2015; Lanotte et al., 2016). The effects on the cells can have knock-on effects on proteins found in the blood (see Section 1.4.4).

From a biotechnological perspective, it was originally thought that mammalian cells are susceptible to the effects of hydrodynamic forces during large scale manufacture (Chisti, 2006; Nienow, 2006). It was thought that this damage could occur due to: impellers physically breaking cells open during fermentation; the high fluid flows generated during mixing (Varley and Birch, 1999); harsh shear flows cells encountered during DSP (Jain et al., 2005). Whilst it has been shown that extensional flows can mechanically deform cells (Bae et al., 2016), Hoare and colleagues subjected mammalian cells to shear flow and did not observe any damage (Hutchinson et al., 2006; Joseph et al., 2016; Rayat et al., 2016). It is instead thought that the bursting of cavities and bubbles during fermentation could be the main cause of damage, which has been alluded to previously (van Reis and Zydney, 2007). The collapse of air cavities in fluids is thought to generate strong extensional flows, which would have the potential to damage cell membranes in bioreactors if not controlled (Hariadi et al., 2015).

In summary, it is thought that cells can be deformed by shear and extensional flows depending on their physiological environment, especially in the case of certain diseases. Whilst mammalian cells are thought to be fairly robust to fluid

flows, they are indeed susceptible to cavity collapse. How *molecules* are affected by flows are the subject of the following sections.

1.4.3 Effects of flow on polymers and DNA

PG de Gennes carried out ground-breaking work on how flows affect non-biological polymers in solution in the 1970's (De Gennes, 1974). This laid out the quantitative framework behind how polymers can undergo coil-stretch transitions, influenced by shear and extensional flows. This work partly led de Gennes being awarded the 1991 Nobel Prize in Physics (NobelPrize.org and AB, 2018). Cathey and Fuller built upon this work, where they made clear that the properties of a solution, in particular its viscosity, have a major impact upon the ease by which a polymer undergoes chain elongation in a shear flow (Cathey and Fuller, 1990). The relationship between extensional flow and temperature on polymer breakage was also examined (Odell et al., 1990). *In-situ* measurements on flexible polymer distortion under shear flow was another technical advancement (Dunstan et al., 2004).

In terms of linear biological polymers, DNA has been used by many groups as a model system. The fact DNA can intercalate fluorescent molecules means it has been exploited for single-molecule studies of its dynamics under extensional (Perkins et al., 1997) and shear flows (LeDuc et al., 1999; Smith et al., 1999). These high-impact studies demonstrated the stretching and tumbling of DNA molecules, sometimes forming dumbbell structures in the flow. This occurs at very low strain and shear rates ($<5 \text{ s}^{-1}$) (LeDuc et al., 1999; Perkins et al., 1997; Smith et al., 1999).

Since plasmid DNA has to be produced at scale for use in biotechnology (Kong et al., 2010), how flow can damage DNA has been another area of study. Hoare and colleagues have investigated how shear and extensional flows can damage (unfold and fragment) supercoiled plasmid DNA, at strain rates ranging from $1.7 \times 10^4 \text{ s}^{-1}$ (Zhang et al., 2007) to $\sim 3 \times 10^5 \text{ s}^{-1}$ (Meacle et al., 2006) and shear rates of $\sim 3.5 \times 10^5 \text{ s}^{-1}$ (Levy et al., 1999). These high strain and shear rates were generated using capillary devices with abrupt contractions (see Figure 1.18b). The authors found that extensional flow seemed to be the dominant factor in causing the observed uncoiling of the DNA, especially after subjecting the samples to multiple flow events (Meacle et al., 2006; Zhang et al., 2007).

The fundamental and applied studies discussed suggest that flow-induced unfolding and aggregation of proteins is feasible. Studies that demonstrate this are described in the final section of this introduction.

1.4.4 Effects of flow on proteins

As mentioned in Section 1.2.4, hydrodynamic forces could be a contributing factor to the damage of biopharmaceutical proteins. More generally, researchers have been interested in understanding if and how shear and extensional flows affect proteins more generally for over 50 years (Bekard et al., 2011; Charm and Wong, 1970a; Thomas and Geer, 2011). More recently, proteins of physiological importance have been the subjects of flow studies. These will be discussed first.

Flow studies on elongated proteins

von Willebrand Factor (vWF) is one of the best studied proteins under flow, described by Timothy Springer as the “Jedi knight of the bloodstream” (Springer, 2016). This several MDa glycoprotein is 10s to 100s nm in length and “feels the force” of its hydrodynamic environment, allowing both the regulation of its length (Lippok et al., 2016; Zhang et al., 2009) and the unfolding of its platelet binding site when the force is sufficient (Aponte-Santamaría et al., 2015; Sing and Alexander-Katz, 2010). This latter event triggers the blood coagulation process. These studies have all shown that this large, elongated protein, is unfolded at relatively low strain rates ($1 \times 10^3 \text{ s}^{-1}$) (Fu et al., 2017; Sing and Alexander-Katz, 2010). Fibrinogen is another blood glycoprotein involved in coagulation. Shearing of this protein at physiologically relevant shear rates led to a loss of clotting activity over time (Charm and Wong, 1970b). The authors proposed that the shear environment *in vivo* could account for regulation of this protein.

Fluid flows have also been used to influence the aggregation of proteinaceous biomaterials. Dragline spider-silk has remarkable strength, making it a desirable material to make recombinantly (Egan et al., 2015). Rammensee et al. developed a microfluidic chip which mimics the spider’s silk gland, where shear flow aligns silk spidroin proteins, which then assemble into β -sheet rich silk under extensional flow (Rammensee et al., 2008). Often, very high concentrations of protein are

used in these devices ($>500 \text{ mg mL}^{-1}$) (Andersson et al., 2017) , which will affect the hydrodynamic forces applied to the proteins.

Flow studies on globular proteins

Charm and Wong subjected catalase, rennet and carboxypeptidase to shear in a viscometer at shear rates up to 1155 s^{-1} for up to 90 minutes. In the absence of an air-water interface, they found that the enzymatic activity of these proteins decreased by up to 50% (Charm and Wong, 1970a). The authors suggested this was due to the “breaking of tertiary structure when the enzyme is appropriately oriented in the shear field” (Charm and Wong, 1970a). Tirrell and Middleman built on these studies, subjecting urease to shear and monitoring its ability to break down urea during and post-shear. They found that this 484 kDa protein did lose activity after shear, but could partially recover, i.e. once the shear flow was removed, the enzyme regained activity (Tirrell and Middleman, 1975).

Conversely, Thomas and Dunhill subjected bovine liver catalase to shear rates varying from 636 to $1 \times 10^6 \text{ s}^{-1}$. They saw little to no loss of activity of the enzyme in this closed system, citing air-water interfaces as a possible major cause of catalase’s activity in other groups’ studies (Thomas and Dunnill, 1979). In 1997, recombinant human (rh) DNAse was shown to be unaffected by shear alone, whilst rh Growth Hormone formed aggregates under the same conditions (Maa and Hsu, 1997). Jaspe and Hagen performed a seminal study in 2006, where they exposed cytochrome c to very high shear rates ($2 \times 10^5 \text{ s}^{-1}$), yet saw no unfolding monitored by *in situ* Trp fluorescence. They proposed that a shear rate exceeding $1 \times 10^7 \text{ s}^{-1}$ would be necessary to unfold a 100 residue protein (Jaspe and Hagen, 2006). Two therapeutic proteins, recombinant human growth hormone and an IgG1 were stressed in a Couette shear device for up to 30 min at a shear rate = $3,840 \text{ s}^{-1}$. No unfolding of either protein was observed by *in situ* CD (Brückl et al., 2016).

In contrast to these studies, many groups have observed protein unfolding under shear flow. Ashton et al. performed *in situ* Raman spectroscopy on lysozyme (Ashton et al., 2009), followed by BSA, insulin, β -lactoglobulin and other proteins (Ashton et al., 2010). Deviation in the Raman spectra of these proteins, from those of the quiescent samples, increased as a function of protein size. Bekard and Dunstan also observed deviations in BSA and insulin structure using *in situ*

methods, such as CD and fluorescence, under shear (Bekard et al., 2012; Bekard and Dunstan, 2009).

Simon et al. subjected three globular proteins (lysozyme, bovine serum albumin (BSA) and alcohol dehydrogenase (ADH)) to stress under shear and extensional flows in a four-roll mill apparatus. They found that extensional flows were more damaging (i.e. brought about more aggregation) to these proteins than shear flows and that larger proteins (BSA and ADH) were more susceptible to flow effects than lysozyme, forming larger aggregates and in greater quantity (Simon et al., 2011). Using the same apparatus, the group have characterised the aggregation of monoclonal antibodies under flow (Wolfrum et al., 2017).

How flow affects mAb aggregation has obvious implications given the discussion of the effects of aggregation in Section 1.2.4. The literature is fairly divided on what the main factors driving hydrodynamic aggregation of mAbs are (Bekard et al., 2011). Bee et al. subjected a concentrated mAb solution (150 mg mL^{-1}) to stress under high shear ($\gamma = 2.5 \times 10^5 \text{ s}^{-1}$), yet saw no aggregation. Duerkop et al. very recently subjected mAbs and other globular proteins to stress under high shear ($\gamma = 1 \times 10^8 \text{ s}^{-1}$). They observed minimal aggregation for these proteins, except when cavitation was present (Duerkop et al., 2018). The Bracewell group have subjected a variety of mAbs (including IgG1s and IgG4s) to defined shear stresses and have observed protein aggregation (Biddlecombe et al., 2007; Biddlecombe et al., 2009; Tavakoli-Keshe et al., 2014). For an IgG4, it was found that the roughness of the surface in the shear device played the largest role in the aggregation of the mAb (Biddlecombe et al., 2009). More recently, this work has been applied to samples under high shear rates, where shear was thought to influence the aggregation (Nanda et al., 2017).

There is thus a clear divide between those who think flow can, a) perturb protein structure and b) induce/influence protein aggregation, and those who do not. The discrepancies between different studies can be put down to the following: different apparatus used to subject the proteins to stress under flow; the purity and origin of the proteins used in the study; topology and size (MW) of the proteins used; the protein concentration and buffer conditions used; the absence or presence of air-water and water-solid interfaces within the device; the time scale of exposure to the flow field in question and the method(s) of detection used to monitor 'damage' to the protein. Addressing these discrepancies form the basis of this PhD thesis, which is outlined in the next section.

1.5 Basis of the study

With these studies in mind, my colleague (now Dr) John Dobson (School of Mechanical Engineering, University of Leeds) designed and built a device which has the ability to subject proteins to extensional flow fields, followed by high shear rates. Extensional flows have generally been neglected, except where reference has been made to polymers or elongated proteins being more sensitive to extensional rather than shear flow, e.g. (Springer, 2016) and the work of Simon et al., 2011. Building a device able to apply known and modulatable extensional flow fields would allow one to validate this hypothesis for a range of model proteins.

The extensional flow device (EFD), referred to as the EFD or 'the device' herein, was designed and built with the following criteria in mind: production of a laminar, non-turbulent flow; generation of extensional flow by an abrupt contraction in the flow; a low volume to minimise the amount of protein needed; low flow rates to minimise the engineering challenge to achieve high strain rates whilst maintaining laminar flow.

Two syringes connected *via* a capillary fit the above criteria (see Materials and Methods); similar to the extensional flow setup shown in Figure 1.18b (lower panel). Molecules entering the contraction would be subjected to strain (i.e. extensional flow) at the point of contraction, followed by high shear rates along the length of the capillary.

BSA was initially selected as a model protein, as it has been shown by many groups to unfold and/or aggregate under both shear and extensional flow fields (Ashton et al., 2010; Bekard et al., 2012; Simon et al., 2011). A wide-array of biochemical (Peters, 1996), structural (Majorek et al., 2012) and aggregation data (Bhattacharya et al., 2011; Bhattacharya et al., 2014; Militello et al., 2003) are also available for this 573 residue, α -helical protein. Once the initial characterisation of the system was performed, these data would be implemented by investigating the behaviour of model mAbs obtained from industrial collaborators.

1.6 Aim and Objectives

The overall aim of this project is to understand how extensional and shear flows affect proteins and induce their aggregation. Specifically, the effects of flow on therapeutic proteins is the focus of this thesis. The objectives to achieve this are:

- To identify robust experimental conditions and analytical methods to monitor the aggregation of BSA, induced by extensional flow. Experimental techniques would be selected from the literature, including spectroscopic, scattering and mass spectrometric methodologies, then combined to understand the aggregation pathway of this model protein.
- To subject model monoclonal antibodies to stress in the EFD, comparing their observed aggregation behaviour to that of BSA and similar proteins in the literature.
- To modify the solvent and flow conditions in the EFD to explore a wide parameter space. Experimental design would be driven by both the fundamental fluid mechanics from above, as well as relevant literature from the biopharmaceutical industry.
- To use an array of different biophysical and biochemical techniques, to develop a predictive, general mechanism for how monoclonal antibodies aggregate under extensional flow.
- To understand how the EFD compares to orthogonal and orthologous assays for biopharmaceutical aggregation. Using a set of previously studied proteins, bioinformatics and statistical approaches will be used to compare different assays and methodologies to the EFD and the data obtained from it.

2. Materials and Methods

2.1. Materials

2.1.1 Equipment

Gel-filtration chromatography (purification) equipment

- ÄKTA Prime (GE Healthcare)
- Superdex 26/60 S200 column (GE Healthcare)
- Superdex 10/300 GL column (GE Healthcare)

Extensional flow equipment

- Breadboard base (Thor Labs)
- Syringe Clamps (Thor Labs)
- Ferrule compression fittings (Hamilton)
- Gilson P10 O-ring (Gilson)
- Borosilicate glass capillaries (Sutter Instruments)
- Ceramic capillary cutter (Sutter Instruments)
- Arduino Microcontroller (Arduino)
- Stepper motor (Haydon Switch and Waterbury CT Instrument Co.)
- Gas-tight 1 mL syringes 1001 RN model (Hamilton)

Analytical instruments and equipment

- miniDAWN TREOS + QELS module (Wyatt)
- Nanosight LM10 (Malvern Panalytical)
- JEM1400 Transmission Electron Microscope (JEOL)
- Eppendorf 5810 Benchtop Centrifuge (Eppendorf)
- Optima TLX Ultracentrifuge (Beckman Coulter)
- TLA100 rotor (Beckman Coulter)
- Ultrospec 2100 Pro UV/Visible spectrophotometer (Amersham Biosciences)
- UV-transparent cuvettes (STARSTEDT)
- Thickwall Polycarbonate ultracentrifuge tubes (Beckman Coulter)
- GenFuge 24D Benchtop centrifuge (Progen)
- Electrospray Ionisation capillaries (Sutter Instruments)
- Synapt High Definition Mass Spectrometer (Waters)
- Gel Documentation UV-transluminator (SynGene)
- QuantaMaster spectrofluorimeter (Photon Technology International)
- 1 cm path length fluorescence cuvette (Hellma Analytics)
- Synapt G2Si Mass Spectrometer (Waters)

- P680 HPLC pump (Dionex)
- ASI-100 Automated sample Injector (Dionex)
- PDA-100 Photodiode Array Detector (Dionex)
- STA-565 Column holder (Dionex)
- Polypropylene HPLC (0.5 and 1.5 mL) vials (ThermoFisher)
- HPLC caps (ThermoFisher)
- G3000 SWXL gel filtration HPLC column (TOSOH)

Gel electrophoresis equipment

- Glass plates (ATTO)
- Magnetic clamps (ATTO)
- Silicone rubber gaskets (ATTO)
- Electrophoresis tank (ATTO)
- Electrophoresis cables (ATTO)
- Powerpac Basic (Bio-Rad)
- Orbital Shaker (AQS Shaker Manufacturing)

Software

- ASTRA 6.1 (Wyatt)
- NTA 2.3 (Malvern Panalytical)
- Image Capture Engine Software 6.02 (AMT)
- Microsoft Excel 2010 (Microsoft)
- Origin (Pro 8.1 and 2017) (OriginLab Corp.)
- MassLynx 4.1 (Waters)
- MATLAB (Version unknown) (Mathworks)
- Genetool Version 4.03 (SynGene)
- CamSol webserver (Section 2.2.17)
- StavroX 3.6.0.1 (University of Halle)
- Chromeleon 6.4 (Dionex)
- ProtPARAM webserver (Section 2.2.17)
- Clustal Ω webserver (Section 2.2.17)
- ABodyBuilder webserver (Section 2.2.17)
- Solubis webserver (Section 2.2.17)
- PyMOL (Version 4.4) (Schrödinger)

2.1.2 Chemicals (all reagent grade unless stated)

Buffer salts and acids

- HPLC-grade Ammonium acetate stock solution (7.5 M) (Sigma-Aldrich)
- L-Arginine (Acros Organics)
- Sodium succinate dibasic hexahydrate (Sigma-Aldrich)
- Sodium acetate (Sigma-Aldrich)
- L-Histidine (Sigma-Aldrich)
- Sodium phosphate monobasic (Sigma-Aldrich)
- Sodium phosphate dibasic (BDH Lab Supplies)
- 4-(2-Hydroxyethyl)piperazine-1-ethanesulphonic acid (HEPES) (Sigma)
- Sodium chloride (Fisher Scientific)
- Tris {(trishydroxymethyl)aminomethane} base (Fisher Bioreagents)
- Guanidine hydrochloride (Sigma-Aldrich)
- Urea (Thermo)
- Hydrochloric acid (37% w/w) (Honeywell/Fluka)
- Nitric acid (70% w/w) (Honeywell/Fluka)
- Glacial acetic acid (95% w/w) (Fisher Scientific)

Dyes (covalent and non-covalent) and covalent cross-linkers

- Alexa-488 (Molecular Probes/ Invitrogen)
- Uranyl acetate (2% w/v) (source unknown)
- 5-{2-(iodoacetamido)Ethylamino}Naphthalene-1-sulphonic acid (IAEDANS) (Molecular Probes/Invitrogen)
- Nile Red (Sigma-Aldrich)
- 8-Anilinonaphthalene-1-sulphonic acid (ANS) (Sigma-Aldrich)
- bis(sulfosuccinimidyl)suberate (BS³) cross-linking reagent (Thermo Scientific)
- BS³-d⁴(deuterated) cross-linking reagent (Thermo Scientific)

Electrophoresis-related chemicals

- Protein Plus Protein Dual Xtra molecular weight ladder (Bio-Rad)
- Instant Blue (Coomassie Brilliant Blue) stain (Expedeon)
- SilverXpress Silver Staining Kit (Invitrogen)
- 10x Cathode buffer (Crystal buffers Tris-Tricine SDS) (Severn Biotech)
- 10x Anode buffer (2M Tris.HCl pH 8.8) (home-made)
- 30% w/v Acrylamide:Bis-acrylamide (37.5:1 ratio) solution (Severn Biotech)

- Tetramethylethylenediamine (TEMED) (Sigma)
- Ammonium persulphate (APS) (Sigma-Aldrich)

Miscellaneous chemicals used

- Bromophenol blue (Sigma-Aldrich)
- Glycerol (Fisher Scientific)
- Dithiothreitol (DTT) (Formedium)
- Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) (Aldrich)

2.2. Methods

2.2.1. Proteins used in the study and their preparation

Preparation of Bovine Serum Albumin (BSA)

BSA was purchased from Sigma Aldrich at $\geq 98\%$ purity. The protein was reconstituted at a concentration of $\sim 35\text{-}55\text{ mg mL}^{-1}$ in $0.22\text{ }\mu\text{m}$ -filtered and de-gassed 25 mM ammonium acetate buffer, pH 5.1. The protein was purified by gel-filtration chromatography using a 26/60 Superdex 200 column (GE Healthcare), pre-equilibrated with 25 mM ammonium acetate buffer, pH 5.1. The protein was loaded onto the column in 3.5 mL injections, with the column operated at 1 mL min^{-1} with an AKTA Prime (GE Healthcare). Fractions were pooled and the purity of the protein validated by SDS-PAGE using a 12% (w/v) SDS-PAGE gel (see Section 2.2.2). The purified, monomeric BSA was stored in frozen aliquots at $-20\text{ }^{\circ}\text{C}$ until needed.

Prior to stress experiments, aliquots were thawed at room temperature and where necessary, concentrated by centrifugal filtration using a 30 kDa cut-off centrifugal filter (Merck Millipore) at $4,000\text{ rpm}$ in an Eppendorf 5810 centrifuge. Samples were then filtered using a syringe-driven $0.22\text{ }\mu\text{m}$ filter (Merck Millipore and JET BIOFIL) and the concentration of the sample determined by UV-Visible spectroscopy (see protein concentration determination below).

Preparation of MedImmune mAb solutions

Antibodies from MedImmune Ltd were supplied as frozen, 1 mL aliquots (following Protein A chromatography). These were defrosted, aliquoted and snap-frozen in liquid nitrogen. The tubes of protein were then stored at -80 °C until further use. The three mAbs procured from MedImmune are all of the IgG1 class: MEDI1912 (called WFL herein), MEDI1912_STT (called STT herein) (Dobson et al., 2016) and Nip109 (called mAb1 herein). For use in flow experiments, aliquots were defrosted on ice, then dialysed for 4 hrs into the desired buffer using 3.5 kDa cut-off GeBA Dialysis tubes (Generon). The buffers used in the thesis were: 150 mM ammonium acetate, pH 6.0; 125 mM L-arginine + 20 mM sodium succinate, pH 6.0; 10 mM sodium acetate, pH 6.0; 10 mM sodium succinate, pH 6.0; 10 mM L-histidine pH 6.0, 10 mM sodium phosphate pH 7.2 and 100 mM sodium phosphate, pH 7.4. All buffers were filtered through a 0.22 µm filter, de-gassed and stored at 4°C. Following dialysis, the proteins were driven through a 0.22 µm filter and their concentration determined using UV-visible spectroscopy.

Preparation of Adimab protein solutions

IgG1 samples from Adimab were supplied frozen on ice in 5 mL aliquots (Jain et al., 2017) (purified by Protein A chromatography). Samples were defrosted, aliquoted, snap-frozen in liquid nitrogen and stored at - 80 °C until required. mAb samples were dialysed into filtered and de-gassed 25 mM HEPES, 150 mM sodium chloride, pH 7.3 in the same manner as above. Following dialysis, these mAbs were also passed through a 0.22 µm syringe filter and their concentrations determined using UV-visible spectroscopy.

Protein concentration determination

Protein concentrations were determined using an Amersham Biosciences Ultrospec 2100 Pro UV/Visible spectrophotometer. UV-transparent cuvettes (STARSTEDT) were used to measure the A_{280} of the reference and protein samples. The molar extinction coefficients (ϵ_{280}) and molecular masses (Mw) for the proteins used in the thesis are shown in Appendix Table 1.1. Once determined, the protein solutions were diluted to the appropriate concentration with buffer prior to use in experiments.

2.2.2 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

Plate setup

Two SDS-PAGE gels were prepared by placing a silicone gasket in between a flat glass plate and a raised spacer plate (ATTO). This creates a 1.5 mm gap in between the plates. This was repeated to make the second gel. The moulds were clamped and then the appropriate gel solution added. In two-component (stacking and resolving gels), the resolving gel was cast, a layer of Milli-Q H₂O added and allowed to set for one hour. The stacking gel was then cast, a 12- or 14- lane comb added to create the wells, then the gel left to set for at least one hour to set. Where no stacking gel was used, the separating gels were cast and the comb inserted immediately, before allowing the gels to set. This setup creates two, 80 mm x 100 mm gels. For all gels, APS and TEMED were added last to each mixture, before vortexing and casting the gel.

'Conventional' SDS-PAGE gels

The components for 12% (w/v) SDS-PAGE gels are detailed in Table 2.1.

Table 2.1. Gel components for 12% (w/v) SDS-PAGE gels.

Component	Resolving gel (mL)	Stacking Gel (mL)
Milli-Q H ₂ O	6.4	3
30% acrylamide:bis-acrylamide solution	8	0.7
1.5 M Tris.HCl, 0.4 % (w/v) SDS pH 8	5.2	1.25
10% (w/v) ammonium persulphate (APS)	0.05	0.1
TEMED	0.05	0.01

Chemical cross-linking SDS-PAGE gels

Initial cross-linking gels were prepared according to Devine, 2016. This is shown in Table 2.2.

Table 2.2. Initial cross-linking gel recipe

Component	Resolving gel (mL)	Stacking Gel (mL)
Milli-Q H ₂ O	0.44	3.72
30% acrylamide:bis-acrylamide solution	7.5	0.83
3 M Tris.HCl, 0.3 % (w/v) SDS pH 8.45	5	1.55
Glycerol	2	-
10% (w/v) ammonium persulphate (APS)	0.1	0.1
TEMED	0.04	0.04

Later gels (as indicated in Chapter 5) were prepared with a revised, 8% (w/v) gel as follows in Table 2.3.

Table 2.3. Lower % (8% w/v) SDS-PAGE gel for later cross-linking experiments

Component	Resolving gel (mL)	Stacking Gel (mL)
Milli-Q H ₂ O	9.45	5.3
30% acrylamide:bis-acrylamide solution	4.27	2
3 M Tris.HCl, 0.3 % (w/v) SDS pH 8.45	2.05	1.3
Glycerol	-	-
10% (w/v) ammonium persulphate (APS)	0.1	0.1
TEMED	0.04	0.04

Semi-denaturing SDS-PAGE gels

For SD-SDS-PAGE, 6% (w/v) gels were prepared with no stacking gel as follows:

Table 2.4. 6% (w/v) SD-SDS-PAGE gel recipe

Component	Resolving gel (mL)	Stacking Gel (mL)
Milli-Q H ₂ O	13	-
30% acrylamide:bis-acrylamide solution	2	-
3 M Tris.HCl, 0.3 % (w/v) SDS pH 8.45	2.6	-
Glycerol	-	-
10% (w/v) ammonium persulphate (APS)	0.1	-
TEMED	0.04	-

Electrophoresis setup

SDS-PAGE samples were either run in tandem or individually with a blank with the equipment listed in Section 2.1. Samples were diluted with either 6x loading buffer (300 mM Tris.HCl, pH 6.8, 300 mM DTT, 6% (w/v) SDS, 0.3 % (w/v) bromophenol blue and 60% (v/v) glycerol) or in SD-SDS-PAGE native loading dye (as above minus SDS and DTT), prepared by Dr Amit Kumar. Samples were loaded into the gels and electrophoresed according to the relevant section herein. In the gel electrophoresis tank, the cathode buffer contained 100 mM Tris.HCl, 100 mM Tricine, 0.1% (w/v) SDS, pH 8.3, whilst the anode buffer contained 400 mM Tris.HCl, pH 8.8.

2.2.3 Extensional flow apparatus and experimental setup

The EFD consists of two, 1 mL Hamilton Syringes (1001 RN SYR), drilled at the end to accommodate a 0.3 mm internal diameter (I.D) borosilicate glass capillary (Sutter Instruments). The capillary length was 75 mm in all experiments, except shear experiments where capillaries were cut to either 50 mm or 37.5 mm long with a ceramic cutter (Sutter Instruments), then flame-finished in a Bunsen burner. Prior to each run, the syringes and barrels were washed with 2% (v/v) Hellmanex-III solution (Hellma Analytics), then Milli-Q water, then the appropriate filtered and de-gassed buffer. Compression fittings (Hamilton) and O-rings (Gilson) were fitted to the capillary and connected to one syringe. The setup was inspected to ensure it was sealed with buffer (i.e. no air-bubbles), then 0.5 mL of protein solution drawn *slowly* into the syringe through the capillary, before removal of any obvious air-bubbles. The syringes were then sheathed in silicone tubing and clamped to the EFD with syringe clamps (Thor Labs). Once connected and sealed, the syringes were driven at a set plunger velocity by a stepper motor (Haydon Switch and Waterbury CT Instrument Co), programmed by an Arduino Microcontroller. Unless stated otherwise, this velocity = 8 mm s⁻¹ (strain rate = 11,750 s⁻¹, shear rate 52,000 s⁻¹, see Section 3.1). The protein solution was shuttled between the two syringes for the desired number of passes, then the motor stopped, the rig disconnected and the solution expelled *slowly* into an Eppendorf tube down the side to avoid cavitation. One pass is defined as the emptying of one syringe and the re-filling of another. For a frame of reference, one pass at 8 mm s⁻¹ takes ~6 s, thus 100 passes takes ~ 10 minutes.

Control protein samples, unless otherwise stated, were incubated at ambient temperature in an Eppendorf tube for the duration of the longest flow experiment prior to analysis. These samples will be referred to as 'native' or 'quiescent' herein. Images of the flow device and apparatus are shown in Figures 2.1 to 2.3.

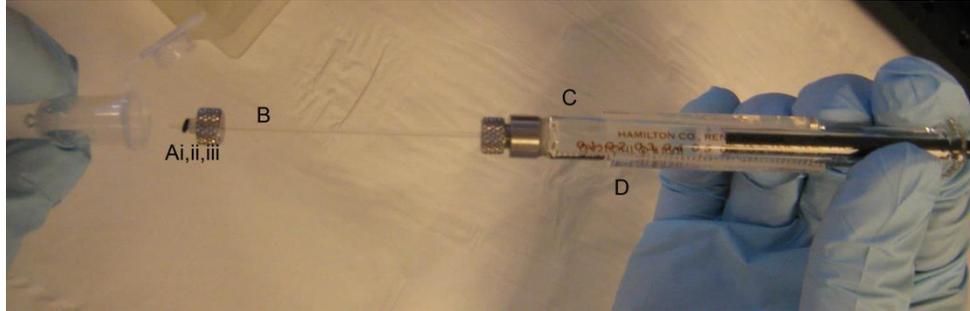


Figure 2.1a. Syringe and capillary used for extensional flow stress experiments. Ai= Rubber Gilson O-ring, ii= Ferrule, iii = Steel syringe cap. B = 0.3 mm (inner diameter) capillary (75 mm length). C = Hamilton syringe, containing filtered protein solution. D = silicone collar.

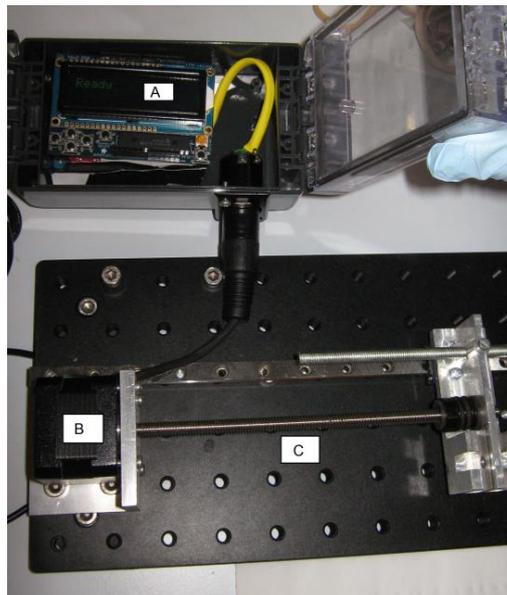


Figure 2.1b. Mechanical driver of the extensional flow device. A = Microcontroller unit. The number of cycles is shown on the display once the rig is started. B = Stepper motor. C= Motor shaft, which drives the syringe plungers.

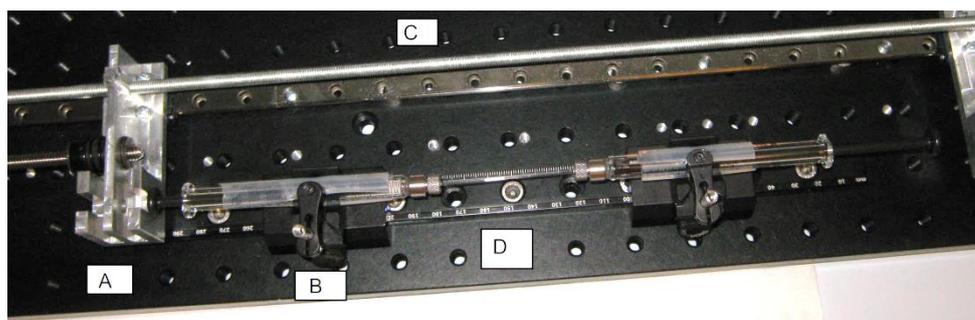


Figure 2.1c. The connected syringes in the extensional flow device. A = syringe driver which moves towards the motor shown in 2.1b (to the left) when the device starts and vice-versa. B = Syringe clamps which hold the syringes in place. C = stainless steel rod which connects the two syringe drivers to the stepper motor. D = Stainless steel baseplate to keep the rig level. The sample syringe is loaded on the right hand side.

2.2.4. Dynamic light scattering (DLS)

Experimental

Quiescent or stressed protein samples were analysed by batch-mode DLS using a Wyatt miniDawn TREOS system, equipped with an additional quasi-elastic light scattering (QELS- synonym for DLS) detector. All mAb samples were analysed without prior dilution. 5 mg mL⁻¹ BSA samples were diluted 1:2 with 25 mM ammonium acetate pH 5.1, whilst 10 mg mL⁻¹ BSA samples were diluted 1:5 with the same buffer prior to analysis. A five-minute buffer baseline was recorded in the Astra 6.1 software, followed by injection of 250 µL sample into the instrument. DLS data were collected for five minutes, before 1 mL of 0.22 µm-filtered and de-gassed buffer was injected and a second, five-minute baseline collected. The flow cell was then cleaned by injection of 1–2 mL (each) of 1M Nitric acid, MilliQ H₂O and then buffer. The buffer was kept on ice to minimise cavitation (air bubble formation) in the instrument. Three-minutes of DLS data were analysed in the Astra 6.1 software.

Theoretical (data processing)

As stated in Section 1.3.4, DLS measures the fluctuation of scattered light as a function of time. A second order correlation function describes these fluctuations thus (Equation 2.1):

$$g^{(2)}(\tau) = \frac{\langle I(t)I(t + \tau) \rangle}{\langle I(t) \rangle^2}$$

Equation 2.1. $I(t)$ is the intensity of the scattered light at a time t (s). $\langle \rangle$ brackets indicate averaging over all time.

The correlation function is based on the delay τ , i.e. the amount the intensity trace shifts from the original prior to averaging. For a monodisperse sample, the correlation function is given in Equation 2.2:

$$g^{(2)}(\tau) = B + Ae^{(-2\Gamma\tau)}$$

Equation 2.2. B = baseline for the correlation function at infinite delay, A is the correlation function amplitude at zero delay and Γ is the decay rate.

Γ is found by fitting a non-least linear squares algorithm to the measured correlation function data (performed by the Astra software). The diffusion coefficient for the particle in question can then be found with Equation 2.3:

$$D = \frac{\Gamma}{q^2} \text{ where } q = \frac{4\pi n_0}{\lambda_0} \sin\left(\frac{\theta}{2}\right)$$

Equation 2.3. n_0 is the refractive index of the solvent, λ_0 is the vacuum λ (i.e. not in solution) of the incident photon and θ is the scattering angle.

Finally, the Stokes-Einstein Equation is used to determine the hydrodynamic radius of the particle (see Equation 1.1, Section 1.3.4). Example correlograms obtained from DLS were fitted to a single-exponential decay in Origin with Equation 2.4.

$$y = y_0 + Ae^{\frac{-(x-x_0)}{\tau}}$$

Equation 2.4. Single Exponential Decay equation fit to correlation function data throughout the report. A is the amplitude, y_0 is the y axis intercept, x_0 is the x -axis intercept and τ is the delay time.

Non-monodisperse samples were further analysed to determine the hydrodynamic radii (average or distributions) of species in solution. DLS data can be subjected to both Cumulants and Regularisation analysis in Astra. The Cumulants analysis permits the calculation of the z-average (mean) radius and width (standard deviation) of a solution (Hanlon et al., 2010). These values can then be used to calculate the polydispersity index (PDI) of the solution according to Equation 2.5 (Roger et al., 2016).

$$PDI = \left(\frac{\sigma^2}{z^2} \right)$$

Equation 2.5. Calculation of Polydispersity index (PDI) from the square of the 'width' of the z-average radius (s.d. in nm²) and the square of the z-average radius (nm²).

The Regularisation algorithm is much more complex (Provencher, 1982), but essentially fits as few single-exponential decays as possible to best describe the DLS data (Hassan et al., 2014; Minton, 2016). The y-axis of a regularisation plot is the differential intensity fraction (log nm⁻¹) against the size of the species in solution (nm). All data were processed in Microsoft Excel and plotted in Origin.

2.2.5 Nanoparticle Tracking Analysis

Experimental

Native and stressed BSA samples (5 and 10 mg mL⁻¹) were diluted to the same extent as in Section 2.2.4, to minimise noise. 250 µL of protein sample was injected against gravity into a Nanosight LM10 instrument (Malvern Panalytical), equipped with a 642 nm laser. Upon checking for expulsion of air, three, 90-second long videos were recorded in the NTA 2.3 software. The instrument settings were as follows: screen gain = 1, detection threshold = 10nm, T = 22°C, η = 0.95 cP and camera brightness = 4–12 (to minimise background noise). Following data capture, the sample was withdrawn by syringe. The apparatus was then disassembled and cleaned stringently with 70% (v/v) ethanol solution. Compressed air was used to remove residual ethanol in the entry and exit ports of the instrument.

Theoretical (data processing)

The three videos (per sample) were batch processed in the NTA 2.3 software, where the 2D motion of the particles in time (from the videos) was used to calculate the apparent diffusion coefficient of the particles and count them (see Section 1.3.4). The particle size is reported as hydrodynamic diameter in nm. The collated data were then averaged in Microsoft Excel 2010 and plotted in Origin Pro.

2.2.6 Transmission Electron Microscopy (TEM)

Experimental

Native and stressed BSA samples (5 and 10 mg mL⁻¹) were diluted to the same extent as in Section 2.2.4, to avoid over-staining the EM grids. A colloidal carbon-coated copper EM grid (prepared by Martin Fuller, University of Leeds) was immersed in 20 µL of protein sample for 45 s. Excess sample was then blotted off onto Whatman filter paper, before the grid was washed and blotted in 3 x 10 µL of Milli-Q H₂O. The grid was stained by immersing the sample in 10 µL of 2% (w/v) uranyl acetate solution, before immediately blotting the sample and allowing it to air-dry (~1 minute). The grid was then stored in a protective holder for at least 24 hours before imaging on a JEOL JEM1400 TEM at 120 kV.

Data processing

Images were recorded at 1,000x and 10,000x magnification using the AMT software provided with the instrument. Images were cropped to size using Microsoft Paint.

2.2.7 Insoluble protein pelleting assay

Experimental

Two x 200 µL of quiescent and stressed protein sample (two tubes per sample) were loaded into Beckmann Coulter Ultracentrifuge Tubes. The samples were centrifuged at 30,000 rpm (~35,000 xg) for 30 mins at 4°C in a Beckmann Coulter Optima ultracentrifuge, equipped with a TLA100 rotor. Following this, 150 µL of supernatant was removed from the tube. For BSA, the remaining pellet was re-suspended in 200 µL of 6M guanidine hydrochloride in 25 mM Tris HCl buffer, pH 6.0. This pellet fraction was then diluted to a final volume of 2 mL and left overnight at room temperature. 50 µL of the supernatant was also diluted to 2 mL final

volume and incubated overnight at room temperature. The A_{280} of each sample was then recorded by UV-visible spectroscopy (Section 2.2.1).

For the mAbs, 200 μL of 6M guanidine hydrochloride in 25 mM Tris HCl buffer, pH 6.0 was added to each pellet fraction, the tubes sealed with Parafilm and incubated at 4°C overnight. 50 μL of the supernatant was added to 200 μL of the same denaturing buffer and incubated overnight at 4°C. The A_{280} of the mAb samples were then measured by UV-Visible spectroscopy as above. In Chapter 4, BSA aggregation landscape data were produced using this mAb protocol.

Theoretical

The A_{280} values for each sample were used to calculate the amount of protein in the pellet and supernatant fractions. The extinction coefficients from 2.2.1 were used for all proteins, except BSA where the theoretical extinction coefficient was used ($43,220 \text{ M}^{-1}\text{cm}^{-1}$) (Gill and von Hippel, 1989). The difference between these was used to calculate the % protein in the pellet fraction (% protein in pellet) according to Equation 2.6.

$$\% \text{ protein in pellet} = \left(\frac{([P] - [S])}{([protein_0])} \right) \times 100$$

Equation 2.6. Equation to calculate the % protein in pellet. [P] and [S] are the molar concentrations of protein in the pellet and supernatant fractions respectively. [Protein₀] is the molar concentration of the native protein stock, e.g. 0.5 mg mL⁻¹ mAb ~ 3.5 μM . N.B. Total recovery of insoluble protein from the pellet fraction was validated by calculating $(([P] + [S])/2)$ and comparing this to the [protein₀] above.

2.2.8 Native mass spectrometry of BSA

Experimental

Stressed or quiescent BSA samples were analysed using native electrospray ionisation-mass spectrometry (ESI-MS) using a Synapt HDMS mass spectrometer (Waters Corporation). Home-made sputter coated capillaries (Sutter instruments) were used to infuse the sample into the spectrometer. The spectrometer was operated in positive ion mode with the following settings: capillary cone voltage 1.2 kV, sample cone voltage 80 V, extraction cone 4 V, source temperature 40 °C, backing pressure 5–6 mBar, trap voltage 20 V,

trap/transfer gas flow 3 mL min⁻¹. The instrument was calibrated using aqueous caesium iodide (CsI) (Sigma Life sciences) ions.

BSA samples were analysed at a concentration of 10 µM, with samples being unclarified or centrifuged at 4°C for ten minutes at 8,000 xg in a bench-top centrifuge (Progen) prior to infusion into the spectrometer (see Section 3.2.)

A series of 25 mM ammonium acetate solutions were also prepared at pH 6.8, 5.1, pH 4 and pH 3, with titrations performed with glacial acetic acid. Native purified BSA was diluted down (in large excess) to a final concentration of 10 µM in these buffers, left to equilibrate for one hour at room temperature and then measured by mass spectrometry.

Data processing

Data were processed in MassLynx 4.1 to determine the mass-to charge ratio of each peak (m/z) and to assign peak masses.

2.2.9 Fluorescence correlation spectroscopy (FCS) of BSA

Experimental

To fluorescently label BSA, a 1 mg mL⁻¹ BSA solution was prepared in 4 M urea in 25 mM Tris.HCl, pH 7.5 (by Dr Amit Kumar). This was mixed 1:10 (mol:mol) with the sulfhydryl reactive dye, Alexa Fluor 488 C5 Maleimide (Molecular Probes, ThermoFisher Scientific) and the conjugation reaction allowed to proceed for 4 hr at room temperature. The labelled protein was purified from unlabelled protein and dye using a Superdex 10/300 GL gel-filtration column (GE Healthcare). Fractions were pooled and stored in black Eppendorf tubes at - 20°C.

In flow experiments, 10 µL labelled BSA (concentration = 1 mg mL⁻¹) was added to 500 µL of the BSA solution to be stressed, or to a quiescent sample. Following stress, the sample was analysed with a home-built FCS apparatus with a 1 fL confocal volume (Gell et al., 2008). 100 µL was loaded into the apparatus and 10 successive 30 s scans were recorded and then averaged to obtain the final correlation function.

Data fitting

Data were analysed by Dr Roman Tuma (University of Leeds) using non-linear least squares fitting using MATLAB scripts (Mathworks) (Tipping et al., 2015). The diffusion coefficients obtained were converted to apparent hydrodynamic radii by calibrating against the free dye. The general analytical expression of the auto-correlation function obtained by FCS is given in Equation 2.7.

$$G_{(\tau)} = \frac{1}{\bar{N}} \left(1 + \frac{\tau}{\tau_D}\right)^{-1} \cdot \left(1 + \frac{\tau}{K^2 \tau_D}\right)^{-\frac{1}{2}} + DC$$

$$\text{Where } \tau_D = \frac{\omega_0^2}{4D} \text{ and } K = \frac{z_0}{\omega_0}$$

Equation 2.7 Equation for the auto-correlation function obtained by FCS. \bar{N} is the average number of fluorophores in the point-spread function (the efficiency of fluorescence collection inside an instrument's confocal volume) at any time, z_0 is the $1/e^2$ sample volume in the direction of the optical axis, ω_0 is the sample volume perpendicular to the optical axis, DC is the value of the autocorrelation as $\tau \rightarrow \infty$ (typically = 1), τ_D is the molecular diffusion time and D is the translational diffusion coefficient of the molecule. Taken from (Gell et al., 2006).

The hydrodynamic radii extracted from the fitting above were processed in Microsoft Excel and plotted in Origin.

2.2.10 IAEDANS labelling of BSA in the presence and absence of TCEP

Experimental

5 mM 5-{2-(iodoacetamido)Ethylamino}Naphthalene-1-sulphonic acid (IAEDANS) (Molecular Probes), from a 167 mM stock prepared in 1 M Tris.HCl, pH 8, was added to 5 mg mL⁻¹ BSA (in 25 mM ammonium acetate, pH 5.1) and stressed for 0–100 passes in the EFD at a plunger velocity of 8 mm s⁻¹. For experiments containing Tris(2-carboxyethyl)phosphine (TCEP), this was either added to the BSA on isolation prior to stress, or added to the dye mixture (final concentration = 0.5 mM from a 250 mM stock) before addition to the BSA as stated above. In *ex situ* labelling experiments, 5 mg mL⁻¹ BSA was stressed in the EFD under the conditions above, then incubated with the dye (with and without TCEP) for the same length of time as the flow experiment. All IAEDANS labelling experiments

were quenched with SDS-PAGE loading dye containing 200 mM dithiothreitol (DTT). The samples were then electrophoresed (Section 2.2.2) on a 12% (w/v) SDS-PAGE gel.

Data processing

The gel was irradiated with UV light using a Syngene Gel Documentation UV-transilluminator. The intensities of the fluorescent bands were quantified using the Gene Tool software supplied with the instrument. The gel was then subsequently stained with Coomassie Brilliant Blue. The gel experiment and images were performed and processed by Dr Amit Kumar.

2.2.11 Nile Red and 8-Anilinonaphthalene-1-sulphonic acid (ANS) fluorescence spectroscopy

Nile Red (Sigma-Aldrich) was dissolved in 1 mL of DMSO (Sigma-Aldrich) and a 180 μM stock prepared using a ϵ_{552} of 19,600 $\text{M}^{-1} \text{cm}^{-1}$ in DMSO (Castro et al., 2005). 0.5 mg mL^{-1} samples of STT and WFL in 0.1 M sodium phosphate, pH 7.4 were stressed for 10 passes at a plunger velocity of 8 mm s^{-1} in the EFD. 10 μM Nile Red (final concentration) was added to 500 μL of the protein stock either before stress (*in situ*) or after (*ex situ*). Quiescent samples were prepared by incubating the dye with protein at room temperature for 2 min. Following incubation, the samples were transferred to 1 cm path length quartz cuvettes (Hellma) and diluted 1:1 with 0.1 M sodium phosphate buffer, pH 7.4. Samples were then analysed using fluorescence emission spectroscopy using a PTI fluorescence spectrometer. Samples were excited at 550 nm and emission recorded at 560–740 nm with 5 nm slit widths. Three spectra were recorded and averaged in Microsoft Excel. The spectrum for the dye in buffer control was subtracted away from all spectra.

For ANS experiments, a 250 μM stock was prepared in 1 mL of Milli-Q H_2O . The stock concentration was determined using a $\epsilon_{350} = 4,900 \text{ M}^{-1} \text{cm}^{-1}$ (Azzi, 1974). Protein samples were prepared as above, except 25 μM ANS was used instead of Nile Red. A BSA reference sample was prepared by adding 25 μM ANS to 6.8 mg mL^{-1} BSA in 25 mM ammonium acetate, pH 5.1. Fluorescence emission spectroscopy was performed using the PTI spectrometer, with 7.5 nm slit widths. Samples were excited at 380 nm, with emission monitored between 400–600 nm. Three spectra were recorded for each sample, as well the dye in buffer reference

(subtracted away from all spectra as above). All spectral data were processed in Microsoft Excel and the final graphs produced in Origin.

2.2.12 Chemical cross-linking of WFL with BS³

Chemical cross-linking

WFL was dialysed into 100 mM sodium phosphate buffer, pH 7.4. A 1:1 mixture of BS³ cross-linker (XL) and BS³-d⁴, its deuterated analogue, was prepared in 20 mM sodium phosphate, pH 7.4. The 12.6 mM stock was incubated on ice. WFL (at a concentration of 0.5 mg mL⁻¹) was stressed for 10 passes with either a 50x or 100x molar excess (~175 or 350 μM) of XL mixture *in-situ* at a plunger velocity of 8 mm s⁻¹. WFL was also stressed for 10 passes, then XL added *ex-situ*. A quiescent WFL sample was also incubated with XL. All samples were incubated with cross-linker for four minutes in total at room temperature, before quenching the reaction with 20 mM (final) Tris.HCl, pH 8.0. Both reducing and native gel samples were prepared by adding protein: dye in a 5:1 ratio. Samples were boiled or unboiled as stated in Chapter 5. 15 μL of the native samples were loaded onto either an 8% or 12% SDS-PAGE gel and electrophoresed at 50 mA overnight. The gels were then stained using Coomassie Brilliant Blue and imaged using the SynGene Gel Documentation instrument.

Cross-link identification by mass spectrometry

Following this, peptides were analysed using Liquid Chromatography & Mass Spectrometry using a Synapt G2si instrument by Rachel George and Dr Patrick Knight in the Mass Spectrometry Facility, University of Leeds. Spectra were processed using StravroX software to identify cross-linked peptides (<https://www.stavrox.com/>) (Götze et al., 2012). The following software settings were used: WFL sequence in FASTA format; Minimum detection threshold of 10 ppm and 0.1 Da mass error; Expected modifications: reduction and alkylation of Cys (+ AcO⁻) and oxidation of Met; BS³ cross-linker with d⁴ also selected ('heavy'); Conjugation between K to KSTY; Identification of intra and dead-end cross-links; Remove pre-score.

2.2.13 High-performance liquid chromatography (HPLC) monomer loss assay

Experimental

WFL, mAb1 and STT were stressed for 0–100 passes at a plunger velocity of 8 mm s⁻¹. The protein concentration was either 0.25 or 0.5 mg mL⁻¹, with all proteins dissolved in 150 mM ammonium acetate, pH 6.0. Following stress, the samples were initially clarified using 0.22 µm syringe filters. 100 µL of filtrate was then diluted 1:1 with a 5x stock of buffer (125 mM L-arginine, 20 mM sodium succinate pH 6.0) and put into HPLC vials and sealed. Later samples were clarified by performing the insoluble protein pelleting assay (Section 2.2.7), then adding 100 µL of supernatant to the quench solution as above. All samples were kept at 4°C in Dionex automated sample injector. 50 µL of each sample was injected onto a TOSOH G3000 SWXL gel filtration column equilibrated with either 150 mM ammonium acetate, pH 6.0 or 125 mM L-arginine, 20 mM sodium succinate, pH 6.0. The column was operated at a flow rate of 0.4 mL min⁻¹. Control samples were loaded in triplicate (technical replicates) to calculate the HPLC loading error. The number of replicates for each sample is given in the text in Chapter 5.

Data processing

HPLC chromatograms were integrated in the Chromeleon software (Version 6.4). The path length of the HPLC UV detector was 1 cm. The peak areas from the integration were converted to the mass of monomer in Microsoft Excel. All data were plotted in Origin Pro.

2.2.14 Semi-denaturing sodium dodecyl sulphate polyacrylamide gel electrophoresis (SD-SDS-PAGE)

WFL, STT and mAb1 samples were stressed for 0–100 passes at a plunger velocity of 8 mm s⁻¹ under the same conditions as Section 2.2.13. Following stress, 50 µL of sample was added to 10 µL of native loading dye (no reducing agent or SDS) and the sample kept on ice. 25 µL of each sample was loaded onto a 6% (w/v) separating SDS-PAGE gel (no stacking gel) and electrophoresed at 45 mA for 3 hours. The gel was then stained with Coomassie Brilliant Blue (Expedeon Instant Blue) or silver stained using the manufacturer's (Invitrogen SilverXpress Silver Staining Kit) protocol. All incubation steps below were performed with shaking on an orbital shaker (AQS Shaker Manufacturing).

Briefly, the gel was washed twice with Milli-Q H₂O. The gel was fixed by adding Milli-Q water, methanol (Fisher Scientific) and acetic acid (Fisher Scientific) in a 6:5:1 (v/v/v) ratio (50 mL final volume), then left for ten minutes. The gel was then sensitised by adding Milli-Q water, methanol and Sensitiser (20:20:1 v/v/v ratio, 25 mL final volume) and incubating for 30 mins. This solution was discarded and the sensitisation step repeated. The gel was then washed for 5 mins in 50 mL of Milli-Q H₂O, the solution discarded and the step repeated. The gel was then stained with Stainer A, Stainer B and Milli-Q H₂O (1:1:98 v/v/v, 25 mL total volume) for 15 mins. Developing solution (1.5 mL of Developer and ~24 mL of ddH₂O) was added carefully to the gel until bands appeared, with constant shaking. Development was terminated by pipetting 5 mL of Stopper solution into the Developing solution, with constant agitation. The gel was left to stop for ten minutes, before the gel was washed three times with 50 mL Milli-Q H₂O washes. The gels were imaged at the end of each staining protocol using the SynGene Gel Documentation instrument.

2.2.15 Batch mode static light scattering

WFL, STT and mAb1 samples were prepared as in 2.2.13. Samples were injected into the Wyatt miniDAWN Treos instrument as in Section 2.2.4. The instrument simultaneously collects static light scattering at the angles 40°, 90° and 131° in addition to DLS. The light scattering data were analysed in the ASTRA 6 software by setting the concentration of the sample and using a dn/dc value of 0.186 g/cm (Section 1.3.4). Data were fit to the Debye mass model as shown in Equations 2.8 and 2.9.

$$\frac{R_{\theta}}{K^*c} = MP(\theta) - 2A_2cM^2P^2(\theta)$$

where $K^* = 4\pi^2n_0^2 \left(\frac{dn}{dc}\right)^2 \lambda_0^{-4}N_A^{-1}$ and

$$P(\theta) \sim 1 - \frac{2\mu^2\langle r^2 \rangle}{3!} \dots \text{ where } \mu = \left(\frac{4\pi}{\lambda}\right) \sin\left(\frac{\theta}{2}\right)$$

Equation 2.8. Reciprocal version of the Zimm equation for static light scattering. c = mass concentration of solute (g mL^{-1}), M = weight-average molar mass (g mol^{-1}), A_2 is the second osmotic virial coefficient (sometimes called B_{22}) (mol mL g^{-2}), K^* is an optical constant, with n_0 being the refractive index of the solvent at the incident wavelength (658 nm), N_A is Avogadro's number ($6.022 \times 10^{23} \text{ mol}^{-1}$), dn/dc is the differential refractive index increment (0.186 mL g^{-1} , Section 1.3.4), $P(\theta)$ is the theoretical form factor (describing the size and shape of the solute), $\langle r^2 \rangle$ is the root-mean square radius and R_{θ} is the excess Rayleigh ratio (cm^{-1}).

The Debye fit plots R_{θ}/K^*c against $\sin^2(\theta/2)$. As $\theta \rightarrow 0$, $P(\theta) \rightarrow 1$.

$$\frac{R_0}{K^*c} = M - 2A_2cM^2 \text{ where (if } A_2 = 0), M = \frac{R_0}{K^*c}$$

Equation 2.9. Debye equation for SLS data. Plotting this yields a straight line of $y = c - mx$. Plotting x and y yields the weight-averaged molecular weight of the sample at the y -intercept and the gradient = $\langle r \rangle$ or root-mean square radius of the solute (Wyatt, 1993).

When operated in batch mode, error in the weight-average mass measurement can be determined thus (Equation 2.10), with the uncertainty defined in Equation 2.11.

$$M_{avg} = \left(\frac{\sum_i M_i \sigma^2 M_i}{\sum_i \sigma^2 M_i} \right)$$

Equation 2.10. Calculation for the uncertainty-weighted average molar mass (M_{avg}), where M_i is the average-molar mass at a time i .

$$\sigma_{M_i} = \frac{1}{\sum_i \sigma^2 M_i}$$

Equation 2.11. Calculation for the uncertainty in the molar mass measurement, as defined by the Wyatt Astra 6.1 software manual.

2.2.16 Seeding and cross-seeding assays

BSA at a concentration of 5 mg mL⁻¹ was prepared as in Section 2.2.1. WFL and STT were also prepared as above to a concentration of 0.5 mg mL⁻¹ in 150 mM ammonium acetate, pH 6.0. Seed material was prepared by stressing each protein for the stated number of passes, then putting this material on ice. Seeding reactions were performed by adding 10% (v/v) seed material to native protein, then stressing the protein for the given number of passes at a plunger velocity of 8 mm s⁻¹. Following stress, the samples, including the quiescent controls, were analysed using the insoluble protein pelleting assay (Section 2.2.7)

For cross-seeding assays, a 10% (v/v) addition of seed material from protein x was added to a native stock of protein y and *vice-versa*. Stress experiments and quiescent samples were prepared and analysed as stated above.

2.2.17 Adimab protein dataset experiments and analyses

Choice of 33 mAbs from Jain et al. 2017 dataset

Of the 133 IgG1s in the Jain et al. 2017 dataset, 56 were chosen by selecting subsets of mAbs from the five clusters in Figure 4 of the Jain et al, 2017 paper (see Chapter 6): 22 from Cluster 1; 10 each from Clusters 2 and 3 and; 7 each from Clusters 4 and 5. V_H and V_L sequences for each of the mAbs were input into ProtParam to calculate their pI and molecular weight (Gasteiger et al., 2005). The intrinsic CamSol score of each V_H-V_L pair at pH 7 was also calculated by inputting protein sequences into the CamSol webserver (Section 1.3.1; <http://www-mvsoftware.ch.cam.ac.uk/index.php/login>) (Sormanni et al., 2015). Initially, five mAbs were selected which had wide separation in the Jain study: Daratumumab (1), Tabalumab (2), Denosumab (3), Ipilimumab (4) and Ixekizumab (5). The other 28 mAbs were selected from the initial subset, with alternatives suggested by Jain et al. depending on the availability of the molecules. This led to a final panel of 33 mAbs (Appendix Table 1.2).

Phylogenetic tree construction with Clustal Ω

The V_H and V_L sequences of the 33 mAbs were input into the Clustal Ω webserver (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) (Sievers et al., 2011). The tree was downloaded and manually annotated according to the Figure legend in Section 6.4.2.

Extensional flow experiments on the Adimab proteins

All Adimab proteins were prepared according to Section 2.2.2. Daratumumab (1) was initially stressed between 0–100 passes (at a concentration of 0.5 mg mL⁻¹) at a plunger velocity of 8 mm s⁻¹, before subjecting the samples to analysis by the pelleting assay. This preliminary assay led to the choice of 200 passes, 8 mm s⁻¹ plunger velocity, 0.5 mg mL⁻¹ protein as the parameters used to test the proteins. Quiescent samples were incubated alongside the stressed samples for 20 mins at ambient temperature prior to analysis with the pelleting assay (Section 2.2.7).

Flow experiments were performed on mAbs 1–5 in duplicate (biological repeats). Proteins from the additional 28 were initially stressed in groups of 6 or 5, with the composition of each group determined randomly. For the repeat of this set, the composition of each group of proteins was again selected randomly. Errors were then propagated according to Section 2.2.18.

Spearman's Rank analysis of the mAbs with and without flow data

Statistical analysis of the mAb data was performed by Tushar Jain (Adimab, MA, USA) according to Jain et al. 2017. To assess the correlation between the relative rankings of the mAbs within each group of assays, Spearman's rank correlations were calculated for all pairwise combinations of antibody characteristics. A p-value was calculated for each pairwise combination to assess the statistical significance of each correlation. In essence, the performance of a mAb in assay x was compared to that of assay y. This was performed on data for each of the 33 mAbs from the Jain et al. paper (Chapter 6), augmented with data from the EFD.

Rank analysis using 'distance from ideal measurement'

For the four groups of assays (Section 6.1), 'ideal' values were derived. For example, the ideal mAb would show 0% aggregation in a monomer loss assay, or 0% aggregation in the pelleting assay from the EFD. The pairwise distances for the 33 mAbs above, including and excluding extensional flow, were performed by Tushar Jain. Briefly, a mAb, i, was assigned a red flag if it exceeded the threshold value for an assay j. This value for a mAb was scaled according to Equation 2.12.

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

Equation 2.12. Equation for the scaled value for a mAb i, in an assay j. y = measured value, Y_{50%}, Y_{80%} and Y_{20%} are the median, 80th and 20th percentile of the 137 or 33 mAbs being analysed.

The assay values were then scaled to allow comparison between assays, omitting SGAC-SINS data. Scaled values were limited between $V_{ij} \pm 2$ scaled units. An ideal value for each assay was then assigned according to Equation 2.13.

$$V_j(\text{ideal}) = \min(V_{ij}) \text{ for each assay, } j$$

Equation 2.13. Equation for the ideal value for an assay j, based on the minimum scaled value in Equation 2.12 for each assay.

These values were then subtracted from one another (to yield the ideal assay value for each group of assays identified from the Spearman's rank analysis to give Equation 2.14.

$$S_{ij} = V_{ij} - V_j$$

$$\therefore S_{ij} = \bar{x}(S_{ij}) \text{ for an assay } j \text{ per group}$$

Equation 2.14. Equation for the ideal value for an assay j, based on the minimum scaled value in Equation 2.12 for each assay.

Finally, the 'distance from ideal' values were calculated for each group of assays (1–4 excluding flow or 1–5 including flow) according to Equation 2.15.

$$d_i = \sqrt{\frac{(\sum S_{ij}^2)}{4 \text{ (or 5 incl. EFD)}}$$

Equation 2.15. Equation to calculate the distance from ideal value, d_i, for each mAb.

The d_i values were hierarchically clustered by Tushar Jain in R. Distance from ideal measurements were converted to ranks by this author. The mAb closest to ideal was defined as having a rank of 1. To assign the rank of the mAbs an integer (1–33), the mAbs were ordered lowest to highest in Microsoft Excel (1 being best, 33 being worst).

scFv homology model construction

V_H-V_L sequence pairs, for each of the 33 mAbs, were input in FASTA format into the ABodyBuilder webserver (<http://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/Modelling.php>) (Leem et al., 2016). Output models were saved in .pdb format and edited in PyMOL.

APR identification in Solubis

APRs were identified within each V_H-V_L sequence pair by inputting the structures from above into the Solubis webserver (Section 1.3.1) (<http://solubis.switchlab.org/>) (Van Durme et al., 2016). This performs TANGO to calculate the aggregation propensity of a particular APR (higher value = more aggregation-prone), as well as running FoldX to determine the contribution of the APR to the thermodynamic stability of the protein ($\Delta G_{\text{contrib}} \leq 0$ = favourable). The Solubis score (essentially a structure-corrected TANGO score) for each APR was calculated using Microsoft Excel. $\Delta G_{\text{contrib}}$ scores for each APR were normalised from 0 to 1 ($\Delta G_{\text{contrib}}$ -5 to +5), then multiplied by the TANGO score for the APR. The solvent accessible surface area (SASA) for each APR was computed in PyMOL. Δ SASA calculations were performed in PyMOL by deleting either the V_H or V_L chain and re-computing the SASA of the APRs (see Section 6.4). Once all calculations were performed in Microsoft Excel, data were plotted using Origin.

2.2.18 Error propagation method

All experiments in this thesis were performed in at least duplicate, on separate days, unless otherwise stated. For samples where N = 2, with each sample yielding two technical replicates or an ensemble average (e.g. insoluble protein pelleting assay and Cumulants analysis, respectively), the final mean and error (standard deviation), $C \pm \delta C$, was calculated according to Equation 2.16:

$$\frac{\delta C}{C} = \sqrt{\left(\frac{\delta A}{A}\right)^2 + \left(\frac{\delta B}{B}\right)^2}$$

Equation 2.16. Error propagation for N = 2 samples. A = mean of replicate A, B = mean of replicate B, δA = s.d from replicate A, δB = s.d from replicate B. The equation is re-arranged and solved for δC (the propagated error). Large deviations between replicates are thus inflated according to the above equation.

3. Characterising protein aggregation induced by extensional flow

3.1 Overview of the extensional flow device (EFD)

As discussed in Sections 1.4 and 1.5, it has long been thought that proteins could have their structures perturbed by hydrodynamic forces, leading to an increase in their aggregation propensity. To date, this link has been tenuous, in part caused by the lack of clarity surrounding the fluid flows present in some studies. How extensional flow, in particular, affects proteins has been largely understudied. To investigate whether or not extensional flow has the ability to induce aggregation in globular proteins, a custom-made flow device was designed and validated by my colleague, Dr John Dobson (see Acknowledgements and Section 2.2.3).

3.1.1. Computational fluid dynamics under standard flow conditions

There are several different methods available to generate extensional flow fields (Section 1.4.1). A syringe-based device, incorporating a reduction in cross-sectional area was chosen for this study, owing to its simplicity, low sample volume, lack of air-water interfaces and laminar flow conditions. Computational fluid dynamics (CFD) allows one to model how the mass, momentum and energy of a fluid changes as a function of time. The basis of CFD are the Navier-Stokes equations, which describe the flow in terms of its pressure and velocity at a given point (Aris, 1962). The fluid is modelled by dividing it into packets or 'cells', which are then combined to form a mesh. The differential Navier-Stokes equations are then solved for each packet in the mesh (Sharma et al., 2011). With the device in mind, CFD was performed, modelling the fluid at a plunger velocity of 8 mm s^{-1} , which was selected as a baseline value. The CFD results are shown as part of a schematic for the device in Figure 3.1.

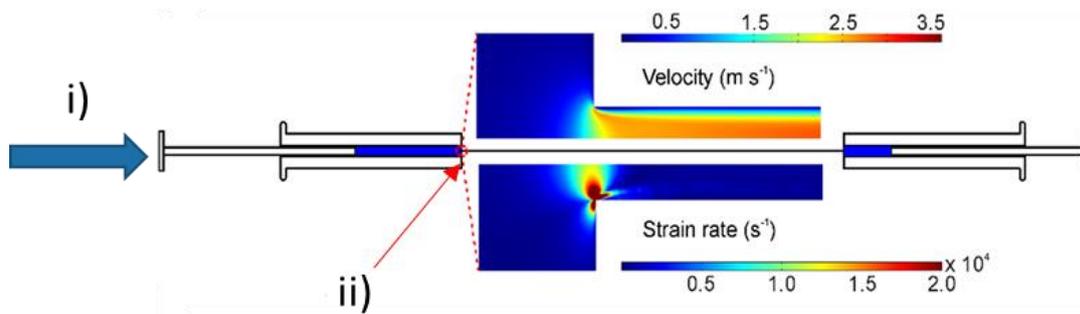


Figure 3.1. CFD results for the device, overlaid on a schematic of the EFD.

i) The direction of flow is shown by the blue arrow. ii) At the point of contraction (red arrow) between the syringe and capillary, the fluid is forced to accelerate.

The CFD profile (plunger velocity of 8 mm s^{-1}) at the top shows the resulting increase in the velocity of the fluid, whilst the bottom profile shows the strain rate at the contraction point. The contraction point thus represents the extensional flow region, whereas the velocity gradient across the width of the capillary represents the high shear region. CFD was performed by Dr John Dobson, University of Leeds, in COMSOL version 4.0.

The results show there is a rapid rate of increase in the velocity of the fluid (~238-fold) at the point of contraction between the syringe and capillary, due to the incompressible nature of the fluid, highlighting the extensional flow region characterised by the strain rate, $\dot{\epsilon}$. The shear rate $\dot{\gamma} = 52,000 \text{ s}^{-1}$ at the capillary wall. The shear stress = 46.3 Pa at the capillary wall. The increases in velocity and strain rate, at the contraction point, take place over a short distance of just 2 mm. This is summarised in Figure 3.2.

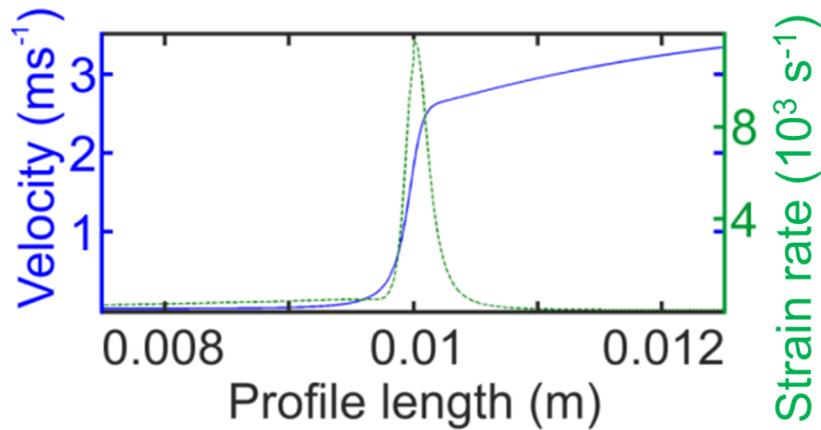


Figure 3.2. CFD result for the central streamline at a plunger velocity of 8 mm s⁻¹. The CFD results show the centre-line strain rate = 11,750 s⁻¹, with the velocity increasing from 8 mm s⁻¹ to 1.9 m s⁻¹. A profile position of 0.01 m corresponds to the point of contraction. Calculations performed by Dr John Dobson, University of Leeds.

Under these standard flow conditions (plunger velocity of 8 mm s⁻¹), during one pass (defined as the complete emptying of one syringe into the other driven by the stepper motor, Section 2.2.3), a protein within the fluid occupies: the barrels of the syringe for ~5 s; the extensional flow region (the contraction point) for 18 x 10⁻⁶ s and the high shear capillary region for 1 x 10⁻³ s. The more passes a protein experiences in the device, the more times it encounters these defined stresses. The Reynolds number was calculated according to Equation 1.5 (Section 1.4.1) for the fluid in the syringe and capillary.

$$Re = \frac{\rho v L}{\eta}$$

Equation 1.5. Equation for Reynolds number, reproduced from Section 1.4.1.

In the CFD model, at 20 °C, $\rho = 997 \text{ kg m}^{-3}$ and $\eta = 1.0016 \text{ N s m}^{-2}$. In the syringe, $L = 4.61 \times 10^{-3} \text{ m}$ and $v = 8 \times 10^{-3} \text{ m s}^{-1}$. In the capillary, $L = 3 \times 10^{-4} \text{ m}$ and $v = 1.9 \text{ m s}^{-1}$. Solving Equation 1.5 with the parameters above, $Re \approx 37$ in the syringe and 570 in the capillary. This means the flow remains laminar throughout the device.

3.1.2. Computational fluid dynamics under variable flow conditions

The EFD syringe plungers can be driven by the stepper motor at a variety of velocities. Maintaining the device geometry described in Section 2.2.3., CFD was performed at a range of plunger velocities (1–25 mm s⁻¹), with the strain rate, shear rate and exposure time to extensional flow being the principal outcomes. These results are shown in Figure 3.3.

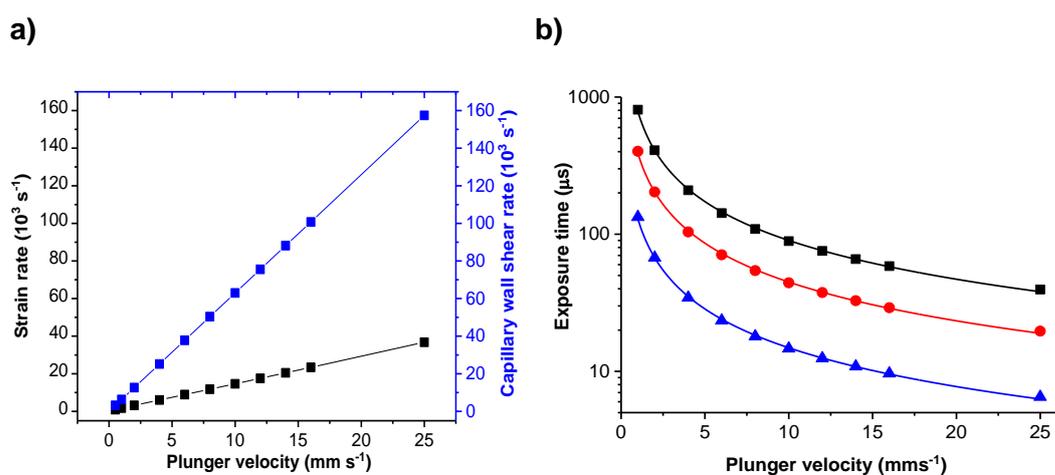


Figure 3.3. CFD results for the EFD at variable plunger velocity. a) Plot of strain rate (black) and capillary wall shear rate (blue) in the EFD at the plunger velocities shown. Linear trend lines ($R^2 = 1$) were fit to the data to guide the eye. b) Plot of exposure time of fluid elements exposed to 50% (black), 75% (red) or 90% (blue) of the maximum centre-line strain rate. Power law fits of the form $y = ax^{-0.94}$ were fitted to the data in Origin ($R^2 > 0.995$ in all cases). CFD data generated and fitted by Dr John Dobson, University of Leeds.

The data show that the centre-line strain rate and capillary wall shear rate vary linearly with plunger velocity. As the velocity increases, the length of time the fluid element is exposed to extensional flow decreases, as shown in Figure 3.3b. Further analysis of the CFD calculations revealed that the flow is no longer laminar in the capillary at plunger velocities exceeding 25 mm s⁻¹. Experiments which subjected proteins to increased strain rates ($> 11,750$ s⁻¹) were thus capped at a plunger velocity of 16 mm s⁻¹, in agreement with the criteria set out in Section 1.5.

3.2 Characterisation of BSA aggregation under defined fluid fields

3.2.1. Characterising BSA aggregate quantity and morphology

With an understanding of the strain and shear rates present within the device at a given plunger velocity, as well as knowledge of the length of time proteins will be exposed to these flow fields, experiments could be performed to examine if and how these flows affect protein aggregation. BSA, purchased from Sigma-Aldrich at 98% purity, was further purified (with Dr Amit Kumar, University of Leeds) using gel filtration chromatography (Section 2.2.1). A representative chromatogram and SDS-PAGE gel are shown in Figure 3.4.

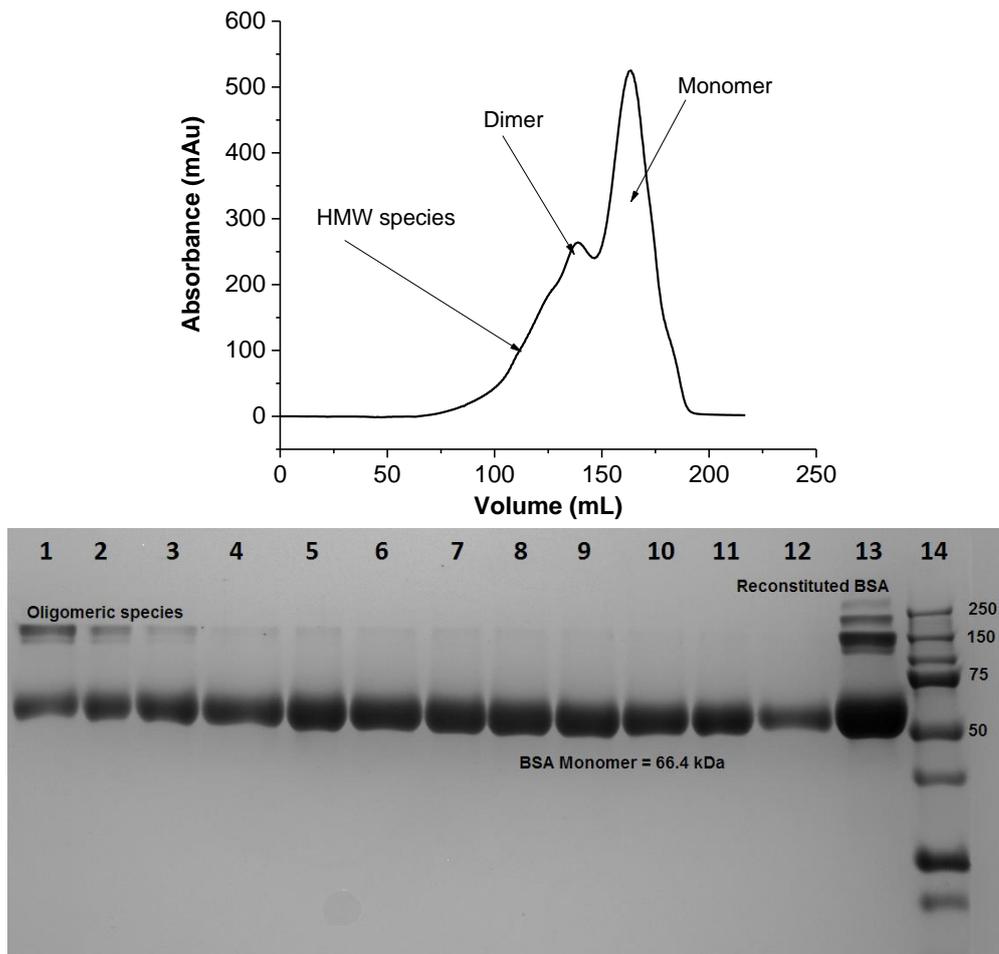


Figure 3.4. Further purification of BSA using gel-filtration chromatography (see Methods Section 2.2.1). (Top) A representative chromatogram for the purification process. (Bottom) 12% (w/v) SDS-PAGE gel for BSA. Lanes 1–3 = higher-molecular weight (HMW) species, 4–9 = dimer shoulder peak, 10–12 = monomer peak. Lane 13 = reconstituted BSA. 14 = MW ladder (standard masses given in kDa). 2 mL fractions were pooled between 156–190 mL as the monomeric protein, with the elution volume corresponding to a ~70 kDa protein.

Purified BSA (Section 2.2.1) was then stressed in the EFD, at a quantify protein loss from solution as a function of pass number (Section 2.2.7). The results are shown in Figure 3.5.

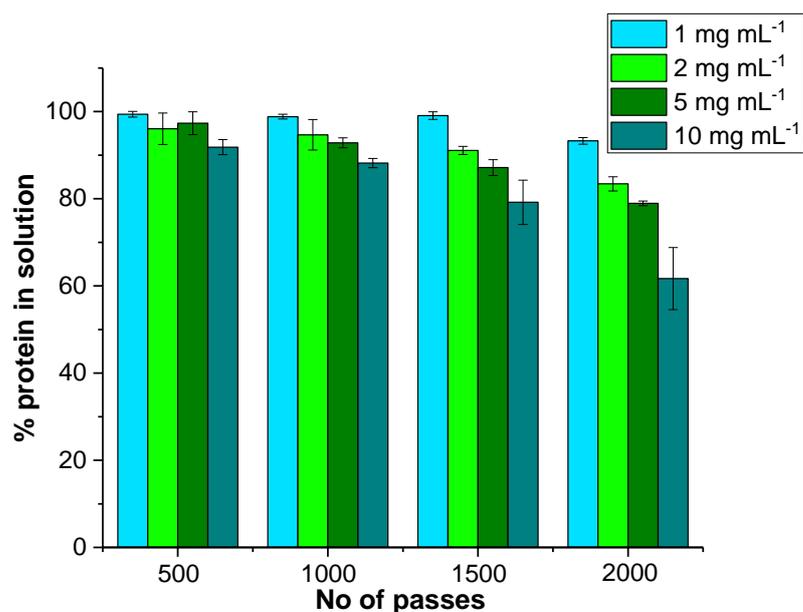


Figure 3.5. Determination of the amount of BSA lost from solution following hydrodynamic stress in the EFD, as determined using the pelleting assay (Section 2.2.7). Samples were stressed for 500–2000 passes at the concentrations shown in the legend. The plunger velocity = 8 mm s^{-1} , $\dot{\epsilon} = 11,750 \text{ s}^{-1}$, $\dot{\gamma} = 52,000 \text{ s}^{-1}$. The data were normalised to the quiescent samples, which showed no aggregation (100% in solution). Errors from two independent experiments were propagated according to Section 2.2.18.

The data show that as the number of passes increases, there is a decrease in the amount of protein left in solution, as quantified with the pelleting assay. The observed effect is sensitive to concentration after 1500 passes. For example, little aggregation (<10% loss of soluble protein) observed at 1 mg mL^{-1} compared to ~35% loss of protein from solution at 10 mg mL^{-1} after 2000 passes. The quiescent protein samples, incubated at ambient temperature for the duration of the flow experiment did not show any loss of protein from solution.

To understand the consequences of aggregation at a coarse level, the morphology of the aggregates formed in the EFD was assessed using TEM, Section 2.2.6. In addition to stressing the samples for long periods of time (500–2000 passes take 50–200 minutes to complete), samples were also analysed by TEM after 50 and 100 passes. EM grids were prepared immediately after stress, prior to analysis of the sample with the pelleting assay. The images from these samples are shown in Figure 3.6.

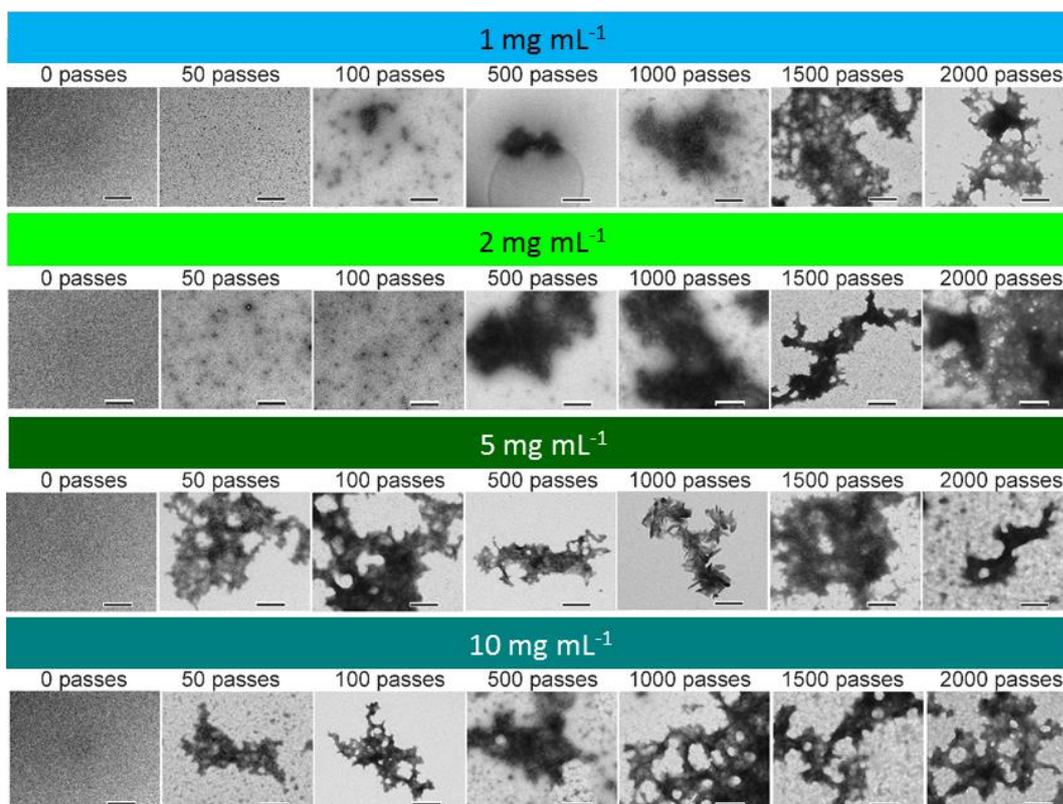


Figure 3.6. TEM images of BSA after subjection to hydrodynamic stress in the EFD at a plunger velocity of 8 mm s^{-1} for the number of passes and concentration stated. The images shown were taken at $10,000 \times$ magnification, with the scale bar = 500 nm .

The images above clearly show that the BSA aggregates which form in the EFD are amorphous in nature and several microns in size. At lower concentrations (1 and 2 mg mL^{-1} , light blue and green bars respectively), the aggregates appear much more disperse across the grids, only forming large amorphous species after 500 passes. Importantly, no amorphous aggregates were observed on the grids prepared from the quiescent samples. Together with the pelleting data, it is clear that following exposure to the hydrodynamic forces found in the EFD, BSA forms amorphous aggregates, which eventually become insoluble.

3.2.2 Measuring aggregate size and dispersity using biophysical techniques

According to the Lumry-Eyring Nucleation Polymerisation (LENP) model (discussed in Section 1.2; Andrews and Roberts, 2007; Li and Roberts, 2009; Lumry and Eyring, 1954), proteins typically aggregate to large, amorphous species along a pathway. The smaller oligomers which form *en route* to insoluble aggregates can be detected using an array of light scattering technologies (Section 1.3.4). DLS and NTA have both been used to size particles and aggregates in the 1 nm to 2 μm size range (Filipe et al., 2010), including BSA aggregates (Borzova et al., 2017).

To monitor the formation and growth of BSA aggregates as a function of pass number, BSA solutions were stressed for 10–2000 passes at a concentration of 1, 2, 5 or 10 mg mL^{-1} at a plunger velocity of 8 mm s^{-1} . Quiescent samples were also prepared as stated in the Methods. The samples were then analysed by DLS (see Section 2.2.4). The autocorrelation functions initially obtained from the DLS analysis assume monodispersity, as only a single exponential decay is fitted. Representative correlation functions for the BSA samples are shown in Figure 3.7.

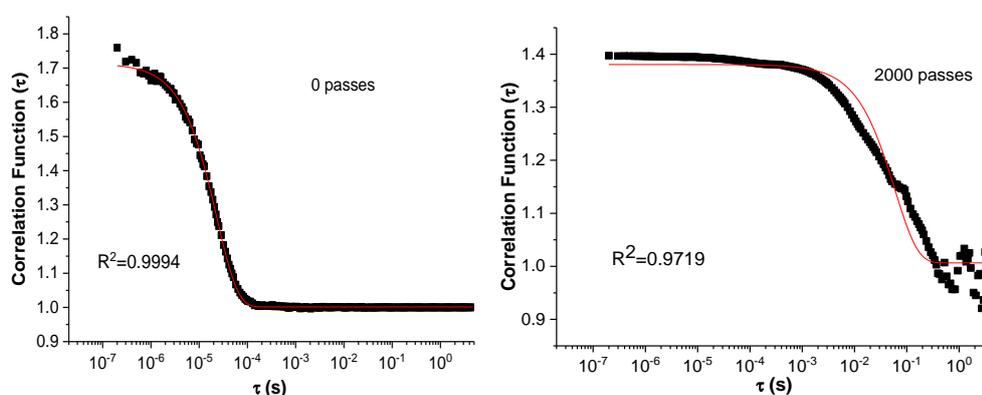


Figure 3.7. Representative correlation functions obtained by DLS for Quiescent BSA (left) and BSA after 2000 passes in the EFD (right). The samples were diluted 1:2 with buffer as stated in Section 2.2.4. The R^2 values shown are for the goodness of fit for a single-exponential decay to the data.

The autocorrelation functions show that quiescent samples are predominantly monodisperse and the species contained within them are small, hence the rapid decay and good fit observed in Figure 3.7. Conversely, where aggregates

dominate the scattering signal, the decay in the correlation function is slow and thus the single exponential fit is poor. This is indicative of high dispersity. The DLS data were thus deconvolved using both the Cumulants and Regularisation methods (see Section 2.2.4). As mentioned in Section 1.3.4, the Cumulants method is the simplest, yielding the z-average radius and width (standard deviation) of the sample. These data are displayed in Figure 3.8.

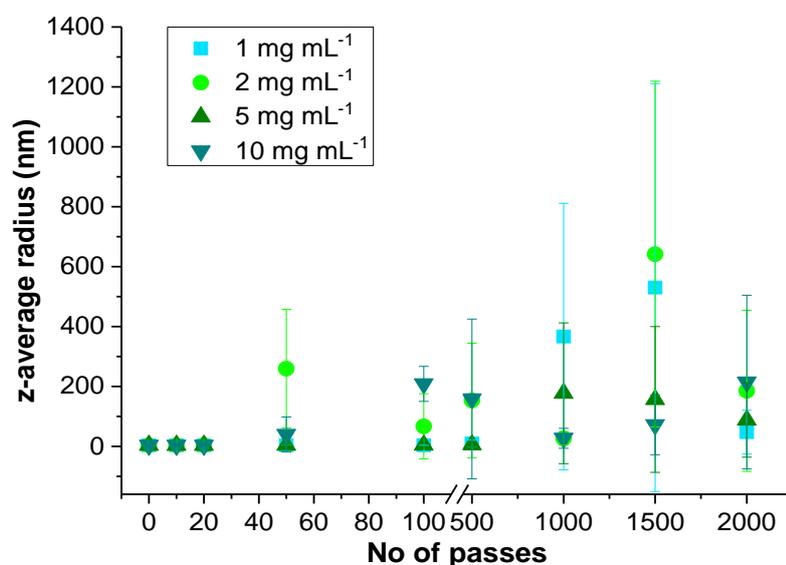


Figure 3.8. Plot of the z-average radii of BSA stressed for 0 (quiescent sample) to 2000 passes at a plunger velocity of 8 mm s^{-1} , obtained by Cumulants analysis of the DLS data. BSA was stressed at the concentrations shown in the legend. Errors from two independent measurements were propagated according to Section 2.2.18, except 2 mg mL^{-1} , 1000 passes ($N = 1$).

The data show that until BSA solutions have been subjected to 50 passes of stress in the EFD, no aggregation is observed, with the measured z-average radii of the samples being in close agreement with the literature value of $\sim 3.5 \text{ nm}$ (Atmeh et al., 2007; Yadav et al., 2011). Following 50 passes, the average size of the aggregates in the samples grows to several hundred nm, over a very broad size range, as indicated by the error bars in each point. Aggregates which are several microns in size (as shown by TEM, Figure 3.6) are possibly too large to diffuse readily into the flow cell, hence the z-average radius does not increase steadily as a function of pass number. The dispersity of the above samples can be quantified by the PDI (Section 2.2.4). The PDI of the samples were calculated and are shown in Figure 3.9.

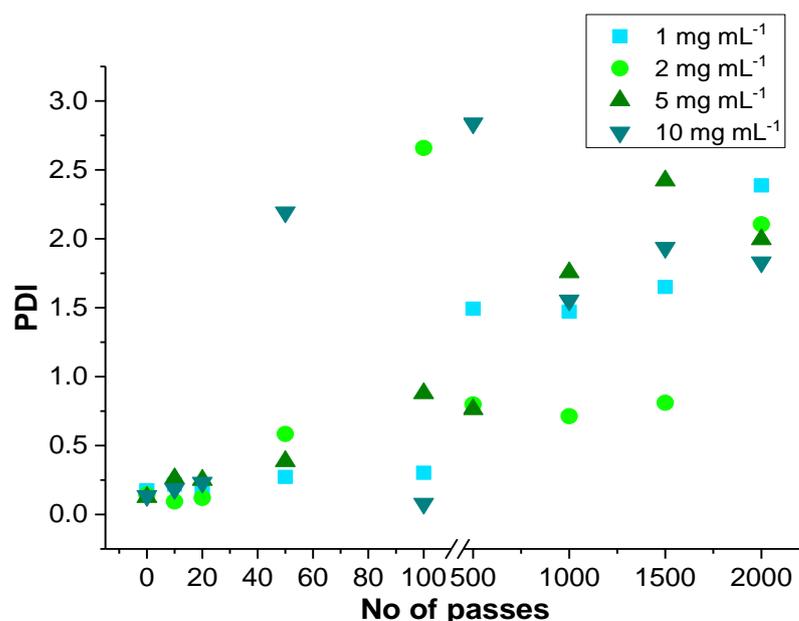


Figure 3.9. Plot of polydispersity index (PDI) values, calculated from the z-average radii measurements (and their widths) obtained from Cumulants DLS analysis of the BSA samples in Figure 3.8, against the number of passes.

The PDI data show that until 50 passes have been performed, the BSA samples remain essentially monodisperse, i.e. their measured PDI is close to 0.1. After more passes threshold has been exceeded, the samples become polydisperse, with the PDI measurements increasing from ~0.4 – 2.7 as a function of pass number. Samples with a PDI >0.6 are considered very disperse, which may compromise the accuracy of the size distributions obtained through fitting of the DLS data (Roger et al., 2016).

The Regularisation algorithm is more sophisticated than Cumulants; its increased resolving power thus allows one to see how the size distribution of aggregates changes as a function of pass number. A representative regularisation plot for a stressed and quiescent BSA sample is shown in Figure 3.10, with a complete set of values presented in Appendix Table 2.

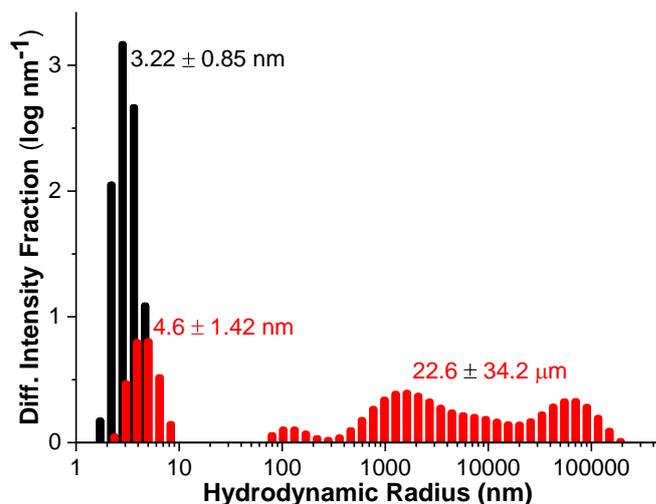


Figure 3.10. DLS regularisation plot for quiescent BSA (black) and BSA after 2000 passes of stress at a plunger velocity of 8 mm s^{-1} . The BSA concentration = 5 mg mL^{-1} .

The regularisation plots show that after stress in the EFD, the BSA aggregates exist over a broad size range of tens of nm to several microns. Their presence greatly reduces the signal intensity from that of the monomeric BSA (the red peak at 4.6 nm). However, DLS lacks the resolution to further distinguish populations in these broad peaks (Filipe et al., 2010). Nanoparticle tracking analysis was thus performed on unstressed and stressed BSA samples in the same manner as discussed above (see also Section 2.2.5). Particle concentration vs size plots for these samples are shown in Figure 3.11.

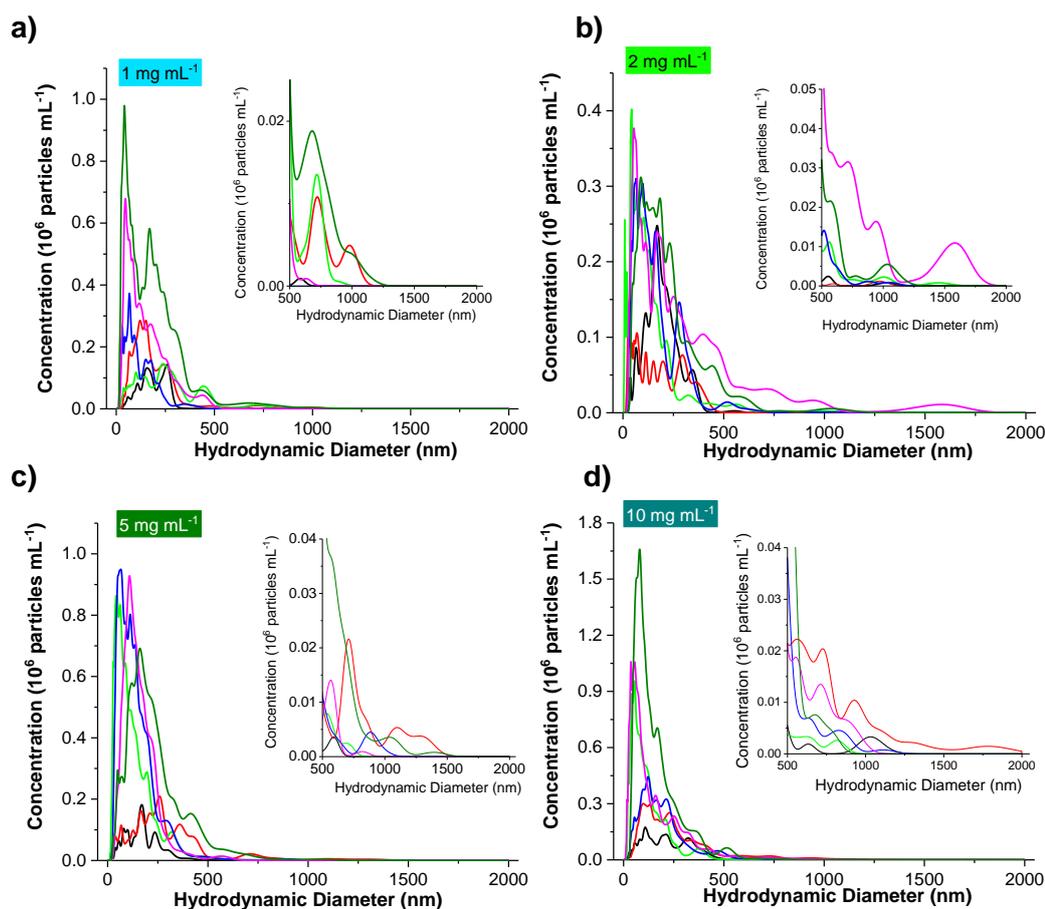


Figure 3.11. Particle number concentration vs size plots for stressed BSA samples analysed using NTA. a) 1 mg mL^{-1} , b) 2 mg mL^{-1} , c) 5 mg mL^{-1} and d) 10 mg mL^{-1} . All samples were stressed for the number of passes indicated at a plunger velocity of 8 mm s^{-1} . The quiescent samples and those stressed for 10 or 20 passes had no aggregates present when analysed. The coloured lines correspond to: 50 passes (—), 100 passes (—), 500 passes (—), 1000 passes (—), 1500 passes (—) and 2000 passes (—) respectively.

The data show that when stressed, BSA forms aggregates over a broad size range, from $\sim 40 \text{ nm}$ to 1800 nm in diameter. Generally, as the number of passes increases, the number of aggregates in the $40\text{--}500 \text{ nm}$ size range increases. As seen in the DLS data, the number of aggregates exceeding 1000 nm in diameter varies widely as a function of pass number, possible due to larger aggregates not diffusing (inset graphs). The quiescent samples, as well as those stressed for 10 or 20 passes, did not have any particles present when analysed by NTA, as any particles present (including monomer) are too small to be detected. This is in agreement with the DLS data in Figures 3.8 and 3.10. The total number of particles in a sample can also be quantified using NTA (Carr and Wright, 2015; Zhou et al., 2015). This analysis was performed to yield the data in Figure 3.12.

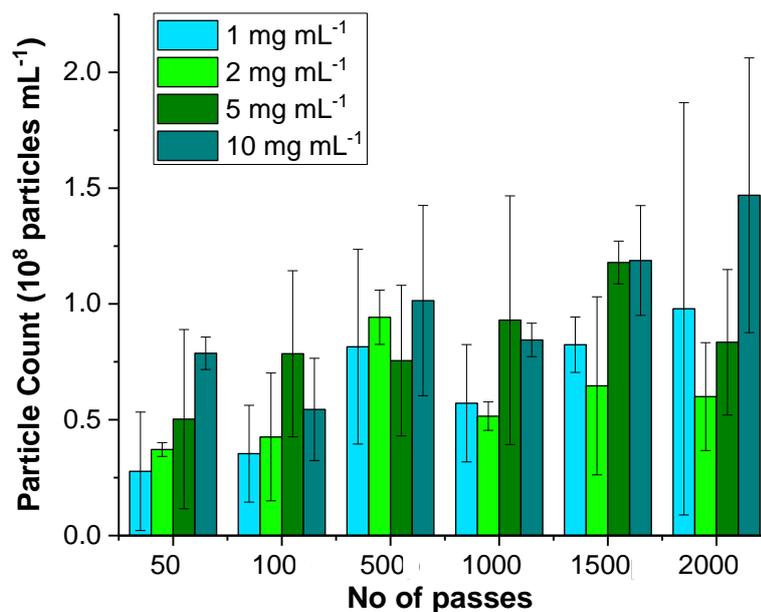


Figure 3.12. Total particle count for the stressed BSA samples detected by NTA in Figures 3.10 a–d. Errors from two independent experiments were propagated according to Section 2.2.18.

The data show that the total number of particles quantified by NTA, following stress in the EFD, ranges from 0.25– 2 x 10⁸ particles mL⁻¹. Whilst the average number (amount) of aggregates detected appears to increase as a function of both pass number and concentration, the dispersity of the samples (hence width of the error bars) makes a firm trend difficult to conclude (see Section 3.5).

Whilst NTA did reveal more information about the aggregates present in the stressed samples, it too was unable to detect aggregates in the quiescent samples and those stressed up to 20 passes, regardless of concentration. To detect whether any aggregates were present during these early events, FCS was exploited, based on its increased sensitivity (Tipping et al., 2015) and its previous use in monitoring BSA unfolding and aggregation in different chemical environments (Pabbathi et al., 2013; Zhang et al., 2015a). BSA fluorescently labelled with Alexa-488 BSA was diluted into unlabelled BSA samples (Section 2.2.9) before stress and the samples measured. The R_H values obtained from the correlation function fitting data are shown in Figure 3.13.

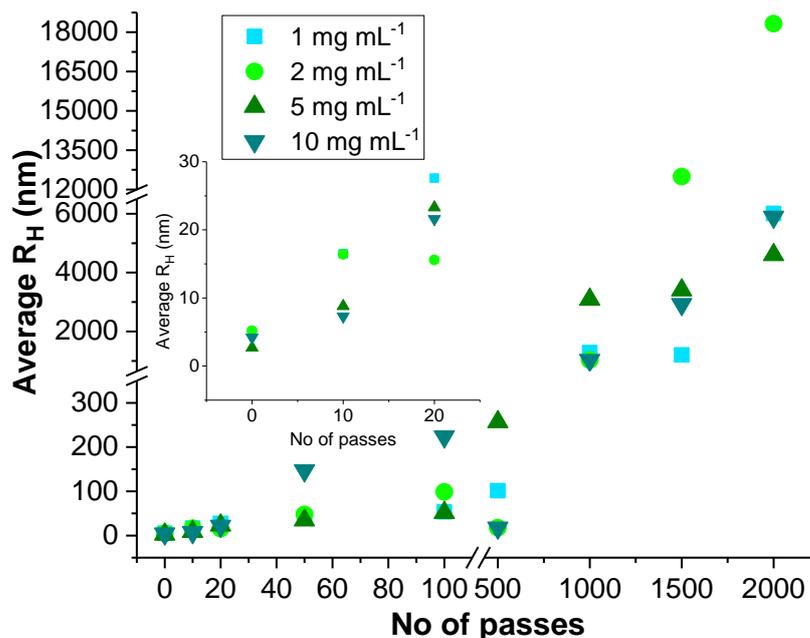


Figure 3.13. Plot of the average R_H values obtained from fitting FCS data for quiescent (0 passes) and stressed (10 – 2000 passes) BSA solutions.

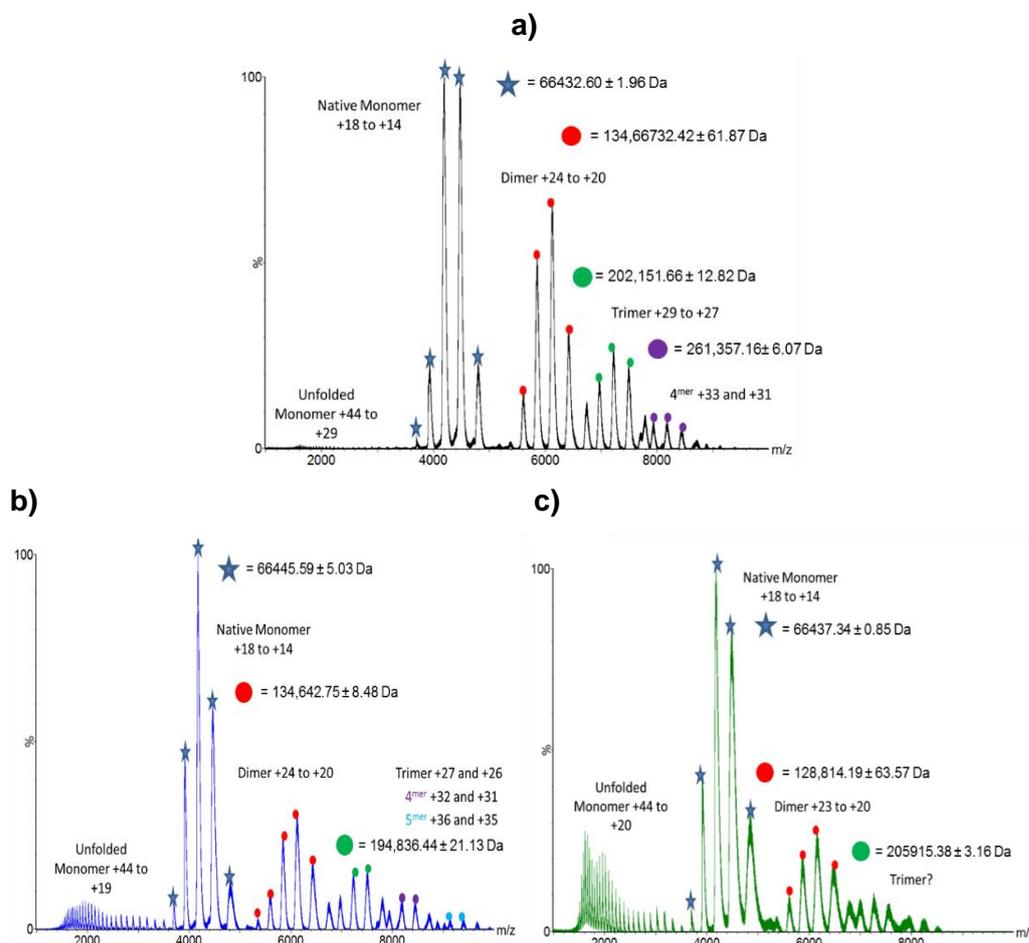
Samples were all stressed at a plunger velocity of 8 mm s^{-1} , at the concentrations indicated, prior to analysis. FCS measurements were performed by Dr Amit Kumar, with analysis performed by Dr Roman Tuma (both University of Leeds).

The data show that after 10 passes, there is a subtle increase in the average R_H of the labelled diffusing species in the sample, increasing from 2.7–5.2 nm (the quiescent sample) to 7.3–16.5 nm. Once again, the most pronounced increase in R_H is observed across all concentrations after 50 passes of stress has been experienced by the samples. After 1000 passes, the aggregates which can still diffuse are several microns in size, in agreement with the TEM and DLS data.

As these values derive from a complex fitting procedure, as well as the averaging of many scans to produce the final result, inferring the stoichiometry of any oligomers which form *en route* to the soluble aggregates and then insoluble species, is difficult. Ammonium acetate was selected as a buffer to facilitate the use of mass spectrometry to attempt to determine such stoichiometry, as has been done for a variety of aggregating protein systems (Knight et al., 2018; Ruotolo et al., 2008; Smith et al., 2010; Young et al., 2015).

BSA samples, subjected to stress or incubated quiescently, were thus subjected to analysis by native mass spectrometry. Initial attempts to do this with unclarified samples were rendered difficult by the presence of broad, overlapping peaks

(data not shown). Samples were thus clarified by centrifugation for 10 minutes at 8,000 g , (except the quiescent sample), before diluting samples down to 10 μ M in the same ratio as the native protein. Spectra for BSA at 5 mg mL⁻¹, stressed for 0, 20 and 100 passes, are shown in Figure 3.14.



Figures 3.14. Native mass spectra for a) quiescent BSA (5 mg mL⁻¹) b) after 20 passes and c) after 100 passes. Spectra were obtained using a 1.2 kV capillary voltage with sample concentration \sim 10 μ M. $N = 1$ for these samples. Masses assigned following processing in MassLynx are displayed next to the peaks as appropriate.

These data seem to show that within the quiescent BSA sample, there are some oligomers which cannot be detected after purification (by SDS-PAGE), nor by any of the other analytical methods discussed above. The high sensitivity of mass spectrometry (Section 1.3.2) may explain why these can be observed. The general population of oligomers (dimer to pentamer) seems to remain the same after 20 passes. Following stress in the EFD, the quality of the spectra obtained, in addition to the amount of unfolded protein observed increases, despite being

introduced to the gas phase under the same conditions as in 3.14a. Possible reasons for this are discussed in Section 3.5. In summary, mass spectrometry did not detect an increase in the amount of oligomer in the sample as a result of stress due to extensional flow. However, deviation was observed from the native sample's spectrum after 20 passes of stress. How therapeutic proteins fare when subjected to extensional flow forms the subject of the next section.

3.3 Assessing the aggregation of therapeutic antibodies under flow

Following the set of experiments performed on BSA in Section 3.2, the effect on extensional flow on other proteins were examined in order to meet the objectives set out in Sections 1.5 and 1.6. Three IgG1 monoclonal antibodies were provided by MedImmune which displayed wide-ranging biophysical properties. MEDI1912 (WFL herein) is a potent inhibitor of nerve growth factor (NGF), binding to its target with a pM affinity. However, it was found to have generally undesirable biophysical properties, such as interacting with chromatography matrices and being cleared rapidly *in vivo* (Dobson et al., 2016). Closer biophysical characterisation, using the SAP algorithm (Chennamsetty et al., 2009) (Section 1.3.1) and HDX-MS (Section 1.3.3), revealed that a solvent-exposed hydrophobic patch was probably the cause of such aberrant biophysical behaviour (Dobson et al., 2016). Rational mutations (W30S, F31T and L56T; all in the V_H domains) were performed on WFL to create MEDI1912_STT (STT herein), abrogating the aggregation-propensity of the protein whilst maintaining its binding affinity for the NGF target. In addition to these highly homologous proteins, a generic IgG1 called Nip109 (mAb1 herein) was studied. This is used by MedImmune as an internal standard, as it has generally favourable biophysical properties and has no known target.

WFL and STT show signs of aggregation (WFL more so) at a concentration ≥ 1 mg mL⁻¹ (Devine, 2016; Dobson et al., 2016), thus the mAbs were all stressed at a concentration of 0.5 mg mL⁻¹ in 150 mM ammonium acetate, pH 6.0 (see Methods Section 2.2.2). Rather than perform an exhaustive set of experiments on these proteins at this stage, it was decided that the samples would be stressed for 20 and 100 passes at a plunger velocity of 8 mm s⁻¹. Following this, DLS, NTA, TEM and the pelleting assay would be used to analyse the samples. Data from the first three techniques above for WFL, STT and mAb1 are shown in Figures 3.15, 3.16 and 3.17, respectively.

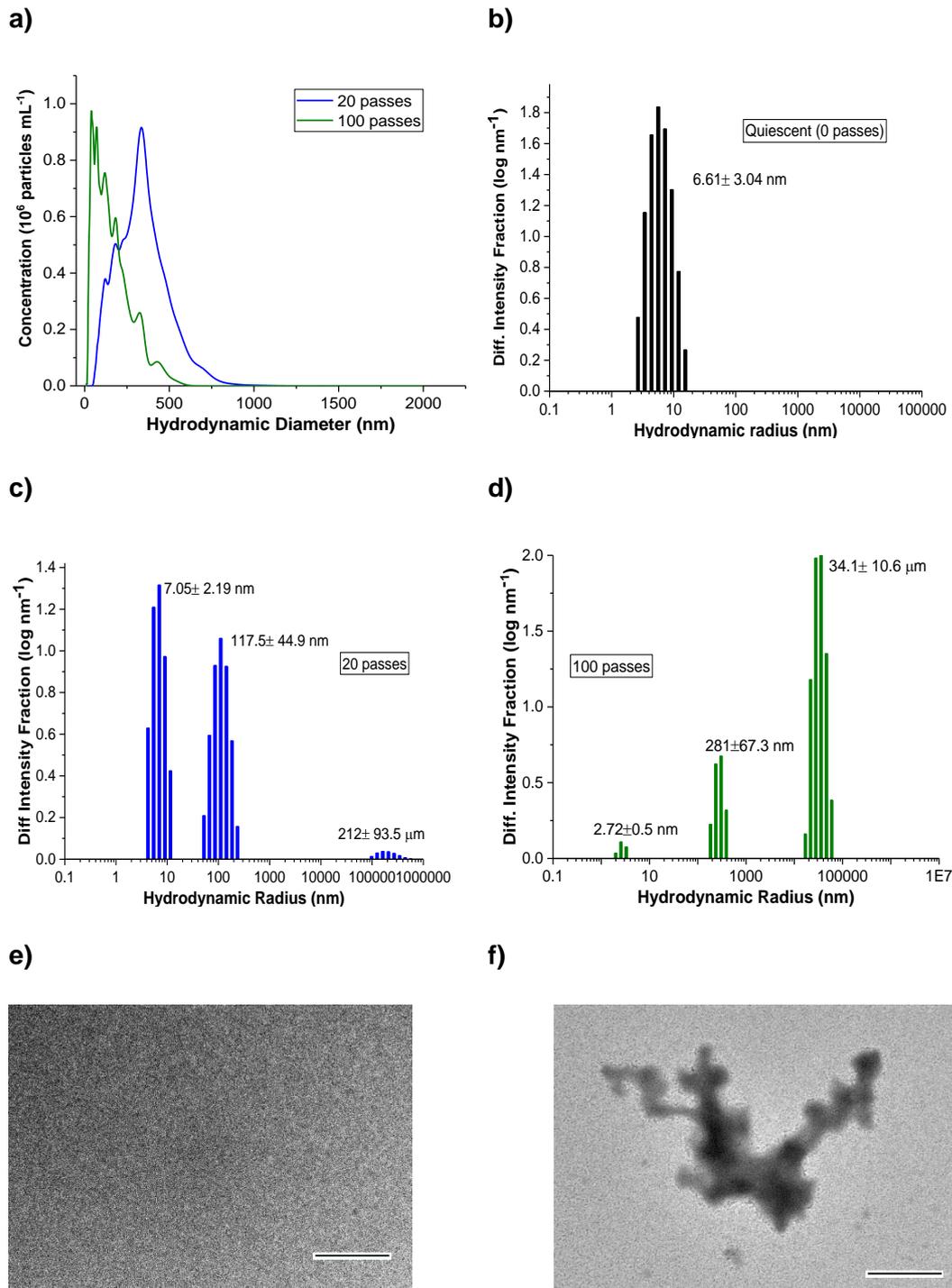


Figure 3.15. Biophysical characterisation of WFL following stress in the EFD at a concentration of 0.5 mg mL^{-1} at a plunger velocity of 8 mm s^{-1} . a) NTA data for WFL after 20 and 100 passes. No particles were present in the quiescent sample. b–d) DLS Regularisation plots for quiescent WFL, 20 passes and 100 passes respectively. e & f) TEM images (taken at 10,000x magnification) for quiescent WFL and after 100 passes respectively. Scale bar = 500 nm.

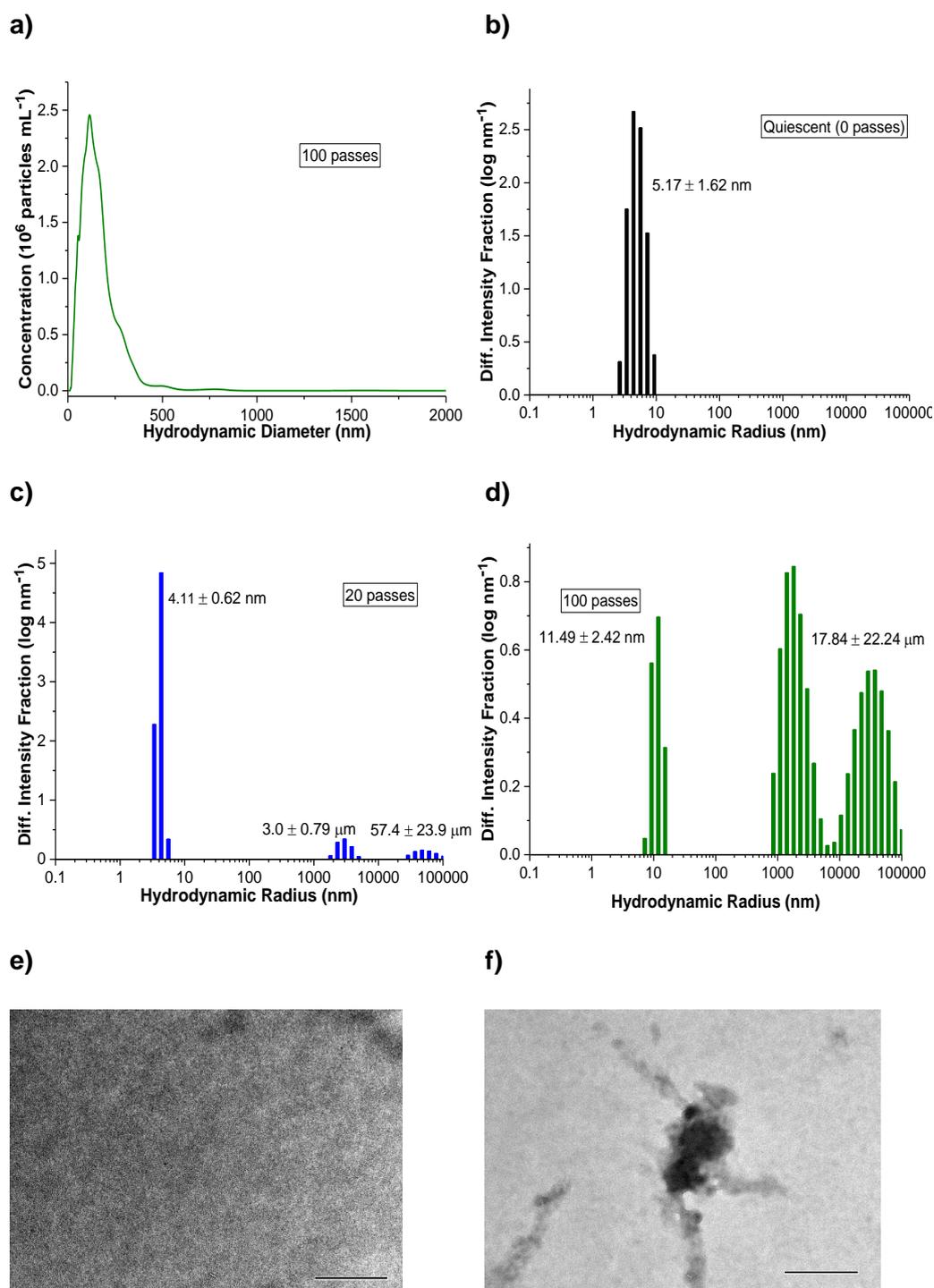


Figure 3.16. Biophysical characterisation of STT following stress in the EFD at a concentration of 0.5 mg mL^{-1} at a plunger velocity of 8 mm s^{-1} . a) NTA data for STT after 100 passes. No particles were present in the quiescent and 20 passes samples. b–d) DLS Regularisation plots for quiescent STT, 20 passes and 100 passes respectively. e & f) TEM images (taken at 10,000x magnification) for quiescent STT and after 100 passes respectively. Scale bar = 500 nm.

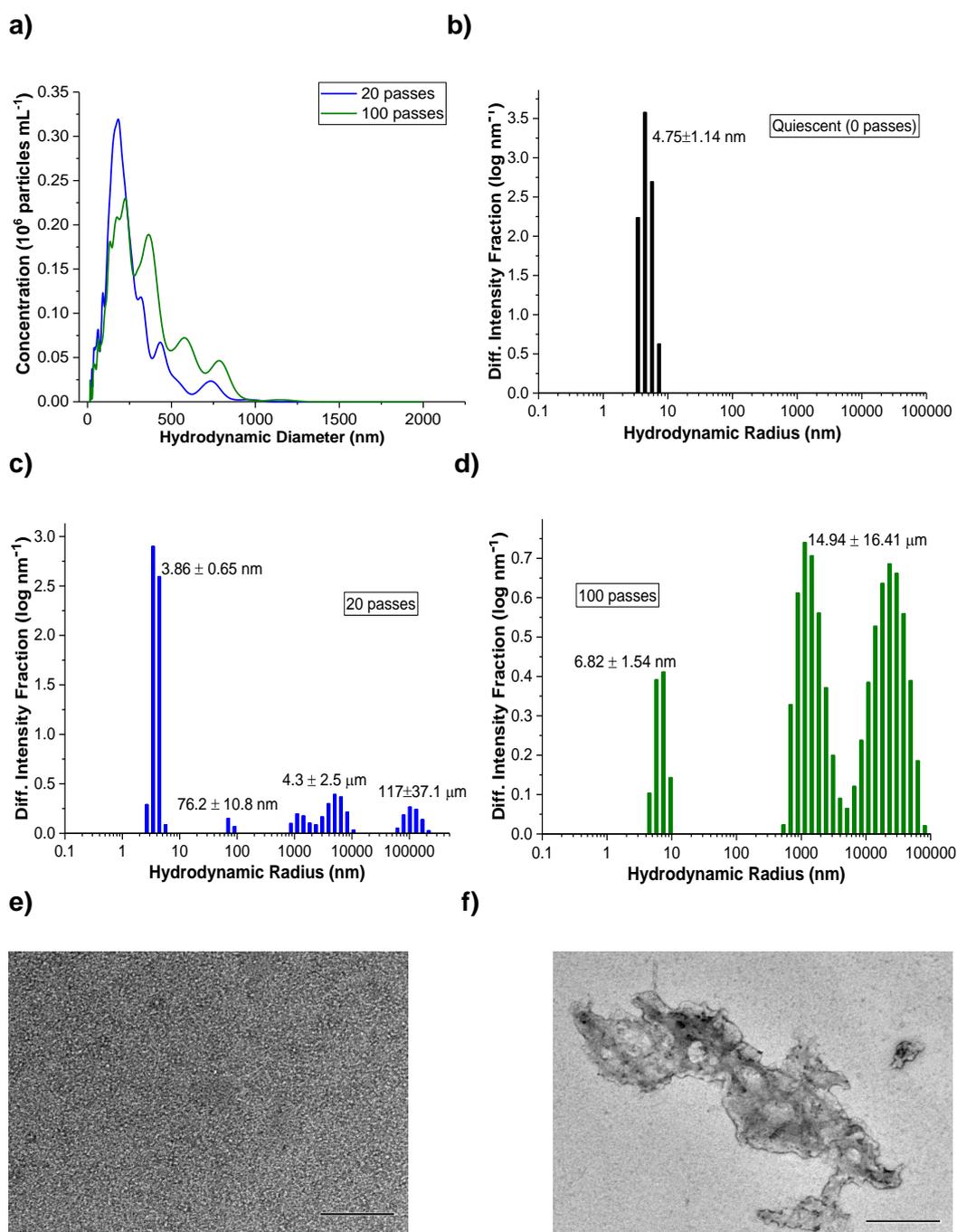


Figure 3.17. Biophysical characterisation of mAb1 following stress in the EFD at a concentration of 0.5 mg mL^{-1} at a plunger velocity of 8 mm s^{-1} . a) NTA data for mAb1 after 20 and 100 passes. No particles were visible in the quiescent samples. b–d) DLS Regularisation plots for Native STT, 20 passes and 100 passes respectively. e & f) TEM images (taken at 10,000x magnification) for quiescent STT and after 100 passes respectively. Scale bar = 500 nm.

The data show that all three mAbs aggregate as a result of being subjected to stress in the extensional flow device. Though WFL and STT differ by just six residues (three per V_H domain), they exhibit distinct aggregation behaviour. WFL begins to aggregate after just 20 passes (Figure 3.15), whereas STT requires 100 passes of stress to begin to form aggregates under the same flow conditions, as evidenced by NTA and DLS (Figure 3.16). No aggregates were present in either quiescent sample, nor the STT sample after 20 passes by NTA. mAb1 appears to have aggregation characteristics akin to WFL's, beginning to aggregate after 20 passes (Figure 3.17). All three mAbs form amorphous aggregates in the EFD, as seen in the TEM images in Figures 3.15 – 3.17. However, performing a ranking of a protein's aggregation propensity from these semi-quantitative data is difficult. Therefore, BSA and all three mAbs were stressed for 20 – 100 passes at the same plunger velocity of 8 mm s^{-1} , then analysed using the pelleting assay. The results are displayed in Figure 3.18.

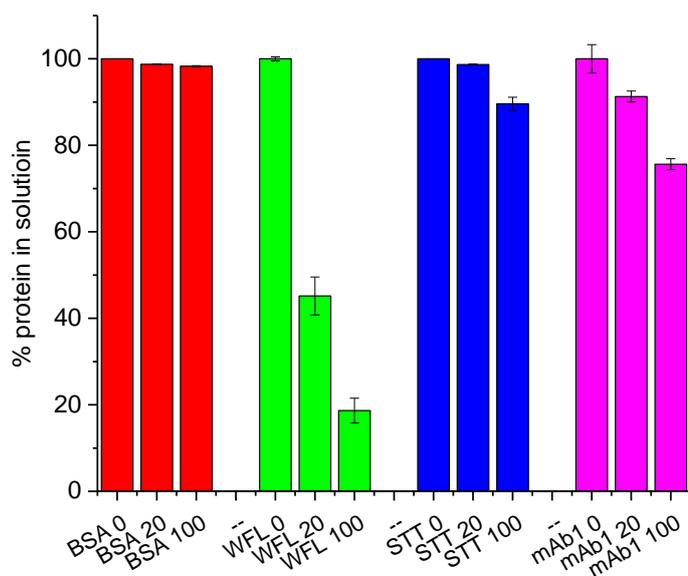


Figure 3.18 Plot of the amount of protein left in solution following 0 (quiescent), 20 or 100 passes of stress in the flow device, then the samples analysed with the insoluble protein pelleting assay. BSA was stressed at a concentration of 5 mg mL^{-1} , whilst the mAbs were stressed at a concentration of 0.5 mg mL^{-1} . The plunger velocity in all cases was 8 mm s^{-1} . The data were normalised to the quiescent samples, which showed no/minimal aggregation ($\sim 100\%$ in solution). Errors from two independent experiments were propagated according to Section 2.2.18.

The data show that BSA appears less sensitive to extensional flow compared to mAbs, as minimal aggregation (< 5% protein in pellet) is observed under the same flow conditions as the mAbs. After 100 passes, little soluble WFL remains (~15%), whereas only 15% of STT is rendered insoluble under the same flow conditions. mAb1 exhibits intermediate behaviour between these two homologous proteins, with around 25% of the protein going into the pellet fraction following 100 passes of stress. Importantly, none of the proteins above showed signs of aggregation when the quiescent samples were analysed by any of the techniques above after 10 minutes incubation under ambient conditions. The fundamental mechanisms behind *why* proteins aggregate in the EFD form the subject of the next section.

3.4 Determining the key drivers of aggregation in the extensional flow device

From the data generated and analysed in Sections 3.2 and 3.3, it is apparent that after exposure to extensional flow, protein aggregation takes place. The extent of aggregation depends on the protein in question, with the heavily disulphide linked protein BSA being much more recalcitrant to the forces found in the EFD than the three mAbs. As predicted, WFL exhibited more pronounced aggregation than its mutant STT, with mAb1 being somewhere in between. Whilst these data clearly show that the EFD can induce aggregation, the mechanism by which it does so was unclear. For example, can the shear flow experienced in the capillary trigger aggregation? How do different plunger velocities, *ergo*, different strain and shear rates affect the observed levels of protein aggregation. Finally, can flow induce the conformational remodelling of proteins?

With these questions in mind, alternative experiments were performed to establish which factors were key to driving aggregation, particularly for BSA given the extensive dataset amassed for this model protein both here and elsewhere, for example Ashton et al., 2010; Bekard et al., 2012; Simon et al., 2011.

3.4.1. Dissecting the role of shear in BSA's aggregation

As stated in Section 3.1, the EFD consists of an extensional flow region (the point of contraction between the syringe and capillary) and a high shear region (along the length of the capillary). To assess the role of shear on the aggregation of BSA, a 5 mg mL^{-1} BSA solution was stressed for 1000 passes at a plunger velocity of 8 mm s^{-1} , using a 37.5 mm capillary instead of 75 mm . This halves the length of time the BSA is exposed to the $52,000 \text{ s}^{-1}$ shear rate in the capillary region. Aggregation was then characterised using the insoluble protein pelleting assay. These data are shown in Figure 3.19.

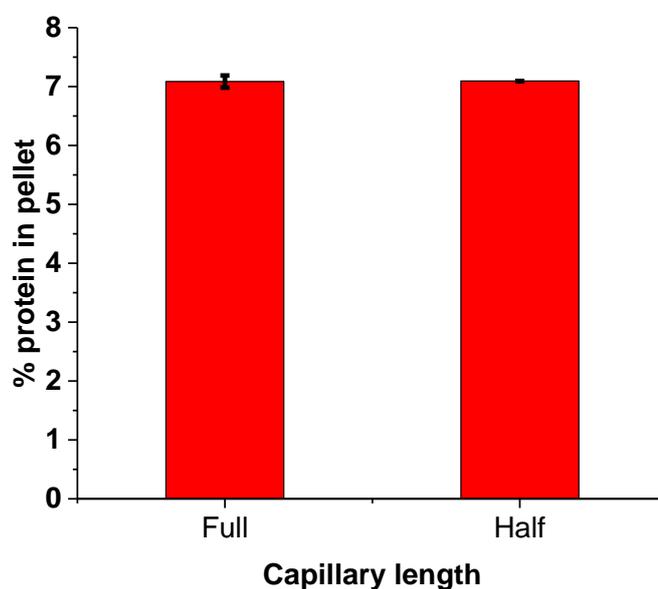


Figure 3.19. Plot of the amount of excess protein in the pellet fraction for 5 mg mL^{-1} BSA following 1000 passes of stress at 8 mm s^{-1} . The full-length capillary = 75 mm long, whereas the half-length = 37.5 mm . Errors from two independent experiments were propagated according to Section 2.2.18.

The data show that when the length of the capillary decreases, the amount of insoluble material formed in the device does not decrease, with $\sim 7\%$ of the material insoluble, regardless of capillary length. Once again, no aggregation was present in the quiescent samples (not shown). This result shows that, for BSA, the high shear flow that the protein is exposed to in the capillary does not dominate the amount of aggregate formed in the device. Instead, it suggests extensional flow dominates BSA's aggregation.

3.4.2. Looking for evidence of particle shedding from the device

As a control to check that none of the aggregates observed by light scattering were a result of shedding of Teflon, glass etc. in the EFD, 25 mM ammonium acetate, pH 5.1, was stressed for 1000 passes at a plunger velocity of 8 mm s^{-1} , then analysed using DLS and NTA. The results are shown in Figure 3.20.

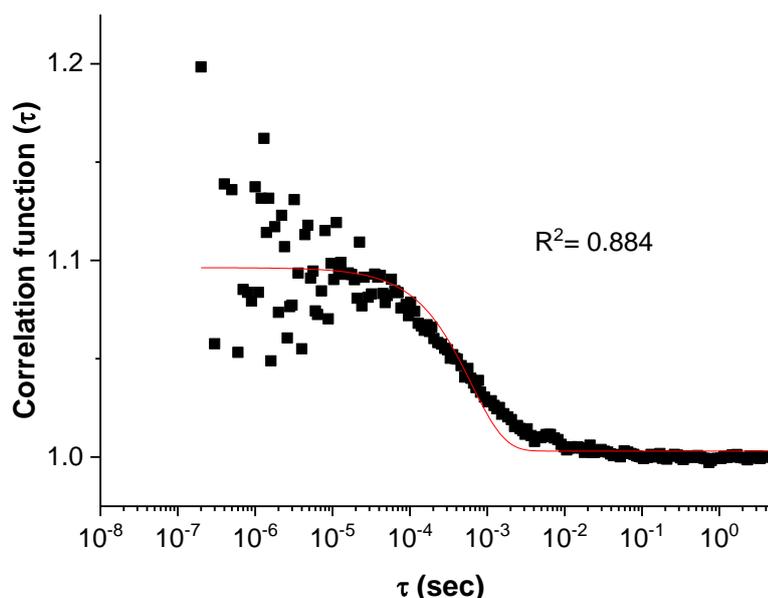


Figure 3.20. Correlation function plot obtained by DLS analysis of 500 μL of 25 mM ammonium acetate, pH 5.1 following 1000 passes of stress in the EFD (plunger velocity = 8 mm s^{-1}). The R^2 value reported is for a single exponential decay fit according to Section 2.2.4.

The data show that no particles are shed from the device following stress. The NTA videos were completely blank, with no particles tracked in any of the videos. The poor fit of a single-exponential decay to the DLS data in Figure 3.20 is indicative of no particles being present in the (very poor signal:noise), with small air bubbles being the likely contributor to any signal present (*cf.* scale in Figure 3.7 for proteinaceous samples). These data suggest that the light scattering data from the collective body of samples measured following stress in the EFD (several hundred throughout the thesis) belong predominantly to protein signal, as opposed to adventitious contaminants.

3.4.3 Conformation re-modelling of BSA under extensional flow

As stated in Section 1.4, many previous studies on the effects of flow on the conformational properties of biomolecules have gathered experimental evidence for such phenomena through means of *in situ* analytics, for example Ashton et al., 2010; Bekard et al., 2012; Smith et al., 1999. The design of the EFD precludes the incorporation of such analytics to probe conformational change at the molecular level. Therefore, an alternative strategy was sought from the literature. IAEDANS has been used to perform cysteine ‘shotgun sequencing’ on red blood cells, probing the dynamics of proteins in given cellular states (Krieger et al., 2011). BSA has one buried cysteine residue (Cys 34), with its other 34 are oxidised as cystines. If extensional flow conformationally remodelled the protein, it was hypothesised that this residue may be labelled with the fluorophore, if the dye and protein were stressed together in the EFD. Unless Cys 34 is naturally exposed due to localised dynamics, the quiescent sample should not label under the same chemical conditions.

To validate this hypothesis, BSA was stressed for 0–100 passes at a plunger velocity of 8 mm s^{-1} in the presence of 5 mM IAEDANS (see Section 2.2.10). Furthermore, to assess whether the 17 disulphide bridges of BSA are made available for labelling as a result of flow, the reducing agent TCEP was also added (0.5 mM final) in some experiments (see Section 2.2.10). These data are shown in Figure 3.21a. In all cases, following stress, samples were analysed using an SDS-PAGE gel (see Methods Section 2.2.2). In another round of experiments, the IAEDANS (with/without TCEP) was added to the BSA solutions *after* stress, to assay the longevity of exposed cysteine residues post-extensional flow. These data are shown in Figure 3.21b.

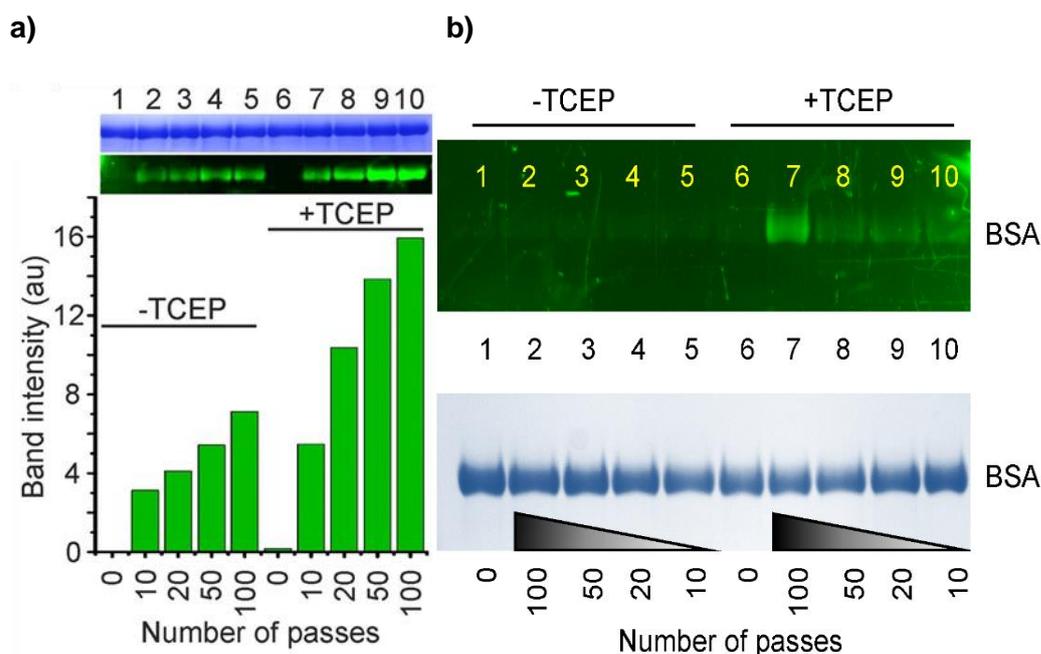


Figure 3.21. SDS-PAGE analyses of 5 mg mL⁻¹ BSA, incubated with 5 mM IAEDANS and/or 0.5 mM TCEP, in the presence and absence of flow. a) Coomassie Brilliant Blue staining (upper) and fluorescence visualisation (lower) and quantitation of BSA labelled with IAEDANS and separated on a 12% (w/v) SDS-PAGE gel. These samples were prepared with the chemical additives *in situ*. **b)** As above, except the chemical additives were added *ex situ*. Lane 1 = Quiescent BSA incubated with IAEDANS. Lane 6 = Quiescent BSA incubated with both TCEP and IAEDANS. In all other lanes, samples were stressed for the number of passes indicated at a plunger velocity of 8 mm s⁻¹. Further detail is provided in Section 2.2.10. Gel electrophoresis, imaging and analysis performed by Dr Amit Kumar (University of Leeds).

The experiments show that in the absence of reducing agent, the greater number of passes experienced by the BSA solution increases the labelling that occurs. In the presence of reducing agent (TCEP), labelling is much more pronounced. The quiescent sample does not label in the absence of flow, with minimal labelling observed in the presence of TCEP (Figure 3.21a). Conversely, when the dye/reducing agent mixture is added to the BSA post-stress, labelling is only observed in the 100 passes sample with TCEP (Figure 3.21b). This could be due to the TCEP's ability to access some Cys residues in the aggregates which form after 100 passes in the device (Sections 3.4.2 and 3.4.3). In summary, these experiments demonstrate that extensional flow can indeed induce the conformational re-modelling of BSA, which then leads to aggregation of the resultant, partially unfolded BSA molecules.

3.4.4 Assessing the role of energy in the EFD and its effects on BSA's aggregation

Jaspe and Hagen produced a 'beads on a string' model to quantify the shear rate required to unfold a globular protein (Jaspe and Hagen, 2006). This model, which assumes the protein under study unfolds *via* a 'dumbbell'-shaped intermediate, was used by John Dobson (University of Leeds) to assess whether or not flow causes partial or complete unfolding of proteins by comparing the forces experienced by a protein in the EFD to those calculated in the model. The model is summarised in Figure 3.22, with the results discussed afterwards.

$$W = \left(\frac{27}{28}\right) \pi \eta d^2 \gamma N^{\frac{7}{3}} \left(\frac{3v_b}{8\pi}\right)^{\frac{1}{3}}$$

(Equation 3.1)

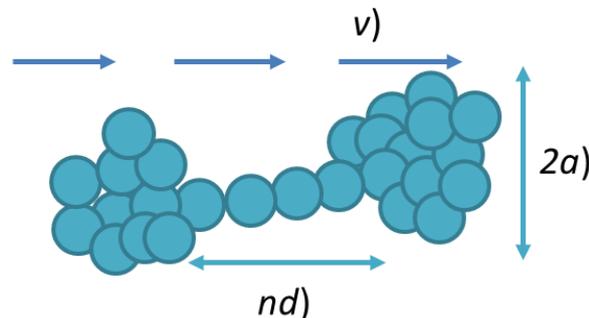


Figure 3.22. Jaspe and Hagen elementary model for proteins unfolding due to flow. These terms were computed by Dr John Dobson, University of Leeds.

The terms in Equation 3.1 are as follows: W = work required to unfold the protein, η = viscosity of the fluid, d = distance between the beads on the string, γ = shear rate, $N = 573$ residues (total in BSA), v_b = volume of one bead (i.e. one amino acid), v = extensional flow vector, a = area of ball of unfolded protein and n = number of residues in the linker.

At the shear rates and strain rates present in the EFD, the Jaspe and Hagen calculations suggest that proteins are likely to only experience forces of ~10 fN. Using the CFD modelling, the global energy in the flow device was calculated by John Dobson, assuming BSA-sized spheres of fluid with diameters = 3.5 nm. These data were then scaled as a rate of energy transfer to the packet of fluid. The data are shown in Figure 3.23.

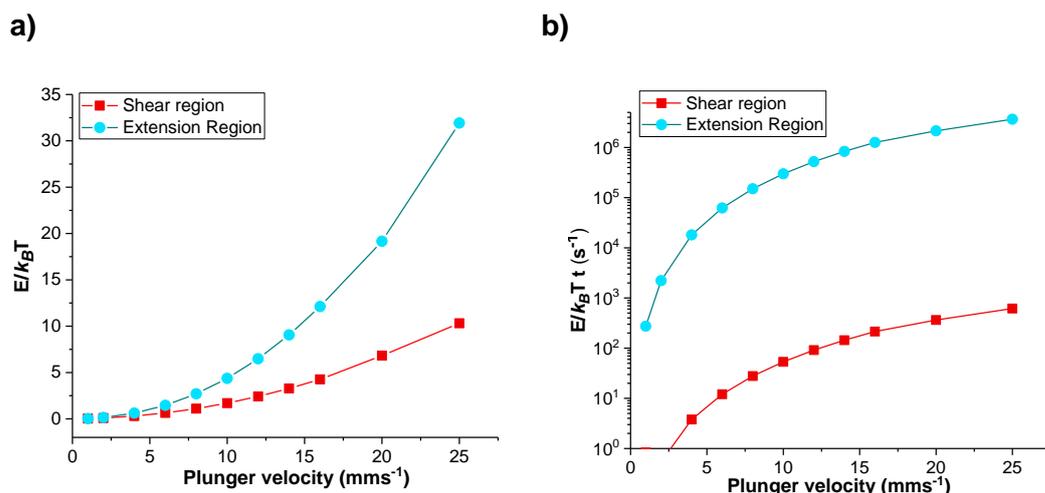


Figure 3.23. Distributions and dissipation of energy in the EFD. a) Energy in the extension (blue) and shear regions (red) of the flow device as a function of plunger velocity. **b)** Rate of energy transfer into the fluid in the extension and shear regions as a function of plunger velocity. The CFD was performed by John Dobson, University of Leeds, with packets of fluid of 3.5 nm in size.

The data in Figure 3.23a show that at plunger velocities $< 10 \text{ mm s}^{-1}$, the thermal energy delivered to the fluid in the shear and extension regions is low (~ 1.1 and $2.7 E/k_B T$ in the shear and extension regions respectively). On a rate scale (Figure 3.23b), because the extensional flow region is very short in the EFD (2 mm, Figure 3.2), energy is delivered rapidly compared to the shear region. As the plunger velocity and strain rate increase, there is an increase in both the extensional flow energy rate and that for shear. However, the magnitude of the extensional flow energy rate remains ~ 1000 -fold higher. As shown in Figure 3.3b, at higher plunger velocities, proteins are exposed to these damaging extensional flow forces for decreasing amounts of time. To see which of energy and time were key to driving BSA's aggregation, a 5 mg mL^{-1} solution of BSA was stressed for 100 passes at plunger velocities of 2-, 4- and 8–16 mm s^{-1} . The samples were then analysed using the pelleting assay. To validate that extension was driving aggregation compared to shear, the 'shear experiment' (i.e. the halving of the capillary length as in Section 3.4.1) was performed at the highest plunger velocity (16 mm s^{-1} , $\dot{\gamma} = 104,000 \text{ s}^{-1}$). The data for these experiments are shown in Figure 3.24.

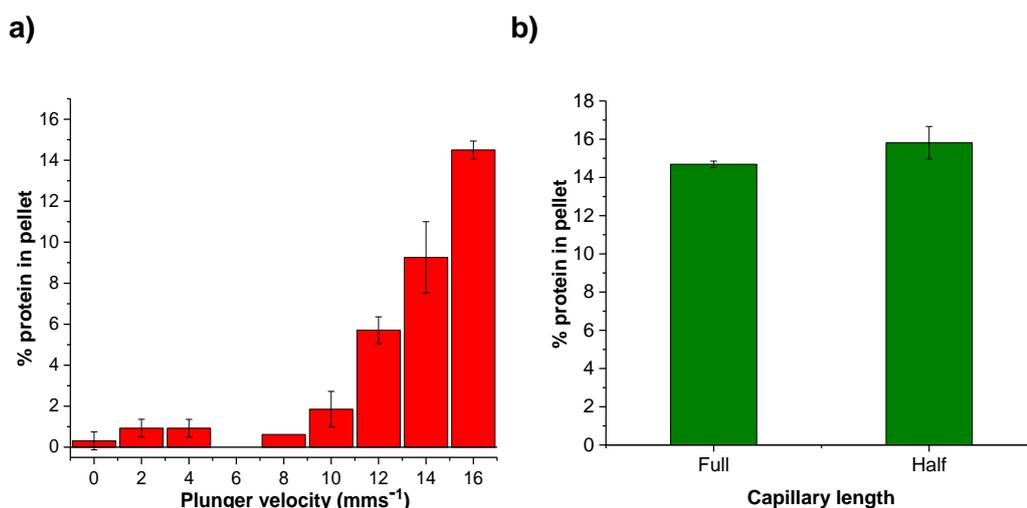


Figure 3.24. Quantification of the amount of insoluble BSA formed when stressed at different strain rates. a) Plot of percentage protein in the pellet fraction after stressing 5 mg mL⁻¹ BSA for 100 passes at the stated plunger velocities. The strain and shear rates are in Figure 3.3. b) Plot of excess BSA in the pellet fraction of 5 mg mL⁻¹ stressed for 100 passes at a plunger velocity of 16 mm s⁻¹ with a full-(75 mm) or half-length (37.5 mm) capillary. Errors were propagated according to Section 2.2.18.

The data show that below a plunger velocity of 12 mm s⁻¹, significant levels of aggregation (i.e. above the noise of the experiment) are not observed (Figure 3.24a). NTA experiments were also performed at these lower velocities (2–6 mm s⁻¹), with no aggregates present in the samples. After 100 passes at 16 mm s⁻¹ ($\epsilon = 23,421 \text{ s}^{-1}$, $\gamma = 104,000 \text{ s}^{-1}$), ~15% of the protein has been lost from solution. In Figure 3.5, 1500 passes were necessary to induce this level of aggregation at a plunger velocity of 8 mm s⁻¹. Despite the very high shear rate in the capillary, reduced exposure of BSA to this flow field did not reduce the amount of aggregation observed in the shear experiment (Figure 3.24b). To summarise, BSA's aggregation is driven by extensional flow energy, with increased force at higher strain rates leading to more pronounced aggregation than that seen following the same number of passes at lower strain rates.

3.5 Discussion

As introduced in Sections 1.4 and 1.5, previous studies which investigated how hydrodynamic flow affects biomolecular conformation have focussed on shear flow, neglecting extensional flow. The device characterised in this section (and the rest of the thesis) subjects proteins to an extensional flow event, followed by high shear along the length of the capillary. Focussing on BSA as a test protein, it is evident that these flow fields have the ability to induce the unfolding and aggregation of proteins, as has been seen by other groups using unrelated devices (Ashton et al., 2010; Bekard et al., 2012; Simon et al., 2011). By using an array of experimental techniques, aggregation could be characterised from the monomeric BSA through to insoluble aggregates ($> 10 \mu\text{m}$).

The aggregates which form in the device cover a broad size range and are highly disperse. Once extensively stressed in the device, the PDI of the samples often exceeds 0.6, with this being a threshold of a highly disperse sample (Roger et al., 2016). The aggregate sizes reported by DLS in such samples will be adversely affected by such dispersity, which may explain the very wide widths observed in Cumulants analysis of the DLS data. The Regularisation algorithm can still detect the monomeric BSA in all of the stressed samples, partly because these species make up the majority of the samples (Fischer and Schmidt, 2016), as evidenced by the insoluble protein pelleting assay.

Whilst NTA allowed for better resolution of the aggregate species, as can be seen by the numerous peaks within a sample curve in Figure 3.11, it too has its problems: the monomer cannot be sized/detected in both stressed and control samples; the traces look 'flat' beyond 1000 nm in many samples, despite larger species being detected in other assays; the total particle count obtained from the averaged samples varies widely. For this last point, the viewing area in NTA is very small ($\sim 40 \text{ nL}$). To obtain a y-axis of $10^8 \text{ particles mL}^{-1}$, the particles counted by the software undergo an extrapolation of ~ 15 million (Zhou et al., 2015). If any particles settled during the analysis, they would not have diffused and may contribute to the poor precision of the NTA measurements.

The increased sensitivity of FCS allowed for the detection of aggregation in BSA after just 10 passes of stress in all samples. Whilst it was thought that the sensitivity of mass spectrometry would aid in the detection of early aggregation events, this was not the case; despite clarification with centrifugation, any residual aggregates impeded the transfer of proteins from the solution phase to the gas phase. The relatively large size of these complexes (a BSA tetramer has a mass ~ 266 kDa) also hindered this process (Ruotolo et al., 2008). The increased levels of unfolded protein observed in the stressed samples could be due to monomer ejection from a complex state, i.e. a partially-unfolded protein dissociates from a complex in the gas phase, hence higher charge states are observed (for the ejected monomer) (Utrecht et al., 2010). At pH 3, it is known that BSA undergoes a structural change to a partially unfolded 'molten globule' state (Bhattacharya et al., 2011). Performing MS on BSA at this pH does indeed lead to many highly charged ions being seen for the monomer in the gas phase (Appendix Figure A3), supporting this observation.

How elongated biomolecules (e.g. DNA, spider silk fibroins and vWF) unfold and undergo conformational change due to flow has been extensively studied (Perkins et al., 1997; Proudfoot et al., 2017; Rammensee et al., 2008; Zhang et al., 2009). It seems logical that these molecules would align in a flow field. The simulations performed by colleagues in the field verify these experimental observations (Sing and Alexander-Katz, 2010; Szymczak and Cieplak, 2007; Szymczak and Cieplak, 2011). How globular proteins might align under flow and respond to the strain imparted on them by the fluid has not been studied in detail other than by Jaspe and Hagen. Their model proposed that very high shear rates, despite the model focussing on strain (extensional flow) would be necessary to fully unfold a globular protein (Jaspe and Hagen, 2006).

The CFD and dumbbell model calculations suggest that proteins in the device are subjected to forces far below those required to fully unfold them (typically tens of pN) as shown from Atomic Force Microscopy (AFM) experiments (Brockwell 2007). Furthermore, the capillary length experiments show that BSA's flow-induced aggregation is not dominated by shear flow. Complete unfolding of proteins is not necessary to cause them to aggregate (Roberts, 2014). This partial unfolding event was captured *in vitro* using IAEDANS to capture the exposure of BSA's usually buried free cysteine residue. This strategy was used previously to capture the dynamics of erythrocyte proteins as the red blood cells undergo shear stress, with more pronounced labelling observed when the shear stress was

increased (Krieger et al., 2011). This agrees with the data in Figure 3.21, with the degree of labelling increasing as a function of pass number. Importantly, the quiescent sample did not label when incubated for 10 mins with the dye; local dynamics of BSA did not account for the labelling. When TCEP was added as a reducing agent, to probe BSA's disulphides during these experiments, labelling was indeed more pronounced. This only occurred in the presence of extensional flow, thus flow must have exposed these recalcitrant residues, which are then reduced, enabling them to react with the dye and form irreversible covalent carbon-sulphur bonds. It is known that reducing agents can influence the aggregation kinetics and conformation of BSA by perturbing the protein's disulphide bridge network (Borzova et al., 2015; David et al., 2008; Katchalski et al., 1957). In the experiments in Section 3.4.3, increased aggregation of BSA was observed in the presence of TCEP as quantified with the pelleting assay (Appendix A2). Together, these experiments provide the key experimental evidence for extensional flow having the ability to partially unfold and induce the aggregation of globular proteins.

Under the limited experimental conditions explored in this chapter, the three mAbs exhibited varied aggregation behaviour once they had been subjected to extensional flow. The largest contrast was seen between WFL and STT; despite differing by just six amino acids in total, WFL showed a ~six-fold increase in the amount of insoluble protein formed under the same flow conditions as its mutant. The 'generic' mAb1 was somewhere in between these two proteins in terms of aggregation behaviour. mAb1 was not as sensitive to extensional flow as WFL, with aggregation was observed after just 20 passes using the four main biophysical techniques from this section (DLS, NTA, TEM and the pelleting assay)- the same threshold as observed for WFL. A striking observation can be made about all three mAbs; their aggregation triggered by hydrodynamic forces was revealed at a concentration of just 0.5 mg mL^{-1} . In molar terms, this is 5x lower than the lowest BSA concentration used (1 mg mL^{-1} BSA $\sim 15 \text{ }\mu\text{M}$), yet a broad dynamic range in response to flow was observed. This could be due to the fact that larger proteins have generally been shown to be more sensitive to hydrodynamic forces than smaller proteins (Simon et al., 2011; Springer, 2016). However, a detailed model for how these proteins, of a completely different topology to BSA, unfold and aggregate under flow is currently missing.

In conclusion, the results presented in this chapter show that the EFD has the ability to induce the partial unfolding and subsequent aggregation of globular proteins. An array of quantitative and qualitative biophysical methods have been identified to characterise the aggregates formed in the device. The defined nature of the flow fields in the EFD, validated using computational fluid dynamics, allow for conclusions to be drawn between the energy imparted into the protein solution on the observed levels of aggregation. Together with the fact that antibodies are sensitive to the effects of flow (aggregation observed at low concentration), this opens up the opportunity to use the flow device to chart wide areas of parameter space. These experiments form the basis of the next chapter.

4. Mapping the aggregation landscapes of proteins under defined fluid fields

4.1. Determining the aggregation landscapes of BSA, WFL, mAb1 and STT

In Chapter 3, protein aggregation induced by extensional flow was characterised using an array of experimental techniques. Of those methods, the insoluble protein pelleting assay proved to be a robust, quantitative measure of the extent of aggregation for a sample under a given set of flow conditions. This assay showed that increasing the number of passes increased the extent of aggregation and that the strain rate and protein also affected aggregation. In addition to these experimentally investigated factors, the physicochemical environment (pH, protein concentration and ionic strength) along with shear flow may also affect aggregation. The parameter space discussed above is summarised in Figure 4.1.

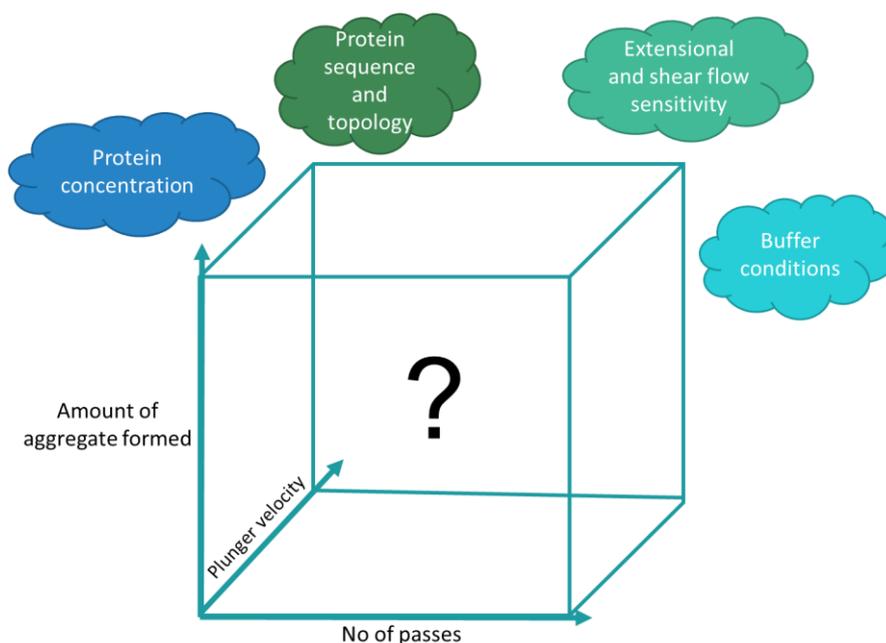


Figure 4.1. Overview of the parameter space that could potentially be explored with the flow device. The number of passes dictates the total time proteins in the EFD are exposed to hydrodynamic stress (extensional flow and high shear rates) in the EFD. The plunger velocity dictates not only how much force will be applied on a protein in the flow, but also how long that force will be applied for. At higher plunger velocities, proteins spend less time in the extension and shear regions and vice-versa. These are thought to be the key operational parameters which will affect the amount of aggregate formed, as measured by the pelleting assay. Plotting the amount of protein aggregate formed when both passes and velocity are varied will produce an ‘aggregation landscape’ or response surface for a given protein. The ‘clouds’ which hang over the landscape, namely: the concentration, sequence and topology of the protein, the buffer components and the sensitivity of the protein to flow. All of these may modulate the landscape.

To reduce the number of factors from Figure 4.1 needed to produce an aggregation landscape, the parameters in the clouds (concentration, buffer composition and capillary length) were initially fixed for each protein. Therefore, BSA and the two contrasting mAbs (WFL and STT) were stressed for different numbers of passes (0–200), at different plunger velocities (2–16 mm s⁻¹), in the EFD (Section 2.2.3). The amount of insoluble protein was then quantified with the ultracentrifugation pelleting assay (Section 2.2.7). The data were then plotted as three-dimensional surfaces. These are shown in Figure 4.2 a–c respectively.

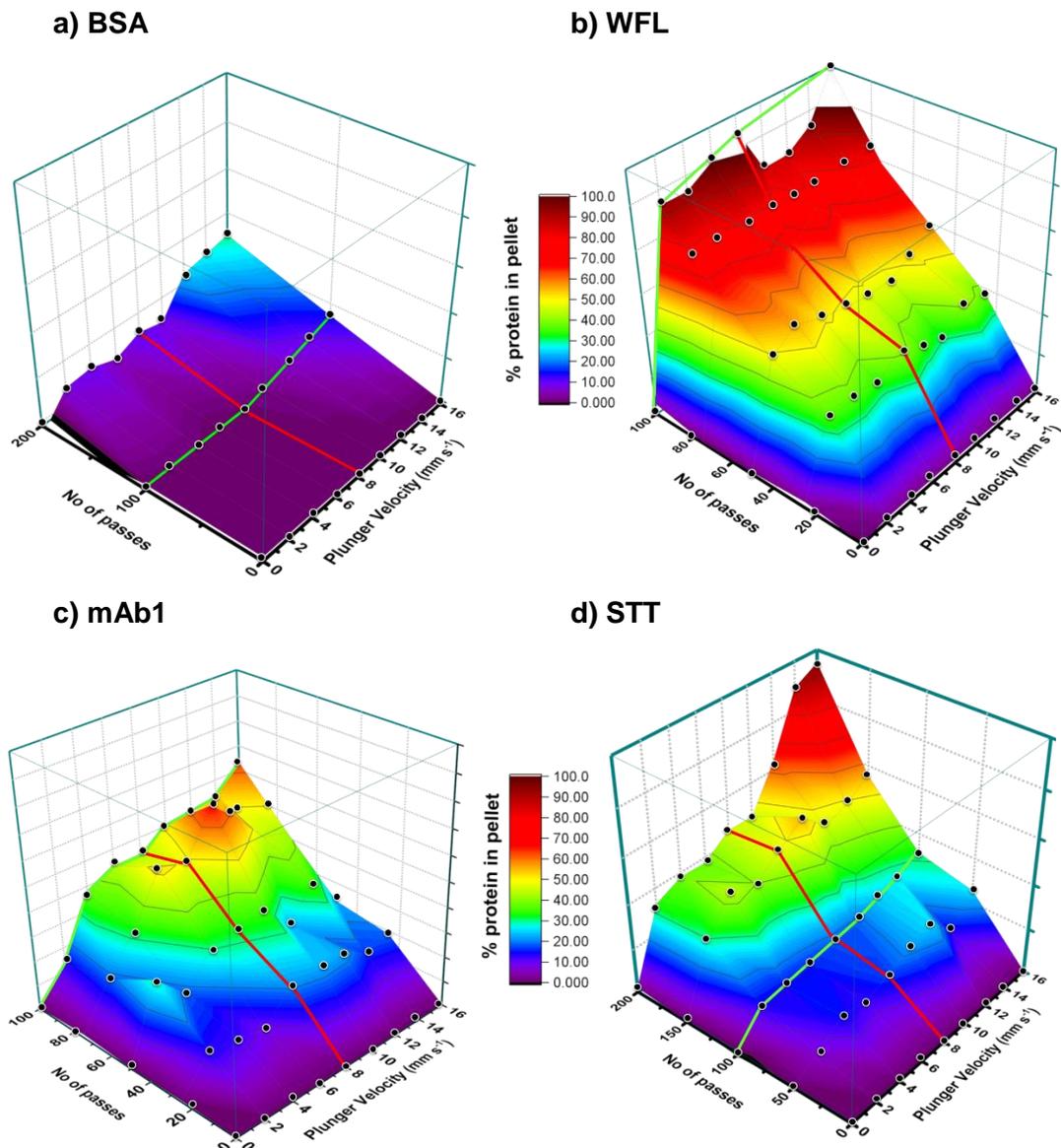


Figure 4.2. Aggregation landscapes for a) BSA, b) WFL c) mAb1 and d) STT. Each three-dimensional aggregation landscape shows the amount of insoluble protein following stress in the EFD for the stated number of passes, at the plunger velocities indicated. a) BSA was stressed at a concentration of 5 mg mL⁻¹ in 25 mM ammonium acetate, pH 5.1. The data for 100 passes are from Figure 3.22a. b) WFL, c) mAb1 and d) STT were each stressed at a concentration of 0.5 mg mL⁻¹ in 150 mM ammonium acetate, pH 6.0. Data points shown are an average of two independent experiments. Green lines guide the eye to a protein's response to different strain rates after 100 passes, whilst red lines guide the eye to pass-dependence at 8 mm s⁻¹. Data in a, b and d) were collected alongside Dr Amit Kumar, with those in c) collected alongside Ms Elizabeth Allen, (both University of Leeds).

The data in Figures 4.2 a–d show that each protein has a unique aggregation landscape when stressed under the stated flow conditions. BSA, which was characterised in Chapter 3, has a relatively flat aggregation landscape (Figure 4.2a). It is clear that, for BSA, when the number of passes doubles from 100 to 200, the amount of aggregate in the pellet fraction also doubles, but only above a strain rate threshold of $14,634 \text{ s}^{-1}$ (equivalent to a plunger velocity of 10 mm s^{-1} , Section 3.4.4). Below this threshold, the landscape has low levels of aggregation (<10% protein in pellet).

Figure 4.2b shows that WFL exhibits a strain-independent, pass-dependent response to aggregation in the flow device. Strain rates as low as $3,184 \text{ s}^{-1}$ can 'activate' WFL molecules for aggregation, with the aggregation reaction reaching completion (80–100% insoluble protein) after 100 passes.

Figure 4.2c shows that mAb1 demonstrates a linear response to increased pass number, which is similar to WFL. However, the extent of aggregation after 100 passes is around half that of WFL. When mAb1 is stressed in a strain-rate dependent fashion, the amount of aggregate observed is reduced at a strain rate of $3,184 \text{ s}^{-1}$ (equivalent to a plunger velocity of 2 mm s^{-1}). Above this threshold, the aggregation of mAb1 is largely independent of strain rate, except when the protein is stressed above a strain rate of $14,634 \text{ s}^{-1}$. Together, these trends show that whilst mAb1 follows similar aggregation trends to WFL, it is not as sensitive to extensional flow as this highly aggregation-prone protein.

STT, on the other hand, possesses a much richer aggregation landscape (Figure 4.2d) to BSA, WFL and mAb1. STT responds to flow in a similar fashion to BSA when stressed for 100 passes or less, with increased aggregation seen when the plunger velocity exceeds 10 mm s^{-1} . When the number of passes doubles from 100 to 200, various levels of aggregation are observed. The 'low aggregation plateau' (plunger velocities of $2\text{--}8 \text{ mm s}^{-1}$) has a shallow gradient, with approximately double the amount of aggregated protein after 200 passes compared to after 100. On the other hand, higher plunger velocities appear to greatly damage STT, with 95% aggregation observed after 200 passes at 16 mm s^{-1} . Whilst differences in flow sensitivity for each protein could be observed in Figure 3.18 (Chapter 3), the landscape analyses reveal how passes and strain affect the observed levels of aggregation following stress in the EFD.

The data for BSA suggests that proteins have a threshold for strain rate below which the protein does not aggregate under flow, i.e. the force per unit time applied by the fluid is not sufficient to conformationally alter the protein. To investigate whether mAbs exhibited similar behaviour, WFL and STT were stressed for 20 and 50 passes, respectively, at a plunger velocity of 0.5 mm s^{-1} ($\dot{\epsilon} = 871 \text{ s}^{-1}$, $\dot{\gamma} = 3,149 \text{ s}^{-1}$). These pass numbers were chosen as they represent the minimum number of passes where aggregation was observed in each protein's respective aggregation landscape. The samples were then analysed with the insoluble protein pelleting assay (Section 2.2.7). The results are shown in Figure 4.3.

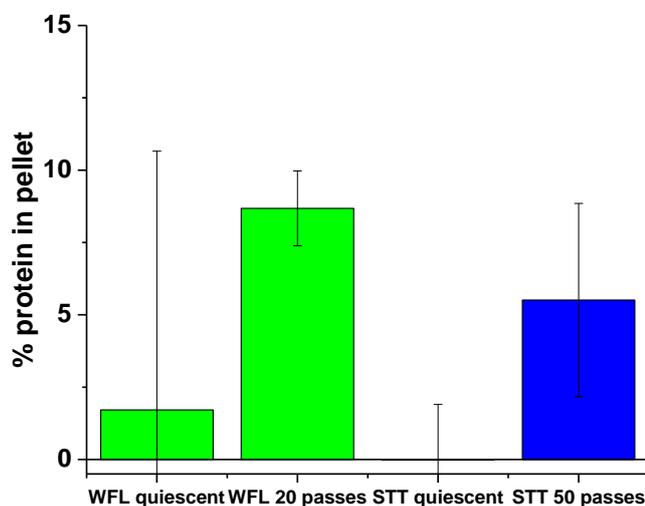


Figure 4.3. Plot of percentage aggregates protein for WFL (green) and STT (blue) following 20 and 50 passes of stress in the EFD (respectively) at a concentration of 0.5 mg mL^{-1} at a plunger velocity of 0.5 mm s^{-1} . Both proteins were stressed in 150 mM ammonium acetate, $\text{pH } 6.0$. The error from two independent experiments were propagated according to Section 2.2.18.

The data show that minimal aggregation is observed ($<10\%$ protein in pellet) for both mAbs. For one replicate of WFL, the prolonged exposure to ambient conditions (the experiment takes ~ 30 mins to complete at 0.5 mm s^{-1}) clearly induced some aggregation in the quiescent sample. In any event, the levels of aggregate seem previously ($>20\%$ protein in pellet) are reduced when this aggregation-prone protein is stressed at this strain rate. In summary, these data show that the three proteins exhibit complex responses to passes and strain rate

in the device. Most strikingly, STT and WFL have very different response surfaces, despite differing by just six residues (out of ~1400).

4.2 The effect of concentration on the extent of flow-induced aggregation

As outlined in Chapter 1, and demonstrated for BSA in Chapter 3, protein aggregation is a concentration-dependent phenomenon. Performing extensional flow experiments at higher protein concentrations, i.e. increasing the protein concentration from 0.5 mg mL^{-1} , is predicted to increase protein aggregation. Furthermore the concentration-dependent aggregation of WFL and STT, studied previously under quiescent conditions (Dobson et al., 2016), could be investigated under flow. To examine the validity of this hypothesis, WFL and STT were stressed for 100 passes at concentrations of 0.5, 1, 2 and 5 mg mL^{-1} at a plunger velocity of 8 mm s^{-1} . mAb1 was only stressed at 1, 2 and 5 mg mL^{-1} for these experiments, thus the 0.5 mg mL^{-1} data from Figure 4.4 are plotted alongside these data as highlighted. The extent of aggregation in the quiescent and stressed samples was quantified using the insoluble protein pelleting assay. The results are shown in Figure 4.4.

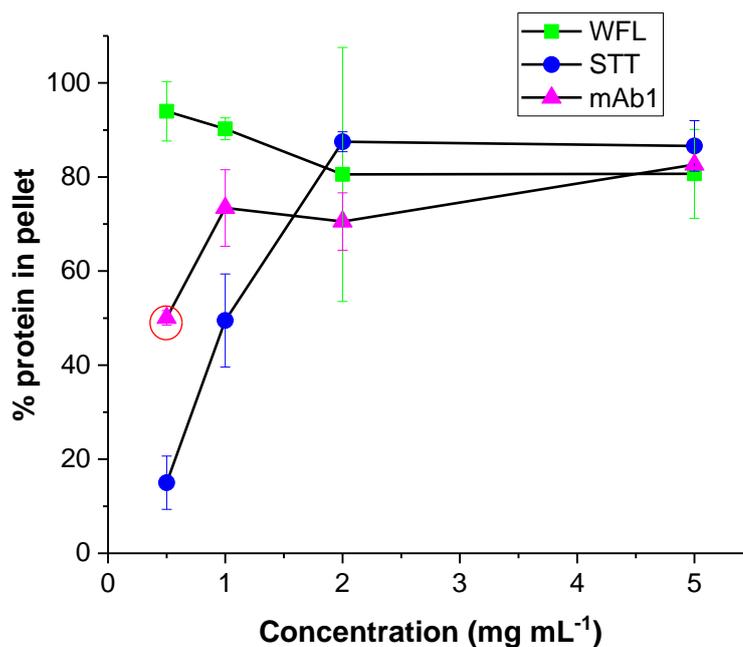


Figure 4.4 Titration plot of % protein in pellet for WFL (green), STT (blue) and mAb1 (magenta) following stress for 100 passes in the EFD at different concentrations. All samples were stressed in 150 mM ammonium acetate pH 6.0 at a plunger velocity of 8 mm s⁻¹. The mAb1 100 passes, 0.5 mg mL⁻¹ data point is from Figure 4.4 (circled in red). Errors from two independent experiments were propagated according to Section 2.2.18. N = 1 for mAb1 (except the circled point).

The data show that the three mAbs have varying sensitivity to concentration over the range investigated. STT is most sensitive, with 88% of the protein rendered insoluble at a concentration of 2 mg mL⁻¹ or higher. mAb1, once again, exhibits intermediate behaviour, with its aggregate extent plateauing at a concentration of 1 mg mL⁻¹ or above. WFL shows almost complete aggregation (ranging from 80–94%), regardless of concentration. The reasons for these concentration effects will be discussed later in this chapter and in Chapter 5.

4.3 Understanding the role of shear in monoclonal antibody aggregation

For BSA, it was found that exposure to high shear rates did not influence aggregation (Chapter 3, Figures 3.19 and 3.24b). As shown in Chapter 3 and the preceding figures here, mAbs are much more sensitive to hydrodynamic flow than BSA. It was thought, therefore, that mAbs may also show sensitivity to shear flows than BSA, despite previous work showing that these molecules were insensitive to this flow field (Bee et al., 2009).

To test this hypothesis, all three mAbs were stressed at a concentration of 0.5 mg mL^{-1} at a plunger velocity of 8 mm s^{-1} . The length of the capillary used was either 75 mm (full-length), 50 mm (2/3 length) or 37.5 mm (half-length). The proteins were subjected to: 20, 50 and 100 passes for WFL, mAb1 and STT respectively. These conditions were chosen as similar levels of aggregation ($\sim 20\text{--}30\%$ protein in pellet) were predicted to be observed from each mAb's respective landscape in Figure 4. Samples were then analysed using the pelleting assay. The data are displayed in Figure 4.5.

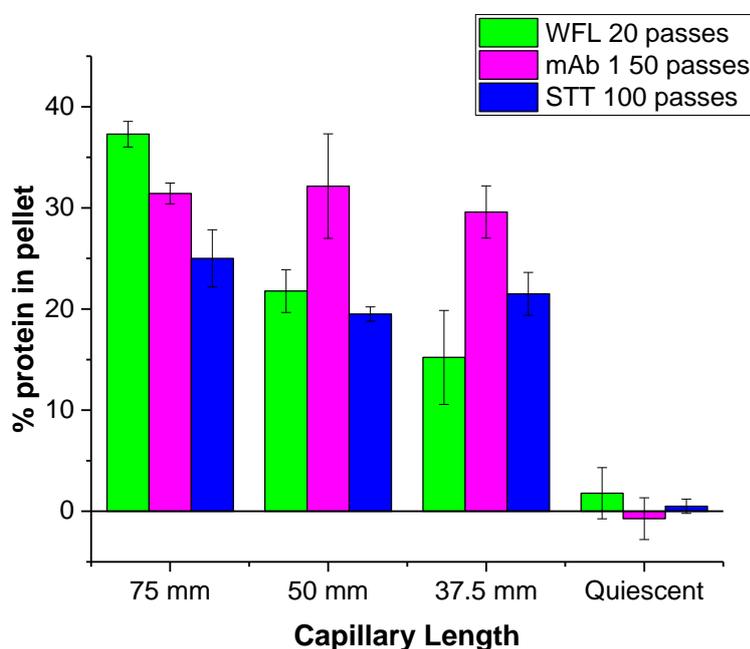


Figure 4.5 Shear-length dependence of the three IgG1s in the EFD. All three mAbs were stressed at a concentration of 0.5 mg mL^{-1} in 150 mM ammonium acetate $\text{pH } 6.0$. The plunger velocity in the device = 8 mm s^{-1} ($\dot{\gamma} = 52,000 \text{ s}^{-1}$). Errors from two independent experiments were propagated according to Section 2.2.18.

The data show that mAb1 and STT are insensitive to shear flow, with the same amount of aggregate forming after flow stress (~ 30% and 22% insoluble protein respectively), regardless of capillary length. WFL on the other hand shows profound shear sensitivity, with the amount of aggregate formed after stress in the EFD approximately halving upon reducing the capillary length from 75 to 37.5 mm. The fact that the EFD initially subjects proteins to an extensional flow event, then shear flow along the capillary may account for the discrepancy with Bee et al. 2009. The result in Figure 4.5 suggests that shear plays a key role in the flow-induced aggregation mechanism of proteins, such as WFL. Proteins such as STT, mAb1 and BSA are not sensitive to shear, instead being much more sensitive to extensional flows as evidenced in Sections 4.1 and 4.2.

4.4 The role of the buffer environment in flow-induced mAb aggregation

As mentioned in Section 1.2.4, the formulation of mAbs into kinetically stable therapeutic dose forms is an area of intensive study in the biopharmaceutical industry (Allmendinger et al., 2015; Goldberg et al., 2011; Goldberg et al., 2017; Wang, 2015). Clearly, the composition of the buffer (especially the buffer salts used, pH and ionic strength) will influence the aggregation behaviour of mAbs. 150 mM ammonium acetate, pH 6.0 has been used previously to study the physicochemical properties of WFL and STT using mass spectrometry (Devine, 2016; Dobson et al., 2016), hence its use thus far in the thesis. To assess how different buffers affect aggregation under flow, STT, mAb1 and WFL were dialysed into four buffers in which the quiescent aggregation of WFL and STT are known (Dobson et al, 2016). 125 mM L-arginine + 20 mM sodium succinate pH 6.0 was also used, as these are the predominant components of the WFL, STT and mAb1 storage buffer (Dobson et al, 2016). The proteins were dialysed into these buffers (Section 2.2.1), then diluted down to a concentration of 0.5 mg mL^{-1} and stressed for 100 passes at a plunger velocity of 8 mm s^{-1} . The resulting aggregation was then quantified with the pelleting assay. The data are shown in Figure 4.6.

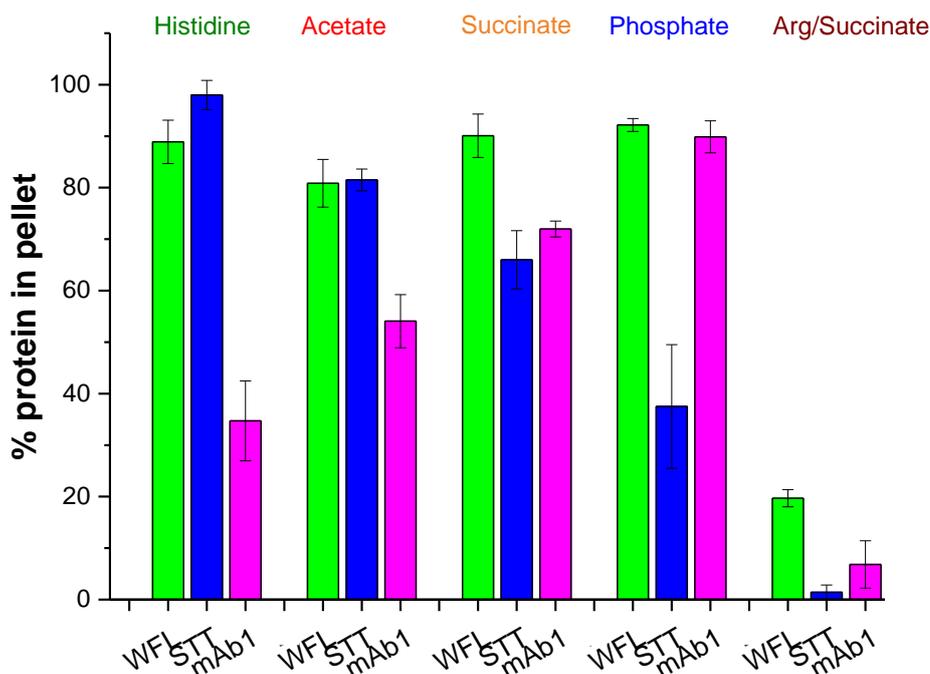


Figure 4.6. Buffer screen of the mAbs with the EFD. WFL (green), STT (blue) and mAb1 (magenta) were dialysed into (left to right): 10 mM L-histidine pH 6.0, 10 mM sodium acetate pH 6.0, 10 mM sodium succinate pH 6.0, 10 mM sodium phosphate pH 7.2 and 125 mM L-arginine + 20 mM sodium succinate pH 6.0. Samples were stressed for 100 passes in the EFD at a plunger velocity of 8 mm s^{-1} . The final mAb concentration was 0.5 mg mL^{-1} in each case. Errors from two independent experiments were propagated according to Section 2.2.18.

The data show that WFL's sensitivity to extensional flow is not ameliorated by any buffer except arginine+succinate. STT behaves worst in L-histidine and best in phosphate and vice-versa for mAb1. The suppression of aggregation for all three mAbs in arginine+succinate is perhaps unsurprising, given its widespread use as an excipient in biopharmaceutical formulations (Baynes et al., 2005; Kim et al., 2016). In previous experiments with WFL and STT in these buffers, DLS was used to detect aggregation (Appendix Figure A4). Whilst stabilising buffers were identified for WFL at concentrations as low as 1 mg mL^{-1} (with phosphate conferring the greatest suppression of aggregation), L-histidine was only identified as being the most stabilising to STT above a protein concentration of 4 mg mL^{-1} . The different results obtained here suggest that the behaviour proteins exhibit in the EFD is very sensitive to the buffer environment. The fact that differences in the aggregation-propensities of the mAbs could be identified at protein

concentrations of just 0.5 mg mL^{-1} highlights the potential for the EFD to be used as a formulation tool in mAb development (see Section 4.6).

4.5 Manipulating the aggregation pathways of WFL and STT

The data presented in this chapter have demonstrated how the strain rate, shear rate and buffer environment significantly affect the aggregation of monoclonal antibodies under flow. From these observations, it was hypothesised that: a) WFL would behave like STT (i.e. to aggregate independent of shear length) when stressed in the presence of arginine, as its WFL motif would be blocked by the excipient and b) STT may be shear-sensitive at high strain rates (as STT seems to behave like WFL at high strain rates (plunger velocity = 16 mm s^{-1})). To test these two hypotheses, shear experiments were performed with each protein. WFL was stressed at a concentration of 0.5 mg mL^{-1} and a plunger velocity of 8 mm s^{-1} for 100 passes in $125 \text{ mM L-arginine} + 20 \text{ mM sodium succinate buffer, pH 6.0}$. The samples were then analysed using the pelleting assay. The results are shown in Figure 4.7.

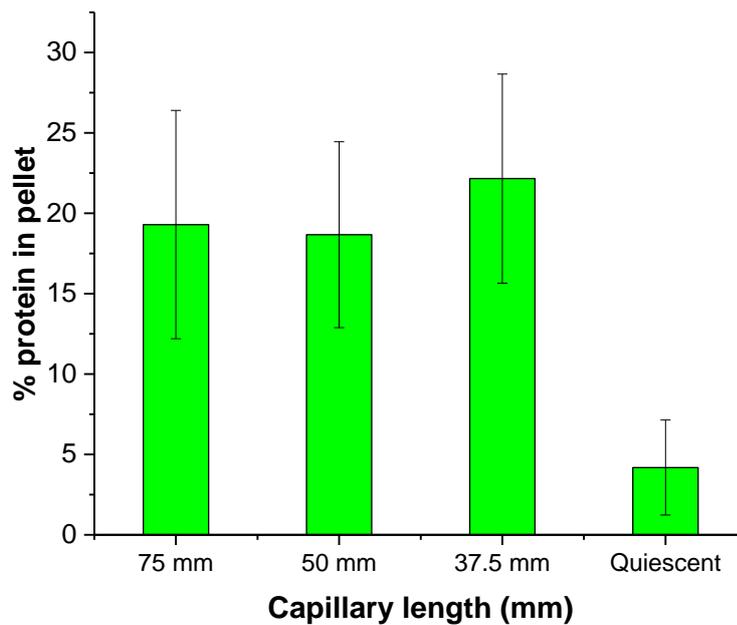


Figure 4.7 Shear-dependence of WFL stressed for 100 passes in arginine+succinate buffer. Samples were stressed at a concentration of 0.5 mg mL^{-1} and a plunger velocity of 8 mm s^{-1} for 100 passes in $125 \text{ mM L-arginine} + 20 \text{ mM sodium succinate buffer, pH 6.0}$. Errors from two independent experiments were propagated according to Section 2.2.18.

The data show that when in the presence of arginine, WFL is no longer sensitive to the high shear rate present in the capillary. After 100 passes in arginine, ~20% of the protein is rendered insoluble, regardless of the capillary length. This is in direct contrast to the data in Figure 4.5, where decreasing the capillary length decreased the levels of aggregation observed. Arginine must therefore affect the aggregation mechanism of WFL by putting the protein into an STT-like state. To investigate whether or not STT aggregates in a WFL-like state at high plunger velocities, the protein was stressed in $150 \text{ mM ammonium acetate}$ for 100 passes, at a plunger velocity of 16 mm s^{-1} . Full-length and half-length capillaries were used in the experiment. The samples were then analysed using the pelleting assay. The data are shown in Figure 4.8.

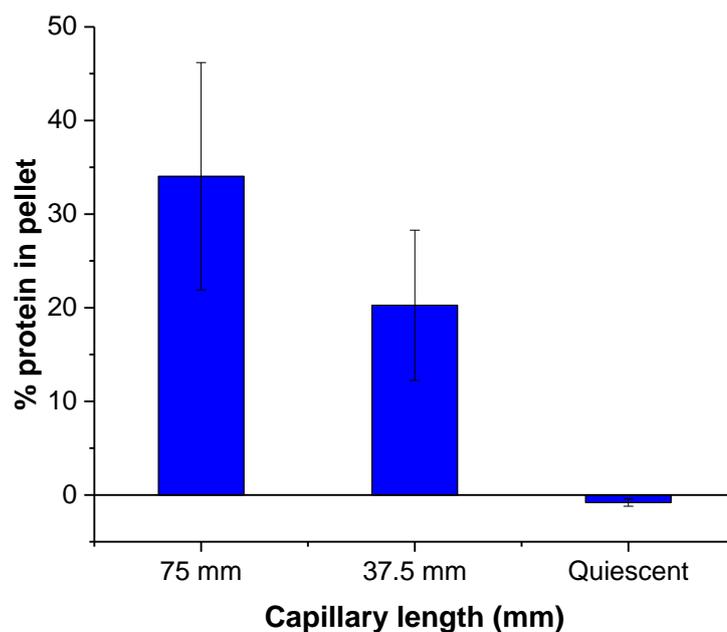


Figure 4.8 Shear-dependence of STT at a plunger velocity of 16 mm s^{-1} ($\dot{\epsilon} = 23,421 \text{ s}^{-1}$, $\dot{\gamma} = 104,000 \text{ s}^{-1}$). STT was stressed in 150 mM ammonium acetate, pH 6.0 at a concentration of 0.5 mg mL^{-1} . Errors from two independent experiments were propagated according to Section 2.2.18.

The data show that, on average, decreasing the length of the capillary decreases the amount of aggregate measured with the pelleting assay for STT (from 34% to 20%). Whilst this difference is not as large (nor as significant) as that observed for WFL in Figure 4.5, these data seem to suggest that STT does behave very differently at this high plunger velocity than under the ‘standard’ flow conditions of 8 mm s^{-1} .

4.6 Discussion

Through varying both the plunger velocity and the number of passes a protein solution is exposed to in the EFD, it is possible to assess how the magnitude and frequency of flow events affect protein aggregation. The variation of these two parameters was principally explored for BSA and the three mAbs, WFL, mAb1 and STT. In Chapter 3, it was shown that BSA has a strain rate threshold above which it begins to aggregate after 100 passes in the EFD. It was also shown that monoclonal antibodies, which have both double the mass and a distinct topology to BSA, are much more sensitive to the effects of hydrodynamic force than BSA. It was thus hypothesised that the strain rate threshold for these molecules would be much lower than for BSA, i.e. lower than $14,634 \text{ s}^{-1}$ (equivalent to a plunger velocity of 10 mm s^{-1}). The data outlined in Figure 4.2a showed this was largely not the case. WFL showed pronounced levels of aggregation, with ~30% of the protein being rendered insoluble after just 20 passes through the device (Figure 4.2b). The strain-rate threshold where aggregation is minimised for this protein is very low indeed, ($\dot{\epsilon} = 871 \text{ s}^{-1}$, Figure 4.3). This may suggest that little structural perturbation, induced by hydrodynamic forces, is required to expose aggregation-prone regions in this protein, with the amount of aggregate increasing as a function of pass number.

mAb1 possesses an 'intermediate' response to flow when compared to WFL and STT (Figure 4.2c). When the number of passes increases, so too does the level of aggregate formed. At lower plunger velocities (2 and 4 mm s^{-1}), the flow-induced aggregation of mAb1 is reduced. At higher plunger velocities, two plateaus of aggregation can be seen ($4\text{--}8 \text{ mm s}^{-1}$ and $10\text{--}14 \text{ mm s}^{-1}$). The 72% sequence identity of mAb1 to WFL and STT may account for the 'intermediate' flow-induced aggregation behaviour possessed by this generic protein.

Conversely, STT's aggregation landscape (Figure 4.2d) has a large region of parameter space where few aggregates form, with less than 30% aggregation observed below 100 passes regardless of strain rate. The landscape analysis reveals regions of equivalence, e.g. 150 passes at low plunger velocities ($2\text{--}6 \text{ mm s}^{-1}$) causes the same amount of aggregation as when the protein experiences 50 passes at high plunger velocities (16 mm s^{-1}). This highlights the complex relationship between these variables and their effect on the observed levels of aggregation. This was observed despite WFL and STT having >99% sequence

identity, differing by just three residues in each V_H domain. The exposure times and magnitudes to shear flow in bioprocessing are reasonably well-known (Bee et al., 2009; Thomas and Geer, 2011). For example, it is thought that mAbs are exposed to similar shear rates and exposure times in the EFD as they are during fill-finish operations (Bee et al., 2009). The equivalent calculations for strain rates are reported less frequently (exceptions include: (Biddlecombe et al., 2009; Tavakoli-Keshe et al., 2014)). The data in Figure 4.2 could inform bioprocessing if these calculations are made for plant-scale equipment and the conditions matched to the data in Figure 4.2. By operating equipment under less damaging conditions, like the zone discussed for STT above, yields of the final product could be maximised by avoiding the extreme, damaging conditions in the top corner of the landscape (200 passes at 16 mm s^{-1}) during processing.

To further flesh out the differences between these three mAbs, the concentration of protein was varied (Figure 4.4). At increased concentrations (1 or 2 mg mL^{-1} for mAb1 and STT, respectively), the amount of aggregate formed in the EFD increases and plateaus. The concentrations explored in this study are low (5 mg mL^{-1} being the highest used) compared to the concentrations typically found for mAb products (Garidel et al., 2017), although they be representative of some points within the bioprocess train (Cromwell et al., 2006). This not only minimised sample consumption, but also ensured that the fluid remained Newtonian, i.e. the viscosity did not change with shear rate (Castellanos et al., 2014), important as the strain rates would change in a non-Newtonian system (King, 2002). This is vital considering the device criteria, set out in the aims (Section 1.6).

When the length of time the mAbs were exposed to shear flow (following extension) was changed, different levels of aggregate were observed (depending on the protein and the buffer conditions used). WFL's aggregation appears to be strongly influenced by shear, with decreased aggregation seen when the capillary length was decreased (Figure 4.5, green bars). From a practical perspective, it was unfeasible to adapt the current EFD to have a capillary smaller than 37.5 mm in order to explore this effect further, i.e. investigate if WFL's flow-induced aggregation can be solely attributed to shear flow. The fact that WFL is sensitive to these conditions is in direct contrast to the data of Bee and others, who have suggested that shear flow does not adversely affect proteins (Bee et al., 2009; Jaspe and Hagen 2006). This result also accentuates the disparity in aggregation behaviour between WFL and STT, despite the two proteins being almost identical in sequence.

This behaviour, as well as the general aggregation propensity of WFL, could be suppressed by performing the flow experiments in the presence of arginine + succinate buffer. It is thought that the aliphatic part of the arginine side-chain can compete for aggregation-prone regions. In addition, the charged moieties of arginine could block polar protein-protein interactions, thus preventing protein aggregation (Kim et al., 2016). Whether this is the mechanism by which mAbs are protected in the EFD is as yet unclear. The fact that the device could identify de-/stabilising buffers using such low concentrations of protein (much lower than those used in the literature (Dobson et al., 2016; Goldberg et al., 2017)) opens up the possibility of the EFD being a formulation tool in the biopharmaceutical industry.

Another result with an unclear explanation is why STT exhibits some shear-sensitivity when stressed at a plunger velocity of 16 mm s^{-1} . It could be that statistically, more STT molecules are put into an activated state following stress under these conditions. Another possibility is that more of the STT molecule is unfolded when it is stressed at this higher strain rate, i.e. the size of the aggregation-prone region and its 'stickiness' becomes more like that found in WFL. These 16 mm s^{-1} activated STT molecules then behave like WFL in the flow device. This suggests that the flow-induced aggregation pathway of a mAb may be dependent on the extent of unfolding it undergoes.

To conclude, through variation of many parameters in the EFD, distinct aggregation behaviours could be identified for BSA and three mAbs. Focussing on the antibodies, conditions were identified whereby these molecules are aggregation-prone or resistant, and the effects of shear, concentration and the buffer environment quantified. The mainly observational data generated in this Chapter, together with the data obtained in Chapter 3, seem to suggest two different pathways through which mAbs aggregate when subjected to extensional flow. A variety of techniques will need to be implemented, not just insoluble pelleting assays, to dissect such an aggregation pathway. This mechanistic work forms the subject of the next chapter.

5. Towards a mechanism of flow-induced monoclonal antibody aggregation

5.1. Preface

Thus far, data have been presented which show that proteins aggregate when they are subjected to extensional and shear flows. In Chapter 3, the aggregation pathway for BSA induced by extensional flow was interrogated. Not only was evidence gathered for conformational remodelling of the monomeric protein, but quantitative data on the end-point of the reaction (amount of insoluble protein found following stress), were also generated. In Chapter 4, a thorough analysis was performed to investigate how the aggregation end-point of mAbs changed in response to changing the wide array of variables present in the EFD. All of these data suggest mAbs may aggregate in the EFD *via* two competing pathways, with the quantity and type of aggregates present along the pathway being protein and condition dependent. These proposed pathways are shown in Figure 5.1.

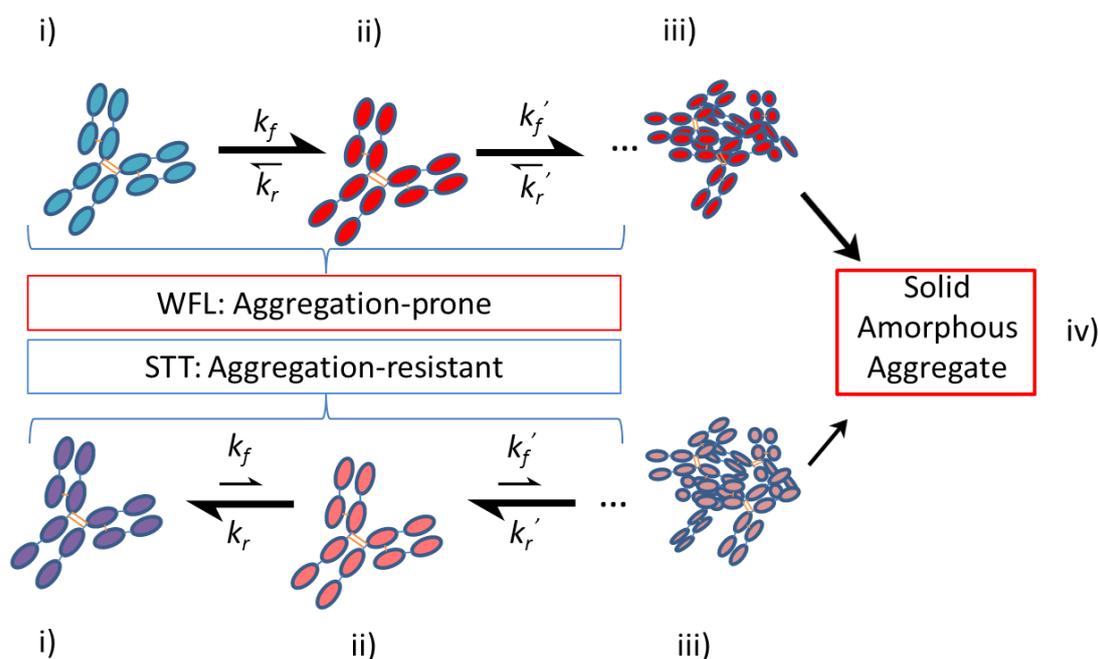


Figure 5.1. Outline of the proposed mechanism of aggregation of mAb aggregation induced by extensional flow. The top pathway is traversed by mAbs which rapidly aggregate through specific interfaces, such as WFL.

The bottom pathway is traversed by mAbs which are more resistant to aggregation. i) The monomeric native mAb is put into an aggregation-prone N^* state by extensional flow. Sensitive proteins such as WFL are put into this state more readily than resistant proteins, such as STT. ii) These N^* species can either re-fold back to the native state, or will self-associate to form soluble aggregates along the aggregation pathway (iii). Once the aggregates reach a certain size, they will no longer be soluble, nor will they be reversible (iv). k_f and k_r are the apparent rate constants for the formation of the N^* state/ its refolding respectively. k'_f and k'_r are the apparent rate constants for the association of N^* species and the dissociation of aggregates respectively.

The mechanism proposed in Figure 5.1 is based on the established aggregation pathways within the LENP model, as discussed in Section 1.1 (see also Figure 1.2). Whilst much is known about the end-point of the antibody aggregation reaction (i.e. the amount of insoluble aggregate formed), there are many unanswered questions surrounding the mechanism *en route* to insoluble protein: At what point is aggregation irreversible? What species (if any) nucleate the aggregation pathway? Do the aggregate size distributions reach equilibrium at a given time point? What is the minimum number of passes needed to trigger

aggregation? What is the nature of the N* state? At what rate do all of the above processes take place?

In this Chapter, data will be presented to attempt to answer the questions above, moving towards a mechanistic understanding of mAb aggregation under flow.

5.2 Using chemical tools to probe the N* state

As outlined in Figure 5.1, it is thought that the initial step in the flow-induced mAb aggregation pathway is the formation of an N* state. As these species underpin the whole pathway, detecting their formation under extensional flow would be highly desirable. For BSA, the free Cys 34 could be exploited to capture conformational re-modelling of the protein using IAEDANS (Chapter 3.4.3). However, all of the Cys residues in WFL, STT and mAb1 form disulphide bridges. An alternative strategy was thus sought from the literature. Recently, Wolfrum et al. (2017) used their four-roll mill device (see Section 1.4.4.) to capture mAb unfolding and aggregation, both *in situ* and *ex situ*, by using the extrinsic fluorophore Nile Red (Wolfrum et al., 2017). The authors used this environmentally sensitive dye to detect the exposure of hydrophobic patches (APRs) upon subjection to flow.

To test if this strategy could be used with the EFD, the extrinsic fluorophores Nile Red and ANS were incubated with WFL and STT samples either before (*in situ*) or after (*ex situ*) stress. No signal change was observed between the quiescent samples and those subjected to hydrodynamic stress (data not shown).

Chemical cross-linking was then used as an alternative probe of the N* state. In previous work, the aggregation interface of WFL under quiescent conditions was identified using chemical cross-linking, followed by mass spectrometry (Devine, 2016; Dobson et al., 2016). The interface identified in this work is summarised in Figure 5.2.

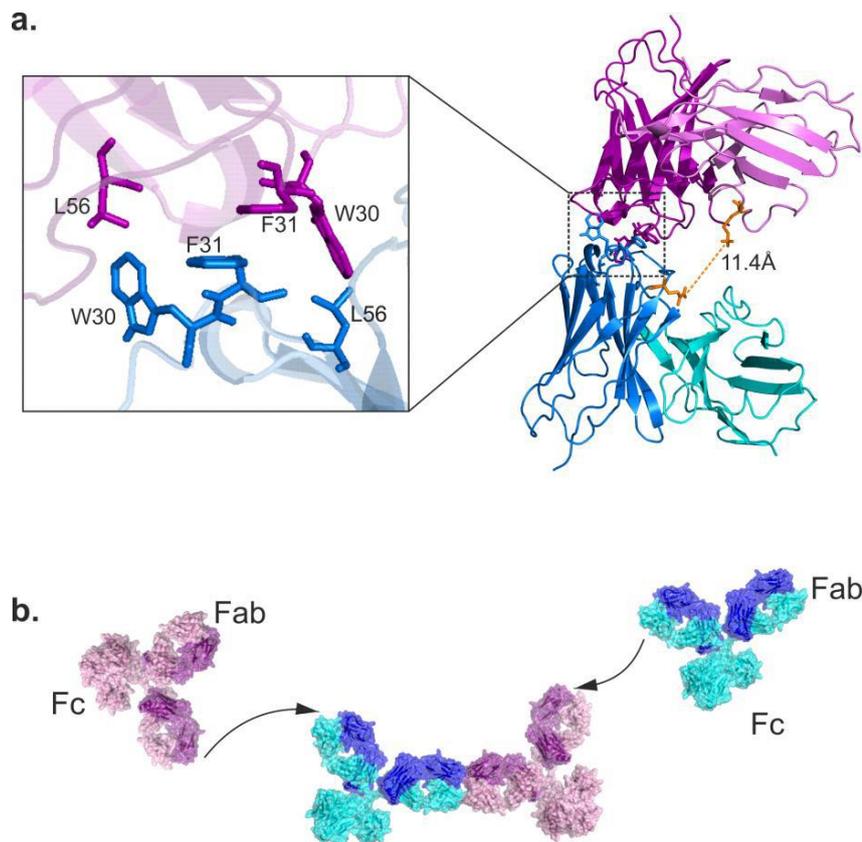


Figure 5.2. Chemical cross-linking of WFL to identify its dimer interface. a) Cross-linking with BS^3 - BS^3 - d^4 identified a cross-link between the V_H N-terminus and K54 in V_L (orange). The key residues in the interface (W30, F31 and L56) are magnified. b) The model proposed by Dobson et al. for WFL aggregation, with monomers self-associating through Fab-Fab interactions. Figure taken from Dobson et al. 2016.

Based on these results, it was hypothesised that if flow altered the conformation of WFL, then a different interface, or many interfaces, would become exposed. These N^* species would then aggregate through these new interfaces. Conversely, identification of the same cross-link would suggest that flow facilitates the formation of the WFL-WFL interaction, thus catalysing the aggregation of the protein. Attempts to identify this cross-link under both quiescent conditions under flow were unsuccessful (data not shown). In summary, under the conditions attempted, the N^* state of WFL (nor STT in the fluorescence experiments) could not be reproducibly or definitively captured by either extrinsic fluorescence or cross-linking (see Discussion). However, a pathway describing how monomers associate to form oligomers, which then become insoluble according to Figure 5.1, can still be constructed using other techniques. These form the basis of the next section of this chapter.

5.3 Following the pathway of mAbs from monomers to aggregates

5.3.1 Using HPLC to determine the loss of monomer as a function of pass number

According to the pathway proposed in Figure 5.1, as is the case for *all* protein aggregation mechanisms, the loss of monomer underpins the whole aggregation pathway (Li and Roberts, 2009; Meisl et al., 2016). This is because, as defined in Section 1.1, an aggregate consists of two or more non-native protein monomers. Measurement of the loss of monomer as a function of time thus means the *exact* amount of protein which has formed aggregate can be measured. To do this, HPLC was exploited, as this is the ‘gold-standard’ technique used by both industrial and academic groups to quantify mAb aggregation (Nicoud et al., 2014; Philo, 2006).

Consequently, WFL, mAb1 and STT were dialysed into 150 mM ammonium acetate, pH 6.0 and stressed for 0–100 passes at a plunger velocity of 8 mm s⁻¹. The protein concentration was kept at 0.5 mg mL⁻¹. Following stress, the samples were clarified by ultracentrifugation and the extent of aggregation quantified using the pelleting assay (Section 2.2.7). Subsequently, 100 µL of supernatant from each tube was combined with 200 µL of quench solution (625 mM L-arginine + 100 mM sodium succinate pH 6.0). The samples were then analysed by HPLC using a TOSOH G3000SWXL gel filtration column (Section 2.2.13). The chromatograms for WFL, mAb1 and STT are shown in Figures 5.3 a–c respectively.

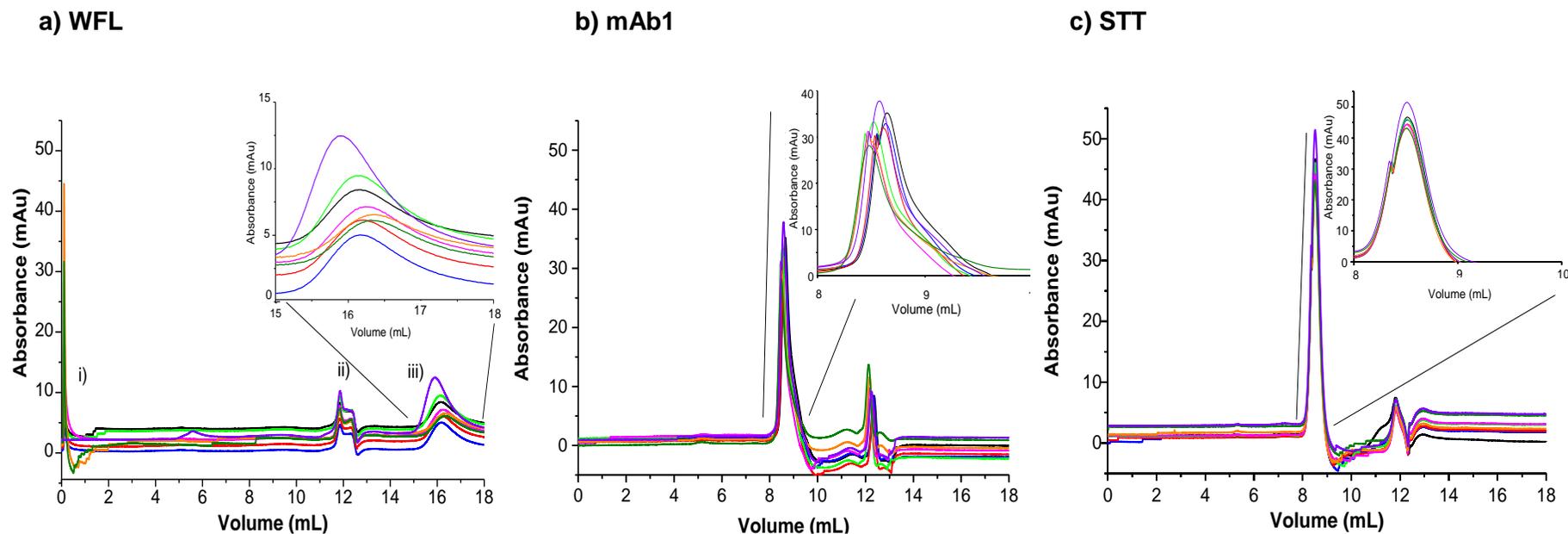


Figure 5.3. HPLC chromatograms for a) WFL, b) mAb1 and c) STT after stress for the given number of passes at a plunger velocity of 8 mm s^{-1} . All protein samples were initially stressed at a concentration of 0.5 mg mL^{-1} in 150 mM ammonium acetate pH 6.0. $50 \mu\text{L}$ injections of quenched sample (Section 2.2.13) were loaded onto a TOSOH G3000SWXL column, equilibrated in 125 mM L-arginine + 20 mM sodium succinate pH 6.0, running at 0.4 mL min^{-1} . One representative trace is shown per sample. Traces are coloured: 0 passes (–), 1 pass (–), 5 passes (–), 10 passes (–), 20 passes (–), 50 passes (–), 80 passes (–) and 100 passes (–). In Figure 5.14a), i) = carry-over peak, ii) = buffer exchange peak, iii) WFL peak.

The data for WFL in Figure 5.3a are very difficult to interpret. Though arginine is thought to minimise the interaction of hydrophobic proteins (like WFL) with gel chromatography matrices (Nicoud et al., 2014), this running buffer was not able to prevent aberrant, non-specific interactions with the column. Significant carry-over was observed from one run to another in some samples, with the quench buffer solvent exchange peak eluting before the WFL mAb (see i, ii and iii in Figure 5.3a). This behaviour has been reported previously for WFL, but in a different running buffer (100 mM sodium phosphate, 100 mM sodium sulphate pH 6.8) (Dobson et al., 2016). This rendered quantification of the loss of WFL monomer futile. On the other hand, both mAb1 and STT show a loss of peak area as a function of pass number, eluting at ~ 8 mL in each case, as expected for a mAb from previous data (Devine, 2016) (Figures 5.3b and c, respectively).

Each peak was integrated, with the resulting areas used to calculate the mass of protein in the injection and thus the monomer concentration. The plot of % monomer lost by HPLC, as well as % loss of supernatant by the pelleting assay (performed on samples from the same experiment), is shown in Figure 5.4.

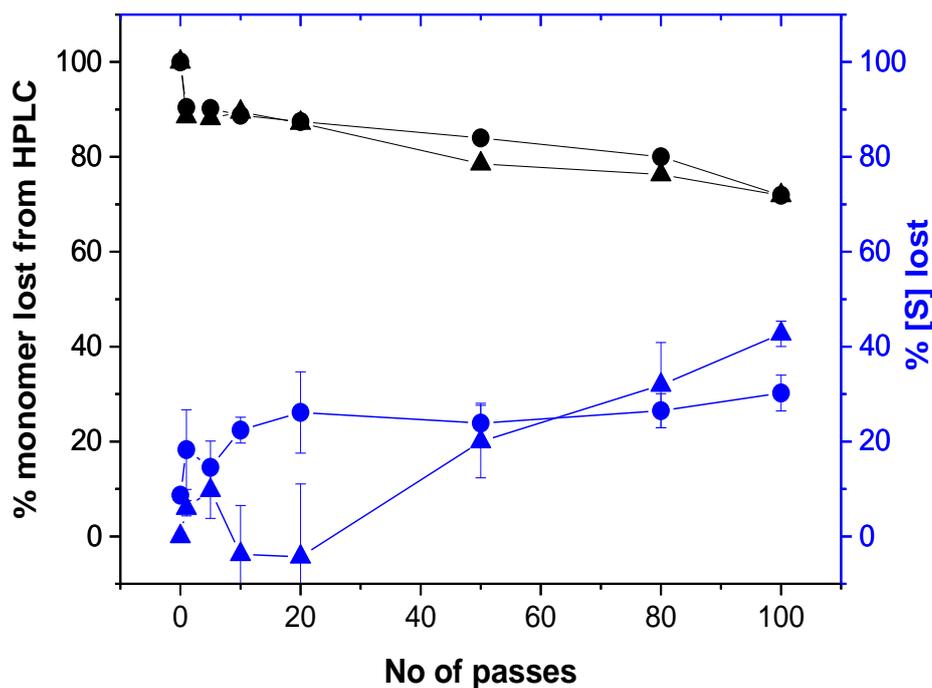


Figure 5.4. Plot of percentage monomer lost as determined by HPLC (left-hand y-axis and black lines) vs loss of protein from the supernatant (right-hand y-axis and blue lines) as a function of the number of passes through the device (x-axis). The data for STT (circles) and mAb1 (triangles) are the same samples as those in Figure 5.3.

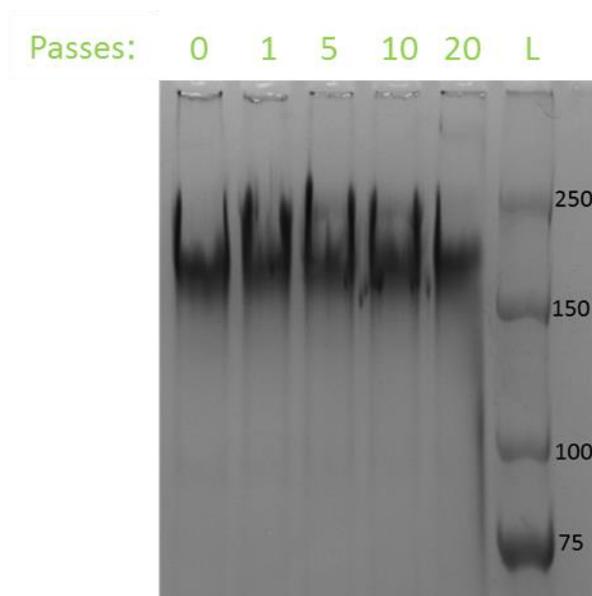
The data seem to show that the amount of monomer lost as measured by HPLC is approximately the same for both proteins after 100 passes (~30%, black symbols). By pelleting, more mAb1 is lost from solution than STT under the same flow conditions (40 and 30% respectively, blue symbols). Given that mAb1 and STT can be readily differentiated by a variety of assays (Chapters 3–5), this identical behaviour in the HPLC assay seems unusual, as well as the shape of the monomer-loss curves. In summary, for this limited number of samples, HPLC could monitor the loss of monomer, induced by extensional flow, as a function of time. However, how these data compare to those from other assays remains inconclusive. Characterising the early events in the aggregation pathway (monomers to oligomers) is the subject of the next section.

5.3.2 Detecting oligomer formation using SD-SDS-PAGE

As shown in Section 5.3.1, it appears that monomeric protein is lost as a function of pass number in the EFD. But where do the monomers go? Determining the stoichiometry of an aggregation reaction is important with regards to any mechanistic models one may wish to construct from the data (Li and Roberts, 2010).

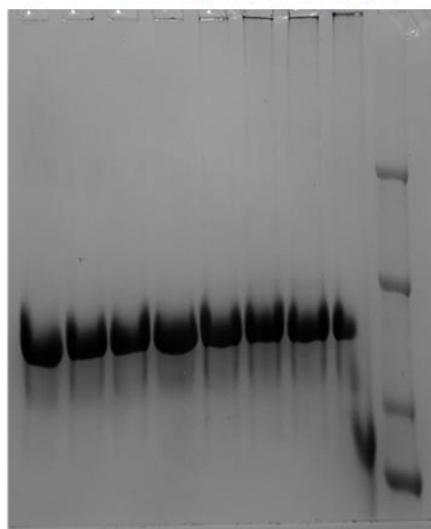
With these observations in mind, WFL, mAb1 and STT were stressed in 150 mM ammonium acetate, pH 6.0, for 0–100 passes at a concentration of 0.5 mg mL⁻¹. The plunger velocity in all experiments was 8 mm s⁻¹. Following stress, Semi-Denaturing (SD) SDS-PAGE samples were prepared according to Section 2.2.14, with no reducing agent or SDS in the loading dye. Gel samples were then loaded onto a 6% (w/v) SDS-PAGE resolving gel, then electrophoresed for 3 hours. Following this, the samples were initially stained using Coomassie Blue (Figures 5.5 a–c for WFL, mAb1 and STT respectively). In later experiments, gels were stained with both SilverXpress and Coomassie Brilliant Blue (Amit Kumar, University of Leeds). Representative silver/blue gels for WFL, mAb1 and STT are shown in Figures 5.6 a–c respectively.

a) WFL



b) mAb1

Passes: 0 1 5 10 20 50 80 100 L



c) STT

L 0 1 5 10 20 50 80 100

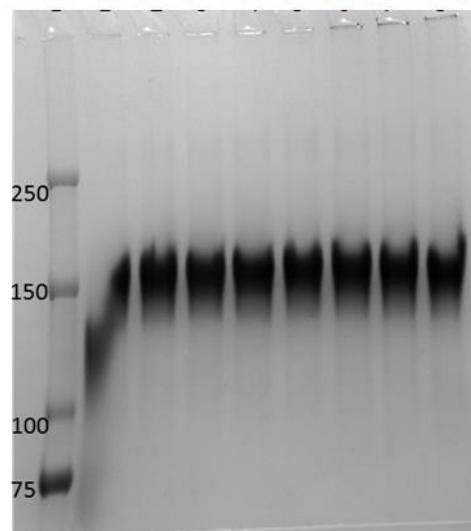
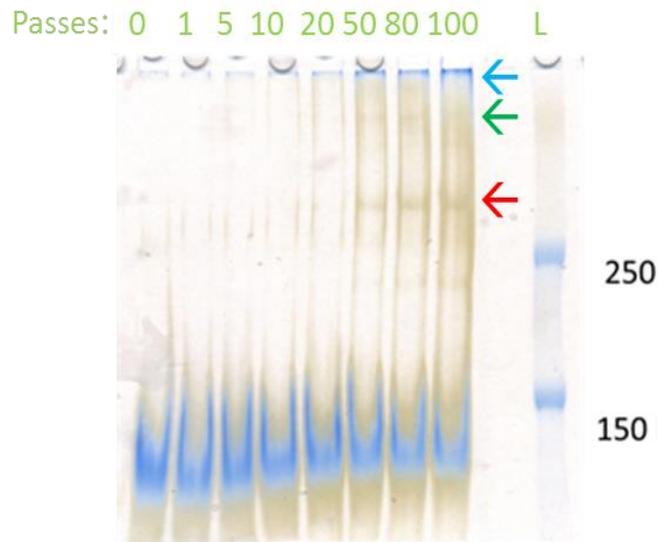


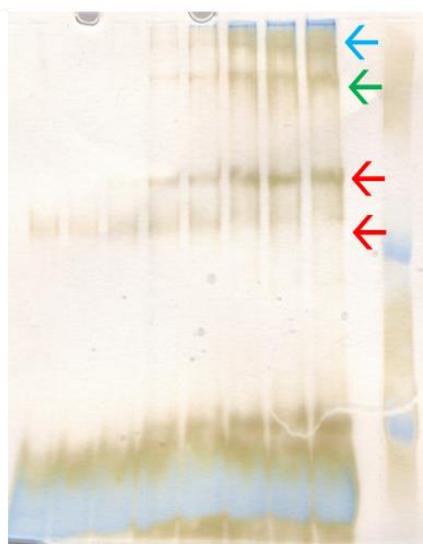
Figure 5.5. 6% (w/v) Semi-denaturing SDS-PAGE of a) WFL (green), b) mAb1 (magenta) and c) STT (blue) samples following stress in the EFD. The number of passes each mAb was subjected to is indicated in each lane. L = Molecular weight marker (MW in kDa). 25 μ L of each sample was loaded (\sim 10 μ g per lane). Stress in the EFD was performed at a plunger velocity of 8 mm s⁻¹. Gels were electrophoresed at 45 mA for 3 hr. Images shown are representative gels from two independent N=2 experiments.

a) WFL



b) mAb1

Passes: 0 1 5 10 20 50 80 100 L



c) STT

0 1 5 10 20 50 80 100 L

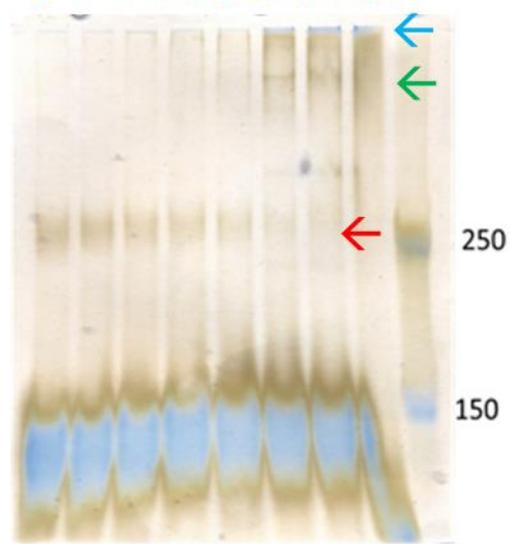


Figure 5.6. Semi-denaturing SDS-PAGE of a) WFL, b) mAb1 and c) STT samples following stress in the EFD. The number of passes each mAb was subjected to is indicated in each lane. L= Molecular weight marker (MW in kDa). 25 μ L of each sample was loaded (\sim 10 μ g per lane). Stress in the EFD was performed at a plunger velocity of 8 mm s⁻¹. Gels shown are representative from at least two independent experiments. Gel preparation, loading and staining was mainly performed by Dr Amit Kumar, University of Leeds. Arrows indicate: dimers (\rightarrow), trimers (\rightarrow) and higher oligomers (\rightarrow).

The data in Figure 5.5 show that oligomerisation is generally difficult to detect in any samples which have undergone hydrodynamic stress in the EFD. When large amounts of aggregate formed, the aggregates did not permeate into the gel, as seen at the top of some lanes, e.g. 50–100 passes in Figures 5.5b and c. Interestingly, the different proteins seem to run as monomers at different levels on the gel; WFL has a 'streaky' monomeric band which runs above the 150 kDa MW marker (Figure 5.5a), mAb1 runs below 150 kDa (Figure 5.5b) and STT at the same level as the marker (Figure 5.5c) (see Discussion).

As the oligomer stoichiometry was the focus of this experiment, when gels were stained by SilverXpress, these gels were run until the 150 kDa marker was ~1 cm from the bottom of the gel (Figures 5.6a–c). Here, the enhanced sensitivity of SilverXpress staining shows that traces of dimers are present in all quiescent samples (red arrows). The dimer population seems to enrich as a function of pass number between 1–20 passes. Following 10 or 20 passes, unique conformations of dimers are particularly prominent on the mAb1 gel (Figure 5.6b). On the STT gel (Figure 5.6a), the dimer band appears to disappear after 20–100 passes. On all three gels, a higher Mw band that appears to migrate ~ half the distance of the 150 and 300 kDa bands is visible, suggestive of a trimer (green arrows). The gel then cannot resolve any higher oligomers (blue arrows). In many samples, especially those of WFL, aggregates can be detected at the top of the gel following stress by both silver and Coomassie staining, with the size of the band increasing as a function of pass number. To summarise, SD-SDS-PAGE seems to be able to report on the conformation of monomers and dimers present in the mAb samples. The dimer population changes as a function of pass number, with the next resolvable oligomer being a trimer. The onset of aggregation appears to be 10 passes. Other orthogonal techniques are necessary to aid the interpretation of this result; these form the basis of the next section.

5.3.3 Using light scattering techniques to track the size and mass of aggregates over time

In Chapter 3, the growth of BSA aggregates and their dispersity over time was tracked using DLS, NTA and FCS. BSA aggregation was detected by DLS after 50 passes, whereas mAb aggregation could be detected after just 20 passes (Section 3.3). DLS is a completely orthogonal method to SD-SDS-PAGE and has been used to measure the onset of aggregation, as well as the size distributions,

of model antibody aggregation systems by many in the field (Arosio et al., 2012; Biddlecombe et al., 2009; Singla et al., 2016). The Wyatt miniDAWN instrument collects static light scattering data, as well as dynamic light scattering data. If the protein concentration of the sample is known, then the ensemble averaged mass of scattering species can be inferred. This strategy has also been used to track mAb aggregation under a range of conditions (Arosio et al., 2012; Barnett et al., 2015; Nicoud et al., 2016).

With this in mind, WFL and mAb1 were stressed for 0–20 passes, whilst STT was stressed for 0–100 passes at a plunger velocity of 8 mm s^{-1} in 150 mM ammonium acetate, pH 6.0. Samples were stressed at a concentration of 0.5 mg mL^{-1} then analysed in the miniDAWN TREOS by both DLS and SLS (see Sections 2.2.4 and 2.2.15). The DLS data were subjected to both Cumulants and Regularisation analysis, as in Chapter 3. However, only the cumulants data (Figure 5.7) and resultant PDI's (Figure 5.8) are displayed. Static light scattering data were fitted according to Section 2.2.15 (Figure 5.10). These data are discussed in Figures 5.7–5.9 respectively.

0.5 mg mL^{-1}

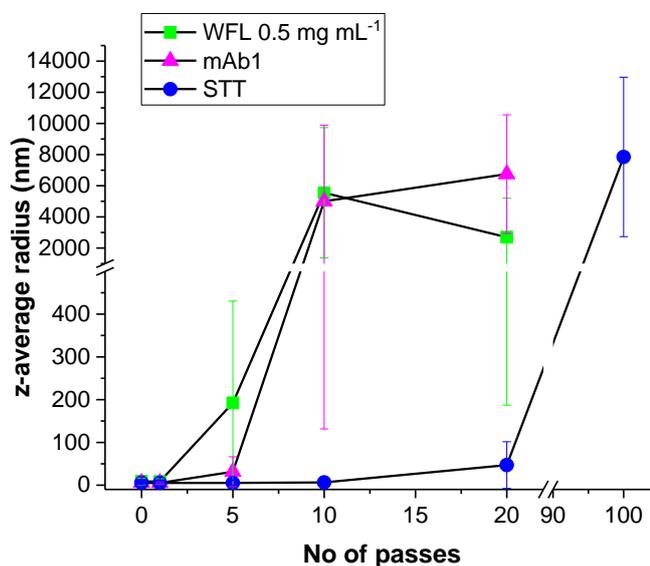


Figure 5.7. Cumulants analysis of DLS data for WFL (green), mAb1 (pink) and STT (blue) obtained following no stress (0 passes) or 1–100 passes in the EFD. Samples were stressed in the EFD at a concentration 0.5 mg mL^{-1} and a plunger velocity of 8 mm s^{-1} . Errors were propagated from $N=2$ experiments according to Section 2.2.18.

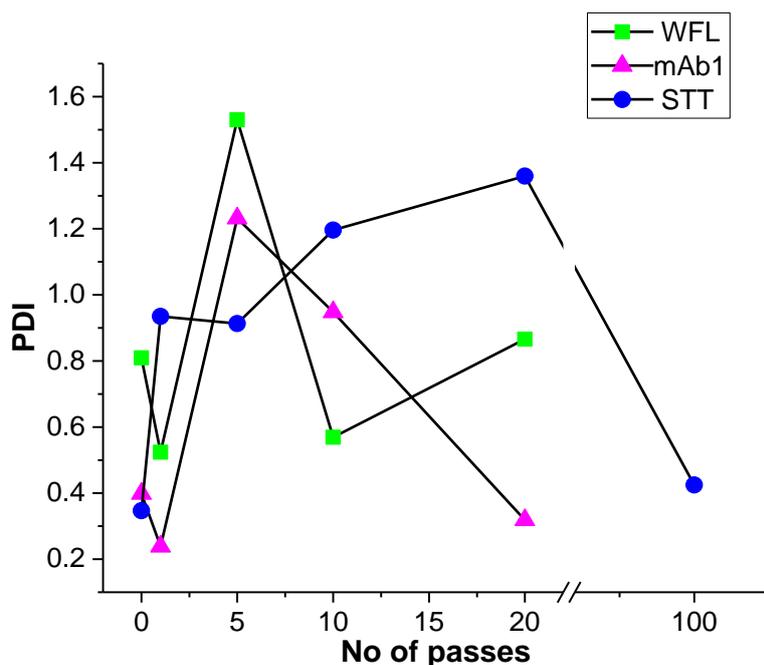


Figure 5.8. Polydispersity index (PDI) calculation of the samples shown in Figure 5.7. PDIs were calculated according to Equation 2.5 in Section 2.2.4.

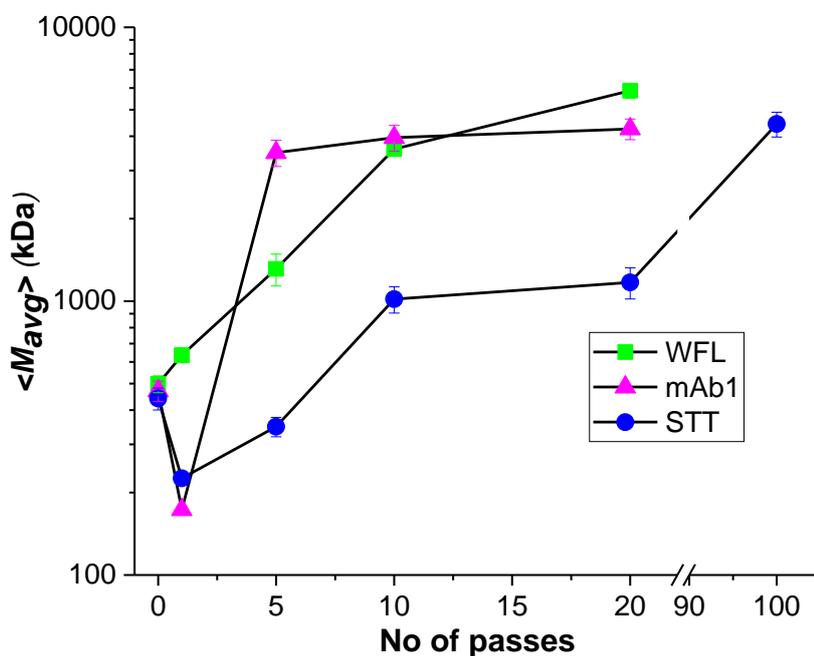


Figure 5.9. Plot of uncertainty averaged molecular weight (M_{avg}) for WFL, mAb1 and STT samples determined using static light scattering. Data were fit to the Debye model according to Section 2.2.15. Errors from two independent experiments were propagated according to Section 2.2.18.

The DLS analysis in Figure 5.7 shows that WFL begins to aggregate after just 5 passes in the device. mAb1 forms micron-sized aggregates after just 10 passes, with STT requiring 100 passes of stress to form such large species. The mAbs in the quiescent samples were sized from 5.2–8.8 nm, in accord with the data in Chapter 3. When the PDI was calculated from the cumulants data (Figure 5.8), the samples below 20 passes were shown to be highly disperse, with many PDI values exceeding the 0.6 threshold discussed in Chapter 3. The dispersity of the final stressed protein samples (20 passes for WFL and mAb1 and 100 passes for STT) appears to decrease to a lower value. This is in contrast to the PDI data for BSA in Chapter 3 (Figure 3.9), which generally increased as a function of pass number. For the SLS data in Figure 5.9, the *average* mass of the scattering species in the sample increases as a function of pass number. This value plateaued after 10 passes of WFL and mAb1 at ~4.5 MDa. STT's aggregates only reach this mass following 100 passes of stress in the EFD. Taken together, these light scattering data show that WFL and mAb1 are more sensitive to flow than STT.

In Chapter 4, it was demonstrated that the amount of protein in the pellet fraction could be tuned for mAb1 and STT by increasing the protein concentration. But what happens at lower concentrations? Are all three mAbs (including WFL) less sensitive to flow if the concentration is decreased? These would be interesting questions to address in future (see Discussion). To understand the kinetics, learning how the size of aggregates changes a function of both time *and* concentration would be necessary. In summary, using DLS and SLS, the onset of aggregation and the growth of aggregates over time can be tracked for all three mAbs as a function of pass number.

5.4 Characterising the size distribution, stability and nucleating ability of flow-induced mAb aggregates

5.4.1 Using DLS to measure aggregate size distributions over time

So far, it has been determined that mAbs appear to aggregate under flow from a pool of oligomers, which grow as a function of the number of passes through the EFD. There will be a critical size/mass above which the aggregates in Figure 5.1 are no longer reversible and become insoluble. Presumably, the 'reversible' species can change size as a function of time until the system reaches equilibrium (according to Figure 5.1). To investigate this further, mAb1 was chosen as the model protein, as it has been shown to behave in an intermediate fashion to WFL and STT throughout this thesis. Regularisation analysis of DLS data generates size distributions of species scattering in the sample, from monomers through to large (1–5 μm) aggregates (Chapter 3). To see if the size distribution of the aggregates changes over time, mAb1 samples were stressed at a concentration of 0.5 mg mL^{-1} in 150 mM ammonium acetate, pH 6.0, for 50 passes. The plunger velocity used was 8 mm s^{-1} . Some samples were clarified by ultracentrifugation, using the same conditions as the protein pelleting assay (see Section 2.2.7) and then the supernatants combined and measured by DLS. Other samples were unclarified and analysed immediately following stress. The quiescent sample was incubated for 5 minutes at room temperature, then measured by DLS. Following analysis, remaining samples were placed on ice and incubated at 4°C for 48 hr. The samples were then measured again by DLS. This was performed twice on separate sets of protein. A representative set of data are shown in Figure 5.10.

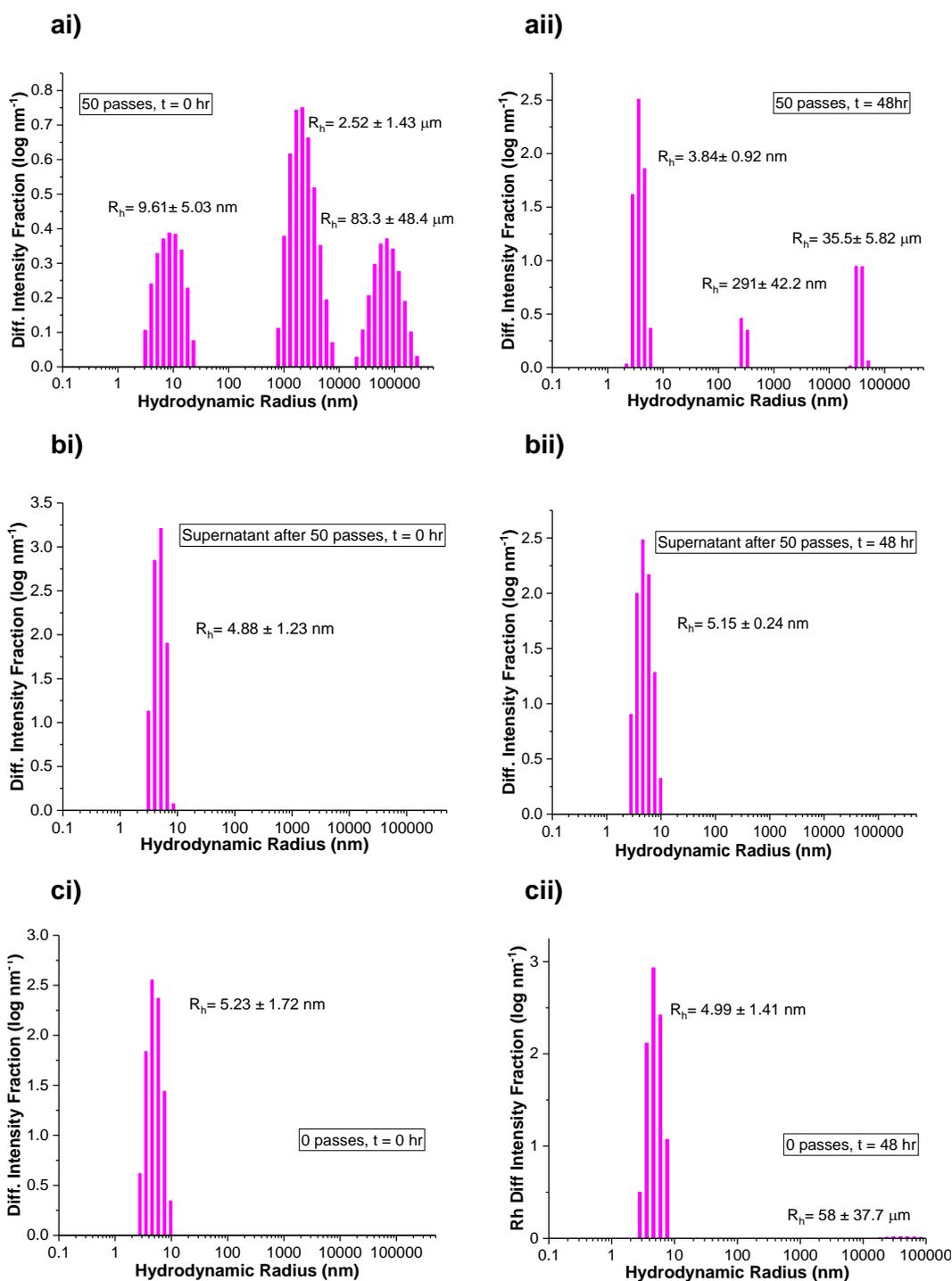


Figure 5.10. Regularisation analysis of DLS data for mAb1 a) after 50 passes, b) supernatant following 50 passes and centrifugation and c) quiescent mAb1. In i) samples were analysed immediately. In ii), samples were analysed after a 48hr incubation at 4°C. All samples were stressed at a plunger velocity of 8 mm s⁻¹. Protein concentration = 0.5 mg mL⁻¹. N= 2 for all samples, with one replicate shown above.

The data in Figure 5.10a show that the aggregates which form after 50 passes cover a large size range, as seen in Chapter 3 and in Section 5.3.2. After 48 hours, the largest aggregates in the tens of μm size range prevail, whilst those in the low μm size range revert to smaller oligomers and monomers. The intensity of the 'monomer' peak increases, as the large aggregates present in the sample no longer dominate the scattering signal. Both the clarified supernatant (Figure 5.10b) and the quiescent mAb1 (Figure 5.10c) samples show no signs of aggregation after 48 hours. These results would suggest that after the cessation of flow, the soluble aggregates which form can revert to smaller species over this timescale. The stressed sample did not appear to contain *more* aggregate after 48 hours, nor did the other samples. This suggests the EFD is acting as a catalyst for aggregation and that flow is necessary to propel the aggregation reaction forwards. If aggregates are present within the sample being stressed, can further aggregation be promoted? This forms the basis of the next section.

5.4.2 Investigating mAb-aggregate nucleation under flow

As mentioned in Section 1.1 and Figure 5.1, nucleation is a common component of protein aggregation mechanisms. It is thought that some proteinaceous aggregates can act as 'nuclei' for aggregation, triggering an acceleration in the rate of aggregate formation or its extent, compared to an aggregate reaction taking place in the absence of such nuclei. For example, it is known that amyloid fibrils can act as a 'template' for monomeric proteins, inducing the aggregation of the latter in the amyloid cascade (Törnquist et al., 2018). The large surface area of aggregates, which catalyse such 'secondary nucleation' events, led to the hypothesis that the larger aggregates, formed due to flow, can accelerate the aggregation reaction.

With this in mind, 0.5 mg mL^{-1} WFL and STT stocks were prepared in 150 mM ammonium acetate pH 6.0 and stressed for 20 and 100 passes, respectively, at a plunger velocity of 8 mm s^{-1} . The samples were then analysed with the pelleting assay, as described previously. $300 \mu\text{L}$ (from the $500 \mu\text{L}$ in the EFD) of one sample was kept on ice following stress, with the other $200 \mu\text{L}$ clarified with the pelleting assay protocol, thus generating clarified (C) and unclarified (U) 'seed' material. Seeding assays were performed by adding a 10% (v/v) spike of the relevant seed material to quiescent protein stock, then stressing the samples under the same conditions as above. These samples were then analysed, along with control samples, with the pelleting assay. In a separate set of experiments,

the supernatants of WFL and STT following flow were characterised using DLS as in Figure 5.10. Data for the above experiments are presented in Figure 5.11a) and b) for WFL and STT, respectively.

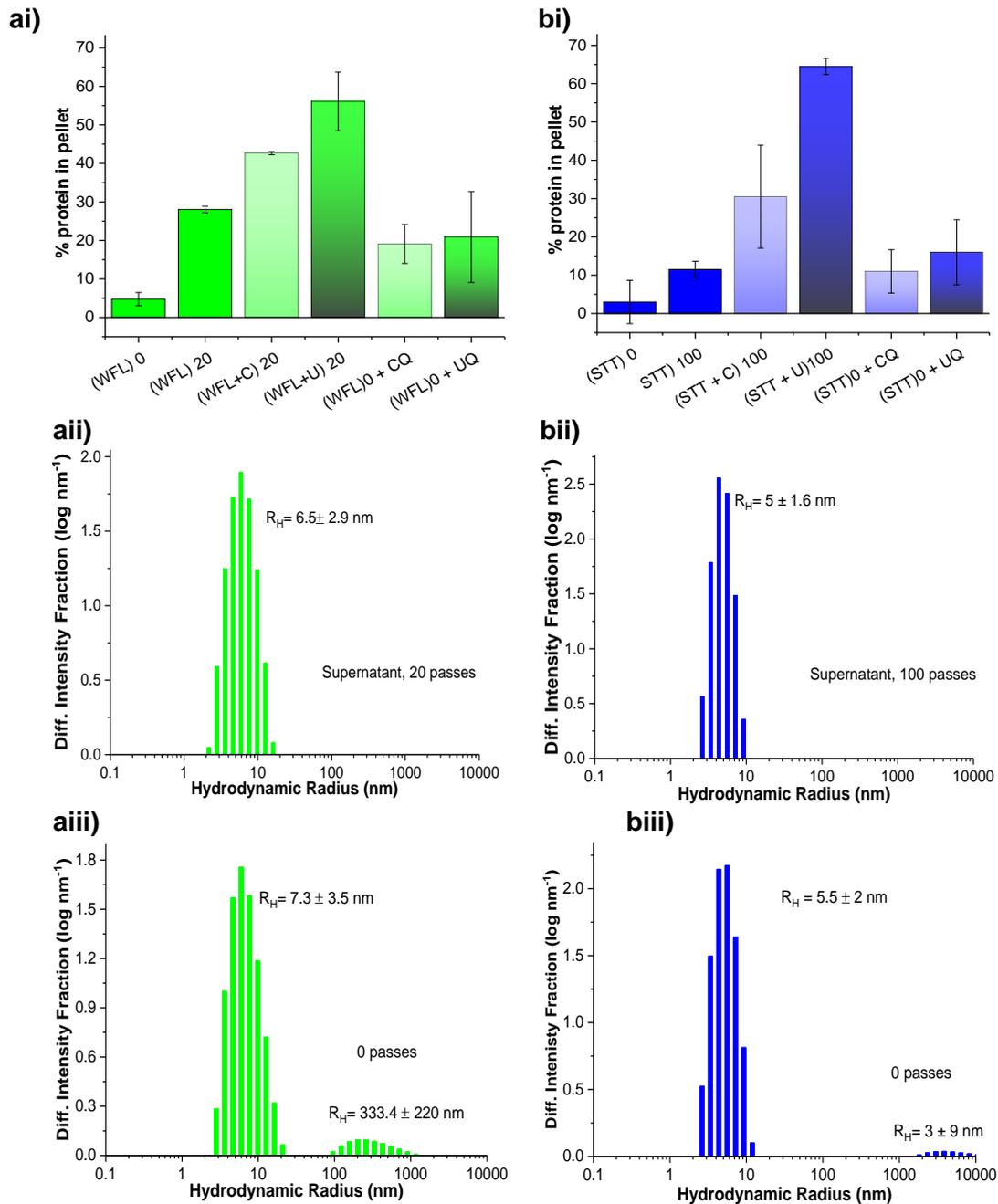


Figure 5.11. Seeding experiment and supernatant characterisation with a) WFL (green) and b) STT (blue). Protein concentration = 0.5 mg mL^{-1} and plunger velocity = 8 mm s^{-1} in all experiments. i) Seeding experiments (Section 2.2.16). No. of passes indicated outside brackets. C = clarified seed material, U = unclarified seed material, Q = quiescent (0 passes). Seed was added (10% v/v) to native protein stocks. All samples were analysed using the pelleting assay ($N = 1$, technical replicate s.d. shown). ii) Regularisation analysis of DLS data for WFL and STT supernatant following stress. iii) As in ii) but for Quiescent samples of WFL and STT. $N = 2$ for all DLS experiments, with one replicate shown. Pelleting data were obtained with Dr Amit Kumar, University of Leeds.

The complex data present in Figure 5.11 point towards which species are involved in nucleating the flow-induced mAb aggregation pathway. When supernatant is added as the 'seed' material to either WFL or STT (3rd bar in Figures 5.11 a and b), the level of aggregate quantified using pelleting approximately doubles. By DLS, this supernatant seed appears aggregate-free, i.e. no soluble aggregates were detected (a and b ii). This could be because the monomers in the supernatant fraction are less conformationally stable than their quiescent counterparts, causing the apparent increase in the amount of insoluble protein. Furthermore, the quiescent sample containing seed material did appear to have 3x more protein in the pellet fraction than the native sample, relative to the ~3% (v/v) aggregate added to the sample. This could indicate that seeds can induce some protein aggregation under quiescent conditions. Most interestingly, for both WFL and STT, when the entire aggregate ensemble is used as the seed material (designated 'unclarified' in the legend), the amount of protein in the pellet fraction appears to double for WFL and ~quadruple for STT (4th bar from left, Figures 5.11a and b respectively). This suggests that the larger aggregates, which form due to hydrodynamic stress, can drive the aggregation reaction further towards completion when compared to un-seeded reactions under the same flow conditions.

This dramatic effect on STT was somewhat unexpected given its general resistance to aggregation compared to WFL. It was hypothesised that the aggregates of either WFL or STT could be 'stickier', i.e. have more of a catalytic effect on the extent of aggregation of each protein. The aggregates from these proteins could also coalesce, as has been observed for some amyloidogenic proteins (Dubey et al., 2014).

To validate this hypothesis, cross-seeding assays were performed, using 'whole' stressed samples as the seed material (see Section 2.2.16). Additional control samples were prepared where each quiescent protein was added to the other, e.g. a 10% (v/v) addition of quiescent native STT was added to WFL and *vice-versa*. This was performed as it is known that some Abs can cross-aggregate (Jacobs et al., 2010). These experiments were performed under the same buffer and flow conditions as above. The data are shown in Figure 5.12.

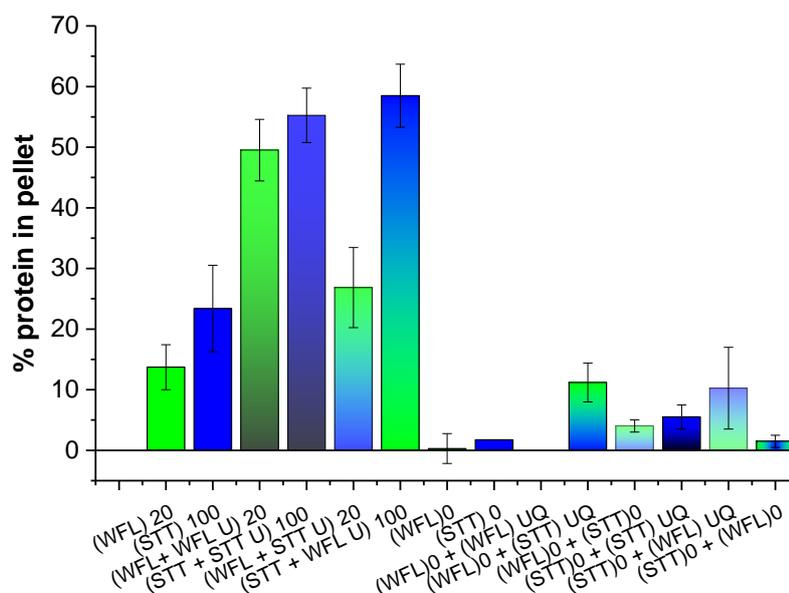


Figure 5.12. Cross-seeding assay of WFL (green) with STT (blue) in 150 mM ammonium acetate buffer, pH 6.0. All passes were performed at a plunger velocity of 8 mm s^{-1} , with a protein concentration of 0.5 mg mL^{-1} . Wherever '(sample) + x' is written, 'x' denotes the species being added 10 % (v/v) to the specified sample. The numbers outside brackets denote the number of passes, U = unclarified seed, Q = quiescent. Errors were propagated from $N=2$ experiments according to Section 2.2.18. Pelleting data were obtained with Dr Amit Kumar, University of Leeds.

The data in Figure 5.12 show that when WFL is seeded with itself {(WFL + WFL U) 20}, as opposed to with STT {(WFL + STT U) 20}, the levels of aggregation observed are 4x higher than without seeding. This is similar to the behaviour observed in Figure 5.11ai. STT appears to be very sensitive to seeding in general, as both WFL and STT seeds increase the amount of aggregate quantified after 100 passes (to ~60% in both cases, 4th and 6th bars from left). This appears to rule out the possibility of WFL seeds being 'stickier' than STT's; instead WFL prefers to self-associate rather than cross-associate. This provides further evidence for a specific flow-induced aggregation pathway for WFL vs a generic pathway for STT.

In Chapter 4, it was demonstrated that arginine+succinate buffer can suppress the flow-induced aggregation of both WFL and STT. WFL behaves in an STT-like fashion in this buffer compared to in ammonium acetate, as the arginine probably competes for WFL's specific binding interface. It follows that WFL's seeding and cross-seeding behaviour may appear like STT's if these experiments were

performed in this buffer. The obvious caveat to this experiment is the fact that observed aggregation is diminished to low levels, thus both proteins were stressed for 100 passes as in Figure 4.6, Section 4.4. The samples were then analysed as above. These data are presented in Figure 5.13, plotted on the same scale as Figure 5.12 to aid comparison.

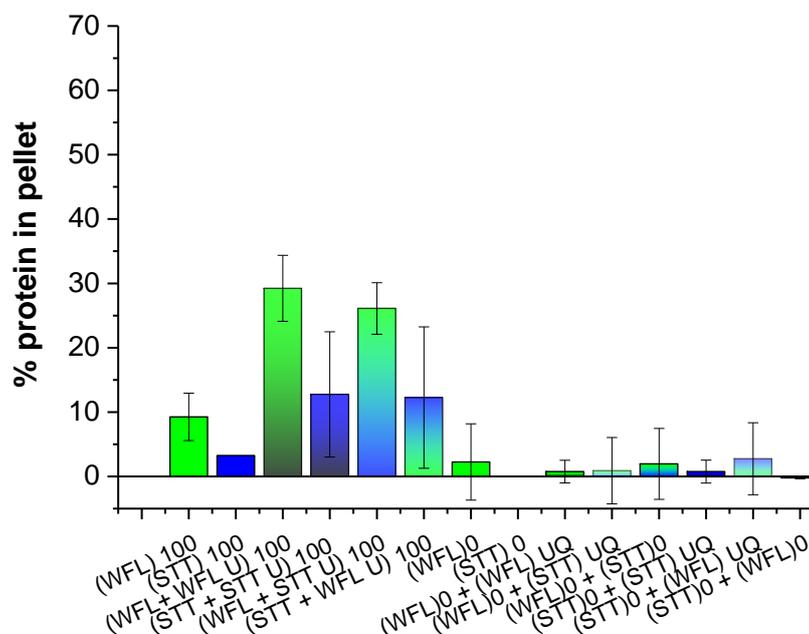


Figure 5.13. Cross-seeding assay of WFL (green) with STT (blue) in 125 mM L-arginine + 20 mM sodium succinate buffer, pH 6.0. All passes were performed at a plunger velocity of 8 mm s^{-1} , with a protein concentration of 0.5 mg mL^{-1} . Wherever '(sample) + x' is written, 'x' denotes the species being added 10% (v/v) to the specified sample. The numbers outside brackets denote the number of passes, U = unclarified seed, Q = quiescent. Errors were propagated from $N=2$ experiments according to Section 2.2.18. Pelleting data were obtained with Dr Amit Kumar, University of Leeds.

The data in Figure 5.13 show that less than 10% of the protein is rendered insoluble following 100 passes of stress in the EFD. When WFL and STT were stressed in the presence of their own, unclarified seed material, the measured levels of aggregate trebled and doubled, respectively, compared to the unseeded conditions. When WFL was cross-seeded with STT aggregates {(WFL + STT U) 100}, in arginine+succinate buffer, the level of aggregation observed is equivalent to the 'self-seeded' samples (~30% protein in pellet). STT's 'cross-seeded' samples are equivalent to its self-seeded (~15% protein in pellet), with both

seeded samples having ~3x the amount of aggregate as the unseeded sample. Little aggregation (<10% protein in pellet) is observed in any of the quiescent samples. These data provide further evidence of the ability of arginine to suppress the flow-induced aggregation of both WFL and STT, in particular, the potential role arginine plays in diverting the aggregation pathway of WFL under extensional flow to one more akin to that of STT.

In summary, through the addition of aggregated protein to quiescent, native samples, the amount of aggregate observed following extensional flow can be augmented. This response is dependent on both the protein(s) involved and the buffer environment. These data demonstrate the role of pre-formed seeds driving the flow-induced aggregation of mAbs.

5.5. Discussion

5.5.1 N*-state characterisation

This Chapter set out to gain experimental evidence for the mechanism of flow-induced aggregation for the model IgG1 antibodies used throughout the study (WFL, mAb1 and STT). The pathways proposed in Figure 5.1 are based on the data obtained for BSA in Chapter 3, where the conformational remodelling of the protein, subsequent aggregation into small oligomers and eventually insoluble aggregates was monitored. The recent study of Wolfrum et al. (2017) built on the work of Simon et al. 2011, using a four-roll mill device to subject mAbs to hydrodynamic forces. The onset of unfolding and aggregation of the protein was monitored using the environmentally sensitive extrinsic fluorophore Nile Red (Wolfrum et al., 2017). Inspired by this, the dye was used to attempt to capture the N* aggregation-prone state of WFL and STT induced by extensional flow. This proved to be very irreproducible, perhaps due to the morphology of the aggregates formed in the EFD inefficiently binding the dye, thus not producing a significant fluorescence increase. Upon closer inspection of the Wolfrum et al., 2017 data, the error bars for their fluorescence measurements are very wide. A gradual decrease in fluorescence was also observed when fluorescence measurements were recorded on the WFL and STT samples. This could have been due to precipitation of the dye and/or the protein from solution, which may also account for the errors in Wolfrum's *ex situ* data (Wolfrum et al., 2017). Furthermore, the failure of ANS to capture this state meant that an alternative approach was necessary.

Chemical cross-linking is gaining ground in the structural mass spectrometry field as a means of capturing transient protein-protein interactions, such as those formed by aggregation-prone proteins (Iacobucci and Sinz, 2017; Martin et al., 2018). The bis-maleimide cross-linking reagent BS³ has been used previously to identify the aggregation interface of WFL in solution under quiescent conditions (Figure 5.2, Devine, 2016; Dobson et al., 2016). If extensional flow causes the conformational remodelling of proteins, then perhaps this interface would be different when the cross-linking reaction was performed under flow. Despite attempts to adapt the protocol to the EFD, the only cross-links identified by tandem mass-spectrometry were either dead-end or *intramolecular*. When this cross-linking strategy was originally applied, attempts to try and improve the abundance of *intemolecular* cross-links included separation of the cross-linked peptides by size-exclusion chromatography and increasing the cross-linker concentration (Devine, 2016). These efforts still led to the identification of just one cross-link between WFL species. As mAbs are large proteins (150 kDa), the isolation of a specific, cross-linked peptide for tandem mass spectrometry analysis is practically difficult. Improved mass detection methods such as the Orbitrap (Makarov et al., 2006), as well as the use of different cross-linking chemistries, may help facilitate the identification of the flow-induced cross link in future (see Chapter 7).

5.5.2 Tracking aggregation as a function of pass number

Learning the minimum number of passes necessary to observe the aggregation of mAbs could be achieved through a range of strategies. Monomer quantification using HPLC is the analytical workhorse of the biopharmaceutical industry (Philo, 2006). This approach has also been used to monitor the loss of monomer (and in some case, oligomer formation) over time to solve aggregation mechanisms of monoclonal antibodies. This approach was not straightforward for the three mAbs explored thus far because: WFL interacted unfavourably with the chromatography matrix, despite the inclusion of arginine + succinate to minimise this possibility; the quality of the columns used deteriorated with time (note the peak shapes obtained for mAb1 in Figure 5.3b vs those for STT in Figure 5.3c); the number of clarification steps to remove the large aggregates which form due to flow prior to analysis; the inclusion of an additional quench step which may have adversely

affected the data quality. These issues have long been of concern in the field (Carpenter et al., 2010). Despite these issues, data for the loss of monomer as a function of time was obtained. The selection of a 'time' axis for extensional flow is difficult for the reasons stated in Chapter 3.1; proteins spend 18 μ s in the extension region, 40 ms in the shear region of the capillary and \sim 5 s in the syringe barrel during one pass. Furthermore, there is a small delay after one pass before the next one takes place (\sim 1 s when the plunger velocity is 8 mm s⁻¹), in addition to mixing when the protein solution enters the empty syringe. Considerations will have to be made about these different time regimes when fitting future kinetic data to models, i.e. whether to define time in region x of the EFD (in seconds) or as the number of passes.

Semi-denaturing agarose gels have been used to determine oligomer stoichiometry (Halfmann and Lindquist, 2008), as well as estimate the molecular weights of large proteins following enzymatic cleavage under shear flow (Lippok et al., 2016). Adapting these strategies for acrylamide gels, SD-SDS-PAGE could identify both the onset of aggregation for the mAbs (10 passes for all three mAbs), and differing electrophoretic mobility amongst the monomers, dimers and trimers observed. All three mAb monomers (WFL, mAb1 and STT) run at different apparent molecular weights on an SD-SDS-PAGE gel. This could be due to the proteins possessing different local conformations. It could also be due to the interactions between the proteins themselves and the polyacrylamide gel matrix (Hames, 1990). The distinction between these two phenomena would require further investigation. The aggregates appear to grow in a linear fashion, i.e. 1^{mer}, 2^{mer}, 3^{mer}...) rather than through dimeric states (2^{mer}, 4^{mer}, 6^{mer}...). This is in agreement with previous data obtained for WFL (Devine, 2016), as well as for other mAbs in the literature (Li and Roberts, 2009; Nicoud et al., 2014; Singla et al., 2016).

The onset of aggregation and its subsequent evolution was tracked in more detail using dynamic and static light scattering. The intensity of light scattered by a macromolecule increases by a 6th power with the size of the scatterer, e.g. a 60 nm protein aggregate would scatter light at one million times the intensity of a 6 nm mAb monomer (Filipe et al., 2010). Therefore, when the z-average radii of the mAbs measured by DLS increase as a function of pass number, the sensitivity of DLS allows the onset of aggregation to be easily identified. mAb1 and WFL were most sensitive to extensional flow, with their measured aggregates plateauing at

~8 μm after 20 passes, compared to STT. This was when the proteins were stressed at a concentration of 0.5 mg mL^{-1} .

Static light scattering has been used by many in the mAb aggregation kinetics field to monitor the mass increase of aggregates as a function of time (Arosio et al., 2012; Barnett et al., 2015; Kalonia et al., 2016). In our case, 'time' can be considered as the number of passes through the EFD. Just as DLS has caveats surrounding larger macromolecules, so too does SLS. The amount of light scattered by a macromolecule scales linearly with mass (see Section 1.3.4). When the uncertainty-averaged molecular mass of a sample is determined, the more aggregates present in the sample, the greater the overall magnitude of this value. At 0.5 mg mL^{-1} , the uncertainty-averaged molecular mass increased as a function of pass number in the same fashion as the DLS data. The end-point masses for all three mAbs varied as a function of concentration: $0.5 \text{ mg mL}^{-1} \approx 4 \text{ MDa} \approx 27^{\text{mer}}$ and $0.25 \text{ mg mL}^{-1} \approx 11 \text{ MDa} \approx 73^{\text{mer}}$ (data not shown). This plateauing effect could be due to: an artefact of the instrument (only three scattering angles are present, compromising the accuracy of the fit compared to a 15-angle goniometer, e.g. (Arosio et al., 2012)); the species measured at this point represent the maximum size of the soluble aggregates present, with the coalescence of these forming insoluble aggregates. The aggregates formed at 0.5 mg mL^{-1} break up under shear flow, as has been proposed for various systems (Hoare et al., 1982; Ren et al., 2015; Sadegh-Vaziri et al., 2018). The Smoluchowski equations describe the kinetics of amorphous aggregate growth (Arosio et al., 2012; Schmitt et al., 2000; Zidar et al., 2018). If more data were obtained tracking flow-induced mAb aggregate growth, then fitting these equations to the data could help obtain a rate-constant for soluble aggregate growth of the mAbs triggered by extensional flow.

The Roberts group in Delaware have developed an array of experimental signatures typical of amorphous proteins when they aggregate, depending on whether or not their aggregation is dominated by unfolding, nucleation and so on (Li and Roberts, 2009). These signatures, identified through monomer loss assays and light scattering measurements, could possibly account for the mechanistic reason behind the observations made in this chapter. Further work is necessary to prove this.

5.5.3 Characterising nucleation in the EFD

The identification of species which nucleate a protein aggregation pathway is of interest to many researchers in the field (Silva et al., 2017; Törnquist et al., 2018). The initial use of mAb1 as a surrogate for WFL and STT, based on its intermediate behaviour, confirmed that the size distributions of aggregates do change over a long time period, with large (tens of microns in size) aggregates remaining stable, whilst smaller aggregates dissociate to smaller species. Most importantly, the presence of aggregates left quiescently in this manner does not induce *more* aggregation in a sample – flow is necessary to catalyse this. Furthermore, these experiments showed that stressed samples clarified using centrifugation do not aggregate over a two day period, nor do the quiescent samples. This could have implications for future experiments (see Chapter 7). The subsequent seeding and cross-seeding experiments performed with WFL and STT showed that aggregated protein can nucleate extensional flow-induced aggregation; this is the first demonstration of this to this author’s knowledge. These experiments also provide further evidence for WFL’s avidity for its hydrophobic ‘WFL’ motif, as its aggregation is influenced to a greater extent by *its* seeds rather than those of its 99.6% sequence relative STT. Arginine could be successfully out-competing WFL monomers for this motif, moving its aggregation pathway from the top route in Figure 5.1 to the pathway traversed by STT.

5.5.4. The EFD mechanism data in the context of established aggregation models

The data in Section 5.2 do not provide conclusive, direct evidence for the induction of protein aggregation under flow in terms of the formation of an aggregation-prone activated N^* state. However, the fact that the quiescent samples are shown to be free of aggregate by many orthogonal assays means that one can still conclude flow acts a catalyst of mAb aggregation. The mAbs aggregate to form oligomers and soluble aggregates following 5–20 passes through the EFD, depending on the protein and the technique used to detect aggregation. The absolute size and mass of these aggregates, as well as how they grow, is possibly concentration-dependent. If so, one or many steps in the aggregation-pathway (containing a k'_{obs} term) would be very sensitive to concentration. Further work would be necessary to understand this.

The more passes the mAb solution experiences, the more the aggregation pathway in Figure 5.1 is driven to the right, i.e. away from monomer towards the formation of soluble and insoluble aggregates. Allowing the system to reach equilibrium, it is clear that soluble aggregate formation on the pathway is reversible. Adding aggregated material in the form of seeds, in the presence of flow, propels the pathway towards the right, as more aggregate is observed in such seeded samples than those without seeds. In addition, it has been proposed that proteinaceous aggregates and other colloids can readily adsorb to gas-liquid and liquid-solid interfaces, with layers then being sloughed from these surfaces due to shear flow (Bee et al., 2011; Bee et al., 2012; Grigolato et al., 2017; Koepf et al., 2018; Sediq et al., 2016). If these aggregates broke up under flow (Sadegh-Vaziri et al., 2018), and these species were able to nucleate the pathway, then the impact on the aggregation kinetics of mAbs in such environments may be profound (Törnquist et al., 2018). These proposed models are summarised in Figure 5.14.

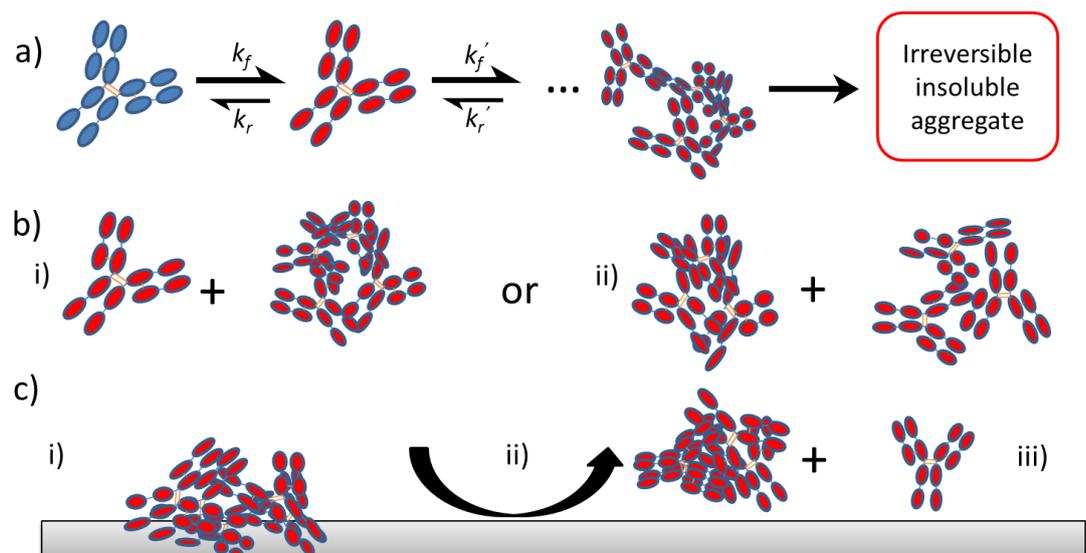


Figure 5.14. Proposed aggregation models in the EFD. a) Extensional flow events promote the formation of aggregation-prone species (red) from native monomers (blue) (cf. Figure 5.1). b) 'Seed' material could incorporate mis-folded monomers (i), or the aggregates could themselves coalesce, accelerating the formation of insoluble aggregate (ii). c) Large aggregates could interact with the glass surfaces of the EFD (i). During an extensional flow event (one pass), aggregates could be sloughed from the surface (ii). This could facilitate the accelerated aggregation phenomena discussed in b (iii).

Many more experiments are necessary to fully understand this pathway. The large body of data gathered using the EFD to look at how hydrodynamic forces induce protein aggregation could be used to inform the development of next generation mAb therapies. This applicability forms the subject of Chapter 6.

6. Screening the aggregation propensity of therapeutic mAbs using the extensional flow device

6.1 Overview of the Jain, Sun et al. *Proc. Natl. Acad. Sci.* (2017) Adimab paper

As discussed in Section 1.2 and throughout this thesis, biopharmaceuticals currently approved or in development, are dominated by the monoclonal antibody scaffold. During the drug development pipeline, the identification of mAbs with desirable biophysical properties is of the utmost importance, for the reasons discussed in Figure 1.6, Section 1.2. For small-molecule therapies, such ‘desirable’ properties were defined in a seminal paper by Lipinski et al., leading to the ‘Rules of five’. These state that an oral medication should typically have: 5>hydrogen-bond donors; Molecular weight < 500 Da; log P< 5; 10> hydrogen-bond acceptors (Lipinski et al., 1997).

However, no such sets of drug discovery ‘rules’ exist for biological therapies, except for guidelines surrounding the ‘quality’ of the molecule under the QbD framework (Section 1.4). Generally, a desirable mAb would be: easy to express in mammalian systems; highly soluble; stable in terms of both colloidal and conformational stability; highly specific for its target and non-immunogenic *in vivo*.

With this in mind, a team of researchers at Adimab LLC (a spin-out from MIT founded by Prof Dane Wittrup) set out to define a ‘Lipinski’s Rules of five’ for mAbs. At the time of writing the paper (accepted Dec 2016), the team at Adimab had characterised the 137 mAb-based therapies that were either approved, in Phase-III or in Phase-II clinical trials. The workflow was as follows: Take the variable sequences (V_H and V_L) of the above mAbs; Graft the variable sequences (V_H and V_L) of the 137 mAbs into a common IgG1 scaffold; Over-express the sequences in HEK293 cells; Purify the mAbs by Protein-A chromatography; Formulate the mAbs in HEPES-buffered saline and finally, characterise the mAbs using twelve different biophysical techniques. These techniques were chosen by Adimab as they have at least one literature reference and use less than 1 mg of material per assay (Jain et al., 2017). The methods selected are summarised in Table 6.1.

Table 6.1 Summary of the twelve techniques used to analyse the 137 mAbs in the Jain et al. paper.

Technique (Key reference where applicable)	Overview	Biophysical property probed
HEK Titre	The amount of mg L ⁻¹ of protein expressed in the HEK293 cell line.	Aggregation-propensity of the sequence.
CIC (Cross-interaction chromatography) (Jacobs et al., 2010)	Polyclonal antibodies conjugated to column matrix. mAbs of interest loaded onto column.	Poor specificity of mAb of interest.
PSR (Polyspecificity Reagent) binding assay (Xu et al., 2013)	Biotinylated membrane proteins used as PSR. IgG-presenting yeast, similar to those in Section 1.2.3 incubated with these samples. MFI used to quantify substrate binding.	Poor specificity of mAb of interest.
CSI-BLI (Clone Self-Interaction by Biolayer Interferometry) (Sun et al., 2013)	Human IgG loaded onto a biosensor. mAb of interest incubated for 5 mins, with association measured using the Octet system.	Aberrant mAb-mAb interactions.
AC-SINS (Affinity Capture-Self-Interaction Nanoparticle Spectroscopy) (Liu et al., 2014)	Gold nanoparticles conjugated to goat anti-human IgG Fc fragments and polyclonal goat mAbs. Particles incubated with mAb of interest. Large wavelength shift gives a negative result.	Aggregation-propensity of the mAbs of interest.
SG (Salt-gradient) AC-SINS (Estep et al., 2015)	Gold nanoparticles (as above) incubated with mAb of interest for 30 mins. Samples then diluted into 0.3–1M ammonium sulphate. 560 nm shift plotted against salt concentration.	Aggregation-propensity of the mAbs of interest.

T_m determination using DSF (Differential Scanning Fluorimetry) (He et al., 2011)	mAb of interest mixes with SYPRO orange. Plate scanned from 40°C to 95°C at 0.25°C min ⁻¹ . Fluorescence followed as a function of time to allow Fab T _m to be assigned in BioRad software.	Thermal stability of the mAb of interest.
SMAC (Standup Monolayer Adsorption Chromatography) (Kohli et al., 2015)	Zenix SEC-300 column features a monolayer of silica. Longer retention times correlate with poor colloidal stability of the mAb.	Colloidal stability of the mAb of interest.
AS (Accelerated stress) SEC Slope	1 mg mL ⁻¹ mAb samples incubated for 30 days at 40°C. Various time points taken, with samples measured by gel-filtration chromatography. The amount of aggregate was quantified and the gradient calculated.	Aggregation-propensity / thermal stability of mAb of interest. Important in shelf-life prediction.
HIC (Hydrophobic Interaction Chromatography) (Estep et al., 2015)	mAb of interest spiked into 1M ammonium sulphate buffer. Samples loaded onto butyl-NP5 HIC column. Retention times noted.	Hydrophobicity of the mAb of interest.
ELISA (Enzyme-linked Immunosorbent Assay) (Mouquet et al., 2010)	The antigens Cardiolipin, keyhole limpet haemocyanin, lipopolysaccharide, ss- and ds-DNA and insulin were coated onto ELISA plates overnight. Blocking and wash steps then performed. mAb of interest added to plate. mAb-antigen binding detected using anti-human IgG-HRP conjugated mAb.	Promiscuous mAb binding.
BVP (Baculovirus particle) assay (Hötzel et al., 2012)	As above, except baculovirus particles were used in place of the antigens listed for the ELISA assay.	Unfavourable pharmacokinetics of the mAb of interest, i.e. rapid clearance <i>in vivo</i> .

Following analysis of the data generated from analysing the 137 mAbs with the methods in Table 6.1, it was apparent that some mAbs performed unfavourably in some assays, with the majority of mAbs performing favourably. When the data obtained for each technique were plotted on a frequency vs measured assay value, most of the techniques yielded a long-tailed distribution, with the tails pointing in the unfavourable direction, e.g. Figure 6.1a. (Jain et al., 2017). The only assays showing a Gaussian distribution of mAbs were the HEK titre and T_m assays (Figure 6.1b). To understand the relationship between each assay, the data were hierarchically clustered and subjected to Spearman's rank correlation analysis. This produced a 'family tree' of assays, grouped by relatedness (Figure 6.1c).

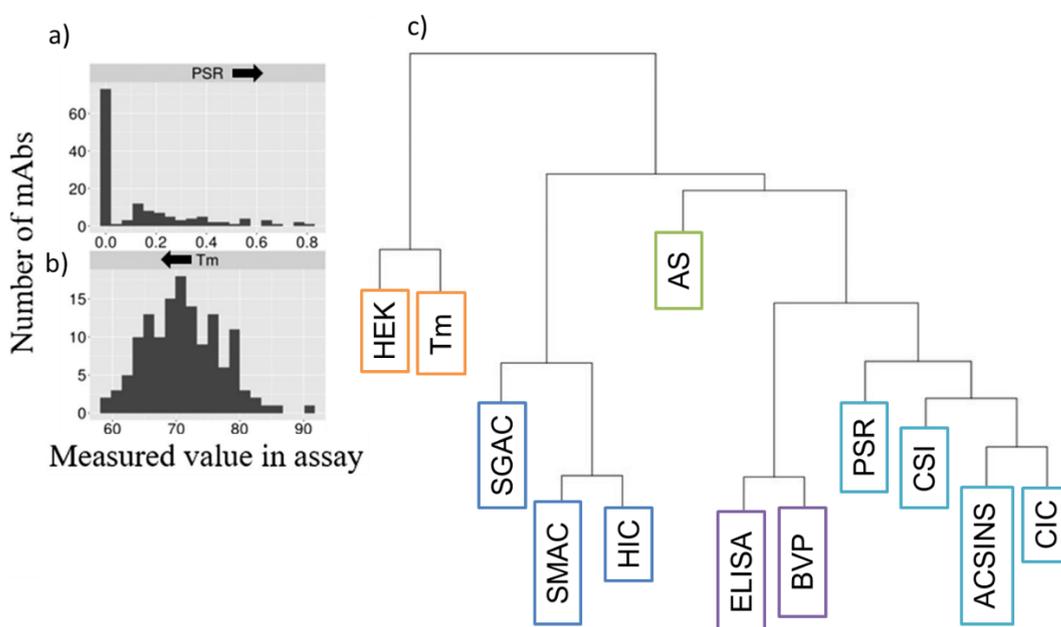


Figure 6.1. Hierarchical clustering tree of biophysical assays in the Jain et al. paper. a) Example of a long-tailed distribution for an assay, in this case PSR. b) Example of a Gaussian distribution for an assay, in this case the T_m . c) The assays grouped together by colour were found to be statistically related (see text). The HEK/ T_m family is most distinct from the PSR family of assays according to their Spearman's rank order correlation.

Adapted from Jain et al. 2017.

The clustering found that mAbs which generally expressed well in HEK had a stable Fab as measured by T_m (orange boxes). The other branches probed different biophysical properties of the proteins: self-association and adsorption to chromatography matrices (dark-blue boxes). Whilst in its own family, accelerated stress (green box) was found to be more closely linked to assays probing non-specific association (purple and light-blue boxes) than self-interaction.

In further analysis, because CHO cells are typically used for mAb production, the orange assay group was discarded. To derive their 'Rules of five' for mAbs, Jain et al. were interested in looking at how the approved mAbs in the dataset (48/137) behaved in the different assays. To do this, each assay was assigned a threshold value which, if exceeded, suggests the protein displays non-ideal behaviour. For example, a retention time in the HIC assay exceeding 11.7 ± 0.6 min would be indicative of a poorly performing mAb (full list of values shown in Appendix Table A3). Exceeding one of the thresholds from any assay within a cluster would lead to a mAb being given a 'red flag'. Jain et al. noted that generally, the approved mAbs possessed fewer red flags than those in Phase-III and Phase-II (Jain et al., 2017). The mAbs were then further scrutinised and compared to one another through computing their pairwise distances (Section 2.2.17) to produce a cluster diagram. This is shown in Figure 6.2

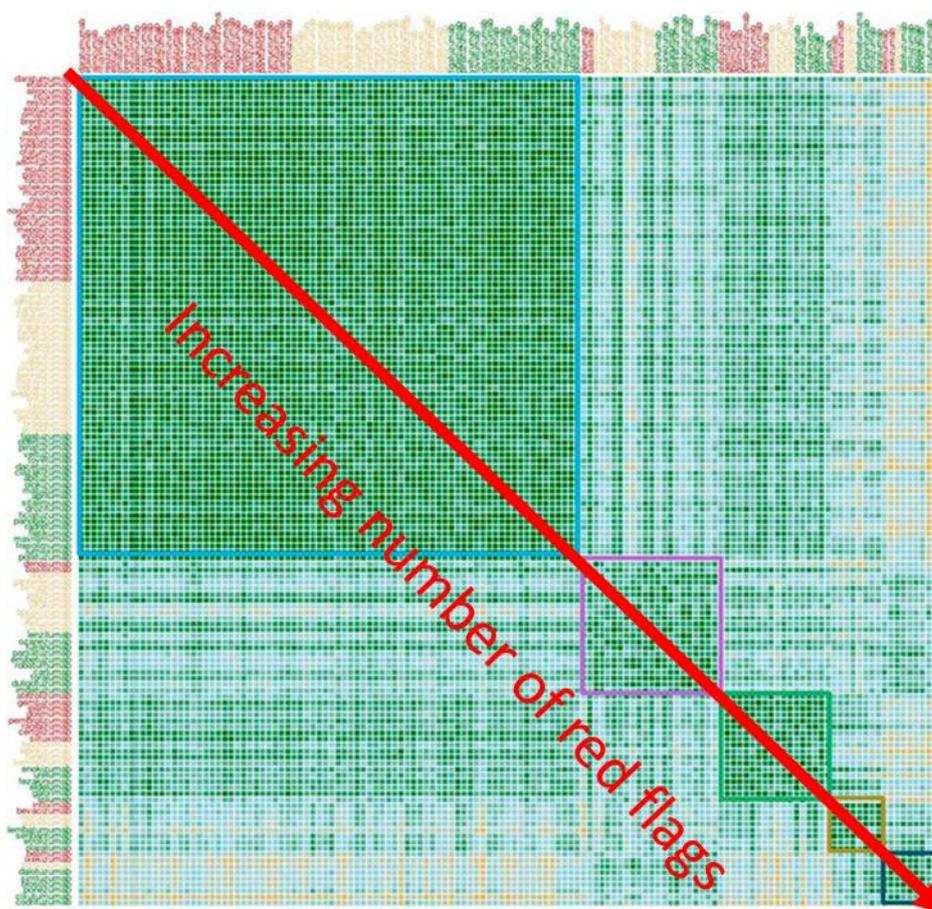


Figure 6.2. Cluster diagram of the 137 mAbs according to their relatedness in the biophysical assays. On the axes, Approved mAbs are labelled in red, Phase-III in yellow and Phase-II in green. Dark green squares in the plot = highly related mAbs, yellow = poorly related. The mAbs group into 5 Clusters (coloured squares). Taken from Jain et al. 2017 and appended with arrows.

The data in Figure 6.2 show that the mAbs form five ‘clusters’ of relatedness. The largest such cluster (top-left corner) contains 81 of the 137 mAbs, which possess the most similar biophysical properties in the four groups of assays. This group includes 34 of the 48 approved mAbs. As one moves along the diagonal indicated by the red arrow, the number of red flags assigned to a mAb increases, whilst the number of approved mAbs in a cluster decreases. In spite of the caveats in this reference dataset (see Discussion, Section 6.5), Jain concluded that the results and analysis presented here will serve as a tool for industrial and academic scientists alike to aid the development of next-generation mAb therapies, based on the cut-offs defined above.

With this reference dataset in mind, a collaboration was established with Adimab to answer the following:

- Can the EFD distinguish between mAbs both within and between clusters, using the % protein in pellet (following ultracentrifugation) as a readout?
- Are approved mAbs more resilient to hydrodynamic forces than those in Phase-II?
- How does the EFD compare to the assays and groups defined in Figure 6.1? Could the EFD replace any/all of the assays used by Jain et al.?
- Can emerging bioinformatics tools help rationalise the aggregation data generated from the EFD on these mAbs?

The answers to these questions form the basis of the rest of this Chapter.

6.2 Selection of a new mAb family from the Adimab dataset

Subjecting all 137 mAbs to stress in the flow device is far beyond the scope of this project. In order to select an appropriate number of mAbs for stress in the flow device, a rank order of the mAbs was produced by Tushar Jain (Adimab, LLC) using a 'distance from ideal' analysis of the red flag data from Jain et al. 2017 (see Section 2.2.17). In brief, an ideal mAb's measured values in the biophysical assays from Figure 6.1 would be the top values in the favourable direction, for example, the lowest retention time on a HIC column. A mAb from the Adimab dataset closest to this 'ideal' mAb would thus attain a rank of 1/137. Conversely, a mAb which performed the worst in the assays would be the furthest away from ideal and thus attain a rank of 137/137. The rank orders produced in this fashion were plotted against the numerical order of each mAb in Figure 6.2. The mAbs were colour-coded according to cluster number, with the data shown in Figure 6.3.

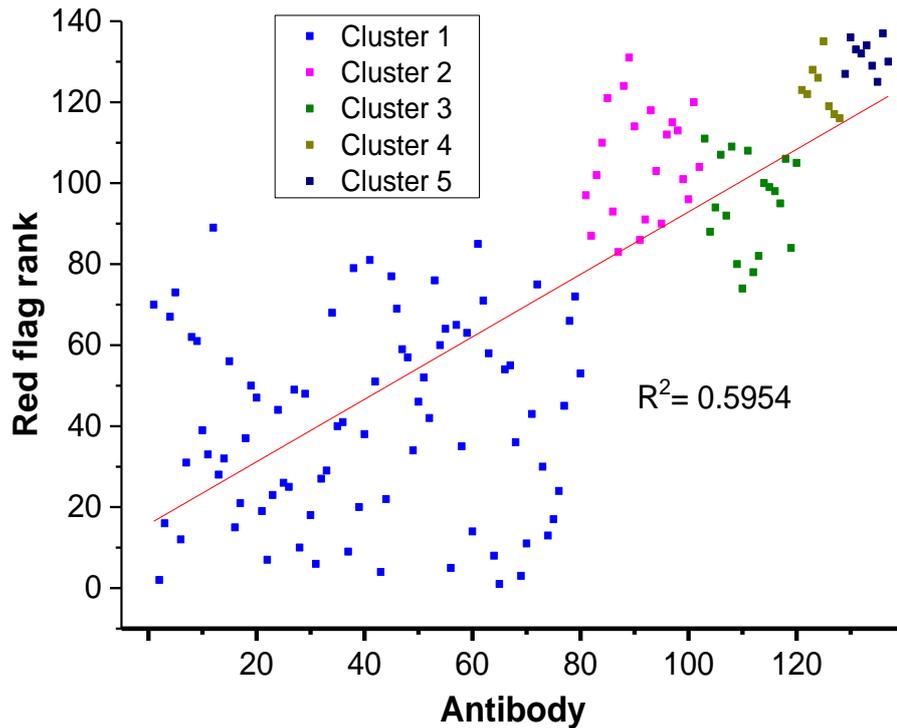


Figure 6.3. Plot of ‘distance from ideal’ red flag rank against protein number from Figure 6.2. The five clusters are colour-coded according to Figure 6.2, with Cluster 1 (blue) being in the top-left corner and Cluster 5 being in the bottom-right corner. A linear trend line was fitted to the data in Origin Pro to guide the eye.

The data show that as the cluster number increases, the red flag rank number assigned to each mAb from the Jain et al analysis also increases. This would agree with the data in Figure 6.2, as the smaller clusters (Clusters 4 and 5) generally possess fewer approved antibodies, more Phase-II antibodies and more mAbs with red flags than the larger clusters (colour-coding on axes of Figure 6.2). There is a large spread in rank amongst the antibodies in Cluster 1, possible due to the large number of mAbs (80/137) in this cluster. Conversely, the spread in rank is very similar between Cluster 2 and 3, as well as between Clusters 4 and 5.

Rather than analyse all 137 mAbs, an initial set of 50 were selected from Figure 6.3 of differing rank and cluster number. The list was sent to Adimab, who selected the final set of 33 mAbs to be tested with the EFD: 15 from Cluster 1, 6 from Cluster 2, 5 from Cluster 3, 4 from Cluster 4 and 3 from Cluster 5. The cluster numbers of the mAbs, alongside bioinformatics data, are in Appendix Table 1.2.

6.3 Screening the Adimab IgGs using the extensional flow device

6.3.1 Generating the EFD pelleting data

In the Jain et al. paper, the mAbs were formulated into 25 mM HEPES + 150 mM sodium chloride, pH 7.3. Whilst some assays required different running buffers etc. to work, the decision was made to stress the Adimab proteins in this formulation buffer. Initially, a series of ‘ranging shot’ experiments were performed with Daratumumab. This was chosen as it is an approved mAb from Cluster 1, thus it was hypothesised that it would be resistant to extensional flow, like STT from Chapters 3–5. To validate this hypothesis, Daratumumab was stressed for 0, 20, 50 and 100 passes at a plunger velocity of 8 mm s^{-1} and a protein concentration of 0.5 mg mL^{-1} . Following this, the samples were analysed using the insoluble protein pelleting assay. The data are shown in Figure 6.4.

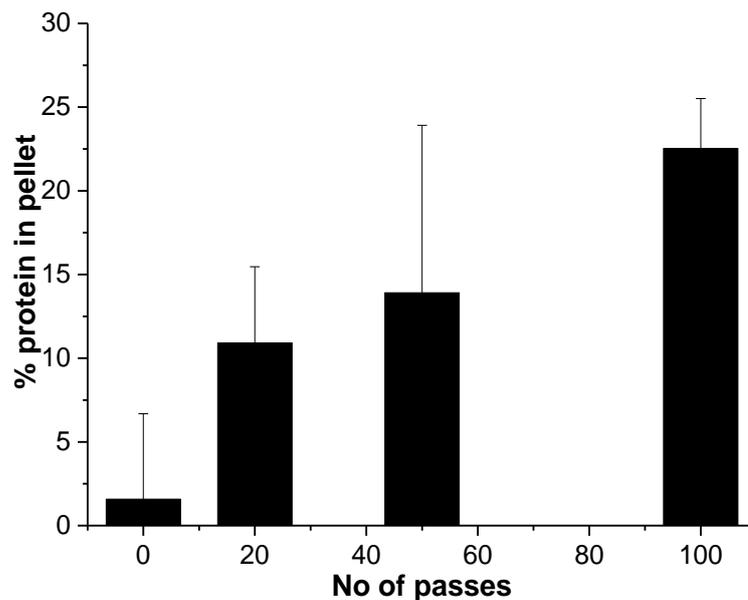


Figure 6.4. ‘Ranging shot’ for Daratumumab. The protein was stressed at a concentration of 0.5 mg mL^{-1} for 0 (quiescent control sample), 20, 50 and 100 passes at a plunger velocity of 8 mm s^{-1} . The samples were then analysed using the pelleting assay. Errors from two independent experiments were propagated according to Section 2.2.18.

The data confirm that Daratumumab is fairly resistant to extensional flow, with ~20% of the protein being rendered insoluble after 100 passes in the EFD. This is similar to how STT behaved in the flow device when stressed in ammonium acetate (~18%, Chapter 4). Assuming that the other mAbs in this cluster would yield similar data, it was decided to double the number of passes to 200, to try and increase the dynamic range observed. At first, Daratumumab, Tabalumab, Denosumab, Ipilimumab and Ixekizumab were stressed for 200 passes under the same flow and buffer conditions as above. The remaining 28 mAbs were then selected at random, according to Section 2.2.17. Quiescent samples were incubated at room temperature for 20 minutes. All experiments were performed in duplicate, with the samples analysed using the insoluble protein pelleting assay. The data for all 33 mAbs are displayed in Figure 6.5.

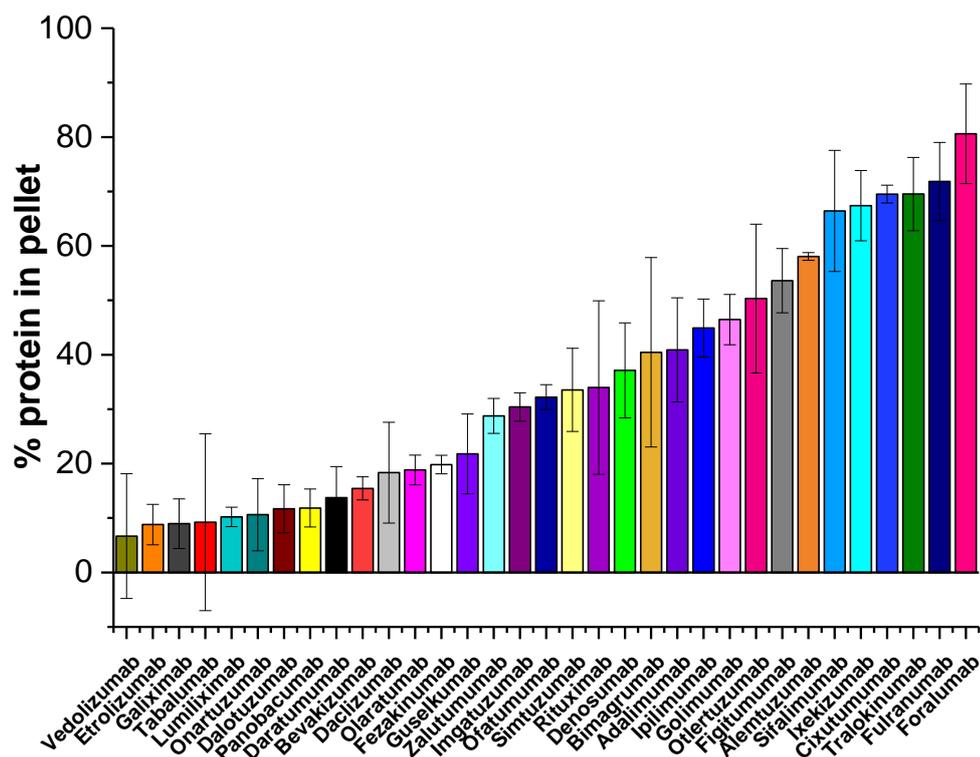


Figure 6.5. Plot of the percentage insoluble protein following stress in the EFD. The 33 Adimab proteins were stressed in the EFD for 200 passes at a plunger velocity of 8 mm s^{-1} in 25 mM HEPES + 150 mM sodium chloride, pH 7.3. Samples were analysed using the insoluble protein pelleting assay. 0–5% aggregate was seen in all quiescent samples (data not shown). Errors from two independent experiments were propagated according to Section 2.2.18. Pelleting data were collected with the assistance of Dr Amit Kumar, University of Leeds.

The data show that the mAbs aggregate to widely differing extents when subjected to stress in the EFD, with Vedolizumab yielding ~10% and Foralumab ~80% protein in pellet, respectively. Visually, there appears to be three different aggregation response levels to stress in the EFD: low (0–20%), medium (20–60%) and high (>60%). In the Jain et al. paper, long-tailed distributions were observed for 10/12 of the assays employed to interrogate the biophysical properties of the mAbs (Section 6.1). To assess the statistical relationships in the pelleting results, the data from Figure 6.5 were plotted as a histogram, with a bin size of 10% protein in pellet. These data are displayed in Figure 6.6. Further statistical analysis was also performed, which is displayed in Figure 6.7.

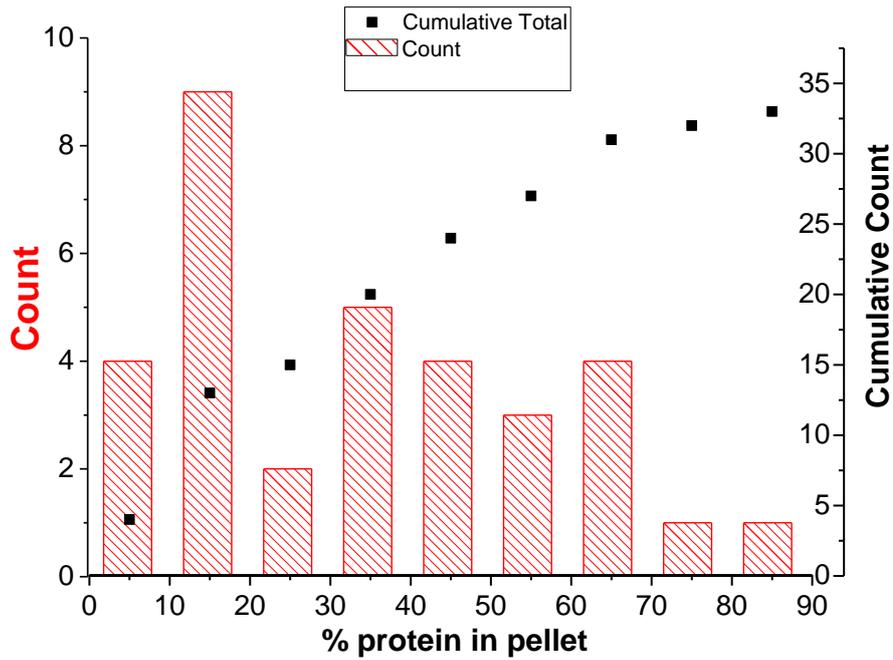


Figure 6.6. Histogram and cumulative frequency diagram for the pelleting data in Figure 6.5.

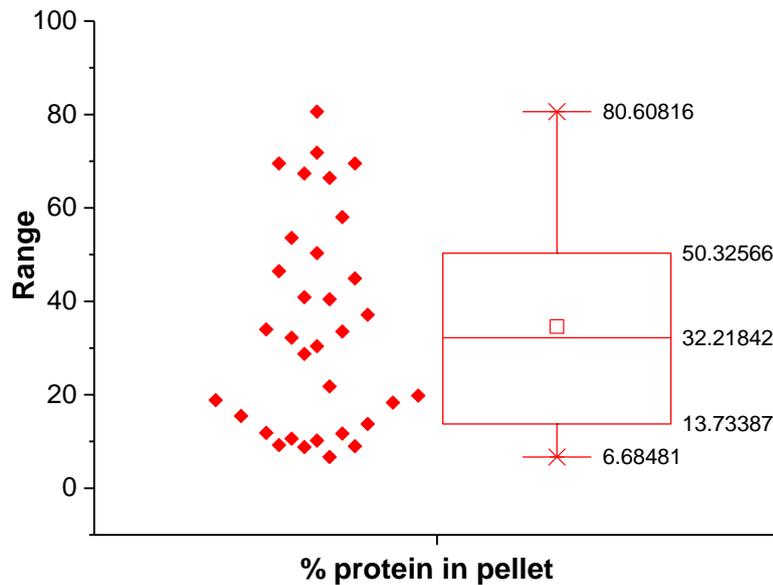


Figure 6.7. Box and whisker diagram showing the common statistical features in the pelleting data from Figure 6.5. The data points from Figure 6.5 are shown as red diamonds on the left. The whiskers of the box represent the 99% (top) and 1% (bottom) quartiles. The box comprises the 75%, median and 25% quartiles. The mean is denoted by the box in the centre.

Statistical analysis of the data shows that the % protein in pellet following stress has a long-tailed distribution, i.e. it is not Gaussian. The approval rating of the mAb appeared to not affect the level of aggregation seen (Appendix Figure A5). Whilst the pelleting data for some mAbs were quite variable between replicates, the average propagated error between replicates is low, at around 7%. This is typical of the error threshold seen for STT, mAb1 and WFL in the previous chapters. In isolation, the data on their own merely show that, under these flow and buffer conditions, those mAbs in the 0–20% range (Vedolizumab to Guselkumab) are more resilient to stress under extensional flow than those in the +60% range (Alemtuzumab to Foralumab). To compare how the EFD data compare to the Adimab dataset, further statistical analysis is necessary.

6.3.2. Statistical comparison of the EFD data to those from twelve alternative biophysical assays.

Following the generation of the EFD data, the values of mean % protein in pellet for each mAb were sent to Tushar Jain (Adimab) for analysis with the statistical clustering methods discussed in Section 6.1. All of the analyses performed on the original 137 mAbs were repeated using the 33 mAbs selected in Section 6.2. These analyses were performed with and without the flow data. The first such analysis was Spearman's rank clustering of the assays themselves. The hierarchical clustering trees for each set of assays (with and without flow) are shown in Figure 6.8.

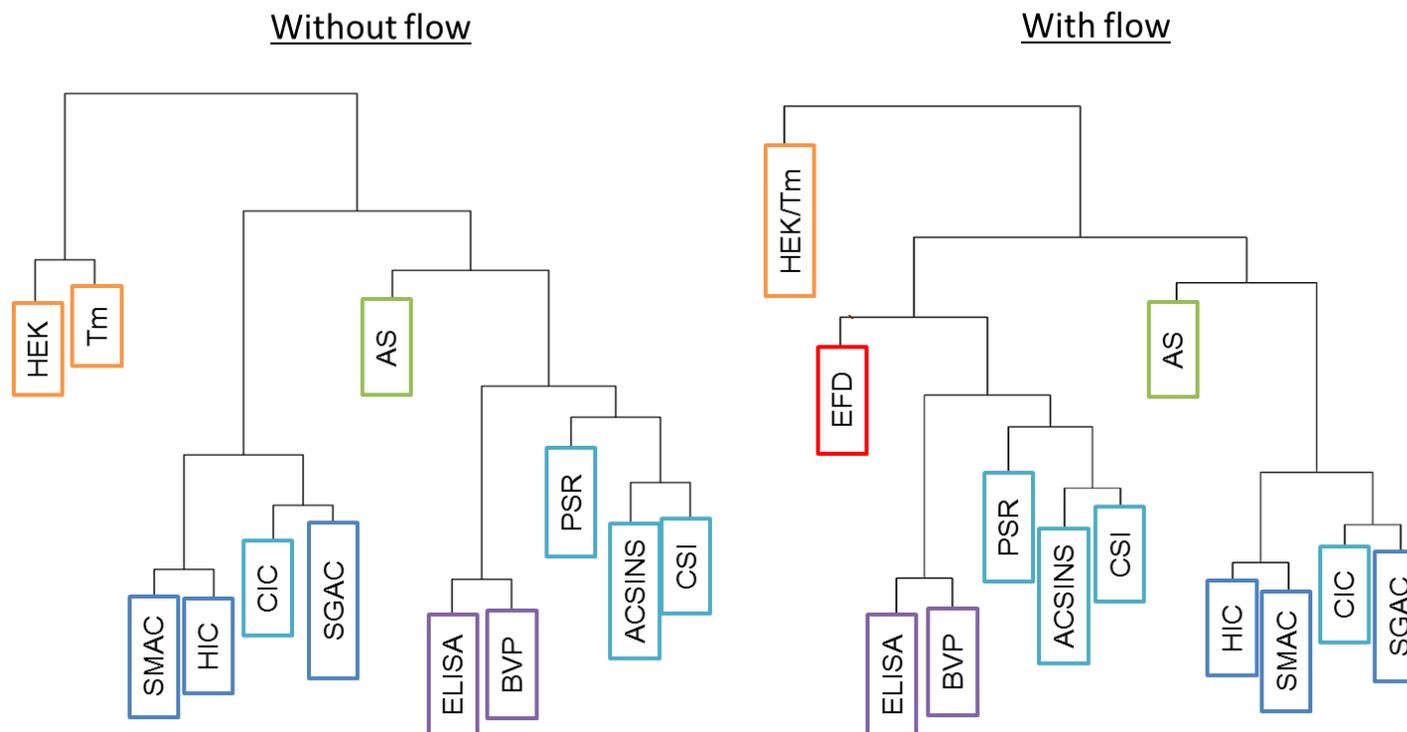


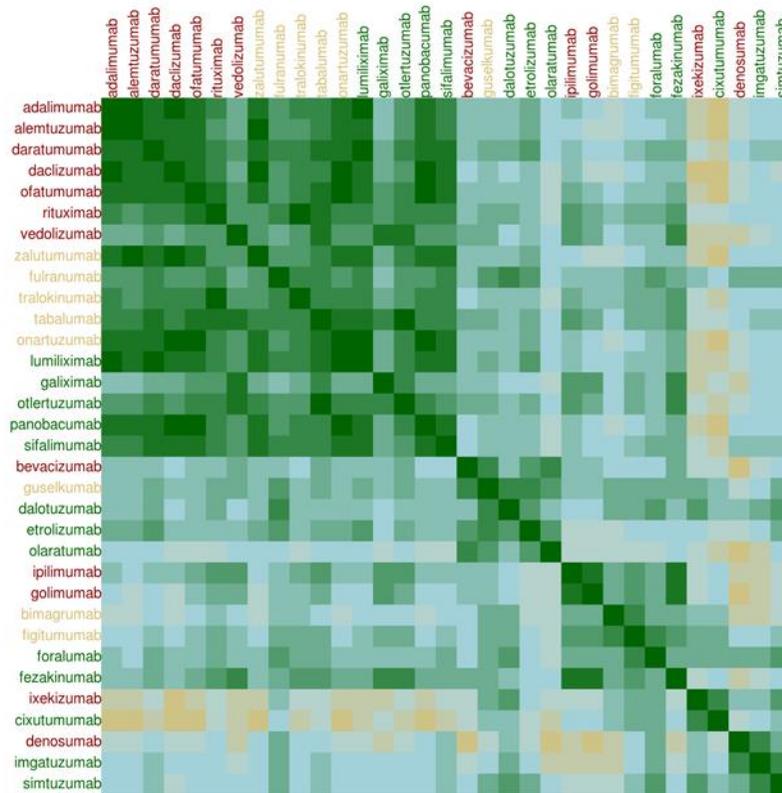
Figure 6.8. Hierarchical clustering ‘Family trees’ of assays used to assay the biophysical properties of the 33 Adimab IgG1s. The tree on the left shows the clustering data for the 33 mAbs selected for this study without the extensional flow data. The tree on the right shows the clustering data for the 33 mAbs when extensional flow is included as an assay (EFD, red box). Analysis performed by

Tushar Jain, Adimab.

The data in Figure 6.8 contain an interesting set of trends. When the data for the 33 mAbs in this study were correlated using Spearman's rank in the absence of flow, the hierarchical clustering deviates from that in Figure 6.1. With the full set of mAbs (all 137), CIC tended to give similar results to PSR, AC-SINS and CSI (light-blue boxes, Figure 6.1). The CIC data for the mAbs selected for this study tended to be more similar to the assays probing self-interaction (dark-blue boxes, left-hand side, Figure 6.8). When extensional flow is included as a '13th' assay, the hierarchical arrangement of the assays changes quite dramatically. The performance by mAbs in the EFD tend to be most related to the assays which probe poly-specificity (purple and light-blue boxes) than the others. The EFD also has its own branch on the tree; it does not pair with another assay directly. This analysis suggests, therefore, that the EFD (and pelleting assay) is a distinct assay to those used by Jain et al. (2017). This means that what the EFD data tell one about the 'developability' of a mAb is distinct to those from ELISAs etc. The implications of this are discussed in Section 6.5.

The hierarchical clustering of the mAbs themselves, from their assay data including and excluding the EFD, was performed as in Figure 6.2. Once again, the HEK titre and T_m data were excluded from the analysis. The two cluster diagrams are shown in Figure 6.9.

Clusters without flow (33 mAbs)



Clusters with flow (33 mAbs)

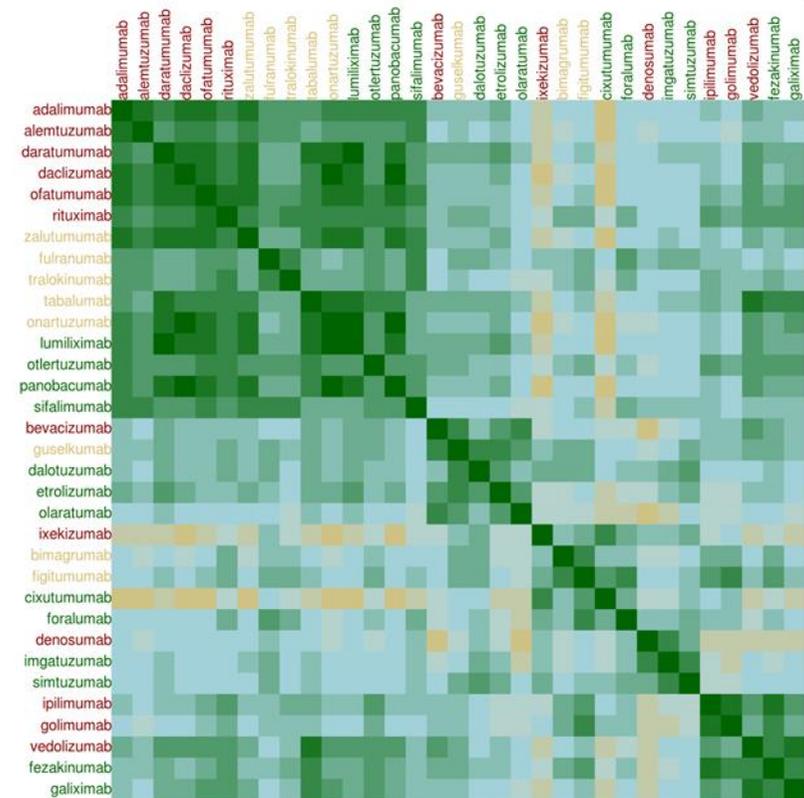


Figure 6.9. Cluster diagrams for the 33 mAbs according to their relatedness in the biophysical assays, both without (left) and with (right) flow data. On the axes, Approved mAbs are labelled in red, Phase-III in yellow and Phase-II in green. Dark green squares in the plot = highly related mAbs, yellow = poor relatedness. Analysis by Tushar Jain, Adimab.

The diagrams in Figure 6.9 show that flow clearly affects how the mAbs are grouped into clusters. The largest group of related mAbs in the top-left corner decreases in size from 17 to 15 when the EFD data are included. The other clusters also appear more distinct (for the 33 mAbs analysed) when the EFD data are included (right-hand plot) than in its absence (left-hand plot). The general trend that more approved mAbs tend to pool in the top left remains, except for when flow is included as an assay, where three approved mAbs group together in the bottom-right cluster. Following this, the relationship between the cluster a mAb finds itself in and the amount of aggregation observed in the EFD assay were correlated with the clusters above. This is shown in Figure 6.10.

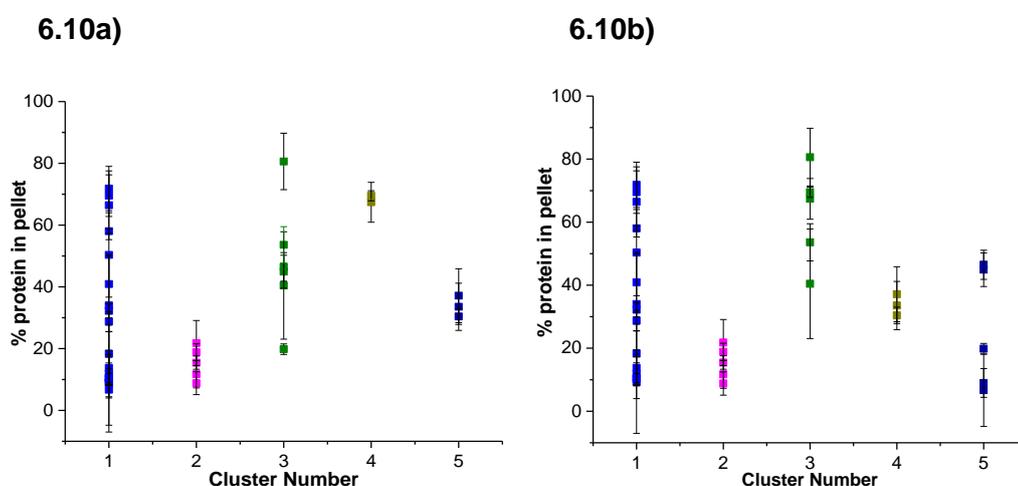


Figure 6.10. Plot of % protein in pellet following stress in the EFD against the cluster number for the Adimab mAbs when a) flow is excluded and b) included in the analysis. Data shown are the same as in Figure 6.5.

The data show that whilst the spread of aggregation behaviour by mAbs in Clusters 1–3 remains steady, both the position and spread of observed aggregation by mAbs in Clusters 4 and 5 changes. For example, the mAbs in Cluster 4 (when flow is not included as an assay in the clustering analysis) in Figure 6.10a tended to all be fairly aggregation-prone. Therefore, when the EFD data are included in the analysis, the overall relatedness to the other mAbs changes, with all three of these mAbs changing cluster. These data show that the EFD affects the grouping of some mAbs more than others. Using the ‘distance from ideal’ red flag analysis, the rank of each mAb including and excluding the EFD data can be calculated to see how the ranks of the mAbs change with the inclusion of this new assay branch. The data are shown in Figure 6.11.

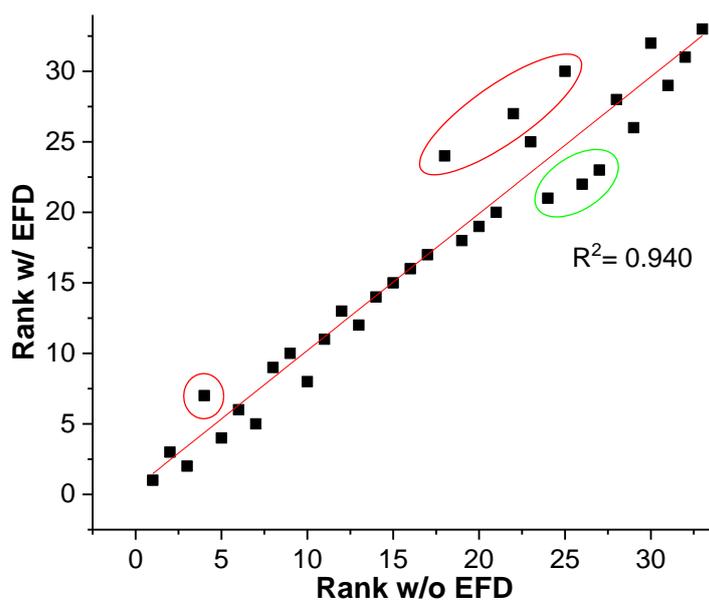


Figure 6.11. Scatter plot of the rank of each mAb including and excluding data from the extensional flow device. A rank of 1 was ascribed to the mAb closest to ideal, whilst the mAb furthest away was ascribed a rank of 33. mAbs circled in red are examples which are ranked far worse with EFD data included. mAbs circled in green are examples which are ranked far better with EFD data excluded. Distance from ideal calculations were performed by Tushar Jain, Adimab LLC.

The data show that some mAbs retain their rank regardless of whether or not flow is included as an assay. Several points have moved position slightly away from the line, indicating that flow has changed their position negligibly, e.g. from 2nd to 3rd. However, some mAbs move several positions when flow is included in the analysis. For example, Alemtuzumab moves from a rank of 4/33 to 7/33 (lower red circle). Some mAbs appear more favourable when the EFD data are included. For example, Denosumab moves from a rank of 26/33 to 22/33. The mAbs circled in red will be discussed further in Sections 6.4 and 6.5. This plot seems to suggest that a single EFD assay cannot replace all the assays used by Jain et al. To confirm this was the case, the mAbs in Figure 6.5 were ranked from best to worst (1–33) and plotted against the distance from ideal rank of the mAbs excluding the EFD from Figure 6.11 (Appendix A6). No correlation was observed, confirming that the EFD augments the array of assays with which one could appraise the ‘developability’ of a mAb. But why do some mAbs behave differently in the EFD to the other assays? Are there any sequence-related features that could help rationalise the data? These questions will be answered in Section 6.4.

6.4. Using bioinformatics to attempt to rationalise the EFD data obtained for the 33 mAbs

6.4.1. Are some families of mAbs more aggregation-prone than others?

So far, it is clear that the 33 mAbs selected from the Adimab dataset cluster together in terms of their relatedness: within an assay, between assays (i.e. amongst one another) and within one assay category (i.e. a branch on the phylogenetic trees in Figure 6.8). With this latter point in mind, it was postulated whether or not a set of mAbs with high sequence homology exhibited similar or diverse levels of aggregation in the EFD. This is particularly pertinent given the data for WFL and STT in Chapters 3–5, where it was demonstrated that despite possessing 99.6% sequence identity, the two mAbs have markedly different aggregation behaviour. The mAb V_H and V_L sequences were input into Clustal Ω (see Section 2.2.17) and a phylogenetic tree constructed. The tree was annotated with the amount of aggregate observed: ● = 0–20%, ■ = 21–60% and ▲ = 61%< protein in pellet after 200 passes. The rank of the protein in the flow assay was also added onto this diagram. Furthermore, to see how the families identified relate to the clusters in Figure 6.9, these too were added to tree. The phylogenetic tree is shown in Figure 6.12.

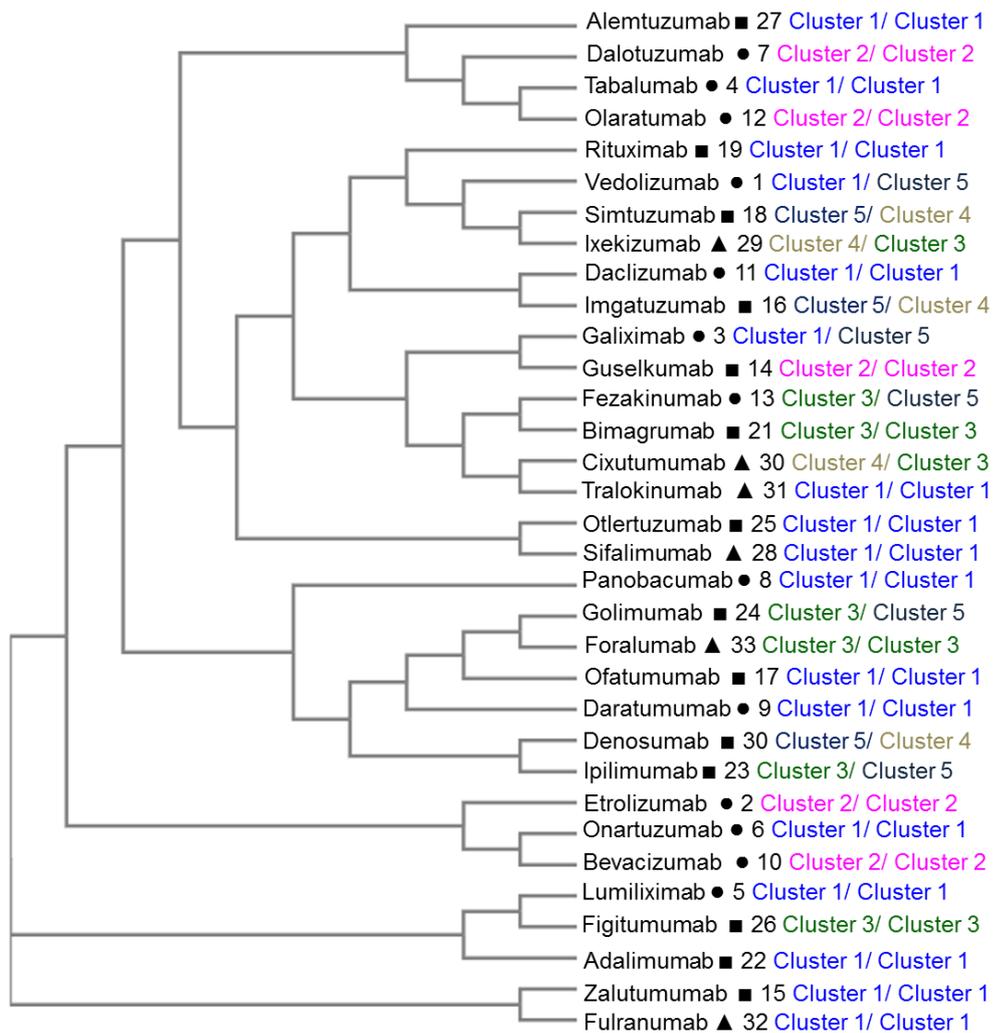


Figure 6.12. Phylogenetic tree for the 33 mAbs. The mAbs at the bottom of the tree are least related to those at the top. ● = 0–20%, ■ = 21–60% and ▲ = 61%< protein in pellet, respectively. Numbers in black denotes the rank in the EFD. Cluster x/ Cluster y = Cluster number without flow/ Cluster number with EFD data, from Figure 6.9. The length of the branch is proportional to the number of mutations away from the ancestral sequence. The tree was constructed using Clustal Ω according to Section 2.2.17, then the tree annotated in Microsoft Powerpoint.

The data in Figure 6.12, whilst complex, show that the majority of the families of mAbs identified possess mAbs which aggregated to different extents in the EFD. Exceptions include Cixutumumab and Tralokinumab (30/33 and 31/33 respectively) and Etrolizumab and Onartuzumab (2/33 and 6/33 respectively). mAbs from the same cluster which are related in sequence, e.g. Zalutumumab and Fularnumab (bottom sequences) can have differing behaviour. This is in

direct agreement with the data obtained previously for WFL and STT. The data show, therefore, that aggregation-propensity in the EFD is not a ‘family trait’.

6.4.2 Do the simple, intrinsic properties of the mAb sequences correlate with the amount of aggregate formed by the EFD?

A significant advantage of the Jain et al. (2017) paper to similar studies is the availability of the V_H and V_L sequences as a dataset for all of the mAbs used in the study. IgG1s have incredibly similar structures, with variability typically limited to the CDRs in the variable domains, as opposed to the framework regions. With this in mind, an array of bioinformatics tools were employed to try and correlate the intrinsic biophysical characteristic of the 33 mAbs with their aggregation behaviour under extensional and shear flow. As stated in Section 6.2, the pI and intrinsic CamSol scores for the V_H and V_L sequences were computed for each mAb (see Section 2.2.17), as these solubility-based calculations may explain why some mAbs are more susceptible to hydrodynamic forces/more aggregation-prone than others (Sormanni et al., 2017). These were plotted against the % protein in pellet for each mAb. The data are shown in Figures 6.13a) and b) respectively.

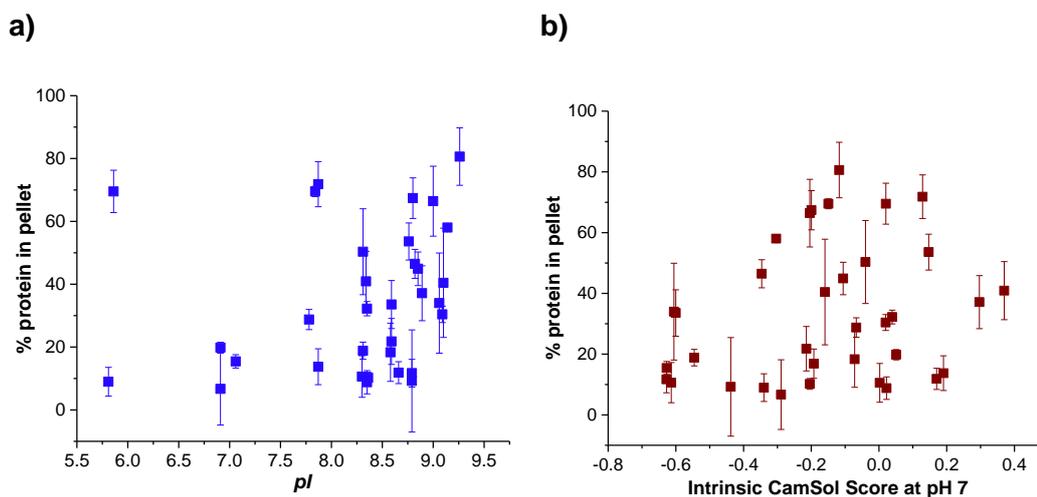


Figure 6.13. Plot of the % protein in pellet (from Figure 6.5) against a) pI and b) Intrinsic CamSol score at pH 7 for the scFv of the 33 mAbs stressed in the study. V_H and V_L sequences were input into ProtParam and CamSol according to Section 2.2.17.

The data in Figures 6.13 show that the trend between the pI (a) or intrinsic CamSol score (b) and the extent of aggregation observed in the EFD for the Adimab proteins, is not simple. The mAbs stressed in the device near their pI (~ pH 7) had low levels of aggregation in the EFD (0–20%) (Figure 6.13a). The most basic mAbs (pI > 8.5) appear to have generally worse aggregation behaviour in the EFD. More sophisticated statistical tests would be needed to interrogate the significance of this observation. A CamSol score below -1 is indicative of a poorly-soluble sequence (Sormanni et al., 2015). None of the V_H+V_L sequence combinations for each mAb exceed this threshold. Despite this, even the sequences predicted to be the most insoluble behaved favourably in the EFD. These data suggest that attractive interactions between native state proteins, coupled with the intrinsic solubility of the binding domains, does not lead to increased levels of aggregation in the EFD. It should be noted that the CamSol score is averaged across the whole sequence, thus localised regions with poor solubility can be masked with such an approach.

The amino acids found in the CDRs are crucial to both antigen binding and resistance to aggregation (Dudgeon et al., 2012; Peng et al., 2014). A recent paper analysed the correlation of the AC-SINS assay with the net charge of the CDRs for the Adimab dataset. The authors found that the more positive the net charge at pH 7, the worse the mAb behaved in the AC-SINS assay (Alam et al., 2018). Additionally, as CDRs drive the affinity of a mAb to its target, it was posited that affinity maturation may render mAbs more aggregation-prone. To examine whether the physico-chemical properties of the CDRs correlate with EFD rank, the net CDRs of the 33 mAbs were calculated in ProtParam, according to the strategy of Alam et al. A literature search for the binding affinities of the 33 mAbs was also performed, with 26 identified. These values were plotted against the EFD aggregation data from Figure 6.5. The data are shown in Figures 6.14a and b) respectively.

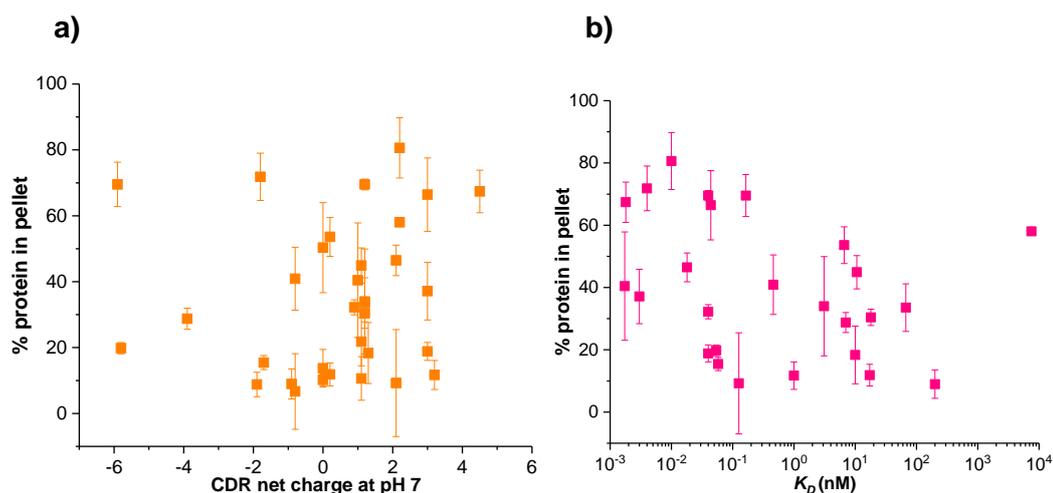


Figure 6.14. Plot of the % protein in pellet (from Figure 6.5) against a) Net CDR net charge at pH 7 for all 33 mAbs and b) K_D for 26/33 mAbs stressed in the study against their targets. CDR Net charges were calculated according to the method of Alam et al., 2018 (Section 2.2.17). K_D values for the 26 mAbs shown above were obtained from the literature, with references give in Appendix Table A4.

These data show, once again, that the correlation between the biophysical characteristics possessed by the mAbs and their apparently complex response to hydrodynamic forces, is not simple. The CDR analysis (Figure 6.14a) shows that even when the CDR charge is positive, which is thought to lead to high degrees of self-association, the flow-induced aggregation data cover the full dynamic range seen in Figure 6.5 (~10%–80%). The same trend is seen when the binding affinity for 26 of the mAbs is plotted against their observed aggregation (Figure 6.14b). The most potent mAbs have pM binding affinities, with their observed EFD aggregation levels covering a large dynamic range. This is similar to WFL and STT in Chapters 3–5, as both mAbs have similar binding affinities (Dobson et al., 2016), yet markedly different aggregation behaviour in the EFD. Once again, further statistical analysis would be necessary to aid interpretation of the data.

6.4.3 Using Solubis to identify APRs in the variable domains of the Adimab proteins

Where are the APRs distributed in the 33 mAbs?

Rather than the *global*, intrinsic physico-chemical properties of the 33 mAb sequences, *local* regions could be driving aggregation in the EFD, as seen for WFL in the previous chapters. As discussed in Section 1.3.1, some APR predictors have recently emerged which allow structural information to be incorporated into the analysis, rather than just sequence. The Solubis algorithm was developed to combine the TANGO APR algorithm with the FoldX force-field to aid the design of thermodynamically stable mAbs (Van Durme et al., 2016; van der Kant et al., 2017). The use of this webserver relies on having structural information for the protein of interest. Clearly, solving the crystal structures for all 33 mAbs used was not practical.

Homology models of mAbs can, however, readily be constructed with reasonable precision using ABodyBuilder (Leem et al., 2016). The 33 mAbs' V_H and V_L sequences were input into the ABodyBuilder webserver to generate homology models for all of the proteins (see Section 2.2.17). These PDB files were then analysed with the Solubis webserver to identify APRs within the proteins. This yields both the TANGO score of the APR (i.e. its stickiness) and its contribution to the thermodynamic stability of the mAb (via the $\Delta G_{contrib}$ term). 175 APRs were identified for the 33 mAbs. The homology models were then annotated with the APRs (identified above) in PyMol. For brevity, the structures of only twelve mAbs of interest are shown. The four mAbs circled in red from Figure 6.11, which are ranked quite differently when flow is included as an assay, are presented in Figure 6.15. The four mAbs which move by two or more clusters when flow is included as an assay are in Figure 6.16. Finally, the four mAbs which did best and worst in the EFD, and best and worst overall (closest and furthest to ideal, respectively) are shown in Figure 6.17.

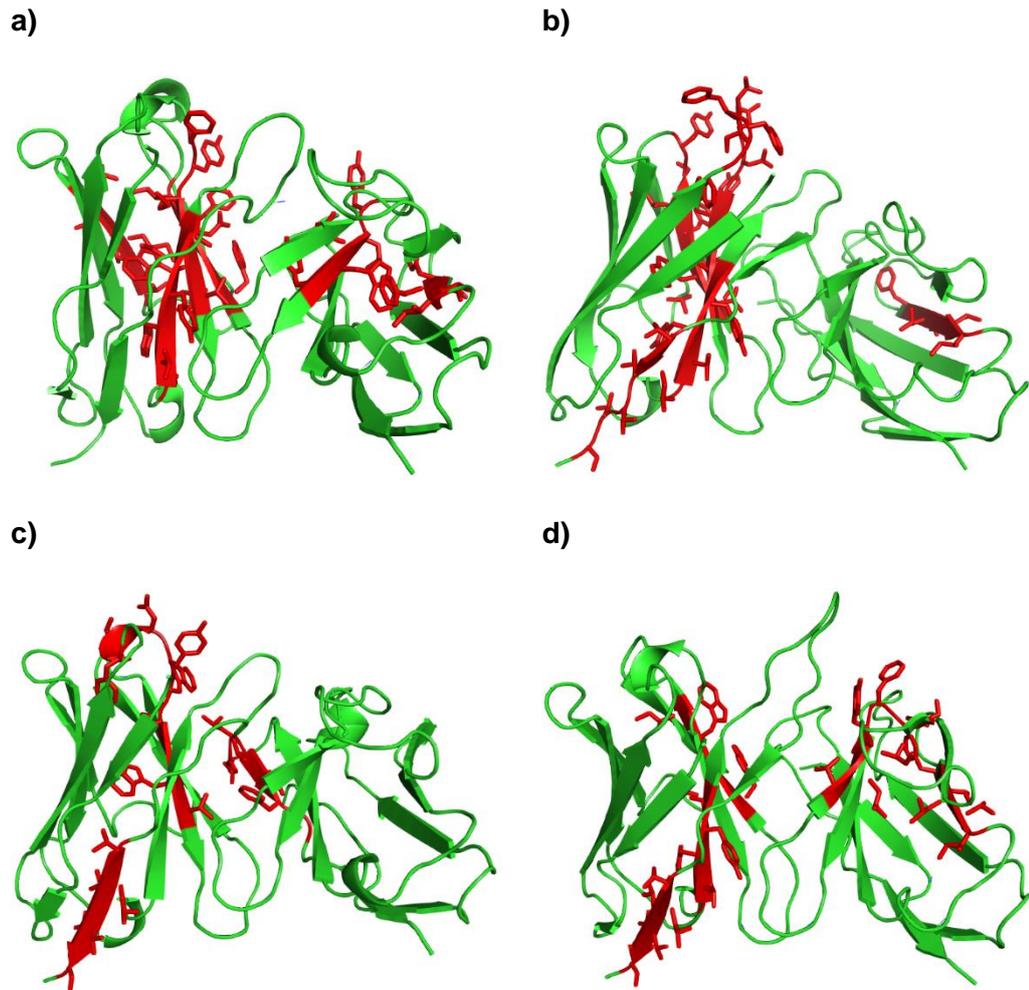


Figure 6.15. scFv homology models (V_H and V_L , left and right respectively) for a) Alemtuzumab, b) Etrolizumab, c) Guselkumab and d) Bevakizumab. APRs identified in Solubis are highlighted in red, with the rest of the structure shown in green.

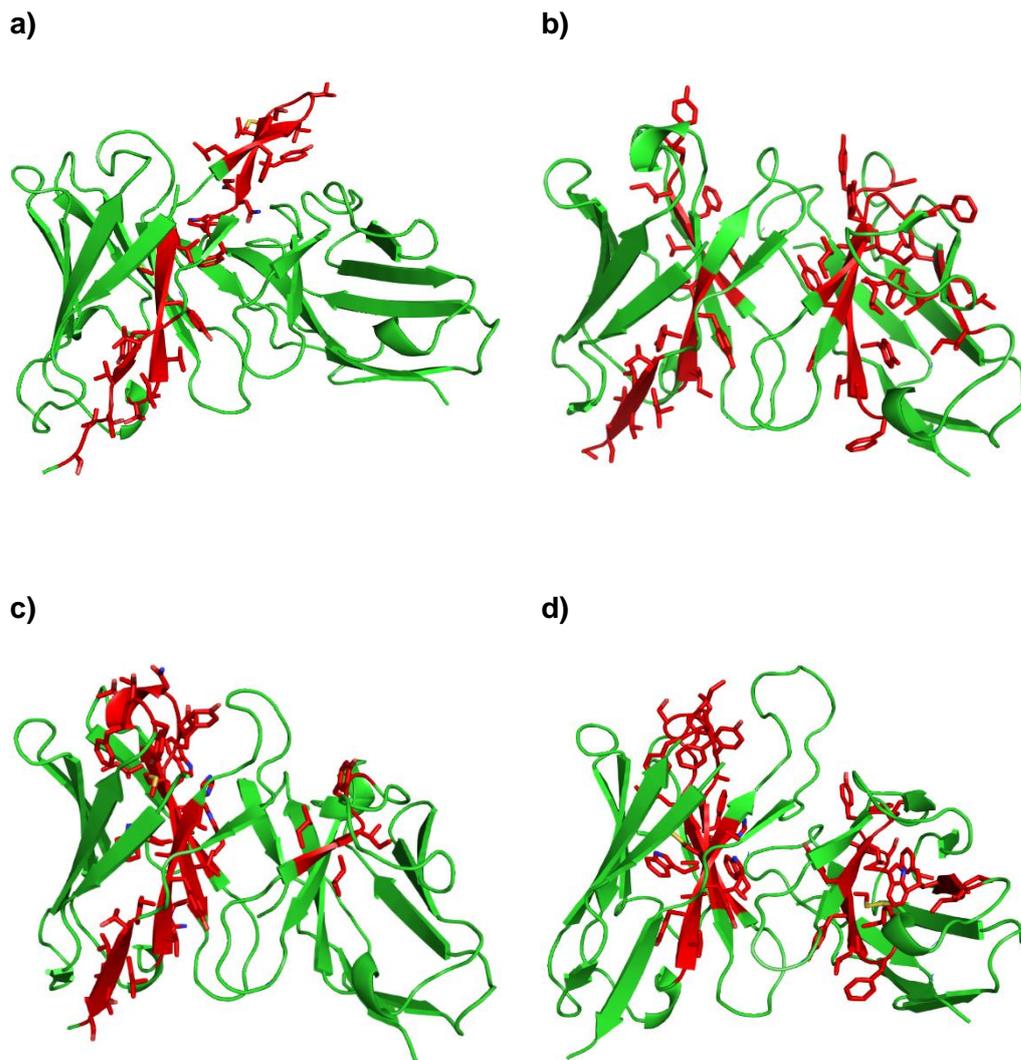


Figure 6.16. *scFv homology models (V_H and V_L , left and right respectively) for a) Galiximab, b) Ipilimumab, c) Fezakinumab and d) Golimumab. APRs identified in Solubis are highlighted in red, with the rest of the structure shown in green.*

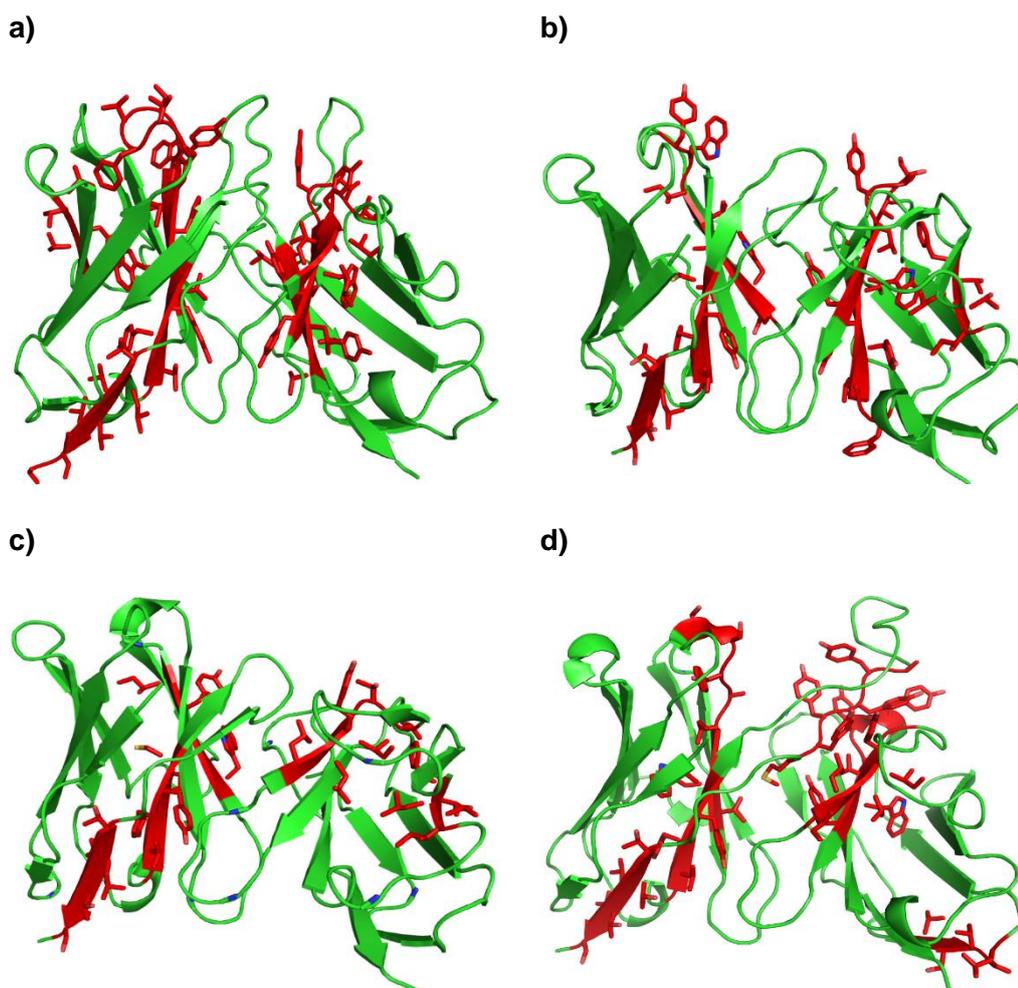


Figure 6.17. scFv homology models (V_H and V_L , left and right respectively) for a) Vedolizumab, b) Foralumab, c) Daclizumab and d) Cixutumumab. APRs identified in Solubis are highlighted in red, with the rest of the structure shown in green.

The models shown throughout Figures 6.15–6.17 present some interesting features. Generally, the APRs in the scFvs tend to be localised along the β -sheets comprising the V_H - V_L interface (discussed later). Some mAbs, e.g. Etolizumab (6.15b), Galiximab (6.16a) and Cixutumumab (6.17d) have particularly long CDRH3 loops, which feature/ are adjacent to APRs. Interestingly, Galiximab (6.16a) does not possess any APRs in its V_L domains, the only protein of the 33 to have this feature.

How aggregation-prone are the APRs identified in the 33 mAbs?

To further understand the aggregation propensity of these regions, the 175 APRs from all 33 mAbs were graphed as a 'stretch plot' (Section 2.2.17), along with the extent of aggregation of the mAbs under flow, using the three levels described previously (Section 6.3.1). This plot is shown in Figure 6.18.

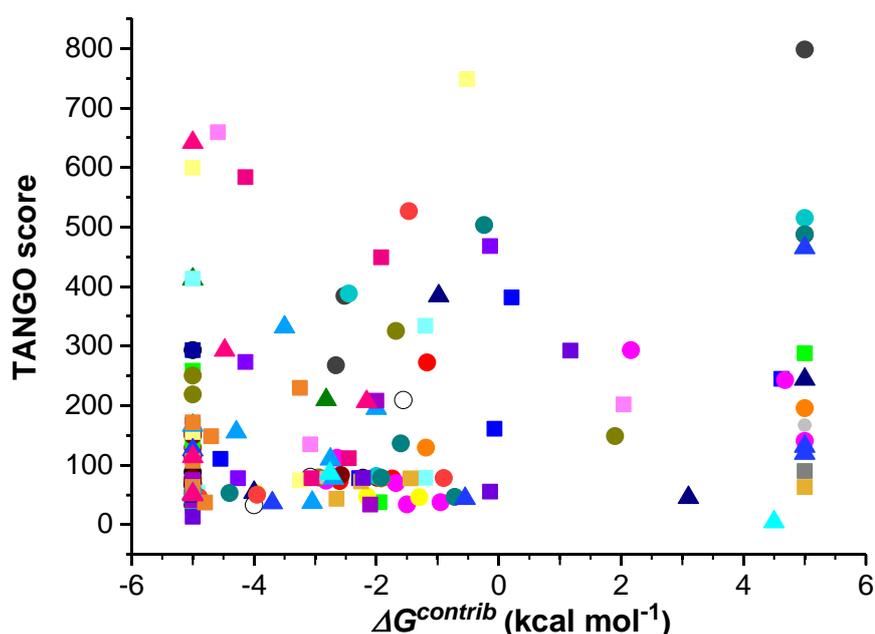


Figure 6.18. 'Stretch plot' of the APRs identified in the 33 mAbs using **Solubis**. The higher the TANGO score, the higher the aggregation-propensity of a given APR. The more negative the $\Delta G^{\text{contrib}}$, the more stabilising the APR is to a mAb (i.e. buried not surface-exposed). Symbols represent the coarse aggregation propensity of the parent mAb that contains that APR (● = 0–20%, ■ = 21–60% and ▲ = 61%+ aggregation under flow respectively). The APRs are coloured according to their parent mAbs from Figure 6.5.

These data show that most of the APRs in the mAbs appear to be of reasonably low aggregation propensity and buried in the hydrophobic core of the V_H and V_L domains (i.e. bottom-left quadrant). If the levels of aggregation observed in the EFD were merely due to surface-exposed, highly sticky APRs, then the top-right

quadrant would contain many triangles. The stretch plot suggests that this is not the case. To validate this, the Solubis score was for each APR within each mAb was summed and plotted against the amount of insoluble protein quantified after the mAb was stressed in the EFD. The data are shown in Figure 6.19.

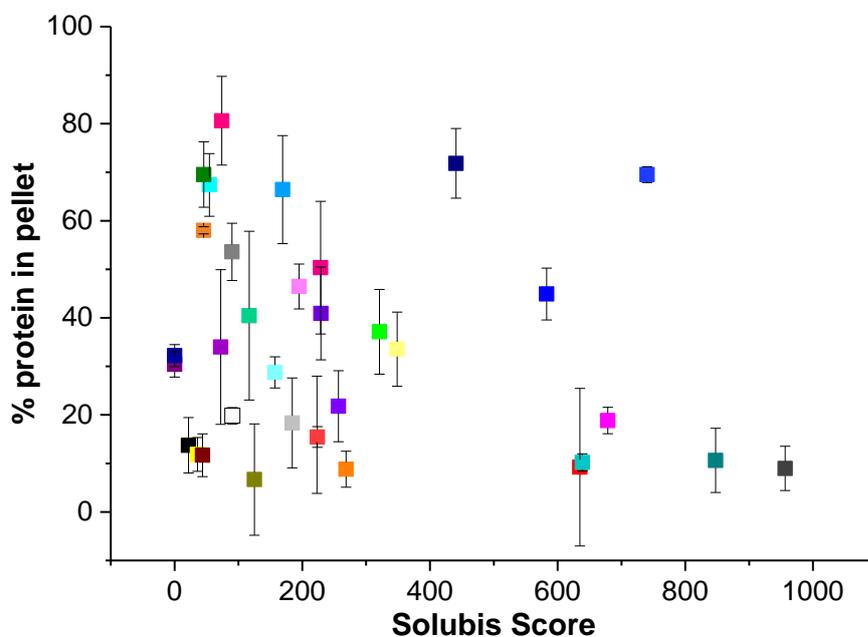


Figure 6.19. Plot of the Solubis score ($\Delta G^{contrib} \times TANGO$ score for the total number of APRs) vs the % protein in pellet formed in the pelleting assay for each mAb after stress. All data are taken from and colour-coded according to Figure 6.5.

The data show that all of the mAbs possess at least one *partially* buried APR, as the total Solubis score for all but two mAbs is greater than zero. Higher Solubis scores correlate with exposed, sticky APRs in a mAb (van der Kant et al., 2017). When this structural correction of the TANGO score takes place, it is even clearer that the APRs which may form the aggregation interfaces in the 33 Adimab proteins stressed in this study are *not* exposed in the native state. If this was the case then Onartuzumab and Galiximab (the teal and dark grey points with Solubis scores ~850 and 960 respectively) would aggregate to the greatest extent in the EFD. In fact, the converse was observed to be true. These data add further evidence to the hypothesis proposed in Chapter 5 that mAbs aggregate through a *partially* unfolded N^* state, triggered by extensional flow.

Does the V_H - V_L interface become exposed under flow, driving mAb aggregation?

Upon closer analysis of the APR distribution plots, Vedolizumab (6.17a) and Foralumab (6.17b) occupy the poles of the extent of aggregation measured in the EFD (best and worst respectively), yet appear to have a similar distribution of APRs in their variable domains, as well as low Solubis scores (125 and 74 respectively). Daclizumab (6.17c) possesses a fairly benign set of APRs (Solubis score = 184) and is the best ranked mAb of the 33, whilst Cixutumumab (6.17d) has a particularly exposed YYYYYM motif, probably accounting for its high Solubis score (740) and worst ranking overall out of 33 proteins. Coincidentally, in the original ranking analysis in Figure 6.3, Cixutumumab is ranked 137/137.

At face value, it appears that Solubis cannot report whether or not a particular protein will be sensitive to the effects of extensional flow. It can however identify potential APR binding interfaces; as stated below Figure 6.17, many of the mAb scFvs have APRs distributed along their V_H - V_L interface. It was thus posited that if this interface was the crucial region where mAbs partially unfold under flow to form the N^* state, then the unfolding of one domain would increase the stickiness of that region by exposing more of this interface to the solvent. The largest difference in solvent accessibility when either V_H or V_L unfolds, leading to the exposure of the 'stickiest' set of APRs as measured by TANGO, may then correlate with the greatest amount of aggregate formed in the EFD. To test this hypothesis, the estimated solvent-accessible surface area (SASA) of each APR in all 33 mAbs was calculated using PyMol. The V_H and V_L domains were deleted in turn, with the SASA re-calculated after the deletion in each case (see Section 2.2.17). The Δ SASA, summed TANGO score of the whole scFv and the amount of aggregate formed after stress in the EFD are presented in Figure 6.20.

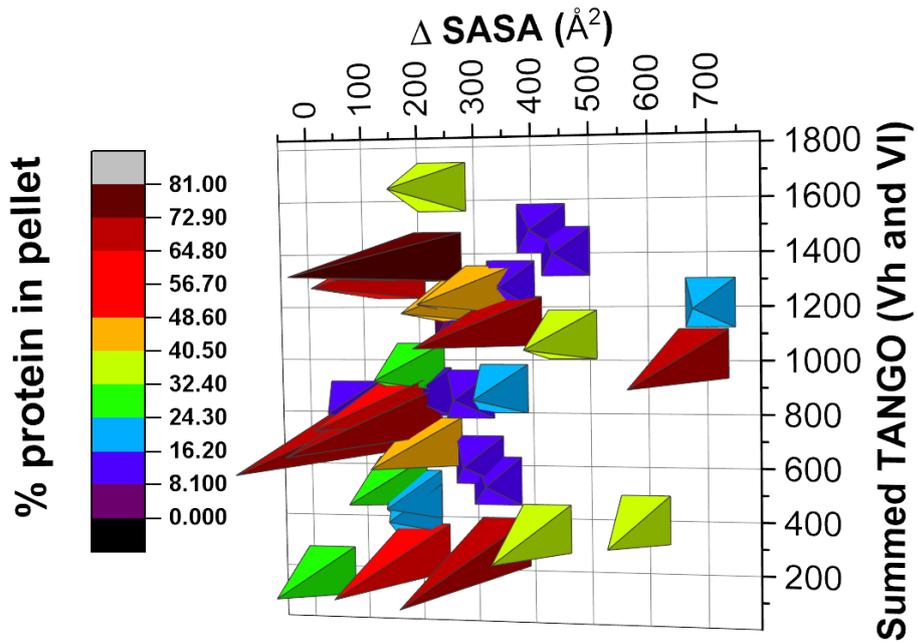


Figure 6.20. Plot of the summed TANGO score against the change in interfacial APR solvent-accessible surface area for the 33 Adimab scFvs. The amount of aggregate formed under flow for each mAb (from Figure 6.5) is mapped onto the plot, with the colour and length of each peak corresponding to the amount of aggregate quantified in the pellet.

The data in Figure 6.20 show that this hypothesis was incorrect. If exposure of this interface under flow led to a greater amount of aggregate being observed after stress in the flow device, then the large red peaks in Figure 6.20 would all be in the top-right corner of the plot. Only Cixutumumab occupies this quadrant, which may once again explain its behaviour in the EFD and the other assays. In conclusion, whilst Solubis can help generate hypotheses as to *why* a particular mAb may performed as it did in the EFD or another assay, it cannot rationalise the data and provide a clear answer for this. Possible reasons for this will be included in Section 6.5.

6.5 Discussion

As introduced in Sections 1.2 and 1.3, in addition to Section 6.1, various assays and methods are employed by the biopharmaceutical industry to determine the ‘developability’ of a mAb. No one assay provides all of the information necessary to determine whether or not a molecule will pose a ‘developability’ issue or not. Many high-throughput techniques are employed by industrial laboratories to screen both large numbers of conditions and molecules for colloidal and conformational stability (Bhirde et al., 2018; Capelle et al., 2007; Chaudhuri et al., 2014; Goldberg et al., 2011; Zhao et al., 2010). Forced-degradation methods are employed to assess the sensitivity of mAbs to different environmental stresses (Luo et al., 2011; Nowak et al., 2017). The assay chosen, the physicochemical environment and the molecule itself will all impact the result of such developability assays- which ultimately dictate whether or not a drug candidate falls out of the development pipeline (Yang et al., 2013).

The generation and analysis of large datasets is gaining attention in the biopharmaceutical industry and beyond, under the umbrella of ‘Big Data’ (Costa, 2014; Szlezák et al., 2014). This is particularly true in today’s era of Quality by Design (Section 1.2.4). Just as the analysis of a large set of molecular data led to Lipinski’s Rules of five (Lipinski et al., 1997), the work of Jain et al. set out with a similar aim for monoclonal antibody therapeutics (Jain et al., 2017). One issue limiting the maximal implementation of Big Data in the mAb development field is the availability of intellectual-property free molecules and their associated biophysical/ colloidal and conformational stability data (Szlezák et al., 2014). The NISTmab, a reference molecule developed specifically for the validation of biopharmaceutical assays (Saro et al., 2015; Schiel et al., 2015), has been employed by many groups to assess different analytical techniques (Hsieh and Wyatt, 2017; Karageorgos et al., 2017). The obvious limitation here is that of dataset size when correlating the aggregation of NISTmAb with one or more attributes from an unrelated mAb (or panel of mAbs).

The Jain et al. (2017) paper made a large step towards alleviating this problem by: (i) using clinically relevant mAbs of which the sequence of the variable domains are known; (ii) using established assays from the literature; (iii) clearly presenting the analytical framework used to cut through the dataset and (most importantly), (iv) making these datasets publicly available. Prior to this chapter,

the only mAbs subjected to stress in the EFD were WFL, mAb1 and STT. These molecules allowed for the extensive exploration of parameter space in the EFD, in addition to the appraisal of the different analytical methods available to this author to characterise aggregation. The translation of this research to industry would be limited with a dataset of just $N = 3$ antibodies. Through systematically choosing a subset of $\sim 1/4$ of the Jain mAbs, it is clear that the EFD can distinguish between robust, intermediate and sensitive mAbs under the conditions selected for this study.

It is highly likely, given the data in Chapters 4 and 5 (for WFL, mAb1 and STT), that changing the buffer, protein concentration and flow conditions would change the levels of aggregation observed for these proteins following extensional flow. There are additional caveats with the molecules themselves, as acknowledged by Jain et al. (2017): some of the IgGs are originally IgG4's, not IgG1s; the mAbs are not in their designed formulation buffers in any of the assays used by Jain or in the EFD (except when defining the cut-offs in the Jain et al. (2017) paper; these molecules are based on clinically relevant mAbs- they must have been 'developable' in the eyes of both the proprietor and the regulator to make it into the clinic.

Despite this, the work presented in this Chapter shows that the EFD appears to be statistically distinct to other orthologous assays used in the field. Through the addition of more assays to the analytical repertoire, it is clear that for certain mAbs, e.g. Alemtuzumab in Figure 6.11, that the incorporation of the EFD would change the ranking of a mAb in the eyes of a developer. Further work could then be performed on the mAb, e.g. formulation screens and landscape generation of the type seen in Chapter 4, to stabilise such mAbs against hydrodynamic stress. Outright 'bad' mAbs such as Cixutumumab may be deemed impossible to develop or manufacture. It would be interesting to see how mAbs lower down the development pipeline fare in both the EFD and the assays employed by Jain et al. (2017). Finally, it is evident that the bioinformatics tools currently available, such as CamSol and Solubis, cannot predict or explain why a particular mAb will aggregate under extensional flow.

These methods do allow for the prediction of APR interfaces which may form under flow, which open up many questions to be addressed in the future. Such questions, in addition to a critical appraisal of this thesis form the subject of Chapter 7.

7. Concluding remarks and future directions

The aim of the project was to understand whether (and if so, how) extensional and shear flows affect proteins to induce aggregation. Specifically, the effects of flow on therapeutic proteins was the main focus of this thesis. Initially, a model protein was sought from the literature where evidence for conformational change and/or aggregation existed. BSA, whilst not a therapeutic protein, was a suitable candidate for this role. Its ready availability and low cost (10g, £50) meant that a wide array of parameter space in the EFD, in addition to a range of analytical methods, could be explored. This allowed for understanding to be gained about how the number of passes in the EFD related to both the amount and size of BSA aggregates formed due to flow. Crucially, the free Cys34 in BSA could be exploited in the IAEDANS assay to show that extensional flow causes conformational remodelling and subsequent aggregation of *globular* proteins, the first demonstration of this event to this author's knowledge.

The flow fields present in the EFD are well-defined and characterised by CFD. By coupling a detailed knowledge of the proteins used in the study to an understanding of the flow environment in the EFD, firmer conclusions could be drawn as to *why* a given protein aggregates following stress under extensional and shear flow. With therapeutic proteins in mind, WFL and STT represent the 'Yin and Yang' of antibodies. Despite differing by only six residues, the two mAbs have been shown to display contrasting aggregation behaviour under quiescent conditions (Devine, 2016; Dobson et al., 2016). The device was demonstrated to be able to differentiate between these model mAbs and the generic mAb1 under a limited set of flow conditions. This initial work (presented in Chapter 3) also showed the heightened sensitivity of the three mAbs to hydrodynamic forces compared to BSA, as they aggregated to a greater extent than BSA after just 100 passes at a concentration of 0.5 mg mL^{-1} , ten-fold lower than BSA.

In Chapter 3, it was observed that as the plunger velocity increases in the EFD, the energy imparted into the fluid also increases. For BSA, it was observed that above an energy threshold, aggregation becomes more pronounced. In Chapter 4, the parameter space in the EFD was explored for BSA, WFL, mAb1 and STT. The insoluble protein pelleting assay, though crude, allowed for quantitation of aggregation as a result of extensional and shear flow. The aggregation landscapes generated for the model proteins revealed complex relationships

between the number of passes in the EFD (the total time exposed to hydrodynamic forces) and the magnitude of the strain and shear rates present, with the levels of aggregation observed. This study provided more evidence for the striking disparity between WFL and STT in particular, with WFL's sensitivity to extensional flow highlighted. Under certain conditions, these proteins were also shown to be sensitive to shear flow, despite numerous studies questioning the role such flow fields have in triggering protein aggregation (Bee et al., 2009; Brückl et al., 2016; Jaspe and Hagen, 2006; Thomas and Dunnill, 1979).

The role the buffer plays in flow-induced aggregation was also investigated for the first time. This study highlighted the power of arginine+succinate buffer to suppress protein aggregation under flow. This screen was performed using buffers for which WFL and STT's aggregation behaviour, under quiescent conditions at higher protein concentrations, was known (Dobson et al., 2016). However, these studies are not perfect- the buffers used were all of low molarity (10 mM) and differing ionic strength, factors which are known to affect the colloidal stability of proteins (Hofmann and Gieseler, 2018). Furthermore, the formulated concentration of therapeutic antibodies are far higher than the 0.5 mg mL⁻¹ investigated in this thesis. Whilst the conditions in Chapter 4 were advantageous for this study, they do not necessarily reflect a 'real' mAb formulation. At higher concentrations, the viscosity of the solution may change, affecting the fundamental fluid mechanics present in the EFD. A more thorough investigation of formulations and how they affect flow-induced mAb aggregation/ physicochemical properties would be of great interest in the future.

Much of the work in Chapter 4 looked at the amount of aggregate formed at the end of the pathway from monomeric protein to insoluble aggregate. In Chapter 3, this pathway could be described for BSA. This was perhaps facilitated by the protein: being well-studied historically; possessing a naturally occurring free Cys residue and being less topologically complex than mAbs. In Chapter 5, such a pathway was sought for the three mAbs. The activated N* state of these proteins was difficult to capture. Extrinsic fluorescent dyes were found to be ineffective in detecting aggregation *in situ*. This was perhaps due to differences between the proteins, flow fields and buffer conditions in this thesis and the work of Wolfrum et al. (Wolfrum et al., 2017).

Chemical cross-linking experiments were also unsuccessful in capturing the dimer aggregation interface of WFL. This is not a trivial exercise- capturing such

interactions between 150 kDa proteins in a complex flow environment makes downstream analysis difficult. Further optimisation of the cross-linking conditions, in addition to implementation of a more sensitive mass spectrometer now available to this author, would make this experiment more feasible. Together, the chance of detecting low-abundance inter-molecular cross-links may increase. It would be highly desirable to have a flow device available where optics etc. could be in place to probe the flow-induced formation of the N* state directly in the device. At present, the metal collar of the Hamilton syringes make such a design impossible. The design and testing of microfluidic extensional flow devices is under current investigation within this author's wider research team (Panagi, 2018).

To attempt to gain an idea of the underlying kinetics of monomer loss caused by stress in the EFD, HPLC assays were incorporated into the workflow. This assay highlighted why WFL was rejected as a mAb candidate by MedImmune; its aberrant interaction with the chromatography matrix (despite arginine being present) rendered kinetic analysis unfeasible. For STT and mAb1, this assay did work in principle, showing a loss of monomer as a function of time (pass number). However, the quality of the data was compromised by the age of the column and the (later shown to be unnecessary) quench step. It would be desirable to perform these experiments again with: a new column, more time points and, more concentrations. This would allow better data to be obtained to which kinetic models could be fitted.

The onset of aggregation after just ten passes could be detected using SD-SDS-PAGE. A detailed understanding of the migration behaviour of mAbs on such gels (such as higher apparent molecular weights of monomers, dimers etc.) would be interesting to investigate in the future. The data from these gels, coupled to those obtained from light scattering, allowed the formation and growth of mAb aggregates as a function of pass number to be monitored for WFL, mAb1 and STT. In future, by obtaining more data at different concentrations and pass numbers, in addition to performing the HPLC experiments described above, a comprehensive mechanistic model describing flow-induced mAb aggregation could be constructed for the first time.

The seeding and cross-seeding experiments presented in this thesis showed that flow-induced mAb aggregation can be nucleated. These experiments highlighted the affinity WFL has for its own aggregates over those of STT. Performing these

experiments at different seed concentrations (e.g. 5% v/v) and measuring the effect this has on the *rate* of aggregation, as opposed to just the end-point, would be an interesting future study. Overall, obtaining the mechanism of flow-induced aggregation would be highly desirable, as the resulting kinetic model then becomes predictive. This would maximise applicability in the biopharmaceutical industry if the aggregation mechanism of a mAb observed in the EFD is the same as that seen in an industrial setting.

Regardless of the mechanism, the ability of any assay or device to distinguish between robust and aggregation-prone proteins is highly pertinent to the development of biopharmaceuticals (Chapter 1). If current predictive assays used by the biopharmaceutical industry were perfect, then mAb aggregation would not be perceived as a problem and much of the work in this thesis would be redundant. Clearly, this is not the case, thus new tools are needed to tackle the aggregation problem. In Chapter 3–5, the EFD was able to differentiate between three model mAbs. How would it fare with a wider panel of mAbs?

The 2017 paper by Jain et al. generated a large reference dataset of 137 clinically relevant mAbs and their biophysical behaviour in twelve different assays. By choosing a subset of 33 mAbs, then subjecting them to stress in the EFD and pelleting the samples, it was shown that some of the mAbs are recalcitrant to the effects of flow (like STT), some are reasonably aggregation-prone (like mAb1) and some are very sensitive (like WFL). When the data generated by the EFD were compared to the data from Jain et al. (2017), it was shown that the EFD is a distinct biophysical tool to those currently used in the biopharmaceutical industry. This means one gains unique, additional information about the biophysical behaviour of a mAb. Bioinformatics analysis of the mAbs' variable sequences and model structures highlighted possible aggregation-prone regions which may become exposed due to extensional flow. However, none of the methods selected provide an 'easy answer' as to *why* a mAb should necessarily aggregate when subjected to stress in the EFD. The 'WFL' motif in MEDI1912 (WFL) ((Dobson et al., 2016)) and the YYYYYM loop in Cixutumumab are the exceptions to this rule.

As the biopharmaceutical industry moves away from the traditional IgG mAb towards more exotic, 'unnatural' modalities (Section 1.2), the aggregation-propensity and manufacturability of such molecules may be a problem in the future, as the 'academic territory' surrounding this is fairly uncharted. In addition, personalised on-demand biologic medicines are coming closer to a reality (Boles

et al., 2017). How hydrodynamic forces impact such molecules at this scale is not known; perhaps the scale of the EFD would be the perfect tool to predict this? For better or worse, hydrodynamic forces affect biomolecules both *in vivo* and *in vitro*. The work presented in this thesis opens many possible avenues of further study. It is the hope of this author that the work will facilitate the manufacture of promising therapeutic proteins into the medicines of the future.

To conclude, the work contained within this thesis shows that:

- Hydrodynamic forces have the ability to induce protein aggregation. The extent of aggregation observed is mainly dependent on the protein in question, the magnitude of the hydrodynamic forces present in the EFD, the amount of times a protein experiences these flow events and the buffer composition.
- Aggregation-prone proteins, or proteins put into highly aggregation-prone states (such as STT at 16 mm s^{-1}), are sensitive to shear flow.
- Flow-induced aggregation likely proceeds through a pathway involving 'activated' N* species, oligomers and soluble aggregates, culminating in the formation of insoluble protein.
- The ability of different bioinformatics and biophysical methods to predict, detect and quantify flow-induced protein aggregation varies depending on the method and the sample in question.
- The aggregation behaviour of mAbs in the EFD is distinct to that observed using alternative biophysical methods.

The implications of this work are:

- Potential for future experiments exploring the mechanisms of flow-induced protein aggregation within this author's wider research group and others in the field, building on the data presented in this thesis.
- The EFD and similar devices could be used to screen mAbs, at small-scale, for their 'manufacturability', i.e. their relative resistance to hydrodynamic forces. Alternative biophysical methods currently used by industry do not provide one with this information, as the EFD has been shown to be distinct to other assays. As mAbs will encounter extensional and shear flows during bioprocessing, identification of robust mAbs earlier on in the development pipeline would be beneficial to the biopharmaceutical industry.
- Knowledge of strain rates and exposure times in the EFD could be compared to those found in industry to inform manufacturing practices which minimise aggregation.
- The EFD could be used as a formulation screening tool to stabilise aggregation-prone mAbs (such as WFL) against harsh hydrodynamic environments in which they would usually aggregate.

8. Appendices

Appendix 1. Ancillary Tables

Table A1.1 Molecular Weights and molar extinction coefficients for the proteins used in the study

Protein	Molecular Weight (Da)	Molar extinction coefficient at 280 nm (ϵ_{280}) ($M^{-1} cm^{-1}$).
BSA	66,430	43,824
WFL	148,422	239,440
mAb1	147,483	207,360
STT	148,107	228,440
Daratumumab	147,903	210,000
Denosumab	147,797	210,000
Tabalumab	149,107	210,000
Ipilimumab	147,933	210,000
Ixekizumab	149,357	210,000
Olaratumab	152,607	210,000
Panobacumab	148,740	210,000
Vedolizumab	149,491	210,000
Fulranumab	148,464	210,000
Imgatuzumab	147,883	210,000
Dalotuzumab	148,852	210,000
Tralokinumab	147,052	210,000
Onartuzumab	149,813	210,000
Ofatumumab	148,574	210,000
Etrolizumab	146,944	210,000
Guselkumab	146,450	210,000
Foralumab	148,164	210,000
Fezakinumab	147,302	210,000
Daclizumab	146,720	210,000
Figitumumab	149,084	210,000
Simtuzumab	148,302	210,000
Zalutumumab	149,121	210,000
Golimumab	149,543	210,000
Galiximab	147,805	210,000
Otlertuzumab	147,295	210,000
Bevakizumab	149,132	210,000
Alemtuzumab	148,652	210,000
Bimagrumab	145,354	210,000
Lumiliximab	150,312	210,000
Sifalimumab	146,838	210,000
Cixutumumab	148,964	210,000
Adalimumab	148,079	210,000
Rituximab	147,130	210,000

Table A1.2 *pl*, Intrinsic CamSol score at pH 7, Light-chain class and Cluster number from Jain et al. 2017 for the Adimab proteins

Name of mAb	<i>pl</i>	CamSol Score (pH 7)	λ or κ ?	Cluster No.
Daratumumab	7.9	0.190311	κ	1
Tabalumab	8.9	0.296595	κ	2
Denosumab	8.8	-0.437674	κ	1
Ipilimumab	8.9	-0.105756	κ	3
Ixekizumab	8.8	-0.199419	κ	5
Olaratumab	8.3	-0.545537	κ	4
Panobacumab	8.7	0.17024	κ	1
Vedolizumab	6.9	-0.289134	κ	1
Fulranumab	7.9	0.128657	κ	2
Imgatuzumab	9.1	0.019702	κ	2
Dalotuzumab	8.8	-0.62771	κ	5
Tralokinumab	5.9	0.020284	λ	1
Onartuzumab	8.3	-0.613459	κ	1
Ofatumumab	8.4	0.038952	κ	1
Etrolizumab	8.4	0.02221	κ	4
Guselkumab	8.6	-0.214467	λ	4
Foralumab	9.3	-0.117484	κ	2
Fezakinumab	6.9	0.050742	λ	3
Daclizumab	8.6	-0.071961	κ	1
Figitumumab	8.8	0.146984	κ	2
Simtuzumab	8.6	-0.600181	κ	5
Zalutumumab	7.8	-0.067162	κ	1
Golimumab	8.8	-0.346518	κ	3
Galiximab	5.8	-0.339799	λ	3
Otlertuzumab	8.3	-0.039988	κ	1
Bevakizumab	7.1	-0.627061	κ	4
Alemtuzumab	9.1	-0.303701	κ	1
Bimagrumab	9.1	-0.15951	λ	2
Lumiliximab	8.4	-0.204455	κ	1
Sifalimumab	9.0	-0.204455	κ	1
Cixutumumab	7.8	-0.149332	λ	5
Adalimumab	8.3	0.36987	κ	1
Rituximab	9.1	-0.605734	κ	1

Table A2: Table of mean hydrodynamic radii (nm) obtained by regularization analysis of DLS data for 1, 2, 5 and 10 mg mL⁻¹ BSA subjected to extensional flow in Chapter 3.2.2. The protein samples were stressed at a plunger velocity of 8 mm s⁻¹ (strain rate = 11750 s⁻¹ and shear rate = 52000 s⁻¹). One representative sample is shown from N = 2 experiments, except where indicated.

Protein Conc. (mg mL ⁻¹)	No. of passes	Peak 1 (nm)	Peak 2 (nm)	Peak 3 (nm)	Peak 4 (nm)
1	0	3.4 ± 0.9			
	10	3.1±0.6	1748±434		
	20	3.3±0.9			
	50	3.2±0.8	5955±2106		
	100	3.5±1.2	310±80		
	500	3.3±0.66	77±25	994±357	
	1000	7.0±5.7	1422511450		
	1500	5.2±24	16490±16460		
	2000	3.0± 0.5	139± 20	1915± 576	140564 ± 172260
2	0	3.3±1			
	10	3.2±0.7			
	20	3.2±0.8			
	50	2.9±0.8	857±812	367715±23735 1	
	100	2.2±0.6	98±32	693±227	16215 ±5780
	500	4.1±1.1	887±705	26911±12802	
N =1	1000	11.2±3.8	58522±129872		
	1500	4.4± 2.2	24678±29920		
	2000	8.4±1.6	1230±571	27912±12189	
5	0	3.2±0.8			
	10	3.4±1.0	506±157		
	20	3.3±0.8	292±76		
	50	3.1±0.6	2146±700	52739±24392	
	100	2.5±0.9	36±9	31158±7720	
	500	3.2±1	352±100	37600±8400	
	1000	4.6±1.4	22636±34198		
	1500	4.2±2.6	13983±20580		
	2000	3.7±0.8	82±21	26683±39806	
10	0	3.3±1.0			
	10	3.3±0.9			
	20	2.7±0.3	15598±4812		
	50	2.4±0.3	520±140	3713±1268	
	100	2.9±1.1	5877±12817		
	500	3.5±0.9	65±9	12383±12494	
	1000	2.8±0.5	26±5	2896±2044	65301 ±23533
	1500	3.7±1.7	531±441	12722±5140	
	2000	5.4±1.3	15623±33429		

Table A3. “Thresholds of biophysical properties derived from analysis on samples corresponding to the 48 approved antibodies”, i.e. ‘red flag’ thresholds from Jain et al., 2017

Group	Assay	Worst 10% threshold	Units (of the flag)
1	PSR	0.27 ± 0.06	N/A
	AC-SINS	11.8 ± 6.2	nm ($\Delta\lambda$ >)
	CSI	0.01 ± 0.02	BLI response units (>)
2	CIC	10.1 ± 0.5	Retention time (mins) (>)
	HIC	11.7 ± 0.6	Retention time (mins) (>)
	SMAC	12.8 ± 1.2	Retention time (mins) (>)
	SGAC-SINS	370 ± 133	Salt concentration (mM) (>)
3	BVP	4.3 ± 2.2	Fold-over-background (>)
	ELISA	1.9 ± 1.0	Fold-over-background (>)
4	AS	0.08 ± 0.03	Monomer loss % day ⁻¹ (>)

Table A4. Table of dissociation constants for selected Adimab proteins with their targets. K_{DS} were extracted from the literature, with the appropriate references highlighted. Many references were sourced from the IUPHAR Guide to Immunopharmacology, available at <http://www.guidetoimmunopharmacology.org/immuno/index.jsp>

mAb Name	K_D (nM)	Reference
Tabalumab	0.126	(Manetta et al., 2014)
Denosumab	0.003	(Casas et al., 2013)
Ipilimumab	10.6	(Ramagopal et al., 2017)
Ixekizumab	0.0018	(Liu et al., 2016)
Olaratumab	0.04	(Loizos and Huber, 2006)
Panobacumab	0.003	(Horn et al., 2010)
Fulranumab	0.004	(Wild Jr. et al., 2004)
Imgatuzumab	18	(Schanzer et al., 2016)
Dalotuzumab	1	(Brana et al., 2014)
Tralokinumab	0.164	(Thom and Minter, 2012)
Ofatumumab	0.04	(Liu, 2011)
Foralumab	0.01	(Dean and Depis, 2014)
Fezakinumab	0.054	(Fouser et al., 2007)
Daclizumab	10	(Queen et al., 1995)
Figitumumab	6.6	(Collens et al., 2014)
Simtuzumab	67	(Smith et al., 2013)
Zalutumumab	7	(Bleeker et al., 2004)
Golimumab	0.018	(Shealy et al., 2010)
Galiximab	200	(Mak and Saunders, 2006)
Bevakizumab	0.058	(Papadopoulos et al., 2012)
Alemtuzumab	7500	(Silicon Kinetics, 2016)
Bimagrumab	0.00173	(Berger et al., 2010)
Sifalimumab	0.044	(Oganesyan et al., 2015)
Cixutumumab	0.04	(McKian and Haluska, 2009)
Adalimumab	0.46	(Tracey et al., 2008)
Rituximab	3.1	(Macklin et al., 2015)

Appendix 2. Ancillary Figures

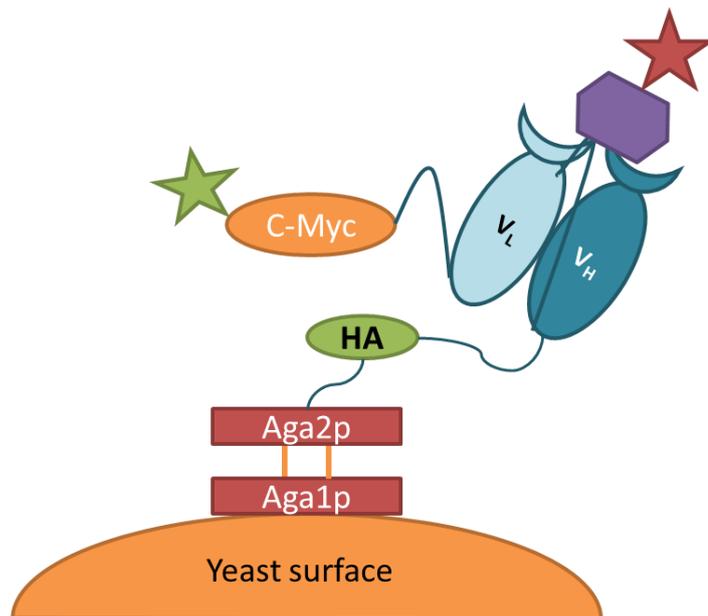


Figure A1. Outline of yeast surface display. The scFv of interest (V_L and V_H) is expressed on the surface of the yeast as a fusion protein with Aga2, linked via hemagglutinin. Orange bars represent disulphide bridges. Blue lines represent flexible glycine-serine linkers. The biotinylated antigen (purple) can have its binding detected by fluorescently labelled Avidin (red star). An epitope tag (c-Myc) can be used to detect expression of the scFv by using a fluorescently labelled antibody (green star). Figure adapted from (Chao et al., 2006).

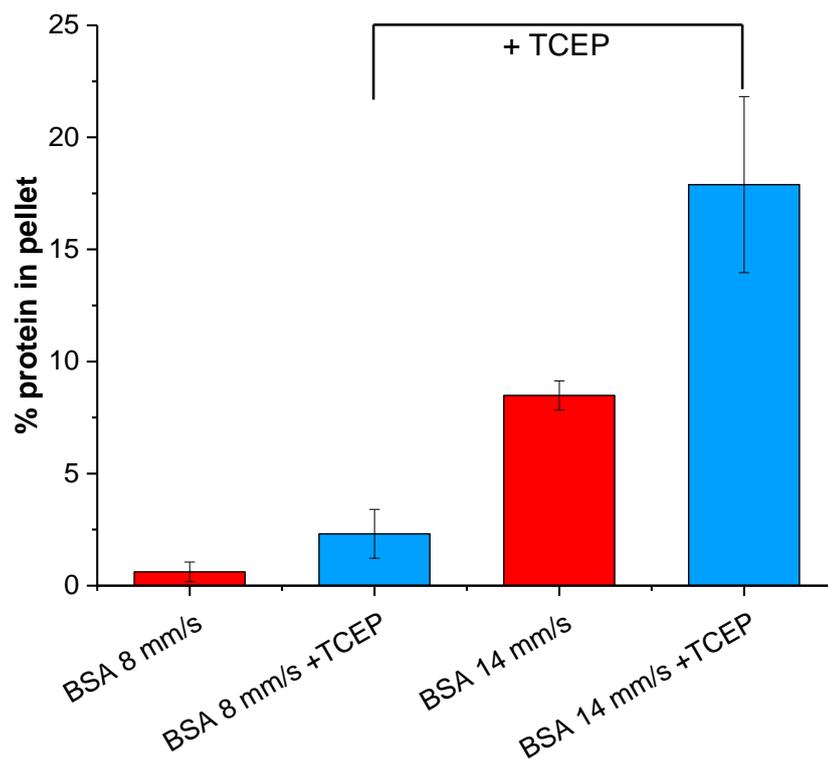


Figure A2. The amount of insoluble protein formed after stressing BSA in the presence and absence of reducing agent (TCEP). 5 mg mL⁻¹ BSA was stressed for 100 passes at a plunger velocity of 8 mm or 14 mm s⁻¹, in the presence/absence of 0.5 mM TCEP. Errors from two independent experiments were propagated according to Section 2.2.18.

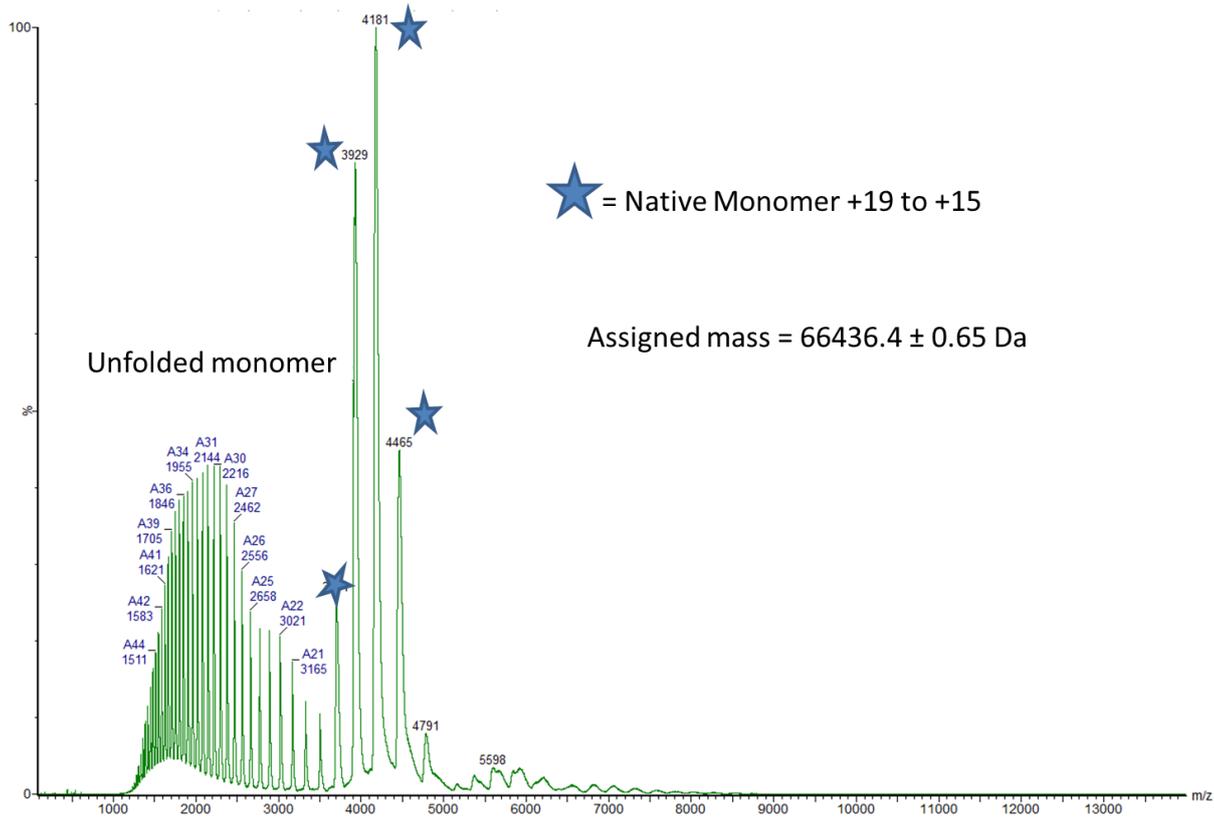


Figure A3. Native mass spectrum of 10 μ M BSA in 25 mM ammonium acetate, pH 3.0. Spectrometer conditions were as stated in Section 2.2.8.

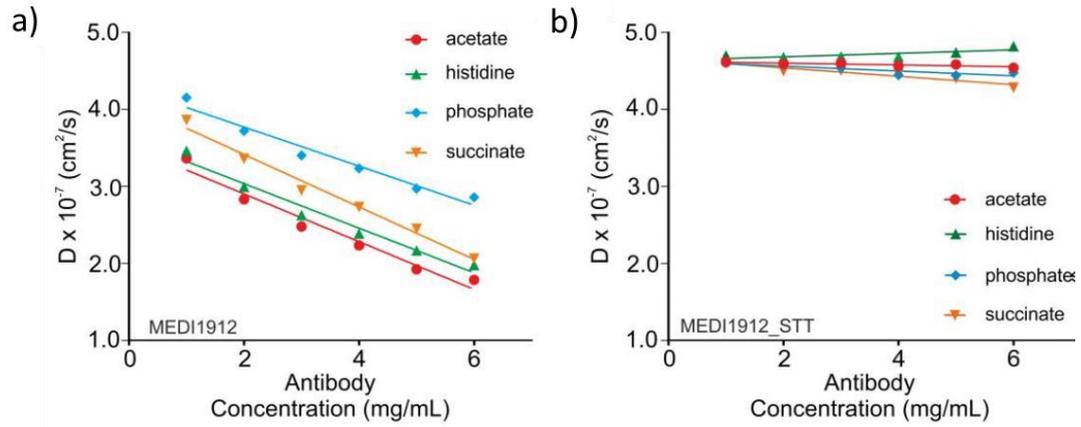


Figure A4. Change in diffusion coefficient (D) against the concentration of a) WFL and b) STT. The four buffers used were 10 mM sodium acetate pH 6.0 (red), 10 mM histidine pH 6.0 (green), 10 mM sodium phosphate pH 7.2 (blue) and 10 mM sodium succinate pH 6.0 (orange). Figure taken from Dobson et al. 2016.

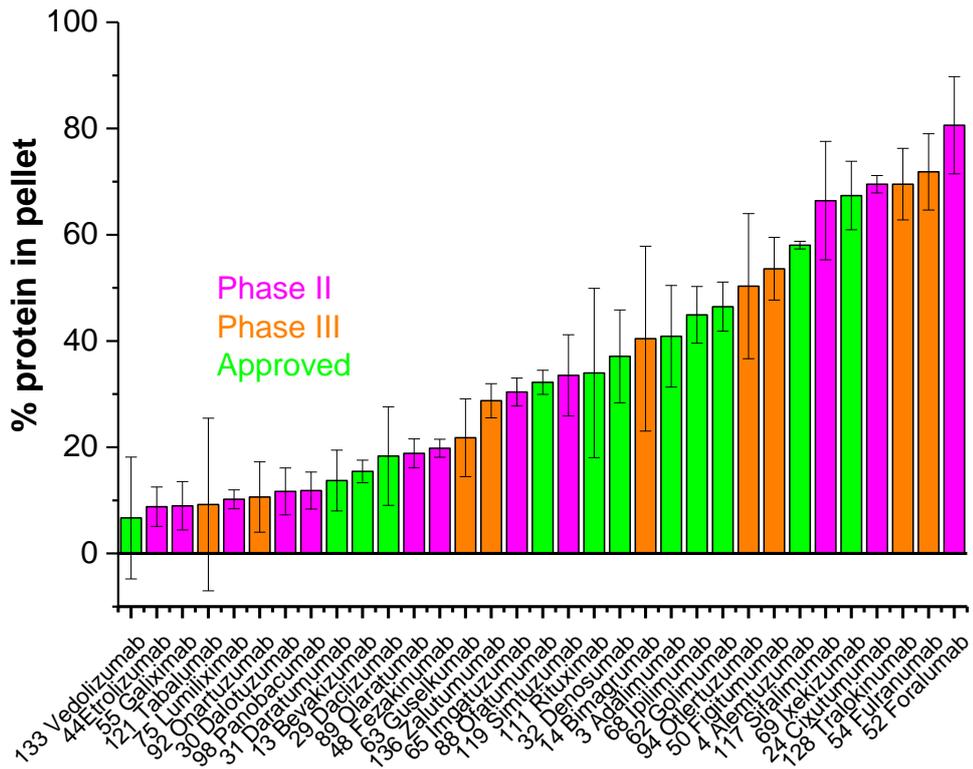


Figure A5. Plot of the amount of insoluble protein formed by the mAbs after stress in the EFD (from Figure 6.5) according to approval rating. Pink = Phase II, Orange = Phase III and Approved mAbs are in Green.

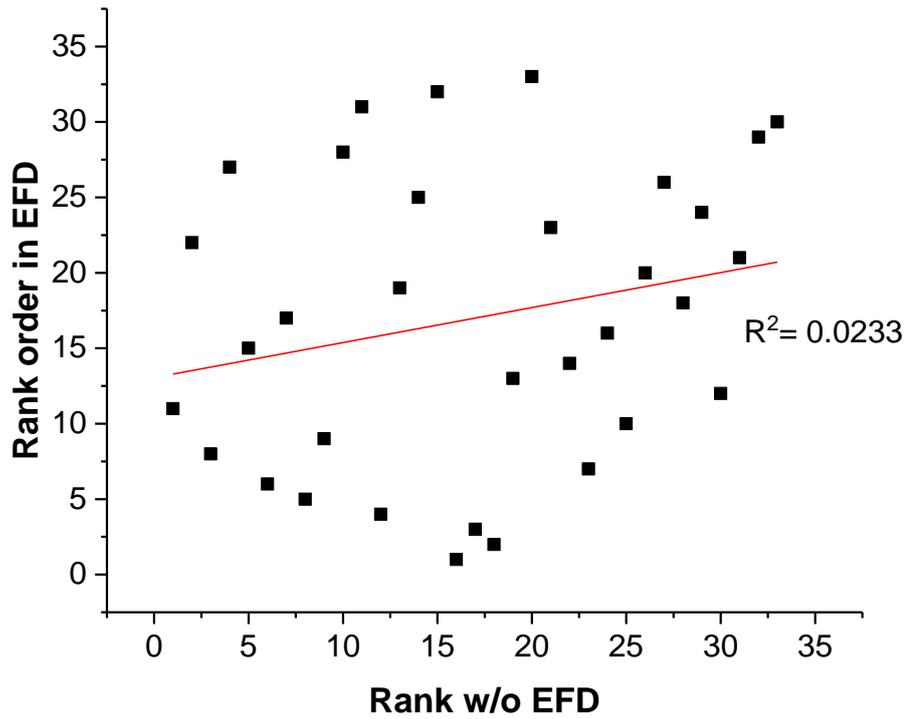


Figure A6. Plot of the rank order in the EFD of the 33 Adimab proteins against the rank derived from the distance from ideal analysis by Jain et al., excluding the EFD data. A linear trend line was fit to the data in Origin 2018 to show the lack of correlation between the EFD rank and that from Jain et al. 2017.

9. References

- Abaci HE, Shen Y-I, Tan S, Gerecht S. 2015. Recapitulating physiological and pathological shear stress and oxygen to model vasculature in health and disease. *Sci. Rep.* **4**:4951.
- Agarkhed M, O'Dell C, Hsieh M-C, Zhang J, Goldstein J, Srivastava A. 2017. Effect of Surfactants on Mechanical, Thermal, and Photostability of a Monoclonal Antibody. *AAPS PharmSciTech* **19**:79–92.
- Aggarwal SR. 2014. What's fueling the biotech engine — 2012 to 2013. *Nat. Biotechnol.* **32**:32–39.
- Akkapeddi P, Azizi S-A, Freedy AM, Cal PMSD, Gois PMP, Bernardes GJL. 2016. Construction of homogeneous antibody–drug conjugates using site-selective protein chemistry. *Chem. Sci.* **7**:2954–2963.
- Alam ME, Geng SB, Bender C, Ludwig SD, Linden L, Hoet R, Tessier PM. 2018. Biophysical and Sequence-Based Methods for Identifying Monovalent and Bivalent Antibodies with High Colloidal Stability. *Mol. Pharm.* **15**:150–163.
- Allmendinger A, Mueller R, Huwyler J, Mahler HC, Fischer S. 2015. Sterile Filtration of Highly Concentrated Protein Formulations: Impact of Protein Concentration, Formulation Composition, and Filter Material. *J. Pharm. Sci.* **104**:3319–3329.
- Andersson M, Jia Q, Abella A, Lee X-Y, Landreh M, Purhonen P, Hebert H, Tenje M, Robinson C V, Meng Q, Plaza GR, Johansson J, Rising A. 2017. Biomimetic spinning of artificial spider silk from a chimeric minispidroin. *Nat. Chem. Biol.* **13**:262–264.
- Andrews JM, Roberts CJ. 2007. A Lumry-Eyring nucleated polymerization model of protein aggregation kinetics: 1. Aggregation with pre-equilibrated unfolding. *J. Phys. Chem. B* **111**:7897–7913.
- Anfinsen. 1973. Principles that Govern the Folding of Protein Chains. *Science.* **181**:223–230.
- Anfinsen CB, Haber E, Sela M, White FH. 1961. the Kinetics of Formation of Native Ribonuclease During Oxidation of the Reduced Polypeptide Chain. *Proc. Natl. Acad. Sci.* **47**:1309–1314.
- Aponte-Santamaría C, Huck V, Posch S, Bronowska AK, Grässle S, Brehm MA,

- Obser T, Schneppenheim R, Hinterdorfer P, Schneider SW, Baldauf C, Gräter F. 2015. Force-Sensitive Autoinhibition of the von Willebrand Factor Is Mediated by Interdomain Interactions. *Biophys. J.* **108**:2312–2321.
- Arakawa T, Philo JS, Ejima D, Tsumoto K, Arisaka F. 2006. Aggregation Analysis of Therapeutic Proteins. *Bioprocess Int.* **4**:42–43.
- Aris R. 1962. Vectors, Tensors, and the Basic Equations of Fluid Mechanics. In: , pp. 82–98.
- Arosio P, Müller T, Rajah L, Yates E V., Aprile FA, Zhang Y, Cohen SIA, White DA, Herling TW, De Genst EJ, Linse S, Vendruscolo M, Dobson CM, Knowles TPJ. 2016. Microfluidic diffusion analysis of the sizes and interactions of proteins under native solution conditions. *ACS Nano* **10**:333–341.
- Arosio P, Rima S, Lattuada M, Morbidelli M. 2012. Population balance modeling of antibodies aggregation kinetics. *J. Phys. Chem. B* **116**:7066–7075.
- Ashton L, Dusting J, Imomoh E, Balabani S, Blanch EW. 2010. Susceptibility of different proteins to flow-induced conformational changes monitored with Raman spectroscopy. *Biophys. J.* **98**:707–714.
- Ashton L, Dusting J, Imomoh E, Balabani S, Blanch EW. 2009. Shear-induced unfolding of lysozyme monitored in situ. *Biophys. J.* **96**:4231–6.
- Atkins P, de Paula J. 2010. Atkins' Physical Chemistry 9th ed. Oxford: Oxford University Press 972 p.
- Atmeh RF, Arafa IM, Al-Khateeb M. 2007. Albumin Aggregates: Hydrodynamic Shape and Physico-Chemical Properties. *Jordan J. Chem.* **2**:169–182.
- Ayyar V, Arora S, O'Kennedy R. 2016. Coming-of-Age of Antibodies in Cancer Therapeutics **37**:1009–1028.
- Azevedo AM, Rosa PAJ, Ferreira IF, Aires-Barros MR. 2009. Chromatography-free recovery of biopharmaceuticals through aqueous two-phase processing. *Trends Biotechnol.* **27**:240–247.
- Azzi A. 1974. The Use of Fluorescent Probes for the Study of Membranes. *Methods Enzymol.* **32**:234–246.
- Bae YB, Jang HK, Shin TH, Phukan G, Tran TT, Lee G, Hwang WR, Kim JM. 2016. Microfluidic assessment of mechanical cell damage by extensional stress. *Lab Chip* **16**:96–103.

- Baek Y, Zydney AL. 2018. Intermolecular interactions in highly concentrated formulations of recombinant therapeutic proteins. *Curr. Opin. Biotechnol.* **53**:59–64.
- Bajgain P, Tawinwung S, Elia LD, Sukumaran S, Watanabe N, Hoyos V, Lulla P, Brenner MK, Leen AM, Vera JF. 2018. CAR T cell therapy for breast cancer: harnessing the tumor milieu to drive T cell activation. *J. Immunother. Cancer* **6**:1–13.
- Balaguru UM, Sundaresan L, Manivannan J, Majunathan R, Mani K, Swaminathan A, Venkatesan S, Kasiviswanathan D, Chatterjee S. 2016. Disturbed flow mediated modulation of shear forces on endothelial plane: A proposed model for studying endothelium around atherosclerotic plaques. *Sci. Rep.* **6**:27304.
- Barnard JG, Singh S, Randolph TW, Carpenter JF. 2011. Subvisible particle counting provides a sensitive method of detecting and quantifying aggregation of monoclonal antibody caused by freeze-thawing: insights into the roles of particles in the protein aggregation pathway. *J. Pharm. Sci.* **100**:492–503.
- Barnett G V., Razinkov VI, Kerwin BA, Laue TM, Woodka AH, Butler PD, Perevozchikova T, Roberts CJ. 2015. Specific-Ion Effects on the Aggregation Mechanisms and Protein–Protein Interactions for Anti-streptavidin Immunoglobulin Gamma-1. *J. Phys. Chem. B* **119**:5793–5804.
- Baynes BM, Wang DIC, Trout BL. 2005. Role of arginine in the stabilization of proteins against aggregation. *Biochemistry* **44**:4919–4925.
- Bee JS, Schwartz DK, Trabelsi S, Freund E, Stevenson JL, Carpenter JF, Randolph TW. 2012. Production of particles of therapeutic proteins at the air–water interface during compression/dilation cycles. *Soft Matter* **8**:10329.
- Bee JS, Stevenson JL, Mehta B, Svitel J, Pollastrini J, Platz R, Freund E, Carpenter JF, Randolph TW. 2009. Response of a concentrated monoclonal antibody formulation to high shear. *Biotechnol. Bioeng.* **103**:936–943.
- Bee JS, Davis M, Freund E, Carpenter JF, Randolph TW. 2010. Aggregation of a monoclonal antibody induced by adsorption to stainless steel. *Biotechnol. Bioeng.* **105**:121–9.
- Bee JS, Randolph TW, Carpenter JF, Bishop SM, Dimitrova MN. 2011. Effects of surfaces and leachables on the stability of biopharmaceuticals. *J. Pharm. Sci.* **100**:4158–4170.

- Bekard IB, Asimakis P, Teoh CL, Ryan T, Howlett GJ, Bertolini J, Dunstan DE. 2012. Bovine serum albumin unfolds in Couette flow. *Soft Matter* **8**:385.
- Bekard IB, Asimakis P, Bertolini J, Dunstan DE. 2011. The effects of shear flow on protein structure and function. *Biopolymers* **95**:733–45.
- Bekard IB, Dunstan DE. 2009. Shear-induced deformation of bovine insulin in Couette flow. *J. Phys. Chem. B* **113**:8453–7.
- Berger C, Herrmann CL, Sheppard K-A, Trifilieff E, Urlinger S. 2010. Compositions and methods for increasing muscle growth. US8388968B2.
- Berkowitz SA, Engen JR, Mazzeo JR, Jones GB. 2012. Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars. *Nat. Rev. Drug Discov.* **11**:527–40.
- Bhattacharya A, Prajapati R, Chatterjee S, Mukherjee TK. 2014. Concentration-dependent reversible self-oligomerization of serum albumins through intermolecular β -sheet formation. *Langmuir* **30**:14894–14904.
- Bhattacharya M, Jain N, Bhasne K, Kumari V, Mukhopadhyay S. 2011. pH-Induced conformational isomerization of bovine serum albumin studied by extrinsic and intrinsic protein fluorescence. *J. Fluoresc.* **21**:1083–1090.
- Bhirde AA, Chiang MJ, Venna R, Beaucage S, Brorson K. 2018. High-Throughput In-Use and Stress Size Stability Screening of Protein Therapeutics Using Algorithm-Driven Dynamic Light Scattering. *J. Pharm. Sci.* **107**:2055–2062.
- Biddlecombe JG, Craig A V., Zhang H, Uddin S, Mulot S, Fish BC, Bracewell DG. 2007. Determining antibody stability: Creation of solid - Liquid interfacial effects within a high shear environment. *Biotechnol. Prog.* **23**:1218–1222.
- Biddlecombe JG, Smith G, Uddin S, Mulot S, Spencer D, Gee C, Fish BC, Bracewell DG. 2009. Factors influencing antibody stability at solid-liquid interfaces in a high shear environment. *Biotechnol. Prog.* **25**:1499–1507.
- Birch JR, Racher AJ. 2006. Antibody production. *Adv. Drug Deliv. Rev.* **58**:671–685.
- Bleeker WK, Lammerts van Bueren JJ, van Ojik HH, Gerritsen AF, Pluyter M, Houtkamp M, Halk E, Goldstein J, Schuurman J, van Dijk MA, van de Winkel JGJ, Parren PWHI. 2004. Dual mode of action of a human anti-epidermal growth factor receptor monoclonal antibody for cancer therapy. *J. Immunol.* **173**:4699–707.

- Bobst CE, Abzalimov RR, Houde D, Kloczewiak M, Mhatre R, Berkowitz SA, Kaltashov IA. 2008. Detection and characterization of altered conformations of protein pharmaceuticals using complementary mass spectrometry-based approaches. *Anal. Chem.* **80**:7473–81.
- Boles KS, Kannan K, Gill J, Felderman M, Gouvis H, Hubby B, Kamrud KI, Venter JC, Gibson DG. 2017. Digital-to-biological converter for on-demand production of biologics. *Nat. Biotechnol.* **35**:672–675.
- Borotto NB, Zhou Y, Hollingsworth SR, Hale JE, Graban EM, Vaughan RC, Vachet RW. 2015. Investigating Therapeutic Protein Structure with Diethylpyrocarbonate Labeling and Mass Spectrometry. *Anal. Chem.* **87**:10627–10634.
- Borzova VA, Markossian KA, Kleymenov SY, Kurganov BI. 2017. A change in the aggregation pathway of bovine serum albumin in the presence of arginine and its derivatives. *Sci. Rep.* **7**:3984.
- Borzova VA, Markossian KA, Kara DA, Kurganov B. 2015. Kinetic regime of dithiothreitol-induced aggregation of bovine serum albumin. *Int. J. Biol. Macromol.* **80**:130–138.
- BPSA Extractables and Leachables Subcommittee. 2007. Recommendations for extractables and leachables testing. *Bioprocess Int.* **5**:36–44.
- Brana I, Berger R, Golan T, Haluska P, Edenfield J, Fiorica J, Stephenson J, Martin LP, Westin S, Hanjani P, Jones MB, Almhanna K, Wenham RM, Sullivan DM, Dalton WS, Gunchenko A, Cheng JD, Siu LL, Gray JE. 2014. A parallel-arm phase I trial of the humanised anti-IGF-1R antibody dalotuzumab in combination with the AKT inhibitor MK-2206, the mTOR inhibitor ridaforolimus, or the NOTCH inhibitor MK-0752, in patients with advanced solid tumours. *Br. J. Cancer* **111**:1932–1944.
- Brekke OH, Sandlie I. 2003. Therapeutic antibodies for human diseases at the dawn of the twenty-first century. *Nat. Rev. Drug Discov.* **2**:52–62.
- Bria CRM, Jones J, Charlesworth A, Ratanathanawongs Williams SK. 2016. Probing Submicron Aggregation Kinetics of an IgG Protein by Asymmetrical Flow Field-Flow Fractionation. *J. Pharm. Sci.* **105**:31–39.
- Brückl L, Schröder T, Scheler S, Hahn R, Sonderegger C. 2016. The Effect of Shear on the Structural Conformation of rhGH and IgG1 in Free Solution. *J. Pharm.*

Sci. **105**:1810–1818.

Buss NAPS, Henderson SJ, McFarlane M, Shenton JM, De Haan L. 2012. Monoclonal antibody therapeutics: History and future. *Curr. Opin. Pharmacol.* **12**:615–622.

Büttel IC, Chamberlain P, Chowers Y, Ehmann F, Greinacher A, Jefferis R, Kramer D, Kropshofer H, Lloyd P, Lubiniecki A, Krause R, Mire-Sluis A, Platts-Mills T, Ragheb JA, Reipert BM, Schellekens H, Seitz R, Stas P, Subramanyam M, Thorpe R, Trouvin J-H, Weise M, Windisch J, Schneider CK. 2011. Taking immunogenicity assessment of therapeutic proteins to the next level. *Biologicals* **39**:100–109.

Campbell I. 2012. *Biophysical Techniques*. Oxford: Oxford University Press 95-106, 151-152 p.

Capela EV, Aires-Barros MR, Freire MG, Azevedo AM. 2017. Monoclonal Antibodies–Addressing the Challenges on the Manufacturing Processing of an Advanced Class of Therapeutic Agents. *Front. Clin. Drug Res. Infect. Vol. 4* **4**:142–203.

Capelle MAH, Gurny R, Arvinte T. 2007. High throughput screening of protein formulation stability: Practical considerations. *Eur. J. Pharm. Biopharm.* **65**:131–148.

Carboni EJ, Bognet BH, Bouchillon GM, Kadilak AL, Shor LM, Ward MD, Ma AWK. 2016. Direct Tracking of Particles and Quantification of Margination in Blood Flow. *Biophys. J.* **111**:1487–1495.

Carboni EJ, Bognet BH, Cowles DB, Ma AWK. 2018. The Margination of Particles in Areas of Constricted Blood Flow. *Biophys. J.* **114**:2221–2230.

Carpenter JF, Kendrick BS, Chang BS, Manning MC, Randolph TW, Carpenter BJB, Kendrick BS, Chang BS, Manning MC, Randolph TW. 1999. Inhibition of Stress-induced Aggregation of Protein Therapeutics. *Methods Enzymol.* **5**:236–255.

Carpenter JF, Randolph TW, Jiskoot W, Crommelin DJ a, Middaugh CR, Winter G. 2010. Potential inaccurate quantitation and sizing of protein aggregates by size exclusion chromatography: essential need to use orthogonal methods to assure the quality of therapeutic protein products. *J. Pharm. Sci.* **99**:2200–8.

Carr B, Wright M. 2015. Nanoparticle Tracking Analysis: A review of the first 1,000

reports of applications and usage of NTA 1-239 p.

- Carter PJ, Lazar GA. 2018. Next generation antibody drugs: Pursuit of the “high-hanging fruit.” *Nat. Rev. Drug Discov.* **17**:197–223.
- Casas A, Llombart A, Martín M. 2013. Denosumab for the treatment of bone metastases in advanced breast cancer. *Breast* **22**:585–592.
- Castellanos MM, Pathak JA, Colby RH. 2014. Both protein adsorption and aggregation contribute to shear yielding and viscosity increase in protein solutions. *Soft Matter* **10**:122–31.
- Castro GR, Larson BK, Panilaitis B, Kaplan DL. 2005. Emulsan quantitation by Nile red quenching fluorescence assay. *Appl. Microbiol. Biotechnol.* **67**:767–770.
- Cathey CA, Fuller GG. 1990. The optical and mechanical response of flexible polymer solutions to extensional flow. *J. Nonnewton. Fluid Mech.* **34**:63–88.
- Chakroun N, Hilton D, Ahmad SS, Platt GW, Dalby PA. 2016. Mapping the Aggregation Kinetics of a Therapeutic Antibody Fragment. *Mol. Pharm.* **13**:307–319.
- Challener CA. 2017. Single-use bioreactors have reached the big time. *BioPharm Int.* **30**:18–23.
- Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, Wittrup KD. 2006. Isolating and engineering human antibodies using yeast surface display. *Nat. Protoc.* **1**:755–768.
- Charm SE, Wong BL. 1970a. Enzyme inactivation with shearing. *Biotechnol. Bioeng.* **12**:1103–1109.
- Charm SE, Wong BL. 1970b. Shear Degradation of Fibrinogen in the Circulation. *Science.* **170**:466–468.
- Chaudhuri R, Cheng Y, Middaugh CR, Volkin D. 2014. High-Throughput Biophysical Analysis of Protein Therapeutics to Examine Interrelationships Between Aggregate Formation and Conformational Stability. *AAPS J.* **16**:48–64.
- Chen G, Warrack BM, Goodenough AK, Wei H, Wang-Iverson DB, Tymiak AA. 2011. Characterization of protein therapeutics by mass spectrometry: Recent developments and future directions. *Drug Discov. Today* **16**:58–64.
- Chen Y, Kim MT, Zheng L, Deperalta G, Jacobson F. 2016. Structural

- Characterization of Cross-Linked Species in Trastuzumab Emtansine (Kadcyla). *Bioconjug. Chem.* **27**:2037–2047.
- Chen Y, Barkley MD. 1998. Toward Understanding Tryptophan Fluorescence in Proteins. *Biochemistry* **37**:9976–9982.
- Chennamsetty N, Voynov V, Kayser V, Helk B, Trout BL. 2009. Design of therapeutic proteins with enhanced stability. *Proc. Natl. Acad. Sci.* **106**:11937–11942.
- Cheung MS, Garcia AE, Onuchic JN. 2002. Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci.* **99**:685–690.
- Chisholm CF, Nguyen BH, Soucie KR, Torres RM, Carpenter JF, Randolph TW. 2015. In Vivo Analysis of the Potency of Silicone Oil Microdroplets as Immunological Adjuvants in Protein Formulations. *J. Pharm. Sci.* **104**:3681–3690.
- Chisti Y. 2006. Bioreactor design. In: Kristiansen, B, Ratledge, C, editors. *Basic Biotechnol.* 3rd ed. Cambridge: Cambridge University Press, pp. 181–200.
- Chiti F, Dobson CM. 2009. Amyloid formation by globular proteins under native conditions. *Nat. Chem. Biol.* **5**:15–22.
- Chiu ML, Gilliland GL. 2016. Engineering antibody therapeutics. *Curr. Opin. Struct. Biol.* **38**:163–173.
- Christie M, Torres RM, Kedl RM, Randolph TW, Carpenter JF. 2014. Recombinant murine growth hormone particles are more immunogenic with intravenous than subcutaneous administration. *J. Pharm. Sci.* **103**:128–139.
- Chudasama V, Maruani A, Caddick S. 2016. Recent advances in the construction of antibody-drug conjugates. *Nat. Chem.* **8**:114–9.
- Clark M. 2006. Basic Biotechnology. Ed. C Ratledge, B Kristiansen 3rd ed. Cambridge: Cambridge University Press 627-655 p.
- Clarke J, Itzhakit LS. 1998. Hydrogen exchange and protein folding. *Curr. Opin. Struct. Biol.* **8**:112–118.
- Collens S, Forbes J, Simms R, Wu T, Drewery J, Hu M, Stephenson K, Burak E, McCarthy T, Valliant J. 2014. In vitro residualization and binding of figitumumab radioimmunoconjugates. *J. Nucl. Med.* **55**:1190.

- Conchillo-Solé O, de Groot NS, Avilés FX, Vendrell J, Daura X, Ventura S. 2007. AGGREGSCAN: A server for the prediction and evaluation of “hot spots” of aggregation in polypeptides. *BMC Bioinformatics* **8**:65.
- Costa FF. 2014. Big data in biomedicine. *Drug Discov. Today* **19**:433–440.
- Cromwell MEM, Hilario E, Jacobson F. 2006. Protein aggregation and bioprocessing. *AAPS J.* **8**:E572-9.
- Daugherty AL, Mersny RJ. 2006. Formulation and delivery issues for monoclonal antibody therapeutics. *Adv. Drug Deliv. Rev.* **58**:686–706.
- David C, Foley S, Mavon C, Enescu M. 2008. Reductive unfolding of serum albumins uncovered by Raman spectroscopy. *Biopolymers* **89**:623–634.
- Davies PF. 2009. Hemodynamic shear stress and the endothelium in cardiovascular pathophysiology. *Nat. Clin. Pract. Cardiovasc. Med.* **6**:16–26.
- Dean Y, Depis F. 2014. Combination Therapies and Methods Using Anti-CD3 Modulating Agents and Anti-IL-6 Antagonists. US20140199299A1.
- Deng W, Xu Y, Chen W, Paul DS, Syed AK, Dragovich MA, Liang X, Zakas P, Berndt MC, Di Paola J, Ware J, Lanza F, Doering CB, Bergmeier W, Zhang XF, Li R. 2016. Platelet clearance via shear-induced unfolding of a membrane mechanoreceptor. *Nat. Commun.* **7**:12863.
- Desmyter A, Transue TR, Ghahroudi MA, Dao Thi M-H, Poortmans F, Hamers R, Muyldermans S, Wyns L. 1996. Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. *Nat. Struct. Biol.* **3**:803-811.
- Devine PWA, Fisher HC, Calabrese AN, Whelan F, Higazi DR, Potts JR, Lowe DC, Radford SE, Ashcroft AE. 2017. Investigating the Structural Compaction of Biomolecules Upon Transition to the Gas-Phase Using ESI-TWIMS-MS. *J. Am. Soc. Mass Spectrom.* **28**:1855–1862.
- Devine PWA. 2016. Defining the mechanism behind the self-association of therapeutic monoclonal antibodies using mass spectrometric techniques; University of Leeds.
- Dill KA, Chan HS. 1997. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**:10–19.
- DiMasi JA, Grabowski HG, Hansen RW. 2016. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J. Health Econ.* **47**:20–33.

- Dobson CM. 2004. Principles of protein folding, misfolding and aggregation. *Semin. Cell Dev. Biol.* **15**:3–16.
- Dobson CL, Devine PWA, Phillips JJ, Higazi DR, Lloyd C, Popovic B, Arnold J, Buchanan A, Lewis A, Goodman J, van der Walle CF, Thornton P, Vinall L, Lowne D, Aagaard A, Olsson L-L, Ridderstad Wollberg A, Welsh F, Karamanos TK, Pashley CL, Iadanza MG, Ranson NA, Ashcroft AE, Kippen AD, Vaughan TJ, Radford SE, Lowe DC. 2016. Engineering the surface properties of a human monoclonal antibody prevents self-association and rapid clearance in vivo. *Sci. Rep.* **6**:38644.
- Doronina SO, Toki BE, Torgov MY, Mendelsohn BA, Cerveny CG, Chace DF, DeBlanc RL, Gearing RP, Bovee TD, Siegall CB, Francisco JA, Wahl AF, Meyer DL, Senter PD. 2003. Development of potent monoclonal antibody auristatin conjugates for cancer therapy. *Nat. Biotechnol.* **21**:778–784.
- Dransart B, Wheeler A, Hong T, Tran C, Abalos R, Quezada A, Wang S, Kluck B, Sanaie N. 2018. Evaluation of a Next-Generation Protein A Chromatography Resin for the Purification of Monoclonal Antibodies. *Bioprocess Int.* **16**:14–19.
- Dubey K, Anand BG, Temgire MK, Kar K. 2014. Evidence of Rapid Coaggregation of Globular Proteins during Amyloid Formation. *Biochemistry* **53**:8001–8004.
- Ducry L. 2012. Challenges in the Development and Manufacturing of Antibody–Drug Conjugates BT - Therapeutic Proteins: Methods and Protocols. In: Voynov, V, Caravella, JA, editors. Totowa, NJ: Humana Press, pp. 489–497.
- Dudgeon K, Rouet R, Kokmeijer I, Schofield P, Stolp J, Langley D, Stock D, Christ D. 2012. General strategy for the generation of human antibody variable domains with increased aggregation resistance. *Proc. Natl. Acad. Sci.* **109**:10879–10884.
- Duerkop M, Berger E, Dürauer A, Jungbauer A. 2018. Impact of Cavitation, High Shear Stress and Air/Liquid Interfaces on Protein Aggregation. *Biotechnol. J.* **13**:1800062.
- Dunbar CE, High KA, Joung JK, Kohn DB, Ozawa K, Sadelain M. 2018. Gene therapy comes of age. *Science.* **359**:eaan4672.
- Dunstan DE, Hill EK, Wei Y. 2004. Direct Measurement of Polydiacetylene 4-Butoxycarbonylmethylurethane Segment Orientation and Distortion in Shear: Semidilute Solutions. *Macromolecules* **37**:1663–1665.

- van Durme J, De Baets G, Van Der Kant R, Ramakers M, Ganesan A, Wilkinson H, Gallardo R, Rousseau F, Schymkowitz J. 2016. Solubis: A webserver to reduce protein aggregation through mutation. *Protein Eng. Des. Sel.* **29**:285–289.
- Egan P, Sinko R, LeDuc PR, Keten S. 2015. The role of mechanics in biological and bio-inspired systems. *Nat. Commun.* **6**:7418.
- Eisenstein M. 2011. Something new under the skin. *Nat. Biotechnol.* **29**:107–9.
- Elgundi Z, Reslan M, Cruz E, Sifniotis V, Kayser V. 2017. The state-of-play and future of antibody therapeutics. *Adv. Drug Deliv. Rev.* **122**:2–19.
- Englander SW, Mayne L. 2014. The nature of protein folding pathways. *Proc. Natl. Acad. Sci. U. S. A.* **111**:15873–15880.
- Erickson HK, Park PU, Widdison WC, Kovtun Y V, Garrett LM, Hoffman K, Lutz RJ, Goldmacher VS, Blättler WA. 2006. Antibody-Maytansinoid Conjugates Are Activated in Targeted Cancer Cells by Lysosomal Degradation and Linker-Dependent Intracellular Processing. *Cancer Res.* **66**:4426 LP-4433.
- Eschbach G, Vermant S. 2008. TFF Membranes for High MAb Concentration. *Bioprocess Intl* **6**:66–69.
- Estep P, Caffry I, Yu Y, Sun T, Cao Y, Lynaugh H, Jain T, Vásquez M, Tessier PM, Xu Y. 2015. An alternative assay to hydrophobic interaction chromatography for high-throughput characterization of monoclonal antibodies. *MAbs* **7**:553–561.
- Fardin MA, Perge C, Taberlet N. 2014. “The hydrogen atom of fluid dynamics”-- introduction to the Taylor-Couette flow for soft matter scientists. *Soft Matter* **10**:3523–35.
- Feige MJ, Buchner J. 2014. Principles and engineering of antibody folding and assembly. *Biochim. Biophys. Acta - Proteins Proteomics* **1844**:2024–2031.
- Feldhaus MJ, Siegel RW, Opresko LK, Coleman JR, Feldhaus JMW, Yeung YA, Cochran JR, Heinzelman P, Colby D, Swers J, Graff C, Wiley HS, Wittrup KD. 2003. Flow-cytometric isolation of human antibodies from a nonimmune *Saccharomyces cerevisiae* surface display library. *Nat. Biotechnol.* **21**:163–170.
- Feng Y, Taraban MB, Yu YB. 2015. Water proton NMR - A sensitive probe for solute association. *Chem. Commun.* **51**:6804–6807.
- Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. 1990. Electrospray ionization-principles and practice. *Mass Spectrom. Rev.* **9**:37–70.

- Ferguson CN, Gucinski-Ruth AC. 2016. Evaluation of Ion Mobility-Mass Spectrometry for Comparative Analysis of Monoclonal Antibodies. *J. Am. Soc. Mass Spectrom.* **27**:822–833.
- Fernandez-Escamilla A-M, Rousseau F, Schymkowitz J, Serrano L. 2004. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* **22**:1302–6.
- Fersht A. 1999. Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding. *W. H. Free. Co.* pp. 615 p.
- Fieder J, Wagner N, Grammatikos S, Hoffmann H, Kaufmann H, Otto RT. 2005. Use of flow-cytometric analysis to optimize cell banking strategies for production of biopharmaceuticals from mammalian cells. *J. Biotechnol.* **120**:111–120.
- Filipe V, Hawe A, Jiskoot W. 2010. Critical evaluation of nanoparticle tracking analysis (NTA) by NanoSight for the measurement of nanoparticles and protein aggregates. *Pharm. Res.* **27**:796–810.
- Fischer K, Schmidt M. 2016. Pitfalls and novel applications of particle sizing by dynamic light scattering. *Biomaterials* **98**:79–91.
- Fitzgibbon S, Spann AP, Qi QM, Shaqfeh ESG. 2015. In Vitro Measurement of Particle Margination in the Microchannel Flow: Effect of Varying Hematocrit. *Biophys. J.* **108**:2601–2608.
- Fleischman ML, Chung J, Paul EP, Lewus RA. 2017. Shipping-Induced Aggregation in Therapeutic Antibodies: Utilization of a Scale-Down Model to Assess Degradation in Monoclonal Antibodies. *J. Pharm. Sci.* **106**:994–1000.
- Fouser LA, Hegen M, Luxenberg DP, O'Toole M. 2007. Methods of using antibodies against human IL-22 - Google Patents. US7811567B2.
- Franzreb M, Müller E, Vajda J. 2014. Cost Estimation for Protein A Chromatography An In Silico Approach to MAb Purification Strategy. *Bioprocess Int.* **12**:44–52.
- Fu H, Jiang Y, Yang D, Scheifflinger F, Wong WP, Springer TA. 2017. Flow-induced elongation of von Willebrand factor precedes tension-dependent activation. *Nat. Commun.* **8**.
- Gagnon P. 2012. Technology trends in antibody purification. *J. Chromatogr. A* **1221**:57–70.
- Ganesan A, Siekierska A, Beerten J, Brams M, Van Durme J, De Baets G, Van Der

- Kant R, Gallardo R, Ramakers M, Langenberg T, Wilkinson H, De Smet F, Ulens C, Rousseau F, Schymkowitz J. 2016. Structural hot spots for the solubility of globular proteins. *Nat. Commun.* **7**:10816.
- Garcia-Seisdedos H, Empereur-Mot C, Elad N, Levy ED. 2017. Proteins evolve on the edge of supramolecular self-assembly. *Nature* **548**:244.
- Garidel P, Kuhn AB, Schäfer L V., Karow-Zwick AR, Blech M. 2017. High-concentration protein formulations: How high is high? *Eur. J. Pharm. Biopharm.* **119**:353–360.
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. In: . *Proteomics Protoc. Handb.*, pp. 571–607.
- Gell C, Brockwell D, Smith A. 2006. Handbook of single molecule fluorescence spectroscopy. Oxford University Press 24-44 p.
- Gell C, Sabir T, Westwood J, Rashid A, Smith DAM, Harris S-A, Stockley PG. 2008. Single-Molecule Fluorescence Resonance Energy Transfer Assays Reveal Heterogeneous Folding Ensembles in a Simple RNA Stem–Loop. *J. Mol. Biol.* **384**:264–278.
- De Gennes PG. 1974. Coil-stretch transition of dilute flexible polymers under ultrahigh velocity gradients. *J. Chem. Phys.* **60**:5030-5042.
- Gerhardt A, McUmber AC, Nguyen BH, Lewus R, Schwartz DK, Carpenter JF, Randolph TW. 2015. Surfactant Effects on Particle Generation in Antibody Formulations in Pre-filled Syringes. *J. Pharm. Sci.* **104**:4056–4064.
- Gill SC, von Hippel PH. 1989. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**:319–326.
- Glish GL, Vachet RW. 2003. The basics of mass spectrometry in the twenty-first century. *Nat. Rev. Drug Discov.* **2**:140–150.
- de Goeij BE, Lambert JM. 2016. New developments for antibody-drug conjugate-based therapeutic approaches. *Curr. Opin. Immunol.* **40**:14–23.
- Goldberg DS, Lewus RA, Esfandiary R, Farkas DC, Mody N, Day KJ, Mallik P, Tracka MB, Sealey SK, Samra HS. 2017. Utility of High Throughput Screening Techniques to Predict Stability of Monoclonal Antibody Formulations During Early Stage Development. *J. Pharm. Sci.* **106**:1971–1977.

- Goldberg DS, Bishop SM, Shah AU, Sathish HA. 2011. Formulation development of therapeutic monoclonal antibodies using high-throughput fluorescence and static light scattering techniques: role of conformational and colloidal stability. *J. Pharm. Sci.* **100**:1306–1315.
- Goldrick S, Joseph A, Mollet M, Turner R, Gruber D, Farid SS, Titchener-Hooker NJ. 2017. Predicting performance of constant flow depth filtration using constant pressure filtration data. *J. Memb. Sci.* **531**:138–147.
- Götze M, Pettelkau J, Schaks S, Bosse K, Ihling CH, Krauth F, Fritzsche R, Kühn U, Sinz A. 2012. StavroX—A Software for Analyzing Crosslinked Products in Protein Interaction Studies. *J. Am. Soc. Mass Spectrom.* **23**:76–87.
- Grigolato F, Colombo C, Ferrari R, Rezabkova L, Arosio P. 2017. Mechanistic Origin of the Combined Effect of Surfaces and Mechanical Agitation on Amyloid Formation. *ACS Nano* **11**:11358–11367.
- de Groot NS, Pallarés I, Avilés FX, Vendrell J, Ventura S. 2005. Prediction of “hot spots” of aggregation in disease-linked polypeptides. *BMC Struct. Biol.* **5**:18.
- Groves MA, Nickson AA. 2012. Affinity maturation of phage display antibody populations using ribosome display. *Methods Mol. Biol.* **805**:163–190.
- Guan Y, Cao KJ, Cantlon A, Elbel K, Theodorakis EA, Walsh DM, Yang J, Shah J V. 2015. Real-Time Monitoring of Alzheimer’s-Related Amyloid Aggregation via Probe Enhancement-Fluorescence Correlation Spectroscopy. *ACS Chem. Neurosci.* **6**:1503–1508.
- Halfmann R, Lindquist S. 2008. Screening for Amyloid Aggregation by Semi-Denaturing Detergent-Agarose Gel Electrophoresis. *J. Vis. Exp.* **17**:11–13.
- Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hammers C, Songa EB, Bendahman N, Hammers R. 1993. Naturally occurring antibodies devoid of light chains. *Nature* **363**:446–448.
- Hames BD. 1990. Gel electrophoresis of proteins : A practical approach. Ed. B. D. Hames, D Rickwood 2nd ed. IRL Press at Oxford University Press 383 p.
- Hamrang Z, Hussain M, Tingey K, Tracka M, Casas-Finet JR, Uddin S, Van Der Walle CF, Pluen A. 2015. Characterisation of Stress-Induced Aggregate Size Distributions and Morphological Changes of a Bi-Specific Antibody Using Orthogonal Techniques. *J. Pharm. Sci.* **104**:2473–2481.

- Hanes J, Schaffitzel C, Knappik A, Plückthun A. 2000. Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nat. Biotechnol.* **18**:1287–1292.
- Hanlon AD, Larkin MI, Reddick RM. 2010. Free-solution, label-free protein-protein interactions characterized by dynamic light scattering. *Biophys. J.* **98**:297–304.
- Hariadi RF, Winfree E, Yurke B. 2015. Determining hydrodynamic forces in bursting bubbles using DNA nanotube mechanics. *Proc. Natl. Acad. Sci.* **112**:E6086–E6095.
- Harn N, Allan C, Oliver C, Middaugh CR. 2007. Highly concentrated monoclonal antibody solutions: direct analysis of physical structure and thermal stability. *J. Pharm. Sci.* **96**:532–46.
- Hassan PA, Rana S, Verma G. 2014. Making Sense of Brownian Motion: Colloid Characterization by Dynamic Light Scattering. *Langmuir* **31**:3–12.
- Haustein E, Schwille P. 2007. Fluorescence Correlation Spectroscopy: Novel Variations of an Established Technique. *Annu. Rev. Biophys. Biomol. Struct.* **36**:151–169.
- Hawe A, Hulse WL, Jiskoot W, Forbes RT. 2011. Taylor dispersion analysis compared to dynamic light scattering for the size analysis of therapeutic peptides and proteins and their aggregates. *Pharm. Res.* **28**:2302–10.
- Hawe A, Kasper JC, Friess W, Jiskoot W. 2009. Structural properties of monoclonal antibody aggregates induced by freeze-thawing and thermal stress. *Eur. J. Pharm. Sci.* **38**:79–87.
- He F, Phan DH, Hogan S, Bailey R, Becker GW, Narhi LO, Razinkov VI. 2010. Detection of IgG Aggregation by a High Throughput Method Based on Extrinsic Fluorescence. *J. Pharm. Sci.* **99**:2598–2608.
- He F, Woods CE, Becker GW, Narhi LO, Razinkov VI. 2011. High-throughput assessment of thermal and colloidal stability parameters for monoclonal antibody formulations. *J. Pharm. Sci.* **100**:5126–5141.
- Hoare M, Narendranathan TJ, Flint JR, Heywood-Waddington D, Bell DJ, Dunnill P. 1982. Disruption of protein precipitates during shear in Couette flow and in pumps. *Ind. Eng. Chem. Fundam.* **21**:402–406.
- Hofmann M, Gieseler H. 2018. Predictive Screening Tools Used in High-

- Concentration Protein Formulation Development. *J. Pharm. Sci.* **107**:772–777.
- Horn MP, Zuercher AW, Imboden MA, Rudolf MP, Lazar H, Wu H, Hoiby N, Fas SC, Lang AB. 2010. Preclinical In Vitro and In Vivo Characterization of the Fully Human Monoclonal IgM Antibody KBPA101 Specific for *Pseudomonas aeruginosa* Serotype IATS-O11. *Antimicrob. Agents Chemother.* **54**:2338–2344.
- Hötzel I, Theil F-P, Bernstein LJ, Prabhu S, Deng R, Quintana L, Lutman J, Sibia R, Chan P, Bumbaca D. 2012. A strategy for risk mitigation of antibodies with fast clearance. *MABs* **4**:753–760.
- Hsieh VH, Wyatt PJ. 2017. Measuring proteins with greater speed and resolution while reducing sample size. *Sci. Rep.* **7**:10030.
- Hu Y, Liu C, Muyldermans S. 2017. Nanobody-Based Delivery Systems for Diagnosis and Targeted Tumor Therapy. *Front. Immunol.* **8**:1442.
- Huber R, Deisenhofer J, Colman PM, Matsushima M, Palm W. 1976. Crystallographic structure studies of an IgG molecule and an Fc fragment. *Nature* **264**:415–420.
- Hutchinson N, Bingham N, Murrell N, Farid S, Hoare M. 2006. Shear stress analysis of mammalian cell suspensions for prediction of industrial centrifugation and its verification. *Biotechnol. Bioeng.* **95**:483–491.
- Iacobucci C, Sinz A. 2017. To Be or Not to Be? Five Guidelines to Avoid Misassignments in Cross-Linking/Mass Spectrometry. *Anal. Chem.* **89**:7832–7835.
- ICH. 2009. Pharmaceutical Development Q8. *ICH Harmon. Tripart. Guidel.* **8**:1–28.
- Imamura H, Sasaki A, Honda S. 2017. Fate of a Stressed Therapeutic Antibody Tracked by Fluorescence Correlation Spectroscopy: Folded Monomers Survive Aggregation. *J. Phys. Chem. B* **121**:8085–8093.
- Jacobs SA, Wu S-J, Feng Y, Bethea D, O’Neil KT. 2010. Cross-Interaction Chromatography: A Rapid Method to Identify Highly Soluble Monoclonal Antibody Candidates. *Pharm. Res.* **27**:65–71.
- Jahn TR, Radford SE. 2008. Folding versus aggregation: Polypeptide conformations on competing pathways. *Arch. Biochem. Biophys.* **469**:100–117.
- Jain A, Graveline A, Waterhouse A, Vernet A, Flaumenhaft R, Ingber DE. 2016. A shear gradient-activated microfluidic device for automated monitoring of whole

- blood haemostasis and platelet function. *Nat. Commun.* **7**:10176.
- Jain M, Paranandi M, Roush D, Göklen K, Kelly WJ. 2005. Using CFD to understand how flow patterns affect retention of cell-sized particles in a tubular bowl centrifuge. *Ind. Eng. Chem. Res.* **44**:7876–7884.
- Jain T, Sun T, Durand S, Hall A, Houston NR, Nett JH, Sharkey B, Bobrowicz B, Caffry I, Yu Y, Cao Y, Lynaugh H, Brown M, Baruah H, Gray LT, Krauland EM, Xu Y, Vásquez M, Wittrup KD. 2017. Biophysical properties of the clinical-stage antibody landscape. *Proc. Natl. Acad. Sci.* **114**:944–949.
- Jaspe J, Hagen SJ. 2006. Do protein molecules unfold in a simple shear flow? *Biophys. J.* **91**:3415–24.
- Jayapal KP, Wlaschin KF, Hu W, Yap MGS. 2007. Recombinant protein therapeutics from CHO cells-20 years and counting. *Chem. Eng. Prog.* **103**:40–47.
- Jing Y, Borys M, Nayak S, Egan S, Qian Y, Pan S-H, Li ZJ. 2012. Identification of cell culture conditions to control protein aggregation of IgG fusion proteins expressed in Chinese hamster ovary cells. *Process Biochem.* **47**:69–75.
- Jiskoot WIM, Randolph TW, Volkin DB, Middaugh CR, Oneich CSCH, Winter G, Friess W, Crommelin DJA, Carpenter JF. 2012. Protein Instability and Immunogenicity: Roadblocks to Clinical Application of Injectable Protein Delivery Systems for Sustained Release. *J. Pharm. Sci.* **101**:946–954.
- Jones PT, Dear PH, Foote J, Neuberger MS, Winter G. 1986. Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature* **321**:522–525.
- Joseph A, Kenty B, Mollet M, Hwang K, Rose S, Goldrick S, Bender J, Farid SS, Titchener-Hooker N. 2016. A scale-down mimic for mapping the process performance of centrifugation, depth and sterile filtration. *Biotechnol. Bioeng.* **113**:1934–1941.
- Joubert MK, Luo Q, Nashed-Samuel Y, Wypych J, Narhi LO. 2011. Classification and characterization of therapeutic antibody aggregates. *J. Biol. Chem.* **286**:25118–33.
- Junutula JR, Bhakta S, Raab H, Ervin KE, Eigenbrot C, Vandlen R, Scheller RH, Lowman HB. 2008. Rapid identification of reactive cysteine residues for site-specific labeling of antibody-Fabs. *J. Immunol. Methods* **332**:41–52.

- Kalonia C, Toprani V, Toth RT, Wahome N, Gabel I, Middaugh CR, Volkin DB. 2016. Effects of Protein Conformation, Apparent Solubility, and Protein-Protein Interactions on the Rates and Mechanisms of Aggregation for an IgG1 Monoclonal Antibody. *J. Phys. Chem. B* **120**:7062–7075.
- van der Kant R, Karow-Zwick AR, Van Durme J, Blech M, Gallardo R, Seeliger D, Aßfalg K, Baatsen P, Compennolle G, Gils A, Studts JM, Schulz P, Garidel P, Schymkowitz J, Rousseau F. 2017. Prediction and Reduction of the Aggregation of Monoclonal Antibodies. *J. Mol. Biol.* **429**:1244–1261.
- Karageorgos I, Gallagher ES, Galvin C, Gallagher DT, Hudgens JW. 2017. Biophysical characterization and structure of the Fab fragment from the NIST reference antibody, RM 8671. *Biologicals* **50**:27–34.
- Katchalski E, Benjamin GS, Gross V. 1957. The Availability of the Disulfide Bonds of Human and Bovine Serum Albumin and of Bovine γ -Globulin to Reduction by Thioglycolic Acid. *J. Am. Chem. Soc.* **79**:4096–4099.
- Kelley B. 2009. Industrialization of mAb production technology: The bioprocessing industry at a crossroads. *MAbs* **1**:443–452.
- Kelly SM, Jess TJ, Price NC. 2005. How to study proteins by circular dichroism. *Biochim. Biophys. Acta* **1751**:119–39.
- Kennedy PJ, Oliveira C, Granja PL, Sarmiento B. 2017. Antibodies and associates: Partners in targeted drug delivery. *Pharmacol. Ther.* **177**:129–145.
- Kerwin BA. 2008. Polysorbates 20 and 80 Used in the Formulation of Protein Biotherapeutics: Structure and Degradation Pathways. *J. Pharm. Sci.* **97**:2924–2935.
- Khodaparast L, Khodaparast L, Gallardo R, Louros NN, Michiels E, Ramakrishnan R, Ramakers M, Claes F, Young L, Shahrooei M, Wilkinson H, Desager M, Mengistu Tadesse W, Nilsson KPR, Hammarström P, Aertsen A, Carpentier S, Van Eldere J, Rousseau F, Schymkowitz J. 2018. Aggregating sequences that occur in many proteins constitute weak spots of bacterial proteostasis. *Nat. Commun.* **9**:866.
- Kim NA, Hada S, Thapa R, Jeong SH. 2016. Arginine as a protein stabilizer and destabilizer in liquid formulations. *Int. J. Pharm.* **513**:26–37.
- King RP. 2002. Introduction to Practical Fluid Flow. *Introd. to Pract. Fluid Flow*.

Oxford: Butterworth-Heinemann 1-54 p.

- Kiss R, Fizil Á, Szántay C. 2018. What NMR can do in the biopharmaceutical industry. *J. Pharm. Biomed. Anal.* **147**:367–377.
- Knight PD, Karamanos TK, Radford SE, Ashcroft AE. 2018. Identification of a novel site of interaction between ataxin-3 and the amyloid aggregation inhibitor polyglutamine binding peptide 1. *Eur. J. Mass Spectrom.* **24**:129–140.
- Koepf E, Eisele S, Schroeder R, Brezesinski G, Friess W. 2018. Notorious but not understood: How liquid-air interfacial stress triggers protein aggregation. *Int. J. Pharm.* **537**:202–212.
- Köhler G, Milstein C. 1975. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**:495–497.
- Kohli N, Jain N, Geddie ML, Razlog M, Xu L, Lugovskoy AA. 2015. A novel screening method to assess developability of antibody-like molecules. *MAbs* **7**:752–758.
- Kola I, Landis J. 2004. Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* **3**:711–715.
- Konermann L. 2017. Addressing a Common Misconception: Ammonium Acetate as Neutral pH “Buffer” for Native Electrospray Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **28**:1827–1835.
- Kong S, Aucamp J, Titchener-Hooker NJ. 2010. Studies on membrane sterile filtration of plasmid DNA using an automated multiwell technique. *J. Memb. Sci.* **353**:144–150.
- Kontermann RE, Brinkmann U. 2015. Bispecific antibodies. *Drug Discov. Today* **20**:838–847.
- Kopp MRG, Arosio P. 2018. Microfluidic Approaches for the Characterization of Therapeutic Proteins. *J. Pharm. Sci.* **107**:1228–1236.
- Krapp S, Mimura Y, Jefferis R, Huber R, Sondermann P. 2003. Structural Analysis of Human IgG-Fc Glycoforms Reveals a Correlation Between Glycosylation and Structural Integrity. *J. Mol. Biol.* **325**:979–989.
- Krieger CC, An X, Tang H-Y, Mohandas N, Speicher DW, Discher DE. 2011. Cysteine shotgun-mass spectrometry (CS-MS) reveals dynamic sequence of protein structure changes within mutant and stressed cells. *Proc. Natl. Acad. Sci.* **108**:8269–8274.

- Kuypers DRJ, Vanrenterghem YFC. 2004. Monoclonal antibodies in renal transplantation: old and new. *Nephrol. Dial. Transplant.* **19**:297–300.
- Lankers M, Munhall J, Valet O. 2008. Differentiation between foreign particulate matter and silicone oil induced protein aggregation in drug solutions by automated raman spectroscopy. *Microsc. Microanal.* **14**:1612–1613.
- Lanotte L, Mauer J, Mendez S, Fedosov DA, Fromental J-M, Claveria V, Nicoud F, Gompper G, Abkarian M. 2016. Red cells' dynamic morphologies govern blood shear thinning under microcirculatory flow conditions. *Proc. Natl. Acad. Sci. U. S. A.* **113**:13289–13294.
- Lauer TM, Agrawal NJ, Chennamsetty N, Egodage K, Helk B, Trout BL. 2012. Developability index: A rapid in silico tool for the screening of antibody aggregation propensity. *J. Pharm. Sci.* **101**:102–115.
- LeDuc P, Haber C, Bao G, Wirtz D. 1999. Dynamics of individual flexible polymers in a shear flow. *Nature* **399**:564–566.
- Leem J, Dunbar J, Georges G, Shi J, Deane CM. 2016. ABodyBuilder: Automated antibody structure prediction with data-driven accuracy estimation. *MAbs* **8**:1259–1268.
- Leitner A. 2016. Cross-linking and other structural proteomics techniques: How chemistry is enabling mass spectrometry applications in structural biology. *Chem. Sci.* **7**:4792–4803.
- Levinthal C. 1968. Are there pathways for protein folding? *J. Chim. Phys.* **65**:44–45.
- Levy MS, Collins IJ, Yim SS, Ward JM, Titchener-Hooker N, Ayazi Shamlou P, Dunnill P. 1999. Effect of shear on plasmid DNA in solution. *Bioprocess Eng.* **20**:7–13.
- Li F, Vijayasankaran N, Shen A, Kiss R, Amanullah A. 2010. Cell culture processes for monoclonal antibody production. *MAbs* **2**:466–479.
- Li J, Wei H, Krystek SR, Bond D, Brender TM, Cohen D, Feiner J, Hamacher N, Harshman J, Huang RY-C, Julien SH, Lin Z, Moore K, Mueller L, Noriega C, Sejwal P, Sheppard P, Stevens B, Chen G, Tymiak AA, Gross ML, Schneeweis LA. 2017. Mapping the Energetic Epitope of an Antibody/Interleukin-23 Interaction with Hydrogen/Deuterium Exchange, Fast Photochemical Oxidation of Proteins Mass Spectrometry, and Alanine Scrambling Mutagenesis. *Anal. Chem.* **89**:2250–2258.

- Li Y, Roberts CJ. 2009. Lumry-Eyring nucleated-polymerization model of protein aggregation kinetics. 2. Competing growth via condensation and chain polymerization. *J. Phys. Chem. B* **113**:7020–7032.
- Li Y, Roberts CJ. 2010. Protein Aggregation Pathways, Kinetics, and Thermodynamics. In: . *Aggreg. Ther. Proteins*. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 63–102.
- Lin C-W, Tsai M-H, Li S-T, Tsai T-I, Chu K-C, Liu Y, Lai M-Y, Wu C-YC-Y, Tseng Y-C, Shivatare SS, Wang C-H, Chao P, Wang S-Y, Shih H-W, Zeng Y-F, You T-H, Liao J-Y, Tu Y-C, Lin Y-S, Chuang H-Y, Chen C-L, Tsai C-S, Huang C-C, Lin N-H, Ma C, Wu C-YC-Y, Wong C-H. 2015. A common glycan structure on immunoglobulin G for enhancement of effector functions. *Proc. Natl. Acad. Sci. U. S. A.* **112**: 10610-10616.
- Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. 1997. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **23**, 3-25.
- Lippok S, Radtke M, Obser T, Kleemeier L, Schneppenheim R, Budde U, Netz RR, Rädler JO. 2016. Shear-Induced Unfolding and Enzymatic Cleavage of Full-Length VWF Multimers. *Biophys. J.* **110**:545–554.
- Liu L, Lu J, Allan BW, Tang Y, Tetreault J, Chow C, Barmettler B, Nelson J, Bina H, Huang L, Wroblewski VJ, Kikly K. 2016. Generation and characterization of ixekizumab, a humanized monoclonal antibody that neutralizes interleukin-17A. *J. Inflamm. Res.* **9**:39–50.
- Liu Q. 2011. Fully human antibodies against human cd20. WO2004035607A2 *.
- Liu Y, Caffry I, Wu J, Geng SB, Jain T, Sun T, Reid F, Cao Y, Estep P, Yu Y. 2014. High-throughput screening for developability during early-stage antibody discovery using self-interaction nanoparticle spectroscopy. *MAbs* **6**:483–492.
- Loizos N, Huber J. 2006. Antibodies against PDGFR α . United States of America US11917890.
- Lonberg N, Taylor LD, Harding FA, Trunstin M, Higgins KM, Schramm SR, Kuo C-C, Mashayekh R, Wymore K, McCabe JG, Munoz-O'Regan D, O'Donnell SL, Lapachet ESG, Bengoechea T, Fishwild DM, Carmack CE, Kay RM, Huszar D. 1994. Antigen-specific human antibodies from mice comprising four distinct genetic modifications. *Nature* **368**:856–859.

- Lumry R, Eyring H. 1954. Conformation changes of proteins. *J. Phys. Chem.* **58**:110–120.
- Luo Q, Joubert MK, Stevenson R, Ketchem RR, Narhi LO, Wypych J. 2011. Chemical modifications in therapeutic protein aggregates generated under different stress conditions. *J. Biol. Chem.* **286**:25134–44.
- Maa YF, Hsu CC. 1997. Protein denaturation by combined effect of shear and air-liquid interface. *Biotechnol. Bioeng.* **54**:503–12.
- Macklin PS, Morris PJ, Knight SR. 2015. A systematic review of the use of rituximab as induction therapy in renal transplantation. *Transplant. Rev.* **29**:103–108.
- Mahler H-C, Friess W, Grauschopf U, Kiese S. 2009. Protein aggregation: pathways, induction factors and analysis. *J. Pharm. Sci.* **98**:2909–34.
- Majorek KA, Porebski PJ, Dayal A, Zimmerman MD, Jablonska K, Stewart AJ, Chruszcz M, Minor W. 2012. Structural and immunologic characterization of bovine, horse, and rabbit serum albumins. *Mol. Immunol.* **52**:174–182.
- Mak TW, Saunders ME. 2006. Ch 28 Allergy and Hypersensitivity. In: . *The Immune Response*. Academic Press, pp. 873–921.
- Makarov A, Denisov E, Lange O, Horning S. 2006. Dynamic range of mass accuracy in LTQ orbitrap hybrid mass spectrometer. *J. Am. Soc. Mass Spectrom.* **17**:977–982.
- Malhotra P, Udgaonkar JB. 2016. Secondary Structural Change Can Occur Diffusely and Not Modularly during Protein Folding and Unfolding Reactions. *J. Am. Chem. Soc.* **138**:5866–5878.
- Manetta J, Bina H, Ryan P, Fox N, Witcher DR, Kikly K. 2014. Generation and characterization of tabalumab, a human monoclonal antibody that neutralizes both soluble and membrane-bound B-cell activating factor. *J. Inflamm. Res.* **7**:121–131.
- Marks JD, Hoogenboom HR, Bonnert TP, McCafferty J, Griffiths AD, Winter G. 1991. By-passing immunization: Human antibodies from V-gene libraries displayed on phage. *J. Mol. Biol.* **222**:581–597.
- Martin EM, Jackson MP, Gamerdinger M, Gense K, Karamonos TK, Humes JR, Deuerling E, Ashcroft AE, Radford SE. 2018. Conformational flexibility within the nascent polypeptide-associated complex enables its interactions with

- structurally diverse client proteins. *J. Biol. Chem.* **293**:8554–8568.
- Maurer-Stroh S, Debulpaep M, Kuemmerer N, de la Paz ML, Martins IC, Reumers J, Morris KL, Copland A, Serpell L, Serrano L, Schymkowitz JWH, Rousseau F. 2010. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat. Methods* **7**:237–242.
- Mazzer AR, Perraud X, Halley J, O'Hara J, Bracewell DG. 2015. Protein A chromatography increases monoclonal antibody aggregation rate during subsequent low pH virus inactivation hold. *J. Chromatogr. A* **1415**:83–90.
- McKian KP, Haluska P. 2009. Cixutumumab. *Expert Opin. Investig. Drugs* **18**:1025–1033.
- Meacle FJ, H. Z, I. P, Ward JM, Titchener-Hooker NJ, M. H. 2006. Degradation of Supercoiled Plasmid DNA Within a Capillary Device. *Biotechnol. Bioeng.* **97**:1148–1157.
- Meisl G, Kirkegaard JB, Arosio P, Michaels TCT, Vendruscolo M, Dobson CM, Linse S, Knowles TPJ. 2016. Molecular mechanisms of protein aggregation from global fitting of kinetic models. *Nat. Protoc.* **11**:252–272.
- Melchore JA. 2011. Sound practices for consistent human visual inspection. *AAPS PharmSciTech* **12**:215–212.
- Mendoza VL, Vachet RW. 2009. Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrom. Rev.* **28**:785–815.
- Merchant AM, Zhu Z, Yuan JQ, Goddard A, Adams CW, Presta LG, Carter P. 1998. An efficient route to human bispecific IgG. *Nat. Biotechnol.* **16**:677–681.
- Meric G, Robinson AS, Roberts CJ. 2017. Driving Forces for Nonnative Protein Aggregation and Approaches to Predict Aggregation-Prone Regions. *Annu. Rev. Chem. Biomol. Eng.* **8**:139–159.
- Militello V, Vetri V, Leone M. 2003. Conformational changes involved in thermal aggregation processes of bovine serum albumin. *Biophys. Chem.* **105**:133–141.
- Minton AP. 2016. Recent applications of light scattering measurement in the biological and biopharmaceutical sciences. *Anal. Biochem.* **501**:4–22.
- Mitragotri S, Burke PA, Langer R. 2014. Overcoming the challenges in administering biopharmaceuticals: Formulation and delivery strategies. *Nat. Rev. Drug Discov.* **13**:655–672.

- Morán Luengo T, Kityk R, Mayer MP, Rüdiger SGD. 2018. Hsp90 Breaks the Deadlock of the Hsp70 Chaperone System. *Mol. Cell* **70**:545–552.e9.
- Morrison SL, Johnson MJ, Herzenberg LA, Oi VT. 1984. Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. *Proc. Natl. Acad. Sci. U. S. A.* **81**:6851–6855.
- Mouquet H, Scheid JF, Zoller MJ, Krogsgaard M, Ott RG, Shukair S, Artyomov MN, Pietzsch J, Connors M, Pereyra F. 2010. Polyreactivity increases the apparent affinity of anti-HIV antibodies by heteroligation. *Nature* **467**:591–595.
- Moussa EM, Panchal JP, Moorthy BS, Blum JS, Joubert MK, Narhi LO, Topp EM. 2016. Immunogenicity of Therapeutic Protein Aggregates. *J. Pharm. Sci.* **105**:417–430.
- Mullard A. 2018. 2017 FDA drug approvals. *Nat. Rev. Drug Discov.* **17**:81–85.
- Nanda T, He J, Haas M, Shpungin S, Rusanov I, Sweder R, Brisbane C, Nesta D. 2017. Aggregation from Shear Stress and Surface Interaction Molecule-Specific or Universal Phenomenon. *Bioprocess Int.* **15**:30–39.
- Nazari ZE, van de Weert M, Bou-Assaf G, Houde D, Weiskopf A, D. Rand K. 2016. Rapid Conformational Analysis of Protein Drugs in Formulation by Hydrogen/Deuterium Exchange Mass Spectrometry. *J. Pharm. Sci.* **105**:3269–3277.
- Nicoud L, Arosio P, Sozo M, Yates A, Norrant E, Morbidelli M. 2014. Kinetic analysis of the multistep aggregation mechanism of monoclonal antibodies. *J. Phys. Chem. B* **118**:10595–10606.
- Nicoud L, Jagielski J, Pfister D, Lazzari S, Massant J, Lattuada M, Morbidelli M. 2016. Kinetics of Monoclonal Antibody Aggregation from Dilute toward Concentrated Conditions. *J. Phys. Chem. B* **120**:3267–3280.
- Nienow AW. 2006. Reactor engineering in large scale animal cell culture. *Cytotechnology* **50**:9–33.
- NobelPrize.org, AB NM. 2018. Pierre-Gilles de Gennes - Facts. 2018. https://www.nobelprize.org/nobel_prizes/physics/laureates/1991/gennes-facts.html.
- Nowak C, K. Cheung J, M. Dellatore S, Katiyar A, Bhat R, Sun J, Ponniah G, Neill A, Mason B, Beck A, Liu H. 2017. Forced degradation of recombinant monoclonal

- antibodies: A practical guide. *MAbs* **9**:1217–1230.
- Odell JA, Muller AJ, Narh KA, Keller A. 1990. Degradation of polymer solutions in extensional flows. *Macromolecules* **23**:3092–3103.
- Oganesyan V, Peng L, Woods RM, Wu H, Dall'Acqua WF. 2015. Structural Insights into the Neutralization Properties of the Fully Human, Anti-interferon Monoclonal Antibody Sifalimumab. *J. Biol. Chem.* **290**:14979–14985.
- Oliva A, Farin B. 2001. Comparative study of protein molecular weights by size-exclusion chromatography and laser-light scattering **25**:833–841.
- Opuni KFM, Al-Majdoub M, Yefremova Y, El-Kased RF, Koy C, Glocker MO. 2018. Mass spectrometric epitope mapping. *Mass Spectrom. Rev.* **37**:229–241.
- Pabbathi A, Patra S, Samanta A. 2013. Structural transformation of bovine serum albumin induced by dimethyl sulfoxide and probed by fluorescence correlation spectroscopy and additional methods. *ChemPhysChem* **14**:2441–2449.
- Pacholarz KJ, Porrini M, Garlish RA, Burnley RJ, Taylor RJ, Henry AJ, Barran PE. 2014. Dynamics of intact immunoglobulin G explored by drift-tube ion-mobility mass spectrometry and molecular modeling. *Angew. Chemie - Int. Ed.* **53**:7765–7769.
- Palmqvist N, Foster T, Tarkowski A, Josefsson E. 2002. Protein A is a virulence factor in Staphylococcus aureus arthritis and septic death. *Microb. Pathog.* **33**:239–249.
- Panagi I. 2018. Transfer Report; University of Leeds.
- Papadopoulos N, Martin J, Ruan Q, Rafique A, Rosconi MP, Shi E, Pyles EA, Yancopoulos GD, Stahl N, Wiegand SJ. 2012. Binding and neutralization of vascular endothelial growth factor (VEGF) and related ligands by VEGF Trap, ranibizumab and bevacizumab. *Angiogenesis* **15**:171–85.
- Pavia DL, Kriz GS, Lampman GM, Vyvyan JR. 2010. Spectroscopy 4th ed. Brooks/Cole 641 p.
- Peng H-P, Lee KH, Jian J-W, Yang A-S. 2014. Origins of specificity and affinity in antibody–protein interactions. *Proc. Natl. Acad. Sci.* **111**:E2656–E2665.
- Perkins T, Smith D, Chu S. 1997. Single polymer dynamics in an elongational flow. *Science.* **276**:2016–2021.

- Peters T. 1996. All about albumin : biochemistry, genetics, and medical applications. Academic Press 432 p.
- Peyvandi F, Scully M, Kremer Hovinga JA, Cataland S, Knöbl P, Wu H, Artoni A, Westwood J-P, Mansouri Taleghani M, Jilma B, Callewaert F, Ulrichs H, Duby C, Tersago D. 2016. Caplacizumab for Acquired Thrombotic Thrombocytopenic Purpura. *N. Engl. J. Med.* **374**:511–522.
- Pharmacopeia U. General Chapter Particulate Matter in Injections | USP-NF. *US Pharmacop.* http://www.pharmacopeia.cn/v29240/usp29nf24s0_c788.html.
- Phillips JJ, Buchanan A, Andrews J, Chodorge M, Sridharan S, Mitchell L, Burmeister N, Kippen AD, Vaughan TJ, Higazi DR, Lowe D. 2017. Rate of Asparagine Deamidation in a Monoclonal Antibody Correlating with Hydrogen Exchange Rate at Adjacent Downstream Residues. *Anal. Chem.* **89**:2361–2368.
- Philo JS. 2006. Is any measurement method optimal for all aggregate sizes and types? *AAPS J.* **8**:E564-71.
- Piantadosi S. 2017. Clinical trials : a methodologic perspective Third edit. John Wiley & Sons 917 p.
- Plenge RM. 2016. Disciplined approach to drug discovery and early development. *Sci. Transl. Med.* **8**:349ps15.
- Poppe L, Jordan JB, Rogers G, Schnier PD. 2015. On the analytical superiority of 1D NMR for fingerprinting the higher order structure of protein therapeutics compared to multidimensional NMR methods. *Anal. Chem.* **87**:5539–5545.
- Postnova. 2018. General Theory - Postnova Analytics GmbH. <http://www.postnova.com/general-theory.html>.
- Powers ET, Balch WE. 2013. Diversity in the origins of proteostasis networks--a driver for protein function in evolution. *Nat. Rev. Mol. Cell Biol.* **14**:237–48.
- Price NC, Nairn J. 2009. Exploring proteins : a student's guide to experimental skills and methods. Oxford University Press 512 p.
- Proudfoot AG, Davidson SJ, Strueber M. 2017. von Willebrand factor disruption and continuous-flow circulatory devices. *J. Hear. Lung Transplant.* **36**:1155–1163.
- Provencher SW. 1982. CONTIN: A general purpose constrained regularisation program for inverting noisy linear algebraic and integral equations. *Comput. Phys. Commun.* **27**:229–242.

- Queen CL, Schneider WP, Selick HE. 1995. Polynucleotides encoding improved humanized immunoglobulins. US5693761A.
- Radford SE, Dobson CM. 1995. Insights into protein folding using physical techniques: studies of lysozyme and alpha-lactalbumin. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **348**:17–25.
- Radford SE, Dobson CM, Evans PA. 1992. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* **358**:302–307.
- Ramagopal UA, Liu W, Garrett-Thomson SC, Bonanno JB, Yan Q, Srinivasan M, Wong SC, Bell A, Mankikar S, Rangan VS, Deshpande S, Korman AJ, Almo SC. 2017. Structural basis for cancer immunotherapy by the first-in-class checkpoint inhibitor ipilimumab. *Proc. Natl. Acad. Sci. U. S. A.* **114**:E4223–E4232.
- Rammensee S, Slotta U, Scheibel T, Bausch AR. 2008. Assembly mechanism of recombinant spider silk proteins. *Proc. Natl. Acad. Sci. U. S. A.* **105**:6590–6595.
- Randolph TW, Carpenter JF. 2007. Engineering challenges of protein formulations. *AIChE J.* **58**: 1902-1907.
- Ratanji KD, Derrick JP, Dearman RJ, Kimber I. 2014. Immunogenicity of therapeutic proteins: influence of aggregation. *J. Immunotoxicol.* **11**:99–109.
- Rayat AC, Chatel A, Hoare M, Lye GJ. 2016. Ultra scale-down approaches to enhance the creation of bioprocesses at scale: impacts of process shear stress and early recovery stages. *Curr. Opin. Chem. Eng.* **14**:150–157.
- van Reis R, Zydney A. 2007. Bioprocess membrane technology. *J. Memb. Sci.* **297**:16–50.
- Ren Z, Harshe YM, Lattuada M. 2015. Influence of the Potential Well on the Breakage Rate of Colloidal Aggregates in Simple Shear and Uniaxial Extensional Flows. *Langmuir* **31**:5712–5721.
- Renner CB, Doyle PS. 2015. Stretching self-entangled DNA molecules in elongational fields. *Soft Matter* **11**:3105–3114.
- Rinas A, Espino JA, Jones LM. 2016. An efficient quantitation strategy for hydroxyl radical-mediated protein footprinting using Proteome Discoverer. *Anal. Bioanal. Chem.* **408**:3021–3031.
- Ripple DC, Dimitrova MN. 2012. Protein particles: what we know and what we do not know. *J. Pharm. Sci.* **101**:3568–79.

- Roberts CJ. 2014. Protein aggregation and its impact on product quality. *Curr. Opin. Biotechnol.* **30C**:211–217.
- Roder H, Colón W. 1997. Kinetic role of early intermediates in protein folding. *Curr. Opin. Struct. Biol.* **7**:15–28.
- Rodgers KR, Chou RC. 2016. Therapeutic monoclonal antibodies and derivatives: Historical perspectives and future directions. *Biotechnol. Adv.* **34**:1149–1158.
- Roger V, Cottet H, Cipelletti L. 2016. A New Robust Estimator of Polydispersity from Dynamic Light Scattering Data. *Anal. Chem.* **88**:2630–2636.
- Rogge P, Müller D, Schmidt SR. 2015. The Single-Use or Stainless Steel Decision Process. *Bioprocess Int.* **13**:10–15.
- Rosati S, Yang Y, Barendregt A, Heck AJR. 2014. Detailed mass analysis of structural heterogeneity in monoclonal antibodies using native mass spectrometry. *Nat. Protoc.* **9**:967–76.
- Rosenberg AS. 2006. Effects of protein aggregates: an immunologic perspective. *AAPS J.* **8**:E501-7.
- Rousseau F, Serrano L, Schymkowitz JWH. 2006. How Evolutionary Pressure Against Protein Aggregation Shaped Chaperone Specificity. *J. Mol. Biol.* **355**:1037–1047.
- Ruotolo BT, Benesch JLP, Sandercock AM, Hyung S-J, Robinson C V. 2008. Ion mobility-mass spectrometry analysis of large protein complexes. *Nat. Protoc.* **3**:1139–1152.
- Russell WC, Graham FL, Smiley J, Nairn R. 1977. Characteristics of a Human Cell Line Transformed by DNA from Human Adenovirus Type 5. *J. Gen. Virol.* **36**:59–72.
- Sadegh-Vaziri R, Ludwig K, Sundmacher K, Babler MU. 2018. Mechanisms behind overshoots in mean cluster size profiles in aggregation-breakup processes. *J. Colloid Interface Sci.* **528**:336–348.
- Sanlioglu AD, Altunbas HA, Balci MK, Griffith TS, Sanlioglu S. 2013. Clinical utility of insulin and insulin analogs. *Islets* **5**:67–78.
- Saphire EO, Parren PW, Pantophlet R, Zwick MB, Morris GM, Rudd PM, Dwek RA, Stanfield RL, Burton DR, Wilson IA. 2001. Crystal structure of a neutralizing human IGG against HIV-1: a template for vaccine design. *Science.* **293**:1155–

1159.

- Saro D, Baker A, Hepler R, Spencer S, Bruce R, LaBrenz S, Chiu M, Davis D, Lang SE. 2015. Developability Assessment of a Proposed NIST Monoclonal Antibody. In: . *State-of-the-Art Emerg. Technol. Ther. Monoclon. Antib. Charact. Vol. 2. Biopharm. Charact. NISTmAb Case Study*, pp. 329–355.
- Schanzer JM, Wartha K, Moessner E, Hosse RJ, Moser S, Croasdale R, Trochanowska H, Shao C, Wang P, Shi L, Weinzierl T, Rieder N, Bacac M, Ries CH, Kettenberger H, Schlothauer T, Friess T, Umana P, Klein C. 2016. XGFR*, a novel affinity-matured bispecific antibody targeting IGF-1R and EGFR with combined signaling inhibition and enhanced immune activation for the treatment of pancreatic cancer. *MAbs* **8**:811–827.
- Schiel JE, Davis DL, Borisov O V. 2015. State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization Volume 2. Biopharmaceutical Characterization: The NISTmAb Case Study. Ed. John E Schiel, Oleg V Borisov, Darryl L Davis. Washington, DC: American Chemical Society., 427p.
- Schmitt A, Odriozola G, Moncho-Jordá A, Callejas-Fernández J, Martínez-García R, Hidalgo-Álvarez R. 2000. Multiple contact kernel for diffusionlike aggregation. *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.* **62**:8335–8343.
- Van Schouwenburg PA, Kruithof S, Votsmeier C, Van Schie K, Hart MH, De Jong RN, Van Buren EEL, Van Ham M, Aarden L, Wolbink G, Wouters D, Rispens T. 2014. Functional analysis of the anti-adalimumab response using patient-derived monoclonal antibodies. *J. Biol. Chem.* **289**:34482–34488.
- Schramm W. 2014. The history of haemophilia - A short review. *Thromb. Res.* **134**:S4–S9.
- Schroeder HW, Cavacini L, Schroeder Jr. HW, Cavacini L, Schroeder HW, Cavacini L. 2010. Structure and function of immunoglobulins. *J. Allergy Clin. Immunol.* **125**:S41-52.
- Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. 2005. The FoldX web server: an online force field. *Nucleic Acids Res.* **33**:W382–W388.
- Scott DJ, Grossmann JG, Tame JRH, Byron O, Wilson KS, Otto BR. 2002. Low resolution solution structure of the Apo form of Escherichia coli haemoglobin

- protease Hbp. *J. Mol. Biol.* **315**:1179–1187.
- Sediq AS, Van Duijvenvoorde RB, Jiskoot W, Nejadnik MR. 2016. No Touching! Abrasion of Adsorbed Protein Is the Root Cause of Subvisible Particle Formation during Stirring. *J. Pharm. Sci.* **105**:519–529.
- Sharma C, Malhotra D, Rathore AS. 2011. Review of Computational fluid dynamics applications in biotechnology processes. *Biotechnol. Prog.* **27**:1497–1510.
- Sharma VK, Kalonia DS. 2010. Experimental Detection and Characterization of Protein Aggregates. In: . *Aggreg. Ther. Proteins*. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 205–256.
- Sharma VK, Patapoff TW, Kabakoff B, Pai S, Hilario E, Zhang B, Li C, Borisov O, Kelley RF, Chorny I, Zhou JZ, Dill KA, Swartz TE. 2014. In silico selection of therapeutic antibodies for development: Viscosity, clearance, and chemical stability. *Proc. Natl. Acad. Sci. U. S. A.* **111**:18601–18616.
- Shealy D, Cai A, Staquet K, Baker A, Lacy ER, Johns L, Vafa O, Gunn G, Tam S, Sague S, Wang D, Brigham-Burke M, Dalmonte P, Emmell E, Pikounis B, Bugelski PJ, Zhou H, Scallon B, Giles-Komar J. 2010. Characterization of golimumab, a human monoclonal antibody specific for human tumor necrosis factor α . *MAbs* **2**:428–439.
- Shukla AA, Thömmes J. 2010. Recent advances in large-scale production of monoclonal antibodies and related proteins. *Trends Biotechnol.* **28**:253–261.
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**:539.
- Silicon Kinetics. 2016. Therapeutic Antibody Drug Monitoring Assay Development Utilizing Nanopore Optical Interferometry 4 p.
- Silva A, Almeida B, Fraga JS, Taboada P, Martins PM, Macedo-Ribeiro S. 2017. Distribution of Amyloid-Like and Oligomeric Species from Protein Aggregation Kinetics. *Angew. Chemie - Int. Ed.* **56**:14042–14045.
- Simon S, Krause HJ, Weber C, Peukert W. 2011. Physical degradation of proteins in well-defined fluid flows studied within a four-roll apparatus. *Biotechnol. Bioeng.* **108**:2914–22.

- Sing CE, Alexander-Katz A. 2010. Elongational flow induces the unfolding of von willebrand factor at physiological flow rates. *Biophys. J.* **98**:L35–L37.
- Singla A, Bansal R, Joshi V, Rathore AS. 2016. Aggregation Kinetics for IgG1-Based Monoclonal Antibody Therapeutics. *AAPS J.* **18**:689–702.
- Sinha R, Le Gac S, Verdonschot N, Van Den Berg A, Koopman B, Rouwkema J. 2015. A medium throughput device to study the effects of combinations of surface strains and fluid-flow shear stresses on cells. *Lab Chip* **15**:429–439.
- Smith DE, Babcock HP, Chu S. 1999. Single-Polymer Dynamics in Steady Shear Flow. *Science.* **283**:1724–1727.
- Smith DP, Radford SE, Ashcroft AE. 2010. Elongated oligomers in beta2-microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **107**:6794–8.
- Smith GP. 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science.* **228**:1315–1317.
- Smith V, Ogg S, Vlasselaer P Van, Barry VE, Marshall D, Holzer AK, Rodriguez H, Oyasu M, McCauley SA, Garcia CA, Biermann DHT. 2013. LOX and LOXL2 inhibitors and uses thereof - Google Patents. US8461303B2.
- Sneader W. 2000. The discovery of aspirin: a reappraisal. *BMJ Br. Med. J.* **321**:1591–1594.
- Snelling M. 2008. Home intravenous therapy- reducing the risk of infection. *Pharm. J.* **15**:16–18.
- Sormanni P, Amery L, Ekizoglou S, Vendruscolo M, Popovic B. 2017. Rapid and accurate in silico solubility screening of a monoclonal antibody library. *Sci. Rep.* **7**:8200.
- Sormanni P, Aprile FA, Vendruscolo M. 2015. The CamSol Method of Rational Design of Protein Mutants with Enhanced Solubility. *J. Mol. Biol.* **427**:478–490.
- Spada S, Honegger A, Plückthun A. 1998. Reproducing the natural evolution of protein structural features with the selectively infective phage (SIP) technology. the kink in the first strand of antibody kappa domains. *J. Mol. Biol.* **283**:395–407.
- Spadiut O, Capone S, Krainer F, Glieder A, Herwig C. 2014. Microbials for the production of monoclonal antibodies and antibody fragments. *Trends Biotechnol.* **32**:54–60.

- Springer T. 2016. von Willebrand factor, Jedi knight of the bloodstream. *Blood* **113**:1412–1425.
- Strohl WR. 2018. Current progress in innovative engineered antibodies. *Protein Cell* **9**:86–120.
- Sun T, Reid F, Liu Y, Cao Y, Estep P, Nauman C, Xu Y. 2013. High throughput detection of antibody self-interaction by bio-layer interferometry. *MAbs* **5**:838–841.
- Szlezák N, Evers M, Wang J, Pérez L. 2014. The role of big data and advanced analytics in drug discovery, development, and commercialization. *Clin. Pharmacol. Ther.* **95**:492–495.
- Szymczak P, Cieplak M. 2007. Proteins in a shear flow. *J. Chem. Phys.* **127**.
- Szymczak P, Cieplak M. 2011. Hydrodynamic effects in proteins. *J. Phys. Condens. Matter* **23**:33102.
- Tamizi E, Jouyban A. 2016. Forced degradation studies of biopharmaceuticals: Selection of stress conditions. *Eur. J. Pharm. Biopharm.* **98**:26–46.
- Tavakoli-Keshe R, Phillips JJ, Turner R, Bracewell DG. 2014. Understanding the relationship between biotherapeutic protein stability and solid-liquid interfacial shear in constant region mutants of IgG1 and IgG4. *J. Pharm. Sci.* **103**:437–444.
- Teska BM, Brake JM, Tronto GS, Carpenter JF. 2016. Aggregation and Particle Formation of Therapeutic Proteins in Contact With a Novel Fluoropolymer Surface Versus Siliconized Surfaces: Effects of Agitation in Vials and in Prefilled Syringes. *J. Pharm. Sci.* **105**:2053–2065.
- Thom G, Minter R. 2012. Optimization of CAT-354, a therapeutic antibody directed against interleukin-13, using ribosome display. *Methods Mol. Biol.* **805**:393–401.
- Thomas CR, Geer D. 2011. Effects of shear on proteins in solution. *Biotechnol. Lett.* **33**:443–456.
- Thomas CR, Dunnill P. 1979. Action of shear on enzymes: Studies with catalase and urease. *Biotechnol. Bioeng.* **21**:2279–2302.
- Thompson AJ, Herling TW, Kubánková M, Vyšniauskas A, Knowles TPJ, Kuimova MK. 2015. Molecular Rotors Provide Insights into Microscopic Structural Changes during Protein Aggregation. *J. Phys. Chem. B* **119**:10170–10179.

- Tian Y, Ruotolo BT. 2018. The growing role of structural mass spectrometry in the discovery and development of therapeutic antibodies. *Analyst* **143**:2459–2468.
- Tiller KE, Tessier PM. 2015. Advances in Antibody Design. *Annu. Rev. Biomed. Eng* **17**:191–216.
- Tipping KW, Karamanos TK, Jakhria T, Iadanza MG, Goodchild SC, Tuma R, Ranson NA, Hewitt EW, Radford SE. 2015. pH-induced molecular shedding drives the formation of amyloid fibril-derived oligomers. *Proc. Natl. Acad. Sci.* **112**:5691–5696.
- Tirrell M, Middleman S. 1975. Shear modification of enzyme kinetics. *Biotechnol. Bioeng.* **17**:299–303.
- Törnquist M, Michaels TCT, Sanagavarapu K, Yang X, Meisl G, Cohen SIA, Knowles TPJ, Linse S. 2018. Secondary nucleation in amyloid formation. *Chem. Commun.* **54**:8667–8684.
- Tracey D, Klareskog L, Sasso EH, Salfeld JG, Tak PP. 2008. Tumor necrosis factor antagonist mechanisms of action: a comprehensive review. *Pharmacol. Ther.* **117**:244–79.
- Trovato A, Seno F, Tosatto SCE. 2007. The PASTA server for protein aggregation prediction. *Protein Eng. Des. Sel.* **20**:521–3.
- Uetrecht C, Rose RJ, van Duijn E, Lorenzen K, Heck AJR. 2010. Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev.* **39**:1633–1655.
- Urquhart L. 2018. Top drugs and companies by sales in 2017. *Nat. Rev. Drug Discov.* **17**:232.
- Vallée MRJ, Schombs MW, Balaban ZJ, Colyer J, Davis BG. 2016. Ready display of antigenic peptides in a protein “mimogen.” *Chem. Commun.* **52**:3014–3017.
- Varley J, Birch J. 1999. Reactor design for large scale suspension animal cell culture. *Cytotechnology* **29**:177–205.
- Vasudev R, Mathew S, Afonina N. 2015. Characterization of Submicron (0.1–1 μm) Particles in Therapeutic Proteins by Nanoparticle Tracking Analysis. *J. Pharm. Sci.* **104**:1622–1631.
- Vidarsson G, Dekkers G, Rispen T. 2014. IgG subclasses and allotypes: From structure to effector functions. *Front. Immunol.* **5**:1–17.

- Vink T, Oudshoorn-Dickmann M, Roza M, Reitsma J-J, de Jong RN. 2014. A simple, robust and highly efficient transient expression system for producing antibodies. *Methods* **65**:5–10.
- Voet D, Voet JG. 2011. *Biochemistry* 4th ed. New York: Wiley 1520 p.
- Vollrath I, Friess W, Freitag A, Hawe A, Winter G. 2018. Does controlled nucleation impact the properties and stability of lyophilized monoclonal antibody formulations? *Eur. J. Pharm. Biopharm.* **129**:134–144.
- Wales TE, Fadgen KE, Gerhardt GC, Engen JR. 2008. High-Speed and High-Resolution UPLC Separation at Zero Degrees Celsius. *Anal. Chem.* **80**:6815–6820.
- Walsh G. 2003. *Biopharmaceuticals: Biochemistry and Biotechnology* 2nd Editio. Chichester: Wiley 1-551 p.
- Walsh I, Seno F, Tosatto SCE, Trovato A. 2014. PASTA 2.0: An improved server for protein aggregation prediction. *Nucleic Acids Res.* **42**:301–307.
- Wang C, Mann F. 2009. Increasing MAb Capture Productivity. *Bioprocess Int.* **7**:56–61.
- Wang W, Wang EQ, Balthasar JP. 2008. Monoclonal antibody pharmacokinetics and pharmacodynamics. *Clin. Pharmacol. Ther.* **84**:548–58.
- Wang W. 1999. Instability, stabilization, and formulation of liquid protein pharmaceuticals. *Int. J. Pharm.* **185**:129–188.
- Wang W. 2015. Advanced protein formulations. *Protein Sci.* **24**:1031–1039.
- Wang W, Nema S, Teagarden D. 2010. Protein aggregation--pathways and influencing factors. *Int. J. Pharm.* **390**:89–99.
- Wang W, Roberts CJ. 2010. *Aggregation of Therapeutic Proteins* Aggregation of Therapeutic Proteins. Wiley., 486 p.
- Wang W, Singh SK, Li N, Toler MR, King KR, Nema S. 2012. Immunogenicity of protein aggregates--concerns and realities. *Int. J. Pharm.* **431**:1–11.
- Watson C, Sharp JS. 2012. Conformational analysis of therapeutic proteins by hydroxyl radical protein footprinting. *AAPS J.* **14**:206–17.
- Wei H, Mo J, Tao L, Russell RJ, Tymiak A a., Chen G, Iacob RE, Engen JR. 2014. Hydrogen/deuterium exchange mass spectrometry for probing higher order

- structure of protein therapeutics: Methodology and applications. *Drug Discov. Today* **19**:95–102.
- Weisser NE, Hall JC. 2009. Applications of single-chain variable fragment antibodies in therapeutics and diagnostics. *Biotechnol. Adv.* **27**:502–520.
- Wild Jr. KD, Treanor JJS, Huang H, Zhang TJ, Martin F. 2004. Human anti-ngf neutralizing antibodies as selective ngf pathway inhibitors. CT/US2004/022876.
- Williamson M. 2012. *How Proteins Work* 1st ed. New York: Garland Science, Taylor and Francis 1-464 p.
- Winter G, Griffiths AD, Hawkins RE, Hoogenboom HR. 1994. Making antibodies by phage display technology. *Annu. Rev. Immunol.* **12**:433–455.
- Wishart DS. 2013. Characterization of biopharmaceuticals by NMR spectroscopy. *TrAC - Trends Anal. Chem.* **48**:96–111.
- Wolfrum S, Weichsel U, Siedler M, Weber C, Peukert W. 2017. Monitoring of Flow-Induced Aggregation and Conformational Change of Monoclonal Antibodies. *Chemie-Ingenieur-Technik* **89**:987–994.
- Wyatt PJ. 1993. Light Scattering and the Solution Properties of Macromolecules. *Anal. Chim. Acta* **272**:1–40.
- Xu G, Chance MR. 2007. Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chem. Rev.* **107**:3514–3543.
- Xu Y, Roach W, Sun T, Jain T, Prinz B, Yu T-Y, Torrey J, Thomas J, Bobrowicz P, Vásquez M. 2013. Addressing polyspecificity of antibodies selected from an in vitro yeast presentation system: a FACS-based, high-throughput selection and analytical tool. *Protein Eng. Des. Sel.* **26**:663–670.
- Yadav S, Shire SJ, Kalonia DS. 2011. Viscosity analysis of high concentration bovine serum albumin aqueous solutions. *Pharm. Res.* **28**:1973–1983.
- Yan X, Shi Q, Bracher A, Miličić G, Singh AK, Hartl FU, Hayer-Hartl M. 2018. GroEL Ring Separation and Exchange in the Chaperonin Reaction. *Cell* **172**:605–617.
- Yang H, Yang S, Kong J, Dong A, Yu S. 2015. Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy. *Nat. Protoc.* **10**:382–396.
- Yang T-C, Langford AJ, Kumar S, Ruesch JC, Wang W. 2016. Trimerization Dictates

- Solution Opalescence of a Monoclonal Antibody. *J. Pharm. Sci.* **105**:2328–2337.
- Yang W-P, Green K, Pinz-Sweeney S, Briones AT, Burton DR, Barbas III CF. 1995. CDR Walking Mutagenesis for the Affinity Maturation of a Potent Human Anti-HIV-1 Antibody into the Picomolar Range. *J. Mol. Biol.* **254**:392–403.
- Yang X, Xu W, Dukleska S, Benchaar S, Mengisen S, Antochshuk V, Cheung J, Mann L, Babadjanova Z, Rowand J, Gunawan R, McCampbell A, Beaumont M, Meininger D, Richardson D, Ambrogelly A. 2013. Developability studies before initiation of process development: Improving manufacturability of monoclonal antibodies. *MAbs* **5**:787–794.
- Young JC, Agashe VR, Siegers K, Hartl FU. 2004. Pathways of chaperone-mediated protein folding in the cytosol. *Nat. Rev. Mol. Cell Biol.* **5**:781–791.
- Young LM, Saunders JC, Mahood RA, Reville CH, Foster RJ, Tu L-H, Raleigh DP, Radford SE, Ashcroft AE. 2015. Screening and classifying small-molecule inhibitors of amyloid formation using ion mobility spectrometry-mass spectrometry. *Nat. Chem.* **7**:73–81.
- Zaccai NR, Serdyuk IN, Zaccai J. 2017. *Methods in Molecular Biophysics* 2nd ed. Cambridge: Cambridge University Press 1-684 p.
- Zahnd C, Amstutz P, Plückthun A. 2007. Ribosome display: Selecting and evolving proteins in vitro that specifically bind to a target. *Nat. Methods* **4**:269–279.
- Zambrano R, Jamroz M, Szczasiuk A, Pujols J, Kmiecik S, Ventura S. 2015. AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures. *Nucleic Acids Res.* **43**:W306–W313.
- Zbacnik TJ, Holcomb RE, Katayama DS, Murphy BM, Payne RW, Coccaro RC, Evans GJ, Matsuura JE, Henry CS, Manning MC. 2017. Role of Buffers in Protein Formulations. *J. Pharm. Sci.* **106**:713–733.
- Zhang H, Kong S, Booth A, Boushaba R, Levy M, Hoare M. 2007. Prediction of shear damage of plasmid DNA in pump and centrifuge operations using an ultra scale down device. *Biotechnol. Prog.* **23**:858–865.
- Zhang H, Cui W, Gross ML. 2014. Mass spectrometry for the biophysical characterization of therapeutic monoclonal antibodies. *FEBS Lett.* **588**:308–17.
- Zhang X, Halvorsen K, Zhang C-Z, Wong WP, Springer TA. 2009. Mechanoenzymatic cleavage of the ultralarge vascular protein von Willebrand factor. *Science*

324:1330–4.

- Zhang X, Poniewierski A, Hou S, Sozański K, Wisniewska A, Wieczorek SA., Kalwarczyk T, Sun L, Hołyst R. 2015a. Tracking structural transitions of bovine serum albumin in surfactant solutions by fluorescence correlation spectroscopy and fluorescence lifetime analysis. *Soft Matter* **11**:2512–2518.
- Zhang Y, Rempel DL, Zhang H, Gross ML. 2015b. An Improved Fast Photochemical Oxidation of Proteins (FPOP) Platform for Protein Therapeutics. *J. Am. Soc. Mass Spectrom.* **26**:526–529.
- Zhang Y. 2015. Understanding the impact of Fc glycosylation on its conformational changes by molecular dynamics simulations and bioinformatics. *Mol. BioSyst.* **11**:3415–3424.
- Zhao HUI, Graf O, Milovic N, Luan X, Bluemel M, Smolny M, Forrer K. 2010. Formulation Development of Antibodies Using Robotic System and High-Throughput Laboratory (HTL). *J. Pharm. Sci.* **99**:2279–2294.
- Zheng K, Bantog C, Bayer R. 2011. The impact of glycosylation on monoclonal antibody conformation and stability. *MAbs* **3**:568–76.
- Zheng S, Puri A, Li J, Jaiswal A, Adams M. 2017. Particle Characterization for a Protein Drug Product Stored in Pre-Filled Syringes Using Micro-Flow Imaging, Archimedes, and Quartz Crystal Microbalance with Dissipation. *AAPS J.* **19**:110–116.
- Zhou C, Krueger AB, Barnard JG, Qi W, Carpenter JF. 2015. Characterization of Nanoparticle Tracking Analysis for Quantification and Sizing of Submicron Particles of Therapeutic Proteins. *J. Pharm. Sci.* **104**:2441–2450.
- Zhou C, Qi W, Lewis EN, Carpenter JF. 2016. Characterization of Sizes of Aggregates of Insulin Analogs and the Conformations of the Constituent Protein Molecules: A Concomitant Dynamic Light Scattering and Raman Spectroscopy Study. *J. Pharm. Sci.* **105**:551–558.
- Zidar M, Kuzman D, Ravnik M. 2018. Characterisation of protein aggregation with the Smoluchowski coagulation approach for use in biopharmaceuticals. *Soft Matter* **14**:6001–6012.
- Zölls S, Tantipolphan R, Wiggernhorn M, Winter G, Jiskoot W, Friess W, Hawe A. 2012. Particles in therapeutic protein formulations, Part 1: Overview of analytical

methods. *J. Pharm. Sci.* **101**:914–935.

Zöls S, Weinbuch D, Wiggernhorn M, Winter G, Friess W, Jiskoot W, Hawe A. 2013. Flow imaging microscopy for protein particle analysis--a comparative evaluation of four different analytical instruments. *AAPS J.* **15**:1200–11.

Zurdo J, Arnell A, Obrezanova O, Smith N, Gómez De La Cuesta R, Gallagher TRA, Michael R, Stallwood Y, Ekblad C, Abrahmsén L, Höiden-Guthenberg I. 2015. Early implementation of QbD in biopharmaceutical development: A practical example. *Biomed Res. Int.* **2015**:605427.