

Use of in silico and biophysical techniques to predict protein stability

A thesis submitted to University College London
For the degree of
DOCTOR OF PHILOSOPHY

By

Victoria Elizabeth Wood

Department of Biochemical Engineering
University College London

2019

I, Victoria Elizabeth Wood confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed,

Victoria Elizabeth Wood

Abstract

A protein's structure dictates its function; which in turn governs its role of maintaining the cascade of activities that characterise all living systems. In solution, proteins are labile structures that move and flex naturally or in response to external stimuli, which is also important for function. However, this movement of proteins, collectively termed protein dynamics, can also have a negative impact on stability. During manufacturing, biotherapeutic proteins are exposed to a variety of stressors that can cause unfolding and lead to aggregation, which can cause loss of activity and increased immunogenicity. Therefore, implementing strategies to increase resistance to stress is paramount to ensuring protein efficacy and safety.

Hydrogen/deuterium exchange (HDX) has become a valuable tool for the study of protein dynamics. The method involves labile hydrogens in the protein backbone exchanging with deuterium atoms when the protein is placed in a deuterium oxide solution. The subsequent increase in protein mass over time can be measured with high-resolution mass spectrometry (MS). The rate of deuterium exchange is a function of solvent accessibility, hydrogen bonding and protein flexibility. Dynamics can be studied at a local level via the inclusion of a peptide digestion step after the labeling reaction to monitor peptic fragment exchange. This thesis explored the use of peptide level HDX-MS to monitor changes in local protein dynamics of the recombinant cytokine, granulocyte colony stimulating factor (GCSF). Three common stabilising strategies were assessed: site-directed mutagenesis, excipient formulation and lyophilisation. Data found all strategies reduced GCSF flexibility in similar regions of the protein, which was linked to improvements in stability and minimal effect on function. Additionally, regions of increased flexibility were linked with protein destabilisation.

Computational design during biopharmaceutical development is becoming routine for both protein conformation and formulation to reduce the number of screening candidates. *In silico* protein-excipient docking and site-directed mutagenesis data was performed alongside HDX-MS experiments for *in vitro* validation of outputs. Consequently, both *in silico* and biophysical analysis methods provided advancements towards the rational manipulation of the physical, chemical and biological stability of proteins.

Acknowledgements

First and foremost I'd like to thank my supervisors, Professor Paul Dalby of University College London (UCL) and Dr Paul Matejtschuk of The National Institute for Biological Standards and Controls (NIBSC), for drawing up the project, setting up some fantastic lab collaborations and allowing me to grow as an independent researcher. I'd also like to thank them for their stimulating project discussions, and their assistance in editing and amending this thesis.

For project funding and providing doctoral training courses I wish to acknowledge the Engineering and Physical Sciences Research Council (EPSRC) Centre for Doctoral Training in Emergent Macromolecular Therapies (CIMEMT).

I'd like to acknowledge the staff at NIBSC who trained me, kept my experiments going at times and made me feel welcome during my visits. Thank you Dr Jackie Ferguson, Chris Bird, Paula Dilger, Kiran Malik, Chinwe Duru and Dr Adrian Bristow.

Acknowledgements also go to Dr Milena Quaglia and her team at LGC, Kate Groves and Dr Dale Shepherd, for teaching me all things mass spec related, permitting me use their HDX-MS equipment on a regular basis and helping to adapt methods to fit my experimental needs.

Thanks also to Professor Elizabeth Topp and her research group in the Department of Industrial and Physical Pharmacy, at Purdue University for taking me through their solid state HDX-MS methods.

For my *in silico* work, I'd like to thank Dr Teresa Silva Barata, Dr Cheng Zhang and Haoran Yu of the Biochemical Engineering department at UCL for their help and guidance.

Additionally I'd like to thank my Biochemical Engineering PhD cohort at UCL, Rachael Wood, Michael Sibley, Roberto Chiochio, Ivano Colao, Thomas Johnson, Sushobhan Bandyopadhyay and Craig Ward for their camaraderie and fantastic trips over the past four years.

Finally, I want to thank my Mam, Dad and two sisters. I'm incredibly lucky to have such a supportive, loving and awesome family.

Impact statement

Proteins were evolved to function in their natural environment and not for use as biopharmaceuticals. During the different stages of manufacturing there are a number of stressful environments imposed on a protein, which can cause conformational changes to the structure and lead to loss of activity or formation of aggregates. Therefore, methods to stabilise a protein for increased stress resistance is a current and ongoing challenge. The work in this thesis focused on the use of biophysical and computational techniques to both monitor and predict the outcomes of three common stability enhancement strategies: re-engineering, excipient formulation and lyophilisation.

Local regions of high protein flexibility, as measured by hydrogen/deuterium exchange mass spectrometry (HDX-MS), were shown to be successful targets for rational mutagenesis. More importantly, mutants with improved thermal stability were shown by HDX-MS to have reduced flexibility compared to the wild type. This finding provided further evidence to the notion that increased stability can arise from rational rigidification and demonstrated the significance of considering protein dynamics during early phase development and characterisation.

HDX-MS was also used to study proteins in the solid state, where computational predictions and ultra-scale down lyophilisation techniques were integrated with current solid state HDX-MS methods to successfully reduce sample requirements and increase the throughput of the technique.

The importance of selecting appropriate deuterium solution pH was highlighted in this work, whereby specific regions of a model protein became more dynamic with increasing pH. Additionally, acidic deuterated solutions were shown to increase the time window of exchange, providing time resolved exchange of unstructured loop regions, not seen previously in physiological pH studies.

Overall, this thesis showcased the capability of advanced biophysical techniques in combination with *in silico* predictions to understand the driving forces of protein stability, which can be leveraged to increase the shelf-life of biopharmaceuticals and be of benefit to both Industry and patients.

Contents

Abstract.....	3
Acknowledgements.....	4
Impact statement.....	5
List of Figures.....	11
List of Tables.....	16
Abbreviations.....	18
1 Introduction.....	21
1.1 Biopharmaceuticals.....	21
1.2 Biotherapeutic proteins	21
1.2.1 <i>Protein conformation</i>	21
1.2.2 Gibbs free energy.....	22
1.2.3 Protein dynamics	22
1.2.4 Protein folding.....	23
1.2.5 Protein stability	24
1.3 Granulocyte colony stimulating factor	27
1.3.1 Biological function.....	27
1.3.2 Production.....	27
1.3.3 Characterisation.....	28
1.3.4 Structure	28
1.3.5 Receptor interaction sites.....	29
1.3.6 Relevance of granulocyte colony stimulating factor to this work	29
1.3.7 Altering GCSF stability.....	31
1.4 Strategies to decrease protein flexibility	32
1.4.1 Excipients	32
1.4.2 Solid state stabilisation	34
1.4.2.3 <i>Ultra-scale down (USD) rational design of stable lyophilised protein formulations</i>	39
1.4.3 Stabilisation by site-directed mutagenesis	40
1.5 Biophysical analysis of protein stability	45
1.5.1 Standard biophysical analysis	46
1.5.2 Advanced biophysical analysis	47
1.5.2.3 <i>Biophysical analysis of protein-excipient interactions in the solid state</i>	52

1.6	Aims and objectives of thesis.....	55
2	Materials and methods	57
2.1	Molecular biology	57
2.1.1	Plasmid sequencing.....	57
2.1.2	Polymerase chain reaction (PCR) based site-directed mutagenesis.....	57
2.1.3	DNA gel electrophoresis.....	58
2.1.4	Transformation	58
2.1.5	Mutant plasmid sequencing.....	58
2.1.6	Glycerol stocks	59
2.2	GCSF production	59
2.2.1	Large scale cell culture	59
2.2.2	Primary separations	59
2.2.3	Solubilisation and refold	60
2.2.4	Purification	60
2.2.5	Formulation.....	61
2.3	Mutant production	61
2.4	Characterisation.....	62
2.4.1	Ultraviolet (UV) absorbance	62
2.4.2	Non-reduced (NR) SDS-PAGE	62
2.4.3	Bioactivity.....	63
2.4.4	High performance liquid chromatography (HPLC).....	64
2.4.5	Accelerated thermal degradation of GCSF mutants	65
2.4.6	Melt midpoint (T_m) and aggregation onset (T_{agg}) temperatures.....	65
2.4.7	Mass measurements by mass spectrometry	66
2.4.8	Collision induced unfolding ion-mobility spectrometry-mass spectrometry (CIU-IMS-MS)	67
2.4.9	Hydrogen deuterium exchange mass spectrometry.....	68
2.4.10	Ultra-scale down lyophilisation.....	76
2.4.11	GCSF solid state formulation screen.....	77
2.5	Solid state formulation analysis.....	78
2.5.1	Residual moisture analysis by thermogravimetric analysis (TGA)	78
2.5.2	Determination of formulation collapse temperatures (T_c) by freeze-drying microscopy (FDM)	78
2.5.3	Differential scanning calorimetry (DSC).....	79

2.5.4	Reconstitution scoring	79
2.6	<i>In silico</i> applications	79
2.6.1	Structure selection	79
2.6.2	Changes in mutant folding energy ($\Delta\Delta G$) predictions.....	79
2.6.3	Excipient docking	81
3	Biophysical and <i>in silico</i> method development.....	82
3.1	Results.....	84
3.1.1	WT plasmid sequencing	84
3.1.2	GCSF production	86
3.1.3	Characterisation.....	88
3.1.4	Development of a peptide-level HDX-MS method for GCSF	92
3.1.5	Effect of pH on GCSF HDX	95
3.1.6	Effect of buffer and protein concentration on thermal stability	101
3.1.7	Optimised peptide level HDX-MS for GCSF.....	102
3.1.8	Structure selection for <i>in silico</i> research	103
3.2	Discussion.....	107
3.2.1	GCSF production	107
3.2.2	Buffer and protein concentration have large effects on GCSF aggregation	107
3.2.3	Characterisation of GCSF	108
3.2.4	Peptide level GCSF HDX-MS.....	109
3.2.5	Effect of exchange solution pH on GCSF HDX-MS.....	110
3.2.6	Structure selection for <i>in silico</i> research.....	112
4	GCSF stability mutants I: <i>in silico</i> design, production and stability analysis.....	114
4.1	Results.....	115
4.1.1	Stabilisation by rational design	115
4.1.2	RosettaDesign target selection	116
4.1.3	Stabilisation by computational design	117
4.1.4	Production and characterisation of mutants.....	121
4.1.5	Hydrophobicity.....	137
	Mutant characteristics summary	138
4.2	Discussion.....	140
4.2.1	RosettaDesign	140
4.2.2	RosettaDesign outcome	141
4.2.3	Rosetta_ddg_monomer	143

4.2.4	Rosetta_ggd_monomer outcome.....	144
4.2.4.1	<i>T38W and L71W</i>	144
4.2.5	Rigidification as a strategy for stabilisation.....	145
4.2.6	Notes for future mutant production.....	146
5	GCSF stability mutants II: advanced biophysical analysis	148
5.1	Results.....	149
5.1.1	Mutant selection.....	149
5.1.2	Mutant identity.....	149
5.1.3	Accelerated thermal degradation.....	151
5.1.4	GCSF Mutant HDX-MS.....	158
5.1.5	Predicting stability by protein flexibility.....	164
	Discussion.....	165
5.1.6	Biophysical methods to compare WT and mutants.....	165
5.1.7	Mutant identity.....	166
5.1.8	Purity by SEC-HPLC.....	166
5.1.9	Accelerated thermal degradation.....	167
5.1.10	Mutant HDX-MS.....	168
5.1.11	Flexibility as a predictor of protein degradation/ stability.....	171
6	Mapping GCSF-excipient interactions in the aqueous state	172
6.1	Results.....	173
6.1.1	Effect of different excipients on GCSF T_{agg}	173
6.1.2	Excipient docking.....	177
6.1.3	Correlating docking and T_{agg} values.....	178
6.1.4	Residue interaction analysis.....	179
6.1.5	Cluster analysis.....	181
6.1.6	Deuterated excipient HDX-MS.....	182
6.1.7	Equilibration tests using intrinsic fluorescence and static light scattering..	193
6.1.8	Effect of internal reference peptide on GCSF stability.....	195
6.2	Discussion.....	196
6.2.1	Excipient selection.....	196
6.2.2	High throughput screening.....	197
6.2.3	Evaluating iGEMDOCK total binding energies.....	197
6.2.4	Evaluation of iGEMDOCK excipient docked poses.....	198
6.2.5	Exploring GCSF-excipient interactions using HDX-MS.....	199

6.2.6	Internal reference peptide reports changes in intrinsic exchange	199
6.2.7	Internal reference peptide does not alter GCSF stability	200
6.2.8	Mechanisms of GCSF stabilisation by excipients	200
6.2.9	Formulation mixing equilibration	208
7	Mapping GCSF-excipient interactions in the solid state	210
7.1	Results.....	211
7.1.1	Ultra-scale down lyophilisation.....	211
7.1.2	GCSF-sucrose ssHDX-MS	219
7.1.3	USD moisture induced cake collapse study	228
7.1.4	USD excipient screen with internal reference peptide.....	228
7.1.5	ssHDX-MS to explore the protective interactions of mannitol and phenylalanine.....	232
7.1.6	GCSF-mannitol ssHDX-MS.....	235
7.1.7	GCSF-phenylalanine ssHDX-MS.....	237
7.1.8	Comparing excipient protection factors	239
7.1.9	Comparing ssHDX-MS and docking H-bond data.....	240
7.1.10	USD lyophilisation	242
7.2	ssHDX-MS with GCSF.....	246
7.2.1	ssHDX-MS implementation on Waters system	246
7.2.2	Improvements observed for the GCSF peptide map coverage.....	247
7.2.3	Moisture induced collapse of samples caused issues during labelling.....	247
7.2.4	Interactions of GCSF-excipients in the solid state	251
7.2.5	Future work.....	255
8	Conclusion	257
8.1	Future development of HDX-MS for lead candidate selection	263
9	References	264
10	Appendix.....	278

List of Figures

Figure 1. The protein folding energy landscape.	24
Figure 2. 3D structure of recombinant human granulocyte colony stimulating factor as determined by NMR.....	28
Figure 3. Structures of frozen material.....	35
Figure 4. Natural amino acid nomenclatur	42
Figure 5. Venn diagram of amino acid characteristics.....	43
Figure 6. Hierarchy of protein biophysical analysis techniques.....	46
Figure 7. HDX-MS process flow diagram.....	50
Figure 8. Schematic of the Waters automated HDX-MS system	52
Figure 9. Schematic of ssHDX-MS sample preparation and analysis.....	53
Figure 10. Result chapter schematic.....	56
Figure 11. Example of SLS GCSF data with determination of the temperature of aggregation onset (T_{agg})	66
Figure 12. GCSF pET21a plasmid schematic.....	84
Figure 13. Alignment of sequenced GCSF gene and original Bristow et al. (2012) DNA sequence	85
Figure 14. BL21 (DE3) <i>E.coli</i> cell fermentation and GCSF expression	86
Figure 15. SEC chromatogram for standard GCSF purification.....	87
Figure 16. Non-reduced SDS-PAGE of SEC fractions.....	87
Figure 17. SEC-HPLC analysis of GCSF fraction pool	88
Figure 18. Intact protein mass analysis of GCSF fraction pool.....	89
Figure 19. Folded 5V MS spectra of GCSF in 50 mM ammonium acetate pH 4.25.....	90
Figure 20. Unfolding of GCSF as measured by collision induced unfolding ion-mobility spectrometry-mass spectrometry (CIU-IMS-MS).....	91
Figure 21. GCSF Peptide maps pre- and post- quench solution optimisation.....	93
Figure 22. Repeatability of GCSF peptide relative uptake as measured by HDX-MS	96
Figure 23. Relative uptake of GCSF peptides over time with differing solution pD and sample pH as measured by HDX-MS.....	97
Figure 24. Differential plots of GCSF peptide uptake in differing pD solutions, corrected to a standard exchange time, as measured by HDX-MS.....	100
Figure 25. Effect of protein and buffer concentration on GCSF thermal denaturation	101
Figure 26. Relative deuterium uptake of GCSF in 10mM sodium acetate pH 4.25 as measured by HDX-MS.....	103

Figure 27. B-factor analysis of GCSF crystal structures.....	106
Figure 28. Average HDX-MS % rate plotted against the average 2D9Q B-factor value for different structured regions of GCSF	115
Figure 29. RosettaDesign predicted mutations and $\Delta\Delta G$ values	116
Figure 30. RosettaDesign top nine mutation locations	117
Figure 31. Point mutation map of GCSF by Rosetta_ddg_monomer	119
Figure 32. Lowest $\Delta\Delta G$ mutant analysis.....	120
Figure 33. Rosetta_ddg_monomer top eleven mutation locations.....	121
Figure 34. Purification yield of WT and mutants	123
Figure 35. Non-reduced SDS-PAGE analysis of GCSF WT and mutant purified samples	124
Figure 36. Comparison of RosettaDesign GCSF mutant bioactivity relative to WT using the GNFS-60 cell proliferation bioassay	125
Figure 37. Comparison of Rosetta_ddg_monomer GCSF mutant bioactivity relative to WT using the GNFS-60 cell proliferation bioassay	126
Figure 38. Fluorescence intensity measurements of RosettaDesign mutants relative to WT GCSF	128
Figure 39. Fluorescence intensity measurements of Rosetta_ddg_monomer mutants relative to WT GCSF 1	129
Figure 40. Fluorescence intensity measurements of Rosetta_ddg_monomer mutants relative to WT GCSF 2	130
Figure 41. Static light scattering of RosettaDesign GCSF mutants in comparison to WT 1.	132
Figure 42. Static light scattering of RosettaDesign GCSF mutants in comparison to WT 2.	133
Figure 43. Static light scattering of Rosetta_ddg_monomer GCSF mutants in comparison to WT 1.....	134
Figure 44. Static light scattering of Rosetta_ddg_monomer GCSF mutants in comparison to WT 2.....	135
Figure 45. Static light scattering of Rosetta_ddg_monomer GCSF mutants in comparison to WT 3.....	136
Figure 46. RP-HPLC of WT, C17A and Q120I	137
Figure 47. Mass difference of GCSF mutants with WT by LC-QToF-MS.....	149
Figure 48. WT GCSF IMS deconvoluted peak example	150
Figure 49. Initial sample content of WT GCSF and mutant samples by SEC-HPLC.	151
Figure 50. SEC-HPLC chromatograms of WT GCSF stored at 45 °C over time	153
Figure 51. Monomer loss for GCSF and mutants at 45 °C incubation	154

Figure 52. Aggregation for GCSF and mutants at 45 °C incubation	155
Figure 53. Dimerisation for GCSF and mutants at 45 °C incubation	157
Figure 54. (NR) SDS-PAGE of thermally degraded GCSF and mutant samples after incubation for 7 days at 45 °C	158
Figure 55. G51R and Q120I peptide differential uptake plots	161
Figure 56. C17A peptide differential uptake plot	162
Figure 57. L71W, Q107Y and Q131F peptide differential uptake plots	163
Figure 58. GCSF mutant average peptide relative differential uptake versus initial monomer loss rate	164
Figure 59. Static light scattering with increasing GCSF concentration.	174
Figure 60. Static light scattering with increasing excipient concentration	175
Figure 61. Highest T_{agg} value obtained for each excipient with 0.5 mg/mL GCSF in 10 mM sodium acetate pH 4.25	176
Figure 62. Average total pose energy of excipients docked with GCSF PDB 2D9Q using iGEMDOCK	178
Figure 63. Correlation between docking results (E_{Dock}) and T_{agg}	179
Figure 64. Average GCSF residue docking energies with excipients using IGEMDOCK	180
Figure 65. 3D representation of excipient clustering on 2D9Q	181
Figure 66. Image of freeze-dried cakes formed during excipient deuteration	182
Figure 67. GCSF peptide map for deuterated excipient HDX-MS experiments	183
Figure 68. Differential uptake plots of GCSF pH 4.25 in the presence of deuterated arginine over time	186
Figure 69. Differential uptake plots of GCSF pH 4.25 in the presence of deuterated mannitol over time	187
Figure 70. Differential uptake plots of GCSF pH 4.25 in the presence of deuterated phenylalanine over time	188
Figure 71. Differential uptake plots of GCSF pH 4.25 in the presence of deuterated sucrose over time	189
Figure 72. Alignment of HDX-MS differential uptake with in silico docking data for GCSF and mannitol	190
Figure 73. Alignment of HDX-MS differential uptake with in silico docking data for GCSF and sucrose	191
Figure 74. Alignment of HDX-MS differential uptake with in silico docking data for GCSF and phenylalanine	192

Figure 75. Alignment of HDX-MS differential uptake with in silico docking data for GCSF and arginine	193
Figure 76 Effect of sample equilibration time on GCSF SLS and fluorescence measurements	194
Figure 77. Thermal stability of GCSF with and without an internal reference peptide.....	195
Figure 78. Schematic presentation of A) preferential binding and B) preferential exclusion	200
Figure 79. Structure of phenylalanine in solution.....	203
Figure 80. Relative uptake of deuterium of GCSF loopCD peptide 123-138 over time.....	206
Figure 81. Two-step preferential binding and exclusion model.	209
Figure 82. Residual moisture of lyophilised cakes within glass vials and varied positions within a 96-well microtiter plates.....	212
Figure 83. Image of lyophilised GCSF formulations within a USD 96-well microtitre plate	213
Figure 84. SEC-HPLC monomer recovery % of GCSF lyophilised with different formulations	215
Figure 85. Sodium acetate GNFS-60 cell proliferation bioassay	216
Figure 86. Sodium citrate GNFS-60 cell proliferation bioassay	217
Figure 87. PBS GNFS-60 cell proliferation bioassay	218
Figure 88. Freeze-drying microscope (FDM) images of GCSF in acetate and 50 mM sodium citrate.....	219
Figure 89. Monomer content of GCSF samples with sucrose pre- and post-lyophilisation	220
Figure 90. Peptide map of reconstituted GCSF in 50 mM citric acid pH 4.25.....	221
Figure 91. Differential plots of lyophilised GCSF uptake with and without sucrose, as measured by ssHDX-MS.....	223
Figure 92. Differential plots of lyophilised GCSF peptide uptake with and without the internal reference peptide, as measured by ssHDX-MS.....	225
Figure 93. Differential plots of lyophilised GCSF and uptake with and without sucrose and including the internal reference peptide in both samples, as measured by ssHDX-MS.....	226
Figure 94 Correlation between differential sucrose uptake values, with and without the internal reference peptide.....	227
Figure 95. Relative deuterium uptake for internal reference peptide lyophilised with different excipient.....	229
Figure 96. DSC of 50 mM citric acid with different excipients.....	230
Figure 97. Freeze-drying microscope (FDM) images of GCSF with different excipients.....	232

Figure 98. Freeze-dried cake appearance during labelling with deuterium using a desiccator at 43% RH.....	233
Figure 99. Monomer content of GCSF with mannitol and phenylalanine pre- and post-lyophilisation.....	234
Figure 100. Differential uptake plots for GCSF lyophilised with mannitol as measured by ssHDX-MS.....	236
Figure 101. Differential uptake plots for GCSF peptides lyophilised with phenylalanine as measured by ssHDX-MS.....	238
Figure 102. ssHDX-MS protected regions aligned with docked regions for GCSF and excipients.....	241
Figure 103. Stabilising of proteins in solution and dry state	252
Figure 104. HDX micro-reactor chip.....	263

List of Tables

Table 1. Freeze-dry cycle for VirTis Advantage dryer with a 1 mL fill within 2 mL glass vials.	72
Table 2. 42h r lyophilisation cycle for 200 μ L filled 2 mL vials using a VirTis Genesis freeze-dryer.....	73
Table 3. 24 hr lyophilisation cycle for with 50 μ L filled 1.4 mL Micronic tubes.....	75
Table 4. Freeze-dry cycle of 96-well microtitre plates.....	76
Table 5. Freeze-dry cycle 96-well microtitre plates with anneal step	76
Table 6. Layout for liquid and solid state samples within a 96-well microtitre plate.....	77
Table 7. Resfile text used for RosettaDesign point mutation modeling.	80
Table 8. Optimising HDX-MS quench solution.....	92
Table 9. Optimising the processing parameters during pepsin digestion step of HDX-MS... 94	
Table 10. HDX reaction condition and exchange time corrected to standard HDX condition of pH 7.4 and 22 $^{\circ}$ C.)	99
Table 11. Effect of protein and buffer concentration on GCSF thermal denaturation linear fitting.....	102
Table 12. All available PDB structures for GCSF on RCSB website in 2016.....	105
Table 13. The nine lowest $\Delta\Delta G$ RosettaDesign mutations, their change in molecular weight (ΔMW) and the location within GCSF secondary structure.	117
Table 14. The eleven lowest $\Delta\Delta G$ Rosetta_ddg_monomer mutations, their change in molecular weight (ΔMW) and the location within GCSF secondary structure	121
Table 15. GCSF mutant characterisation summary table..	139
Table 16. Curve fitting values for Figure 5A GCSF and mutant monomer loss at 45 $^{\circ}$ C.	154
Table 17. Curve fitting for Figure 6A GCSF and mutant aggregation rates at 45 $^{\circ}$ C.	155
Table 18. Line fitting for Figure 6A WT GCSF aggregation rate at 45 $^{\circ}$ C.....	156
Table 19. Line fitting for Figure 7A GCSF and mutant dimerisation rates at 45 $^{\circ}$ C incubation.	157
Table 20. Peptide map summary for WT and mutant samples	159
Table 21. Line fitting for GCSF mutant differential uptake versus initial monomer loss rate.	164
Table 22. Range of excipient concentrations added to formulation screen samples.	173
Table 23. Excipient summary	177
Table 24. Relative deuterium uptake for internal reference peptide with different excipient formulations.....	184

Table 25. Categorisation for sample positions within a 96-well microtiter plate	211
Table 26. Reconstitution scores for GCSF in different lyophilised formulations	214
Table 27. Relative deuterium uptake for an internal reference peptide with sucrose solid state formulations.....	222
Table 28. Sum of all peptide differential values for each excipient at different incubation times	239

List of abbreviations

3D	Three dimensional
ACN	Acetonitrile
Amp	Ampicillin
ASM	Auxiliary solvent manager
AUC	Analytical ultracentrifugation
BCM	Barycentric mean fluorescence intensity
BEH	ethyl-bridged hybrid
BSM	Binary solvent manager
CD	Circular dichroism
CDR	complementarity-determining loop
CEX	Cation exchange
CHO	Chinese hamster ovary
CIU	Collision induced unfolding
CMD	Computational molecular design
CV	Column volume
D₂O	Deuterium oxide
Da	Dalton
DCI	Deuterium chloride
DF	Digestion factor
DIA	Data independent acquisition
DLS	Dynamic light scattering
DoE	Design of experiments
DSC	Differential scanning calorimetry
EDTA	Ethylenediaminetetraacetic acid
ESI	Electrospray ionization
FA	Formic acid
FDA	Food and Drug Administration
FDM	Freeze-drying microscopy
fMet	Formylmethionine
FTIR	Fourier transform infrared spectroscopy
GCSF	Granulocyte-colony stimulating factor
GCSFR	Granulocyte-colony stimulating factor receptor

Gdn.HCl	guanidine hydrochloride
h	Hour
H-Bond	Hydrogen bond
HCP	Host cell protein
HDX	Hydrogen-deuterium exchange
HOS	Higher order structure
HPLC	High performance liquid chromatography
HSA	Human serum albumin
IB	Inclusion body
IFL	Intrinsic fluorescence
IgG	Immunoglobulin
IMS	Ion-mobility spectrometry
IPF	Intrinsic protein fluorescence
IPTG	Isopropyl β -D-1-thiogalactopyranoside
IRP	Internal reference peptide
LB	Luria Bertani
LC-MS	Liquid chromatography mass spectrometry
mAb	Monoclonal antibody
MD	Molecular dynamics
min	Minute
MS	Mass spectrometry
MW	Molecular weight
NH	Amide hydrogen
NIBSC	The National Institute for Biological Standards and Control
NIR	Near infrared spectroscopy
NMR	Nuclear magnetic resonance
NR	Non-reduced
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PAM	point accepted mutation
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDA	Protein design automation
PDB	Protein data bank

PES	Polyethersulfone
PhD	Doctor of philosophy
PLGS	ProteinLynx Global Server
QbD	Quality by design
QtoF	Quadruple time of flight
RMSF	Root-mean-square fluctuation
RP	Reverse phase
RT	Room temperature
S	Second
SAXS	Small angle X-ray scattering
SDS	Sodium dodecyl sulphate
SEC	Size-exclusion chromatography
SLS	Static light scattering
ssHDX-MS	Solid-state hydrogen-deuterium exchange mass spectrometry
T_{agg}	Onset of aggregation temperature
TB	Terrific broth
T_c	Collapse temperature
TCEP	Tris(2-carboxyethyl)phosphine
T_e	Eutectic point temperature
T_g'	Glass transition temperature
TGA	Thermo-gravimetric analysis
TK	Transketolase
T_m	Midpoint of unfolding temperature
T_{onset}	Onset of unfolding temperature
UPLC	Ultra-performance liquid chromatography
USD	Ultra-scale down
UV	Ultraviolet
V	Volts
VDW	van der Waals
w_g'	Maximum freeze concentration
WT	Wild type
ΔG	Gibbs free energy
$\Delta\Delta G$	Difference in Gibbs free energy

1 Introduction

1.1 Biopharmaceuticals

Biotechnology describes the use of biological processes, manipulated or modified in some way by modern science. A major industrial application of biotechnology is in the development and preparation of biopharmaceutical products from natural sources using recombinant techniques. Such products offer promising treatments for many previously incurable diseases and include cytokines, growth factors, hormones, interferons and other regulatory peptides and biotherapeutic proteins, as well as products from novel cell lines (WHO 2018). As a consequence, biopharmaceuticals have a large presence in the global healthcare landscape. In 2013, the global biologics market totalled £262.2 billion (\$200.6 billion) and, at a compound annual growth rate of 10.6, is expected to reach £505.4 billion (\$386.7 billion) by the end of 2019 (Highsmith 2015).

1.2 Biotherapeutic proteins

Biotherapeutic proteins are an important class of large biologic molecules that are classified more generally as biomacromolecules or biopolymers. Proteins are highly complex structures, made up of 20 different naturally occurring monomer units called amino acids. The amino acids are linked together by peptide bonds along the polypeptide backbone. The uniqueness in the linear arrangement of amino acids is a key element for the structural complexity of proteins, and gives rise to a diverse range of functionalities.

1.2.1 *Protein conformation*

Proteins are three dimensional (3D) structures. There are four different levels to the 3D structure, the first is the amino acid sequence, which is known as the primary structure, the second is the local folding patterns in the primary structure, forming α -helix and β -sheets created by the hydrogen bonding (H-Bond) between the carbonyl and amide groups on the polypeptide chain, also known as the secondary structure. The third level is the higher level of folding between secondary structures to obtain the final 3D structure of a single polypeptide chain, which is known as the tertiary structure. Finally, the fourth level is the interaction of two or more identical or different polypeptide chains, which is known as the quaternary structure.

During protein folding to its 3D native structure, various weak noncovalent bonds form the secondary structures, including ionic, dipoles (H-Bonds), nonpolar (hydrophobic effect), and van der Waals (VDW) interactions. These weak bonds involve the interactions of amino acid side chains, as well as elements of the polypeptide backbone, particularly the amide hydrogens. While individually these interactions are weak, their large number and the cooperative manner in which they form (during the folding process) provide the necessary enthalpic and entropic, release of structured water via the hydrophobic interactions between amino acids such as leucine and isoleucine, driving force to override the large unfavourable decrease in entropy that occurs as a protein folds into its native conformation (Berkowitz and Houde 2015). Primary bonds can also form between the folded regions to provide additional stabilisation of the 3D structure. The most common bond is the disulphide bond, which is formed between cysteine residues, and can occur within a polypeptide chain (intrachain) and/or between two different chains (interchain).

1.2.2 Gibbs free energy

The Gibbs free energy is a way of measuring the stability of a folded protein by calculating the difference in enthalpy and entropy. It follows the 2nd law of thermodynamics whereby the total entropy of a system and its surroundings always increases for a spontaneous process. The Gibbs free energy (ΔG) of a folded protein can be calculated using the equation:

$$\Delta G = \Delta H - T\Delta S \quad \text{Equation 1}$$

where ΔH is the sum of the energy of the bonds and interactions holding the structure in place (H-Bonds, VDW, electrostatic, disulphide, etc.), and $T\Delta S$ is the chain conformational entropy energy (water and hydrophobic effects), in kJ/mol. If a process is advantageous i.e. a protein is folded, the change in Gibbs free energy is negative ($\Delta G < 0$).

1.2.3 Protein dynamics

A delicate balance between stabilising and destabilising contributions causes a stability of most globular proteins in solution to be in the range of around 10 to 70 kJ mol⁻¹ (Privalov 1979; Nölting 2006). Thus, the stabilisation of a folded protein by its bonds is only marginal. When comparing Gibbs free energy against the average thermal energy content of a protein molecule (which is equal to kT , where k = Boltzmann constant and T = temperature) and the distribution of this energy in terms of the amount of thermal energy per molecule,

varying numbers of these weak secondary bonds can be broken as a function of time (Berkowitz and Houde 2015). This breaking of bonds, over time, creates different levels of fluctuations within a protein's higher order structure and leads to dynamic structural conformations. The intrinsic dynamic conformation of proteins has been shown to play an important role in their function, allowing them to change and respond to the presence of other molecules and/or other stimuli (Teilum et al. 2009). Biological and biochemical processes including signal transduction, antigen recognition, protein transport and enzyme catalysis rely on this ability to change conformation or to adapt to change.

1.2.4 Protein folding

The concept of "protein folding funnels" was described by Leopold et al. (1992) as a kinetic mechanism for understanding the self-organizing mechanism of proteins based on a sequence-structure relationship. The vertical axis of the funnel represents the internal free energy. At the top of the funnel is the unfolded protein with high folding energy and entropy and at the bottom is the fully folded protein in its native state with the lowest folding energy and entropy. The native state is a collection of conformations which differ by a small amount of free energy. The unfolded state is entropy dominated which eventually decreases and the protein flows towards the bottom of the funnel, with *in vitro* folding times from microseconds to seconds (Kamerzell and Engen 2008).

Protein folding can also be explained by the "protein folding energy landscape" concept, first described by Frauenfelder et al. (1991). The protein folding energy landscape is also funnel-like but includes residual "ruggedness" where multiple folding routes and sub-ensembles exist but have a preferred direction of flow (Kamerzell and Middaugh 2008). A particular energy landscape is resigned to specific temperature, pressure and solvent conditions. Figure 1 shows a one dimensional cross-section of the protein folding energy landscape concept, showing the hierarchy of protein dynamics and energy barriers. States are defined as a minimum in the energy surface and a transition state is the maximum between the wells. The states are split into three tiers based on their folding energy (ΔG), where lower tiers have faster fluctuations between large numbers of closely related substrates. The populations of the states A and B (P_A and P_B) are defined as Boltzmann distributions based on their difference in free energy (ΔG_{AB}). The barrier between these states referred to as ΔG^\ddagger , which determines the rate of interconversion (k). Changes to the system, as seen by the transition of lines from dark blue to light blue or from dark blue to

light blue, causes a change in the ratio of the populations, and can be caused by changes in external conditions or mutation (Henzler-Wildman and Kern 2007).

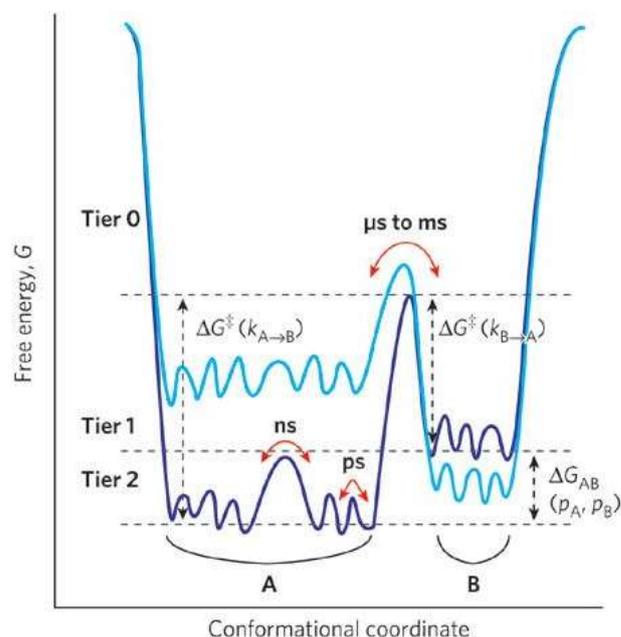


Figure 1. **The protein folding energy landscape.** Taken from Henzler-Wildman and Kern 2007.

Figure 1 also defines the amplitude and timescale of protein dynamics where the slow fluctuations for Tier 0 are in the amplitude of μs to ms , between distinct states that are separated by large energy barriers. Typically these are larger amplitude collective motions between relatively small numbers of states (Henzler-Wildman and Kern 2007). The fluctuations such as those for Tier 1 and 2 have fast timescale dynamics, and are typically a large ensemble of structurally similar states separated by small energy barriers with more local small amplitude (ps to ns) fluctuations.

1.2.5 Protein stability

Proteins were evolved to function in their natural environment and not for use as biopharmaceuticals. During the different stages of manufacturing there are a number of stressful environments imposed on a protein, which can cause conformational changes to the structure and lead to loss of activity or formation of aggregates. Therefore, methods to stabilise a protein for increased stress resistance is a current and ongoing challenge.

There are two main types of protein instability: chemical and physical. Chemical instability involves the covalent modification of a protein or amino acid residue to generate a new molecule. This can include bond cleavage or formation, rearrangement and substitution.

Physical instability occurs when the integrity of the 3D structure is altered, and includes conditions such as denaturation, aggregation (insolubility), precipitation and adsorption (Manning et al. 2010). In some cases, the chemical and physical instability pathways are synergistic where a chemical reaction may trigger a physical reaction, such as when oxidation is followed by aggregation (Patel et al. 2011). Irreversible protein aggregation is a major issue in the biotechnology industry, as it can be encountered throughout all stages of production including refold, purification, sterilisation, shipping, and long-term storage. Aggregation can have a number of deleterious effects on biotherapeutic proteins including the loss of efficacy, increased immunogenicity, altered pharmacokinetics, and reduced shelf-life.

1.2.5.1 *Aggregation*

Protein aggregation denotes the process by which individual (monomeric) protein molecules assemble into stable complexes composed of two or more proteins. Aggregates vary in size and can be soluble, insoluble, covalent, noncovalent, reversible or irreversible; they may be visible or may be present as subvisible particles (Maggio 2016).

1.2.5.1.1 Covalent

Covalently linked protein aggregation is caused by chemical reactions such as β -elimination, disulphide bond exchange or transamidation. Disulphide bond exchanged aggregates can be identified using detergents such as sodium dodecyl sulphate (SDS), where the aggregates dissociate and the high performance liquid chromatography HPLC peak or SDS-polyacrylamide gel electrophoresis (SDS-PAGE) band observed during analysis is lost. A study by Ribarska et al. (2008), studied the aggregation of glycosylated granulocyte-colony stimulating factor (GCSF; Lenograstim) under physiological conditions, where they observed a loss in higher molecular mass band under reducing conditions. As such they demonstrated that the soluble aggregates observed during their study were disulphide bonded.

1.2.5.1.2 Non-covalent

Non-covalent aggregation can be caused by interactions between exposed hydrophobic residues of unfolded or misfolded protein molecules (Maggio 2016), or by cross β -sheet formation. Generally, protein aggregation involves β -sheets due to their weaker dipole moment compared to α -helices (Querol et al., 1996; Wang 2005). β -strand swapping was shown to cause aggregation in a study by Das et al. (2011). They found aggregation of

human γ D-crystallin was caused by successive domain swapping at the C-terminal β -strands and alanine substitutions (an amino acid of low β -sheet propensity) of the hydrophobic residues in those aggregation-prone β -strands prevented domain swapping.

1.2.5.2 Conformational stability

Large changes in the secondary and tertiary structures of a protein are typically followed by aggregation. In addition, intermediates that are structurally expanded compared to the native state have been found to precede aggregation (Krishnan et al. 2002; Raso et al. 2005). As propensity of a protein to unfold is governed by its conformational stability, non-covalent aggregation can be accelerated by stressors such as elevated temperature or low pH to perturb the conformational structure of the protein structure.

1.2.5.2.1 Colloidal stability

The propensity of proteins to aggregate once unfolded is known as non-native colloidal stability. During low colloidal stability, intermolecular interactions are sufficiently attractive to condense unfolded proteins into non-native aggregates. For high colloidal stability, the intermolecular interactions are repulsive, or insufficiently attractive, and the unfolded protein will fold reversibly without aggregating (Perchiacca and Tessier 2012). Colloidal stability is commonly measured by the second osmotic virial coefficient with subscripts denoting protein–protein interactions (B_{22} value):

$$B_{22} = \frac{2\pi}{M^2} \int_0^{\infty} r^2 (1 - e^{-\frac{u(r)}{kT}}) dr \quad \text{Equation 2}$$

where M is the protein molecular weight, π is osmotic pressure, r is the intermolecular separation distance, $u(r)$ is the interaction potential, k is the Boltzmann constant, and T is the absolute temperature. The interaction potential, $u(r)$, describes all of the interaction forces between two protein molecules, which include H-bond, electrostatic interactions, VDW, and all other short-range interactions. A positive B_{22} values indicates high colloidal stability, where protein–solvent interactions are favored over protein–protein interactions. Whereas, negative B_{22} values reflect low colloidal stability, with protein–protein interactions being favored (Chi et al. 2003). Non-ionic surfactants are widely used in the formulation of biotherapeutic proteins to prevent colloidal based aggregation, as well as minimise surface absorption (Maggio 2016). However, a common issue with the use of non-ionic surfactants is that they contain polyoxyethylene moieties, which can auto-oxidize to

produce reactive peroxides along with other chemically reactive species causing an increase in unwanted protein immunogenicity (Maggio 2016). An alternative to the use of surfactants is re-engineering of structures to reduce hydrophobic interactions, which may be a safer option for long-term storage of biopharmaceuticals.

1.3 Granulocyte colony stimulating factor

1.3.1 Biological function

Granulocyte-colony stimulating factor (GCSF) belongs to a group of secreted glycoproteins called colony stimulating factors (CSFs) that bind to receptor proteins on the surfaces of hemopoietic stem cells to activate intracellular signaling pathways, which in turn cause the cells to proliferate and differentiate into a specific kind of blood cell (Welte 2012). The clinical use of GCSF is in chemotherapies to increase white blood cell count following reduction (neutropenia) caused by cytotoxic therapeutic agents (Bishop et al. 2001).

1.3.2 Production

The first commercially available recombinant, human GCSF (r-huGCSF) product was produced by Amgen Inc. under the trade name Neupogen® and was approved by the United States Food and Drug Administration (FDA) for use in chemotherapy induced neutropenia in 1991 (Herman et al. 1996). Following its patent expiry, the first GCSF biosimilar was approved by the FDA in May 2015 for ZARXIO (Sandoz Ltd). Both Neupogen® and ZARXIO are produced via biotechnological methods due to low availability of GCSF from natural sources.

r-huGCSF can be produced in a number of expression systems including mammalian cells such as Chinese hamster ovary cells (CHO; referred to as Lenograstim) and microbial cells such as *Escherichia coli* (referred to as Filgrastim). Both systems have their processing advantages and disadvantages and each of their final products differs from the naturally occurring protein. For CHO, Lenograstim has different glycosylation patterns and for *E.coli*, Filgrastim has an additional N-terminal methionine residue and the absence of glycosylation. These changes to the protein caused by the type of expression system showed no detrimental effects to the functionality of the protein, and both products have been shown to be capable of supporting myeloid proliferation and differentiation in biological assays (Bendall and Bradstock 2014).

1.3.3 Characterisation

The human GCSF protein contains 177 amino acids, and has a molecular weight of 19,600 Da. It is also O-glycosylated at a single site (T133). Filgrastim contains 175 amino acids (including the N-terminal methionine, M) and has a molecular weight of 18,800 Da due to the absence of glycosylation. Both human GCSF and Filgrastim have five cysteine residues, four of these form two intermolecular disulphide bridges at residues C36-C42 and C64-C74, which are of essential importance for the activity of the protein. The remaining cysteine, C17, is free but is only partially solvent accessible (Arvedson and Giffin 2012).

1.3.4 Structure

The secondary and tertiary structure of Filgrastim was determined using nuclear magnetic resonance (NMR) spectrometry with ^{15}N and ^{13}C labelling (PDB 1GNC; Zink et al. (1994) determined both). Four main helices form between residues 11-41 (αA), 71-95 (αB), 102 - 125 (αC), and 145-175 (αD), which configure into an antiparallel helix bundle with long overhead connecting loops (Figure 2). A short helix is also located between the loops joining αA and αB . Of the α -helices, αA and αB are aligned parallel to each other (up up), and antiparallel to αC and αD (down down). Because αA and αB both extend in the same direction, the connecting peptide loop spans the length of the protein and is called the “cross over” region (Arvedson and Giffin 2012). The helical cytokine family is classified into subfamilies based on either short- or long-chain cross over regions. GCSF is part of the long-chain subfamily because the crossover region passes behind αD (Arvedson and Giffin 2012).

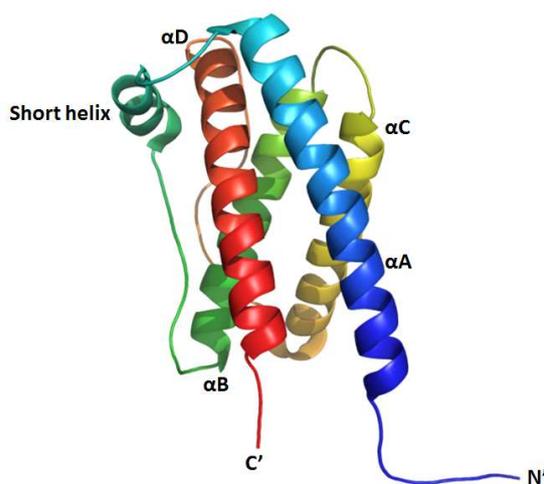


Figure 2. **3D structure of recombinant human granulocyte colony stimulating factor as determined by NMR.** The image was prepared using the PDB 1GNC (Zink et al. 1994) and the visualisation software PyMOL (Schrödinger, NY, USA), where the different α -helix were coloured with a rainbow and labelled from A-D.

1.3.5 Receptor interaction sites

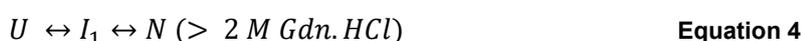
GCSF binds to its receptor (GCSFR) via a major site (II) and minor site (III) in a 2:2 homodimeric complex (Tamada et al. 2012). The major site of interaction is found between residues in αA and αC , and the minor site between residues in αD and the short helix.

1.3.6 Relevance of granulocyte colony stimulating factor to this work

Filgrastim was selected as a model protein for the study of protein dynamics and stability in this thesis due to its small size and basic structure and the availability of an expression system to facilitate mutational analysis. For simplicity it will be referred to as GCSF from hereon in.

1.3.6.1.1 Unfolding of GCSF with denaturants

The folding kinetics of GCSF was determined by Brems et al. (2002) using tryptophan (W) fluorescence and circular dichroism (CD). Using the denaturant, guanidine hydrochloride (Gdn.HCl) and it was found to follow two different mechanisms depending on the concentration of guanidine used:



where U is the unfolded protein, I_1 is the first intermediate state, I_2 is a second intermediate state and N is the native, folded state. I_1 is characterised as having approximately one-half of the native helical structure and none of the native fluorescence. I_2 has 100 percent of the native helical structure and most of the W118 and minimal W58 native fluorescence.

1.3.6.1.2 Aggregation

Aggregation and chemical degradation are the predominant mechanisms for destabilisation of GCSF (Herman et al. 2002). The mechanism of GCSF aggregation is:



where N is the native protein and N' is an aggregation prone conformation (Raso et al. 2005). Only the N' state aggregates and the rate of aggregation is increased with temperature. The aggregation mechanism differs from the unfolding mechanism described

by Brems (2002), as N' is not an intermediate in the folding pathway, but rather a structurally “expanded transition state” (Raso et al. 2005). The structurally expanded transition state was previously shown by Krishnan et al. (2002) to be monomeric with a 15% increase in surface compared to the fully folded state. The tendency of GCSF to assume altered conformations with varying pH has been well established (Kolvenbach et al. 1997). At low pH, GCSF is resistant to aggregation. At physiological pH, the free energy of unfolding is substantially decreased and as such a second mechanism of aggregation occurs which is more similar to the GCSF unfolding mechanism:



where U is the unfolded protein, I is an intermediate and N is the native protein (Raso et al. 2005).

1.3.6.1.3 Thermal stability

The most common way to measure conformational stability is via thermal denaturation. Thermodynamic stability, given by the Gibbs energy difference, allows for a read-out of protein conformational stability, such as the start (T_{onset}) or midpoint temperature of unfolding (T_m) and the onset of aggregation (T_{agg}) which is practical and allows for a ranking of stability of samples (Brader et al. 2015). Using these thermal stability readouts, the role of conformational and colloidal stability in the aggregation of GCSF was explored by Chi et al. (2003) in various solution conditions. They found that high T_m formulations showed no aggregation, whereas low T_m samples coincided with irreversible protein aggregation. Chi et al. (2003) also studied formulation pH and found that the levels of unfolded GCSF at different pH solutions were comparable, however aggregation only occurred in those of high pH (6 and 7). Aggregation at low pH only occurred with high salt (150 mM), which is explained by the decrease in B_{22} value and molecules increasing in attractive interactions. As such, both conformational and colloidal stability can affect GCSF aggregation, and either of the two can be rate limiting (Chi et al. 2003). Conformational and colloidal stability of GCSF was also studied by Robinson et al. (2017) who used 32 formulation designs, and a series of single, double, and triple GCSF mutants to investigate the relationship between T_m and T_{agg} and degradation kinetics. They found the use of the non-ionic surfactant, Tween 80, provided a significant increase in the colloidal stability of samples and increased T_{agg} when included; however, it had no real effect on degradation rates. T_m was found to be a

better predictor of GCSF aggregation at 37 °C storage, indicating the aggregation kinetics of GCSF at 37 °C were strongly linked to unfolding.

1.3.7 Altering GCSF stability

1.3.7.1 *Conformational rigidity*

Conformational dynamics has been previously shown to be linked with protein stability. A lattice model by Tang and Dill (1998) showed that protein dynamics change with varying temperature, and indicated that those with more stable structures had fewer fluctuations. In another study, two external perturbations (temperature and pH) were used to modulate the flexibility and stability of an IgG1 monoclonal antibody (mAb) in an attempt to better understand the possible correlations between flexibility and stability (Kamerzell and Middaugh 2008). The most dynamic mAb was seen at pH 4 and the most rigid at 6. The effect of pH appeared to couple the mAb dynamics to the solvent fluctuations, which controlled its dynamics and stability. Additionally, Spokova et al. (1998) used time resolved fluorescence spectroscopy, to measure changes in the rotational motions of specific fluorophores, as well as differences in the distribution of fluorescence lifetimes, as a function of pH in domain III of annexin V. They found increased fast rotational motions of the single T residue of Domain III and increased excited state heterogeneity as the pH was lowered.

Previous comparative studies have shown that proteins from thermophilic organisms have higher T_m values than their mesophilic organism counterparts (Razvi and Scholtz 2009). Conformational rigidity was also found in proteins from thermophilic organisms when compared to their mesophilic counterparts. Proteins from thermophilic organisms were found to be more compact and polar residues were scarce, whereas charged residues were abundant (Mamonova et al. 2013). From molecular dynamics (MD) simulations, Mamonova et al. (2013) found proteins from thermophilic organisms had shorter mobile loop regions than homologous proteins from mesophilic organisms. Mamonova et al. (2013) also noted significant increases in the number of the long-lived salt bridges in the thermophile proteins indicating that salt bridges may play an important role in the thermal stability. An increase in thermostability by altering amino acid residues towards those found in a thermophilic organism was demonstrated by Morris et al. (2016) who altered the cofactor-binding loops of transketolase (TK) towards those found in *Thermus thermophilus* at equivalent positions and improved both T_m and T_{agg} . A further study by Yu et al. (2017) on TK used a rigidifying flexible sites (RFS) strategy to rigidify flexible loops and found three

single-variants (I189H, A282P, D143K) more thermostable than wildtype (WT). These results suggest efforts to reduce the mobility of proteins, through rigidification, could significantly increase the stability.

1.4 Strategies to decrease protein flexibility

1.4.1 Excipients

Excipients are inactive substances that can be added to a solution to stabilise the protein structure and/or prevent aggregation. The excipient categories for use to stabilise proteins range from small molecular weight ions such as salts, and buffering agents to intermediate sized solutes such as amino acids and sugars, to larger molecular weight compounds such as polymers and other proteins (Kamerzell et al. 2011). With regards to small molecular weight ions, these excipients are commonly used to maintain solution pH and buffer-ion specific interactions with the protein. For the intermediate sized solutes, amino acids typically include histidine, arginine, glycine, proline, lysine or methionine, and are included to specifically interact with the protein and provide buffering and tonicity modifications. Sugars commonly used as excipients include sucrose, trehalose, sorbitol, mannitol, glucose and lactose, and are included as protein stabilisers against environmental stress as well as tonicity modifiers. The larger molecular weight compounds such as proteins and polymers include human serum albumin (HSA), gelatin, polyethylene glycol (PEG) and polyvinylpyrrolidone (PVP). The inclusion of protein and polymers as excipients provides competitive inhibitors of protein adsorption.

The intermediate sized solutes have been previously shown to increase ΔG either by: increasing the free energy of the unfolded state or decreasing the free energy of the native state. The sugars, sucrose and trehalose, are osmolytes, used in nature to stabilise microorganisms under harsh environmental conditions such as high temperatures and low water environments (Lentzen and Schwarz 2006). These disaccharides are thought to stabilise proteins by preferential exclusion at high concentrations in the liquid state (Kaushik and Bhat 2003; Kendrick et al. 1997).

1.4.1.1 *Formulation screening*

During protein formulation development, the effects of formulation variables on defined critical parameters are examined to optimise protein stability. The selection of excipients for protein stabilisation is currently performed empirically through screening studies of multiple excipients under a number of different conditions and concentrations. High-throughput screening through the use of robotics, automated instrumentation and

improved data capture and analysis has enhanced the number of experiments compared to manual experimentation (with a given amount of material over a given amount of time (Kamerzell et al. 2011)). However, formulation strategies are still often based on trial and error and/or formulator experience. The new challenge for protein formulation is to achieve the long term goal to rationally design stable formulations based on a thorough understanding of protein stability at the molecular level. The FDA encourages the implementation of quality by design (QbD) in the development of all pharmaceutical products, including generic drugs (Lawrence 2008). QbD is based on building quality into the final product by understanding and controlling formulation and manufacturing variables (Jiang et al. 2011B).

1.4.1.2 *In silico* formulation design

Advances in predicting *in vivo* performance of drug products has the potential to change how drug products are developed and reviewed. Modeling and simulation methods are now more commonly used in drug product development and aid scientists in designing rational screening strategies (Jiang et al. 2011B).

The *in silico* prediction of protein and excipient interactions is performed by a process called molecular docking. Molecular docking software can perform either rigid docking, where the conformation of the molecules is not altered during the docking process; or flexible docking, where the conformation of part or whole molecules change during fitting. There are two stages in a molecular docking protocol: sampling and scoring (Novikov et al. 2009). During sampling, a range of different orientations of a ligand (excipient) in respect to the binding surface of the target (protein), are generated. In the second step, these poses are scored and ranked according to a mathematical algorithm used to predict the binding affinities. Empirical scoring functions focus on binding free energy optimisation. These functions deconvolute the free energy between receptor and ligand (ΔG) into a sum of the weighted (ω_i) free binding energies for each component (electrostatic interactions, H-bonds) that contribute to the interaction (ΔG_i ; Barata et al. 2016). There are a number of software packages currently available for molecular docking including Dock (Ewing et al. 2001), AutoDock (Goodsell et al. 1996, iGEMDOCK (Yang et al. 2004), and GLUE (Goodford et al. 1985).

A recent study by Barata et al. (2016) developed the protein-excipient docking methods for GLUE and iGEMDOCK and compared results with previously published data by Shukla and

Trout (2011). The group studied the preferential sites of interaction between *Drosophila* Su(dx) protein (WW34) and arginine. The results obtained with the GLUE software package identified two preferential sites for arginine interaction, while iGEMDOCK solutions were all in the same site. However, the lowest energy solution for both software packages was found to be in the same site and was the same as previously reported. Barata et al. (2016) also correlated the energy of binding output from iGEMDOCK protein-excipient interactions with T_m with reasonable negative linear correlation (the lower the energy the higher the T_m). The group screened a number of common commercial excipients including amino acids, saccharides, polysorbates and sugar alcohols with Fab A33. Their experimental results showed that the success rate of iGEMDOCK was 78% (root-mean-square derivations below 2.0 angstrom) on 305 protein-compound complexes. Consequently, the data from the study highlighted the potential of computational docking of protein and excipients to reduce the number of formulation excipient candidates for *in vivo* screening, and to more importantly, rationally inform protein formulations, which is in alignment with QbD.

1.4.2 Solid state stabilisation

Even with the addition of excipients, a large number of chemical instabilities can occur in aqueous solution formulations because the presence of water promotes chemical and physical degradation (Crommelin 2013). Freeze drying, also known as lyophilisation, is widely used for pharmaceuticals to improve the stability and long-term storage stability of labile biotherapeutic proteins (Tang and Pikal 2004). Although capital costs are high for equipment and process development, there are reduced costs with lyophilised material due to the removal of cold chain storage, and lighter products costing less to transport. In addition, uncontrollable shipping and transportation stressors such as agitation, high and low temperatures and freezing pose less of an issue with solid state material. These advantages are directly seen in the increase in the number of products requiring lyophilisation where 50% of approved biopharmaceutical drugs on the list of the FDA and of the European Medicines Agency are freeze-dried (Constantino et al. 1998; Fissore et al. (2018).

The choice of drying method is dependent on the economics of the drying and the intended route of administration. Spray dried powders are commonly prepared with the intended route of administration via inhalation as control over particle size can be obtained (Abdul-Fattah et al. 2007), whereas parenteral administration is usually associated with freeze dried material and requires reconstitution prior to administration (Crommelin 2013).

1.4.2.1 Lyophilisation

Lyophilisation involves the removal of solvent (usually water) from a sample, by lowering the temperature and pulling a vacuum to a level where the product shows significantly increased stability. The product is dried without excessive heating therefore the risk of thermal degradation is minimised. Furthermore, storage vessels such as vials can be sealed under vacuum or under inert gas which protects against oxidative damage.

A typical freeze-drying process consists of three stages. The first is freezing, the second is primary drying where most of the ice water is sublimed, and the third is secondary drying where residual water is further removed until a solid state remains. A freezer dryer has three main components, a drying chamber where the samples are contained upon a temperature-controlled fluid-filled shelf, an ice condenser chamber containing refrigeration coils, and a two-stage vacuum pump.

1.4.2.1.1 Freezing

There are three states to describe the morphology of the frozen solids: crystalline, polycrystalline and amorphous (Figure 3). In the crystalline state, the substances are arranged in an ordered and repeating pattern, with the molecules regularly connected by specific interactions. By contrast, in the amorphous state, the molecules are stochastically arranged (Zhang 2017). When a solution consists of multiple crystalline solutes, they could form into a polycrystalline state if those solute molecules could not form a unified crystalline structure. Amorphous forms are typically favoured over crystalline solids for protein biopharmaceuticals, as they provide a higher dissolution rate and solubility, improved mechanical properties, and also better preservation of the tertiary structure of proteins (Yu 2001; Zhou et al. 2002).

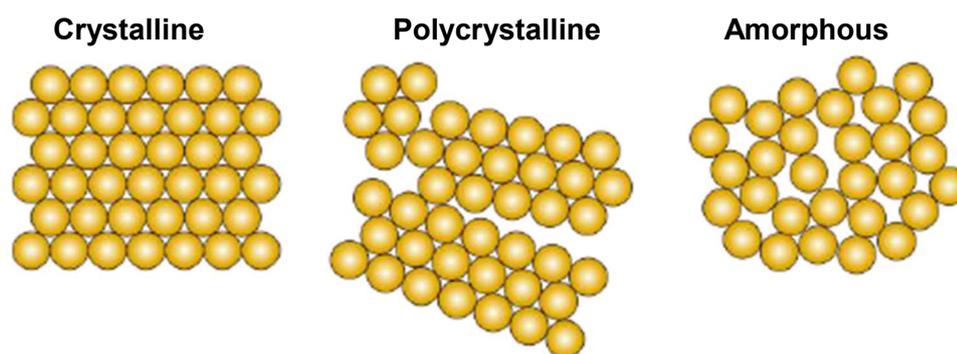


Figure 3. Structures of frozen material. Taken from Zhang (2017)

The speed and method of freezing is important because it influences the frozen matrix composition that in turn dictates the flow of vapour from the product during the drying steps. The larger, wider and more uniform the ice crystals are, the better the flow of vapour, and the smaller the crystals the more impedance on the vapour flow, and the higher moisture residence time, which in turn increases the drying time and the risk of cake collapse.

Annealing is often used during a freeze-drying process, where the temperature is briefly increased above the final freezing temperature, followed by cooling at a slow rate. This can transfer the material into a more stable structure by allowing the crystallisation of bulking agents such as mannitol or glycine. Annealing can also be used to encourage large ice crystal formation. In addition, a hold step at the final freeze temperature can be useful to ensure that the product is completely frozen.

1.4.2.1.2 Critical temperatures

There are a number of critical temperatures, which should be known when designing a freeze-drying process. The first is the glass transition temperature (T_g') and eutectic point (T_e), which are the characteristic temperatures for amorphous and crystalline states, respectively. At the end of freezing, the temperature should be well below the T_g' or T_e (Tang & Pikal 2004). T_g' and T_e can be determined using differential scanning calorimetry (DSC), where a phase change is measured by its change in heat capacity. The second is the microscopic collapse temperature (T_c). The T_c indicates the temperature above which an amorphous product loses its macroscopic structure (Tang & Pikal 2004) due to a decreased viscosity (Liu and Kuhlman 2006). It is typically 2 °C above the T_g' (Colandene 2007). For a mixture of amorphous and crystalline states, the T_c of microscopic collapse occurs between T_g' and T_e , while the T_c of macroscopic collapse is equivalent to T_e . The T_c of a formulation can be determined using freeze-drying microscopy (FDM), where freeze-drying conditions are simulated and observed using frame by frame image analysis such that collapse can be attributed to a certain temperature.

1.4.2.1.3 Primary drying

During primary drying, the material is dried by maintaining the temperature 2-5 °C below the T_g'/T_e so that it does not collapse or melt, while a vacuum is pulled until the pressure-temperature balance is such that the ice sublimates directly into vapour without melting (Abdul-Fattah et al. 2007). The sublimed water vapour migrates out of the drying-chamber

and is collected on the refrigerator coils within the condenser chamber. The process of sublimation leads to evaporative cooling of the product, lowering the product temperature and reducing the rate of sublimation. To avoid this and to prevent process slowing, heat is applied to the shelf to maintain a constant temperature, but this heat must not increase the temperature of the product above the T_c . During primary drying, heat enters the product via three mechanisms: by direct contact between the container base and shelf across the container and through the frozen mass to the drying front (conduction), by convection between product and gas molecules in the chamber and by radiation from vial to vial and from the chamber door. Drying of a product progresses from the surface to the base, meaning that the rate of sublimation decreases over time as the dry layer develops in thickness and the mass transfer resistance increases, decreasing the migration of water vapour. Oppositely, as the dry layer develops, the heat transfer resistance decreases, as the transport distance is conversely reduced by the retreating interface.

1.4.2.1.4 Secondary drying

During secondary drying, the shelf temperature is further increased so that residual water within the product can diffuse to the surface and evaporate. At the beginning of secondary drying, the residual moisture contained within the product is usually around 5-10% and is reduced to a unique optimum % identified during initial studies and product activity characterisation (Matejtschuk 2007). Conventionally, the content of residual water is expressed as a mass ratio relative to the product. Sample containers can be sealed when taking them out from the chamber after the process cycle, however, if sterile conditions are necessary, stoppers can be partially inserted into the containers when loading and then fully inserted post-lyophilisation by bringing down the hydraulic freeze-dryer shelf above them. This also avoids re-absorption of moisture from the atmosphere (Zhang 2017).

1.4.2.2 *Solid state formulation design*

During the freezing stage of lyophilisation different excipients can be included to non-specifically stabilise against freeze induced damage, also known as “cryoprotection”. Cryoprotectants stabilise the protein via the mechanism of preferential exclusion, which effectively increases the free energy barrier between the native and denatured state (Timasheff 1998).

Preservation of the protein's native structure is dependent on interactions between protein residues and the surrounding water, also known as its "hydration shell". Therefore, if the protein is dehydrated it will unfold. Excipients that protect a protein from dehydration are known as "lyoprotectants". Lyoprotection is thought to occur by one of two hypothesised mechanisms. The first is "water replacement", whereby the excipients H-Bond to the dried protein in place of the lost water, maintaining the folded structure. The second is "vitrification", whereby formation of an amorphous glass provides stability through reduced molecular motion and physical separation of protein molecules, reducing aggregation propensity (Ohtake et al. 2011). Both mechanisms agree on the stabilisation of protein in the solid state caused by decreased mobility. Water replacement was shown by Carpenter and Crowe (1989) using Fourier transform infrared spectroscopy (FTIR), where a band in the spectrum for lysozyme carboxylate groups H-Bonds to water was present in the spectrum when lyophilised in the presence of trehalose or sucrose. Prestelski et al. (1993) also demonstrated the replacement of water by sugars. By titrating sucrose with increasing amounts of protein they found a decreased in the amount of residual water following lyophilisation. This finding suggested that effects of sugars on proteins in the solid state are not due to the presence of increased amounts of water as the level of water in formulations dried with the sugars was as low as that for the protein lyophilised from just buffer or water. Reduced unfolding and aggregation is hypothesised to occur via the spatial separation between protein molecules within the amorphous glassy matrix (vitrification). A study by Suzuki et al. (1998) found a high degree of stabilisation of lactate dehydrogenase with sucrose in the amorphous state where H-bonds were formed with the lyophilised protein.

Whilst amorphous stages are required for cryo- and lyoprotection, if a product has a relatively low mass of protein, often it requires a bulking agent in the formulation to prevent the protein from being lost during drying and to form the product cake (Carpenter et al. 2002). Mannitol and glycine are examples of bulking agents, which can also serve as tonicity modifiers that usually crystallise during lyophilisation (Pikal 1990; Carpenter and Chang 1997).

1.4.2.3 *Ultra-scale down (USD) rational design of stable lyophilised protein formulations*

There are a number of other variables to screen for in a solid state formulation including pH, ionic strength, buffer type, excipients, and protein concentration. This leads to a large number of samples to prepare and analyse. As rapid formulation development has important financial ramifications (Carpenter et al. 2002), methods to reduce sample volumes and processing times are of great value. In 2009, Grant et al. (2009) demonstrated the use of an ultra-scale down (USD) 96-well microtitre plate, teamed with a factorial design of experiments (DoE), to successfully optimize a stable solid state formulation, which was later verified during scale-up to glass vials. The USD technique offered rapid process cycles due to the low sample volumes, which is of benefit during early-stage formulation development. A disadvantage to the USD method was the differing rate of sublimation for wells in different locations on the USD microtiter plate. It was found that USD wells on the outer edge of the plate had a smaller volume of water remaining compared to the inner USD wells due to thermal transfer via radiation from the sides of the dryer (Grant et al. 2009).

1.4.2.3.1 USD moisture analysis

Because water is often involved in protein degradation, its presence in the final product can be deleterious in preserving the potency and stability of product (Krasucka et al. 2012). Therefore, for implementation of USD methods, an accurate way of measuring moisture of cakes in wells is required. There are a variety of analyses that can be used for the determination of water content loss on drying, such as thermogravimetry, near-infrared spectroscopy, gas chromatography and Karl Fischer Titration, which have been discussed in literature (Krasucka et al. 2012), however most methods are non-applicable to USD lyophilised cakes due to the small amount of material available for analysis. In his thesis, Grant (2011) discussed the unsuccessful use of Karl Fischer titration to measure the moisture content of freeze-dried materials in USD wells due to difficulty in weighing out accurate quantities of material. In addition, he suggested the use of thermo-gravimetric analysis (TGA) for more effective analysis of well cakes due to the low material requirements.

Another USD moisture issue was documented by Robinson (2016) in his PhD thesis, where it was noted that during accelerated degradation storage, lyophilised cakes within

microwells shrank whereas those in vials did not. A study by Ullrich et al. (2015) on lyophilised amorphous cake shrinkage and cracking found an inverse correlation between cake shrinkage and cracking during freeze-drying. They studied different disaccharide excipients including trehalose, maltose and sucrose and found the degree of shrinkage was greater with disaccharides that have higher levels of unfrozen water at the point of maximum freeze concentration (w_g'). This observation was seen across all excipient concentrations and suspected to be caused by the loss of non-frozen water from the solid left behind after sublimation of the ice. The data from this study suggested that USD cake shrinkage could be due to differences in w_g' , however to confirm this, development of a method to measure USD lyophilised cake moisture is required.

1.4.2.4 *GCSF in the solid state*

There are only a small number of studies covering lyophilisation of GCSF. During development of the commercial formulation of Neupogen[®], Herman et al. (2002) found no improvement in shelf-life stability could be demonstrated by using a solid state formulation. They also found GCSF had sensitivity to freeze-thawing when co-formulated with mannitol. The substitution for sorbitol removed this sensitivity due to it freezing in an amorphous rather than crystalline state. Grant et al. (2012) screened seven different excipients including trehalose, HSA, mannitol, sucrose, arginine, Tween 20 and phenylalanine, and two buffers (30 mM histidine pH 7.0 and 50 mM sodium phosphate pH 7.0) in a DoE screen. They found HSA and Tween 20, significantly influenced the preservation of biological activity during the freeze drying. Gao et al. (2012) lyophilised GCSF within an undocumented buffer and 4% (w/v) trehalose and confirmed their freeze-dried material was stable by RP-HPLC. Finally, in a study to develop their 2nd International reference standard of GCSF, Wadhwa et al. (2011) described a lyophilisation formulation of GCSF with arginine, phenylalanine, trehalose, HSA and Tween 20 (NIBSC 2013). Results from GCSF lyophilisation studies suggest there is scope to study the protein in different formulations with varied survival effects.

1.4.3 Stabilisation by site-directed mutagenesis

When formulation fails to sufficiently stabilise, amino acid residues can also be substituted to modify the 3D structure to improve structural interactions and ΔG . The rational design and construction of novel proteins using structural information and manipulation of the protein sequence is known as protein engineering (Fersht and Serrano 1993).

1.4.3.1 *Increased structural rigidity*

Proteins have a number of different modes of motion and flexibility such as tumbling, breathing, side-chain rotation, and shear and hinge motions (Fersht 1999). In 2012, Bishop et al. (2012) demonstrated that replacing glycine residues with alanine residues in GCSF led to increased alpha helical propensity and packing, which in turn increased thermal stability. This result suggests mutations replacing flexible residues to improve interactions or packing is an effective strategy for increasing conformational stability.

The loop regions of proteins that generally connect the main secondary structures have been shown to be particularly flexible, and has led to a number of attempts to rigidify these regions for increased stability. Perchiacca and Tessier (2012) reported mutations within relatively large complementarity-determining loops (CDR; 5–20 residues) on the surface of variable antibody domains can alter the stability.

In 1987, Alber et al. (1987) generated temperature stable mutations of lysozyme gene of bacteriophage T4, where they found the altered amino acids all had low crystallographic thermal factors and low solvent accessibility side chains. This suggested X-ray data could be used to identify flexible regions in proteins for rigidification. B-Factors are values assigned to protein residues that are obtained from X-ray data, which indicate smearing of atomic electron densities with respect to their equilibrium position (Yu et al. 2017). B-factors have been previously used to target regions for engineering of a number of thermostable enzymes. For example, by modifying only sequence regions with the highest B-factors, increased thermostability was observed for a lipase (Reetz and Carballeira 2006) and enzyme CalB (Kim et al. 2010). As well as the use of B-factor values, a study by Zhang et al. (2018) used root-mean-square fluctuation (RMSF) measurement values from MD simulations to select conformationally flexible residues of A33 Fab for mutation. MD simulations are designed to mimic biochemical experimental environment of proteins in water solution at room temperature and the RMSF is the deviation over time between an atom or residue position and its reference position. The RMSF was shown to increase significantly in MD simulations at residues with increased flexibility (van der Kamp et al. 2010) and by using these values, Zhang et al. (2018) found local dynamics of the heavy chain of the C-terminus played a key role in A33 Fab aggregation.

1.4.3.2 Mutant design

1.4.3.2.1 Amino acid selection

Once a region for stabilisation has been identified, the selection of which residue to include for increased stability, whilst minimising detrimental effects on activity, causes new challenges. There are 19 alternative natural amino acid residues to select from during point mutation and their nomenclature is summarised in Figure 4. As amino acids often share common properties, several classifications have been proposed. A classification that explains mutation data through correlation with the physical, chemical and structural properties of amino acids was presented by Taylor (1986) based on the point accepted mutation (PAM) matrix by Dayhoff et al. (1978), where the major factor was size of the side chain, closely followed by hydrophobicity (Figure 5). As experimental mutagenesis of all possible amino acids at multiple sites across a protein would be incredibly laborious, *in vitro* evolution and *in silico* computational designs are currently the most popular approaches for selecting mutational candidates in a time and effort efficient manner.

G	Glycine	Gly	P	Proline	Pro
A	Alanine	Ala	V	Valine	Val
L	Leucine	Leu	I	Isoleucine	Ile
M	Methionine	Met	C	Cysteine	Cys
F	Phenylalanine	Phe	Y	Tyrosine	Tyr
W	Tryptophan	Trp	H	Histidine	His
K	Lysine	Lys	R	Arginine	Arg
Q	Glutamine	Gln	N	Asparagine	Asn
E	Glutamic Acid	Glu	D	Aspartic Acid	Asp
S	Serine	Ser	T	Threonine	Thr

Figure 4. Natural amino acid nomenclature. Amino acid nomenclature listed includes the single letter (first column), full name (second) and three letter name (third). Taken from DbBrowser (2018).

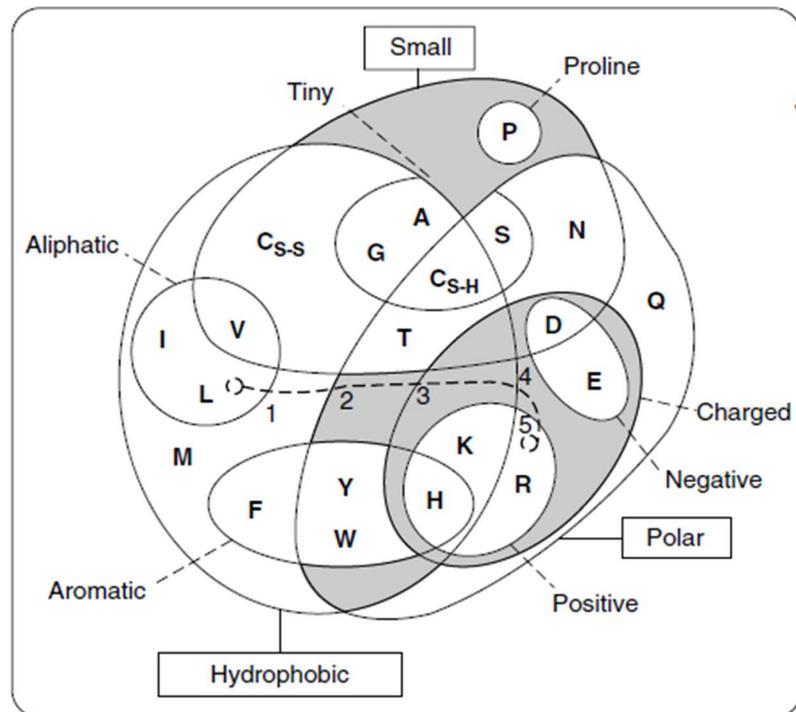


Figure 5. Venn diagram of amino acid characteristics. Amino acids are positioned based on multidimensional scaling of Dayhoff's mutation matrix, and grouped by common physico-chemical properties. Size is subcategorised into small and tiny. Affinity for water is described by polar and hydrophobic, which overlap, and charged, which is divided into positive and negative. Sets of aromatic and aliphatic amino acids are also marked. Taken from Betts and Russell (2003).

1.4.3.2.2 *In vitro* stabilisation

To stabilise a protein by *in vitro* evolution, libraries of variants are created by spontaneous mutations by error-prone PCR, cassette/saturation mutagenesis or DNA shuffling (Wunderlich et al. 2005). Stabilised mutants are identified from such libraries using selection techniques for stability and/or function. *In vitro* evolution can be laborious and time-consuming. It is also limited by library size and therefore only a small number of positions (around ten) can be randomised simultaneously (Wunderlich et al. 2005; Luo et al. 2002). If a large number of positions are required for engineering, there are a number of computational approaches for *in silico* high-throughput screening.

1.4.3.2.3 *In silico* stabilisation

An important subset of rational engineering methods consists of approaches that utilise high-resolution 3D structure information (Marshall et al. 2003). Advances in protein structure determination techniques have allowed the developments of modeling methods to predict the difference in Gibbs free energy, $\Delta\Delta G$, due to mutation. The Rosetta algorithm is the most successful current method (Rohl et al. 2004). A study by Kim et al. (2010)

combined B-factor experimental data with the *in silico* re-engineering program RosettaDesign, to identify and increase CalB thermostability. RosettaDesign is a freely accessible Rosetta online server. Backbone coordinates of the target structure are required in the form of a .pdb file, as well as specifics of which residues to re-engineer, also known as a resfile, in the form of a .txt file. The server returns the sequences, coordinates and energies of the engineered proteins. RosettaDesign uses Monte Carlo optimisation with simulated annealing to search for amino acids that pack well on the target structure and satisfy H-Bond potentials (Liu and Kuhlman 2006). Instead of selecting a small number of protein mutants, high-throughput sequencing using the Rosetta_ddg_monomer application allows the construction of millions of protein variants and their respective $\Delta\Delta G$ *in silico*, termed “deep mutational scanning” (Shin and Cho 2015). A recent comparison indicated that the Rosetta_ddg_monomer program generally provided more accurate results than other methods (Yu et al. 2017), where the use of a minimisation method involving limited backbone minimisation after repacking of all the side chains was found to generate the greatest experiment-prediction correlation coefficient of 0.69 (Kellog et al. 2011). Due to the large output dataset the downstream analysis for data interpretation is then processed into a mutational table or map (Shin and Cho 2015).

1.4.3.2.4 Immunogenicity

Along with the maintenance of activity, re-engineering of WT biotherapeutic proteins should also aim to preserve or reduce the immunogenicity of the molecule. Increased immunogenicity can lead to neutralisation and loss of efficacy, and/or elicit severe side effects. Consequently, assessment of the immunogenic response of final biotherapeutic candidates via techniques such as radioimmunoprecipitation assays (RIPA), enzyme-linked immunosorbent assays (ELISA) and neutralising antibody bioassays (Kessler et al. 2006), should be performed before progression to *in vivo* and clinical trials.

1.4.3.2.5 Engineering history of GCSF

GCSF lends itself well to mutagenesis studies due to its simple structure and low number of potential residues to mutate. Consequently, the engineering of GCSF has been explored for a number of different physiochemical traits including identification of residues involved with GCSFR interaction (Kuga et al. 1989), alpha helical folding propensity (Bishop et al. 2001), and reduced clearance from the body via increased receptor binding affinity (Sarkar et al. 2002). *In vitro* evolution of GCSF was performed by Buchanan et al. (2012) using error-prone PCR. The final mutant (C17G, W58R, Q70R, and F83L) had improved soluble

expression with a thousand fold increase in yield to WT as screened by incubation with DTT followed by an activity assay. Computational redesign of GCSF was performed by Luo et al. (2002) using Protein Design Automation (PDA) to score the fit of sequences to a 3D GCSF structure using physical-chemical potential functions. A combination of 10 mutations (C17L, G28A, L78F, Y85F, L103V, V110I, F113L, V151I, V153I, and L168F) improved T_m by 13 °C, relative to WT. The group achieved this by targeting the core residues of GCSF without risking change to GCSFR binding sites on the surface of the molecule.

1.4.3.2.6 Available GCSF 3D structures

In the computational redesign of GCSF by Luo et al. (2002), the 3D structure input was designed using homology modeling of a human GCSF sequence mapped onto the bovine GCSF crystal structure (1BGC; Lovejoy et al. 1993), because the human GCSF structure had not been solved. In the past 16 years since this work was published, there are now multiple different high-resolution crystal structures available for recombinant, human GCSF. As a result, there is great potential for the development of new GCSF stability mutants using more accurate 3D structures of the protein, which in combination with more high-throughput computational programs could encompass the entire protein. Additionally, the homodimer interaction of GCSF with GCSFR was reported by Tamada et al. 2012, allowing for the exact residues involved with GCSFR interaction to be filtered out of selection and ensure mutations will not affect the efficacy of the molecule.

1.5 Biophysical analysis of protein stability

As established at the beginning of this introduction, protein stability is inherently linked to its mobile 3D structure provided by constant breaking and formation of bonds. To study this phenomenon, biophysical analysis is required. Biophysics is an interdisciplinary science that uses the methods of physics to study biological systems. There are currently a wide range of biophysical analytical techniques to measure global and local stability of proteins ranging from standard to advanced techniques (Figure 6).

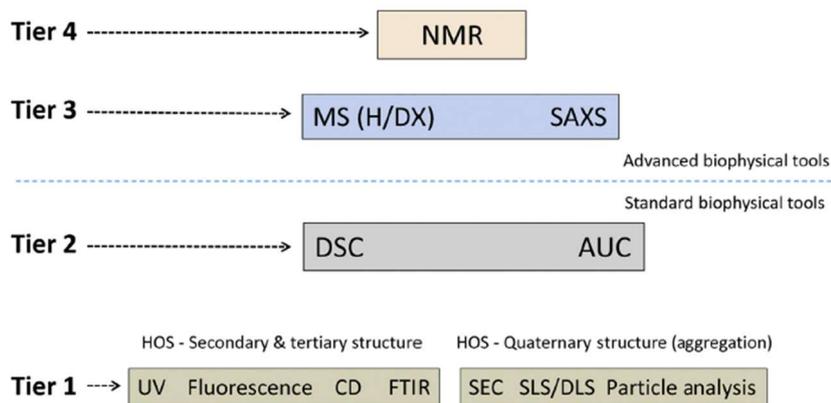


Figure 6. **Hierarchy of protein biophysical analysis techniques.** Analysis techniques are arranged into four tiers depending on their resolution, with Tier 4 having the highest resolution. Tiers are also split into two groups based on their standard or advanced use in Industry. Technique abbreviations are as follows: nuclear magnetic resonance (NMR), hydrogen-deuterium exchange with mass spectrometry MS (H/DX), small angle X-ray scattering (SAXS), differential scanning calorimetry (DSC), analytical ultracentrifugation (AUC), higher order structure (HOS), circular dichroism (CD), Fourier transform infrared spectroscopy (FTIR), size-exclusion chromatography (SEC), static light scattering/dynamic light scattering (SLS/DLS). Taken from Houde and Berkowitz (2015A).

1.5.1 Standard biophysical analysis

The standard biophysical analysis techniques are comprised of Tiers 1 and 2 (Figure 6). The term “standard” refers to the straightforward, relatively inexpensive and low resolution of these techniques. Consequently, these techniques also require the least amount of time to acquire data and the least amount of user expertise. As previously discussed with the measurements of T_m and T_{agg} , they generally provide a global snapshot measurement that characterises the overall secondary and tertiary structure of proteins. Tier 2 contains differential scanning calorimetry (DSC) and analytical ultracentrifugation (AUC). Tier 1 is additionally split into the measurement of either secondary and tertiary structure, or quaternary structure. For the measurement of secondary and tertiary structure the techniques include UV, fluorescence, circular dichroism (CD), and Fourier transform infrared spectroscopy (FTIR). For the measurement of quaternary structure, the techniques include size-exclusion chromatography (SEC), static light scattering/dynamic light scattering (SLS/DLS) particle analysis. Whilst low resolution, these techniques that measure quaternary structure are incredibly important for the characterisation of high molecular weight aggregates.

1.5.2 Advanced biophysical analysis

The advanced biophysical analysis techniques are comprised of Tiers 3 and 4 (Figure 6). Techniques in the advanced group are those that are used currently in a limited capacity, which is a reflection of their high resolution requiring relatively expensive machinery and a high level of user expertise (Houde and Berkowitz 2015A). Tier 3 contains hydrogen-deuterium exchange with mass spectrometry (HDX-MS) and small angle X-ray scattering (SAXS), and Tier 4 contains nuclear magnetic resonance (NMR) spectroscopy.

1.5.2.1 *NMR*

Nuclear magnetic resonance (NMR) spectroscopy is often the method of choice for investigating protein motions and dynamic processes due to its superior detail and resolution at the single amino acid level. The use of NMR to characterise the dynamics of proteins was reviewed in detail by Palmer III (2004). NMR spectroscopy enables measurements of many different dynamic processes from rapid bond vibrations to slow conformational transitions (Henzler-Wildman and Kern 2007). Additionally, with the development of multi-dimensional NMR, various new pulse sequences and relaxation techniques has allowed access to protein dynamic processes that occur on timescales of picoseconds to seconds (Henzler-Wildman and Kern 2007). However, this increased level of resolution at shorter time scales comes with major technical disadvantages. NMR has issues with high molecular weight samples, where experiments on proteins larger than 40 kDa are difficult (Konermann et al. 2011). NMR typically requires high sample concentrations, which can be an issue with limited sample availability and stability. NMR has also shown issues in measuring protein aggregation and insolubility (Houde and Berkowitz 2015A). Finally, a major challenge for NMR analysis is the complex observation and assignment of hundreds of peaks across multiple samples (Coales et al. 2010).

1.5.2.2 *HDX-MS*

In recent years, the study of protein dynamics has largely shifted toward the use of hydrogen/deuterium exchange mass spectrometry (HDX-MS). HDX-MS can provide both global and local structural information on proteins in liquid and solid states. Advantages of HDX-MS over NMR include its greater sensitivity, requiring lower sample concentrations, virtually unlimited protein size range, and the capability to detect co-existing protein conformers (Konermann et al. 2011).

1.5.2.2.1 HDX

The use of HDX to study protein dynamics was first demonstrated by Linderstrøm-Lang (1958). By using density gradient tubes, they found hydrogen atoms in OH, NH, and SH groups readily exchanged for deuterium when exposed to a deuterated buffer and increased the mass of the protein by ~ 1 Da per hour. Every amino acid residue with the exception of proline and the first amino acid in the chain have an amide NH group. The backbone amide hydrogens involved in weak H-Bonds or located at the surface of the protein can exchange rapidly with deuterium, whereas those buried in the interior or those involved in stabilising H-bonds exchange more slowly (Wei et al. 2014). The exchange of amide hydrogens, can therefore provide information on protein flexibility, conformational distribution, H-Bond patterns, and structure (Zhang et al. 2015).

During HDX, amide hydrogens exchange with solvent hydrogens through acid-, base- or water-catalysed reactions (Coales et al. 2010):

$$k_{ch} = k_H[H^+] + k_{OH}[OH^-] + k_{water} \quad \text{Equation 7}$$

where the acid-catalysed reaction $k_H[H^+]$ is faster at low pH (< 2.5) and the base-catalysed reaction $k_{OH}[OH^-]$ increases at higher pH (> 2.5). Rate constant values were previously derived by Bai et al. (1993), using NMR to study polyalanine peptide exchange rates in labelling solutions of varying pD. The amide hydrogens of proteins in the native, folded state exchange according to the equations:



$$k_{ex} = k_{op} * k_{ch} / (k_{cl} + k_{ch}) \quad \text{Equation 9}$$

where k_{op} is the rate at which amide hydrogen converts from the closed state into the open state. Conversely, k_{cl} is the rate amide hydrogen converts from the open state into the closed state. For most proteins at or below neutral pH, amide HDX occurs by an EX2 mechanism, where $k_{cl} \gg k_{ch}$ and equation 9 becomes:

$$k_{ex} = k_{op} * \frac{k_{ch}}{k_{cl}} = k_{ch} * K_{op} \quad \text{Equation 10}$$

where the ratio of the measured HDX rate in the folded protein, k_{ex} , and the calculated intrinsic rate (k_{ch}) yields the K_{op} . K_{op} values have some correlation with the extent of

secondary/tertiary structure and resulting dynamic characteristics around the corresponding amide hydrogen.

1.5.2.2.2 HDX coupled with mass spectrometry

HDX is a temperature and pH dependent reaction. Above pD 4, k_{ch} increases by one order of magnitude with each pD unit, reaching values on the order of $10^3 s^{-1}$ at pD 9.34 (Konermann et al. 2011). Because of this, HDX can be quenched by placing the labelled protein in a low pH, low temperature solution. The quenching of labelled proteins, combined with the development of electrospray ionization (ESI), a process for the transfer of biomolecular analytes into the gas phase as intact ions, coupled with reverse-phase liquid chromatography (LC) facilitated the rapid improvements in HDX-MS methodology to the automated systems currently in use (Konermann et al. 2011).

1.5.2.2.3 HDX-MS workflow

For HDX-MS, the first step is the labelling of the protein. Most HDX-MS strategies employ a continuous labelling strategy where exposure to deuterium is monitored as a function of exposure time (Figure 7). This exposure time can range from seconds to hours followed by a quench step. Consequently, measurement of exchange over time is a measure of structural dynamics rather than actual structure (Konermann et al. 2011). Quenching the HDX reaction facilitates local conformational analysis; by using a protease that can retain catalytic activity at pH 2.5, HDX can be localised to a short peptide/ single amino-acid level (Kaltashov et al. 2013). The acidic protease most commonly used in HDX-MS experiments for digestion is pepsin, because it is acid stable and digests non-specifically and reproducibly. Reproducibility is vital as the peptides produced must be the same each time the sample is run, both with deuterated and non-deuterated experiments. A study by Ahn et al. (2013) found aspartic protease pepsin digestion was the least specific and the most reproducible under the same digestion conditions. Additionally, Ahn et al. (2012) developed an on-line pepsin column for HDX-MS experiments. They immobilised pepsin on ethyl-bridged hybrid (BEH) particles and showed that the pepsin column could withstand continuous high-pressure at 10,000 psi. The implementation of the column into the HDX-MS workflow increased pepsin digestion redundancy through increased overlapping peptides, helping to refine the location of deuterium in peptide-level HDX. Redundancy is obtained through overlapping peptides, which can help to refine the location of deuterium down to the amino acid level (Wei et al. 2014).

The addition of reducing and/or denaturing agents in the quench solution can also improve the efficiency of digestion. Following digestion, peptides are separated using a fast but efficient, chromatographic separation at a low temperature around 0 °C to control the deuterium back exchange (Figure 7; Wales et al. 2008).

Measurements of HDX can be performed at the intact protein level to reveal the overall deuterium incorporation or at the peptide level to reveal localised exchange information (Wei et al. 2014). Global analysis measures deuterium uptake on the intact protein and is often performed as an initial step to ensure changes in protein structure occur in response to an effector can be monitored. Local analysis, at the peptide level, usually follows to determine specific sequences that exchange more readily than others (Waters 2018). Non-deuterated control experiments are also performed to identify the non-deuterated mass of all the peptic peptides produced during enzymatic digestion. Relative uptake HDX levels of individual peptides are determined by ESI-MS as a function of labelling time (t) according to the equation:

$$\text{deuteration level } (t) = m(t) - m_0 \quad \text{Equation 11}$$

where m(t) is the centroid mass of the peptide of interest, and m0 is the centroid mass of the non-labelled peptide.

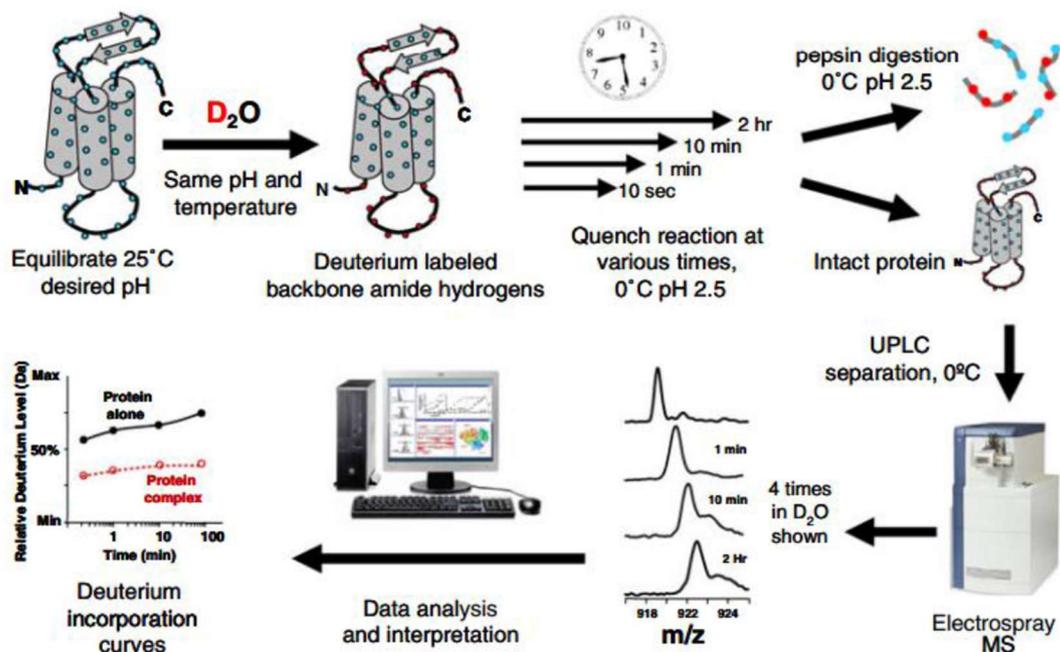


Figure 7. HDX-MS process flow diagram. Taken from Lohitha (2017)

1.5.2.2.4 Commercialised HDX-MS setup

Historically, a HDX-MS lab set up was made up from a combination of individual parts from different manufacturers (Bou-Assaf and Marshall 2015) or custom built. A complete HDX-MS system containing auxiliary (ASM) and binary solvent managers (BSM) coupled to a temperature controlled HPLC column storage unit (HDX manager) and Synapt G2-S mass spectrometer is now commercially sold by Waters Corporation (Milford, MA). Waters collaborated with Leap Technologies (Carrboro, NC) to also automate sample handling using a Leap PAL™ robot controlled by HDX Director Software. The schematic for the Waters automated HDX-MS system is displayed in Figure 8. The LEAP HDX automation manager is made up of two dual heads (PrepPAL and InjPAL) which perform the sample mixing and incubation with D₂O (or non-deuterated control) and quench solutions, and also places samples into the HDX manager injection loop. After the sample loop is loaded, the samples either bypasses the pepsin column for global analysis or enters the pepsin column for peptide-level analysis. For pepsin digestion, the ASM supplies eluent to carry the sample to the digestion column (Waters 2018). The HDX manager, as seen in Figure 8, contains two temperature controlled chambers, which are critical for HDX experiments. The first is the sample chamber, which contains the injection valve, trap valve, analytical column and trapping column. The temperature of the sample chamber is controlled at 0 °C to minimise back exchange. The second is the digestion chamber, which contains the online BEH pepsin column for protein digestion. The digestion column sits in-line between the injection valve and the trap valve and the chamber temperature is usually optimised for different protein digestions to obtain optimal peptide coverage and redundancy. Following digestion, eluted peptide fragments flow on to a VanGuard™ trapping column, which washes unwanted solutes to waste. Trapped peptides are then eluted from the trapping column onto the analytical reverse phase HPLC column (RP-HPLC). A gradient elution from the μBSM elutes the peptides from the trap to the analytical column and from the analytical column into the mass spectrometer. Finally, MS data is acquired and analysed by both ProteinLynX Global SERVER™ (PLGS) and DynamX™ software.

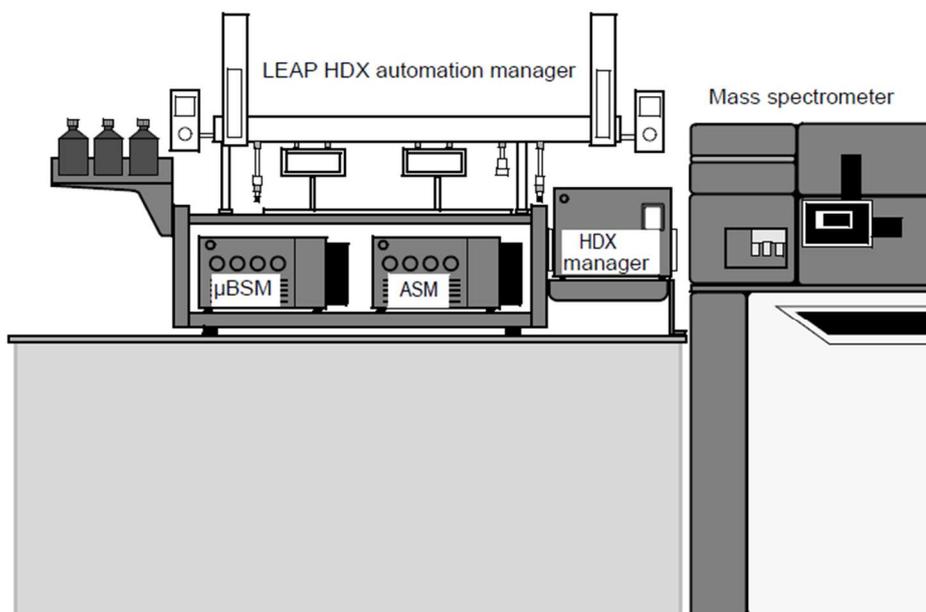


Figure 8. Schematic of the Waters automated HDX-MS system. Taken from Waters (2018)

1.5.2.2.5 HDX-MS and G-CSF

GCSF has been studied in a number of HDX-MS experiments. In 2012, Wei et al (2012), studied the conformational changes between GCSF and PEG-GCSF where they found no significant differences in deuterium uptake and no gross conformational rearrangement upon pegylation. Following this, Mo et al. (2013) used HDX-MS to characterise the disulphide linkage and scrambling within GCSF. A study by Tsuchida et al. (2014) looked at the effects on GCSF stability upon exposure to three different stressors: heating, photoirradiation and oxidation. They found no significant difference in structure for heating and photo-irradiation; however reduced activity in the oxidised sample was confirmed by HDX-MS conformational changes around the GCSFR binding site and C-terminal region. Additionally, GCSF structural stability was studied by Zhang et al. (2015) in the presence of sucrose and benzyl alcohol. It was found that sucrose globally protected GCSF from deuterium uptake, whereas benzyl alcohol increased deuterium uptake in α -helical bundle regions. The work by Zhang et al. (2015) demonstrated the use of HDX-MS to provide mechanistic insights into stabilisation through formulation developments, as well as identification of areas for stability re-engineering.

1.5.2.3 *Biophysical analysis of protein-excipient interactions in the solid state*

During lyophilisation proteins can be dried in conformations susceptible to degradation during long-term storage; therefore the characterisation of proteins within the solid state is highly valuable. There are a number of different techniques to measure protein

conformational changes in the solid state, as reviewed by Moorthy et al. (2015), including FTIR, Raman spectroscopy and near infrared spectroscopy (NIR), as well as fluorescence spectroscopy to monitor tertiary structural changes. In addition, solid-state nuclear magnetic resonance (ssNMR) and neutron scattering are commonly employed to study protein dynamics in solid matrices and DSC can be used to characterize protein stability via determination of T_m . The majority of these techniques have multiple disadvantages as described for NMR earlier, such as complexity of sample preparation, lack of commercially available equipment, high levels of user expertise and/or level of resolution.

1.5.2.3.1 Solid-state hydrogen-deuterium exchange mass spectrometry (ssHDX-MS)

Solid-state hydrogen-deuterium exchange mass spectrometry (ssHDX-MS) is a relatively new technique, adapted by Professor Liz Topp's lab (Moorthy et al. 2015), capable of mapping protein structure and conformations in lyophilised solids. Exposure of the protein to deuterium occurs in the vapor phase via the use of desiccants within a sealed glass desiccator (Figure 9) allowing for labelling of the molecule based on both protein structure and solid state matrix composition. As with solution HDX-MS, the inclusion of reconstitution, quench and an optional pepsin digestion steps, followed by LC-MS, allows uptake to be measured at both a global and peptide level. A number of recent studies have compared ssHDX-MS data with FTIR and found solids providing protection from exchange by ssHDX-MS also show retention of structure by FTIR and, conversely, that solids with loss of secondary structure by FTIR, also show reduced protection from exchange (Li et al. 2007; Li et al. 2008; Sinha et al. 2008; Sophocleous et al. 2012). These results validated ssHDX-MS as an alternative biophysical technique to measure different protein conformations within the solid state that vary with differing formulations.



Figure 9. Schematic of ssHDX-MS sample preparation and analysis. Taken from Moorthy et al. (2015).

1.5.2.3.2 Differences between solution and solid state HDX

The uptake of deuterium by protein amide groups in lyophilised solids differs from that in solution in several respects. Firstly, the concentration of D₂O in amorphous solids is less than in than aqueous solutions and varies with relative humidity (RH), this can slow the rate of uptake to days rather than the hours required for solution HDX-MS labelling. Secondly, water replacement occurs in the solid state rather than preferential exclusion, meaning exposure to deuterium can occur through the local interaction of the amide group with sorbed D₂O or through conduction of deuterium through H-Bond networks in the solid. Due to the fixed state, some dynamic modes available in solution and seen in HDX-MS are not attainable and the rates of opening events and exchange may be altered. Finally, there is a risk that D₂O sorption and diffusion process may affect the rate of ssHDX, however, a study by Sophocleous et al. (2012) found moisture sorption in ssHDX was complete in a period of hours for sucrose and/or mannitol containing mAb formulations, and had minimal contribution to exchange kinetics beyond this time. This suggested the rate and extent of exchange is not a measure of D₂O adsorption (Moorthy et al. 2015)..

1.5.2.3.3 Solid state formulation predictions

As with aqueous state formulations, the ability to select stabilising excipients in a short amount of time is critical for reducing development timelines. Attempts to predict and understand protein-excipient interactions in the solid state *in silico* is still in early phase development within academic groups and mostly unpublished. In 2011, Roughton et al. (2011; unpublished) used computational molecular design (CMD) to study protein-excipient interactions and compared results to experimentally derived data from ssHDX-MS studies. Several proteins were included (calmodulin, lysozyme, myoglobin, and b-lactoglobulin) along with different excipients (mannitol, sucrose, trehalose, and raffinose), where during a blind docking experiment, the excipients were allowed to be flexible while the amino acid residues were held rigid. The amino acids that interacted with an excipient's docked conformation were recorded for fifty total docked conformations for each protein-excipient pair. It was found that regions of reduced exchange in HDX data matched well with regions of frequent protein-excipient interaction. Similarly, in a study by Tarar (2012; unpublished), Autodock was used to compare myoglobin-excipient docking and ssHD-MS data, also found regions which were protected by the excipients in the solid state corresponded to residue-excipient pairs showing high frequency of interactions. This preliminary data shows docking methods as described for proteins in solution (Section 1.4.1.2) has potential to be routinely applied to protein formulations in the solid state.

1.6 Aims and objectives of thesis

The overall aim of this thesis was to use both *in silico* and biophysical techniques, separately or in combination, to predict GCSF stability. This aim was inherently open ended, as stability can be measured in a number of ways in a number of states. As depicted in (Figure 10), result chapters were separated by the three major strategies of biopharmaceutical stabilisation: site-directed mutagenesis, and formulation by excipients in the aqueous and solid states.

Chapter 3 covered the production of GCSF, initial testing and development of biophysical analysis and *in silico* structure selection. Results laid the groundwork for the rest of the thesis.

Chapter 4 took HDX-MS and B-factor data for GCSF in its native, folded state, obtained in chapter 3, and designed stability mutants within regions with increased relative flexibility. The design of mutants was performed using the computational prediction application Rosetta, using the change in folding energy upon point mutation, $\Delta\Delta G$, to rank candidates. The stability mutants were subjected to different characterisation techniques to determine the effect of mutation on GCSF purity, bioactivity and thermal stability.

Chapter 5 continued the work from chapter 4. Mutants with a significant change in thermal stability were analysed for their shelf-life stability using accelerated thermal degradation. Additionally, the impact of mutation on local flexibility relative to WT was assessed by peptide level HDX-MS and correlations made between flexibility and thermal degradation.

Chapter 6 assessed stabilisation by commonly used excipients in aqueous formulation. Protein-excipient interactions were predicted computationally using the docking program iGEMDOCK. Deuterated excipient HDX-MS was performed with the inclusion of an internal reference peptide to monitor changes in intrinsic exchange. Sites with changes in uptake relative to a GCSF without excipients were identified and compared with the docking data to provide insights into the mechanism of excipient stabilisation.

Chapter 7 applied the learnings from the study of GCSF stabilisation in the aqueous state in Chapter 6 to study stabilisation in the solid state. Suitable solid state formulations were screened using previously developed ultra-scale-down (USD) methods. Excipient studied during aqueous HDX-MS in result chapter 4 were included in solid state samples and analysed using a HDX-MS technique developed by the Topp lab at Purdue University, USA (Moorthy et al. 2015), employing deuterium labelling in the vapour phase. Sites with

changes in relative uptake were identified and compared with docking data to again understand the mechanism of stabilisation.

The use of *in silico* and biophysical methods for measuring and predicting GCSF stability were employed throughout all chapters and the rigidification of the higher order structure ultimately linked the results together.

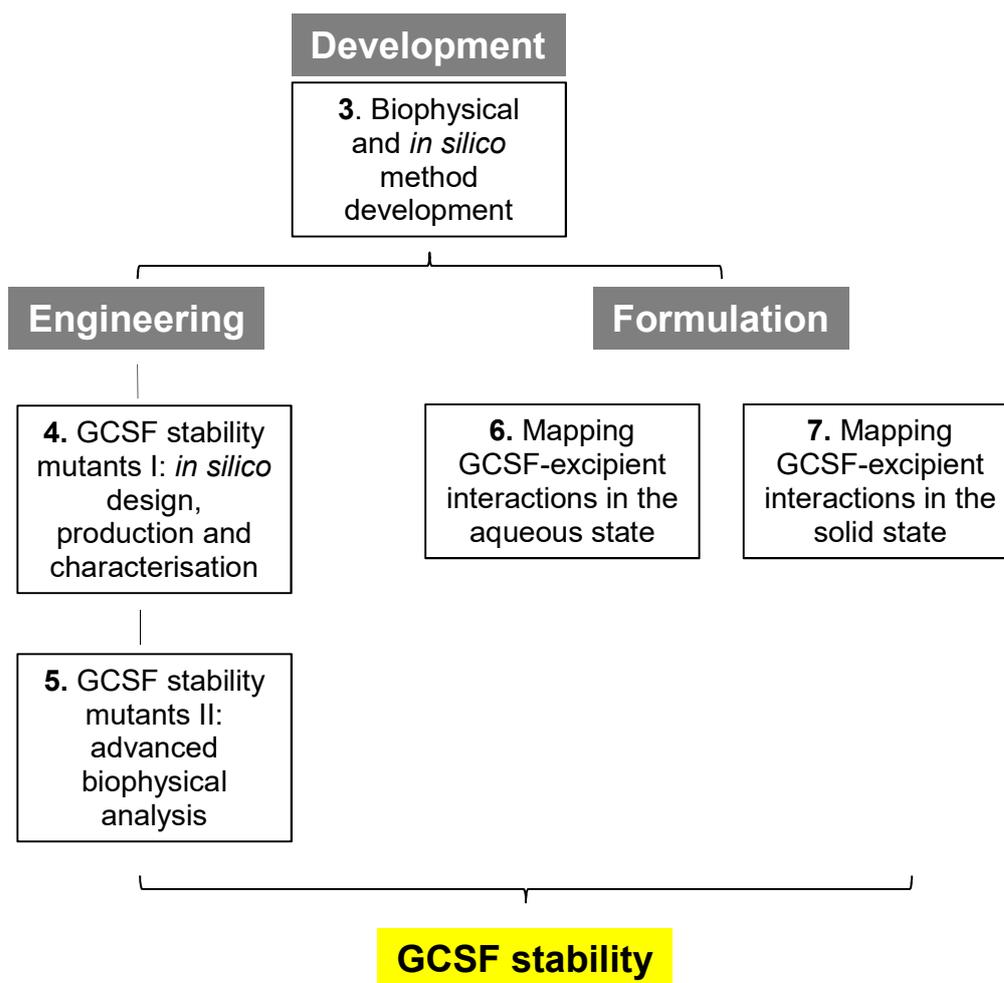


Figure 10. **Result chapter schematic.** The five results chapters (3 to 7) were split into three main categories. Chapter 1 included the development of protein production, biophysical characterisation and methodology. Chapters 2 and 3 were linked in the design, production and characterisation of stability mutants. Finally chapters 4 and 5 were linked in the exploration of excipients to stabilise the protein structure in aqueous and solid states, respectively. All five chapters employed *in silico* and biophysical measurements to further the understanding of the driving forces for GCSF stability.

2 Materials and methods

All chemicals were from Sigma–Aldrich (Sigma-Aldrich Company Ltd, Dorset, UK) unless otherwise noted. Purite distilled water ($18\text{ M}\Omega\text{ cm}^{-1}$) was used in all methods and all sterile filtration was performed using Millex-GP 0.2 μM , 33 mm, polyethersulfone (PES) sterile syringe filters (Millipore, Hertfordshire, UK).

2.1 Molecular biology

The BL21 (DE3) strain of *Escherichia coli* containing the human granulocyte colony stimulating factor (GCSF; Filgrastim) gene on a pET21a plasmid was provided by Dr Adrian Bristow from the National Institute for Biological Standards and Control (NIBSC; Bristow et al. 2012).

2.1.1 Plasmid sequencing

The pET21a plasmid was extracted from cells grown in a 5 mL culture of BL21 (DE3) *E. coli* cells overnight at 250 rpm, 37 °C, in Luria Bertani (LB) media containing 1 mM ampicillin (Amp), using a QiaSpin® Miniprep kit (Qiagen, Hilden, Germany) as per manufacturer’s instructions. The final elution step was changed from 50 μL to 30 μL to ensure a high concentration sample of DNA was obtained for sequencing. The concentration of eluted DNA was measured using a Nanodrop (Thermo Fisher Scientific Inc., Wilmington, USA) UV/Visible spectrophotometer, based on A_{260} values. The wild type (WT) GCSF gene within the extracted plasmid was sequenced by Source Bioscience (Source BioScience UK Limited, Nottingham, UK) using the standard T7 promoter and terminator primers. The returned forward and reverse sequences were aligned to obtain the final sequence.

2.1.2 Polymerase chain reaction (PCR) based site-directed mutagenesis

Manipulation of the GCSF plasmid sequence was performed using polymerase chain reaction (PCR) based site-directed mutagenesis in accordance with the QuikChange Lightning site-directed mutagenesis kit instruction manual (Agilent Technologies, Inc., Santa Clara, CA, USA). PCR was performed using a C1000 Touch™ Thermal Cycler (Bio-Rad, Hercules, CA, USA). Primers were designed using the freely available primer design software recommended by the mutagenesis kit and synthesised by Eurofins Genomics services (Wolverhampton, United Kingdom). A table of mutant designed primers is shown in Appendix 6. For each PCR reaction 125 ng of oligonucleotide primers were added to 50 ng of WT GCSF plasmid template DNA. The extension time for the control and sample reaction was 2.5 mins and 3 mins, respectively. Post-PCR amplification the non-mutated

parental DNA was digested with Dpn1 endonuclease provided with the site-directed mutagenesis kit.

2.1.3 DNA gel electrophoresis

To confirm plasmid amplification, a restriction digest was set up using EcoRI-HF and 1X NEBuffer (New England BioLabs Inc, Ipswich, US). Samples were incubated at 37 °C for 1 hr prior to mixing with 1X loading dye (New England BioLabs Inc, Ipswich, US). Samples were loaded alongside a 1 kb DNA ladder (New England BioLabs Inc, Ipswich, US) into the wells of a 1% (w/v) agarose gel with 1X Invitrogen SYBR Safe staining reagent (Thermo Fisher Scientific Inc., Wilmington, USA). Electrophoresis was carried out at a constant voltage of 80 V for 1 hr in 1X TAE (Tris, Acetic Acid and EDTA) running buffer. Bands were visualised using a Geldoc 2000 (Bio-Rad, Hercules, CA, USA) with UV light, where a single band at approximately 5.9 Kbp confirmed the success of the PCR reaction.

2.1.4 Transformation

Following PCR, 2 µL of each plasmid sample was used to transform 45 µL of XL10-Gold ultracompetent cells (Agilent Technologies, Inc., Santa Clara, CA, USA) in accordance with the instruction manual, using the pUC18 plasmid alongside as a control. Due to its compositional similarity with NZY⁺ broth, super optimal broth with catabolite repression (SOC) medium (Bioline Reagents Ltd, London, UK) was used instead to aid recovery of the competent cells. Transformed cells were plated on LB /Amp agar plates containing 80 µg/ml X-gal and 20 mM IPTG (Generon Ltd, Maidenhead, UK) spread over the surface. Additionally, 2.5 µL of cells from transformation control (pUC18) reaction and 10 µL of pWhitescript mutagenesis control were also plated. The plates were incubated at 37 °C overnight and the number of colonies counted.

2.1.5 Mutant plasmid sequencing

As described in section 2.1.1, cells were cultured and plasmids were extracted. The resulting DNA was sent to either Source BioScience (Source BioScience UK Limited, Nottingham, UK) or Eurofins (Wolverhampton, United Kingdom) for sequencing using standard T7 forward and reverse primers. Forward and reverse nucleotide sequences were aligned and the confirmed DNA sequence translated into amino acid sequence for comparisons to WT.

2.1.6 Glycerol stocks

E. coli BL21(DE3) competent cells (New England BioLabs Inc, Ipswich, US) were transformed according to the instruction manual using 250 ng of mutant plasmid DNA. Several 10-fold serial dilutions were completed with SOC medium followed by overnight incubation at 37 °C on LB/amp agar plates. A single colony was picked and grown overnight in 10 mL of LB/amp at 37 °C, 250 rpm. These cells were added to a 50% (v/v) sterile filtered glycerol solution, at a ratio of 1:1 bacterial cells to glycerol solution, and stored at -80 °C.

2.2 GCSF production

2.2.1 Large scale cell culture

All media was autoclaved for 20 mins at 120 °C, and all additions post autoclaving were filter sterilised. An initial seed culture was generated by culturing a glycerol stock in a 50 mL falcon tube containing 10 mL Terrific Broth (TB)/ Amp. The falcon tube was incubated overnight with shaking at 37 °C, 250 rpm. The 10 mL culture was seeded into sterile 2 L baffled flasks containing 500 mL TB/Amp for 3 h before sterile transfer into a 7.5 L bioreactor (New Brunswick, NJ, USA) containing 5 L of TB. A stock solution of magnesium sulphate heptahydrate/Amp solution was added to the bioreactor through the sterile injection port to yield a final concentration of 1 mM. Dissolved oxygen was controlled at 30% (v/v) via agitation up to 600 rpm followed by maintenance via an oxygen gas cylinder. Temperature was controlled at 37 °C and the pH at 7 using phosphoric acid and ammonium hydroxide. Expression was induced twice, once at mid-exponential and a second at stationary phase growth ($OD_{600} = 10$ and 35 , respectively), by spiking in a 1 M IPTG, sterile filtered, solution through the sterile injection port to a final concentration of 1 mM.

2.2.2 Primary separations

Cells were harvested 3.5 h post-induction by centrifugation at 7080 x g, 20 mins, and 4 °C (Avanti J-20 XPI; Beckman Coulter, Inc., Fullerton, CA, USA). The cell pellets were washed in 10 mM phosphate buffered saline (PBS; Severn Biotech Ltd, Kidderminster, UK) and centrifuged into small 4 g (WCW) pellets within 50 mL falcon tubes at 7,728 x g, for 30 mins, 4 °C. The pellets were stored at -20 °C for storage purposes and to aid with cell lysis. Each cell pellet was defrosted for 30 min at room temperature (RT) followed by re-suspension in 10 mM PBS at 1% (w/v). The suspension was lysed by a single pass through an APV LAB40 high pressure homogeniser at 1000 Bar and stored on ice. Sodium deoxycholate was added at a concentration of 1 mg/mL and the lysate rolled for 15 min, RT, before 20 µL of Benzonase® nuclease (25 U/mL; Merck Millipore, Billerica,

Massachusetts, USA) was added and rolling continued for another 15 mins. The lysate was centrifuged at 17,700 x g, 30 min, 4 °C (Avanti J20 XPI; Beckman Coulter, Inc., Fullerton, CA, USA) to pellet the GCSF inclusion bodies (IB). After removal of the supernatant, the IB pellet was washed twice to remove host cell impurities. In all steps the pellet was resuspended in wash buffer at 1:40 (w/v) ratio at RT using a hand-held food blender and repelleted via centrifugation 17,700 x g, 30 min, 4 °C. Wash A contained 50 mM Tris pH 8, 5 mM EDTA and 2% Triton X-100 (g/v), and Wash B contained 50 mM Tris pH 8, 5 mM EDTA and 1 M NaCl.

2.2.3 Solubilisation and refold

Pellet solubilisation was achieved using a pH shift procedure, which included re-suspension in 10 mL of 4 M urea and pH adjustment to pH 12 using strong NaOH, followed by rolling for 30 min at RT. Refold was achieved by 20X dilution into 1 M Arginine.HCl buffer pH 8.25 via dropletting into the stirred buffer, followed by rolling for > 12 h, RT. Refold was quenched by pH adjustment to 4.25 using strong glacial acetic acid. The refold was clarified by centrifugation at 17,700 x g, 20 min, 4 °C, (Avanti J-20 XPI; Beckman Coulter, Inc., Fullerton, CA, USA), the supernatant was retained and concentrated to a final volume of 10 mL using successive centrifugal filtration steps using Amicon Ultra-15 10 kDa cut off membrane centrifugal filters (Merck Millipore, Billerica, Massachusetts, USA) at 1,389 x g and 4 °C. The final 10 mL sample was sterile filtered and stored at 4 °C.

2.2.4 Purification

Size-exclusion chromatography (SEC) was performed on an ÄKTA™ Avant (GE Healthcare Life Sciences, Germany) using a HiLoad® 26/60 Superdex® 200 prep grade column (GE Healthcare Life Sciences, Germany; 2.6 cm internal diameter; i.d., 60 cm bed height, 320 mL column volume; CV). The 10 mL sample was loaded onto the column using a 10 mL sample injection loop, and eluted isocratically in 50 mM Sodium Acetate pH 4.25 at a flow rate of 3 mL/min as described by Bristow et al. (2012). The fraction collector temperature was set to 6 °C and fractionation was triggered when the A₂₈₀ reading was greater than 5 mAU. The fraction volume was set to 3 mL and fractions were collected using the system's chilled fraction collector directly into 15 mL falcon tubes containing 12 mL Milli-Q H₂O to dilute the final sodium acetate buffer concentration to 10 mM. Fractions were analysed by non-reduced (NR) SDS-PAGE and UV absorbance as described in sections 2.4.2 and 2.4.1 respectfully, where fractions with > 95% purity and > 0.1 mg/mL concentration were pooled and concentrated to a final stock concentration of 0.6 mg/mL using Amicon Ultra-15 10 kDa cut off membrane centrifugal filters at 1890 x g and 4 °C.

2.2.5 Formulation

2.2.5.1 *Dialysis*

Buffer exchange of samples was performed using Slide-A-Lyzer Dialysis cassettes with a 10 kDa cut-off (Fisher Scientific, Leicestershire). Cassette volumes varied depending on the experimental requirements. A sterile syringe and needle was used to inject the sample into the dialysis membrane chamber following the manufacturer's protocol. The dialysis cassette was floated in a beaker containing buffer > 200X the volume of the sample, at 4 °C with stirring. The dialysis buffer was changed twice, once after two hrs and the other after four hrs. The final dialysis buffer and cassette were left stirring overnight and the sample removed using a fresh syringe/needle in the morning.

2.2.5.2 *Studies with excipients*

Excipients and purified GCSF samples were prepared as two separate stock solutions at 2X concentration. Excipient solutions were sterile filtered, and mixed at a 1:1 ratio with the GCSF stock solution to obtain desired 1X concentration of both.

2.3 Mutant production

Lab scale cell culture was performed in parallel for GCSF mutants and WT in order to remove processing variability. Glycerol stocks containing either WT or mutant plasmids were cultured overnight in 50 mL falcon tubes containing 10 mL TB/ Amp with shaking at 250 rpm, 37 °C. The 10 mL cultures were seeded into sterile 2 L baffled flasks containing 400 mL TB/Amp supplemented with 1 mM magnesium heptahydrate. Expression was induced at mid-exponential growth ($OD_{600}=0.6$) by spiking in a sterile filtered 1 M IPTG solution to a final concentration of 1 mM.

Primary separations and IB washing was performed as described in section 2.2.2, however centrifugation during the washing steps was performed at 5,410 x g, for 30 min, at 4 °C rather than 7,728 x g due to unavailability of a more powerful centrifuge. Furthermore, due to the centrifuge rotor having a capacity of 6 tubes per rotor, samples were split into two batches per centrifugation. Refold, clarification and concentration of IBs was performed in accordance with section 2.2.3. For the same reason as the IB wash step, clarification by centrifugation was at 5,410 x g.

Mutant and WT clarified refold samples were stored at 4 °C and concentrated one at a time to a final volume of 10 mL using successive centrifugal filtration steps using Amicon Ultra-15 10 kDa cut off membrane centrifugal filters (Merck Millipore, Billerica, Massachusetts,

USA) at 1890 x g, and 4 °C. The filters were thoroughly washed with 20% ETOH and MilliQ H₂O (Merck Millipore, Billerica, Massachusetts, USA) between samples. Once concentrated, samples were 0.2 µM filtered and progressed to purification whilst the next sample started concentration. Preparatory SEC was performed as described in 2.2.4. Cleaning of the column between samples consisted of a water flush for 1 column volume (CV), 0.2 M NaOH sanitisation for 1 CV, a second water flush for 1 CV and finally equilibration for 1 CV.

2.4 Characterisation

2.4.1 Ultraviolet (UV) absorbance

Protein concentration was calculated using ultraviolet (UV) spectroscopy at 280 nm (A₂₈₀) on a Nanodrop-2000 using the Beer-Lambert Law:

$$c = \frac{A}{\epsilon l} \quad \text{Equation 12}$$

where A is the absorbance value, ϵ is the extinction coefficient, l is the path length of cuvette and c is the unknown protein concentration. The extinction coefficient of WT GCSF was 0.86 (Herman et al. 1996). Where necessary the samples were diluted to produce an A₂₈₀ below 1.0.

2.4.1.1 *Mutant extinction coefficients*

Extinction coefficients for mutants were calculated from the numbers of tryptophan (W), tyrosine (Y) and cysteine (C) residues. Calculations were made using the Gill and von Hippel law derived from Edelhoch (1967) data:

$$E_{protein} = nW \cdot \epsilon W + nY \cdot \epsilon Y + nC \cdot \epsilon C \quad \text{Equation 13}$$

where ϵW is 5690 M⁻¹cm⁻¹, ϵY is 1280 M⁻¹cm⁻¹ and ϵC is 120 M⁻¹cm⁻¹ (Gill and von Hippel 1989).

2.4.2 Non-reduced (NR) SDS-PAGE

Non-reduced (NR) polyacrylamide gel electrophoresis in sodium dodecyl sulphate (SDS-PAGE) was used to assess protein yield and purity during expression, purification and degradation. SDS-PAGE was performed using Novex™ NuPage® 15-well 4-12% Bis-Tris precast gels with 1 x NuPage® MES running buffer (Thermo Fisher Scientific Inc., Wilmington, USA). For cell culture analysis, a 1 mL sample of cell suspension was

centrifuged for 5 min at 13,523 x g and the supernatant decanted. The cell pellets were resuspended in 1X Novex™ NuPage® LDS (Thermo Fisher Scientific Inc., Wilmington, USA) at a volume of 400 µL. For purified protein analysis, samples were mixed with Milli-Q® H₂O and LDS to obtain a sample concentration of 0.1 mg/mL and a 1 x concentration of LDS.

Samples were heated at 70 °C for 10 min, centrifuged at 13,523 x g for 5 min followed by 10 µL loading of the supernatant. A PageRuler™ Prestained Protein Ladder (Thermo Fisher Scientific Inc., Wilmington, USA) was added to the first (and sometimes last) lane of the gel at a load volume of 6 µL, with a molecular weight (MW) marker range of 10 to 180 kDa. Gels were run at the recommended constant voltage of 200 V for 35 min, followed by staining with InstantBlue™ (Expedeon Ltd, Cambridgeshire, UK) for > 1 hr and de-staining with distilled H₂O. Gel images were taken using an Amersham Imager 600 (GE Healthcare Bio-Sciences, PA, USA).

2.4.3 Bioactivity

The biological activity of the pooled sample was determined using a cell proliferation bioassay as developed by Whadawa et al. (2011). The bioassay uses the CellTiTer 96AqueousOne® solution with murine GNFS-60 cells which proliferate in response to GCSF.

Falcon® 96-well sterile, clear, TC-treated polystyrene microplates were used for analysis (Corning Life Sciences B.V., Amsterdam, The Netherlands). The test samples and the NIBSC 2nd international reference standard for GCSF (Wadhwa et al. 2011) were diluted to a concentration of 2 ng/mL with growth medium, and 100 µl loaded either across row A or down column 1 depending on if serial dilution was to be performed down or across the plate, respectfully. Growth medium was added, at 50 µl, to each well within rows B to H (serial dilution down the plate) or Columns 2 to 12 (serial dilution across the plate). Using either an 8- or 12-channel multichannel pipette depending on the direction, serial two-fold dilutions of samples were performed. The final 50 µl from the second last row (or column) was discarded, leaving the final column 12 or row H containing 50 µl media as a blank control. The experimental control and NIBSC 2nd international reference standard were included on all plates.

GNFS-60 cells were grown at 37 °C for 2 to 3 days, in T75 flasks, containing 20 mL of RPMI-1640 Medium, 2 ng/mL r-HuGCSF (Amgen, Uxbridge, UK), 0.5% (v/v) penicillin–streptomycin, and 5% (v/v) foetal bovine serum. The exponentially growing

GNFS-60 cells were washed three times by spinning them down in 50 mL falcon tubes at 250 x g for 10 mins and resuspension in 20 mL RPMI-1640 medium to remove any residual GCSF. Cell count was performed with a Countess Automated Cell Counter (Invitrogen, Life Technologies Corp, Paisley, UK). Cell viability and density was counted using a Countess® automated cell counter (Invitrogen™, Life Technologies Corp, Paisley, UK). A 11 µL sample of cells was added to 11 µL of 0.4% Trypan blue at RT and added immediately to a cell counting chamber slide with two 10 µL chambers. The cells were diluted to a final concentration of 2×10^5 cells/mL and 50 µL added to each well, diluting each sample 2X, as such the concentration of samples ranged from 1 ng/mL to 0.001 ng/mL across the plate and from 1 ng/mL to 0.007 ng/mL down the plate. Plates were covered and incubated at 37 °C for 48 h followed by the addition of 20 µL of CellTiter 96AqueousOne® Solution (Promega, UK) to each well and further incubation for 3-4 h at 37 °C. Absorbance of wells was measured at 490 nm using a plate reader (Spectramax 340PC, Molecular Devices LLC, Wokingham, UK) to determine the proliferation of GNFS-60 cells, with 5 s of shaking before reading.

2.4.4 High performance liquid chromatography (HPLC)

A ThermoScientific Ultimate3000 HPLC system with autosampler (Thermo Fisher Scientific Inc., Wilmington, USA) was used for all HPLC related experiments.

2.4.4.1 *Size exclusion (SE)*

Due to the non-glycosylation of GCSF, acidic size exclusion chromatography is performed to assess the identity and purity of GCSF samples. A TSK3000 swxl column, (300 mm x 7.8 mm, i.d. 5 µm particle size, Tosoh Life Sciences) was used with an isocratic mobile phase consisting of 0.1 M phosphate pH 2.5, as validated by Codevilla et al. (2004) and Herman et al. (1996), at a flow rate of 0.5 mL/min over 40 min. The column was blocked prior to sample injection using a number of injections of the NIBSC 2nd international reference standard for GCSF (Wadhwa et al. 2011), followed by 25 µL sample injections. Each sample was analysed in triplicate with a buffer blank. UV absorbance was measured at 214 nm and 280 nm and peak analysis was performed in Chromeleon (Thermo Fisher Scientific Inc., Wilmington, USA).

2.4.4.2 *Reverse phase (RP)*

a YMC-Pack C4 (Butyl) reverse phase (RP) column (YMC CO., LTD; 150 x 4.6 mm, i.d. 5 µM) was used to analyse mutant samples for variations in hydrophobicity. A gradient elution

was run in accordance with Herman et al. (1996) from 36% to 72% (v/v) acetonitrile (ACN) in 1% (v/v) formic acid (FA), at a flow rate of 0.8 ml/min over 30 mins. Buffers were 0.2 μ M filtered and helium sparged prior to use. Samples were injected in duplicate at a concentration of 1 mg/mL with an injection volume of 25 μ l. UV absorbance was measured at 215 nm and 280 nm wavelengths.

2.4.5 Accelerated thermal degradation of GCSF mutants

Shelf-life was assessed by accelerated thermal degradation. Mutant and WT samples were formulated at 0.2 mg/mL in sterile screw cap, 2 mL tubes and incubated at 37 °C and 45 °C. Samples were taken over a 7 day period and stored at -70 °C prior to analysis. Day 0 samples were also stored at -70 °C (rather than 4 °C) to remove any differences that may be caused by the freeze-thaw process. Samples were defrosted for 20 mins at RT, clarified by centrifugation to remove insoluble material, followed by analysis by SEC-HPLC and (NR) SDS-PAGE as described sections 2.4.4.1 and 2.4.2, respectfully. OriginPro 8.6 (OriginLab Corporation, MA, USA) was used to apply exponential fits to the degradation data using the equation:

$$y = y_0 + A \exp^{kt} \quad \text{Equation 14}$$

where A is the current value, k is the rate constant and t is time.

2.4.6 Melt midpoint (T_m) and aggregation onset (T_{agg}) temperatures

Tandem intrinsic protein fluorescence (IPF; 266 nm excitation, 280 to 450 nm emission scan) and static light scattering (SLS) at 266 and 473 nm were measured for samples using the UNit (Unchained Laboratories, UK). Three micro cuvette arrays (MCAs) containing 16 x 8 μ L sample cuvettes (total 48) with rubber seals were used for each run to maintain the same temperature exposure and sampling times. A temperature ramp of 20 to 90 °C was applied at a rate of 1 °C/30s. A hold at 20 °C was placed at the start of the run for 30 s as well as a 30 s hold at 90 °C at the end of the run. Each sample measurement was replicated five times. The barycentric mean fluorescence intensity (BCM), the intrinsic fluorescence (IFL) and peak height for IPF, and 266 nm counts for SLS, data were exported. Data was plotted in OriginPro 8.6 (Origin Lab Corp., Northampton, MA, USA). The aggregation onset temperature (T_{agg}) was determined, as described by Robinson et al. (2017), where linear fits were applied to the baseline SLS data at the lower temperatures of the curve, and the T_{agg} taken to be the point at which 10% increase in the light scattering at 266 nm occurs relative to the low-temperature baseline (Figure 11; Chakroun et al. 2015).

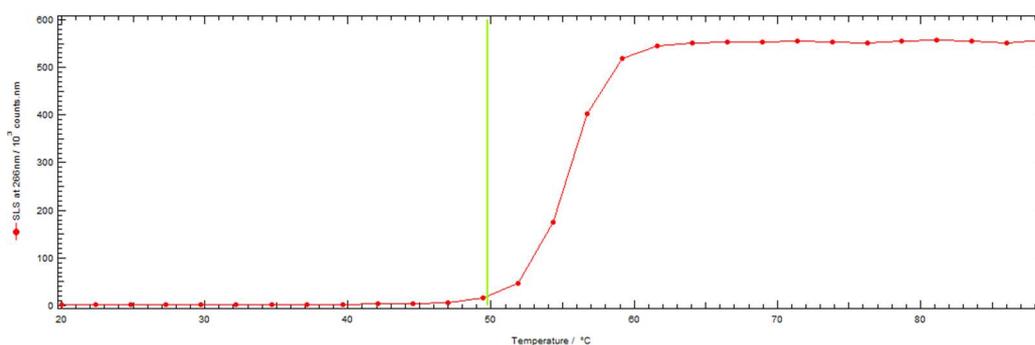


Figure 11. Example of SLS GCSF data with determination of the temperature of aggregation onset (T_{agg})

The thermal transition midpoint temperature (T_m) of variants was determined from IFL measurements by fitting the data to a two-state model using the following equation:

$$I_T = (I_N + aT) + \frac{(I_D + bT - I_N + aT)}{1 + \exp(T_m - T/m)} \quad \text{Equation 15}$$

where I_T is the observed signal, I_N and I_D are the native and denatured baseline intercepts, a and b are the native and denatured baseline slopes, T is the temperature, and T_m is the midpoint of the thermal transition (Chakroun et al. 2016).

2.4.7 Mass measurements by mass spectrometry

2.4.7.1 WT GCSF

The MW of expressed and purified GCSF WT and mutant proteins were measured using an Acquity ultra-performance liquid chromatography (UPLC) system coupled a Xevo G2XS quadrupole time of flight (QToF) with Acquity sample and quaternary solvent managers (Agilent, CA, USA). Separation was performed using a BEH C4 300 (0.1 mm x 50 mm, i.d. 1.7 μ m pore size; Waters Corp., Milford, MA, USA) at 50 °C. The LC mobile phase solvent A consisted of 0.1% (v/v) aqueous formic acid (FA; Thermo Fisher, Hemel Hempstead, UK) and solvent B consisted of acetonitrile (LGC Standards, Teddington, UK) with 0.1% (v/v) FA. The mass spectra were acquired using the QToF/MS Scan parameters: 100 to 3,000 m/z at 3 spectra/s over 333.3 ms/spectrum, 3,225 transients/spectrum and 0 V collision energy.

Samples at 0.3 mg/mL were diluted 2-fold in 10 mM sodium acetate pH 4.25, and 0.3 μ L injected in triplicate. Blank runs were carried out before and after sample injections. Chromatograms were processed on MassLynx V 4.1 software (Waters Corp., Milford, MA, USA) by smoothing the elution peak by the Savitzky Golay method (window size = 1, Number of smooths = 2), and extraction of half height, full width of the peak. Deconvolution was performed to obtain the mass values using the MaxEnt1 function with a

range of 17,000-20,000 and resolution of 0.10 Da/channel. The Uniform Gaussian damage model with a width at half height set at 0.5 Da was applied, with minimum intensity ratios set at 33% left and right, and iterations were allowed until convergence. The main deconvoluted peak mass was compared with the predicted average mass of 18,803 Da (minus 4 hydrogen atoms for S=S, = 18,799 Da) for WT GCSF. For mutants, the difference between the mutant and WT average mass of main deconvoluted peaks was calculated and compared to the theoretical amino acid mass to confirm mutant identity.

2.4.8 Collision induced unfolding ion-mobility spectrometry-mass spectrometry (CIU-IMS-MS)

The unfolding of GCSF with increased collision voltage (CIU) was performed using ion-mobility spectrometry-mass spectrometry (IMS-MS). GCSF at a concentration of 0.2 mg/mL was dialysed into 50 mM ammonium acetate pH 4.25 using Micro Bio-Spin 30 columns (Bio-Rad, Hercules, CA, USA) as per the manufacturer's instructions. The column was washed four times prior to applications to ensure 99.9% buffer exchange from the storage buffer and sample buffer. The sample was applied in a 50 μ L aliquot to the column twice to ensure complete dialysis.

The electrospray ionization-ion mobility-mass spectrometry was performed with a SYNAPT G2 High Definition Mass Spectrometer (HDMS) equipped with a NanoLock ZSpray Exact Mass Ionization Source and MassLynx data processor (all Waters Corp., Milford, MA, USA). The dialyzed GCSF samples were infused to the standard electrospray (z-spray) source using a copper capillary apparatus at an infusion rate of 10 μ L/min. The capillary of the ESI source was held at 3 kV, with the source operating in positive ion mode. The sample cone was operated at 5 V, which was required to avoid gas-phase unfolding and preserve non-covalent interactions. The trap travelling wave collisional cell, was pressurised to 3×10^{-2} mbar. The ion mobility separator, containing nitrogen gas at 0.45 mbar and ambient temperature, employed a series of DC voltage waves (7 V wave height traveling at 280 m/s) to achieve conformational separation. The TOF-MS was operated over the scanning range of m/z 500–5000 at a pressure of 1.8×10^{-6} mbar. Collision energy was added to the ions in the traveling-wave-based ion trap situated prior to the IM separator to perform GCSF CIU.

Tandem-MS (quadrupole selection) mode was used, in which protein ions at m/z values at 2350.7 m/z were selected for. Folded 5 V MS spectra was acquired for 1 min at the start,

middle and end of CIU experiments. The collision voltage was ramped from 5 to 27 V in 3 V increments, with an acquisition time of 2 mins. IM data was recorded for MS-isolated ions at each collision voltage and the IM data for only those ions corresponding to the intact m/z originally isolated were compiled to create CIU figures.

Mass spectra and drift times were processed with MassLynx V 4.1 software (Waters Corp., Milford, MA, USA). Drift time chromatograms were smoothed by the Savitzky Golay method (window size = 1, Number of smooths = 2). GCSF CIU “fingerprints” were plotted as described by Tian et al. (2015) and Eschweiler et al. (2015) using the Protein Unfolding for Ligand Stabilisation and Ranking (PULSAR) software (Allison et al. 2015), with collision voltage hierarchy, and a mass fitting of 18799.1, Zavg of 8 and zwidth of 0.2. Fitting was minimised and exported to PULSAR followed by the creation of unfolding plots with arrival time on the x-axis. Fingerprint plots were prepared using a viridis colour map, where slices were normalised but not smoothed.

2.4.9 Hydrogen deuterium exchange mass spectrometry

2.4.9.1 *General setup*

All HDX-MS experiments utilised a LEAP PAL™ sample handling and mixing system (LEAP Technologies, Carrboro, NC, USA) along with the ACQUITY UPLC M-Class System with HDX Manager (Waters, Manchester, UK), coupled to a SYNAPT® G2-Si HDMS™ (Waters, Manchester, UK).

The Waters HDX director software controlled the LEAL PAL™ automation platform, providing a graphic display of runtime schedules and applying an efficient scheduling algorithm to minimise sample preparation times. Sample handling and mixing were performed using the LEAP PAL™ system set. For each run, a 15 µL protein sample aliquot was diluted 1 in 10 in a buffer solution in either H₂O or D₂O. When preparing D₂O solutions, pH adjustments were made using 99% deuterium chloride (DCl), and corrections were made for pH meter anomalies in D₂O, where $pD\text{ corr} = pD\text{ read} + 0.4$ (Glasoe and Long 1960).

For D₂O exchange experiments, samples were incubated at RT and quenched by a 1:1 dilution in a pH 2.5 solution at 4 °C for 30 s before loading onto an Enzymate™ BEH 5 µm online pepsin column, 2.1 × 30 mm (Waters, Manchester, UK). The temperature was maintained at 25 °C within the controlled digestion column compartment of the HDX

manager. The digested peptides were eluted in 0.05% FA at a flow rate of 80 $\mu\text{l min}^{-1}$ followed by entrapment on a reverse phase VanGuard™ Pre-Column (Waters, Manchester, UK). The flow was diverted by switching valves, and trapped peptides were eluted into an ACQUITY UPLC BEH C18 (1.0 mm x 100 mm, 1.7 μm particle diameter, (Waters, Manchester, UK), at 0 °C within the HDX manager. Peptides were chromatographically separated using a linear gradient starting at 8% ACN, 0.1% FA, and increasing over 7 min to 35% with a flow rate of 100 $\mu\text{L}/\text{min}$, also maintained within the HDX manager at 0 °C. The eluent was directed into the mass spectrometer with electrospray ionization and post-acquisition lock mass-corrected using the 2+ charge state of [Glu1]-fibrinopeptide B, which was infused at a concentration of 100 fmol/ μL at 90° to the analytical sprayer (Cryar et al. 2017).

2.4.9.2 *Peptide maps*

GCSF peptide maps were collected with the mass spectrometer operating in the data independent acquisition (DIA) mode, MS^E. For measurements, the quadrupole was used in rf-only mode, with the quadrupole tuned such that only ions with m/z between 50 to 2000 were transmitted. DIA data was processed using the ProteinLynx Global Server (PLGS) ver 3.02 software (Waters, Manchester, UK). The PLGS processed MS^E data was then database-searched against the accurate mass (± 20 ppm) of the GCSF amino acid sequence, obtained as described in section 2.4.9.2. The Uniprot amino acid sequence of porcine pepsin (P00791) was also included and the enzyme specificity was set to nonspecific (Cryar et al. 2017). Protein identification was set to detect at least 2 fragments per protein, 7 fragments per peptide, and at least 1 peptide per protein. Oxidation of methionine was set as the only variable modification. The peptide map output files were imported into DynamX™ Software v 3.0 (Waters, Manchester, UK) for peptide map finalisation. In general, five LC-MS/MS injections were made per peptide map and the threshold for peptide retention within the map was set to four. A minimum of 0.01 products per amino acid was also set, along with a precursor mass error of < 20 ppm.

2.4.9.3 *Peptide deuterium uptake*

For full HDX-MS experiments, samples were incubated in the deuterium solution for 30 s, 2 min 36 s, 5 min, 15 min, 1 h, 4 h and 8 h. When a large number of samples were to be run, the number of labelling time points was reduced in order to fit samples within the 48 sample holder in the LEAP-PAL™ quench and labelling trays. If sample numbers were still

too large then consecutive runs were performed and a WT GCSF control included in each run to monitor intra-run variability.

HDX-MS data acquisition was performed in MS mode only. To process the MS data, the amount of deuterium uptake for each peptide as a function of time was determined by measuring the centroid of the isotopic distribution using the DynamX 3.0 (Waters, Manchester, UK). Deuterium exchange measurements were analysed with default settings and all data were manually validated and curated if required. Corrections for back-exchange were not made because the data consisted of comparisons between wild type and mutants, or contained an internal reference peptide as described in section 2.4.9.9.

2.4.9.3.1 Differential uptake analysis

To analyse differences between HDX-MS experiment samples, the relative differential uptake of deuterium was calculated by:

$$\text{Relative differential uptake (t)} = m_{\text{experimental}}(t) - m_{\text{control}} \quad \text{Equation 16}$$

where $m_{\text{experimental}}$ is the average mass of the experimental peptide and m_{control} is the average mass of the same peptide in the control solution at a given labelling time (t).

Differential plots were prepared with the y-axis showing the calculated relative differential uptake at a set labeling time point, and the x-axis showing the GCSF peptides produced for the particular experiment. In the interest of space, the ordinal number of peptides was listed on the x-axis, and the corresponding amino acid sequence and residue number range included in the Appendix. Peptides including residues involved in α -helix structures, as identified by Tamada et al. (2012), were also highlighted on differential plots. As some peptides included residues overlapping both loop and α -helix, a cut off of > 50% sequence coverage (last residue number minus half the total number of residues in a peptide) was applied.

2.4.9.4 GCSF peptide map optimisation

The peptic digestion of GCSF required optimisation to obtain a high (> 95%) sequence coverage and redundancy. A range of concentrations of guanidine hydrochloride (Gnd.HCl) from 2 M to 4 M, and reducing agent tris(2-carboxyethyl)phosphine (TCEP; Thermo Fisher, Hemel Hempstead, UK) from 50 mM to 1 M within the quench solution were tested. The

pepsin column parameters were kept constant with a flow rate of 80 $\mu\text{l min}^{-1}$, 25 °C and a capillary tube with a length of 152 mm, and 2 μm i.d. Samples were run in triplicate with blank injections (50 mM sodium acetate, pH 4.25) performed between each sample set to confirm the absence of peptide carryover from previous runs. The pepsin column operation optimisation included the assessment of backpressure, where peptide maps with and without the peek capillary tube were assessed.

2.4.9.4.1 Peptide map analysis

PLGS analysis of peptide maps was performed as described in section 2.4.9, and the output exported to DynamX™ to obtain peptide map coverage %, cleavage rate and redundancy values. The digestion efficiency (DE) for each quench solution peptide map was calculated using the equation:

$$DE = 8y_1 + 5y_2 + 0.6y_3 \quad \text{Equation 17}$$

where y_1 is coverage %, y_2 is cleavage rate% and y_3 is redundancy (Ahn 2013).

2.4.9.5 *Measuring the effect of sample pH on GCSF HDX-MS*

The effect of pH on GCSF conformational stability was assessed by HDX-MS using two different pH buffers: 50 mM sodium acetate, pH 4.25 and 10 mM PBS, pH 7.4 (Severn Biotech Ltd, Kidderminster, UK). The GCSF sample was dialysed into PBS (Severn biotech ref), pH 7.4 as described in section 2.2.5.1. Full HDX-MS was performed as described in 2.4.9.3. Two condition solutions were prepared for each experimental condition: an undeuterated and a deuterated solution. The undeuterated solutions contained the buffers in 100% H_2O . The deuterated solutions contained the buffers in 99.9% deuterium oxide (D_2O) at pD 4.25 ($\text{pD} = \text{pH}_{\text{read}} + 0.40$) and (D_2O) pD 7.4 ($\text{pD} = \text{pH}_{\text{read}} + 0.40$). Seven labelling time points were used and including 30 s, 2.6 min, 5 min, 15 min, 1 h, 4 h and 8 h. Differing condition and sample solution pH HDX-MS experiments were ran consecutively with a total of three runs performed per experimental condition to monitor reproducibility. Results were processed and analysed as described in section 3.3.8.1.

2.4.9.6 *Deuterated excipient HDX-MS*

To observe GCSF-excipient interactions, peptide-level HDX-MS was performed with GCSF and different excipients formulated within a 10 mM sodium acetate in 99.9% deuterium oxide (D_2O) pD 4.25 solution. Excipients were deuterated prior by dissolving the solids in 99.9% D_2O . The deuterated excipient solutions were aliquoted into 2 mL Schott glass vials

(VCDIN2D, Adelphi Tubes Ltd, Haywards Heath, UK) with a 1 mL fill volume and igloo rubber stoppers placed inside, leaving a gap for vapour passage. The vials were subjected to three consecutive freeze-drying cycles (Table 1) to ensure full deuteration using a VirTis AdVantage freeze-dryer (Biopharma, Winchester, UK) as described by (Kendrick et al. 1997). Between each cycle the samples were rehydrated to the original volumes except for after the final freeze-drying cycle where the vials were backfilled with nitrogen under 500 Bar and stoppered using the hydraulic stoppering of the freeze-dryer shelves. Stoppered vials were quickly removed from the freeze-dryer and snap frozen in liquid nitrogen followed by storage at -80 °C until use in HDX-MS experiments.

Table 1. Freeze-dry cycle for VirTis Advantage dryer with a 1 mL fill within 2 mL glass vials. The freeze-dry stages are split into steps describing the temperature, time, vacuum applied (vac) and whether the step is a thermal ramp (R) or hold (H).

Stage	Step	Temp (°C)	Time (min)	Vac (mTorr)	Ramp/Hold
Freeze/Anneal	1	4	30		H
	2	-47	90		R
	3	-47	60		H
	4	-10	60		R
	5	-10	60		H
	6	-47	60		R
	7	-47	60		H
Primary drying	1	-47	60	100	H
	2	-35	30	100	R
	3	-35	480	100	H
Secondary drying	1	20	180	100	R
	2	20	300	100	H
	3	20	0	20	H

Deuterated excipient vials were defrosted for 10 mins, RT, prior to HDX analysis and reconstituted in 10 mM sodium acetate 99.9% D₂O solution pH_{read} 3.85 (pD 4.25) adjusted with DCl. A total of ten vials were pooled per excipient, totalling four 10 mL D₂O deuterated excipient solutions and one control solution of 10 mM sodium acetate 99.9% D₂O stock solution. The pH of the deuterated excipient solutions was adjusted where necessary to 3.81 (pD = 4.21) using a 10% (v/v) DCl solution.

2.4.9.6.1 HDX-MS system adjustments for deuterated excipient experiments

To perform different reagent vial sampling instead of the standard single pots for D₂O and H₂O, the HDX-manager was re-programmed to allow selection of the four 10 mL vials within

the vial holder as one of the parameters during sample scheduling. The reagent vial at position #2 was fixed throughout the run as the non-deuterated solution and reagent vial at position #1 as the control deuterated solution, leaving reagent vials at positions #3 and #4 as variables for deuterated excipient solutions.

2.4.9.7 Solid state HDX-MS

Formulations of GCSF were filled at 200 µL in 2 mL glass vials with igloo rubber stoppers and placed on the Virtis Genesis 25EL freeze-dryer shelf along with excipient only versions of the formulations (no GCSF). The excipient only samples were used to monitor sample temperature during freeze-drying via the insertion of thermocouples. A two day freeze-dry cycle was applied as described in Table 2. Post-lyophilisation vials were backfilled with nitrogen and stoppered.

Table 2. 42h r lyophilisation cycle for 200 µL filled 2 mL vials using a VirTis Genesis freeze-dryer

Stage	Step	Temp (°C)	Time (min)	Vac (mTorr)	Ramp/Hold
Freeze	1	20	30		H
	2	-45	120		R
	3	-45	240		H
Primary drying	1	-45	30	150	H
	2	-45	30	70	H
	3	-25	60	70	R
	4	-25	1200	70	H
Secondary drying	1	30	480	70	R
	2	30	420	20	H

2.4.9.7.1 Labelling in the solid state

Immediately prior to labelling, stoppers were removed and the lyophilised GCSF vials placed around the edges of three 2.4 L borosilicate glass DURAN® desiccators with flat flanges and knobbed lids. The desiccators were prepared the day before and contained 70 mL 99.9% D₂O and potassium carbonate (VWR Chemicals, Leicestershire, UK) in excess for a water activity (a_w) of 0.43 (43% RH). Samples were incubated in triplicate for different labelling time points ranging from 30 min, to 120 hrs, before removal by sliding the glass lid enough to remove individual vials. Vials were stoppered and immediately flash frozen in liquid nitrogen before storage at -70 °C. Vials for 0 s labelling (non-deuterated) were not placed in desiccators and instead immediately stored at -70 °C post-lyophilisation.

2.4.9.7.2 HDX-MS system adjustments for solid state experiments

For ssHDX-MS the HDX manager method was reprogrammed to remove all LEAP PAL™ automated sample preparation up until injection of quenched sample into the sample loop of the LC-MS system. A single vial position was selected for injection, which was triggered once a quenched sample vial was in position 1 within the quench vial tray. The pre-cleaning of the syringe was also removed from the method to reduce the time from vial to injection.

2.4.9.7.3 Reconstitution of ssHDX-MS samples

Samples were removed from -70 °C storage and placed on dry ice prior to LC-MS analysis. A single sample was removed from dry ice, defrosted by hand so the stopper could be removed. The sample was reconstituted in 2 mL of ice cold 0.2% (v/v) FA, followed by vortexing for 10 s. Once reconstituted, 50 µL of the sample was mixed with 50 µL ice cold quench solution within a high recovery HPLC vial and placed in position for injection and the HDX manager instructed to inject the sample.

Blank samples containing 0.2% (v/v) FA were injected between sets of samples. Once injected the samples were processed by LC-MS as per the standard HDX-MS method 2.4.9.3 with differential uptake analysis performed for samples with and without excipients as described in 2.4.9.3.1.

2.4.9.8 *Ultra scale-down solid state HDX-MS*

Formulations were filled at 50 µL in to 1.4 mL Micronic tubes in a 96-well format with plug style TPE Lyo Caps-96 placed on top (MP53099; Micronic, Lelystad, Netherlands). The tubes were placed in an aluminium 96-hole holding plate (Biopharma, Winchester, UK) in the centre of a shelf in the LyoBeta 15 freeze-dryer and subjected to a 24 hr freeze-dry cycle due to the low fill volume (Table 3). At the end of the cycle the chamber was backfilled with nitrogen and the tubes stoppered to reduce moisture absorption of the cakes.

Table 3. 24 hr lyophilisation cycle for with 50 µL filled 1.4 mL Micronic tubes

Stage	Step	Temp (°C)	Time (min)	Vac (mTorr)	Ramp/Hold
Freeze	1	20	10		H
	2	-45	90		R
	3	-45	60		H
Primary drying	1	-45	30	150	H
	2	-45	30	70	H
	3	-25	30	70	R
Secondary drying	1	-25	720	70	H
	2	30	240	70	R
	3	30	240	20	H
Post drying	1	30	60	20	H

Samples were labelled as described in section 2.4.9.7.1. Reconstitution and analysis was also performed as described in section 2.4.9.7.3, with the exception of the reconstitution volume of 0.2% (v/v) FA was changed to 500 µL, rather than 2 mL, maintaining the 10X dilution.

2.4.9.9 *Internal reference peptide (IRP)*

Internal reference peptides (IRP) are important to include within HDX experiments to monitor changes in back exchange as well as the differences in intrinsic HDX rates for different formulation conditions. A synthetic tetrapeptide, [NH₂] PPPI [COOH], as described by Zhang et al. (2012) for the monitoring of intrinsic HDX rates, was ordered by Kate Groves (LGC, Teddington, UK) from ThermoFischer Scientific (Thermo Fisher Scientific Inc., Wilmington, USA). The small 422.53 Da IRP was formulated at 1:1 MEOH:ACN in 10 µL aliquots and frozen at -20 °C. When required a fresh aliquot was removed from the freezer, defrosted, diluted 1:8 in sample buffer and spiked into samples, where specified, at 1:100 for a final concentration of 3 µM.

For uptake analysis in DynamX™ ver 3.0, the IRP was manually added by selecting to add a new protein alongside GCSF and assigning the IRP to the protein by inputting the retention time of 6.11 min and a [M+H]⁺ of 423.2602. The maximum uptake of the peptide was 1, and the charge state was +1.

2.4.10 Ultra-scale down lyophilisation

For ultra-scale down (USD) lyophilisation of GCSF, 96-well flat-bottom plates (Greiner Bio-one Ltd, Gloucestershire, UK) with their skirts custom removed were used, as described in Grant et al. (2011). Lyophilisation was performed with the LyoBeta 15 freeze-dryer applying two different cycles depending on the formulation. For formulations not containing mannitol, the cycle in Table 4 was applied, and for formulations containing mannitol the cycle with an additional anneal step in Table 5 was applied.

Table 4. Freeze-dry cycle of 96-well microtitre plates

Step	Stage	Temp (°C)	Vac (mBar)	Time (HH.MM)
1	Shelf temperature	20		
2	Freezing	-50		01.30
3		-50		01.00
4	Chamber Vacuum		0.1	
5	Primary drying	-50	0.1	00.30
6		-25	0.1	00.30
7		-25	0.1	10.00
8	Secondary drying	30		05.00
9		30		05.00

Table 5. Freeze-dry cycle 96-well microtitre plates with anneal step

Step	Stage	Temp (°C)	Vac (mBar)	Time (HH.MM)
1	Shelf temperature	20		
2	Freezing	-50		01.30
3	Anneal	-10		01.00
	Freezing	-50		01:00
4	Chamber Vacuum		0.1	
5	Primary drying	-50	0.1	00.30
6		-35	0.1	00.30
7		-35	0.1	10.00
8	Secondary drying	25		05.00
9		25		05.00

Post-lyophilisation, the plates were stoppered with plug style TPE Lyo Caps-96 placed on top (Micronic, Lelystad, NED) inside the freeze-dryer.

2.4.11 GCSF solid state formulation screen

WT GCSF at 0.3 mg/mL was dialysed into one of three buffers (50 mM sodium citrate pH 4.2, 50 mM sodium acetate pH 4.2, 10 mM PBS pH 7.4) as described in 2.2.5.1 and added to formulations containing either trehalose, sorbitol, phenylalanine, mannitol or no excipient (control) in the corresponding buffers as described in 2.2.5.2. The total number of formulations was 15 and each formulation was loaded in duplicate to the central wells of two 96-well plates with their skirts custom removed, at a volume of 180 µL per well, as depicted in Table 6. One of the two plates was lyophilised as described in section 2.4.10 using the cycle in Table 5. The second plate was stored in the aqueous phase at 4 °C, serving as a pre-lyophilisation sample.

Table 6. **Layout for liquid and solid state samples within a 96-well microtitre plate.** Formulations are coded by two letters, the first is the buffer used: acetate (A), citrate (C) and PBS (P), the second is the excipient used: trehalose (T), sorbitol (S), mannitol (M), phenylalanine (P) and control (no letter).

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B												
C				AT	AT	CT	CT	PT	PT			
D				AS	AS	CS	CS	PS	PS			
E				AP	AP	CP	CP	PP	PP			
F				AM	AM	CM	CM	PM	PM			
G				A	A	C	C	P	P			
H												

Lyophilised samples were reconstituted with 180 µL ultrapure water (NIBSC, Hertfordshire, UK) and the retention of GCSF monomer during lyophilisation was determined by SEC-HPLC as described in section 2.4.4.1. The monomer retention % was calculated using the equation:

$$\text{Monomer retention \%} = \frac{\text{Peak area}_{\text{post-lyophilisation}}}{\text{Peak area}_{\text{pre-lyophilisation}}} \times 100 \quad \text{Equation 18}$$

2.5 Solid state formulation analysis

2.5.1 Residual moisture analysis by thermogravimetric analysis (TGA)

To assess the residual moisture, TGA was performed on a Perkin Elmer Pyris 1 system (PETA Solutions, Seer Green, Bucks, UK) fitted with an Accupik autosampler. Samples were placed in a nitrogen filled dry glove bag with a relative humidity (RH) of less than 10%. Stoppers were removed and crumbled lyophilised samples packed into hermetically sealed pre-weighed pans (Perkin Elmer N537 0464). Pans were crimped and punctured by the Accupik before loading onto the TGA balance to determine sample weight. Multiple samples were panned for each formulation and samples of Aquamicron solid water standard (A1-Envirosciences, Blyth, UK) were analysed at the beginning and end of each set as a calibrant. A temperature ramp of 30-300 °C at 15 °C/ min was applied and off-gas analysis was performed on a Hiden HRC-20 quadrupole mass spectrometer using helium as the sparge gas. Data was analysed using the Pyris and Hiden software and the water content % (w/v) determined as the change in weight (Y) at the time (min) the water came off the sample.

2.5.2 Determination of formulation collapse temperatures (T_c) by freeze-drying microscopy (FDM)

The FDCS 196 stage mounted on a BX51 Olympus optical microscope with Linkam control units: TMS 94, VC 94 and LNP was used to measure thermal collapse (T_c) of freeze-dried material. The system was connected to a liquid nitrogen Dewar for freezing of samples along with an Edwards vacuum pump (E2M1.5) to pull a vacuum during drying. Samples were pipetted onto a quartz glass crucible with a metal shim and covered with a 13 mm glass coverslip. The crucible was placed in a sample holder and slid into the microscope stage/ immersion lens oil. The holder was moved until the microscope image covered the sample edge with a 20 x lens. The samples were subjected to a freeze step to -50 °C at a ramp rate of 10 °C/ min, followed by a hold for 5 mins at -50 °C. A vacuum of 0.1 mBar was pulled for 5 mins at -50 °C followed by a temperature ramp of 5 °C/min up to 25 °C. Photo frequency was set to every 5 s during freezing and every 2 s during drying. The T_c of each sample was determined by reviewing the run images and working back from total collapse to the point of initiation. The crucible and shim were washed in water and ethanol between samples and a fresh glass coverslip used each time.

2.5.3 Differential scanning calorimetry (DSC)

Modulated differential scanning calorimetry was used to determine thermal events for the formulations used in the USD excipient screen study. 80 μL of a sample was added to an individual pan with lid and O-ring, and crimped. Pans were weighed before and after sample addition and the net sample weights listed in the software along with pan position. Each formulation was analysed in triplicate in separate pans. DSC was performed on a Q2000 DSC (TA instrument, Crawley, Surrey, UK) together with an empty crimped reference pan. Samples were held isothermally for 2 mins, followed by cooling at 10 $^{\circ}\text{C}/\text{min}$ to -90°C . Modulation was then applied at $\pm 1^{\circ}\text{C}$ every 60 s with a sampling interval of 1 s per point. Heating was ramped at 3 $^{\circ}\text{C}/\text{min}$ to 25 $^{\circ}\text{C}$. The glass transition values were determined from the transition midpoints using Universal Analysis 2000 software (TA Instruments, New Castle, US). Crystallisation points were determined by large exothermic dips in heat flow.

2.5.4 Reconstitution scoring

Reconstituted lyophilised samples were scored +1 for immediate dissolution with no particulates, or 0 for dissolution within 2 mins or with pipetting/mixing, and -1 for non-dissolution, presence of particulates/ opalescence.

2.6 *In silico* applications

2.6.1 Structure selection

The RRSCB protein databank (<http://www.rcsb.org/pdb/home/home.do>) was used to search for solved GCSF structures (Rose et al. 2012). Available structure files (.pdb) were downloaded and edited to contain only a single GCSF structure using PyMOL visualisation software (Schrödinger LLC, NY, USA). Information such as amino acid sequence and B-factors (from crystal structures) were extracted from the files. Where available, B-factor values were averaged for each amino acid, not including hydrogen atoms, and normalised. The PDB files were evaluated against purified GCSF where amino acid sequence alignment, structure resolution, R-factor % and completeness, and R^2 values from HDX-MS, B-factor correlations were compared.

2.6.2 Changes in mutant folding energy ($\Delta\Delta G$) predictions

2.6.2.1 *RosettaDesign*

The PDB 2D9Q (Tamada et al. 2012) was cleaned using the “clean_pdb.py script in Rosetta and used as the input file for RosettaDesign (<http://rosettadesign.med.unc.edu/>; Liu and

Kuhlman 2006). The clean protocol renumbered the amino acids to start from 1, so for 2D9Q the start amino acid, number 7, became number 1 and so on and so forth.

Individual resfiles containing instructions on the amino acid position for mutation were prepared for each residue within the targeted amino acid sequence of GCSF (Table 7A). Resfiles for residues involved in GCSFR interactions were not generated. For each prediction, the 2D9Q PDB and the specific resfile were input into the server. A second resfile to obtain WT total energy was also prepared (Table 7B).

Table 7. Resfile text used for RosettaDesign point mutation modeling. A) Single position mutation, where “n” is the numerical position of the amino acid to be mutated. B) WT

A	B
NATAA	NATAA
EX 1	EX 1
USE_INPUT_SC	USE_INPUT_SC
start	
n A ALLAA EX 1 EX 2 EX 3	

Ten independent trajectories were selected for each position, resulting in ten output files per mutation. From each output file, the protein pose total energy, change in residue and the residue total energy information was extracted and averaged for each mutation. The relative change in folding free energy due to point mutations relative to WT, $\Delta\Delta G$, was calculated using the minimum values in the equation:

$$\Delta\Delta G = \Delta G_{Mutant} - \Delta G_{WT} \quad \text{Equation 19}$$

As convention, a negative $\Delta\Delta G$ value was predicted to be a stabilising mutation and a positive $\Delta\Delta G$ was predicted to be a destabilising mutation (Zhang et al. 2018).

2.6.2.2 *Rosetta_ddg_monomer*

The cleaned 2D9Q PDB from the RosettaDesign experiment was minimised and used as the input for *Rosetta_ddg_monomer*. The files associated with the structure preparation were generated by using several python scripts, and the jobs were submitted to UCL Legion High Performance Computing Facility (Legion@UCL) with Rosetta Version 2015.31.58019 as described by Zhang et al. (2018). The high resolution algorithm iterations were set at 50 per mutation and the $\Delta\Delta G$ was calculated from the difference between the mean top three scoring WT and mutant structures. In total 3,192 structures with single mutations were

created. The total predicted ddg files were exported, results combined and a mutational map generated with a 3-colour scale.

2.6.3 Excipient docking

The flexible docking software Generic Evolutionary Method for Molecular DOCKing (iGEMDOCK; Yang et al. 2004) was used to scan for excipient interaction regions across the GCSF 3D structure. The selected GCSF PDB file was edited to remove two GCSFR structures leaving a single GCSF molecule for analysis. The GCSF molecule was protonated for pH 4 by generating a PQR file using the online server www.pbd2pqr.org and applying this to the structure on PyMOL (Schrodinger, LLC, New York, NY, USA) via the Adaptive Poisson-Boltzmann Solver (APBS) plugin tool. 3D excipient molecules were drawn in Maestro 10.3 (Schrodinger, LLC, New York, NY, USA) as described by Barata et al. (2016) using pH 4 pKa information from www.chemicalize.org. A short minimisation protocol was performed using Macromodel 10.0 (Schrodinger, LLC, New York, NY, USA) and the final structures saved as MOL files.

2.6.3.1 *iGEMDOCK*

iGEMDOCK v2.1 graphical environment for recognizing pharmacological interactions and virtual screening (Yang and Chen 2004) was applied to identify regions of GCSF-excipient interactions. Accurate docking (very slow) was selected as the default setting (population size 800, generations 80, and number of solutions 10). The iGEMDOCK docked poses and post-screening analysis tools were used for interaction characterisation and docking energy calculations. The 10 best_pose solution files were visualised using Discovery Studio 4.0 (Biovia, San Diego, CA, USA). Using the Docked Poses/ Post-Screening Analysis tab, clustering of poses was analysed using the “set interaction and atom composition” drop down with 4 interaction clusters and 4 atom composition clusters. Interacting residues were divided in the iGEMDOCK software into two interacting groups (main and side chain) and types of interaction (electrostatic interaction, H-Bond, or VDW).

3 Biophysical and *in silico* method development

This chapter aimed to develop biophysical techniques for the characterisation of GCSF stability and use the data obtained to identify a suitable 3D structure for *in silico* research. The methods developed would also serve as a platform on which further experimental research with structural mutations and formulations could build upon.

In order to produce GCSF, a method first developed by Dr Adrian Bristow at the National Institute for Biological Standards and Control (NIBSC; Bristow et al. 2012) and adapted by Dr Mathew Robinson at University College London (UCL; Robinson et al. 2017) was used. The GCSF plasmid within host BL21 (DE3) *E. coli* cells was first extracted and sequenced to confirm the amino acid sequence. The protein was subsequently expressed using large scale cell culture, refolded and purified. Purity of the final product was measured using different biochemical analysis techniques such as SDS-PAGE, and HPLC, most of which are described in either the European Pharmacopoeia (1999) 6.3 for Filgrastim, or the original characterisation and formulation chapter for Neupogen® by Herman et al. (1996) published in the book “Formulation, Characterisation and Stability of Drugs”. The intact mass of GCSF was also determined using intact protein mass spectrometry and compared to the theoretical mass to confirm the identity of the protein.

Biophysical methods to measure stability were explored in this chapter. Global stability was determined using thermal denaturation with tandem fluorescence and static light scattering (SLS) measurements for a snapshot of stability with varying buffer and protein concentrations. Collision induced unfolding ion-mobility spectrometry mass spectrometry (CIU-IMS-MS) was also explored as an orthogonal, biophysical technique for global stability analysis.

Local stability measurements were investigated by peptide level hydrogen deuterium exchange mass spectrometry (HDX-MS), however, due to the high complexity of operation, significant method development was also required. Firstly, peptide digestion of GCSF required optimisation for > 95% sequence coverage and high levels of peptide redundancy. The concentration of the chaotropic agents, Gnd.HCl and TCEP, in the quench solution were varied as well as the inclusion of a peek flow restrictor to increase back pressure to enhance digestion were investigated. Secondly, repeatability of HDX-MS experiments was assessed by the completion of three consecutive identical experiments and the average relative uptake of peptides assessed. Thirdly, GCSF peptide uptake was compared with pH 4

and 7 solution conditions to determine if differences observed in uptake were intrinsic in nature or changes to protein dynamics had occurred. For HDX-MS experiments, previous work on GCSF had been performed at physiological pH due to the intrinsic rate of exchange being 1000-fold faster at pH 7 than 4. However, GCSF has been shown to slowly but extensively aggregate when incubated at 37 °C, pH 7 (Raso et al. 2005), as such, demonstration of a low pH HDX method was of benefit to the study of this specific protein and its stability.

The final section of this chapter compared the current available solved 3D structures for GCSF to find a suitable candidate for computational modeling. The GCSF protein data bank (PDB) files were analysed for their amino acid sequence alignment, completeness, resolution and fit with the experimental characterisation data for purified GCSF.

3.1 Results

3.1.1 WT plasmid sequencing

The GCSF plasmid donated by Dr Adrian Bristow (NIBSC, Hertfordshire, UK) was extracted from host BL21 (DE3) *E. coli* cells and sequenced. It consisted of a constitutively induced pET21a(+) plasmid with T7 promoter and terminator sequences (Figure 12; Bristow et al. 2012). The gene sequence contained bases 133-654 of the human GCSF precursor accession code M1770, derived from the MIA PaCa-2 cell line, inserted between NdeI and EcoRI restriction sites (Figure 12; Bristow et al. 2012; Devlin et al. 1987). When aligned with the other CHU-2 sequence freely available on UniProt (Gene: *CSF3*) the DNA sequence differed by three amino acids. The *CSF3* gene had a 9 base pair insertion of the amino acids valine, serine and glycine prior to the cysteine at position 36 (UniProt 2017). This alternative gene sequence resulted from differential splicing of transcripts from a single GCSF gene in the CHU-2 cell line (Devlin et al. 1987).

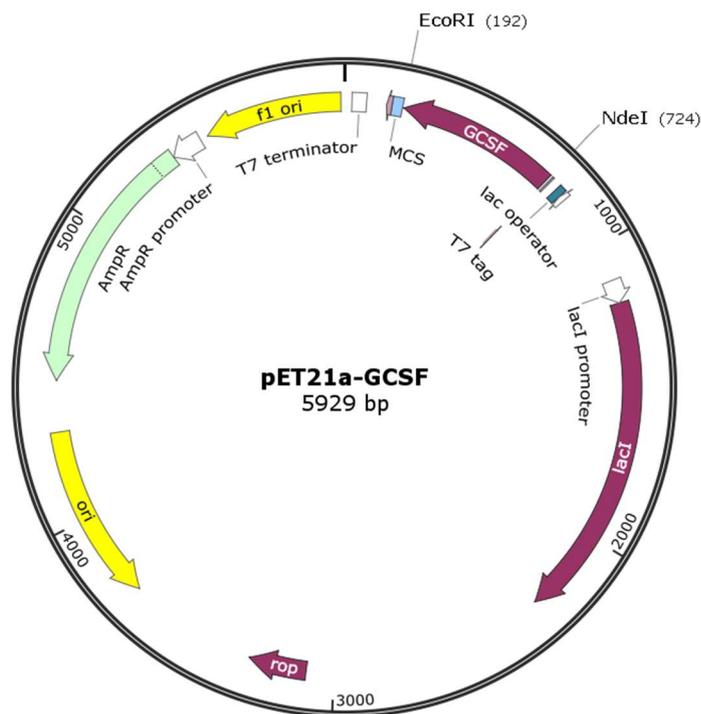


Figure 12. **GCSF pET21a plasmid schematic.** The insertion location of GCSF gene was between EcoRI and NdeI restriction sites, as labelled. Prepared using SnapGene™ (GSL Biotech LLC, IL, USA).

The GCSF gene DNA sequence was aligned with the original Bristow et al. (2012) DNA sequence and yielded a 99% match. The 1% discrepancy was due to three point mutations highlighted in Figure 13, however the single base mutations did not alter the overall GCSF amino acid sequence.

Score	Expect	Identities	Gaps	Strand
948 bits(513)	0.0	519/522(99%)	0/522(0%)	Plus/Plus
Query 1	ACACCTCTAGCCCTGCCAGCTCCCTGCCCCAGAGCTTCTGCTCAAGTGCTTAGAGCAA	60		
Sbjct 1	ACACCTCTAGCCCTGCCAGCTCCCTGCCCCAGAGCTTCTGCTCAAGTGCTTAGAGCAA	60		
Query 61	GTGAGGAAGATCCAGGGCGATGGCGCAGCGCTCCAGGAGAAGCTGTGTGCCACCTACAAG	120		
Sbjct 61	GTGAGGAAGATCCAGGGCGATGGCGCAGCGCTCCAGGAGAAGCTGTGTGCCACCTACAAG	120		
Query 121	CTGTGCCACCCCGAGGAGCTGGTGCTGCTCGGACACTCTCTGGGCATCCCCCTGGGCTCCC	180		
Sbjct 121	CTGTGCCACCCCGAGGAGCTGGTGCTGCTCGGACACTCTCTGGGCATCCCCCTGGGCTCCC	180		
Query 181	CTGAGCAGCTGCCCCAGCCAGGCCCTGCAGCTGGCAGGCTGCTTGAGCCAACTCCATAGC	240		
Sbjct 181	CTGAGCAGCTGCCCCAGCCAGGCCCTGCAGCTGGCAGGCTGCTTGAGCCAACTCCATAGC	240		
Query 241	GGCCTTTTCTCTACCAGGGGCTCCTGCAGGCCCTGGAAGGGATCTCCCCGAGTTGGGT	300		
Sbjct 241	GGCCTTTTCTCTACCAGGGGCTCCTGCAGGCCCTGGAAGGGATCTCCCCGAGTTGGGT	300		
Query 301	CCCACCTTGGACACACTGCAGCTGGACGTCGCCGACTTTGCCACCACCATCTGGCAGCAG	360		
Sbjct 301	CCCACCTTGGACACACTGCAGCTGGACGTCGCCGACTTTGCCACCACCATCTGGCAGCAG	360		
Query 361	ATGGAAGAAGTGGGAATGGCCCTGCCCTGCAGCCCACCCAGGGTGCCATGCCGGCCTTC	420		
Sbjct 361	ATGGAAGAAGTGGGAATGGCCCTGCCCTGCAGCCCACCCAGGGTGCCATGCCGGCCTTC	420		
Query 421	GCCTCTGCTTTCCAGCGCCGGGACAGGAGGGGCTGGTGCCTCCCATCTGCAGAGCTTC	480		
Sbjct 421	GCCTCTGCTTTCCAGCGCCGGGACAGGAGGGGCTTAGTGCCTCCCATCTGCAGAGCTTC	480		
Query 481	CTGGAGGTGTCGTACCGGTTCTACGCCACCTTGCCCAGCCC	522		
Sbjct 481	CTGGAGGTGTCGTACCGGTTCTACGCCACCTTGCCCAGCCC	522		

Figure 13. **Alignment of sequenced GCSF gene and original Bristow et al. (2012) DNA sequence.** Identical nucleotides are displayed with a line linking the two, whereas differing nucleotides are left with a blank space between and have been circled in red.

Due to bacterial expression an additional formylmethionine (fMet) residue was added to the start of the protein sequence. The confirmed amino acid sequence of GCSF is displayed below:

(M)TPLGPASSLPQSFLKCLEQVRKIQGDGAALQEKLCAATYKLCHEPEELVLLGHSLGIPWAPLSS
 CPSQALQLAGCLSQLHSGFLYQGLLQALEGISPELGPTLDTLQLDVADFATTIWQQMEELGM
 APALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVSRYRHLAQP

3.1.2 GCSF production

GCSF IBs were expressed during a 7.5 L *E.coli* fermentation. Sampling during the run showed OD₆₀₀ increased up to a maximum of 33.75 at 283 mins post inoculation (Figure 14). Two IPTG spikes were introduced during the fermentation, as highlighted, the first at 163 mins (OD₆₀₀ = 7.9) and the second two hours later at 283 mins (OD₆₀₀ = 33.75). Cells were harvested four hours after the first induction where the OD₆₀₀ had plateaued. (NR) SDS-PAGE of fermentation samples had non-leaky expression, as non-induced cells showed no obvious GCSF band at 18,800 Da, whereas samples after the two induction time points and harvest showed large bands at the same MW as the GCSF reference sample (Figure 14). In addition, no obvious degradation of the product was observed post first induction.

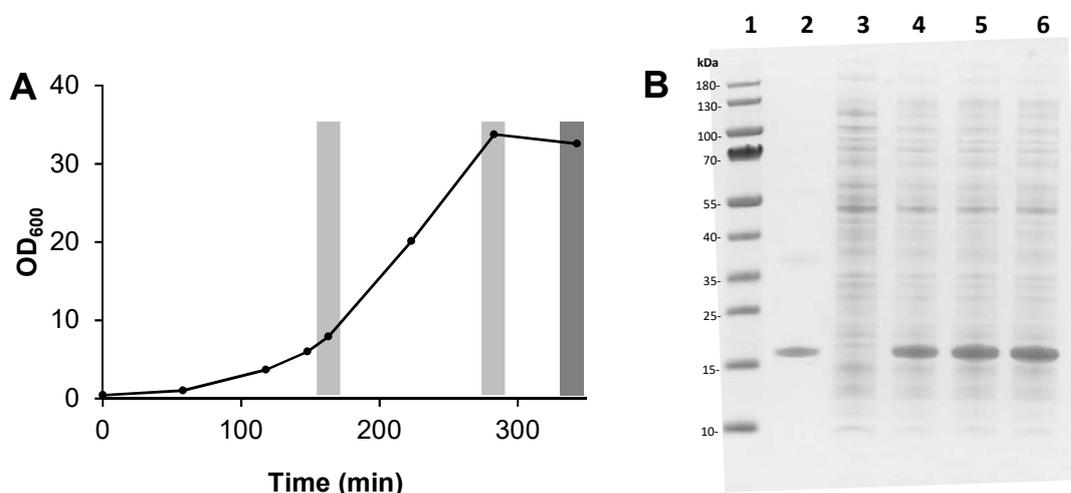


Figure 14. **BL21 (DE3) *E.coli* cell fermentation and GCSF expression.** A) OD₆₀₀ of cells during 7.5 L batch fermentation. Cells were grown on terrific broth supplemented with ampicillin and magnesium sulphate heptahydrate, and IPTG induced at two stages highlighted in light grey. Cells were harvested 4 hours after the first induction as highlighted in dark grey. B) Non-reduced SDS-PAGE of GCSF fermentation samples. From left to right the lane samples are: 1) molecular weight marker, 2) 0.1 mg/mL GCSF reference standard, 3) uninduced cells 148 mins into fermentation, 4) cells 1 hr post induction, 5) cells 2 hrs post induction, 6) cells 4 hrs post induction (harvest).

During SEC purification GCSF eluted at 240 mL with a peak height of 760 mAU (Figure 15). The refold solution containing 1 M arginine.HCl eluted in the column dead volume at 290 mL as seen by the increase in conductivity. The GCSF peak was collected in 3 mL fractions into 15 mL falcon tubes containing chilled Milli-Q® H₂O (Merck Millipore, Billerica, MA, USA) to dilute the sample buffer 5X to 10 mM sodium acetate pH 4.25.

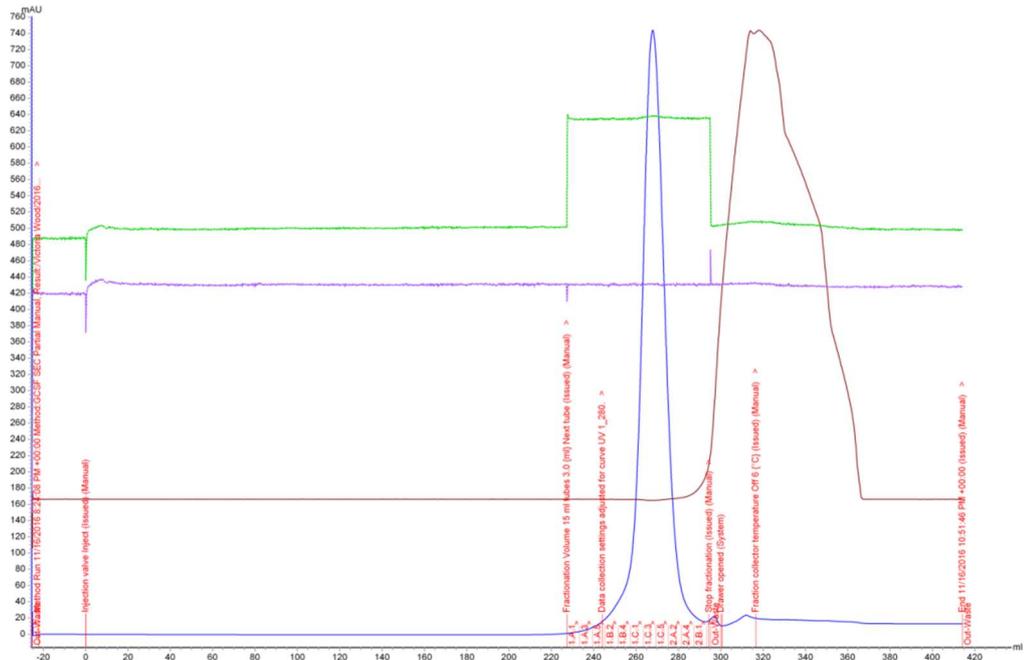


Figure 15. SEC chromatogram for standard GCSF purification. Curves displayed include A₂₈₀ absorbance in mAU (blue), conductivity (brown), pre-column pressure (green), DeltaC pressure (purple), and fractions and run log (red). Fractions were collected in serpentine order starting at 1A1.

SEC fractions were analysed by (NR) SDS-PAGE to inform fraction pooling. From Figure 16 the higher concentration samples contained lower MW species as seen by a band above the main GCSF band. Fractions equal to or greater than 0.1 mg/mL were selected for pooling. A GCSF reference standard ran alongside fractions had dimer and trimer bands present caused by the old age of the standard. All fractions in lanes 6-13 were pooled (Figure 16).

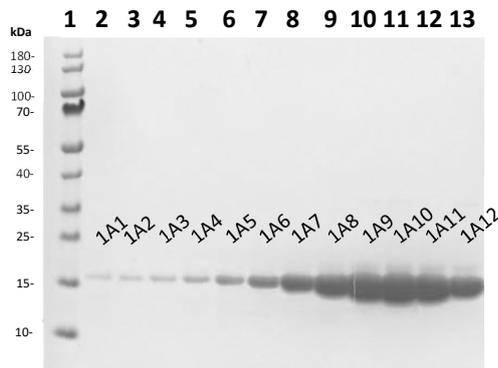


Figure 16. Non-reduced SDS-PAGE of SEC fractions. Fractions were analysed neat with 1 x LDS (Invitrogen) The first lane contains the PageRuler™ prestained molecular weight marker. Lanes 2 to 13 contain the fractions collected from 1A1 to 1A12.

3.1.3 Characterisation

3.1.3.1 Purity and identity of GCSF

By SEC-HPLC analysis, the pooled GCSF fractions contained peaks identified as multimeric, dimeric and monomeric GCSF as well as the formulation buffer (Figure 17A). The monomer eluted at 23 min and was the most abundant species making up 98% of the sample (Figure 17B).

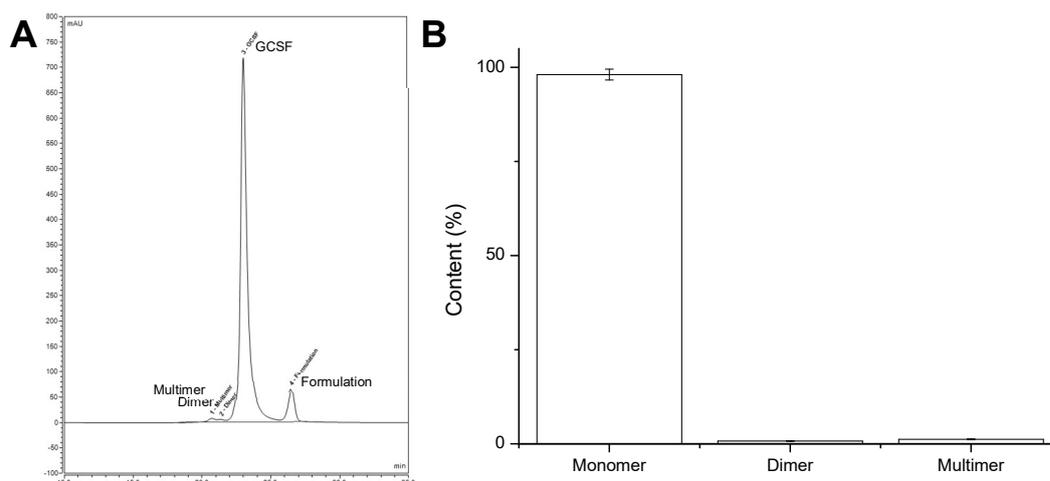


Figure 17. SEC-HPLC analysis of GCSF fraction pool. Analysis performed on a TSK3000 swxl column with an isocratic buffer containing 0.1 M phosphate pH 2.5 and a flowrate of 0.5 mL/min. GCSF sample injected at 25 μ L sample volume at 0.6 mg/mL. Absorbance was measured at A280 and A214. A) Full SEC-HPLC chromatogram. B) Average peak area as a % of the total sample content not including the formulation buffer.

By intact protein analysis, peaks observed by SEC-HPLC were further separated into subsets of different MW species. Monomeric GCSF was deconvoluted to obtain a MW of 18,799 Da, which matched the theoretical mass exactly (Figure 18). This species made up 70% of the sample. The second largest species was also monomeric GCSF with a MW of 18,820 Da. This was +21 Da greater than the theoretical mass and suspected to be monomeric GCSF with an ESI molecular ion adduct such as $M+ACN+2H$, which yields a mass increase of +21.52 (Novatia 2017). Three MW species were also identified in the dimeric 37,598 Da region making up around 3.6% of the sample (Figure 18), which was higher than that determined by SEC-HPLC.

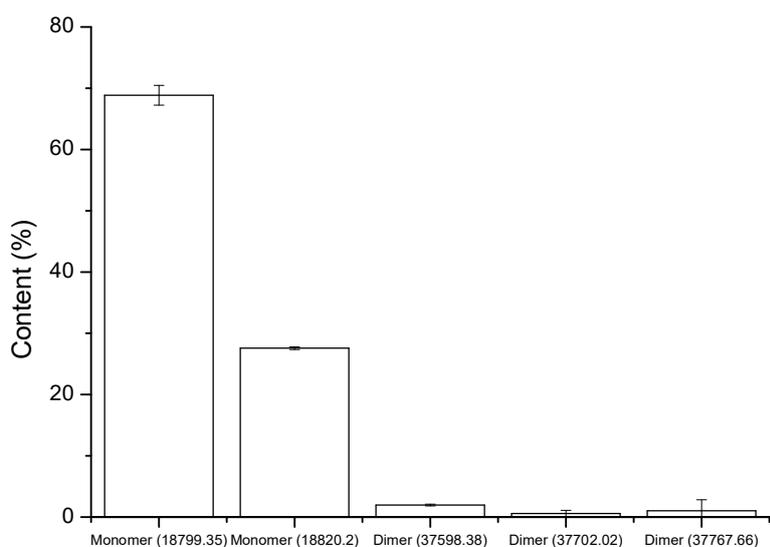


Figure 18. **Intact protein mass analysis of GCSF fraction pool.** Molecular weight species within the GCSF fraction pool sample were identified using an ultra-performance liquid chromatography (UPLC) system coupled with a quadrupole time of flight (QTOF) mass spectrophotometer (MS) instrument with an electrospray ionisation source. 1 μ L GCSF at 0.2 mg/mL, 10 mM sodium acetate pH 4.25 was injected and the mass spectra acquired at 100–3000 m/z with 3 spectra/s, over 333.3 ms/spectrum, 3225 transients/spectrum, and 0 V collision energy. The TIC peak was extracted in 2 min windows from 6-8 min. Each window spectra was deconvoluted, m/z to MW, to identify monomeric and dimeric masses

3.1.3.2 *Unfolding using mass spectrometric methods*

Collision induced unfolding of GCSF was performed in order to develop an orthogonal unfolding characterisation method to fluorescence. MS readings of the sample were taken at 5 V (folded) for 1 min acquisitions pre- and post- CIU-IMS-MS to determine changes to the folding of the protein over time (Figure 19). For GCSF in 50 mM ammonium acetate pH 4.25 the charge state distribution was relatively small indicating good folding of the intact protein. In general, tightly folded proteins will carry only a small number and a narrow range of charges during electrospray ionization, whereas unfolded states produce a heterogeneous highly charged population (Eyles and Kaltashov 2004). In addition, the distribution of charges did not change from the start to end of the experiment meaning any changes during unfolding experiments was not caused by sample degradation (Figure 19).

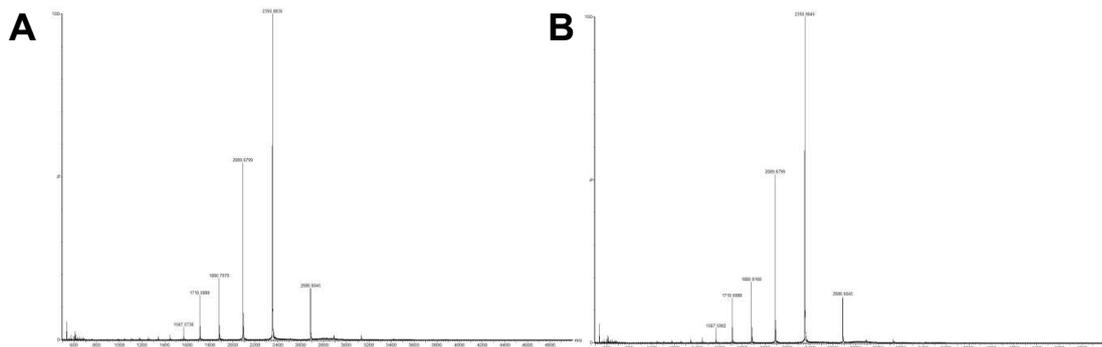


Figure 19. **Folded 5V MS spectra of GCSF in 50 mM ammonium acetate pH 4.25.** Spectra was obtained using a SYNAPT G2 High Definition Mass Spectrometer (HDMS) equipped with a NanoLock ZSpray Exact Mass Ionization Source and MassLynx data processor. The GCSF samples at 0.3 mg/mL were infused to the standard electrospray (z-spray) source using a copper capillary apparatus at a rate of 10 μ L/min. The capillary of the ESI source was held at 3 kV, with the source operating in positive ion mode. The sample cone was operated at 5 V. The TOF-MS was operated over the scanning range of m/z 500–5000 at a pressure of 1.8×10^{-6} mbar. A) Pre-, B) Post- CIU-ESI-MS experiments.

During CIU-IMS-MS, GCSF was activated in the collision cell by a 3 V stepwise increase of the acceleration voltage from 5 V to 27 V. MSMS data from each step was collected in 2 min acquisitions and the applied voltages plotted against the drift time to visualise unfolding and eventually dissociation of the protein (Figure 20A). At 5 V the protein was relatively folded and had an average drift time of 9.7 ms, as trap voltage increased the average drift time increased up to a maximum of 13.56 ms at 27 V. Additionally, there was an intermediate unfolding peak with a drift time around 12 ms. An equilibrium of folded, intermediate and unfolded species occurred at 21 V, where the intermediate species was the most abundant with an equal distribution of folded and unfolded species at either side (Figure 20A). At the highest voltage applied GCSF had not fully unfolded as the intermediate peak was still present but the folded peak had disappeared. This suggested a higher voltage above 27 V should have been applied to see full GCSF unfolding.

Using the drift time data displayed in Figure 20A, a 3D contour plot of ion intensity as a function of activation voltage and drift time, termed a collision induced unfolding (CIU) fingerprint CIU was prepared and displayed in Figure 20B.

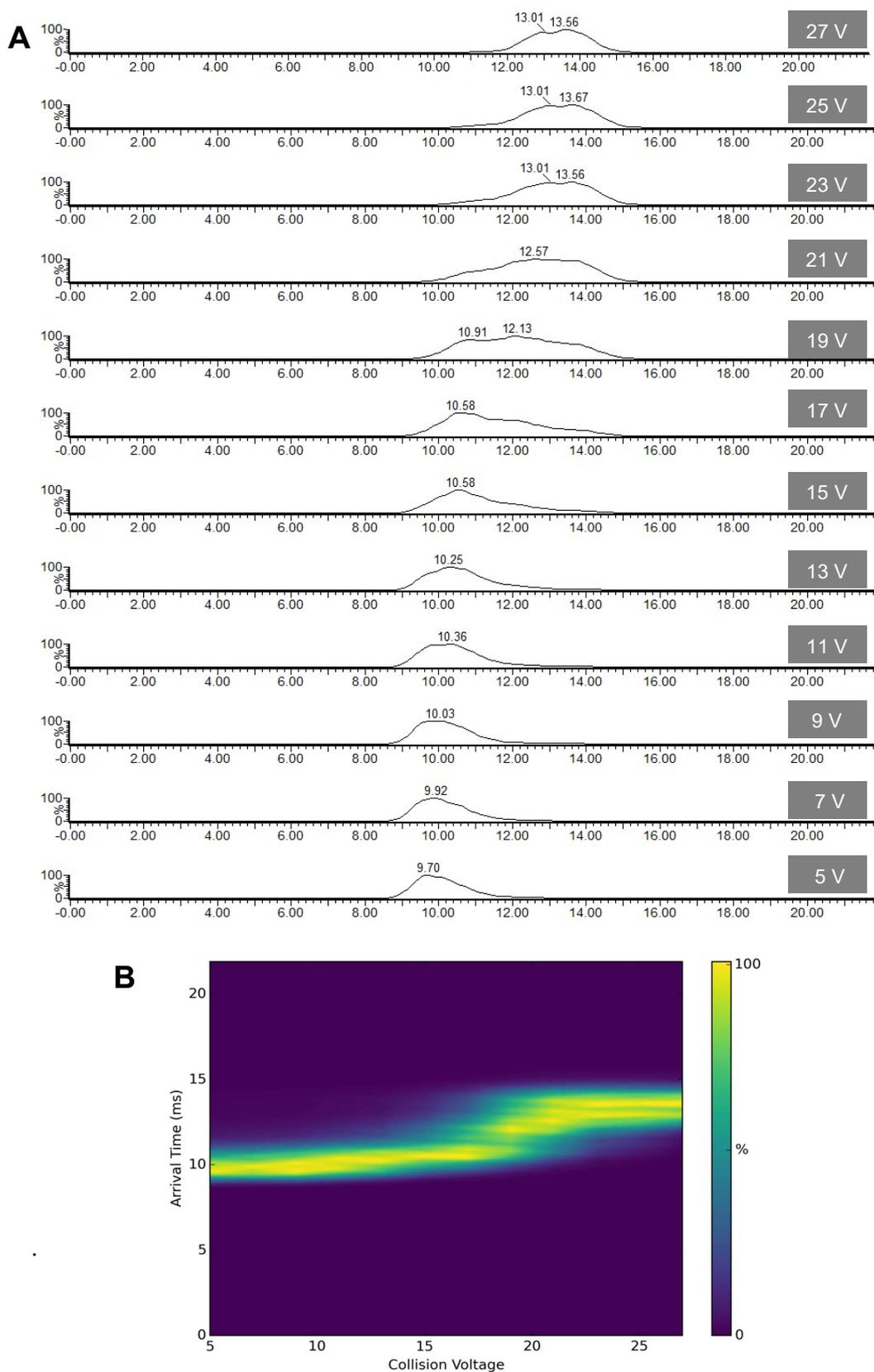


Figure 20. **Unfolding of GCSF as measured by collision induced unfolding ion-mobility spectrometry-mass spectrometry (CIU-IMS-MS).** A) Shift in drift time of 0.2 mg/mL GCSF in 50 mM ammonium acetate pH 4.25 during a stepwise increase in collision voltage. The collision voltage was ramped from 5 to 27 V in 3 V increments (dark grey box). Intact mass data were recorded for MS-isolated ions at each collision voltage, with an acquisition time of 2 mins, and the resulting chromatograms were smoothed by the Savitzky Golay method. The x-axis represents drift time in ms, and the drift time of GCSF at each V is labelled. B) GCSF CIU fingerprint plotted using the Protein Unfolding for Ligand Stabilisation and Ranking (PULSAR) software (Allison et al. 2015), with collision voltage hierarchy, and a mass fitting of 18799.1, Z_{avg} of 8 and $zwidth$ of 0.2 and a viridis colour map.

3.1.4 Development of a peptide-level HDX-MS method for GCSF

In order to perform peptide-level HDX-MS of GCSF, the digestion step conditions needed to be optimised to generate a peptide map with > 95% sequence coverage.

3.1.4.1 Peptide map optimisation

The varying concentration of gdn.HCl and TCEP in the quench solution produced a range of peptide map sequence coverage % ranging from 68 to 97% (Table 8). Digestion factors were calculated from the coverage percentage, cleavage rate percentage and redundancy and compared to identify the optimal quench solution. It was found that both denaturant and reducing agent were required in the quench solution as absence of either produced low DFs, and this was most evident during the absence of reducing agent TCEP where there was no coverage around the disulphide bond forming cysteines at positions 37, 44, 65 and 75 (Figure 21A). The highest DF, 11.65, contained 4 M Gnd.HCl and 0.6 M TCEP and resulted in sequence coverage of 96.6% (Figure 21B).

Table 8. **Optimising HDX-MS quench solution.** Varying the concentration of chaotropic agents, TCEP and Gnd.HCl, in quench solution (10mM Sodium Phosphate pH 2.5) to find the optimal digestion factor (DF). The DF was calculated using the equation $DF = 8y_1 + 5y_2 + 0.6y_3$, where y_1 is coverage, y_2 is cleavage rate and y_3 is redundancy (Ahn 2013). Highlighted are the two samples taken forward for further optimisation.

Rank	DF	TCEP (mM)	Gnd.HCl (M)	No. peptides	Coverage (%)	Cleavage rate (%)	Redundancy
1	11.65	600	4	63	96.6	38.8	335.4
2	11.51	1000	1	67	92.6	41.1	341.1
3	11.27	50	3	62	93.1	37.7	322.3
4	11.23	50	4	59	95.4	34.9	309.1
5	10.79	600	3	62	88	36	325.7
6	10.68	1000	0	60	89.1	37.1	282.3
7	10.65	600	0	59	89.1	37.1	276.6
8	10.37	300	2	55	88	34.3	269.1
9	10.23	0	4	59	82.9	30.9	343.4
10	9.99	50	1	54	84.6	33.1	261.7
11	9.89	600	1	54	82.9	33.1	268
12	9.58	0	3	52	79.4	30.3	284.6
13	9.26	50	0	44	82.9	29.1	195.4
14	7.65	0	0	37	68	22.9	178.3
15	7.65	0	1	37	68	22.9	178.3

From the DF analysis, the concentration of chaotropic agents in the quench buffer appeared to not have an optimal value but rather plateaued. From the data in Table 8 it was observed that a high DF of 11.27 could be obtained from a more conservative concentration of chaotropic agents at 0.05 M TCEP and 3 M Gnd.HCl. As such, the highest DF and the conservative DF quench solutions were taken forward for further digestion optimisation experiments.

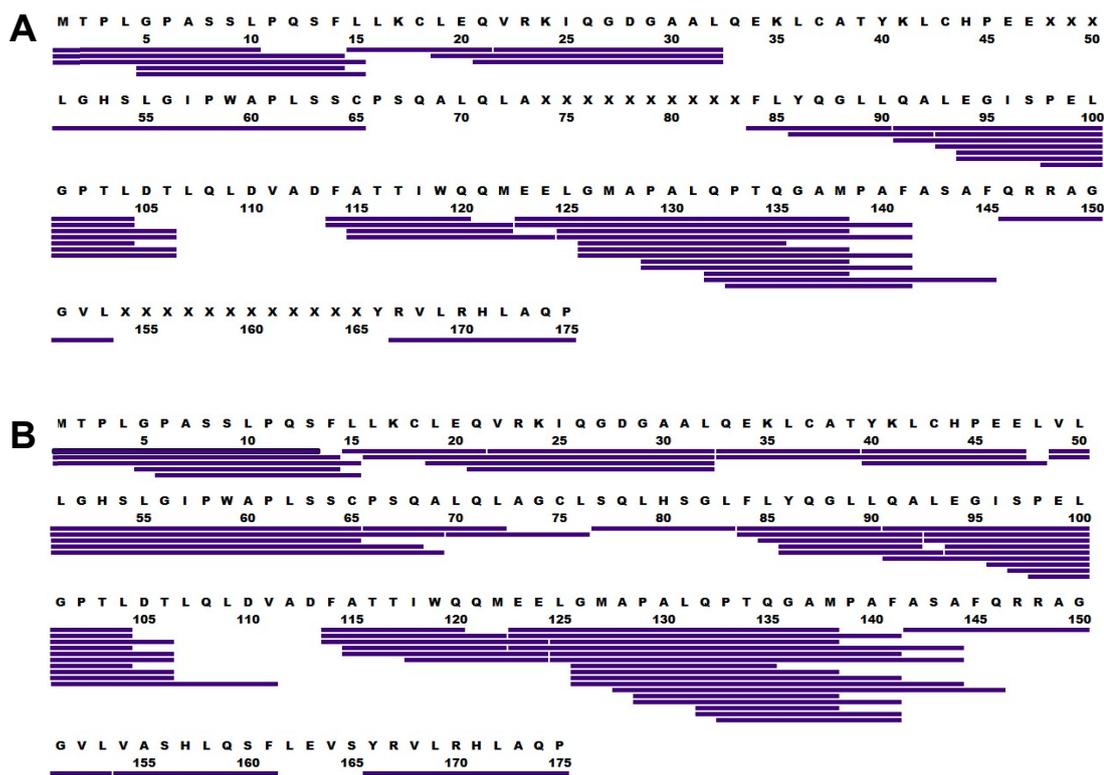


Figure 21. **GCSF Peptide maps pre- and post- quench solution optimisation.** The top line is the amino acid sequence of GCSF with additional methionine residue at the N-terminal, second line is residue number, and third line is the exact positions of peptides produced. A) No chaotropic agents added to quench solution, B) Optimised quench solution containing 4 M Gnd.HCl and 600 mM TCEP.

3.1.4.2 Quench buffer and column pressure optimisation

It had been previously reported that TCEP is unstable in phosphate buffers for more than 24 hrs (Houde and Engen 2013). Consequently, old (> 24 h) and fresh (< 1 h) preparations of quench buffer containing 0.05 M TCEP and 3 M Gnd.HCl in 100 mM sodium phosphate pH 4.25 were compared. Peptide maps were generated three times for each preparation and the average result recorded in Table 9. It was found, as expected, that the old sample

produced a lower DF than the fresh preparation of quench solution. From this result it was decided to change the quench solution buffer from phosphate to sodium acetate.

Table 9. **Optimising the processing parameters during pepsin digestion step of HDX-MS.** Comparison of optimum and conservative concentrations of chaotropic agents in quench solution. Sample A represents an old sample (>24 h) and sample B represents a fresh sample, both in 100 mM sodium phosphate pH 2.5 buffer. Sample C represents a fresh sample in 50 mM sodium acetate buffer pH 2.5 with and without flow restriction. Digestion factor (DF) for each quench solution was calculated using the equation $DF = 8y_1 + 5y_2 + 0.6y_3$, where y_1 is coverage, y_2 is cleavage rate and y_3 is redundancy (Ahn 2013).

Sample	TCEP (M)	Gnd.HCl (M)	Restrictor length (mm)	Coverage (%)	Cleavage rate (%)	Redundancy	DF
A	0.05	3	152	84	30	204	9.43
B	0.05	3	152	90	32	255	10.35
C	0.6	4	152	86	33	280	10.17
C	0.6	4	0	97	35	309	11.35

Peptide maps were next generated using a quench solution containing sodium acetate buffer and 0.6 M TCEP, 4 M Gnd.HCl, with and without a 152 mm capillary tube at the BEH column outlet, effectively adding and removing flow restriction. The addition of flow restriction increased back pressure. Although previously thought to improve digestion, backpressure decreased DF values for GCSF as seen in Table 9, sample C. Without the flow restrictor the sodium acetate buffer and high concentration of chaotropic agents produced the highest DF with an average sequence coverage of 97% and were used for all subsequent HDX-MS experiments.

3.1.5 Effect of pH on GCSF HDX

The effect of exchange and sample solution pH on GCSF was explored by comparing pH 4 and 7. HDX-MS experiments were performed in triplicate for each pH to also assess repeatability. A total of 59 peptides were identified by the GCSF peptide map as common to both pH 4 and 7 experiments. The relative uptake was calculated for each peptide for the different labelling time points. For repeatability assessment, the sum of all time point uptake values per peptide was calculated and displayed in Figure 22A for pH 4, and Figure 22B for pH 7. The standard deviation of each peptide is also displayed and was calculated by taking the sum of the standard deviations calculated for time point triplicates within each run. The data showed good repeatability between values for both experimental conditions. The level of total relative uptake was higher for the pH 7 data set, as expected, due to the increase in the intrinsic rate of HDX. In general, the total relative uptake was highest for the first run of both data sets and decreased with consecutive runs. As the same deuterium solution was used for each run, the data indicated the concentration of available deuterium ions in solution decreased over time. According to the manufacturer recommendation (Merck KGaA, Darmstadt, GE), deuterium oxide should be stored under an inert gas like nitrogen to prevent hydrogenated water uptake. From this information it was presumed that deuterium ions were lost via extended exposure to air, which caused the decrease in total relative uptake over time.

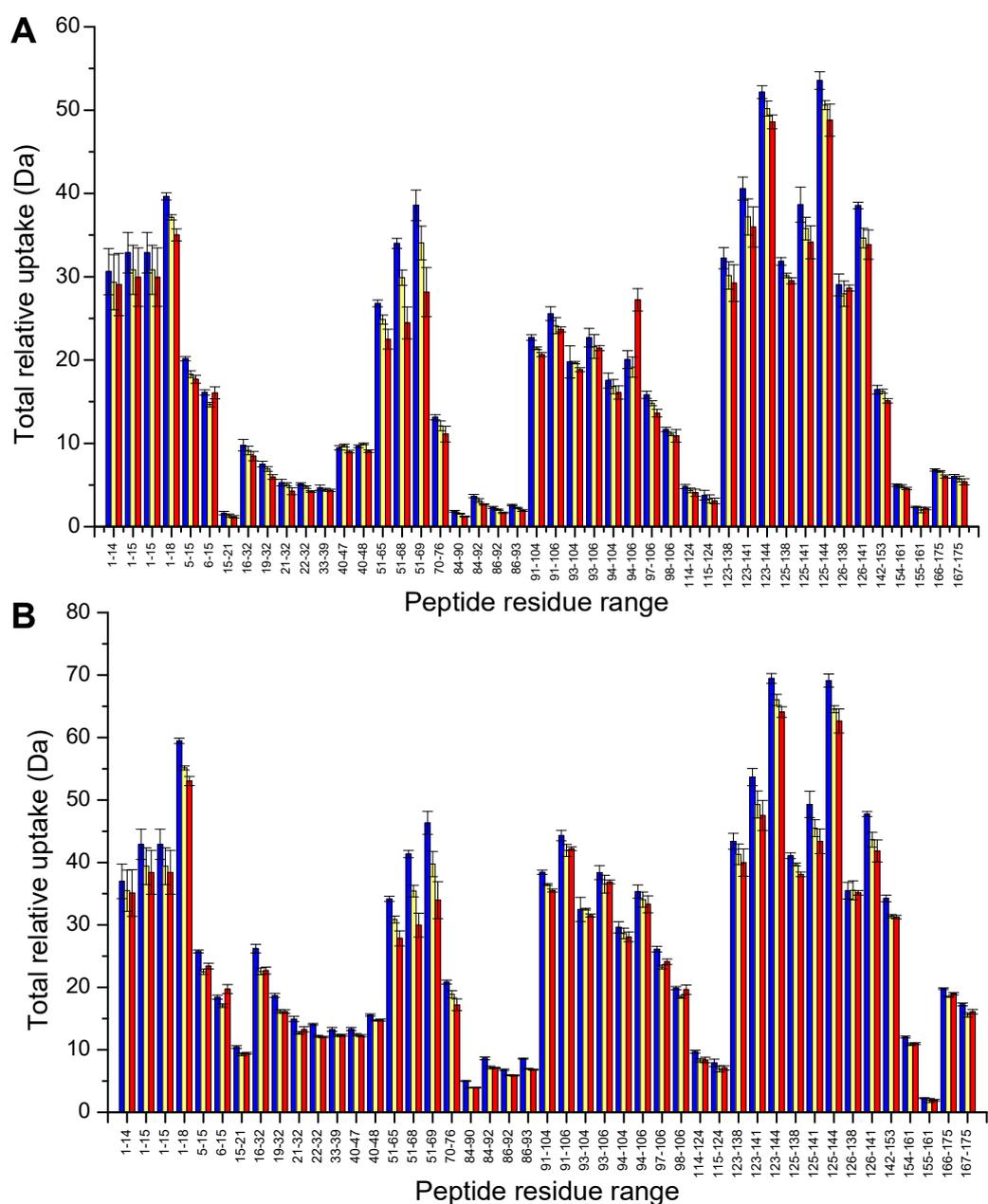


Figure 22. **Repeatability of GCSF peptide relative uptake as measured by HDX-MS.** Peptide-level HDX-MS was performed with GCSF, where the sample and deuterium solution pH/pD was varied A) 50 mM sodium acetate pH/pD 4.25 and B) 10 mM PBS pH/pD 7.4. Each pH/pD experiment was repeated three times to assess repeatability of measurements: run 1 (blue), run 2 (yellow) and run 3 (red). The y-axis denotes the sum of the relative uptake calculated from the change in mass between each undeuterated and deuterated GCSF peptide for seven incubation times (20s, 2.6 min, 5 min, 15 min, 1 hr, 4 hr, 8 hr). The standard deviation was calculated by taking the sum of the standard deviations calculated for time point triplicates within each run. The x-axis denotes the ordinal peptide number, a sequential arrangement of 54 peptides of GCSF by the midpoints in the sequence. Appendix 4 contains the GCSF peptide sequence, residue numbers and locations.

The relative uptake of GCSF peptides in deuterium over time from the second run of both exchange conditions is displayed in Figure 23. For pH 4.4 the uptake rate of peptides was slower, and therefore measurable in the loop regions of the protein, whereas the structured helical regions did not exchange as seen by a flat line across all time points. Conversely, for pH 7.4, the uptake rate was measurable for the structural helical regions, whereas the loop regions reached their maximum exchange level by the first labelling time point as seen by an overlay of uptake at all time points in these regions. The complementarity of the pH conditions, showing changes in uptake rate for the different types of GCSF structures, highlights the power of using different pH solutions to increase the time window for HDX.

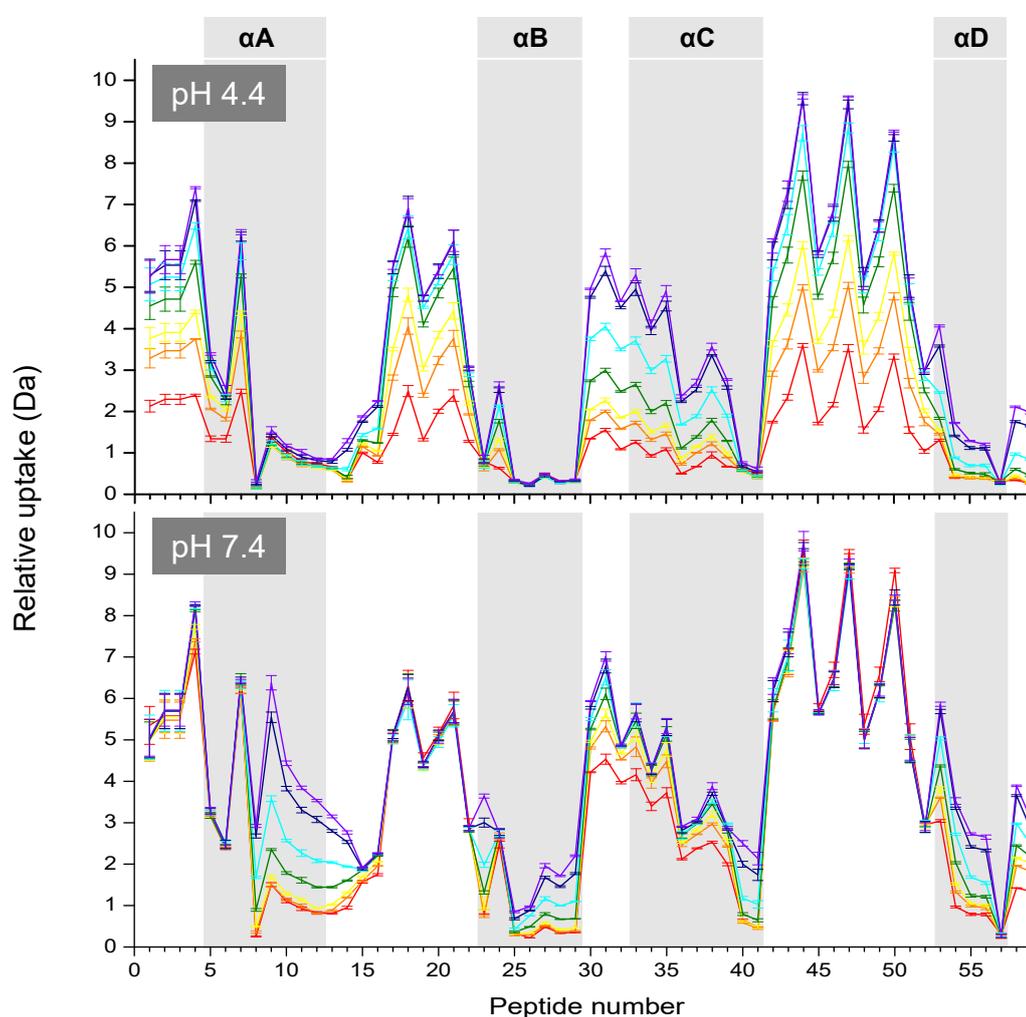


Figure 23. **Relative uptake of GCSF peptides over time with differing solution pD and sample pH as measured by HDX-MS.** The average deuterium uptake for each of 54 peptides of GCSF for seven exchange times (20s, 2.6 min, 5 min, 15 min, 1 hr, 4 hr, 8 hr) in 10 mM sodium acetate 99% deuterium solution with differing pD is displayed. Time points are displayed in rainbow colour, with the earliest time point in red and the latest in violet. Each uptake data point is an average of three independent H/D-exchange experiments with GCSF in either 50 mM sodium acetate pH 4.25 or 10 mM PBS pH 7.4 . The y-axis denotes the relative uptake calculated from the change in mass between the undeuterated and deuterated peptides. The x-axis denotes the ordinal peptide number, a sequential arrangement of 54 peptides of GCSF by the midpoints in the sequence. The different helical regions of GCSF are coloured in the background. The non-coloured regions represent the connecting loop regions. Appendix 4 contains the GCSF peptide sequence, residue numbers and locations.

To determine if the changes to uptake for different pH solutions was purely due to changes in intrinsic exchange or if changes to protein dynamics also occurred, the exchange time for pH 4.4 samples was converted into the corresponding exchange time for pH 7.4 at 22 °C (Table 10). Converting the time window of the HDX saw the longest labelling time of 8 hrs at pH 4.4 was almost equivalent (28.8 s) to the shortened labelling time of 30 s at pH 7.4. As only the first and last time points of the two sets of data were within the same converted time window, differentials between the two time point data sets were calculated for each peptide, for each of the three runs (Figure 24). In theory, after exchange time conversion, if GCSF protein dynamics does not change, the differential uptake value should be zero. From the differential plots in Figure 24, values significantly differed between the two pH conditions indicating changed protein dynamics. Regions significantly affected concurred across all three HDX runs and included α A, LoopAB (including the short helix), loopBC, α C, loopCD, α D and loopD. Run 1 showed a difference to the other two run differentials around the LoopAB region, where there was a significant positive differential indicating high uptake in the pH 7.4 samples at this region during this run. The majority of other GCSF peptides showed a negative differential across all runs, indicating the pH 4.4 sample had a higher relative uptake of deuterium. Regions showing the largest negative differential included loopAB, α C, and loopD. Interestingly, the loopCD region showed an increased differential.

Table 10. **HDX reaction condition and exchange time corrected to standard HDX condition of pH 7.4 and 22 °C.** Time in seconds was corrected to pH 7.4 at 22 °C using the method described by Coales et al. (2010).

Exchange time corrected to standard condition of pH 7.4 and 22 °C (s)	HDX reaction condition	
	pH 4.4 at 22 °C (s)	pH 7.4 at 22 °C (s)
0.03	30	
0.15	150	
0.3	300	
0.9	900	
3.6	3600	
14.4	14400	
28.8	28800	
30		
150		150
300		300
900		900
3600		3600
14400		14400
28800		28800

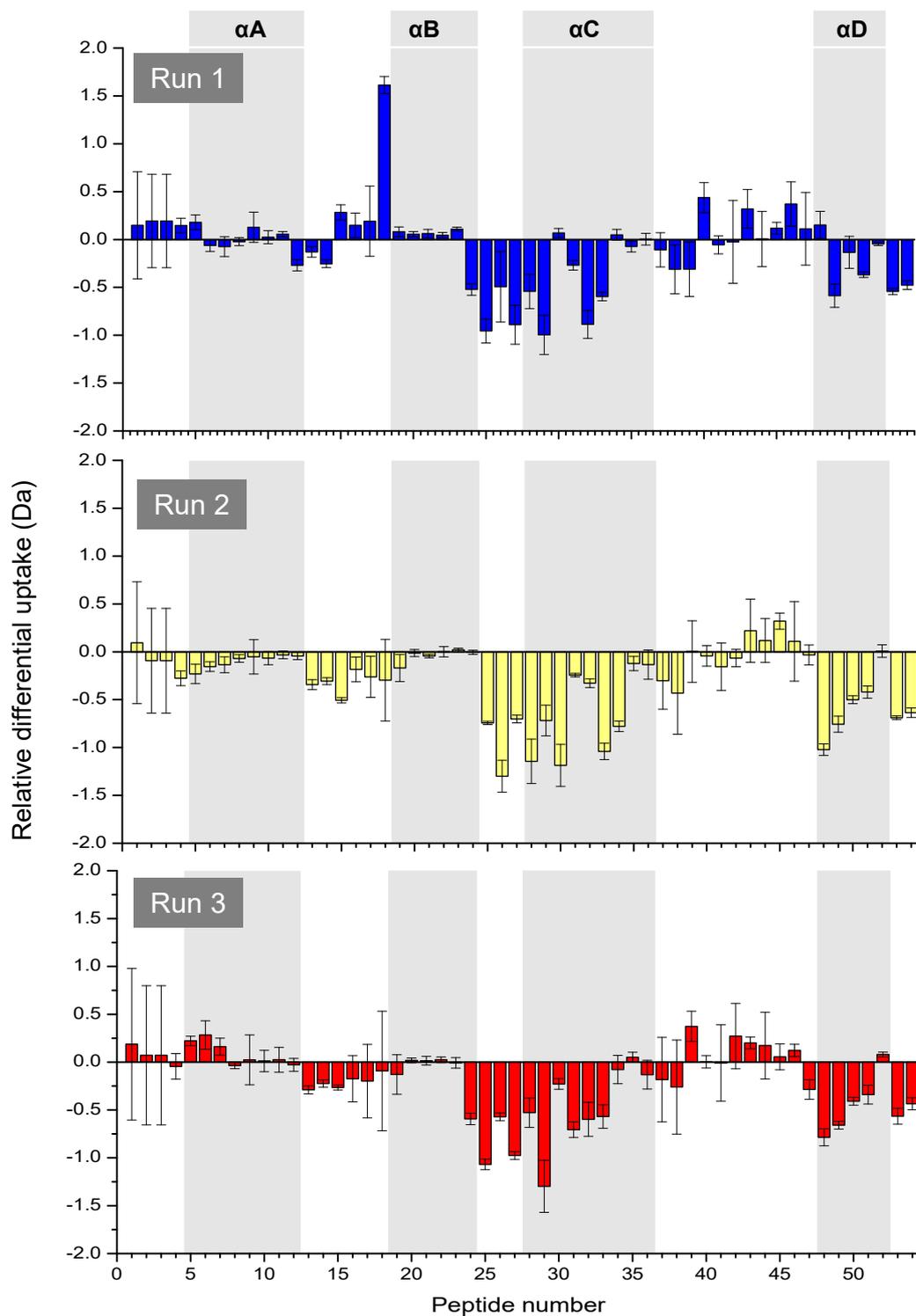


Figure 24. **Differential plots of GCSF peptide uptake in differing pD solutions, corrected to a standard exchange time, as measured by HDX-MS.** The relative uptake values were taken from GCSF peptide-level HDX-MS with pD 4.4 deuterium solution with 8 hrs exchange time and pD 7.4 deuterium solution with 30 s exchange time. The y-axis denotes $\Delta D(t) = m_{pD\ 4.4}(t) - m_{pD\ 7.4}(t)$, where m denotes the mass of the peptide as a function of deuterium exposure time, t . Each figure is an independent HDX-MS experiment comparing the two pD solutions as shown by the run order displayed in the dark grey box in the top left hand corner of each figure. The x-axis denotes the ordinal peptide number, a sequential arrangement of 54 peptides of GCSF by the midpoints in the sequence. The different helical regions of GCSF are coloured in the background. The non-coloured regions represent the connecting loop regions. Appendix 4 contains the GCSF peptide sequence, residue numbers and locations.

3.1.6 Effect of buffer and protein concentration on thermal stability

Buffer and protein concentration were varied to determine the effects on GCSF aggregation onset (T_{agg}) and melt (T_m) temperatures during thermal denaturation. Three different concentrations of sodium acetate at pH 4.25 were selected ranging from 10 to 100 mM, and four concentrations of protein ranging from 0.25 to 1.00 mg/mL. During the dialysis into 50 mM and 100 mM, samples were diluted below 1 mg/mL, therefore only the lower three concentrations could be analysed for these buffer concentrations. Values of T_m and T_{agg} were plotted in Figure 25A and B, respectively and a linear fit for each buffer concentration calculated (Table 11). For T_{agg} , a low concentration of acetate was found to have a higher T_{agg} with R^2 values of 0.99 (Table 11). The standard deviation at 10 mM was also lot lower than the 100 mM samples. This suggested that GCSF had higher thermal stability at the lower buffer concentration. Protein concentration was also shown to have an effect on T_{agg} , where as the concentration of protein increased, the T_{agg} decreased. For the 50 mM and 100 mM buffer concentrations, as the protein concentration increased the difference in T_{agg} values also decreased.

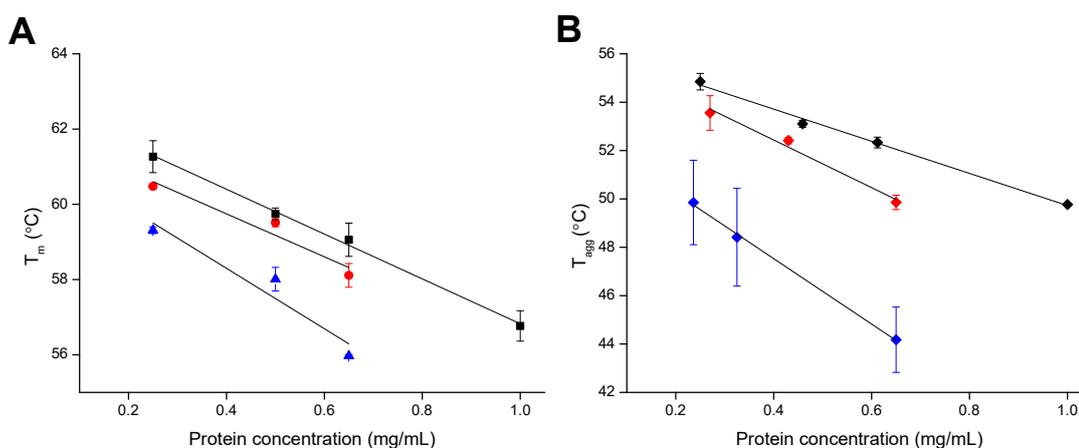


Figure 25. **Effect of protein and buffer concentration on GCSF thermal denaturation.** Different concentrations of GCSF were formulated in different concentrations of sodium acetate pH 4.25 buffer: 10 mM (black), 50 mM (red) and 100 mM (blue). A thermal ramp was applied to samples at 1 °C/ min and fluorescence (peak height) and SLS (at 266 nm) measured to determine A) T_m and B) T_{agg} , respectively.

Table 11. Effect of protein and buffer concentration on GCSF thermal denaturation linear fitting

Figure	Buffer conc. (mM)	Intercept		Slope		Statistics
		Value	Standard error	Value	Standard error	Adj. R-Square
A	10	62.78	0.147	-5.96	0.224	0.996
	50	62.03	0.717	-5.71	1.449	0.879
	100	61.51	1.102	-8.03	2.227	0.857
B	10	56.37	0.215	-6.65	0.336	0.992
	50	56.37	0.598	-9.84	1.257	0.968
	100	52.94	0.241	-13.5	0.547	0.997

Analysis of T_m was found to have the same pattern as for T_{agg} when varying buffer and protein concentrations. T_m values were found to be higher than T_{agg} , as T_{agg} measures aggregation onset whereas T_m is the midpoint of the melt curve. The variability of T_m measurements for samples at the highest buffer concentration was significantly larger than the other concentrations studied indicating a large degree of GCSF instability. This is consistent with previous GCSF studies where aggregation was dependent on diffusion-limited biomolecular collision (Krishnan et al. 2002; Robinson et al. 2017).

3.1.7 Optimised peptide level HDX-MS for GCSF

Using the optimised HDX-MS digestion method and stability analysis of GCSF, peptide-level HDX-MS of GCSF formulated in 10 mM sodium acetate pH 4.25 was performed. The peptide map generated 67 unique peptides. The HDX at seven different labelling time points was performed with a pH 4.25 D₂O solution and the relative uptake plotted for each peptide in Figure 26. From Figure 26, it can be seen that there is an increase in relative deuterium intake over time for most regions of GCSF, apart from the central regions of α helices. The centres of the helices do not generally undergo exchange as they form part of the hydrophobic core and are completely solvent inaccessible. The level of relative uptake increases with time, as expected. The loop regions showed a higher initial uptake compared to the helices. This was also expected as the loops are unstructured regions and fully solvent accessible, whereas the helices have structure, which reduces the availability of hydrogens for exchange. Whilst the helices had a lower exchange rate to loops, there were noticeable differences the level and rate deuterium uptake between the four α -helices. α A underwent the least amount of exchange compared to the other helices suggesting this region is more protected, or rigid. The N-terminal regions of α C and α D, and to some extent α B, were flexible, exchanging overtime to a high level of relative uptake. This

increased exchange is most likely due to disruption by the high exchanging, highly flexible loops they are attached to.

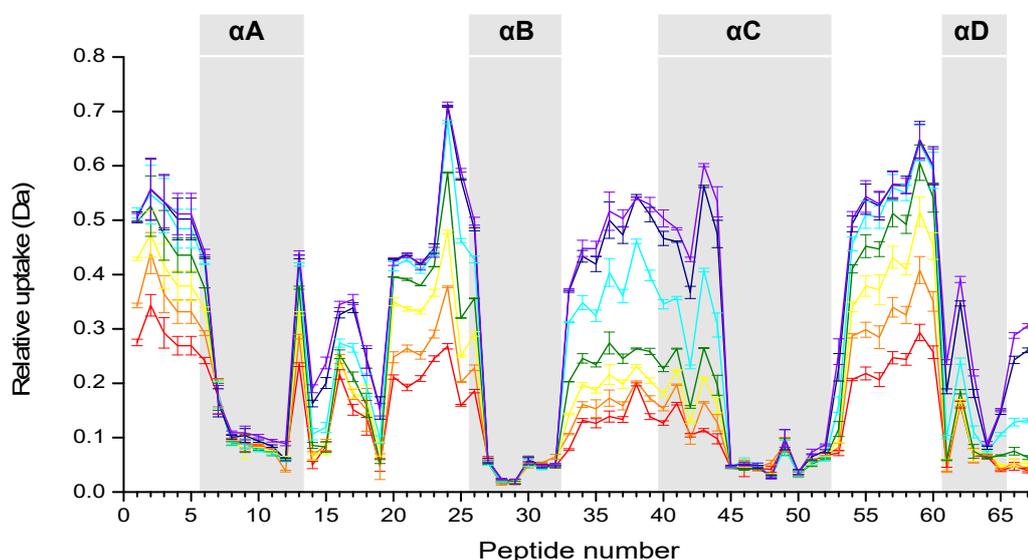


Figure 26. **Relative deuterium uptake of GCSF in 10mM sodium acetate pH 4.25 as measured by HDX-MS.** The average deuterium uptake for each of the 67 peptides of GCSF for seven exchange times (20s, 2.6 min, 5 min, 15 min, 1 hr, 4 hr, 8 hr) in 10 mM sodium acetate 99% deuterium solution, pD 4.25 is displayed. Time points are displayed in rainbow colour, with the earliest time point in red and the latest in violet. Each uptake data point is an average of three independent H/D-exchange experiments with GCSF in 10 mM sodium acetate pH 4.25. The y-axis denotes the relative uptake calculated from the change in mass between the undeuterated and deuterated values. The x-axis denotes the ordinal peptide number, a sequential arrangement of 67 peptides of GCSF by the midpoints in the sequence. The different helical regions of GCSF are coloured in the background. The non-coloured regions represent the connecting loop regions. Appendix 5 contains the GCSF peptide sequence, residue numbers and locations.

3.1.8 Structure selection for in silico research

A RCSB search for “granulocyte-colony stimulating factor” (08 Dec 2016; www.rcsb.org/pdb/home/home), identified eight solved GCSF structures for evaluation (Table 12). The amino acid sequences for each PDB were extracted and aligned to the sequenced GCSF amino acid sequence for similarity % assessment, where it was found that 2D9Q, 1CD9, 1PGR and 1RHG had 100% similarity.

Assessment of structures produced via X-ray diffraction was performed by comparing the resolution, structure completeness, and R-factor %. Resolution is the minimum spacing of crystal lattice planes that still provide measurable diffraction of X-rays (Wlodawer et al. 2008). The lower the Å value the better the resolution, as the smaller spacing increases the number of independent reflections available to define the structure. All structures for GCSF, apart from 1PGR, were within the range of 1.7 to 2.8 Å, which is considered medium-

to-high quality. 1BGC had the best resolution with 1.7 Å, followed by 1BGD and 1BGE, however all three of these structures were not human derived GCSF. The best resolution for human derived GCSF was 1RHG at 2.2.

R-factors measure the global relative discrepancy between the experimentally obtained structure factor amplitudes and the calculated structure factor amplitudes obtained (Wlodawer et al. 2008). Well-refined structures are expected to have an *R-factor* < 20%. 1BGE had the lowest *R-factor* at 19.3%, 1PGR had the highest at 31.7%, and of the human derived GCSF 1RHG had the lowest *R-factor* at 21.5%.

B-factor data was also extracted from the X-ray diffraction structures, averaged for each amino acid (not including H) and normalised. The individual normalised residue values for each GCSF structure is displayed in Figure 27A along with the locations of residues within the GCSF secondary structures. An average of all individual normalised residue B-Factors, not including 1BCG, is displayed in Figure 27B. Regions of high and low average B-Factor values agreed between all structures: high values appear at the ends of the protein as well as around the AB loop and CD loop. The lowest values were located within the hydrophobic core of all four α -helices.

Table 12. All available PDB structures for GCSF on RCSB website in 2016. PDBs were evaluated against one another for their type of structure (method), resolution (res) and R-factor values for crystal structures, similarity to the Bristow et al. (2012) GCSF amino acid sequence and finally their correlation between the average B-factor values and HDX-MS data (Appendix 3).

PDB	Title	Method	Res. (Å)	R-Factor (%)	Completeness	Similarity (%)	R ²	Reference
2D9Q	Crystal structure of the Human GCSF-Receptor Signalling Complex. <i>E.coli</i>	X-ray diffraction	2.8	25.1	96.0	100	0.38	Tamada et al. (2006)
1CD9	2:2 Complex of G-CSF with its receptor	X-ray diffraction	2.8	23.7	97.7	100	0.27	Aritomi et al. (1999)
1PGR	2:2 Complex of G-CSF with its receptor	X-ray diffraction	3.5	31.7	93.1	100	0.32	Aritomi et al. (1999)
1GNC	Structure and dynamics of human granulocyte colony-stimulating factor determined by NMR spectroscopy. Loop mobility in a four-helix-bundle protein	Solution NMR	N/A	N/A		98		Zink et al. (1994)
1RHG	The structure of granulocyte-colony-stimulating factor and its relationship to those of other growth factors.	X-ray diffraction	2.2	21.5	82.3	100	0.00	Hill et al. (1993)
1BGC	Crystal structure of canine and bovine granulocyte-colony stimulating factor (G-CSF)	X-ray diffraction	1.7	21.3	90.3	87	0.10	Lovejoy et al. (1993)
1BGD	Crystal structure of canine and bovine granulocyte-colony stimulating factor (G-CSF)	X-ray diffraction	2.3	20.7		87		Lovejoy et al. (1993)
1BGE	Crystal structure of canine and bovine granulocyte-colony stimulating factor (G-CSF)	X-ray diffraction	2.2	19.3		87		Lovejoy et al. (1993)

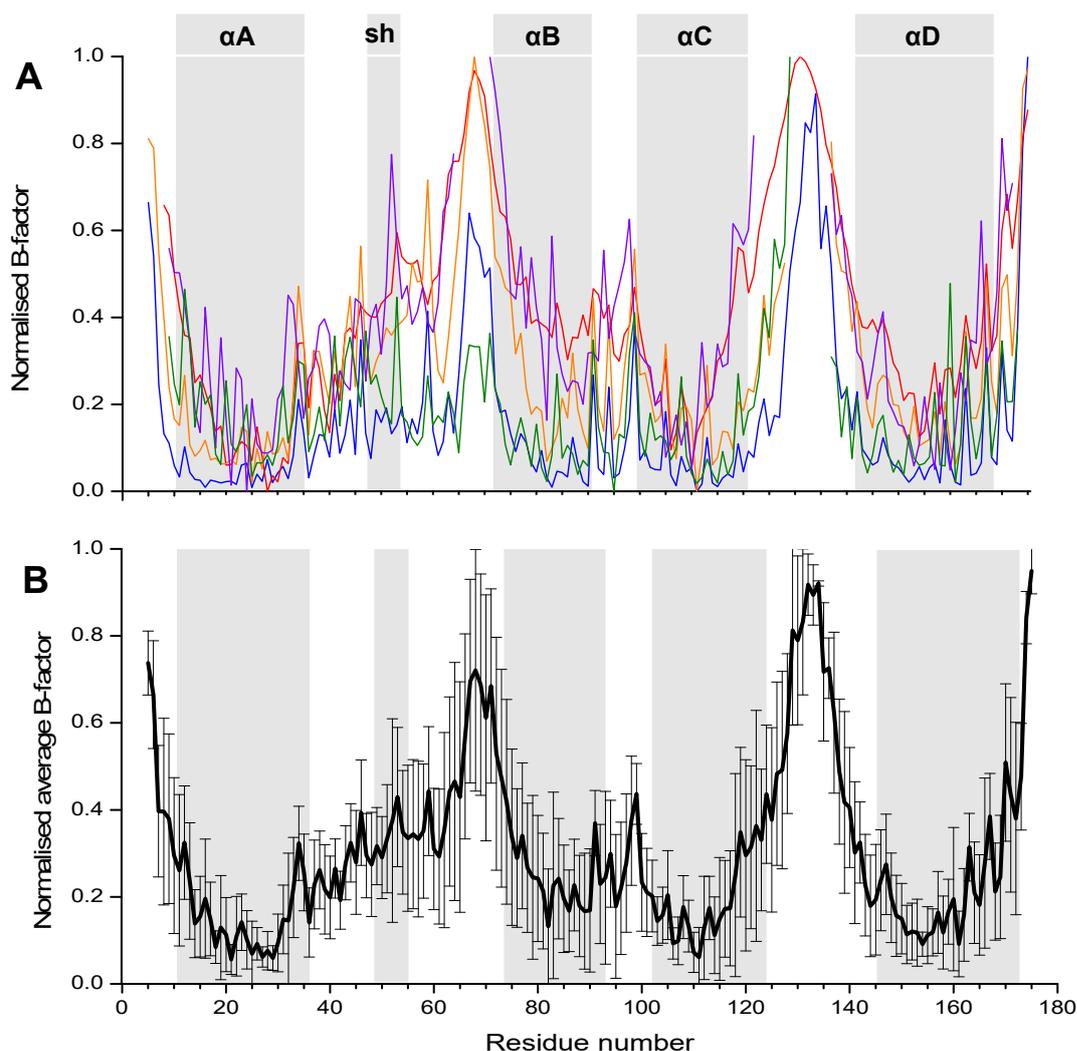


Figure 27. **B-factor analysis of GCSF crystal structures.** A) Overlaid normalised average B-factor values for GCSF residues. PDBs included 1CD9 (blue), 2D9Q (orange), 1PGR (purple) and 1RHG (green). B) Average normalised average residue B-factor value for GCSF residues. GCSF PDB B-factor data was obtained from 1CD96, 2D9Q and 1PGR. 1BCG was not included.

As seen earlier in this chapter, local HDX-MS data is also an indicator of solvent accessibility and flexibility. The average normalised residue B-factor values for each structure were plotted individually against the average HDX-MS rate data for each individual amino acid residue as determined by the experiment in section 3.2.5.1 (see Appendix 3 for individual correlation figures). A linear trend was fit to the data and the R^2 value taken to identify structures with similar dynamics (Table 12). In general, a residue with a high B-factor value also has a high HDX-MS uptake rate, however the correlation was weak. The PDB 2D9Q had the highest R_2 value suggesting it is the most biophysically alike to the purified GCSF prepared in this chapter.

3.2 Discussion

3.2.1 GCSF production

The first aim of this chapter was to express and purify WT GCSF. By SDS-PAGE it was demonstrated that expression was high after both IPTG inductions of the *E.coli* 7.5L fermentation. The second addition of IPTG was included in the method due to the concern that not enough IPTG was added for the number of cells within the fermenter after the first addition. Harvest after four hours had previously been shown to be optimal for protein recovery (Bristow et al. 2012). Due to the high expression levels and reducing intracellular environment GCSF is expressed as inclusion bodies (IBs). IBs are highly amorphous aggregates containing polypeptides that are generally unfolded and partially insoluble and the primary separation process typically includes lysis and disruption of the cells followed by centrifugation. Purification IBs post lysis is relatively simple as a large proportion of the cell pellet is comprised of IBs. An advantage to IB purification is a number of wash steps with detergents, usually too harsh for properly folded proteins, can be applied to remove lipid membranes, lipopolysaccharides (LPS) and other cell debris/contaminants. The GCSF IBs were refolded by urea and high pH solubilisation followed by dilution into 1M arginine.HCl (Rudolph 1996; Kang et al. 1995; Bristow et al. 2012). Arginine.HCl has previously described as a “folding enhancer” and stabilises the native state of the protein by preferential hydration (Rudolph 1996). No reducing reagents, such as DTT, were included in the method as the high pH applied during the solubilisation broke down any inter-chain disulphide bonds, as observed by a change from an opaque to a transparent solution. In addition, only one chromatography step was used for the purification of GCSF, which is unusual as at least two orthogonal methods are typically applied; however, a study by Vemula et al. (2015) found a one-step SEC purification adequate for industrial scale purification of GCSF.

Methods with increased yields of purified GCSF have been reported since the Bristow et al. (2012) study, however the aim of this thesis was not to improve protein yield, and as such the method used was adequate to produce a successful amount of purified, active GCSF material with a relatively low impurity content.

3.2.2 Buffer and protein concentration have large effects on GCSF aggregation

During early stages of GCSF production and experimentation it was found that storage of pooled fractions in the 50 mM sodium acetate pH 4.25 caused visible precipitation of the protein over short periods of time (< 1 month) at 4 °C. As investigated in this chapter using

thermal denaturation, increasing buffer concentration decreased both the T_m and T_{agg} . Whilst the difference between 10 mM and 50 mM was not as pronounced as between 10 mM and 100 mM, the 50 mM concentration decreased thermal stability, especially at high protein concentration. A similar study on GCSF stability with formulation was completed by Ablinger et al. (2012) who explored sodium acetate buffer concentration effects at 4 mM, 24 mM and 44 mM, and at different pH values ranging from 4 to 5. They found the most stabilising formulation was at the lowest buffer concentration and pH. A recent formulation assessment by thermal stress was performed by Alebouyeh et al. (2016), where it was found that different commercial GCSF products with high acetate buffer concentrations (approximately 40 and 70 mM) were not as stable as Neupogen® in 10 mM sodium acetate. Results from this work were used to modify both GCSF purification and HDX-MS methods. During SEC purification, the elution method was modified to elute into fraction tubes containing sterile water to dilute the buffer from 50 to 10 mM sodium acetate and during HDX-MS the sample and D₂O solution were changed from 50 mM to 10 mM to ensure GCSF molecules were stable.

3.2.3 Characterisation of GCSF

The pool of GCSF fractions from SEC purification was analysed by different biophysical methods to A) confirm the identity of WT GCSF, and B) develop methods for subsequent experimental chapters. By SEC-HPLC and IMS it was shown that the SEC fraction pool was 96-98% monomeric and contained no soluble aggregates, which is within the acceptable limit of 1% (Herman et al. 1996). The data also aligned with previous reports on GCSF characterisation, such as the work completed by Skrlin et al. (2010) directly comparing Neupogen® to a biosimilar. The IMS confirmed the intact mass was 18,799 Da (minus 4 Da for the two disulphide bonds), which matched the theoretical mass of the amino acid sequence and confirmed the identity of WT GCSF.

3.2.3.1 *Collision induced unfolding shows a single intermediate unfolding state*

Collision induced unfolding was used to study the unfolding of native GCSF by changing of the trap voltage and was the first record of using this technique for the characterisation of GCSF. During CIU, GCSF transitioned through three different states: folded, partially unfolded and fully unfolded state. At 21 V, an equilibrium of folded, intermediate and unfolded species occurred at 21 V, which is analogous to the T_m of the sample during thermal denaturation studies measuring protein fluorescence.

Previously the folding kinetics of GCSF was determined by Brems (2002) by mixing GCSF with the denaturant, Gdn.HCl and measuring folding and unfolding by tryptophan (W) fluorescence and far-UV circular dichroism (CD). It was found GCSF had two intermediate states prior to folding. Unlike the data from Brems (2002), the CIU data from this chapter observed only one distinct intermediate. The loss of the second intermediate could be due to the different mechanism for unfolding by denaturing agent versus thermal denaturation. Alternatively, due to the stepwise increase of 3 V, not all voltages were applied to the sample. It could be that smaller voltage increments would have increased resolution and allow observation of the two unfolding species.

3.2.4 Peptide level GCSF HDX-MS

3.2.4.1 *Quench solution needs both denaturant and reducing agent for high sequence coverage*

For peptide level HDX-MS analysis the minimum required peptide map sequence coverage is 95%. Quench solution optimisation can significantly increase the sequence coverage % during online GCSF pepsin digestion. As there has been no published study to explore optimisation of the GCSF quench solution digestion factor, all digestion variables were explored.

The value of optimising the quench solution was demonstrated in this chapter by the increase in sequence coverage from 68% to 96.7%. Including chaotropic agents in the quench solution can improved digestion efficiency as the secondary and tertiary structure of the protein are broken down allowing easier access for the pepsin. The use of the denaturant Gdn.HCl was shown to increase the relative number of peptides across the protein, which increased redundancy, whereas the inclusion of the reducing agent TCEP increased the number of peptides around the disulphide bond forming cysteine residues. The optimal concentrations of both reagents was found to be 4 M and 600 mM respectively, however the digestion optimisation plateaued around 3 M guanidine.HCl and 50 mM TCEP. This finding agrees somewhat with Zhang et al. (2015) who used 4 M guanidine.HCl, and 500 mM TCEP (pH 2.4), but was dissimilar to that of Wei et al. (2012) and Tsuchida et al. (2014) whose HDX-MS digestion steps for GCSF used a lower guanidine hydrochloride concentrations at 1.5 M with 500 mM TCEP. In 2017, Brokx et al. (2017) studied pegylated Filgrastim using 8 M guanidine hydrochloride and 200 mM TCEP in the quench solution. The lower concentrations of chaotropic agents recorded may have occurred due to the sample pH being higher than that used during this chapter. All

previous HDX-MS experiments with GCSF formulated the sample at a physiological pH, which may have contributed to the destabilisation of structure, allowing for lower concentrations of chaotropic agents to be effective.

3.2.4.2 *Pepsin column backpressure does not improve sequence coverage*

High pressure has been used to increase digestion efficiency in various mass spectrometry and proteomics applications (Ahn et al. 2012). High pressure promotes protein denaturation, as it mechanically stretches proteins to expose more cleavage sites. As such, the use of a flow restrictor post pepsin column has been shown to increase column back pressure and improve the peptide map sequence coverage for proteins such as human growth hormone and mAb IgG2 (Fang et al. 2015). In this chapter, the effect of backpressure was explored for GCSF peptide map sequence coverage, where it was found it decreased sequence coverage. One possible explanation is the pressure created GCSF aggregates via the exposure of the hydrophobic core, another is that the pressure encouraged the formation of new disulphide bonds on-column involving the four newly free and one originally free cysteine residues (Roessl et al. 2012). A study by Fekete and Guillarme (2015) explored the effects of pressure and temperature on the retention time of a number of different proteins including GCSF using RP-HPLC. The group found both factors caused protein denaturation and increased column retention time. As such, it's possible a method, such as increased backpressure, which works for the similarly structured hGH, could have a very different effect on GCSF. Based on the results of this chapter, a flow restrictor was not included in any subsequent HDX-MS experiments with GCSF.

3.2.5 Effect of exchange solution pH on GCSF HDX-MS

A major disadvantage of low pH HDX-MS is that the low pH will reduce the intrinsic rate of exchange and cause slow exchanging regions, such as those within α -helices, to exchange even slower. On the other hand, a study by Coales et al. (2010) showed that reducing the pH can be exploited to increase the window of exchange measurements for unstructured fast exchanging amide hydrogens, such as those within loop regions. By calculating the relative time window of a low pH HDX solution, exchange rates can be corrected for the change in intrinsic exchange, allowing for comparisons between low and physiological pH solutions.

The labelling times of GCSF at pH 4.4 were corrected using the equation for base-catalysed amide hydrogen exchange reaction combined with the Arrhenius equation (Coales et al. 2010):

$$k_{ch} \sim k_{OH}[OH^-] = A \exp\left(\frac{E_a}{RT}\right)[OH^-] \quad \text{Equation 20}$$

where k_{ch} is the calculated intrinsic rate in minutes, A is the frequency factor, T is the temperature in K, R is the gas constant at 1.987, E_a is the activation energy of the base-catalysed amide hydrogen exchange reaction at 17 kcal/mol (Bai et al. 1993), and $k_{OH}[OH^-]$ is the base-catalysed reaction, which increases at higher pH (> 2.5). Values for the base-catalysed reaction were obtained from polyalanine HDX studies performed by Bai et al. (1993) using $^1\text{H-NMR}$.

The change in protein dynamics, or K_{op} , in different solution conditions can be determined by plotting deuterium uptake curves with time window correction (Coales et al. 2010). If the data overlays and is continual, then protein dynamics can be assumed to be unchanged. Following correction, it was found that the longest time point for the pH 4.4 (28.8 s) matched the shortest time point at pH 7.4 (30 s). Differential analysis of GCSF peptide uptake values from the overlapping time points of both pH HDX experiments found values did not overlay and a large number showed significant positive and negative differentials. Uptake was found to be significantly higher for the pH 4.4 sample in most GCSF regions, apart from loopCD, which had significantly higher uptake at pH 7.4. The study by Coales et al. (2010) describing time window corrections for pH and temperature, found negligible variations for all amide hydrogens in cytochrome c for pH 5 to 10 and temperatures 0 °C and 23 °C. On the other hand, the same study found variations in protein dynamics for human growth hormone (hGH) at four different solution pHs indicating that changes to dynamics with pH may be protein specific. Interestingly, hGH is part of the same four-helix bundle class of proteins as GCSF (Mott and Campbell 1995) and GCSF has previously been shown to be sensitive to pH. When incubated at 37 °C, pH 7, GCSF aggregated slowly but extensively in a protein concentration manner as measured by Raso et al. (2005) but did not occur in pH solutions less than 5 (Herman et al. 2002). The instability data of GCSF at pH 7 combined with the difference in dynamics seen during HDX-MS differential analysis, ultimately indicates that HDX-MS experiments should be

performed using the low pH exchange for GCSF for both sample stability and measurement integrity.

3.2.5.1 *GCSF local dynamics in native conditions*

Peptide level HDX-MS was performed to identify regions of GCSF that are highly dynamic in the native state to inform stability engineering and formulation. In 10 mM sodium acetate pH 4.25, the loop regions were highly accessible with most regions fully exchanged by the longest time point studied. The four α -helices were easily identified from the relative uptake graph, where their centres showed baseline uptake levels which didn't increase with labelling time. This indicates that they were not as accessible as the loop regions. This was to be expected as their backbone NH is involved in the formation of H-Bonds for the 3D structures, and are therefore protected from HDX. The terminals of the helices showed measurable exchange over time which was caused by either a large peptide that spans both loop and helix structures, or fraying (Quint et al. 2010).

As well as informing on differences between structured and unstructured regions, the HDX data provided information on differences between the same structure types, which has potential for use in rational stabilisation strategies. For GCSF, peptides within helix α C and α D showed higher exchange levels compared to those within the α A and α B helices. The high level of exchange for α D was also identified by Zhang et al. (2015) who suggested this could be due to stronger structural constraints on helices α B and α C. The instability of α C and α D could also be due to the close proximity to the highest exchanging Loop region, loopCD, causing fraying. Alternatively, they could contain a lower proportion of hydrophobic residues around central regions, reducing the hydrophobic core strength. Consequently, focusing on these regions for structural stabilisation could be more impactful than stabilisation on an already rigid (low exchange) region such as α A.

3.2.6 Structure selection for *in silico* research

For *in silico* research a solved structure of GCSF was required; a caveat of most computational protein stabilisation strategies. As GCSF is so well studied, there were eight structures available for use. To identify the most suitable one, a comparison was performed and four were identical to the amino acid sequence of the purified GCSF from Bristow et al. (2012). Further analysis of crystal structures found the resolution and *R-Factor* values of 2D9Q were reasonable at 2.8 Å and 25.1%, respectively but were succeeded by 1CD9 with a

lower *R-factor* of 23.7% and a completeness of 95.7%. The other two structures, 1PGR and 1RHG, had large loop sections missing from the structure.

Whilst B-Factor and RMSF values are the two most common ways to identify flexible residues within a protein, the use of peptide-level HDX-MS data also lends itself to the identification of flexible regions of proteins from their increased relative rate of deuterium uptake over time. B-factor and MD simulations require three-dimensional (3D) structure knowledge to perform their analysis and this can cause major hurdles in terms of time and money. On the other hand, HDX-MS requires small sample quantities and no prior knowledge of the 3D structure, suggesting it could be a more efficient alternative for the identification of flexible regions within globular proteins. Additionally, HDX-MS is complementary to X-ray crystallography because it bridges the gap between the static “snapshot” of the protein crystal (solid) and its dynamic properties in solution (Bou-Assaf and Marshall 2014). The GCSF local dynamics data obtained in this chapter allowed for further analysis of the GCSF PDBs. The B-factors of the four PDBs were correlated with HDX-MS relative uptake rates, which allowed for identification of the structure most dynamically similar to experimentally expressed and purified GCSF via linear fit R^2 values. 1CD9 B-factor data did not fit the HDX data for GCSF as well as 2D9Q as seen by the lower R^2 value. Consequently, the structure 2D9Q was selected as the most appropriate structure for *in silico* GCSF work.

4 GCSF stability mutants I: *in silico* design, production and stability analysis

The aim of the following two chapters was to explore the relationship between GCSF flexibility and stability. In this chapter, the GCSF biophysical data obtained in chapter 3 was used to inform high-throughput *in silico* mutagenesis of GCSF, produce a number of stability mutants, and use biochemical and biophysical techniques to evaluate predictions.

HDX local dynamics and B-factor data for GCSF residues in chapter 3, was taken forward to inform regions for stabilisation. Two *in silico* parallel approaches were used to predict stabilising mutations. The first used RosettaDesign, an online server, which required PDB and text instruction files to create optimal amino acid substitutions based on folding energy. The high quality crystal structure of GCSF, 2D9Q, as identified in chapter 3 was used as the structure input for RosettaDesign, and for the text instructions, flexible and non-flexible regions were selected.

The second strategy used a High Performance Computing Facility (HPCF) to run the Rosetta application “Rosetta_ddg_monomer”. The HPCF allowed for faster processing of multiple predictions at the same time, consequently, the entire 2D9Q sequence was input and all possible single point mutations mapped.

Both RosettaDesign and Rosetta_ddg_monomer predict amino acid substitutions based on an increase in protein folding energy ($\Delta\Delta G$). GCSF mutants with the highest ranked folding energies (lowest $\Delta\Delta G$) from both methods were produced using polymerase chain reaction (PCR) based site-directed mutagenesis, followed by expression and purification alongside WT GCSF for direct comparison. Characterisation of the novel GCSF mutants included yield, assessments of purity by (NR) SDS-PAGE, bioactivity by GNFS-60 cell proliferation bioassay, thermal denaturation for T_m and T_{agg} and changes to surface hydrophobicity by RP-HPLC.

The comparison of mutant stability and location within the 3D structure of GCSF allowed for the assessment of the effectiveness of rational design based on flexibility.

4.1 Results

4.1.1 Stabilisation by rational design

The 2D9Q B-factor and HDX-MS average % uptake rate data obtained in chapter 3 were grouped into structured (helices) and unstructured regions (loops) and plotted against each other in Figure 28. A weak positive correlation between B-factor and HDX-MS % uptake rate was found.

There were two main groupings of structures as seen in Figure 28: high B-factor, high HDX (High-High) and low B-factor, low HDX (Low-Low). High-High indicated an agreement of high flexibility and Low-Low indicated an agreement of low flexibility, by both HDX and B-factor data. High-High grouped regions included loops A, ABII, BC, CD and D, as well as the short helix, whereas regions within the Low-Low group included the α -helix A, B, C and D, as well as LoopABI. Of the helices, α A had the lowest average B-factor value and the short helix, had the highest. Helices α C and α D had relatively high B-factor and HDX-MS uptake rates suggesting they are the most flexible. Of the loops, loop CD had the highest average B-factor followed by LoopABII. For HDX-MS data, both α B and α C had very high variability in exchange suggesting they contain small regions that are high exchanging and small regions that are low exchanging, which may be caused by the joining loop regions. Helix B had a high average B-factor value but a low HDX-MS uptake rate.

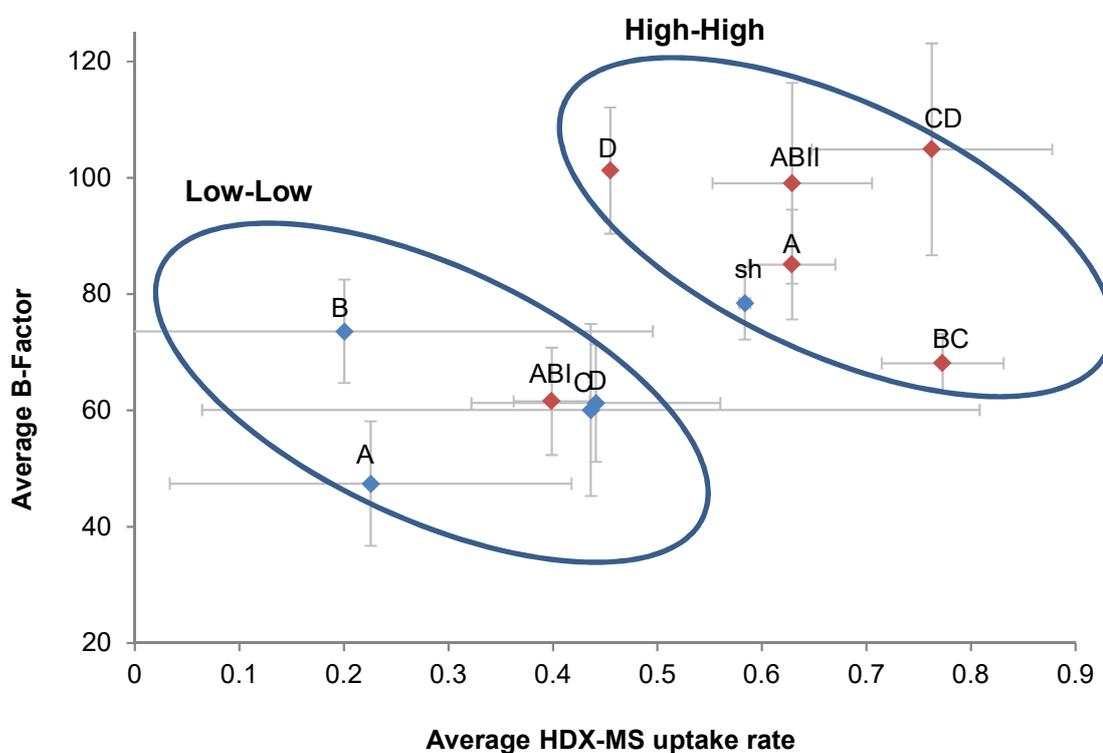


Figure 28. Average HDX-MS % rate plotted against the average 2D9Q B-factor value for different structured regions of GCSF. α -helices are coloured in blue and loops in orange. Points are grouped into those that display high values on both axis (High-High) and with low values on both axis (Low-Low).

4.1.2 RosettaDesign target selection

From the data in Figure 28, the α C helix and loopCD were selected as target regions for RosettaDesign, as both had the highest HDX % uptake rate and variability compared to the other respective structures. The helix α A was also selected for the opposite reason: it showed both a low HDX uptake % rate and low average B-factor values. Although it had the highest combined average B-factor and HDX values, the sh was not selected due to its small size and proximity to GCSFR binding regions.

Following *in silico* mutagenesis, the RosettaDesign average protein pose total energy, and amino acid sequence was extracted from WT and mutant output files. The average total energy of the WT and mutant structures was calculated from ten independent trajectories and the $\Delta\Delta G$ calculated (Figure 29). Some residues were unchanged and were discounted from final results. The residues Lys16, Leu19, Gln20, Arg22, Lys108, Asp109 and Asp112 were also not included due to their interaction with GCSFR (Tamada et al. 2012).

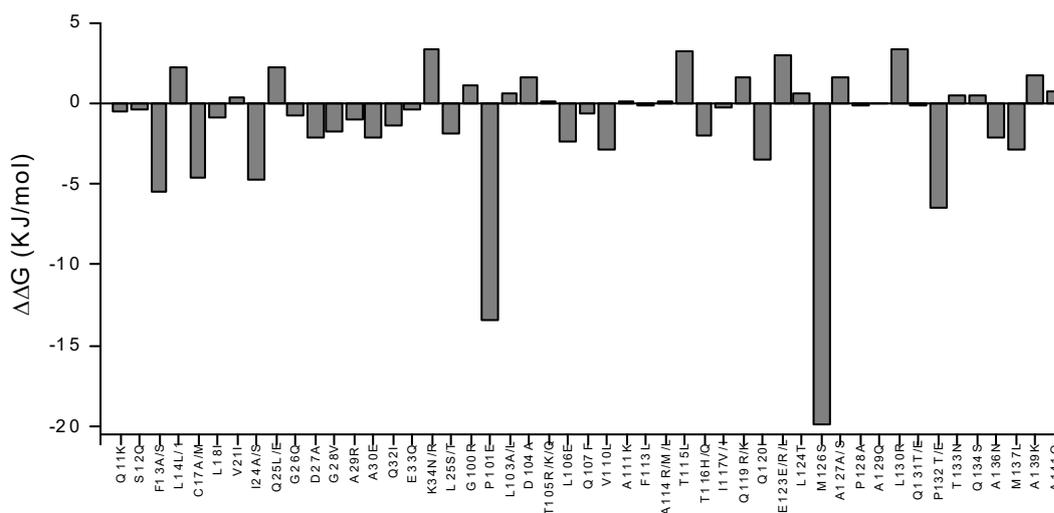


Figure 29. **RosettaDesign predicted mutations and $\Delta\Delta G$ values.** The lowest energy yielding mutation for each GCSF target residue is shown along the x-axis and the average of 10 $\Delta\Delta G$ values plotted along the y-axis, as predicted by RosettaDesign with the GCSF 3D structure PDB 2D9Q (Tamada et al. 2012).

The RosettaDesign $\Delta\Delta G$ values from Figure 29, were ranked from lowest to highest and the top nine selected for experimental mutagenesis (Table 13). From their location shown in Figure 30, the top nine candidates covered all three targeted regions.

Table 13. The nine lowest $\Delta\Delta G$ RosettaDesign mutations, their change in molecular weight (ΔMW) and the location within GCSF secondary structure.

Rank	Mutation	ΔG	ΔMW (Da)	Location
1	M126S	-19.92	-34	Loop CD
2	P101E	-13.47	+32	αC
3	P132T/E	-6.46	+32	Loop CD
4	F13A/S	-5.45	-76	αA
5	I24A/S	-4.68	-42	αA
6	C17A	-4.66	-32	αA
7	Q120I	-3.48	-15	αC
8	M137L	-2.86	-18	Loop CD
9	V110L	-2.83	+14	αC

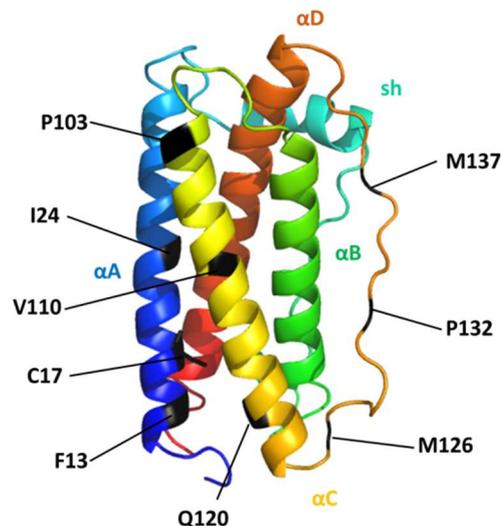


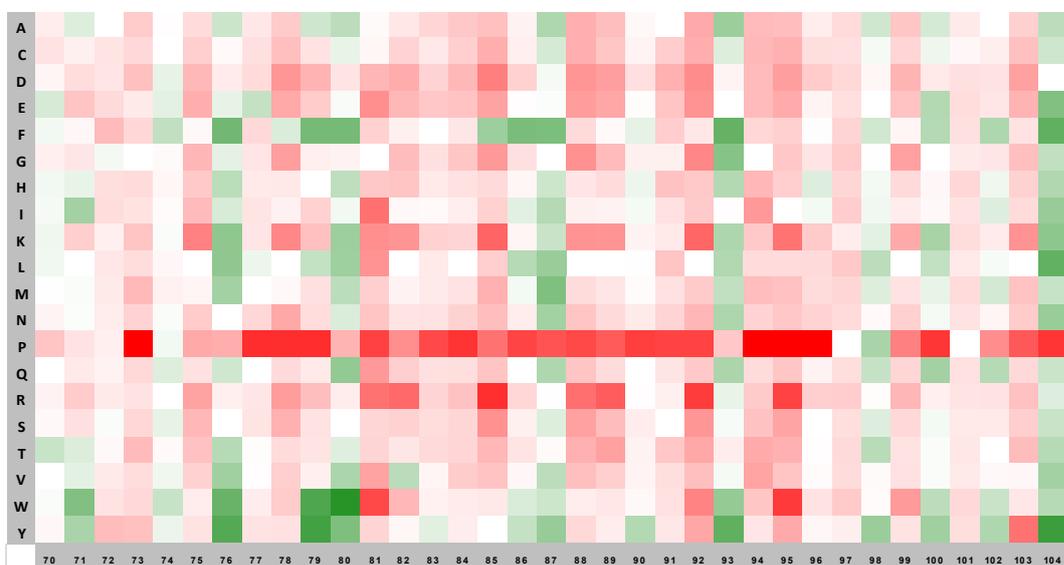
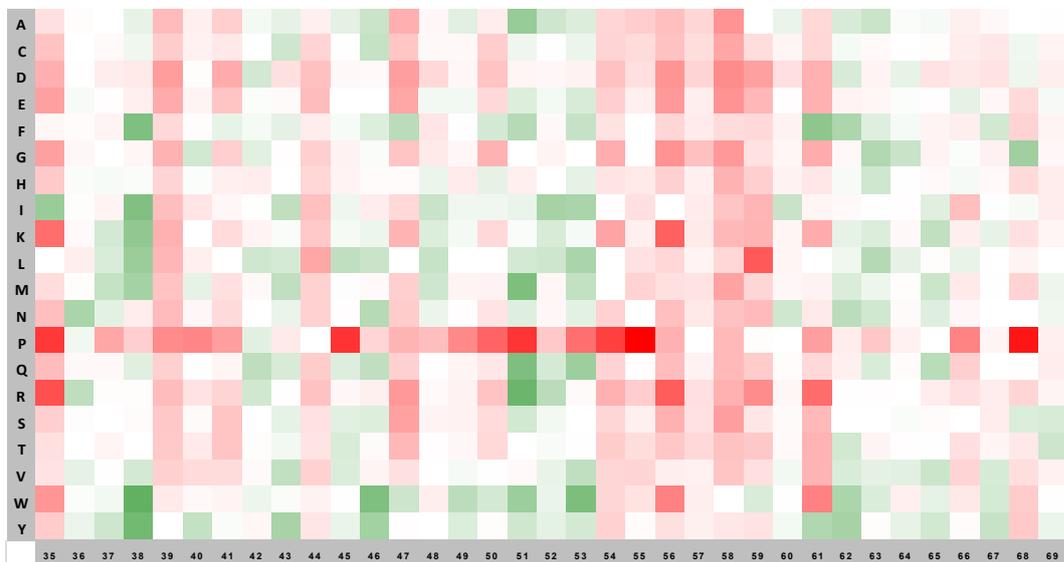
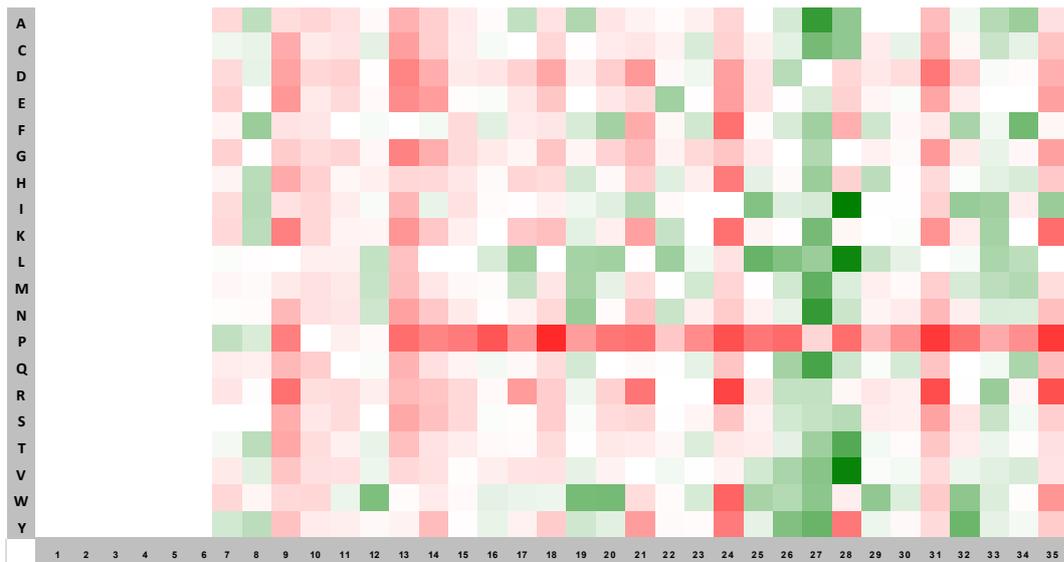
Figure 30. RosettaDesign top nine mutation locations. Mutation locations were highlighted on the GCSF PDB 2D9Q (Tamada et al. 2012) prepared using PyMOL Molecular Graphics System (Schrödinger, USA).

4.1.3 Stabilisation by computational design

4.1.3.1 Rosetta_ddg_monomer

Due to the higher throughput of the Rosetta_ddg_monomer program, all residues within 2D9Q were selected for mutation and all possible amino acid substitutions were predicted. The mutation map of the Rosetta_ddg_monomer $\Delta\Delta G$ output in Figure 31, shows destabilising mutations in squares coloured red and stabilising mutations in green squares. The $\Delta\Delta G$ values for the first six residues of GCSF were not predicted as they were not present in the 2D9Q PDB.

Two main trends were apparent from the Rosetta_ddg_monomer mutation map in Figure 31. Firstly, certain residues were more amenable to substitution, as seen as a column of green squares, and secondly, proline was a common destabilising substitution, as seen by the intense red squares along the proline row of each map segment.



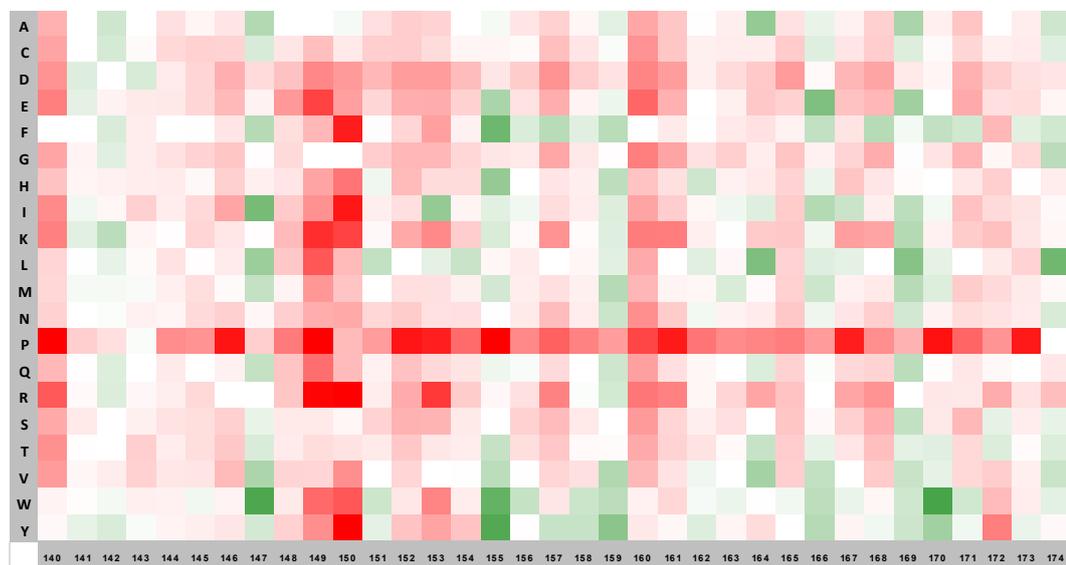
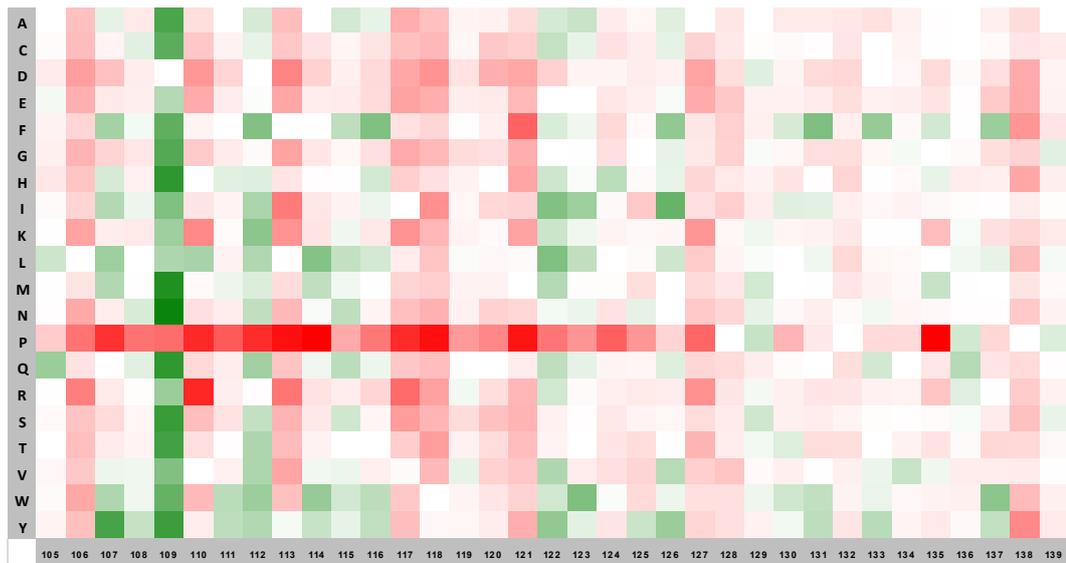


Figure 31. **Point mutation map of GCSF by Rosetta_ddg_monomer.** The x-axis represents the protein residue number and the y-axis represents the possible amino acids. The colour key represents the value of $\Delta\Delta G$, where the green colour indicates negative values, the white colour indicates a value of 0 or no mutation, and the red colour indicates a positive value.

The lowest $\Delta\Delta G$ mutation for each residue was taken forward for further analysis. From the mutant count in Figure 32A, 73% of residues were negative $\Delta\Delta G$ values indicating that the majority of GCSF residues were amenable to substitution. Three mutants had predicted $\Delta\Delta G$ values less than -5. The average $\Delta\Delta G$ of each amino acid type within GCSF was calculated and displayed in Figure 32B, where those with negative values represent mutation amenability whereas those with a positive $\Delta\Delta G$ represent destabilisation upon any mutation. Residues with definitive negative values included cysteine, aspartic acid, glutamic acid, lysine, methionine, glutamine, arginine, serine, threonine and valine. Residues with a positive average $\Delta\Delta G$ included isoleucine and tryptophan. The frequency of amino acids substituted within the lowest $\Delta\Delta G$ mutation at each residue is displayed in Figure 32C where the most commonly introduced residues were tyrosine, tryptophan, phenylalanine, leucine and isoleucine.

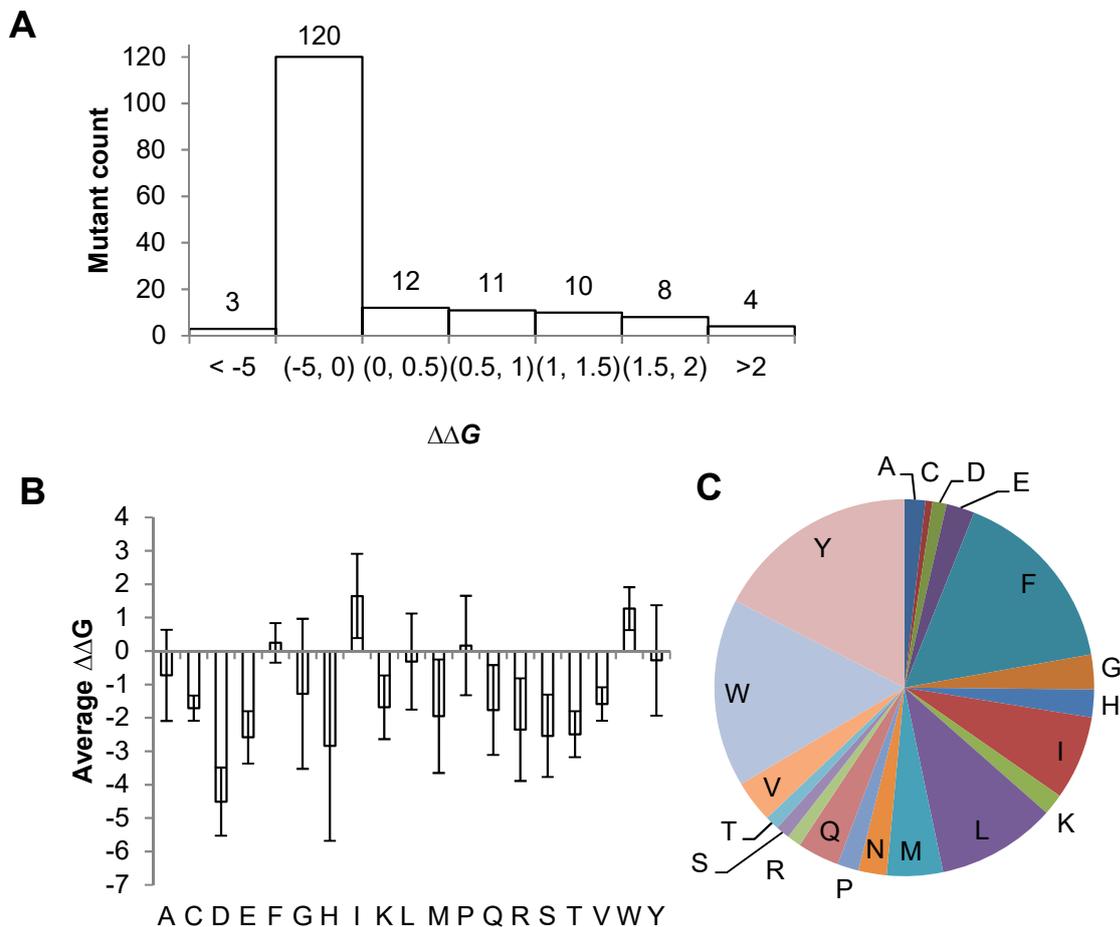


Figure 32. **Lowest $\Delta\Delta G$ mutant analysis.** Values obtained from the lowest $\Delta\Delta G$ mutant per residue of GCSF as predicted by Rosett_dgd_monomer. A) Histogram for the mutant frequency distribution based on $\Delta\Delta G$, B) Average $\Delta\Delta G$ of amino acid residues within the lowest $\Delta\Delta G$ mutants C) Selection frequency for amino acids within the lowest $\Delta\Delta G$ mutants.

The Rosetta_ddg_monomer $\Delta\Delta G$ values were ranked from lowest to highest and the top eleven selected for experimental mutagenesis (Table 14). From Figure 33, the top mutations are located in all α -helices, and loops ABI, ABII and CD, suggesting Rosetta_ddg_monomer showed no preference for any particular regions of GCSF.

Table 14. The eleven lowest $\Delta\Delta G$ Rosetta_ddg_monomer mutations, their change in molecular weight (ΔMW) and the location within GCSF secondary structure

No.	Mutation	ΔMW (Da)	$\Delta\Delta G$	Region
1	S12W	+99	-3.02	αA
2	G51R	+99	-3.45	sh
3	L71W	+73	-2.95	LoopABII
4	D104Y	+48	-4.61	αC
5	Q131F	+19	-3.00	LoopCD
6	S164L	+26	-3.03	αD
7	G28I	+56	-5.66	αA
8	T38W	+85	-3.68	LoopABI
9	S80W	+99	-5.05	αB
10	Q107Y	+35	-4.33	αC
11	S155Y	+76	-4.04	αD
12	WT	0	0	N/A

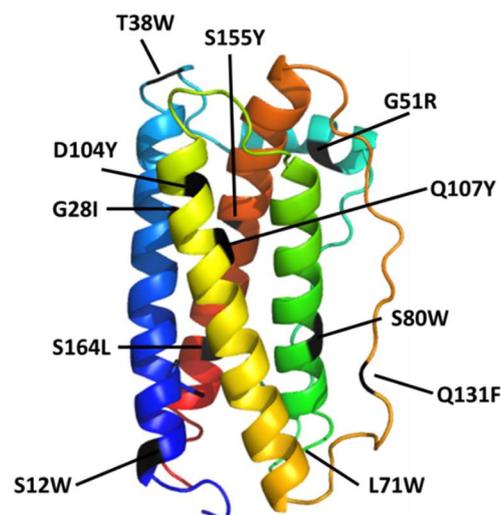


Figure 33. Rosetta_ddg_monomer top eleven mutation locations. Mutation locations were highlighted on the GCSF PDB 2D9Q (Tamada et al. 2012) prepared using PyMOL Molecular Graphics System (Schrödinger, USA).

4.1.4 Production and characterisation of mutants

Cloning of plasmids containing the top ranked RosettaDesign GCSF mutants was performed by UCL undergraduate student Lok Man Wong, and the Rosetta_ddg_monomer GCSF mutant plasmids by UCL MSc student Luyan Kong, under supervision. Both students used a polymerase chain reaction (PCR) based site-directed mutagenesis. The work also included DNA gel electrophoresis to confirm amplification of both the pET21a plasmid and mutant GCSF genes, transformation of ultracompetent cells, plasmid sequencing, and finally transformation into *E.coli* BL21 (DE3) cells for expression. Success of mutagenesis was assessed via alignment of the sequenced mutant amino acid sequence with WT (Appendix

7). Six out of the eight selected RosettaDesign mutant plasmids listed in Table 13 were successfully generated, however, there were issues creating P101E and V110L as they had unwanted mutations in other regions. All eleven of the Rosetta_ddg_monomer mutant plasmids listed in Table 14, were successfully produced.

Each design stream of mutants along with a WT plasmid, were produced in parallel with the same conditions in order to compare them directly. Unfortunately, during the refold stages, the bottle containing the Rosetta_ddg_monomer WT sample leaked and a substantial amount of material was lost. The remaining material was processed; however, the yield was a lower than it should have been.

4.1.4.1 Protein yield

The yield for all mutants and WT samples was determined by the SEC peak height divided by the difference in extinction coefficient (EC) between the mutant and WT (ΔEC ; Figure 34; Appendix 8). A lower yield suggested the mutant did not refold properly and lost during clarification. From the RosettaDesign stream, C17A was the only mutant with a higher yield than WT. Mutants M126S and M137L had similar yields to WT. Mutant Q120I had slightly lower and F13A, I24A and P132E had very low yields. From the Rosetta_ddg_monomer stream, as WT material was lost during refold, the relative yield of mutants could not be assessed, however the average mutant yield (36.13 ± 17 mAU) was used for relative yield determination. Mutant L71W had the highest relative yield. Mutants G51R, Q107Y and S155Y had above average relative yields. Mutants S12W, T38W, D104Y and Q131F had close to average yields and finally, mutants S164L, G28I, and S80W had lower than average relative yields. Mutant G28I had almost no protein present during elution suggesting complete destabilisation and loss of all material during refold.

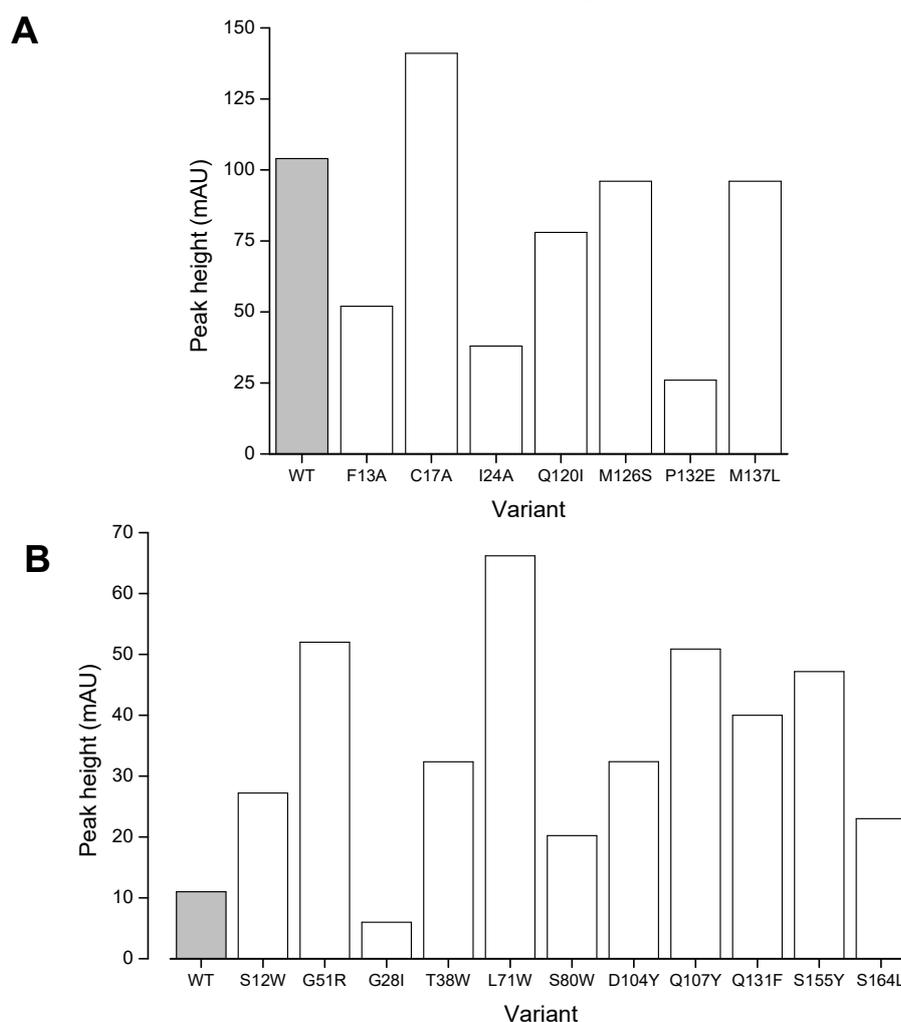


Figure 34. Purification yield of WT and mutants. The peak height of the monomer peak during size exclusion chromatography (mAU) was determined and divided by the calculated extinction coefficient in Appendix 8. A) RosettaDesign design stream, B) Rosetta_ddg_monomer design stream.

4.1.4.2 Purity

The purity of mutant and WT samples was determined by (NR) SDS-PAGE (Figure 35). From the RosettaDesign stream mutants (Figure 35A), all samples had a band at 18.9 kDa indicating successful expression, refold and purification. Bands were also of similar size and intensity to each other and the reference, indicating that protein concentration was comparable. Mutants F13A, C17A Q120I, M126S, M137L and WT appeared as one single band, whereas Mutants I24A and P132E had MW both higher and lower, indicative of multimeric and fragmented species, respectively. The larger MW band was around 25 kDa indicating it is not dimer which would be approximately 36 kDa. From the Rosetta_dgg_monomer stream (Figure 35B), the WT GCSF sample had a faint dimer band. All mutant samples had a major band around 18.9 kDa, however, migration increased for mutants S155Y, S80W, S164L, Q131F and G51R suggesting these mutations caused a reduction in MW. Mutants T38W, S164L and D104Y had dimer in their samples, and mutant G28I had two high MW bands and a smearing of low MW bands suggesting significant degradation occurred during processing.

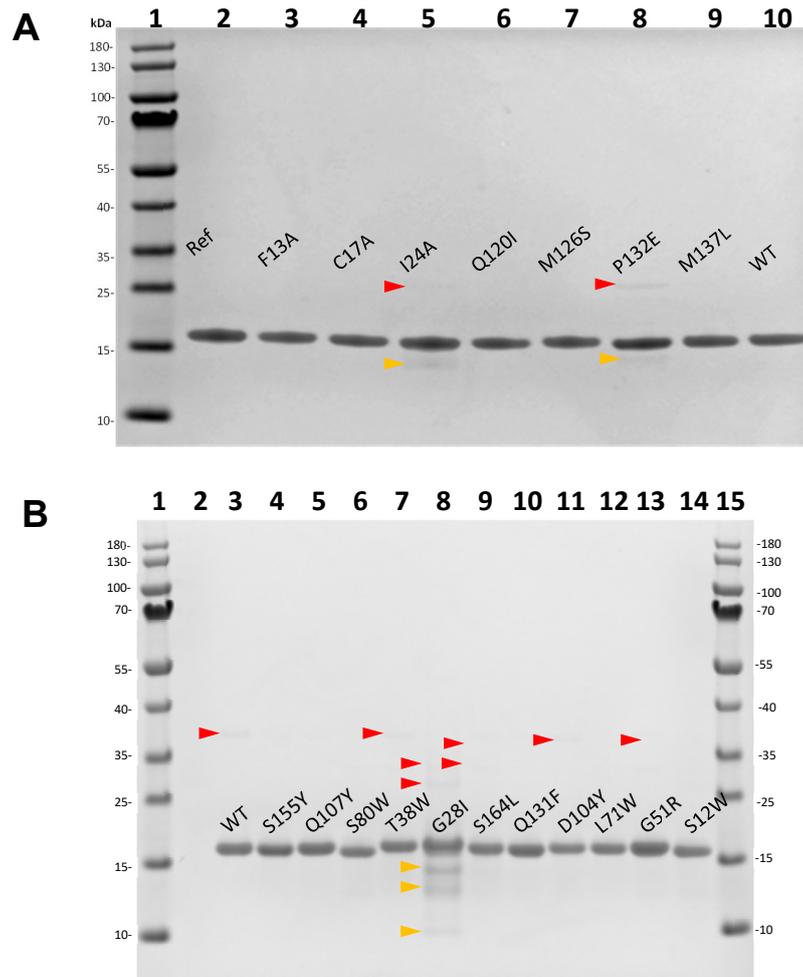


Figure 35. **Non-reduced SDS-PAGE analysis of GCSF WT and mutant purified samples.** A) RosettaDesign stream samples, B) Rosetta_dgg_monomer stream samples. Higher molecular weight impurities are highlighted with a red arrow and lower molecular weight impurities are highlighted with an orange arrow.

4.1.4.3 Bioactivity

The bioactivity of mutants relative to WT was assessed using the GNFS-60 cell proliferation bioassay with parallel line analysis. The cell response to the log dose of each sample was plotted for each sample. A response curve lower than the WT was indicative of a loss of activity due to mutation. From the Rosetta Design stream, no mutants had complete loss of activity, however, mutant G28I was purified in such a low amount that there was not enough sample available for analysis. From Figure 36, mutants C17A, Q120I, M126S and M137L had comparable activity to WT, whereas there was a significant decrease in response for mutants I24A, F13A and P132E. From the Rosetta_ddg_monomer stream mutants in Figure 37, S12W, T38W, G51R, L71W, D104Y, Q107Y, S155Y, Q131F and S164L all had similar response curves to WT. Mutants G51R and Q131F had higher response curves relative to WT at the lower end of the log dose. Mutant S80W had a significant decrease in response compared to all other Rosetta_ddg_monomer mutants.

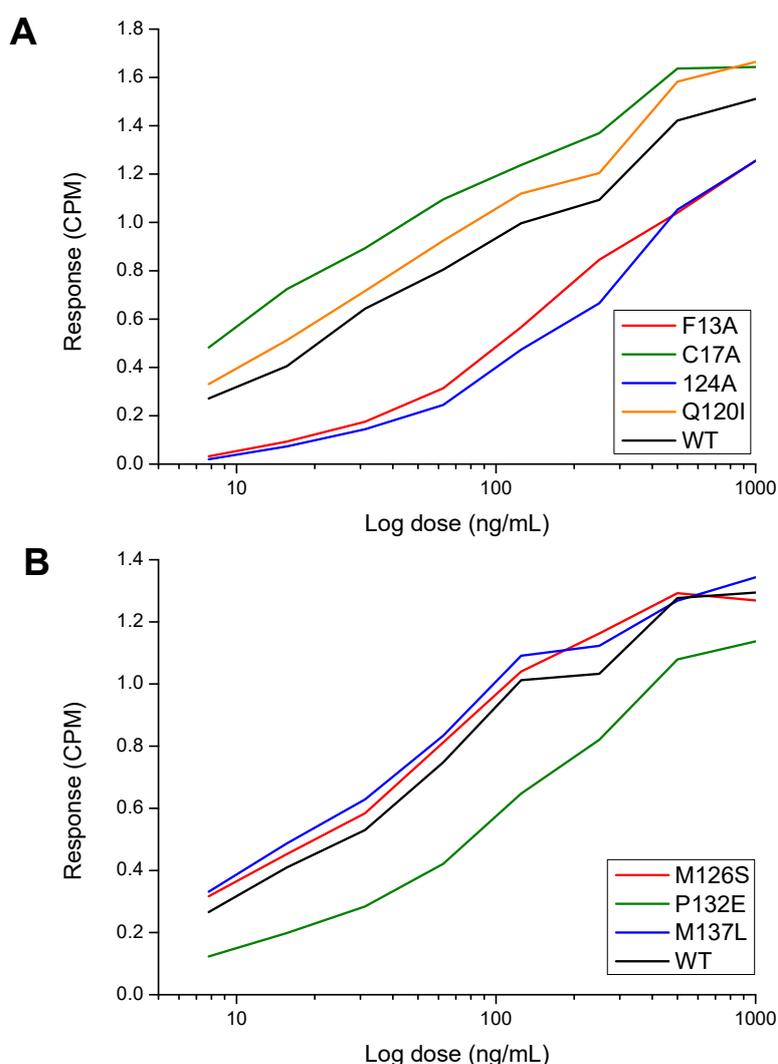


Figure 36. Comparison of RosettaDesign GCSF mutant bioactivity relative to WT using the GNFS-60 cell proliferation bioassay. A and B represent the two different microtitre plates used with the mutant samples displayed in the individual figure legends. A sample of WT GCSF was included on both plates.

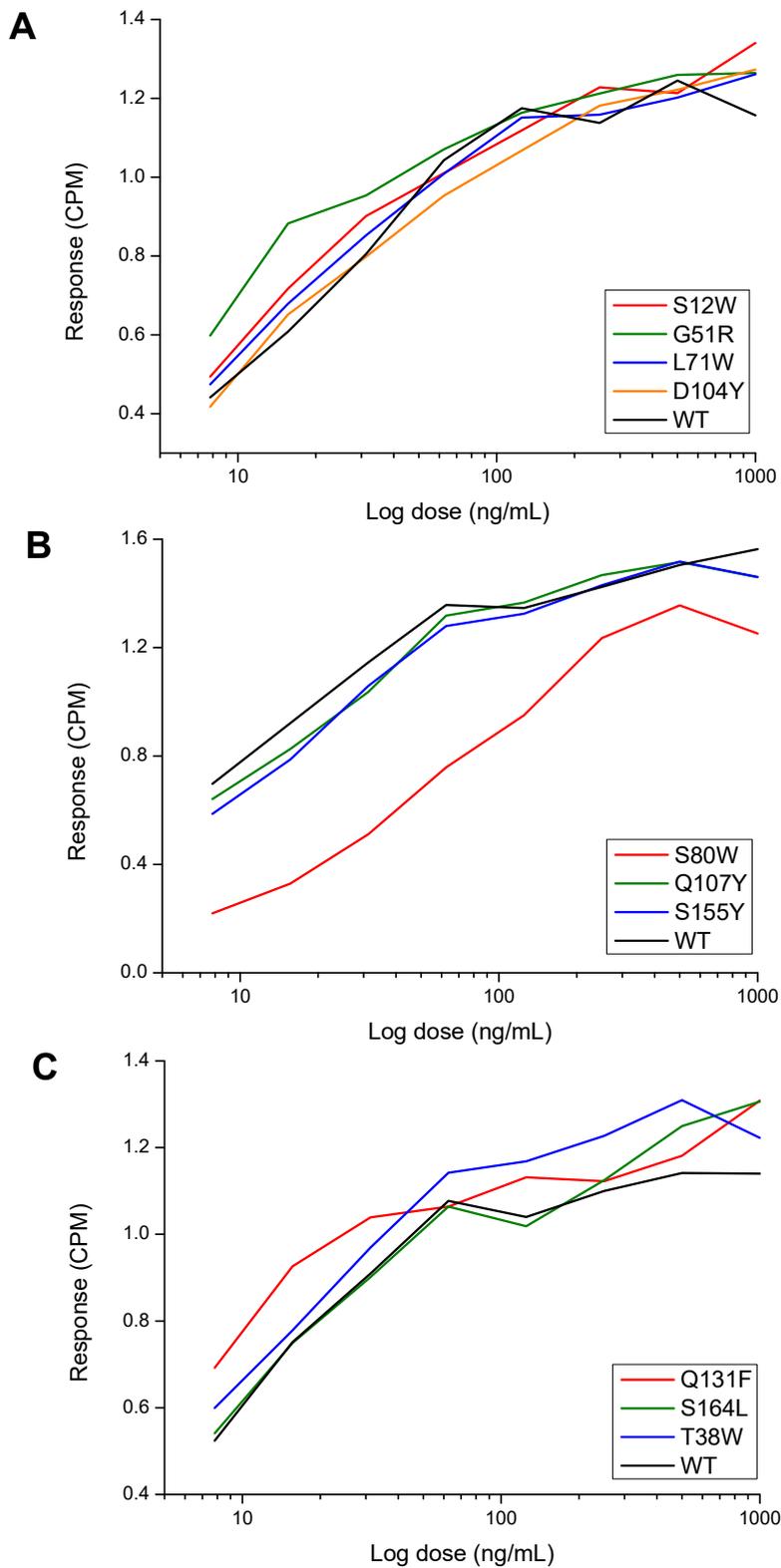


Figure 37. **Comparison of Rosetta_ddg_monomer GCSF mutant bioactivity relative to WT using the GNFS-60 cell proliferation bioassay.** A, B and C represent the three different microtitre plates used with the mutant samples displayed in the individual figure legends. A sample of WT GCSF was included on all plates.

4.1.4.4 *Thermal stability analysis*

To assess mutant thermal stability, a heat ramp was applied to samples and the temperature of unfolding (T_m) and aggregation onset (T_{agg}) determined from the tandem fluorescence intensity and SLS curves using the van't Hoff equation in section 2.4.6., and displayed in the summary table (Table 15).

During sample preparation, mutants I24A and P132E from the RosettaDesign stream could not be concentrated to 1 mg/mL and visible particulates were present in the samples, therefore the WT sample was diluted with 10 mM sodium acetate to allow for comparisons with the two mutants at 0.3 mg/mL. Due to the low sample availability, mutant G28I from the Rosetta_ddg_monomer stream was not analysed.

4.1.4.4.1 Fluorescence intensity

The average fluorescence intensity (FLI) profiles for mutants and WT sample of the same concentration were overlaid for each individual mutant in Figure 38 to Figure 40. The WT samples from both design streams followed the same FLI pattern, starting with a midrange FLI, which decreased with increasing temperature from 20 °C to 45 °C, followed by a sharp increase in FLI up to a maximum around 60 °C which decreased again from 60 to 80 °C. The sharp increase in FLI around 45 °C is indicative of the melting of the protein, where buried tryptophan residues are exposed, resulting in an increase in fluorescence.

The majority of RosettaDesign stream mutants followed the same FLI profile as WT. Mutants F13A, M126S and I24A had a lower starting FLI at 20 °C, whereas all other mutants were comparable to WT. The melt onset temperature (sharp increase in FLI) was lower for all mutants with the exception of M137L, whose FLI overlaid with WT. The maximum intensity of FLI around 60 °C was higher for mutants F13A, C17A, I24A and P132E compared to WT, whereas it is comparable for Q120I, M126S, and M137L.

The majority of Rosetta_ddg_monomer stream mutants also followed the same FLI profile as WT, however mutants G51R, T38W and S80W had a high starting FLI at 20 °C. Mutants S12W, L71W, Q131F, S155Y and D104Y had moderately high starting FLI at 20 °C compared to WT, whereas mutants Q107Y and S164L had a lower starting FLI at 20 °C.

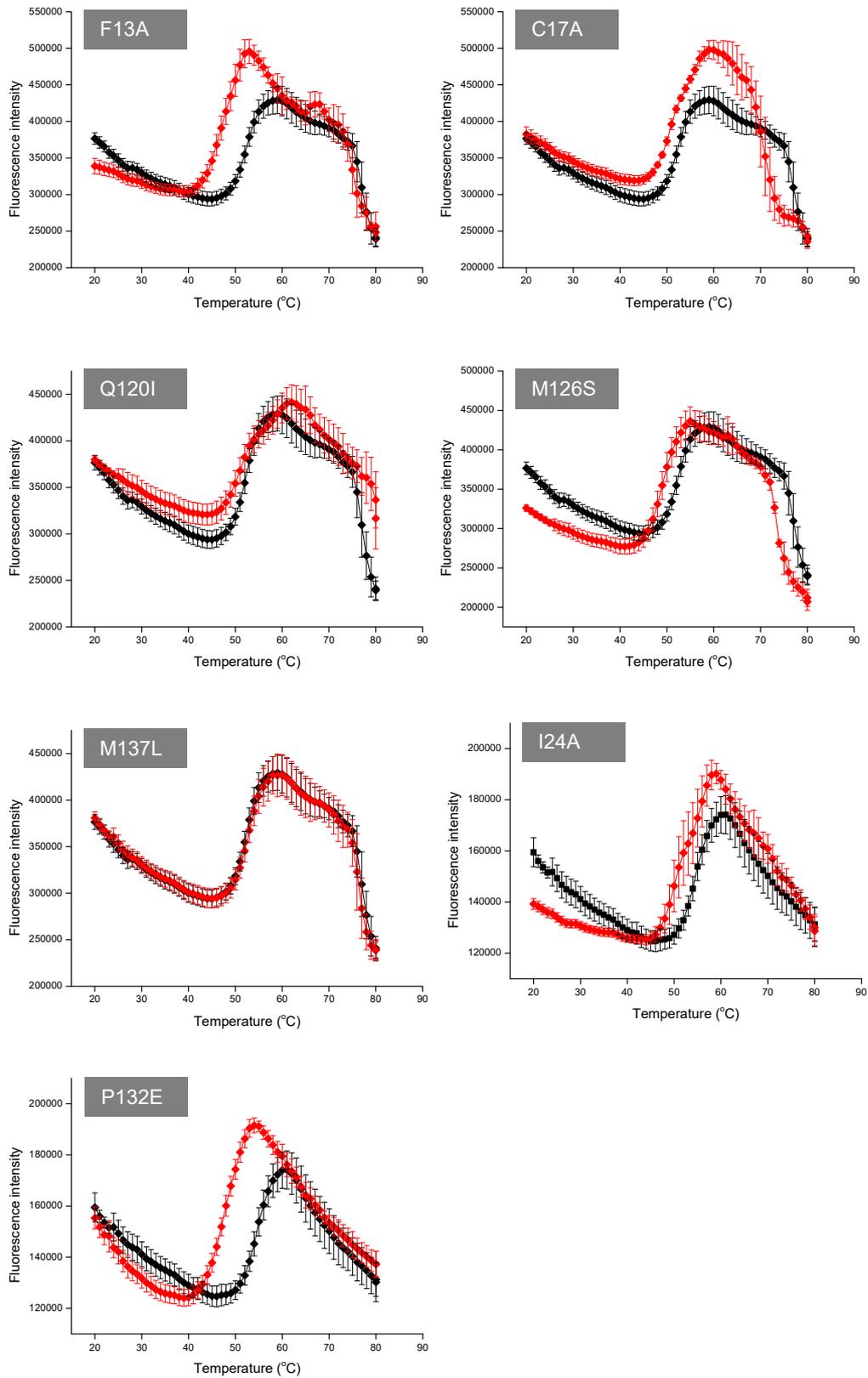


Figure 38. **Fluorescence intensity measurements of RosettaDesign mutants relative to WT GCSF.** Mutant analysed in each figure are named in the grey boxes. Fluorescence was during a thermal ramp from 20 to 80°C at a linear rate of 1 °C/min. All mutant (red) and WT (black) samples were formulated at 1 mg/mL in 10 mM sodium acetate and analysed five times each, apart from I24A and P132E whose samples were formulated at 0.3 mg/mL due to sample concentration issues.

The melt onset temperature as seen by a sharp increase in FLI was higher for mutants L71W, D104Y, Q107Y and S164L, comparable for G51R, T38W and S155Y, and lower than WT for S12W and Q131F. The maximum intensity level reached at ramp temperatures above 50 °C was higher for only mutant Q131F whereas it was comparable for S12W, G51R and S155Y and lower for T38W, L71W, S80W, Q107Y, D104Y and S164L.

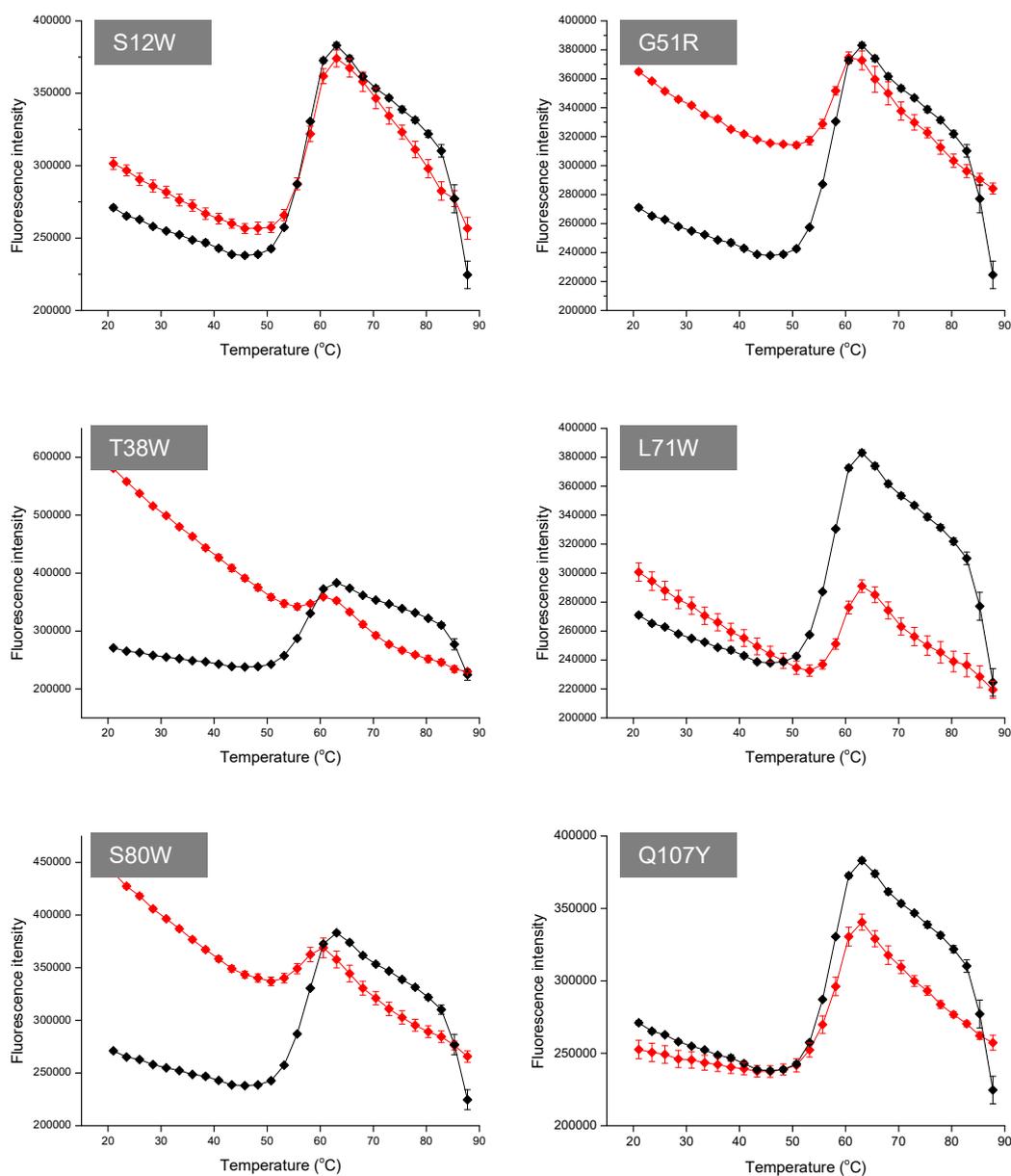


Figure 39. Fluorescence intensity measurements of Rosetta_ddg_monomer mutants relative to WT GCSF 1. Mutant analysed in each figure are named in the grey boxes. Fluorescence was during a thermal ramp from 20 to 80°C at a step rate of 1 °C/min. Mutant (red) and WT (black) samples were formulated at 1 mg/mL in 10 mM sodium acetate and analysed five times each.

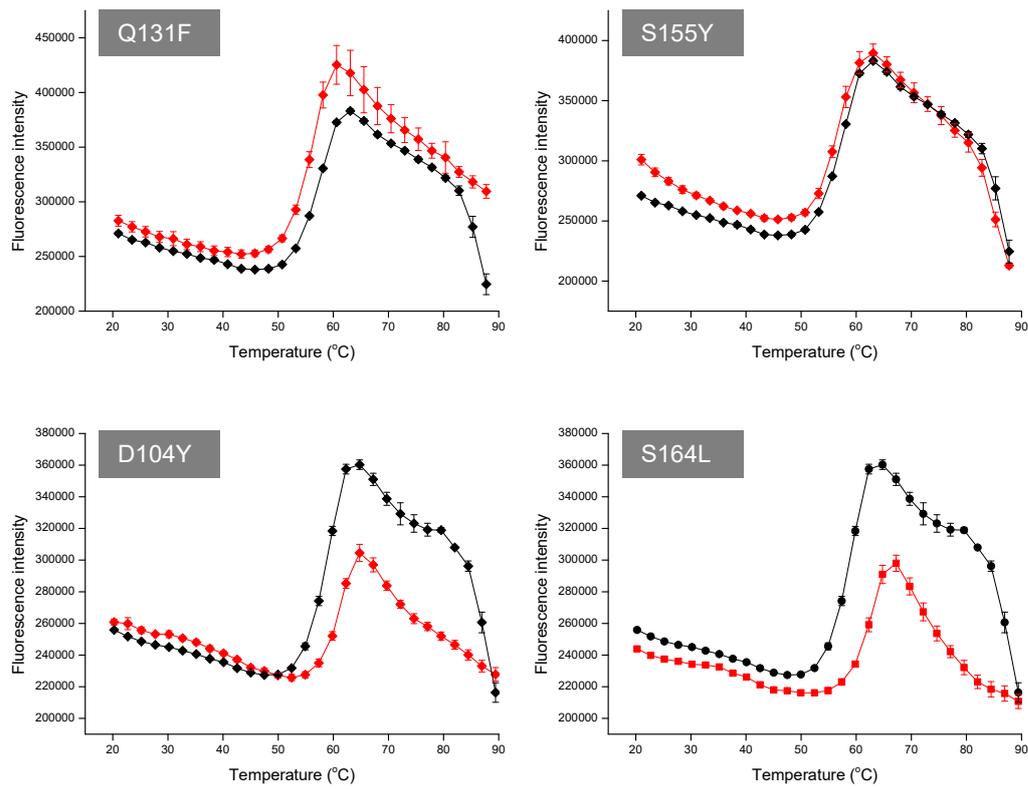


Figure 40. Fluorescence intensity measurements of Rosetta_ddg_monomer mutants relative to WT GCSF 2. Mutant analysed in each figure are named in the grey boxes. Fluorescence was during a thermal ramp from 20 to 80°C at a step rate of 1 °C/min. Mutant (red) and WT (black) samples were formulated at 1 mg/mL in 10 mM sodium acetate and analysed five times each.

4.1.4.4.2 Static light scattering at 266 nm

The average static light scattering (SLS) signal for each mutant versus WT at the same sample concentration is displayed in Figure 41 to Figure 45. The SLS signal intensity is proportional to the mean solute particle mass, as such an increase in SLS signal intensity reflected an increase in particle mass weight by aggregation. SLS allowed for both aggregation onset (T_{agg}) as well as extent of aggregation (intensity) to be determined. The WT GCSF sample aggregated around 48 °C followed by an exponential scatter signal and plateau at 75 °C, and the maximum intensity was around 22,000. For the RosettaDesign mutants in Figure 41 and Figure 42, all mutants, apart from C17A, also reached the same level of SLS intensity as WT. Mutant C17A had a higher maximum intensity than WT which suggested the sample concentration was higher than 1 mg/mL. Mutant C17A also had two stages of aggregation, the first around 46 °C and the second around 75 °C, where the SLS increased above 20,000. The mutant Q120I also displayed a two stage aggregation but had a higher T_{agg} compared to WT. As with the FLI, the SLS profile for M137L overlaid with WT. All other RosettaDesign mutants had a lower T_{agg} than WT. Finally, the SLS profiles for mutants I24A and P132E were noisy due to the low protein concentration of samples.

For the Rosetta_ddg_monomer stream SLS profiles in Figure 43 to Figure 45, the mutant S164L increased SLS intensity relative to WT, mutants T38W and S155Y were comparable to WT and mutants S12W, G51R, L71W, S80W, Q120I and Q131F decreased SLS intensity. For changes to the T_{agg} , the mutants G51R, L71W, S80W and Q120I showed shifts in the SLS curve the right, increasing T_{agg} . Interestingly, these mutants along with Q131F caused a two stage aggregation profile as seen with C17A and Q120I, whereas WT had a single T_{agg} around 48 °C. Mutants S12W, T38W and S155Y had no effect on T_{agg} , whereas mutant Q131F showed a clear decrease in T_{agg} . Mutants D104Y had a high level of noise suggesting protein concentration was lower than recorded during A_{280} analysis. In addition, mutant D104Y had a high starting SLS signal, which could have been caused by larger MW species present in the starting sample.

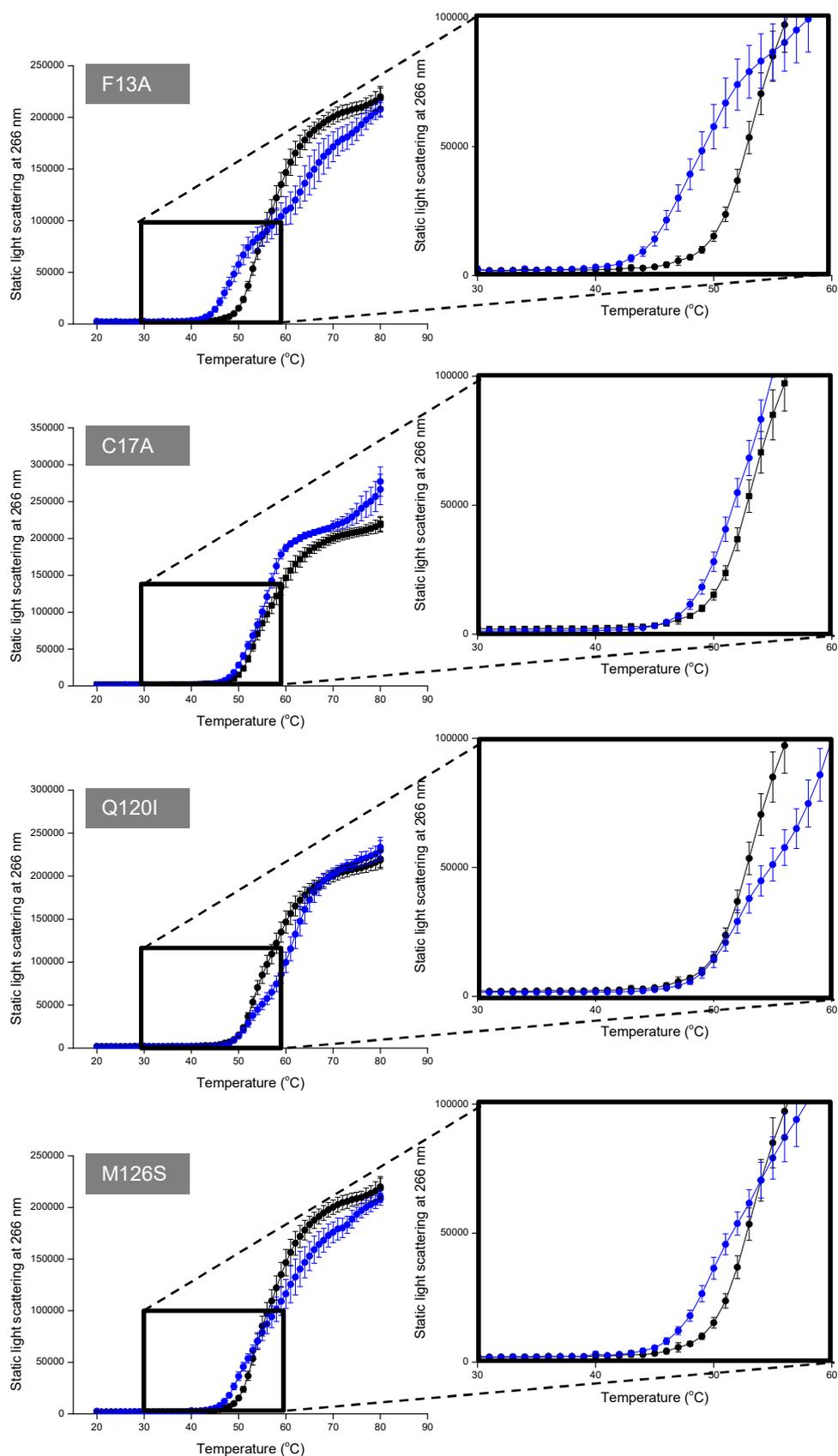


Figure 41. **Static light scattering of RosettaDesign GCSF mutants in comparison to WT 1.** Mutant analysed in each figure is named in the grey box. Static light scattering measured at 266 nm during a thermal ramp from 20 to 80°C at a step rate of 1 °C/min. Mutant (blue) and WT (black) samples were formulated at 1 mg/mL in 10 mM sodium acetate were analysed five times.

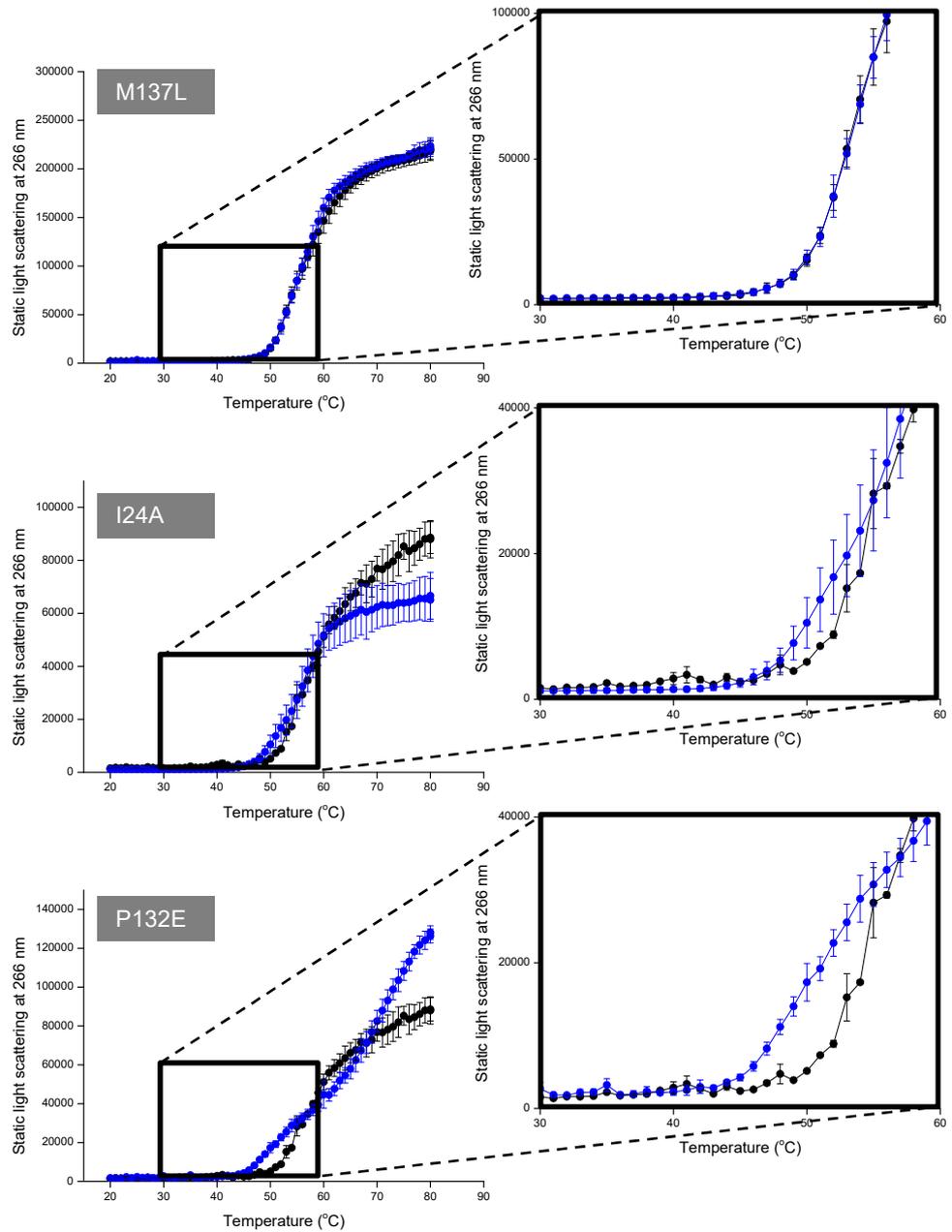


Figure 42. Static light scattering of RosettaDesign GCSF mutants in comparison to WT 2. Mutant analysed in each figure is named in the grey box. Static light scattering measured at 266 nm during a thermal ramp from 20 to 80°C at a step rate of 1 °C/min. Mutant (blue) and WT (black) samples were formulated at 1 mg/mL in 10 mM sodium acetate for mutant M137L and 0.3 mg/mL for I24A and P132E.

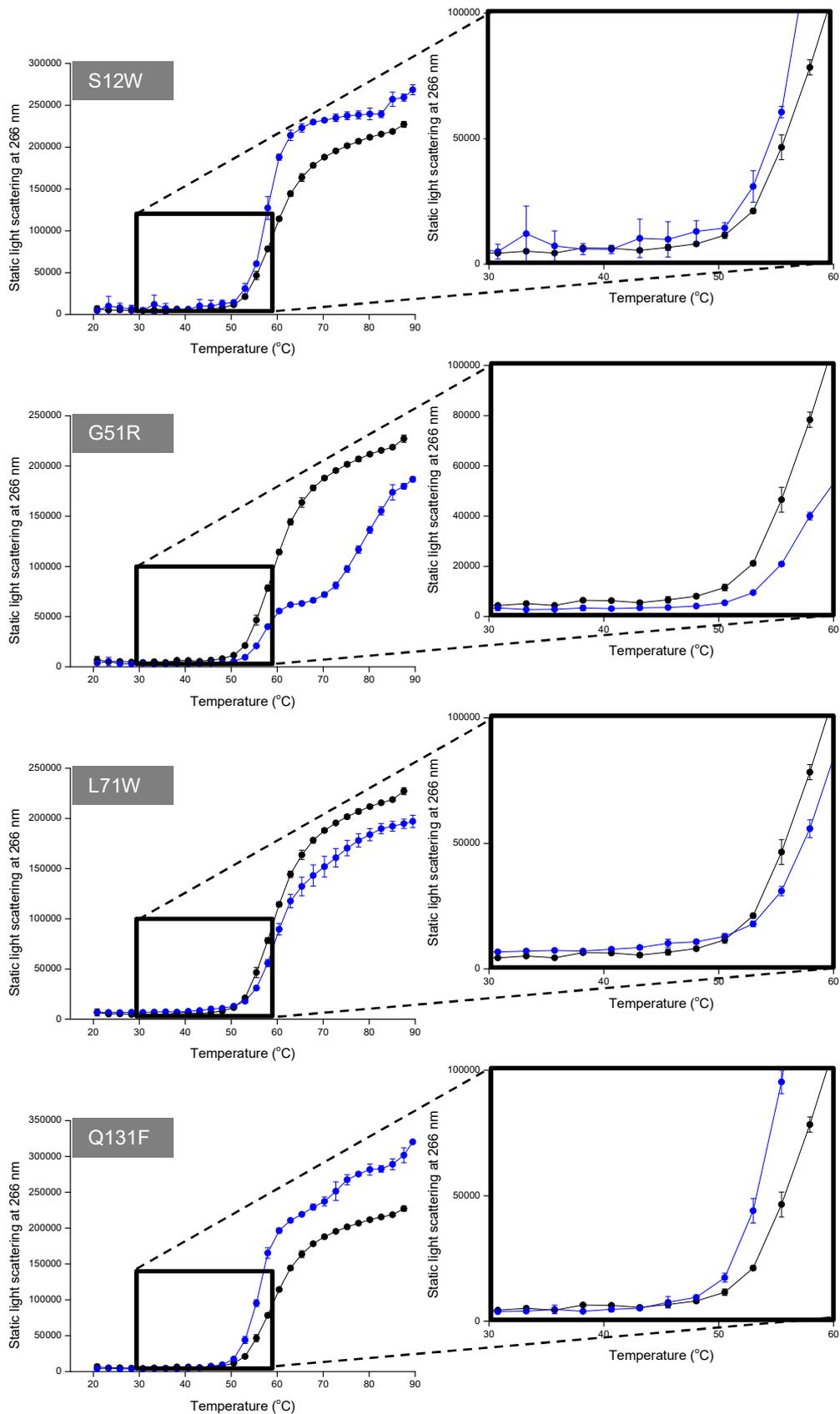


Figure 43. Static light scattering of Rosetta_ddg_monomer GCSF mutants in comparison to WT 1. Mutant analysed in each figure is named in the grey box. Static light scattering measured at 266 nm during a thermal ramp from 20 to 80°C at a rate of 1 °C/min. Mutant (blue) and WT (black) samples were formulated at 1 mg/mL in 10 mM sodium acetate were analysed five times.

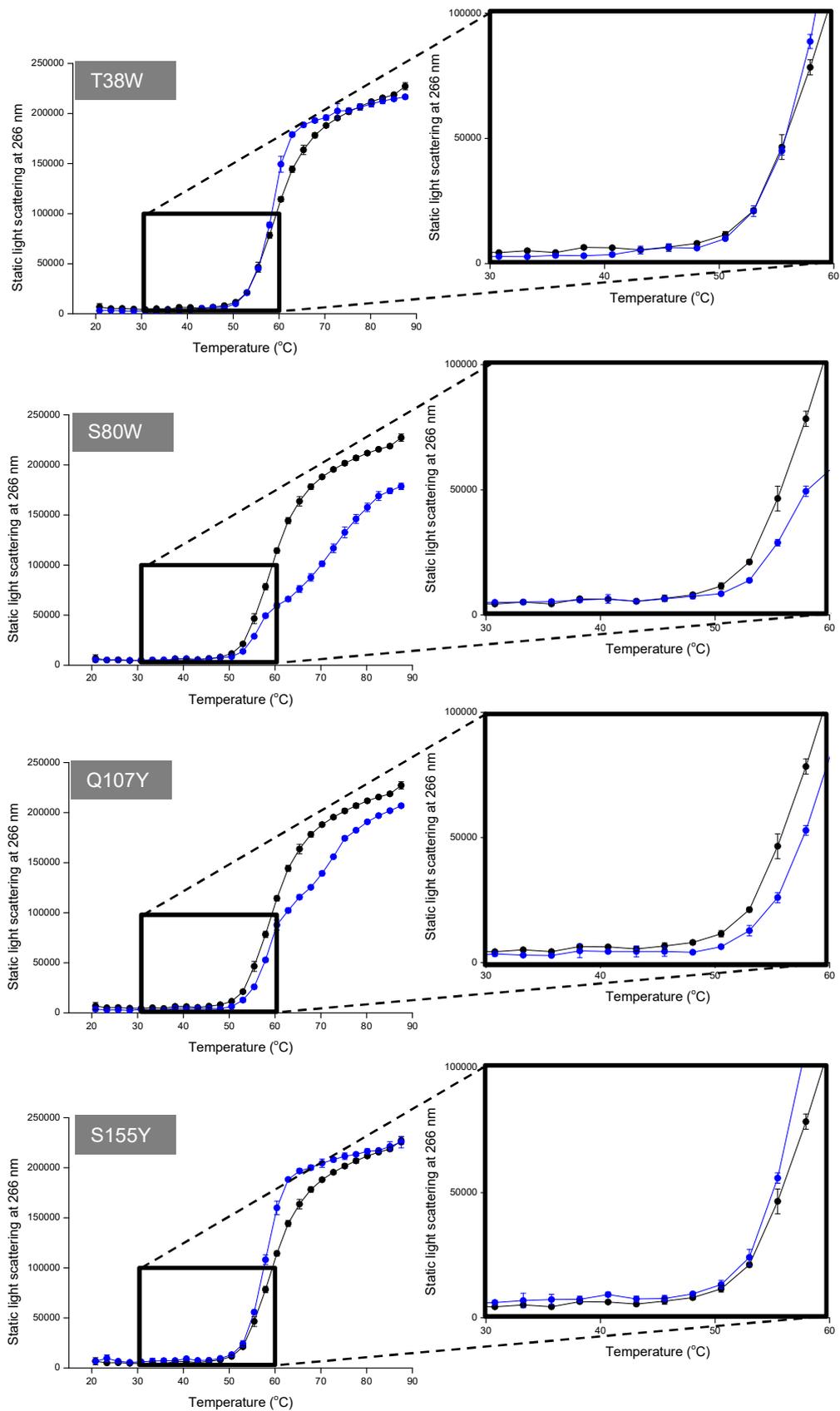


Figure 44. Static light scattering of Rosetta_ddg_monomer GCSF mutants in comparison to WT 2. Mutant analysed in each figure is named in the grey box. Static light scattering measured at 266 nm during a thermal ramp from 20 to 80°C at rate of 1 °C/min. Mutant (blue) and WT (black) samples were formulated at 1 mg/mL in 10 mM sodium acetate were analysed five times.

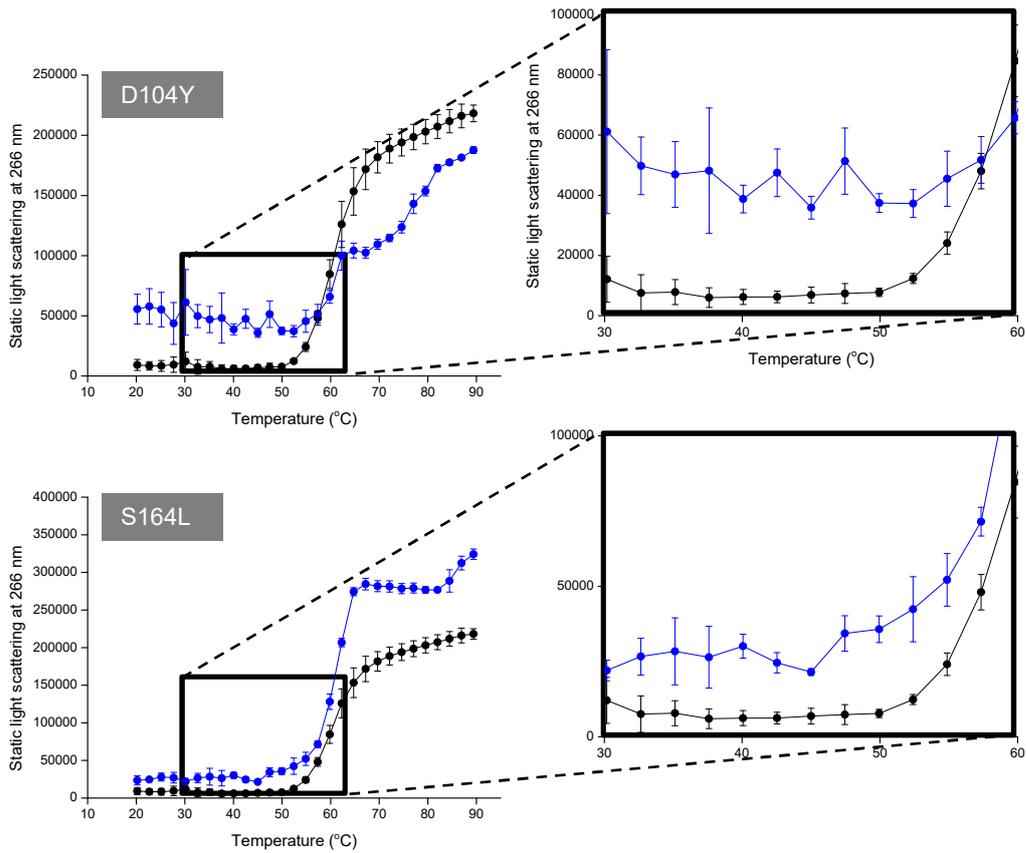


Figure 45. Static light scattering of Rosetta_ddg_monomer GCSF mutants in comparison to WT 3. Mutant analysed in each figure is named in the grey box. Static light scattering measured at 266 nm during a thermal ramp from 20 to 80°C at a rate of 1 °C/min. Mutant (blue) and WT (black) samples were formulated at 1 mg/mL in 10 mM sodium acetate were analysed five times.

4.1.5 Hydrophobicity

RosettaDesign stream mutants C17A and Q120I were analysed along with WT, by reverse-phase HPLC to determine changes to surface hydrophobicity upon point mutation. From overlaid chromatograms in Figure 46, WT GCSF eluted at the same time as C17A indicating no changes in hydrophobicity upon mutation of the free cysteine. In contrast, Q120I eluted later than WT and C17A, requiring a lower polarity environment for detachment from the column, which was attributable to increased surface hydrophobicity.

Post-translational modifications were also observed from the RP-HPLC chromatogram data, where all samples analysed had a large unmodified protein peak and a smaller tailing peak (F-met; Herman et al. 2002). Mutant Q120I was almost identical to WT in peak distribution with an unmodified protein content of 94.4 and 94.6% respectively, whereas for C17A the unmodified protein content was higher than WT at 95.9%.

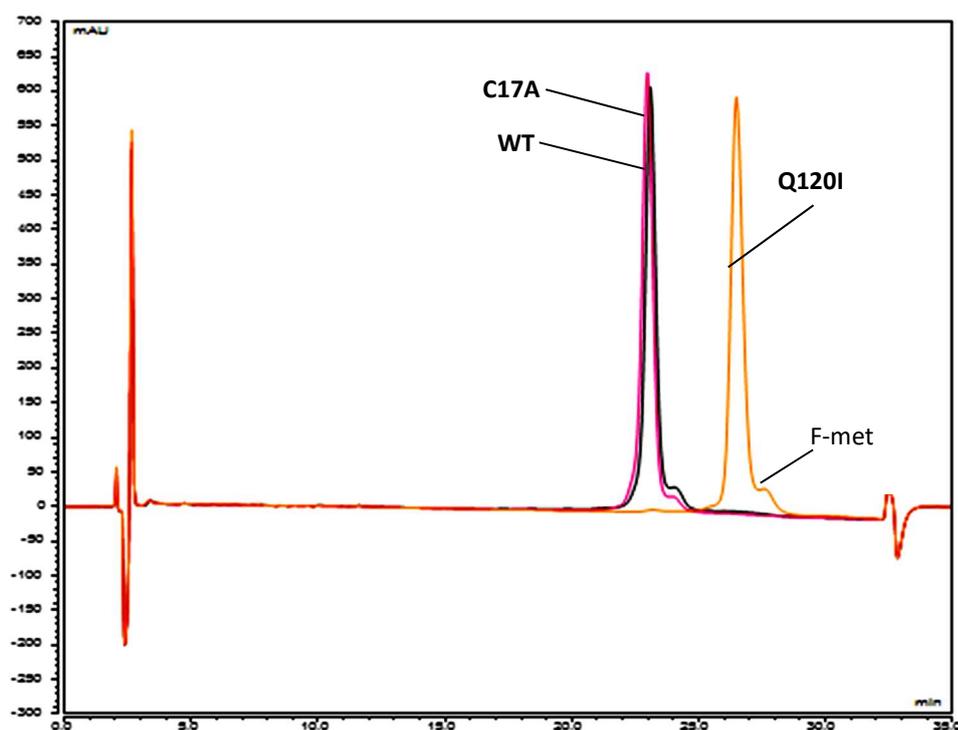


Figure 46. RP-HPLC of WT, C17A and Q120I. Samples were injected at 1 mg/mL in a volume of 20 μ L onto a C4 RP-HPLC column and eluted with an increasing gradient of acetonitrile in 0.1% formic acid. Mutants C17A and Q120I are represented by the colours pink and orange, respectively and WT in black. The main elution peaks show unmodified species and the tail represents F-met GCSF as highlighted.

Mutant characteristics summary

From the characterisation performed on mutant samples, critical values were obtained and summarised Table 15. Values were relatively colour coded within design streams with a red_yellow_green to visualise destabilising (red) and stabilising (green) qualities.

For the RosettaDesign stream, the majority of mutants generated were of relatively high yield and purity, and bioactivity was not affected by mutation. Two mutants displayed poor yield, reflected by the presence of multiple bands during SDS-PAGE analysis. These mutants also had reduced bioactivity relative to WT, but not a loss of total bioactivity, which suggests the decrease in monomer by degradation, could be the cause of the reduced bioactivity. The low yield, purity and bioactivity were also echoed in the negative values for both ΔT_m and ΔT_{agg} and as such these mutants are confirmed to destabilise GCSF. Of the mutants with high yield, purity and relative bioactivity, two mutants had a positive ΔT_m (C17A and M137L), and two mutants had increased ΔT_{agg} (Q120I and M137L) totalling three novel thermostable GCSF mutants.

For the Rosetta_ddg_monomer stream, the majority of mutants generated were also of relatively high yield and purity, and bioactivity was not affected by mutation. Four mutants, S12W, S80W, D104Y and S164L had a lower than average yields but had one band during SDS-PAGE analysis indicating high purity. Mutant L71W had the highest yield of the Rosetta_ddg_monomer stream mutants by a large margin, its bioactivity was comparable to WT, however it provided no improvement to thermal stability. The low yielding G28I was so destabilising that there was not enough material for further analysis post SDS-PAGE where 7 bands were present. Mutant S80W showed high purity but was the only mutant to have lower bioactivity relative to WT. The mutant also showed a decrease in T_m but oddly, an increase in T_{agg} . A similar phenomenon of high T_{agg} and low T_m was seen with the RosettaDesign mutant Q120I. The ΔT_{agg} data was not available for D104Y and S164L due to high noise in the samples obscuring the aggregation onset point. Of the mutants with high yield, purity and relative bioactivity, three mutants had positive ΔT_m values (L71W, D104Y and S164L), and six mutants had increased ΔT_{agg} (T38W, G51R, L71W, Q107Y, and S155Y) totalling seven novel thermostable GCSF mutants.

Table 15. **GCSF mutant characterisation summary table.** Values were calculated relative to the WT sample purified at the same time as the mutant and unavailable data is listed as N/A. Size exclusion chromatography (SEC) data was used to determine yield % ($\text{Value}_{\text{Mutant}} / \text{Value}_{\text{WT}} * 100$). Non-reduced SDS-PAGE gel images were used for purity determination. The GNFS-60 cell proliferation bioassay was used to determine bioactivity relative to WT. Fluorescence intensity (FLI) at 280-420 nm and static light scattering (SLS) at 266 nm were performed in tandem to determine T_m and T_{agg} values, respectively ($\Delta\text{Value} = \text{Value}_{\text{Mutant}} - \text{Value}_{\text{WT}}$).

Stream	Sample	SEC1		(NR) SDS-PAGE2		GNFS-60 cell proliferation bioassay3		FLI		SLS at 266 nm	
		Yield	Purity	Activity	ΔT_m	SD	ΔT_{agg}	SD			
RosettaDesign	F13A	50%	1	-1	-3.44	0.21	-5.31	0.35			
	C17A	136%	1	1	1.55	0.23	-1.01	0.31			
	I24A ¹	37%	3	-1	-1.64	0.23	-2.09	0.99			
	Q120I	75%	1	1	-2.28	0.74	0.31	0.34			
	M126S	92%	1	1	-2.75	0.15	-2.57	1.24			
	P132E ¹	25%	3	-1	-6.98	0.23	-2.73	0.55			
	M137L	92%	1	1	0.62	0.13	0.31	0.40			
	S12W	88%	1	1	-0.99	0.28	-1.45	3.35			
	G28I	14%	7	N/A	N/A	N/A	N/A	N/A			
	T38W	105%	2	1	-0.07	0.49	0.69	1.16			
Rosetta_ddg_monomer	G51R	124%	2	1	-1.66	0.32	3.28	1.12			
	L71W	215%	1	1	0.19	0.28	0.61	1.12			
	S80W	66%	1	-1	-2.15	0.56	2.07	0.38			
	D104Y ¹	84%	1	1	2.04	0.32	N/A	N/A			
	Q107Y	131%	1	1	-0.26	0.31	2.25	0.92			
	Q131F	95%	1	1	-2.71	0.07	-2.68	1.57			
	S155Y	122%	1	1	-0.66	0.28	4.49	1.34			
	S164L ¹	55%	1	1	3.69	0.31	N/A	N/A			

¹ Determined by comparing the height of monomer elution peak and calculating the difference between either the mutant and WT for RosettaDesign stream and between the mutant and the average mutant peak height for Rosetta_ddg_monomer stream. ² Number of bands observed on a non-reduced SDS-PAGE 4-12% Bis Tris gel with MES running buffer, stained with InstantBlue™. ³ Statistical difference between mutant and WT GNFS-60 cell proliferation assay curves as determined by parallel line analysis. ⁴ Samples were run at a lower concentration of 0.3 mg/mL and compared to a WT sample also at 0.3 mg/mL. ⁵ Samples were run separately to the other mutants in the stream. WT was included.

4.2 Discussion

The design, selection and production of GCSF mutants was performed in two streams in this chapter. The first looked to use the online server RosettaDesign, requiring the setup of each mutation prediction one at a time. A combination of B-Factor and HDX-MS data was used to target a small number of regions for RosettaDesign input. Regions were selected based on higher than average flexibility values. The second stream used a High Performance Computing Facility which removed the need to target specific small regions. All possible mutations spanning the entire length of the protein were generated in a number of days using a single job script.

The work described in this chapter, was the first to use the crystal structure of human GCSF (2D9Q) as a template for *in silico* screening. Previous GCSF mutagenesis screening work performed by Luo et al. (2002) used the PDB 1BCG (bovine GCSF) as their template structure. As shown by the alignment of PDB amino acid sequences in the previous chapter, 1BCG was only 87% similar to human GCSF, meaning Luo et al. (2002) had to replace a number of bovine residues with the human residues for those positions, followed by side chain optimisation and structure minimisation. Whereas, starting with the exact structure which has also been shown to dynamically align with the experimental protein in this work reduced the risk of design errors.

4.2.1 RosettaDesign

The RosettaDesign online server was selected for the first design stream due to ease of access and simplicity of user interface; however a drawback of the method was the manual input of individual resfiles for each residue. To complete this for the entire protein with all possible substitutions would require 3307 individual resfiles. As this is not time or energy feasible, specific regions of the protein were selected for rational stabilisation, and the resfiles designed so that the output provided the most stabilising mutation rather than exploring all possible amino acid substitutions.

4.2.1.1 Target selection

Three regions were selected for input into RosettaDesign, one unstructured and two structured. LoopCD was the unstructured region selected as it had the highest average B-factor and HDX rate, indicating it was the most dynamic region of the protein. This rationale was in agreement with NMR work by Zink et al. (1994), who found small nuclear overhauser effect (NOE) and long transverse relaxation time (T_2) values for many residues in the CD loop region during ^1H and ^{15}N NMR spectroscopy meaning they excerpted rapid

motion in the protein backbone. Furthermore, O-glycosylation of GCSF (Lenograstim) was shown by Oheda et al. (1988), by NMR, to be linked to the residue T133, which is found within LoopCD. Glycosylation of GCSF was shown by Gervais et al. (1997) to increase physical stability by reducing local mobility around the glycosylation site, suggesting this region could benefit from rigidification to increase stability. The structured region, α C, was selected due to its relatively high B-factor and HDX rates, as well as high variability. The high variability of α C demonstrated a structural discrepancy between residues within the helix worthy of RosettaDesign stabilisation. The structured region, α A, was included for the opposite reasons to α C - it had relatively low B-factor and HDX rate. By NMR α A was shown to have high NOE and low T_2 (Zink et al. 1994), the opposite to loopCD. The inclusion of a relatively immobile structure, as seen by both low HDX % rate and low average B-factor, served to verify if flexibility is a key determinant of stability.

4.2.2 RosettaDesign outcome

In this chapter, the RosettaDesign stream produced seven GCSF mutants, 71% had high yield and high purity with comparable relative activity to WT and 43 % showed increased thermal stability. The three mutants with increased thermal stability are discussed individually below.

4.2.2.1 C17A

Of all GCSF mutants produced, C17A was the only non-novel substitution (Kuga et al. 1989; Luo et al. 2002; Raso et al. 2005; Jiang et al. 2011A). The mutation sees the removal of the free cysteine, and substitution of a small non-polar alanine residue. During purification, C17A had a 35% increase in yield relative to WT, and was shown to exist as a single pure band by SDS-PAGE, indicating successful refold and purification. A study by Raso et al. (2005) found no presence of cross-linked aggregates during SEC analysis of C17A degradants with the addition of SDS suggesting the mutant increased yield due to the reduction of intra-molecular disulfide bond mismatches during the refolding process.

The bioactivity of C17A was also higher than WT. Whilst this could be due to variability of biological based assays, a study by Jiang et al. (2011A) found their mutant "GCSFa", containing C17A, induced a higher level of absolute neutrophil counts (ANC) than WT did when examining neutrophil levels *in vivo*. They suggested the increase in bioactivity was also due to the reduction of inter-molecular disulphide bonds.

C17A was shown to have a higher T_m but a lower T_{agg} compared to WT. The lower T_{agg} could be caused by a higher concentration of the C17A sample relative to WT, where, as seen in chapter 3, the higher the GCSF concentration the lower the T_{agg} . The C17A study by Raso et al. (2005) found no significant difference in stability of C17A and WT molecules during equilibrium denaturation. However, within the same study, Raso et al. (2005) showed the aggregation kinetics for C17A was slower than WT at the same concentration but they ultimately aggregated to the same extent. On the other hand, Luo et al. (2002), who also investigated the mutagenesis of GCSF, included C17A in their computational re-engineering of more stable molecules. They created a mutant with both C17A and G28A, which improved T_m by 5 °C. However, they did not separate the two mutations into single mutations and test for increases in T_m . It could have been that the G28A caused the increase in stability; however, it suggests C17A could be a synergistic mutation, where increasing refold efficiency yields more protein and a more stable structure.

4.2.2.2 *Q120I*

Mutant Q120I is the first novel, thermostable GCSF mutant produced by the RosettaDesign stream. The mutation substituted a polar glutamine residue for an aliphatic/hydrophobic isoleucine residue, which returned the largest improvement to T_{agg} but had minimal effect on T_m . A study by Koide et al. (1992) found an isoleucine residue at position 23 in human epidermal growth factor (hEGF) was exposed on the protein surface and directly interacted with a hydrophobic pocket of receptor. As shown by the RP-HPLC in this chapter, Q120I eluted later than WT GCSF in a higher concentration of organic phase. The increase in retention time indicates the residue Q120I is partially surface exposed within the 3D structure of GCSF, as such it is hypothesised that the increased hydrophobicity of the substitution decreased intermolecular associations around the region.

4.2.2.3 *M137L*

The mutant M137L is the second, novel thermal stable GCSF mutant produced in this chapter. The mutation substituted a hydrophobic methionine residue with an aliphatic/hydrophobic leucine residue within loopCD. During purification, M137L had a similar yield to WT and the bioactivity was identical, confirming the mutation did not affect biological function. Unlike C17A, the mutant M137L increased T_m and T_{agg} . Although the side chains are non-polar for both methionine and leucine, the solvent transfer free energy for leucine was reported by Fauchere and Pliska (1983) to be 0.6 kcal/mol greater than methionine. This suggests improved folding occurred by substitution with the leucine residue, which increased T_m . A study by Lipscomb et al. (1998) found the substitution of M→L increased

the stability of T4 lysozyme but only for methionine residues within the interior or partially buried within the protein due to the reduced entropic cost of holding a leucine side chain in a defined position. As M137L in GCSF caused improved thermal stability this suggests the residue M137 could be partially buried.

4.2.3 Rosetta_ddg_monomer

The Rosetta_ddg_monomer application was used on Legion@UCL to predict the $\Delta\Delta G$ of all possible single point GCSF mutations. Runs were queued, rather than inputting each one manually and the HPCF ran multiple Rosetta_ddg_monomer runs in parallel reducing the processing time to just under a week. The Rosetta_ddg_monomer on the HPCF also ran 50 trajectories per mutation, rather than the 10 obtained during RosettaDesign.

Rosetta_ddg_monomer output analysis found 73% of residues were amenable to a stabilising mutation. Analysis also found any substitution of the amino acids isoleucine and tryptophan was destabilising to the folded structure. Tryptophan is unique in terms of chemistry and size, meaning that often replacement would most likely destabilise the structure (Betts and Russel 2003) and isoleucine is bulky and hydrophobic, suggesting it is an integral residue to the hydrophobic core of the protein. Tyrosine, tryptophan and phenylalanine were the amino acids selected most frequently for as a stabilising mutation, which is again most likely due to Rosetta predicting that their hydrophobic properties would increase protein core packing. Proline was shown to have a large destabilisation effect for most residue positions. Zhang (2017) also found proline substitutions were least preferred by Rosetta when studying A33 Fab. Proline is a very rigid amino acid and has to be near the surface, or in a loop region, in order to have minimal disruptive influence on its surrounding residues, which limits its incorporation. Out of the lowest $\Delta\Delta G$ mutations listed, proline was only selected only 3 times (average stabilising effect of -1.19 ± 0.26) at positions S7, A123P and A133. These residues were located at the N-terminal loop, and ends of the loopCD. As discussed earlier, loopCD has been shown to be highly mobile, however, small NOE and maximum T_2 values were also found to be at the N-terminus (Zink et al. 1994). The N-terminus has also been shown to not have significant electron density to obtain a crystal structure, reflecting its dynamic behaviour (Lovejoy et al. 1993; Hill et al. 1993; Tamada et al. 2006). As such, the addition of proline at these positions could in theory stabilise GCSF by reducing the flexibility of these regions.

4.2.4 Rosetta_ggd_monomer outcome

In this chapter, the Rosetta_ddg_monomer stream produced eleven novel GCSF mutants, 64% had high yield and purity. All mutants, except for S80W had comparable relative activity to WT. In addition, 64% showed increased thermal stability.

4.2.4.1 *T38W and L71W*

Mutants T38W and L71W saw the residues threonine and leucine substituted with a tryptophan residue. The addition of another aromatic tryptophan residue increased the fluorescence intensity reading of the mutants compared to WT at low temperature; however this was taken into account by the calculation of the extinction coefficients during concentration determination. Even with the change in fluorescence accounted for, mutant L71W was the highest yielding candidate of the Rosetta_ddg_monomer collection. Threonine, leucine and tryptophan are all hydrophobic residues, however tryptophan has additional aromatic characteristics, is capable of non-covalent aromatic ring stacking (π - π interactions), as well as weakly polar interactions (Burley and Petsko 1988). These additional characteristics suggest tryptophan creates a better fit for increased GCSF structural stability. High tryptophan content has been previously described for stabilisation of membrane proteins, where tryptophan residues are concentrated in particular regions and form H-Bonds with carbonyl oxygens of the main chain (Schiffer et al. 1992).

4.2.4.2 *G51R*

The mutant G51R provided the second highest T_{agg} improvement of the Rosetta_ddg_monomer stream. The mutation substituted a non-polar, glycine residue with one of the most complex amino acids, arginine. The residue G51 is located within the 11 residue, short helix and in close proximity with GCSFR binding residues, yet activity was shown to not be affected. The substitution also caused an increase in yield relative to other mutants. A crystallography study by Lovejoy et al. (1999) comparing recombinant bovine GCSF and recombinant human GCSF, found a sharp kink at residues 51 and 50, which disrupted the α -helical main chain H-Bond pattern of both proteins. The mutant G51R was also unusual in the production of a two stage aggregation profile during thermal ramp SLS experiments, compared to the standard one stage aggregation for WT. The intensity of the G51R aggregation profile was also a lot lower than WT. As G51R increased T_{agg} and decreased T_m relative to WT and other mutants studied, results indicate that the short helix may be a hotspot for protein aggregation.

4.2.4.3 *D104Y and Q107Y*

The mutants D104Y and Q107Y were both located in α C, which was shown to be a highly flexible region by HDX-MS and B-factor analysis. The mutations included the removal of polar residues aspartic acid and glutamine with the polar residue tyrosine. As with G51R, D104Y and Q107Y changed the SLS aggregation profile to two distinct aggregation stages during thermal stability assessment. The mutants also had lower aggregation intensity. The mutant D104Y was shown to provide the second largest improvement in T_m , behind S164L, found in α D. As discussed for tryptophan substitutions, tyrosine also has polar and hydrophobic properties, providing additional residue interactions, which presumably provided increased structural rigidity.

4.2.4.4 *S155Y and S164L*

The amino acid serine was replaced by tyrosine and leucine in the mutants S155Y and S164L, respectively. Serine is a small, polar residue and the substitution to the polar/hydrophobic tyrosine residue presumably increased neighbouring residue interactions, increasing structural stability as previously described. For leucine, the amino acid residue has hydrophobic and aliphatic properties, which again suggests the substitution of serine at this position to a more hydrophobic residue was the reason for the stabilisation. Both mutants were located in α D and S164L was found to have the largest improvement in T_m whereas S155Y had the greatest improvement in T_{agg} . This suggests that this particular helix could be hotspot for GCSF instability.

4.2.5 Rigidification as a strategy for stabilisation

Three mutants increased stability of GCSF within the RosettaDesign stream: C17A, Q120I and M137L. With regards to their position C17A was located in α A, Q120I was in α C and M137L in loopCD. As C17A is found within α A, this suggests flexibility information may not be useful for rational design for increased stability, as α A was shown to have low flexibility, however, the removal of the free cysteine is a common, simple stabilisation strategy to prevent the formation of unwanted intermolecular or intramolecular disulfide bonds (Marshall et al. 2003). This suggests that the stabilisation at this position was not due to rigidification of a non-flexible region and would have improved stability if positioned at any region within the protein sequence. The stabilisation of Q120I and M137L in the two flexible targeted regions of GCSF suggests selecting flexible regions for rational target selection could lead to more likely improvements in stability, however the sample size was too small for it to be significant on its own.

Seven mutants increased stability of GCSF within the Rosetta_ddg_monomer stream: T38W, G51R, L71W, D104Y, Q107Y, S155Y and S164L. With regards to their positions, T38W was within the loopABI, G51R in the short helix, L71W in loopABII, D104Y and 107Y α C, and S155Y and S164L in α D. Of these locations, the loopABII and short helix were in the High-High group for HDX-MS uptake % rate and B-factor values. In addition, α C and α D were found to have the highest HDX/B-factor values of the α helices and no mutants produced with substitutions within α A or α B showed increased thermal stability. These results combined with the RosettaDesign mutations in α C and loopCD suggest high flexibility correlates with increased thermal stability upon $\Delta\Delta$ G based mutation predictions.

4.2.6 Notes for future mutant production

A major issue for mutant production during this work was yield. Lab-scale expression and processing was required to generate sufficient characterisation material, however the large scale centrifuge available for this work could only hold six bottles at a time. This directly affected the number of shake flask cultures and IB wash samples that could be processed at the same time. For both design streams the number of combined mutant and WT samples was more than six, meaning cell culture, lysis, washing and refold were split into two processing groups with a day apart. Another major bottleneck in the purification of the mutants was during preparatory size exclusion chromatography step. Due to the large column volume and low maximum flow rate only two mutants could be feasibly purified per day. One purification was performed early in the morning and the other in the evening, with column cleaning during the day and overnight. Previous work incorporating high throughput mutagenesis have studied enzymes, which do not require as intense purification, as bioactivity and stability can be measured within a semi-clean lysate or after His Tag purification. These strategies were not possible for GCSF as during *E.coli* expression the protein is insoluble and needs refolding. Furthermore, bioactivity was determined using a cell proliferation bioassay, which requires high sample purity.

Improvements to the mutant production process could have been achieved by a) having analysis methods that require smaller quantities of material to be generated, b) including C17A in all mutants to reduce loss during refold, or c) having a higher-throughput chromatography step such as cation exchange chromatography (Bhambure and Rathore 2013; Babaeipour et al. 2015), which employs smaller columns, can tolerate higher flow rates and does not require sample concentration prior to loading.

Fortunately, even with the production issues, the method described in this thesis for producing pure, active GCSF was used to successfully produce 18 GCSF mutants and two lots of WT, where the majority of mutants and both WT samples were shown to be relatively pure by a single band visualised by (NR) SDS-PAGE.

5 GCSF stability mutants II: advanced biophysical analysis

The aim of this chapter was to apply the HDX-MS method optimised for GCSF in chapter 3, to a selection of GCSF stability mutants produced in chapter 4.

The selection of mutants produced from design streams (RosettaDesign and Rosetta_ddg_monomer) was based on both the stability and availability of each mutant. The first stage of further analysis confirmed the identity of the selected mutants using intact mass spectrometry. By comparing experimental mass to the predicted mass from the mutant amino acid sequence, determination of single point mutation success was achieved.

The second stage of further analysis explored the effect of mutation on GCSF shelf-life. Accelerated degradation is a commonly used technique to predict the shelf-life of a biotherapeutic drug within a shorter timescale than what could be years for a real-time study. Thermal stress was applied to the WT and mutant samples and comparisons between the degradation rates at different time points were assessed by SEC-HPLC and SDS-PAGE.

The final stage of further analysis used low pH, peptide level HDX-MS to compare local changes in mutant flexibility relative to WT. High flexibility has been previously correlated with protein instability, as such it was hypothesised that stabilising mutants would have increased structural rigidity, as seen by a reduction in the rate of deuterium uptake. Mutant HDX-MS data was then compared with shelf-life stability data and T_m and T_{agg} values obtained in chapter 4 to determine if rigidification was the main cause of increased thermal stability.

5.1 Results

WT GCSF and mutants were prepared in chapter 4. Due to losses during production, there was insufficient material for further analysis of WT GCSF from the Rosetta_ddg_monomer stream; as such WT GCSF from the RosettaDesign stream was used as a control for both sets of mutants.

5.1.1 Mutant selection

Mutants chosen for further characterisation during this chapter included WT, C17A and Q120 from the RosettaDesign group, and G51R, L71W, Q107Y and Q131F from the Rosetta_ddg_monomer group. Selection was based on thermal stability, purity and availability of samples, with the exception of C17A, which was selected as a mutant of interest due to the inconsistency in previous literature on the effect of this particular mutation on GCSF conformational stability.

5.1.2 Mutant identity

To confirm the identity of the mutants, IMS was performed in triplicate for each mutant and WT protein samples. The experimental mass difference between mutant and WT was plotted against the predicted mass difference for each particular amino acid substitution in Figure 47. The intact protein mass of WT GCSF was $18,799.7 \text{ Da} \pm 1.02$, which matched the theoretical mass exactly. The mass also coincided with that measured in chapter 4 for WT GCSF, demonstrating consistency between different production batches.

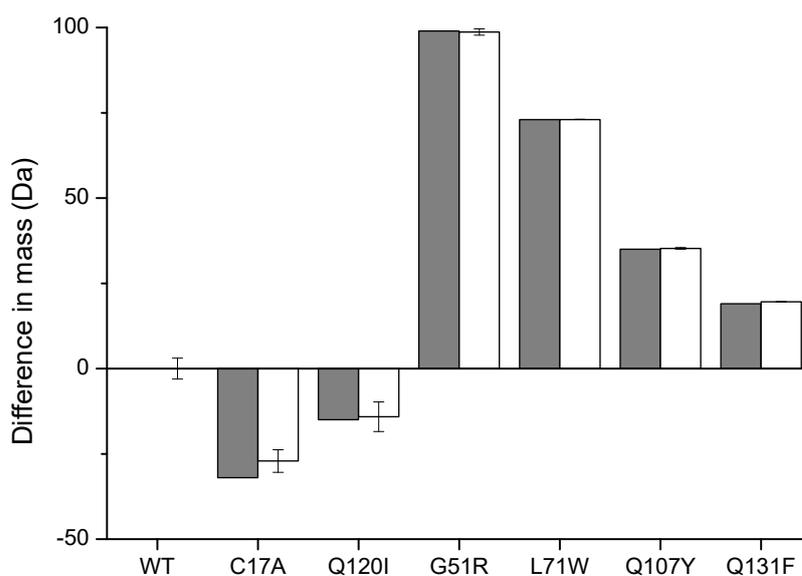


Figure 47. **Mass difference of GCSF mutants with WT by LC-QToF-MS.** Expected intact mass difference is displayed in grey and the average of three experimental intact mass difference in white with standard deviation.

All mutant intact protein mass differences were within 1 standard deviation of the predicted mass difference, apart from C17A, which had a smaller mass than predicted. This suggested no mis-incorporation of other amino acids, other than the one intended, had occurred for all mutants apart from C17A.

During deconvoluted peak analysis, along with the main peak, all samples had two other peaks present: one with a larger MW and one smaller. An example of this was shown in Figure 48 where the average difference in mass of the smaller peak was -18.13 Da and the larger peak was +17.72 Da for all samples. As both peaks were present in all samples, including WT, the species were most likely something occurring during intact protein mass analysis and can be disregarded.

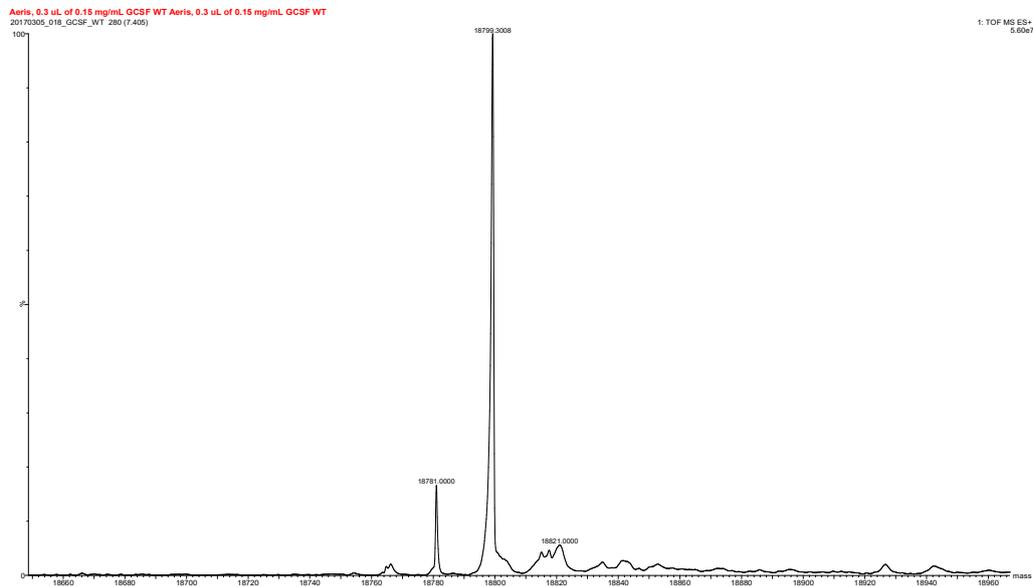


Figure 48. **WT GCSF IMS deconvoluted peak example.** The main three peaks are labelled with their deconvoluted masses in Da.

5.1.3 Accelerated thermal degradation

Mutant and WT samples were stored at elevated temperatures (37 °C and 45 °C) and analysed at days 0, 1, 2, 3, 6 and 7 for their different MW species content using SEC-HPLC.

5.1.3.1 Initial sample content

The different MW species present in the initial sample (time 0) as a % of the total sample is shown in Figure 49. All samples were predominantly monomeric (>80%), however, the dimer and/ or soluble aggregate level varied. The WT sample contained 96% monomer and 4% dimer, as did C17A, suggesting sample similarity. The other mutant from the RosettaDesign stream, Q120I, contained the largest amount of dimer of all samples at 8%, and a low amount of aggregate. Mutant G51R, from the Rosetta_ddg_monomer stream, contained the lowest amount of monomer at 91%, with 6.1% of the sample dimer and 2.9% aggregate. Mutant L71W contained 6.7% aggregate and minimal dimer, whereas mutant Q107Y had a low amount of aggregate of with 0.4% and a dimer content of 3.8%. Finally, mutant Q131F contained a low amount of dimer at 1.9% but the second largest amount of aggregate at 4%.

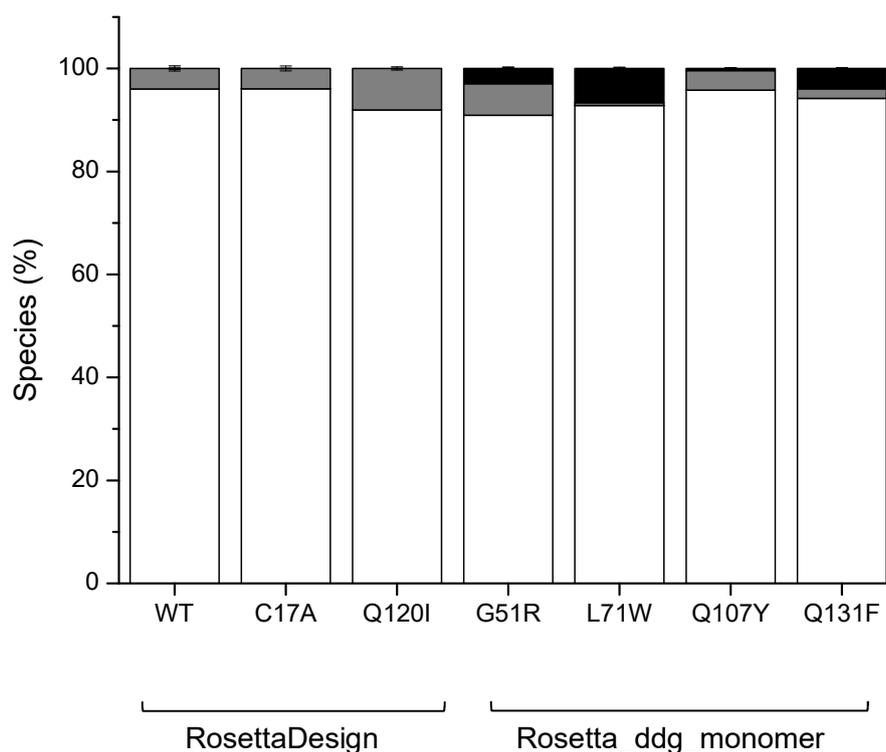


Figure 49. **Initial sample content of WT GCSF and mutant samples by SEC-HPLC.** Samples were formulated at 0.2 mg/mL in 10 mM sodium acetate pH 4.25, frozen at -70 °C followed by defrosting for 20 mins prior to analysis. SEC-HPLC was performed using a TSK3000 SEC column with 0.1 M phosphate buffer pH 2.5 over 40 mins. Different species were aggregate (black), dimer (grey) and monomer (white).

5.1.3.2 *Monomer loss with thermal degradation*

During incubation at 37 °C there was no uniform degradation for samples. As such, only the 45 °C data was included in results analysis. An example of the degradation progression of WT at 45 °C over time is displayed in Figure 50. By the final day (day 7) the degraded sample contained 4 main peaks and in order of retention time (rt) they were identified as aggregate, dimer, monomer and sodium acetate buffer. Across all samples, the aggregate peak had an average rt of 9.72 min \pm 2.77. Monomer had an average rt of 19.88 min \pm 0.07, and the dimer, which sat as a leading shoulder to the monomeric peak, had a variable average rt of 16.33 \pm 5.4. The sodium acetate buffer peak had an average rt of 24.16 min \pm 0.01 and an average area of 42.99 mAU*min \pm 0.87. The consistent area for the formulation peak in samples confirmed loss in monomeric protein was an effect of degradation and not sample loss.

The monomer loss as a % of original starting amount on day 0 for samples is shown in Figure 51A. An exponential trend was fit to the data using OriginPro 8.6 and the resulting values and statistics are displayed below Figure 51A in Table 16. The exponential fit had a high adjusted R² value for each data set, indicating a good fit. The initial rate of degradation displayed in Figure 51B and was calculated from the initial value (A) and exponential decay (R0) values from Table 16. The mutants Q120I and Q107Y were the only two studied with lower degradation rates than WT GCSF. Mutant Q120I had the highest amount of monomer left over after 7 days, with 79% of the original amount remaining. Mutant Q120I also had the lowest initial rate of degradation. Mutant Q107Y had the second highest level of monomer retained after 7 days with 69% remaining, and the second lowest initial rate. The mutant C17A had a similar initial rate of monomer loss to WT, however, the similarity diverged after days 6 and 7 and C17A ended with a lower monomer %. This suggests the mutant is the most alike to WT in terms of thermal degradation. Mutants L71W and Q131F had similar initial rates of degradation to WT, however, by day 7 they had significantly lower amounts of monomer suggesting they are destabilising mutations. Mutant G51R had a significantly higher initial degradation rate than WT and the lowest amount of monomer by day 7, at 37%, indicating it was the most destabilising mutant studied.

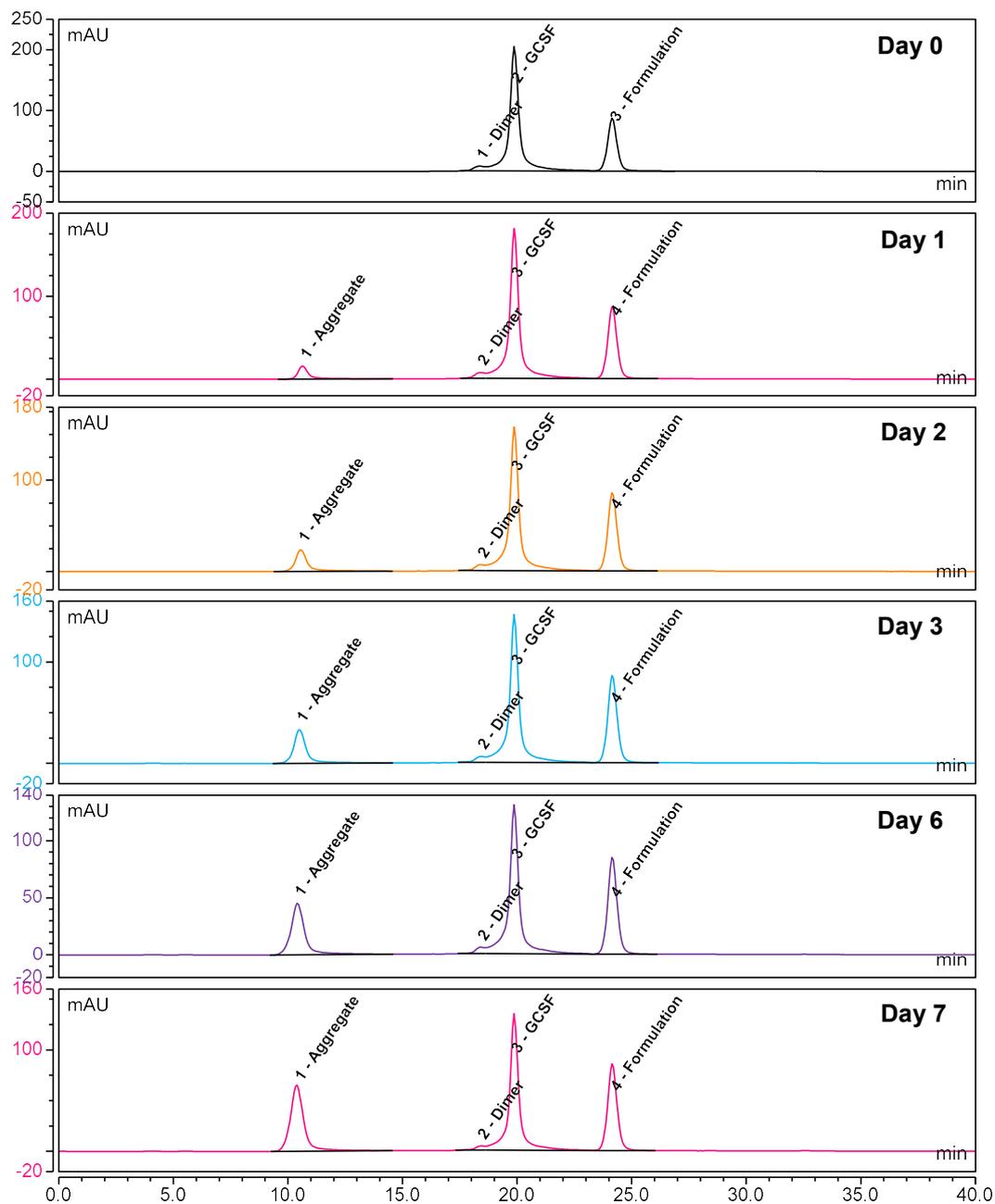


Figure 50. SEC-HPLC chromatograms of WT GCSF stored at 45 °C over time. WT GCSF was formulated at 0.2 mg/mL in 10 mM sodium acetate pH 4.25 and stored at 45 °C. Samples were taken over a 7 day period, frozen at -70 °C followed by defrosting for 20 mins prior to analysis. SEC-HPLC was performed using a TSK3000 SEC column with 0.1 M phosphate buffer pH 2.5 over 40 mins. Different species peaks were identified by their different retention times.

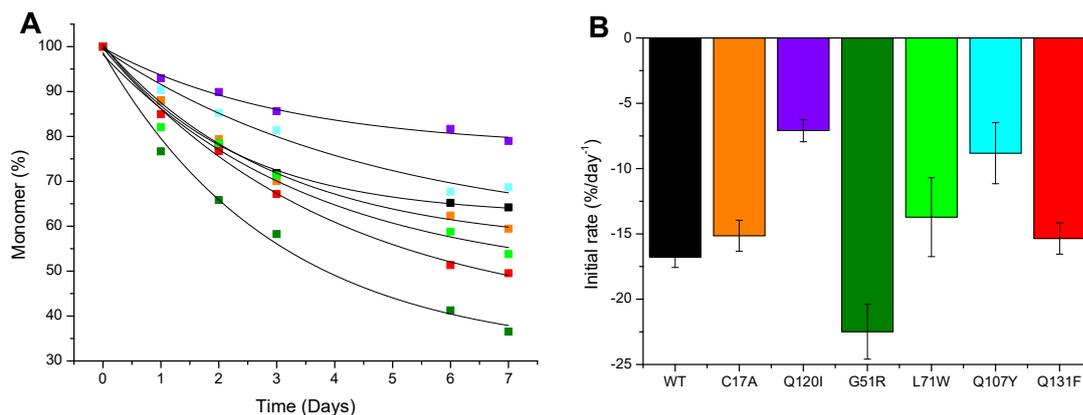


Figure 51. **Monomer loss for GCSF and mutants at 45 °C incubation.** A) Monomer loss as % of original starting amount. Exponential fit applied to data using the equation $Y = y_0 A \exp(R_0 * x)$. B) Initial rate of monomer loss calculated using R_0 and A values from Table 16

Table 16. **Curve fitting values for Figure 51A GCSF and mutant monomer loss at 45 °C.**

	y_0		A		R_0		Statistics	
	Value	SE	Value	SE	Value	SE	Reduced Chi-Sqr	Adj. R-Square
WT	62.289	0.972	38.124	1.107	-0.440	0.033	0.712	0.996
C17A	55.596	2.245	44.766	2.183	-0.338	0.044	1.832	0.993
Q120I	77.344	1.759	22.400	1.667	-0.317	0.062	0.912	0.985
G51R	31.410	4.060	67.264	3.928	-0.334	0.052	5.773	0.990
L71W	47.714	7.295	50.513	6.671	-0.272	0.090	9.261	0.967
Q107Y	58.106	6.535	41.421	5.952	-0.213	0.067	2.999	0.981
Q131F	38.459	3.065	61.219	2.78441	-0.251	0.027	1.224	0.997

5.1.3.3 Aggregation rates with thermal denaturation

The percentage of soluble aggregate present in each sample relative to the initial sample concentration (day 0) of monomer for WT and mutants is shown in Figure 52A. The percentage of aggregate present in the initial sample was also taken into account. Exponential trends were fit to the data using OriginPro 8.6 and the resulting values and statistics are displayed in Table 17, where an exponential fit was appropriate for all samples apart from WT. All exponential fits were found to have high adjusted R^2 values apart from G51R and Q107Y, however, their R^2 values still suggested a reasonably good fit. As the exponential fit did not succeed for WT, a linear fit was applied and the slope of the line

taken as the initial rate (Table 18). The initial rate of aggregation was calculated from the A and R_0 values for all mutants (Table 17). and from the slope of the line for WT (Table 18) is displayed in Figure 52B. WT had the lowest initial aggregation rate of all samples, however, it increased linearly and by day 7 the amount of aggregate was the second largest behind G51R. Mutants C17A and Q131F had joint second highest initial rates of aggregation but reached relatively low final levels of aggregate % after 7 days. Mutants L71W and Q107Y also had similar initial aggregation rates and final levels of aggregate at day 7 which was lower than WT, C17A and Q131F. Mutant Q120I had the lowest initial rate of aggregation and the lowest final level of aggregate, indicating it was the most aggregation resistant.

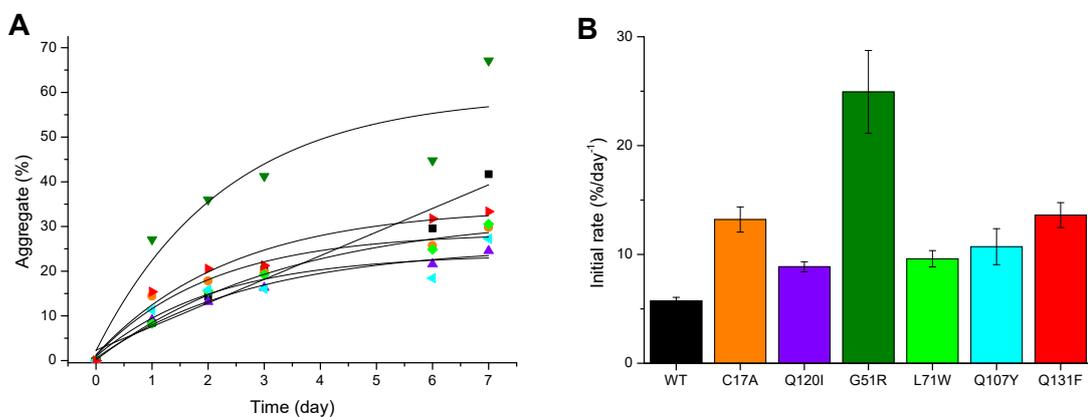


Figure 52. **Aggregation for GCSF and mutants at 45 °C incubation.** A) Aggregation as % of original starting monomer content minus aggregate at time 0. Exponential fit applied to data using the equations $Y = y_0 A \exp(R_0 * x)$ and $y = ab + c$. B) Initial rate of aggregation calculated using R_0 and A values from

Table 17. Curve fitting for Figure 52A GCSF and mutant aggregation rates at 45 °C.

	y₀		A		R₀		Statistics	
	Value	SE	Value	SE	Value	SE	Reduced Chi-Sqr	Adj. R-Square
WT	-10426	2544470	10428.46	254446	0.00	0.12	12.82	0.94
C17A	28.61	2.77	-27.50	3.35	-0.48	0.15	7.03	0.93
Q120I	25.61	1.79	-25.10	1.78	-0.35	0.07	1.33	0.98
G51R	59.52	10.71	-57.31	12.10	-0.43	0.24	83.98	0.83
L71W	32.81	3.81	-32.65	3.53	-0.29	0.08	3.33	0.97
Q107Y	23.84	4.13	-22.92	4.90	-0.47	0.26	14.72	0.82
Q131F	34.26	3.49	-32.96	3.81	-0.41	0.13	7.86	0.95

Table 18. Line fitting for Figure 52A WT GCSF aggregation rate at 45 °C

	Intercept		Slope		Statistics
	Value	SE	Value	SE	Adj. R-Square
WT	0	--	5.733	0.321	0.981

5.1.3.4 Dimerisation with thermal degradation

The percentage of dimer present in each sample over time relative to the initial sample concentration (day 0) of monomer for samples is shown in Figure 53A. Dimer present in the initial sample was also taken into account. An exponential fit did not succeed for the majority of mutants or WT and as such linear trends were fit to the data using OriginPro 8.6 and the resulting values and statistics are displayed in Table 19. The linear fitting was found to have low R² value for most mutants suggesting a poor fit. The slope of the line was taken as the initial rate, displayed in Figure 53B, where values are negative as dimer decreased over time.

From Figure 53B, mutant L71W had the lowest rate of loss of dimer, which can be accounted for the fact the mutant also had the lowest level of dimer present in the initial

sample. In contrast, mutant Q120I had the highest level of dimer in the initial sample, and this level remained constant over time, as seen by the almost horizontal linear trend line. Mutant G51R had the second highest level of dimer in the initial sample, however unlike mutant Q120I, G51R dimer was lost rapidly during accelerated thermal denaturation as seen by the highest rate of dimer loss in Figure 53B. Mutants C17A and Q107Y had similar initial rates of dimer loss to WT, however, mutant Q107Y also had similar levels of dimer to WT over time, whereas C17A had a lower level of dimer throughout the 7 days. Mutant Q131F was the only mutant to not fit a linear trend line as seen by the large standard error (SE) for the linear slope, as well as the low adjusted R² value (Table 19). This is due to the sudden increase in dimer at days 6 and 7 as seen in Figure 53A.

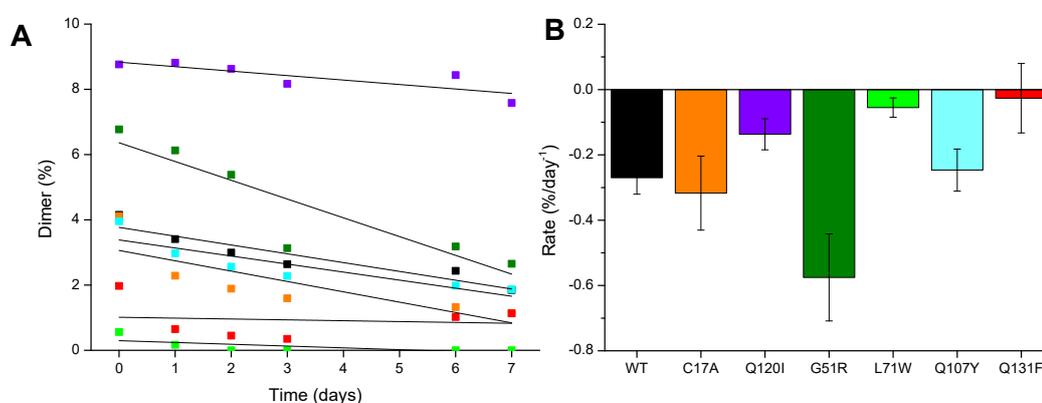


Figure 53. **Dimerisation for GCSF and mutants at 45 °C incubation.** A) Dimer as % of original starting monomer content minus aggregate at time 0. Exponential fit applied to data using the equations $Y = y_0 A \exp(R_0 * x)$ and $y = ab + c$. B) Initial rate of dimerisation calculated using R_0 and A values from Table 19.

Table 19. **Line fitting for Figure 53A GCSF and mutant dimerisation rates at 45 °C incubation.**

	Intercept		Slope		Statistics
	Value	Standard Error	Value	Standard Error	Adj. R-Square
WT	3.768	0.205	-0.270	0.050	0.847
C17A	3.062	0.460	-0.317	0.113	0.578
Q120I	8.832	0.194	-0.137	0.048	0.591
G51R	6.363	0.541	-0.576	0.133	0.779
L71W	0.296	0.120	-0.055	0.030	0.330
Q107Y	3.384	0.262	-0.246	0.064	0.732
Q131F	1.013	0.432	-0.026	0.106	-0.231

5.1.3.5 (NR) SDS-PAGE of day 7 thermal degradation samples

To visualise the degradants of WT and mutant samples, the day 7 samples were also analysed by non-reduced SDS-PAGE (Figure 54). For all samples there was an increase in MW species creating a ladder effect from monomer at 18.9 kDa, to dimer at 36 kDa, to trimer at 62 kDa and so on. The mutant Q120I had a fainter degradant ladder compared to the other mutant and WT samples, and had no clear dimer or trimer bands confirming its status as the most thermally stable mutant.

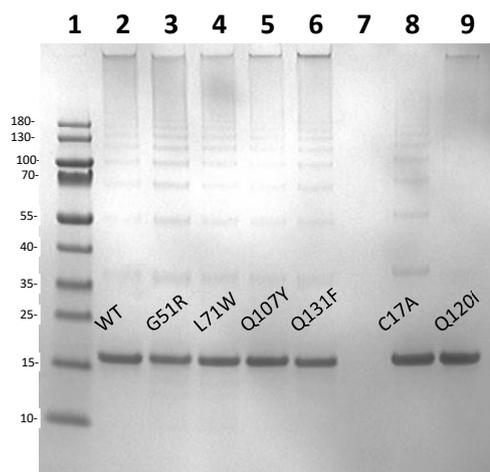


Figure 54. (NR) SDS-PAGE of thermally degraded GCSF and mutant samples after incubation for 7 days at 45 °C. Samples were prepared at 0.1 mg/mL in LDS and loaded into each well of a 4-12 % Bis Tris gel with MES running buffer. Electrophoresis was performed for 30 mins at 200 V followed by InstantBlue coomassie staining. Sample identities are labelled on the gel, and the PageRuler™ MWM ladder was loaded in lane 1.

5.1.4 GCSF Mutant HDX-MS

To explore changes in local protein flexibility following point mutation, low pH peptide level HDX-MS was performed. Mutant and WT samples were incubated in 10 mM sodium acetate, 99.9% D₂O, pH 3.82 (pD = 4.25) for three time points: 2.6 min, 4 hrs and 8 hrs and peptide uptake calculated relative to an undeuterated sample. The early time point of 2.6 min was included to monitor any fast exchanging regions, and the longer time points of 4 hrs and 8 hrs were selected due to the low pH solution reducing the intrinsic exchange rate, which increased the time required to observe differences between slow exchanging regions. Due to the lack of sample space within the HDX-MS LEAP PAL sample holder (max = 53), a total of three HDX-MS runs were required to analyse all mutants. WT was included in each run, as high variability in labelling between runs has been previously reported (Moroco and Engen 2015) and could affect relative comparisons.

5.1.4.1 *Mutant peptide maps*

Peptide maps were generated for WT and all mutants as point mutation can change the peptides generated during digestion. Five peptide maps were generated for each sample and the final peptide map was formed of peptides present in a minimum of three out of the five peptide maps. The final peptide map coverage for each mutant and WT is summarised in Table 20.

Table 20. Peptide map summary for WT and mutant samples

Sample	No. of peptides	Redundancy	Coverage (%)
WT	105	6.96	97.7
C17A	94	6.06	98.9
G51R	95	6.42	97.7
L71W	103	7.02	97.7
Q107Y	119	8.11	97.7
Q120I	90	5.74	97.7
Q131F	106	6.47	97.7

All peptide maps had >95% coverage, which is optimal for HDX-MS experiments. Interestingly, all samples apart from mutant C17A had 97.7% coverage, and the same residues were missed each time: 162-166 (LEVSY), along with two other single amino acids at variable positions. The 162-166 region was covered in the C17A peptide map and the region 49-51 (VLL) was missed.

5.1.4.2 *Mutant analysis in DynamX*

MS files from the same run were imported into DynamX 3.0 and using the WT peptide map from run 1, peptides were identified and the stacked spectral plots analysed together. As certain peptide amino acid sequences were different between WT and mutants, this resulted in a different number, size and retention time of peptides around the site of mutation. As a result, mutant peptides around the site of mutation required manual addition by inputting the sequence, residue start number and rt data generated from the mutant peptide maps.

5.1.4.3 *Differential exchange*

The DynamX 3.0 analysis software could not calculate the differential between WT and mutant peptides containing the site of mutation due to differences in amino acid letter sequence, resulting in large gaps in the DynamX differential plots and heat maps. As a result, raw uptake data was exported and the average relative uptake differential of peptides with the same residue number range calculated.

The overall differential value for each peptide was calculated by the sum of the average relative uptake differential for each time point. Mutants G51R and Q120I were included in run 1 and the differentials with WT displayed in Figure 55. The mutant C17A was included in run 2 and the differential with WT displayed in Figure 56. The mutants L71W, Q107Y and Q131F were included in run 3 and the differentials with WT displayed in Figure 57. The y-axis range of all differential figures was set to the highest mutant range to compare relative differential values across all mutants. Mutant G51R showed the highest level of positive differential values for a large number of peptides compared to the other mutants (Figure 55). A positive differential was indicative of increased uptake from increased flexibility. Two regions, loopAB and α D, had particularly large differential values for G51R. In contrast, mutant Q120I showed an overall mid-level decrease in the differential uptake for most peptides, suggesting a global stabilisation of the protein. Peptides including the mutation showed no difference in exchange as seen by a value around 0, although peptide number 55 (residues 114-120) next to the site of mutation had an increase in exchange (positive differential). There was also an increase in uptake in peptides around the N-terminal of the protein including residues 5 to 15.

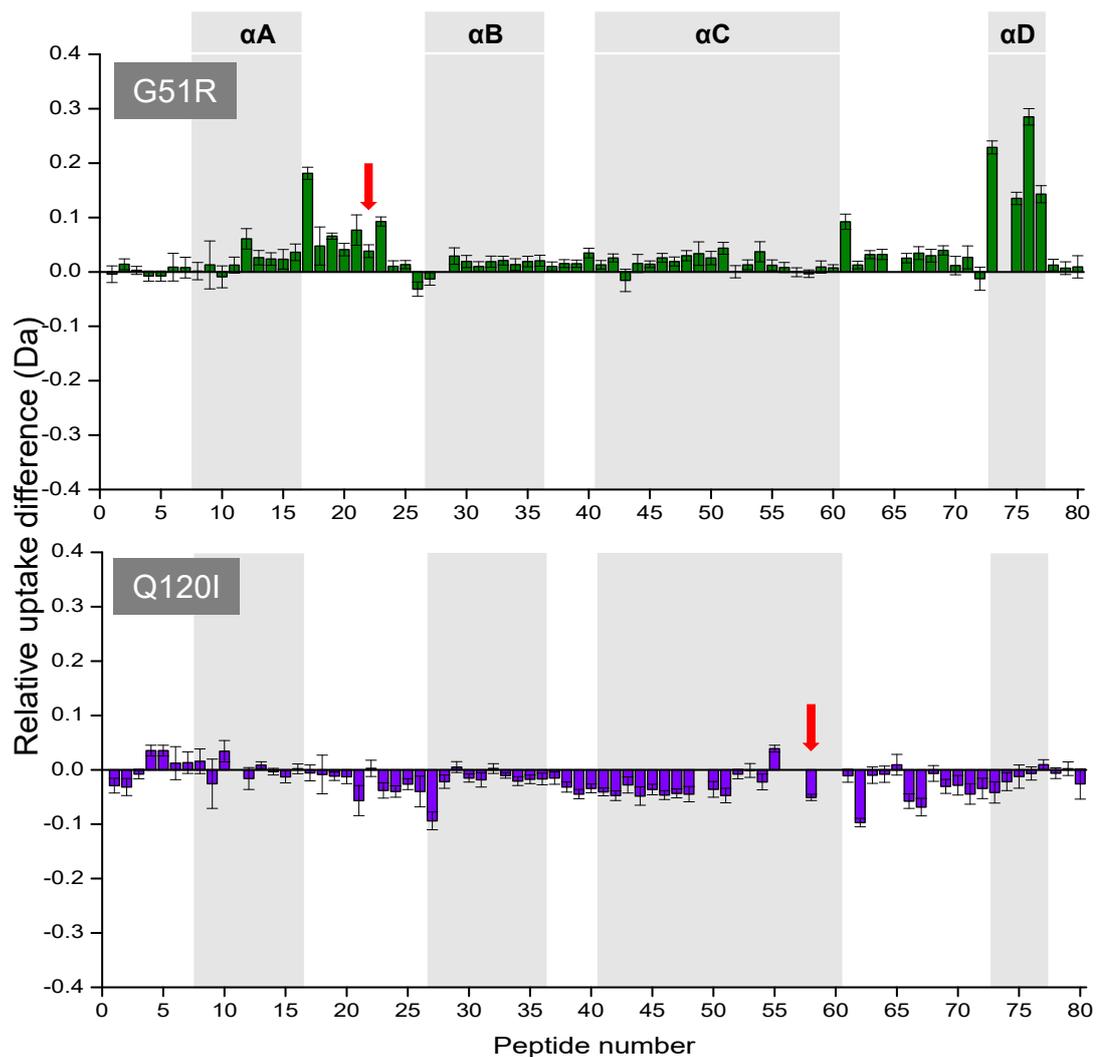


Figure 55. **G51R and Q120I peptide differential uptake plots.** Individual HDX-MS time point relative uptake values were added together and differential calculated (WT uptake minus mutant uptake). The red arrow represents the region containing the site of mutation. The y -axis denotes the relative uptake calculated from the change in mass between the undeuterated and deuterated values. The x -axis denotes the ordinal peptide number, a sequential arrangement of 80 peptides of GCSF by the midpoints in the sequence. The different helical regions of GCSF are coloured in the background. The non-coloured regions represent the connecting loop regions. Appendix 10 contains the GCSF peptide sequence, residue numbers and locations.

For mutant C17A, peptides around the site of mutation showed a low level increase in uptake, as seen by the positive differential values for peptides in Figure 56. A second region of increased uptake also occurred around the LoopCD region but in general the level of differential across C17A was low compared to mutants G51R and Q120I in Figure 55, suggesting low level destabilisation of the structure.

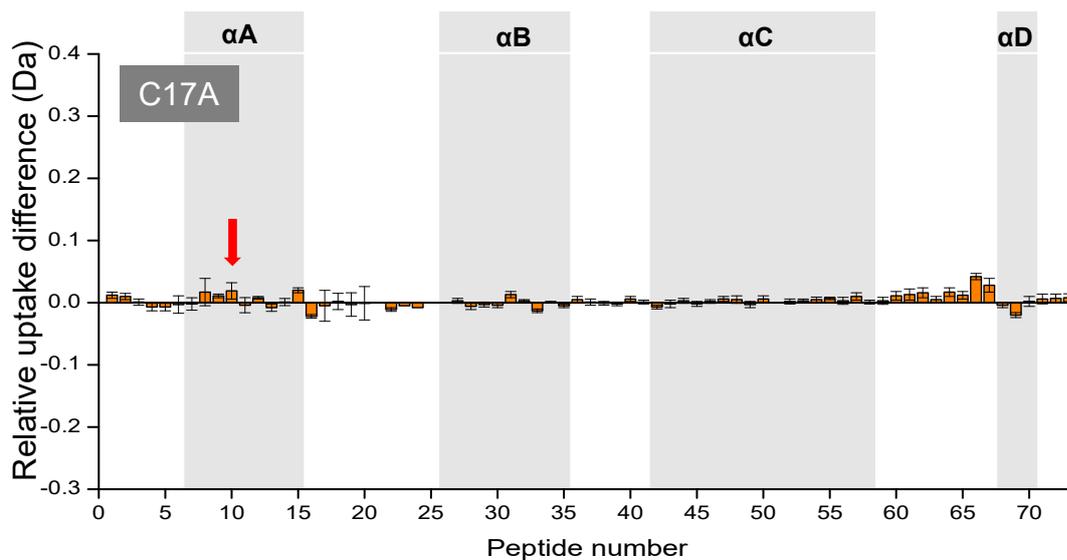


Figure 56. **C17A peptide differential uptake plot.** Individual HDX-MS time point relative uptake values were added together and differential calculated (WT uptake minus mutant uptake). The red arrow represents the region containing the site of mutation. The x-axis denotes the ordinal peptide number, a sequential arrangement of 80 peptides of GCSF by the midpoints in the sequence. The different helical regions of GCSF are coloured in the background. The non-coloured regions represent the connecting loop regions. Appendix 11 contains the GCSF peptide sequence, residue numbers and locations

Mutant L71W showed a mix of positive and negative peptide differential values for peptides suggesting increased flexibility of the protein in certain regions and increased rigidity in others (Figure 57). The C-terminal loopD (residues 166-175) was particularly affected by increased flexibility as seen by the peptides in this area having a relatively large positive differential values. On the other hand, peptides including the site of mutation, located in loopAB, showed a decrease in deuterium uptake, as well as peptides from the N-terminal and loopA region, indicating increased rigidity of these sites.

The majority of mutant Q107Y peptides had negative differential values, especially around the site of mutation (Figure 57). The peptides next to the site of mutation had a unique pattern of alternating low and baseline level differentials for a span of 12 peptides. This indicated specific regions within this area were protected from exchange due to the mutation. A large number of peptides covering the site of mutation showed a low negative differential suggesting decreased flexibility around the region.

The mutant Q131F showed low level, negative differential values for the majority of peptides (Figure 57). A number of peptides were discounted from the analysis around the site of mutation, hence the lack of values in that area. This was due to extremely large variations in uptake between mutant and WT peptides. Peptides flanking the site of mutation showed low level, but significant rigidification. This was prominent around the α D

helix, where the differential value was lowest, suggesting this region was particularly stabilised by the mutation.

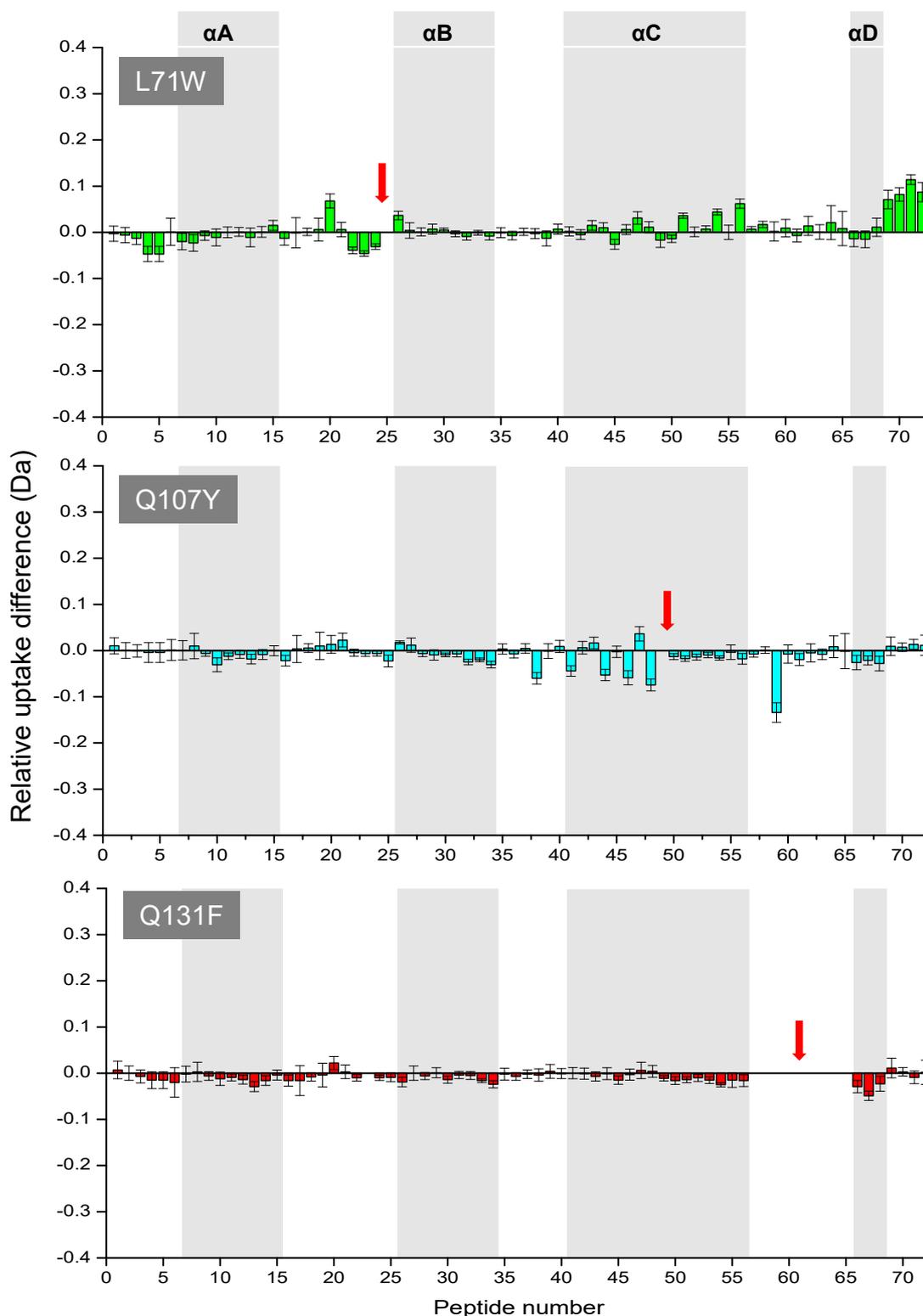


Figure 57. **L71W, Q107Y and Q131F peptide differential uptake plots.** Individual HDX-MS time point relative uptake values were added together and differential calculated (WT uptake minus mutant uptake). The red arrow represents the region containing the site of mutation. The x-axis denotes the ordinal peptide number, a sequential arrangement of 80 peptides of GCSF by the midpoints in the sequence. The different helical regions of GCSF are coloured in the background. The non-coloured regions represent the connecting loop regions. Appendix 12 contains the GCSF peptide sequence, residue numbers and locations

5.1.5 Predicting stability by protein flexibility

During HDX-MS differential uptake data analysis, a trend between monomer loss rate during accelerated thermal degradation experiments and a majority of peptides with positive or negative differentials was observed. The average HDX-MS peptide differential was calculated and plotted against the initial rate of monomer loss for each mutant. From Figure 58A, there was a moderate positive correlation between the two measurements, offset by the mutant Q131F. Mutant Q131F had a high initial monomer loss rate but an overall negative differential, which caused it to outlie from the data set. Of all mutants studied, Q131F had a significant loss of coverage around the site of mutation during HDX-MS analysis. This region most likely would have contained differences in exchange compared to WT, suggesting HDX-MS data without this region is not a valid representation of Q131F flexibility. Consequently, removal of mutant Q131F, as seen in Figure 58B, changed the correlation from a moderate to very strong positive correlation, as seen by the change of R^2 value from 0.60 to 0.96 (Table 21).

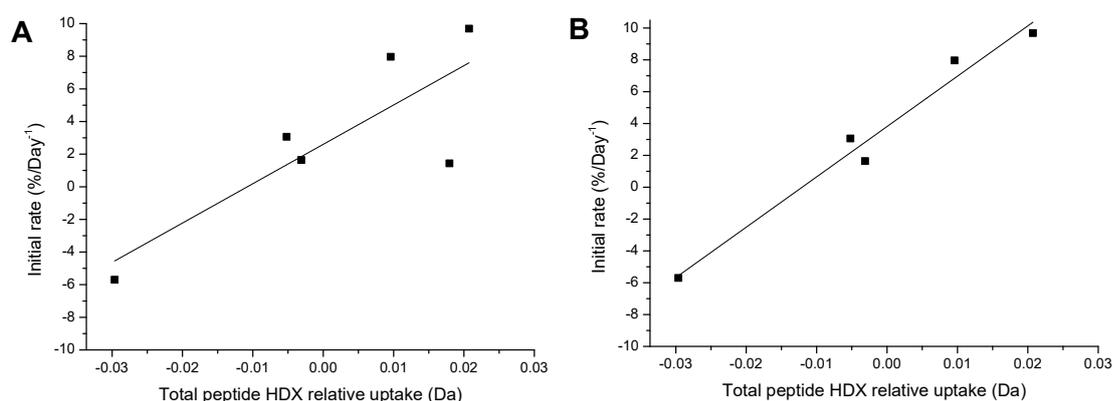


Figure 58. **GCSF mutant average peptide relative differential uptake versus initial monomer loss rate.** A) including mutant Q131F, and B) removing mutant Q131F.

Table 21. **Line fitting for GCSF mutant differential uptake versus initial monomer loss rate** A) Figure 12A (including mutant Q131F) and B) Figure 12B (removing mutant Q131F).

	Intercept		Slope		Statistics
	Value	SE	Value	SE	Adj. R-Square
A	2.600	1.424	241.146	83.225	0.597
B	3.815	0.514	316.256	30.354	0.964

Discussion

5.1.6 Biophysical methods to compare WT and mutants

The aim of this chapter was to apply advanced biophysical analysis to GCSF stability mutants produced in chapter 4 in order to further characterise the mutation effects on the flexibility and stability.

The characterisation of mutation on GCSF higher order structure was previously attempted by Bristow et al. (2012). The group produced helix destabilising single, double and triple point mutations (A→G) and found the GNFS-60 bioassay was more effective at distinguishing between mutants and WT than circular dichroism, which couldn't distinguish any of the mutants. Further study on the three single and a double mutants, was performed by Aubin et al. (2015), who compared their effects on higher order structure by NMR. They found all four mutants produced distinctive 2D spectra including the disappearance of the alanine amide signal and appearance of a new amide signal for the glycine. The mutations also produced a number of chemical shift perturbations of the amide resonances of surrounding residues. This change in local magnetic environment at the GCSF mutation site demonstrated the feasibility of NMR to monitor effects of mutations on higher order structure; however, the study also found sample conditions for GCSF NMR were not stable for triple mutants. The NMR samples contained 0.017 to 0.4 mM GCSF protein equating to 0.32 to 7.5 mg/mL. As discussed in chapter 3, high concentration decreases the T_m and T_{agg} of GCSF. Moreover, as discussed at the end of chapter 4, mutants and WT GCSF were produced in shake flask cultures and the overall yield was in the range of 20 mg per sample. As such, using NMR to study GCSF mutants produced in chapter 4 was unfeasible as the technique would rapidly consume material, and potentially stress the samples such that reliable causes of mutation on higher order structure could not be established.

Previous studies to compare WT and mutant structures by HDX-MS have also been successful in elucidating effects on stability and activity. The mutation of the residue Y407 in the CH3 domain of an IgG1 mAb was shown by Rose et al. (2012) by HDX-MS to produce a more exposed interface around the site of mutation, which was linked to increase sialylation, galactosylation, and branching of the N-linked glycans in the CH2 domain. In addition, allosteric structural effects in the CH2 domain and in the CH2–CH3 interface were identified, providing a possible explanation for dramatic changes in glycosylation (Rose et al. 2012). The approach was also used by Brier et al. (2006) to identify the binding region of

inhibitors targeting the human mitotic kinesin Eg5. The group saw decreased uptake in two regions interacting with inhibitors, which upon mutation resumed the normal rate of exchange of the control sample without inhibitors. Consequently, HDX-MS was able to help identify regions of proteins involved in ligand binding.

As regions of GCSF flexibility were initially determined by HDX-MS in chapter 3, and the technique is non-intrusive and requires low sample concentrations, the next logical step was to repeat HDX-MS analysis with GCSF mutants and compare differences in uptake with WT to determine differences in local flexibility.

5.1.7 Mutant identity

During this chapter it was established by intact mass spectrometric experiments that the mutants studied matched the theoretical mass change, apart from C17A which had a lower experimental mass. This suggests that C17A may have contained another unintentional point mutation, which lowered the molecular mass. Further information on mutant identity also arose during HDX-MS where peptide maps of mutants and WT were generated. The PLGS software searches for peptides from a theoretical sequence input and all mutant peptide maps covered at least 97.7% of the sequence. The mutant C17A peptide map covered a higher percentage than this at 98.9% and the missing region was identified as the amino acid residues 49-51 (valine leucine and leucine) suggesting this could be the region containing another point mutation. Further tests to determine the identity of C17A should be performed before future work is completed and the biophysical analysis performed in this chapter should be viewed with the unconfirmed identity of C17A in mind.

5.1.8 Purity by SEC-HPLC

In the previous GCSF mutant chapter, samples were analysed by SDS-PAGE and shown to be pure by the presence of a single band. From SEC-HPLC of initial sample content, samples were also confirmed to be predominantly monomeric species. Whilst there was varying levels of dimer and aggregate per mutant, which will be discussed in more detail later, the variability should not have affected HDX-MS analysis as measurements are an average of the sample species rather than specific molecules. Although, it should be noted that the presence of different MW species at different ratios may cause discrepancies during thermal denaturation analysis, as the species may provide a “seed” for aggregation (discussed in further detail in the next section).

5.1.9 Accelerated thermal degradation

Accelerated degradation methods are commonly applied by formulation scientists to predict the shelf-life of drugs in a more time efficient manner. Accelerated thermal denaturation is generally applied to simulate accidental exposure of biopharmaceuticals to non-recommended conditions during production, storage, handling and administration (Tamizi and Jouyban 2016). However, the resistance to thermal degradation can also be indicative of enhanced shelf-life stability as thermostable proteins were correlated by Maddux et al. (2014) to have longer shelf-lives.

The most typical experimental parameter used to accelerate aggregation rates is temperature (Weiss et al. 2009). According to Chan (2016) a starting point for thermal degradation incubation temperatures is usually at least 10 °C below the T_m . The thermal degradation temperatures selected for the GCSF mutant comparisons was 37 and 45 °C and the amount of monomer, dimer and aggregate of samples when held at these elevated temperatures was measured using SEC-HPLC. At 37 °C there was no uniform degradation of WT or mutant samples over time. A study by Herman et al. (1996) studied GCSF stored at 29 °C and found the amount of aggregate only exceeded specification after 6 to 8 months. As 37 °C is greater than 10 °C below the WT T_m identified in chapter 3, this explains why one week monitoring of samples was not long enough to see degradation at 37 °C. At 45 °C the incubation storage temperature was 7.36 °C below WT T_m and exponential loss of monomer and linear gains of aggregate were observed over 7 days.

From the characterisation data obtained in chapter 3, mutant Q131F was expected to perform poorly during accelerated thermal degradation. However, G51R, L71W and C17A also displayed a higher rate of monomer loss and increased aggregation relative to WT. The mutant G51R, surpassed Q131F as the mutant with the highest monomer loss and aggregation rates, whereas L71W and C17A had degradation rates only slightly lower than WT. Mutants Q107Y and Q120I were the only two mutants with lower degradation rates relative to WT confirming their contribution to increased thermal stability. The T_m and T_{agg} values identified in chapter 3 gave little indication of the performance of the mutants during accelerated thermal degradation. By T_m , mutants C17A, and L71W were more thermostable than WT and by T_{agg} Q120I, G51R and Q107Y were more aggregation resistant than WT.

The lack of correlation between the accelerated thermal denaturation results in this chapter and the T_m/T_{agg} values from the previous chapter may have been caused by the variation in

MW species within the samples prior to accelerated thermal denaturation. This may have affected degradation kinetics via seeding. A seed is a soluble aggregate previously formed during processing or storage of the samples which can cause changes to degradation kinetics as upon introduction to stress (temperature, pH, etc.; Weiss et al. 2009). Mutants G51R, L71W and Q131F had significantly higher levels of initial aggregate in the samples compared to WT, which may have enhanced their aggregation rates. Data suggests the initial aggregate content was a further indication of mutant stability following storage at 4 °C; however SEC-HPLC data prior to storage would be required to determine if this is true.

A study by Thiagarajan et al. (2016) compared the molecular features of nine mAbs with their performance during accelerated and long-term storage by SEC-HPLC. They found colloidal stability, self-association propensity and conformational characteristics (exposed tryptophan) provided reasonable prediction of accelerated stability, with limited predictive value at 2–8 °C stability, and no correlations to stability behavior were observed with onset-of-melting temperatures or domain unfolding temperatures as measured by DSC. This finding combined with data from this chapter, suggests that snapshot T_m/T_{agg} values may not be entirely useful for prediction of GCSF mutant degradation kinetics during storage at 45 °C.

5.1.10 Mutant HDX-MS

Approaches to predict protein degradation based on local unfolding or conformational changes in otherwise folded monomers has been demonstrated by HDX coupled with NMR or mass spectrometry (Kendrick et al. 1997; Tobler and Fernandez 2002; Tobler et al. 2004). With the additional attractive feature of potentially identifying specific regions of a protein that can be targeted for protein engineering strategies to minimise the inherent aggregation propensity (Weiss et al. 2009). In this chapter, differences in local HDX of six GCSF mutants relative to WT were successfully determined to further understand the effect of specific point mutations on conformational stability.

5.1.10.1 C17A

The mutant C17A substituted the free cysteine residue for a small basic alanine residue. A lack of activity and stability difference between C17A and WT was shown previously by Raso et al. (2002) and was most likely due to the fact the cysteine side chain is partially buried (Herman et al. 1996; Buchanan et al. 2012). Furthermore, during RP-HPLC in the previous chapter, C17A eluted at the exact same time as WT confirming they have the

same surface hydrophobicity. From all of this information, C17A was expected to show minimal differences to WT by HDX-MS.

During HDX-MS analysis the mutant C17A showed a low level of increased uptake around the site of mutation and in loopCD, however, the level of increase was low relative to the other mutants studied. During accelerated degradation, C17A had a slightly higher rate of monomer loss to WT suggesting the mutation was destabilising. This combined with the HDX-MS data suggests C17A increased flexibility, which caused a decrease in thermal stability. A similar pattern was observed for all other mutants studied. Due to the unconfirmed identity of C17A, it is possible another point mutation occurred during production, which caused the reduction in stability.

5.1.10.2 *Q120I*

By HDX-MS the dynamics of Q120I was measured relative to WT where it was shown that overall the majority of peptides decreased in uptake, which correlated with the decrease in degradation rate during accelerated thermal degradation, suggesting the increase in stability is caused by a decrease in protein flexibility. Interestingly, residues adjacent to the site of mutagenesis increased in exchange and an allosteric decrease in uptake of residues at the N-terminal of GCSF also occurred. The N-terminal of GCSF has previously been shown to be highly mobile by NMR (Zink et al. 2004) and not necessary for activity (Kuga et al. 1989). By RP-HPLC in chapter 4, the change in Q120I retention time confirmed the surface exposure of the residue as well as the increased hydrophobicity of the molecule. From the data it is hypothesised the increase in hydrophobicity of the substitution decreased intermolecular associations at the site of mutation and adjoining helices causing an increase in stability, and the increased uptake of adjacent residues occurred because of displacement from the isoleucine residue.

5.1.10.3 *G51R*

Although, G51R increased GCSF T_{agg} in chapter 4, data from this chapter suggests, that G51R was actually destabilising as it had the highest amount of initial aggregate in the sample during SEC-HLC analysis and had the highest rate of aggregation during accelerated thermal degradation. By HDX-MS, G51R created an increase in exchange around the site of mutation (short helix), as well as in part of the loopABII, loopCD and α D, in-keeping with the observation that increased flexibility causes increased instability. As G51R had the largest MW change of all mutants studied at +98.7 Da, it is hypothesised that the

destabilisation was caused by the increased size of arginine causing disruption to the GCSF 3D structure and ultimately causing a new route for aggregation.

5.1.10.4 *L71W*

A discrepancy between increased T_{agg} and T_m , and accelerated thermal degradation was observed for L71W, as the mutant had a higher rate of aggregation than WT. As with G51R, L7W also had a high level of aggregate in the initial sample prior to accelerated thermal degradation. As leucine is a relatively large amino acid, L71W was the second largest increase in mass for the mutations studied, with +73 Da. L71 is located at the end of loopABII close to αB and as seen with G51R and Q120I, there was an increase in exchange around the site of substitution towards the αB terminal, presumably due to the increased size of the substituted tryptophan residue. There was also a decrease in the exchange along the length of loopABII. As L71W showed both increased and decreased flexibility by HDX-MS, results suggest the increased flexibility at the site of mutation had a larger influence on stability.

5.1.10.5 *Q107Y*

Mutant Q107Y had a lot of features in common with mutant Q120I: mutant Q107Y reduced the rate of degradation of GCSF during accelerated thermal degradation, increased T_{agg} and decreased T_m . In these cases, T_{agg} was a good predictor of accelerated thermal degradation results. Located in αC , Q107Y showed an increase and decrease in exchange at both sides of the mutation site as seen with L71W; however the mutant had an overall decrease in exchange for most peptides indicative of decreased flexibility.

5.1.10.6 *Q131F*

Mutant Q131F substituted the amino acid, glutamine with phenylalanine within the middle of loopCD. This substitution is not too dissimilar from Q120I, changing a flexible glutamine residue with a larger, hydrophobic residue. The only differences were residue region and MW, where Q120I decreased GCSF MW whereas Q131F increased it. During accelerated thermal degradation, Q131F had a high rate of monomer loss. This was expected as Q131F was shown in the previous chapter to have a lower T_m and T_{agg} . By HDX-MS, the mutant showed a general low level decrease in exchange for most peptides and a significant decrease in exchange in the αD region. On average the differential value was negative indicating an overall stabilisation of the structure. As the overall differential value was negative but the thermal stability data indicated the mutant was destabilising, this caused mutant Q131F to sit outside of the general trend found in this chapter that increased

thermal stability by mutation is caused by decreased flexibility. A lot of peptides around the site of mutation were lost during analysis suggesting perhaps an increase in flexibility could have occurred in this region (in keeping with other mutant data) but was not identified.

5.1.11 Flexibility as a predictor of protein degradation/ stability

When mutant Q131F was removed from the data set, a strong correlation between degradation rate relative to WT and the average HDX-MS differential value was observed. This suggested flexibility, as measured by differences in peptide HDX-MS is generally a strong predictor of thermal stability. Whilst the data set is small, the change in HDX-MS for stability mutants was obvious. In addition, results from this study go some way to confirm that targeting of flexible regions for rigidification is an effective strategy to improve thermal stability as both thermostable mutations identified in this chapter were located in regions identified as flexible by B-Factor and HDX-MS data.

6 Mapping GCSF-excipient interactions in the aqueous state

The aim of this chapter was to use both standard and advanced biophysical techniques to assess the use of computational protein-excipient docking methods to A) accurately predict the stabilising effect of different excipients on GCSF aggregation and B) characterise the interactions.

For stability predictions, excipients were experimentally assessed for their effect on GCSF thermal stability using high-throughput screening and ranked by their improvements to T_{agg} . This ranking allowed for a comparison to binding energies generated by the computational docking software iGEMDOCK between the same excipients and the GCSF 3D structure, where a correlation provided evidence that computational predictions could substitute experimental screening processes in formulation development.

For location of interactions, HDX-MS analysis of GCSF was performed with a select number of excipients included within the deuterium exchange solution. The change in uptake was compared to a non-excipient-containing control solution to identify regions of interaction. By comparing the predicted sites of interaction with the experimental HDX-MS data, validation of the computational predictions was performed and discussed in detail.

6.1 Results

In this chapter the excipients, L-arginine monohydrochloride, D-mannitol, D-sorbitol, L-phenylalanine, D-(+)-trehalose dihydrate and D-sucrose were studied for their interactions with GCSF. They are referred to by their name minus their relative configurations from hereon in.

6.1.1 Effect of different excipients on GCSF T_{agg}

Thermal degradation analysis was conducted with GCSF in the presence and absence of excipients to quickly and qualitatively assess the effect on T_{agg} . Excipients studied included two amino acids (arginine and phenylalanine), two saccharides (sucrose and trehalose) and two isomeric sugar alcohols (mannitol and sorbitol), at three different concentrations (Table 22). The mid-range concentrations were selected based on the typical range used in the formulation of biopharmaceuticals and the high and low range concentrations were selected around the mid-range value. Detergents, such as the polysorbate 80 (Tween 80), were not included in the study due to their difficulty to computationally model and the lack of interaction/ change in higher order structure GCSF in solution as measured by ^{15}N -NMR (Aubin et al. 2015)

Table 22. Range of excipient concentrations added to formulation screen samples. All samples contained GCSF at 0.3 mg/mL in 10mM sodium acetate pH 4.25

Sample	Concentration (mM)		
	Low	Mid	High
Control	0	0	0
Arginine	50	100	200
Mannitol	50	100	200
Phenylalanine	25	50	100
Sorbitol	50	100	200
Sucrose	50	100	200
Trehalose	50	100	200

Two different GCSF concentrations of 1 mg/mL and 0.5 mg/mL in 10 mM sodium acetate, pH 4.25 were also included in formulations to assess the sensitivity of SLS to measure

aggregation. The effect of increasing GCSF concentration on SLS is shown in Figure 59. The maximum scatter intensity was lower for the lower GCSF sample concentration of 0.5 mg/mL, which is caused by a decrease in the amount of GCSF molecules available for aggregation. The T_{agg} of the control GCSF sample also decreased from $50.4\text{ }^{\circ}\text{C} \pm 0.4$ to $49.1\text{ }^{\circ}\text{C} \pm 0.2$ with increasing protein concentration, as seen by a shift in the scatter curve to the left in Figure 59. This may have occurred because of increased scattering sensitivity at the higher protein concentration.

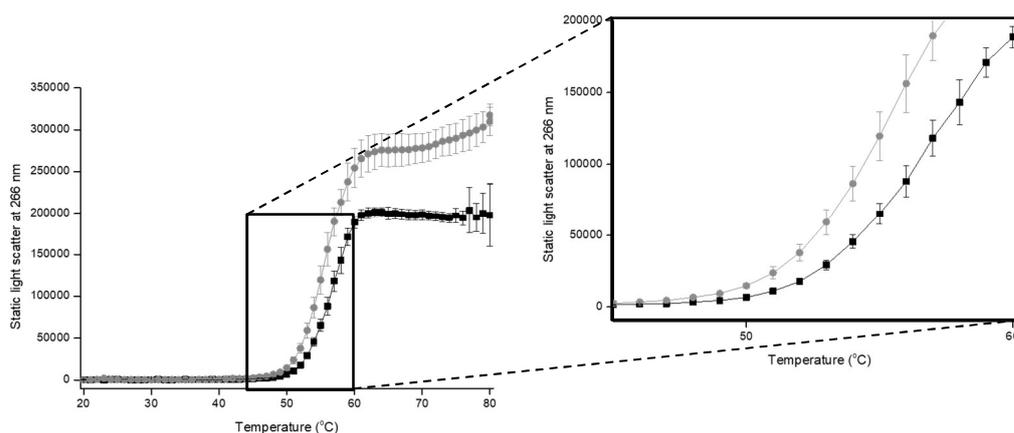


Figure 59. **Static light scattering with increasing GCSF concentration.** Static light scattering measured at 266 nm during a thermal ramp from 20 to 90°C at a rate of 1 °C/min. GCSF in 10 mM sodium acetate was analysed five times and averaged at two different concentrations: 0.5 mg/mL (black square), and 1 mg/mL (grey diamond).

Upon addition of the same concentration of excipients it was found that the difference in T_{agg} between the two concentrations of GCSF was consistent across all formulations (data not shown). As such, concentration was deemed to have no effect on excipient ranking and the high GCSF concentration excipient data (1 mg/mL) were omitted from subsequent results to avoid repetition. The SLS profiles for 0.5 mg/mL GCSF with varying types and concentrations of excipient compared to the control are shown in Figure 60.

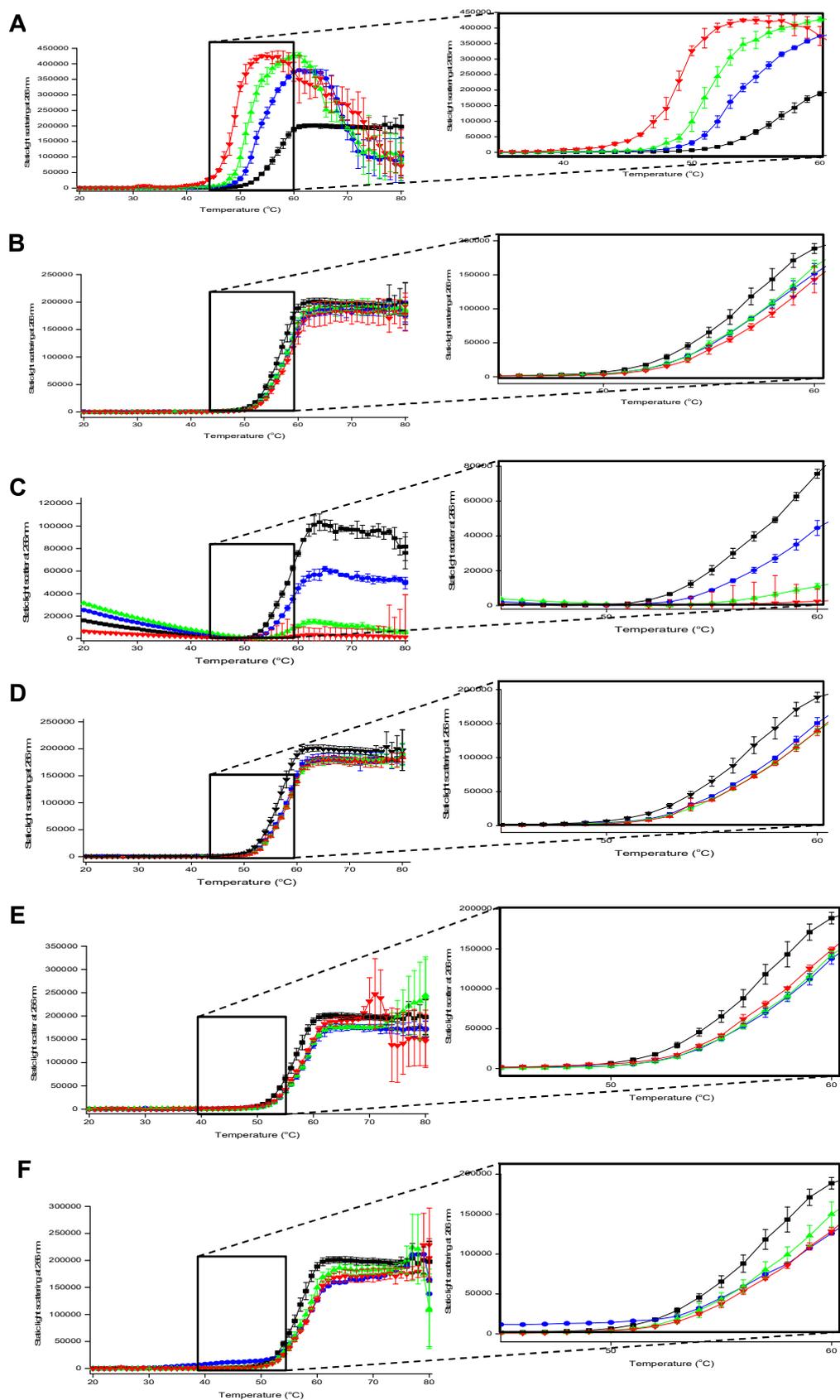


Figure 60. **Static light scattering with increasing excipient concentration.** Static light scattering measured at 266 nm during a thermal ramp from 20 to 90 °C at a rate of 1 °C/min. GCSF in 10 mM sodium acetate was analysed three times per formulation and averaged. Excipient concentrations included a control with none added (black square), low (blue circle), mid (green triangle), and high (red triangle). Excipients analysed included A) arginine, B) mannitol, C) phenylalanine, D) sorbitol, E) sucrose, and F) trehalose.

In general, the addition of an excipient increased the thermal stability of GCSF as seen by a shift in the scatter curve to the right, indicating a higher temperature was reached before aggregation commenced. This was true for all excipients except for arginine (Figure 60A) where a lower temperature caused earlier and more extensive GCSF aggregation as indicated by a curve shift to the left and scatter intensity increase with increasing concentration. For the sugars, sucrose (Figure 60E) and trehalose (Figure 60F), the increase in excipient concentration saw an initial improvement in aggregation onset temperature but increasing concentration had a negligible further improvement to aggregation onset or scatter intensity, whereas for mannitol (Figure 60B), phenylalanine (Figure 60C) and sorbitol (Figure 60D), an increase in concentration saw an incremental shift to the right with increasing concentration as well as a decrease in scatter intensity. This suggested the high excipient concentration was not yet optimal. The stabilising effect on aggregation with increasing concentration was especially clear with the phenylalanine scatter (Figure 60C) where there was a large shift to the right in the scatter curve and decreasing intensity with increasing concentration. At the high phenylalanine concentration, the increase in scatter due to aggregation was almost non-existent; however, the standard deviation was relatively large when compared to all other samples.

The highest T_{agg} value for each excipient is shown in Figure 61. Increased T_{agg} of GCSF was observed for all excipients studied apart from arginine, which was in concurrence with the SLS profile analysis. Phenylalanine at 100 mM provided the highest increase in T_{agg} to 53.33°C.

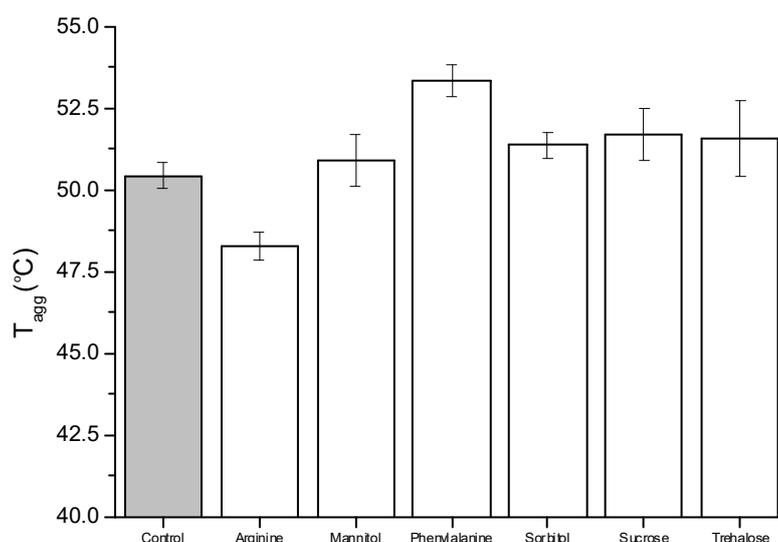
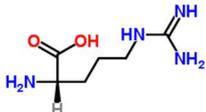
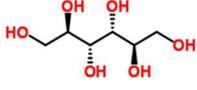
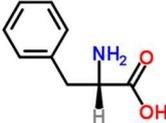
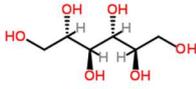
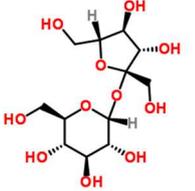
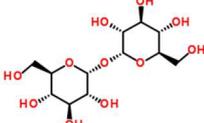


Figure 61. Highest T_{agg} value obtained for each excipient with 0.5 mg/mL GCSF in 10 mM sodium acetate pH 4.25. Error bars are the standard deviation of the mean.

6.1.2 Excipient docking

Docking between GCSF and the excipients was performed using iGEMDOCK. The structures for the six excipients, as well as their pKa, is summarised in Table 23.

Table 23. **Excipient summary.** pKa values obtained from Chemicalize.org (2016), structures obtained from RSC ChemSpider (2016)

Excipient	Type	pKA	Structure
L-arginine	Amino acid	2.35 (-CO ₂ H) 9.87 (-NH ₃ ⁺)	
D-mannitol	Polyol, Sugar alcohol	12.59	
L-phenylalanine	Amino acid	2.20 (-CO ₂ H) 9.31 (-NH ₃ ⁺)	
D-sorbitol	Polyol, Sugar alcohol	12.59	
D-sucrose	Polyol, Sugar	12.6	
D-(+)-Trehalose	Polyol, Sugar	11.91	

iGEMDOCK generated 10 “best docking” poses for each excipient and GCSF in the form of PDB files. The binding energies of the protein-excipient PDB files were visualised in the iGEMDOCK Docked Poses/ Post-Screening Analysis, exported and averaged (Figure 62). Where the lower the binding energy, the stronger the interaction between protein and excipient. The total binding energy was summation of all potential non-covalent bonds including H-Bond, VDW and electrostatic interactions. Of the excipients studied, none showed any electrostatic interactions; as such the total binding energy was a combination

of H-Bonds and VDWs only. Sucrose, followed closely by trehalose, had the lowest total binding energy and sorbitol was found to have the highest binding energy. By breaking down the total binding energy into its components, sucrose had the highest VDW value, whereas sorbitol had the lowest. Mannitol has the lowest H-Bond value, whereas sorbitol has the highest.

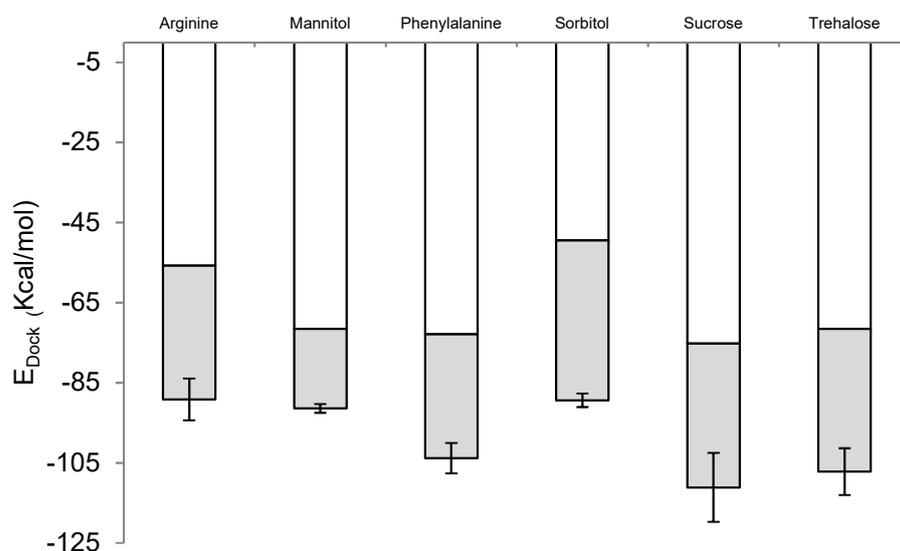


Figure 62. **Average total pose energy of excipients docked with GCSF PDB 2D9Q using iGEMDOCK.** Standard deviation calculated as the sum of the standard deviation of the Hbond (grey) and VDW values (white).

6.1.3 Correlating docking and T_{agg} values

Protein-excipient T_m values have previously been shown to correlate with iGEMDOCK total binding energy outputs, where the higher the T_m the lower the total binding energy (Barata et al. 2016). To determine if this was true for GCSF-excipients and T_{agg} , the highest T_{agg} value for each GCSF-excipient formulation was plotted against the average total binding energy (Figure 63). The correlation showed that iGEMDOCK predicted total binding energies described 76% of the experimental space for the lower concentration of GCSF (0.5 mg/mL). Note: arginine was removed from correlations due to its unique destabilising effects causing it to outlie from the rest of the excipient data.

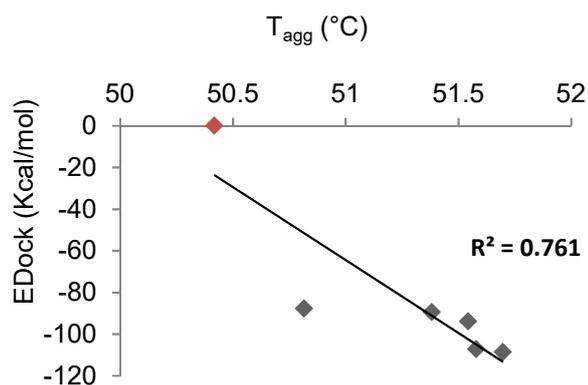


Figure 63. Correlation between docking results (E_{Dock}) and T_{agg} .

6.1.4 Residue interaction analysis

iGEMDOCK Post-Screening Analysis provided an in-depth analysis of the GCSF residues involved with the excipient interactions. Excipient interacting residues were divided into two interacting groups (main and side chain) and type of interaction (electrostatic interaction, H-Bond, or VDW). The binding energy for each GCSF residue and the different excipients is displayed for H-bond and VDW in Figure 64, where interactions were split into main (M) and side chain (SC) groups. From both H-bond and VDW binding energy figures, trehalose, sucrose and sorbitol interacted with more GCSF residues, and by H-bonds trehalose had the most negative average binding energy with the side chain of L69, whereas by VDW sorbitol had the most negative average binding energy with the main group of G70. Both H-Bond and VDW figures showed large standard deviation of the average binding energies suggesting significant comparisons between excipient binding energies to the same residue can't be made, however comparisons between regional differences can still be made.

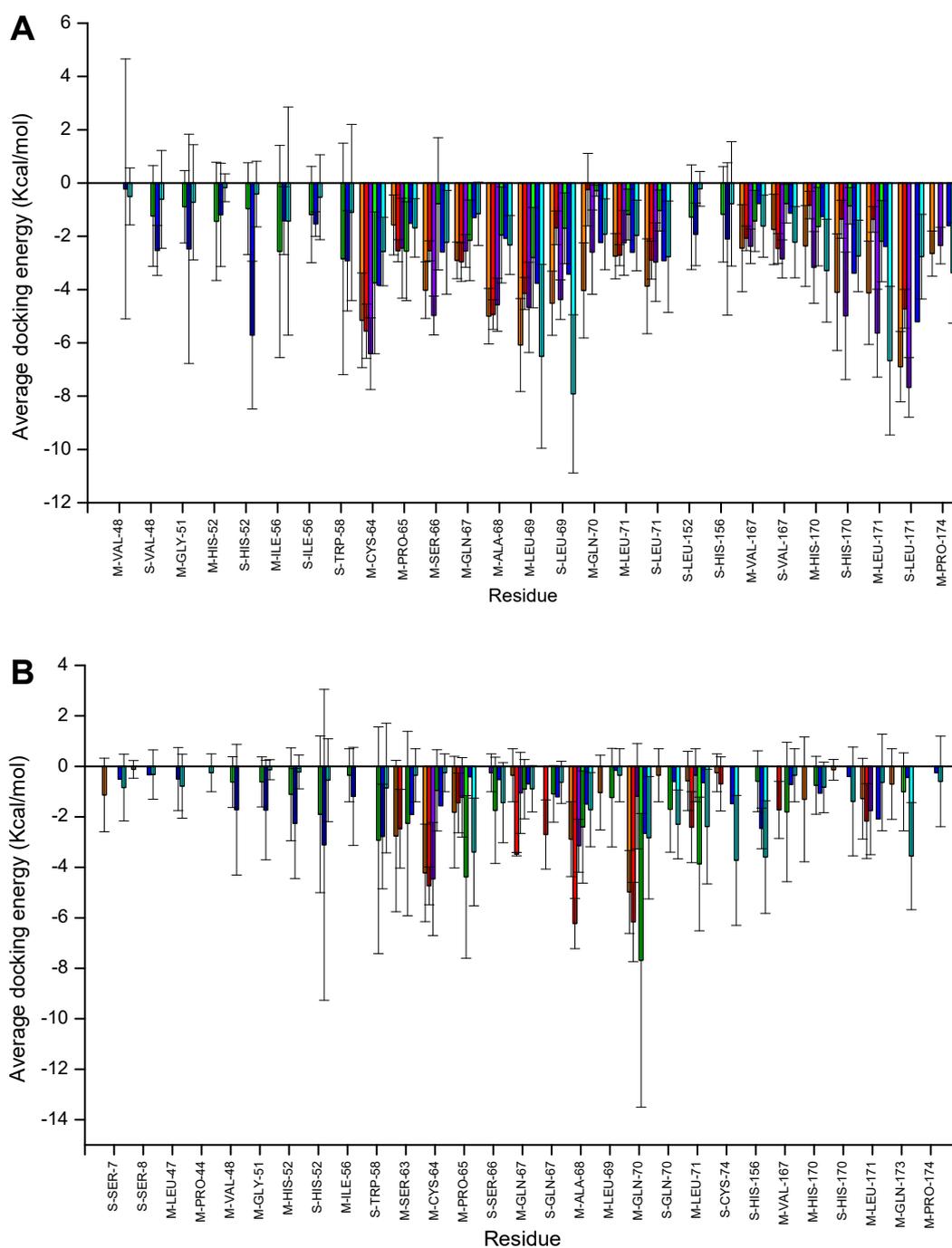


Figure 64. **Average GCSF residue docking energies with excipients using IGEMDOCK** A) H-bond interactions, and B) van der Waal (VDW) interactions. Excipients docked energy calculated from 10 docked positions with the GCSF PDB 2D9Q. Excipient structures included arginine (orange), mannitol (red), phenylalanine (purple), sorbitol (green), sucrose (dark blue), Trehalose (cyan). Side chain (S) and main chain (M) residue interactions labelled along with 3 letter amino acid residue and position number.

6.1.5 Cluster analysis

From the iGEMDOCK cluster analysis, all excipients docked at the same region of GCSF around the C-terminal and start of loopAB2 (Figure 65). From the residue interaction and cluster analysis, the amino acids arginine and phenylalanine as well as mannitol docked exclusively at this region, whereas sorbitol, sucrose and trehalose also docked in a second region around the short helix, end of loopABII and start of α A.

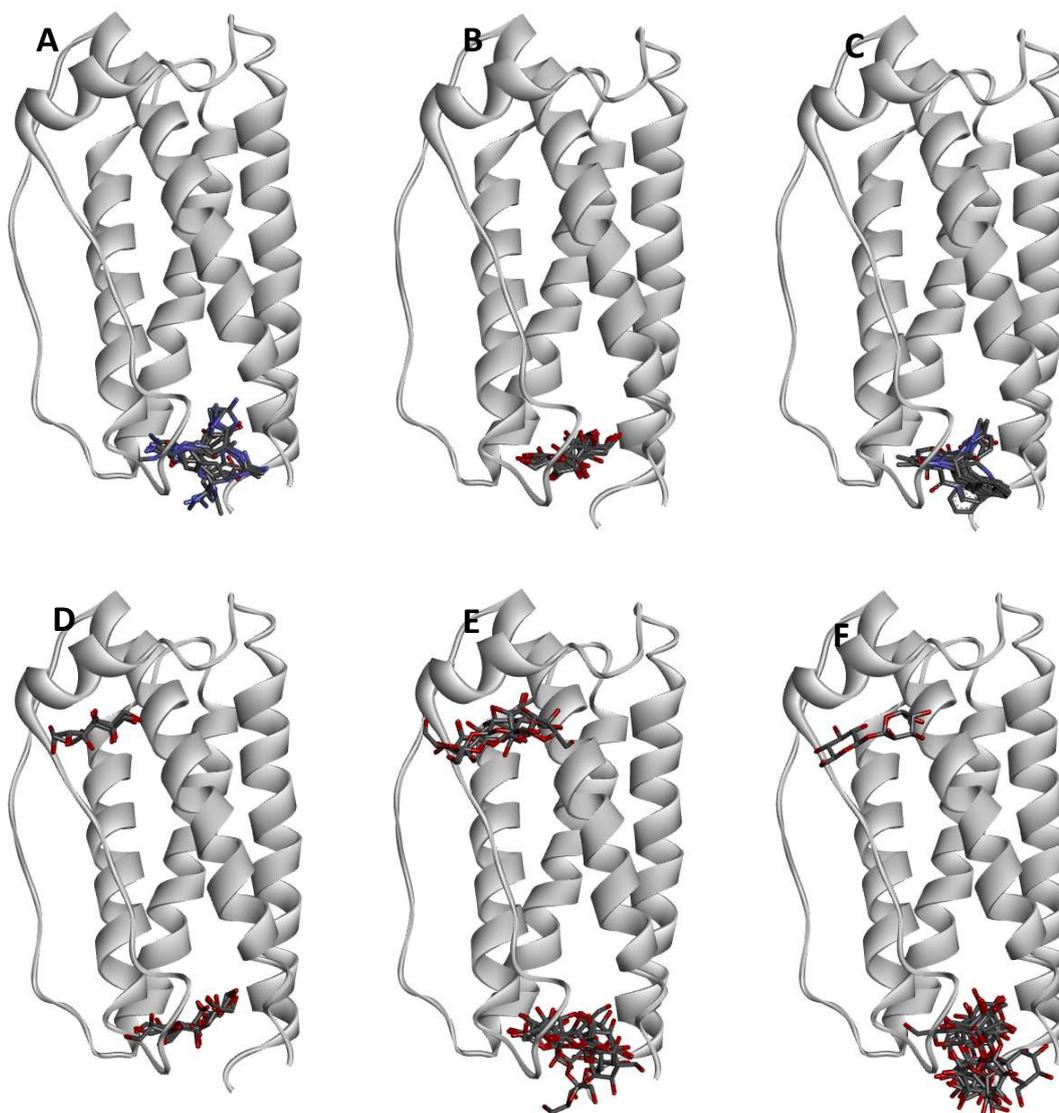


Figure 65. **3D representation of excipient clustering on 2D9Q.** 10 best poses displayed A) Arginine, B) Mannitol, C) Phenylalanine, D) Sorbitol, E) Sucrose, F) Trehalose. Using Biovia discovery studio 2016

6.1.6 Deuterated excipient HDX-MS

6.1.6.1 Deuteration of excipients

Excipients selected for HDX-MS experiments included arginine, mannitol, sucrose and phenylalanine. These excipients were chosen as they are easily freeze-dried and represent different classes of excipients: amino acids (arginine and phenylalanine), saccharides (sucrose) and sugar alcohols (mannitol). Sorbitol was not selected as it created a gel in the bottom of the vial during previous freeze-drying attempts.

Excipient concentrations were maximised to ensure changes in uptake would be measurable. Mannitol and sucrose were dissolved at a concentration of 10% (w/w), and phenylalanine and arginine were dissolved at concentrations of 100 mM and 50 mM respectively. Deuterated excipients were dissolved in D₂O, freeze-dried and reconstituted in D₂O three times. An image of the freeze-dried cakes formed is shown in Figure 66, where arginine was the only excipient to not form a solid white cake.



Figure 66. **Image of freeze-dried cakes formed during excipient deuteration.** Excipients were dissolved in 99% deuterium oxide and filled in 2 mL glass vials to a volume of 1mL. Vials were freeze-dried, stoppered with rubber stoppers and capped. The top of the vials was coloured based on the excipient, 10 % (w/v) mannitol (red), 50 mM arginine (black), 10 % (w/v) trehalose (green), 10 % (w/v) sucrose (blue) and 100 mM phenylalanine (yellow).

6.1.6.2 HDX-MS

Due to a limitation in the number of vials the LEAP-PAL™ quench and labelling trays can accommodate, HDX-MS experiments were performed in three consecutive runs. The first run generated the GCSF peptide map, the second included mannitol and sucrose, and the third included arginine and phenylalanine. The control D₂O solution used to prepare the excipient solutions was also included in all runs as the non-excipient control. Labelling time points were selected to measure fast exchange (2.6 min), mid exchange (1 hr) and slow exchange (8 hrs). Blank samples (10 mM sodium acetate, pH 4.25) were injected between each sample set to monitor peptide carryover.

6.1.6.3 Peptide map

The GCSF peptide map generated from peptides identified in four out of five on-line pepsin column digests is displayed in Figure 67, where a sequence coverage of 96% and a redundancy of 3.81 was achieved.



Figure 67. **GCSF peptide map for deuterated excipient HDX-MS experiments.** Five separate injections of 0.3 mg/mL GCSF were digested by an on-line pepsin column followed by LC-MS analysis. Of the 755 peptides produced, 58 peptides survived filtering: maximum MH⁺ Error (ppm) was set to 20 and the minimum products per amino acid was 0.01, as well as a file threshold of 4 was also applied. The top line is the amino acid sequence of GCSF, second line is residue number, and third line is the exact positions of peptides produced.

6.1.6.4 Internal reference peptide

An internal reference peptide (IRP) was included in the GCSF sample to monitor and correct for any changes to the intrinsic rate of uptake caused by differences in labelling solutions. The peptide was analysed alongside GCSF in the HDX .raw files using the DynamX 3.0 software by inputting the retention time of 6.31, amino acid sequence “PPPI” and maximum uptake of 1.0. The average relative uptake over time for the IRP in the control and deuterated excipient samples is displayed in Table 24. The differential between control and experimental relative uptake was calculated to obtain correction values for GCSF differential data. In general, the difference in intrinsic exchange between solutions and the control was slightly lower for the sugars (Table 24A) and negligible for the amino acids (Table 24B) suggesting excipients were almost fully deuterated. Once corrected for these small differences in intrinsic exchange, any changes in HDX kinetics induced by the addition of excipients are attributable to changes in the backbone flexibility of GCSF.

Table 24. **Relative deuterium uptake for internal reference peptide with different excipient formulations.**
 A) Run 1, B) Run 2. Differential calculated as excipient uptake minus control uptake.

A

State	Exposure (min)	Uptake (Da)	Uptake SD	Differential correction (Da)	Differential correction SD
Control	0.0	0.000	0.000		
	2.6	0.163	0.006		
	60.0	0.800	0.005		
	480.0	0.836	0.004		
Mannitol	0.0	0.000	0.000	0.000	0.000
	2.6	0.148	0.006	-0.015	0.012
	60.0	0.811	0.003	0.011	0.008
	480.0	0.849	0.007	0.013	0.010
Sucrose	0.0	0.000	0.000	0.000	0.000
	2.6	0.144	0.004	-0.019	0.010
	60.0	0.792	0.007	-0.008	0.012
	480.0	0.850	0.004	0.014	0.008

B

State	Exposure (min)	Uptake (Da)	Uptake SD	Differential correction (Da)	Differential correction SD
Control	0.0	0.000	0.000		
	2.6	0.143	0.005		
	60.0	0.792	0.007		
	480.0	0.836	0.002		
Arginine	0.0	0.000	0.000	0.000	0.000
	2.6	0.152	0.007	0.009	0.012
	60.0	0.793	0.008	0.001	0.015
	480.0	0.834	0.008	-0.002	0.010
Phenylalanine	0.0	0.000	0.000	0.000	0.000
	2.6	0.143	0.005	0.000	0.010
	60.0	0.796	0.007	0.004	0.014
	480.0	0.833	0.005	-0.003	0.007

6.1.6.5 Charge state uptake discrepancies

Multiple charge states are beneficial to estimate deuterium uptake more accurately, however, during manual review of all sample stacked spectral plots, it was found that there was a disagreement between charged states causing large standard deviation values when averaged. This shouldn't normally occur as the mass of the peptide should be the same at

each charge state, however, Kate Groves (LGC, Teddington, UK) completed further investigations into this (data not shown) and found peptides showed higher amounts of back exchange for +2 compared to +1. As such all +2 spectra, where +1 was available, were discounted from analysis.

6.1.6.6 *Effects of excipients on GCSF HDX-MS*

Relative uptake values for each peptide were calculated by taking m/z average value of the isotope cluster at each labelling time point, and subtracting the average m/z at 0s (non-deuterated peptide). The differential uptake for each peptide was obtained by subtracting the control relative uptake from the experimental relative uptake. Values were then corrected for differences in intrinsic exchange as calculated by the IRP differential values (Table 24). The differential uptake plots for each excipient versus the control were displayed in Figure 68 to Figure 71, where a positive differential indicated lower deuterium uptake in the control sample and vice versa.

For the fast exchange differential values (2.6 min) the majority of the GCSF peptides showed positive differential uptake values with mannitol, phenylalanine and sucrose. This indicated that the protein had higher exchange when these excipients were present in the formulation compared to the buffer only control solution. After slow exchange (8 hrs) with the same three excipients, the majority of the GCSF peptides showed negative differential uptake values, indicating that the protein had a lower exchange when these excipients were present in the formulation compared to the buffer only control solution. The mid exchange (1 hr) showed a mid-point between the fast and slow exchange data sets with the differential values close to 0 (no difference) for mannitol and sucrose, and a weak negative differential for phenylalanine. As such, when displayed together, all time points showed a clear transition from high to low deuterium uptake over time when incubated with excipients. An exception to this trend came with the addition of arginine, which caused a negative differential uptake value across the majority of the GCSF peptides for all time points (Figure 68).

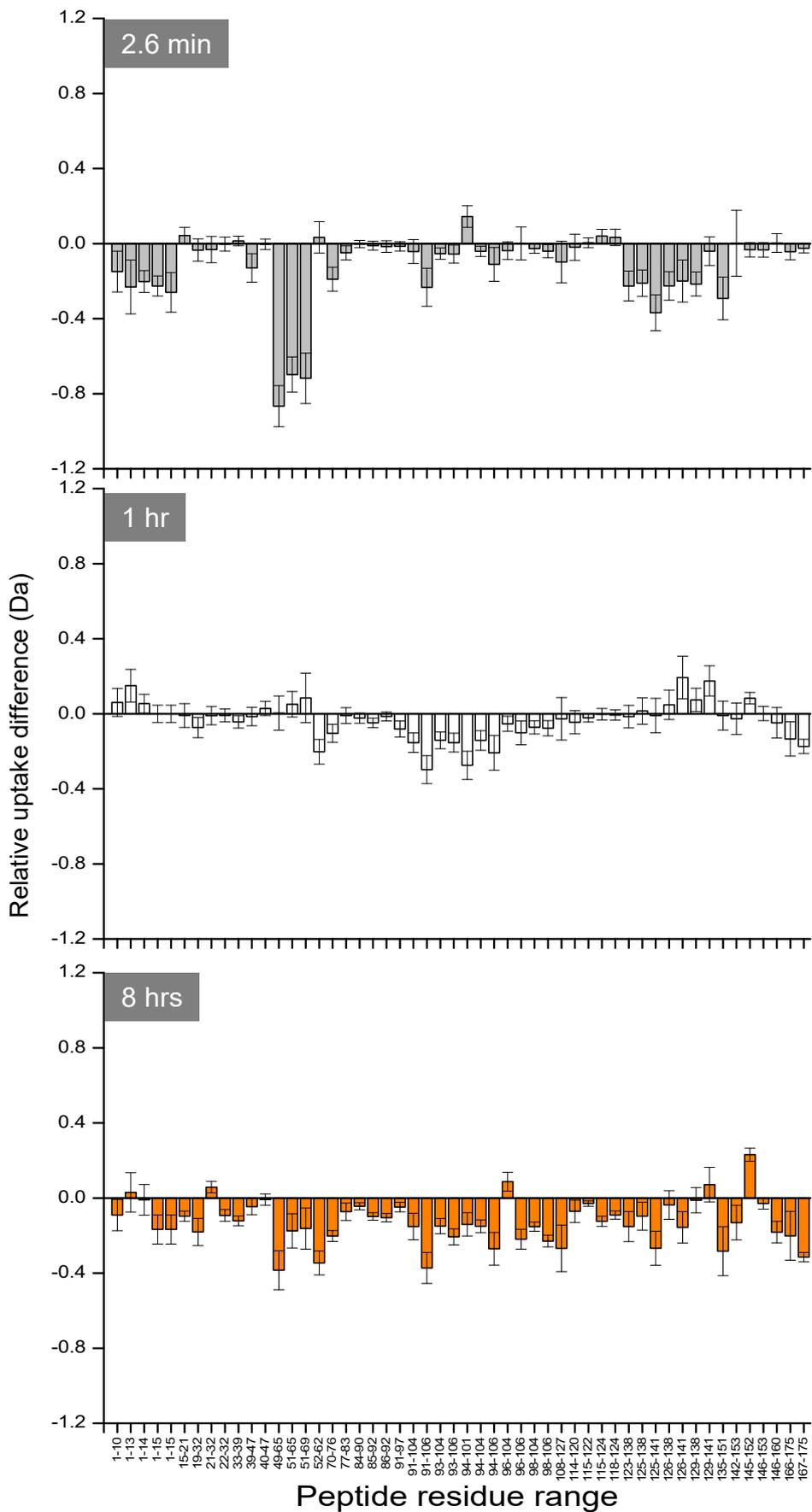


Figure 68. **Differential uptake plots of GCSF pH 4.25 in the presence of deuterated arginine over time.** Relative uptake difference is shown for all peptides; positive values indicate excipient addition caused an increase in deuterium uptake, whereas a negative value indicates decreased uptake. Difference plots are shown for different incubation times (grey boxes) ranging from fast (2.6 min), mid (1 hr) and slow (8 hrs).

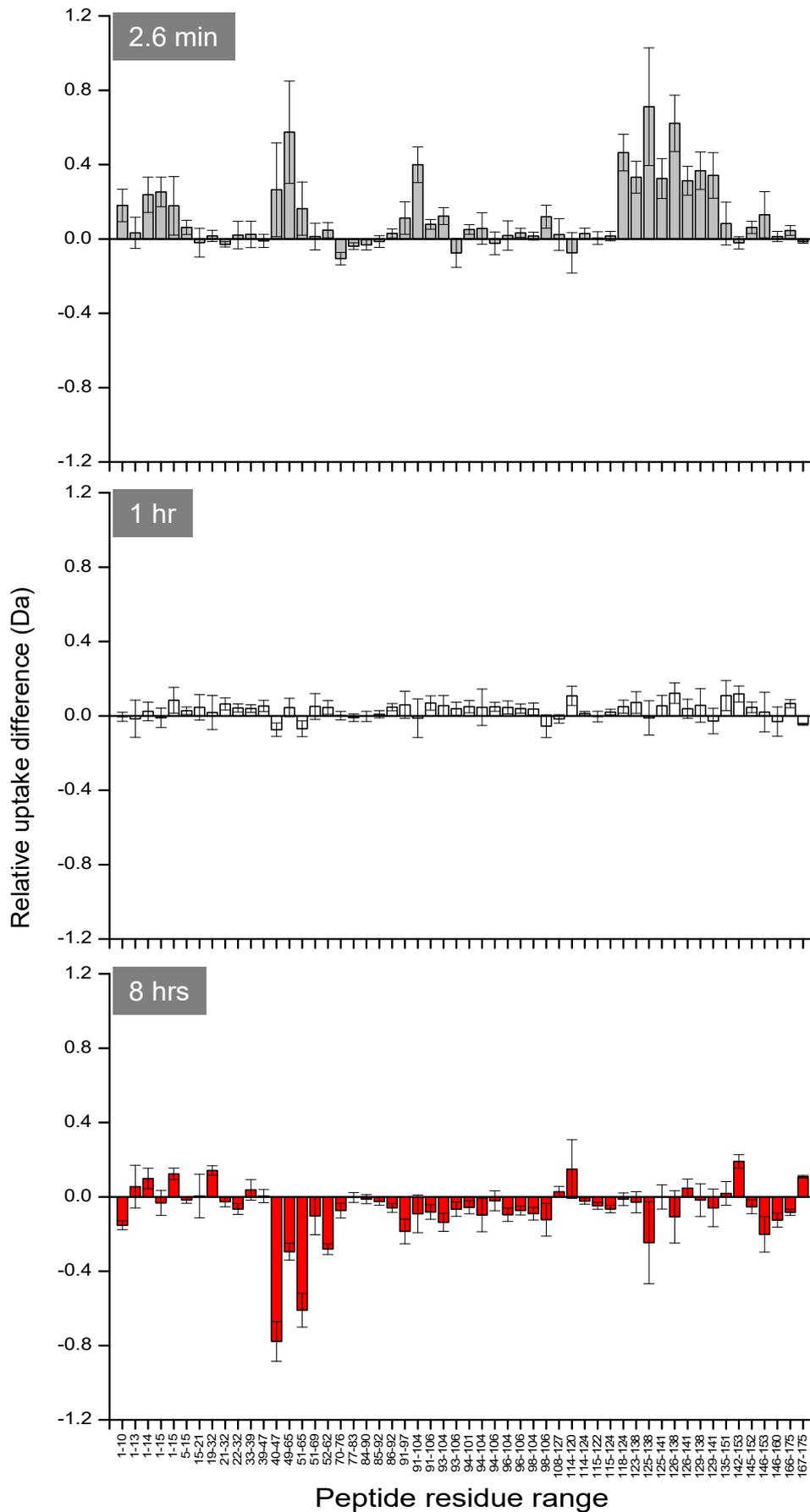


Figure 69. **Differential uptake plots of GCSF pH 4.25 in the presence of deuterated mannitol over time.** Relative uptake difference is shown for all peptides; positive values indicate excipient addition caused an increase in deuterium uptake, whereas a negative value indicates decreased uptake. Difference plots are shown for different incubation times (grey boxes) ranging from fast (2.6 min), mid (1 hr) and slow (8 hrs).

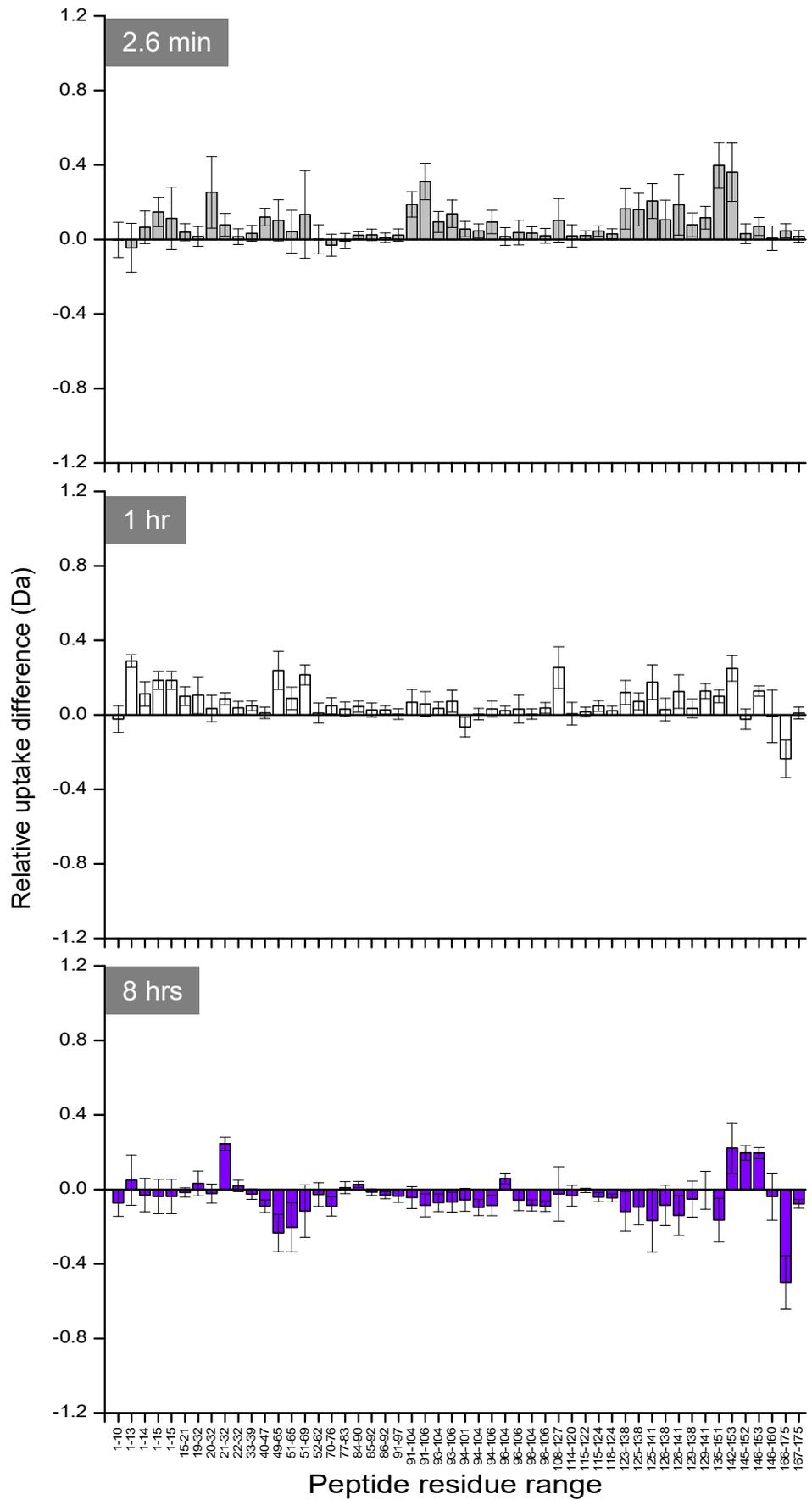


Figure 70. **Differential uptake plots of GCSF pH 4.25 in the presence of deuterated phenylalanine over time.** Relative uptake difference is shown for all peptides; positive values indicate excipient addition caused an increase in deuterium uptake, whereas a negative value indicates decreased uptake. Difference plots are shown for different incubation times (grey boxes) ranging from fast (2.6 min), mid (1 hr) and slow (8 hrs).

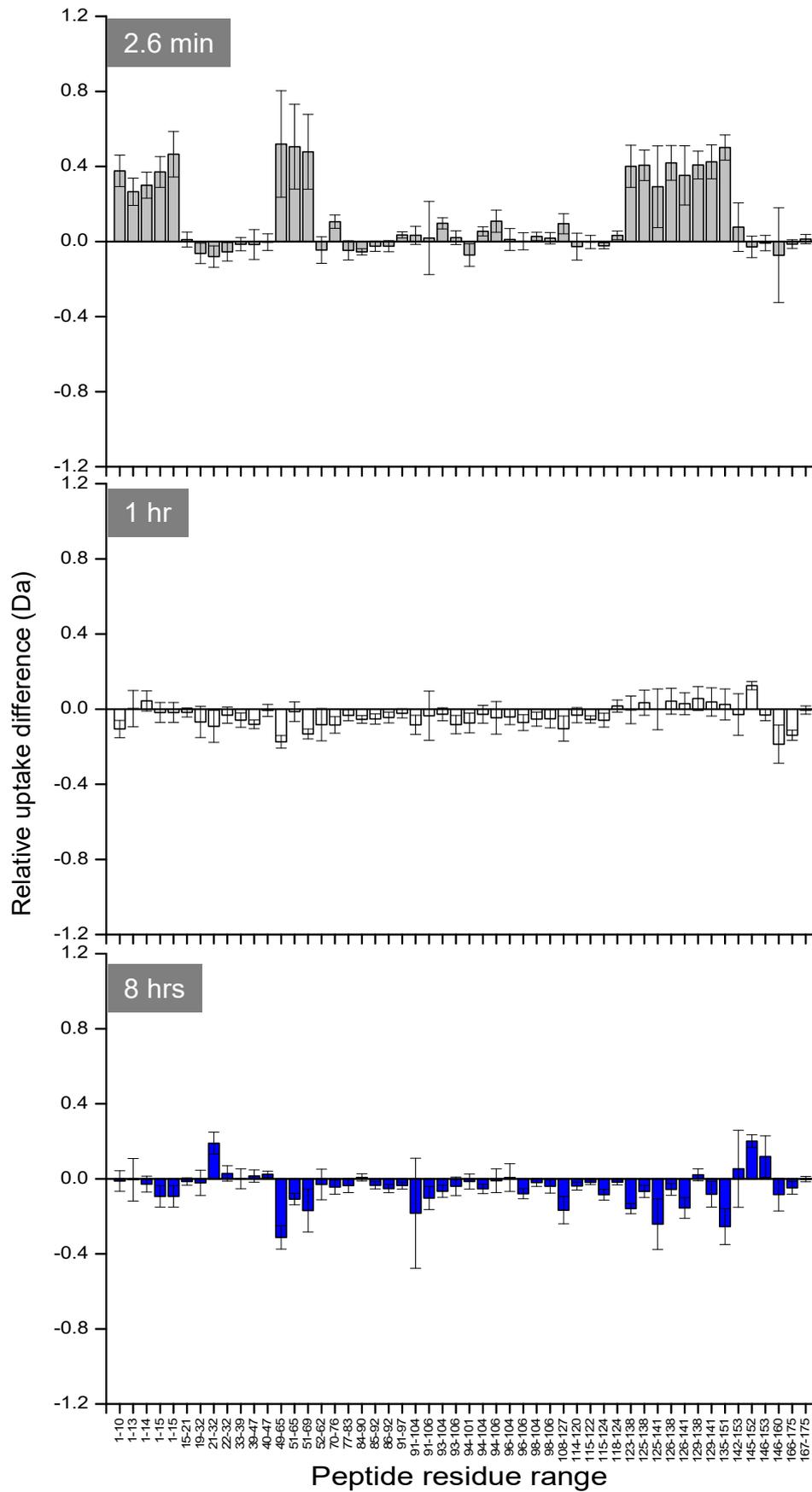


Figure 71. **Differential uptake plots of GCSF pH 4.25 in the presence of deuterated sucrose over time.** Relative uptake difference is shown for all peptides; positive values indicate excipient addition caused an increase in deuterium uptake, whereas a negative value indicates decreased uptake. Difference plots are shown for different incubation times (grey boxes) ranging from fast (2.6 min), mid (1 hr) and slow (8 hrs).

6.1.6.7 Aligning HDX protection with predicted binding

As the slow exchange time point data presented regions with the most obvious excipient protection effects (negative differential uptake) this data was taken forward for validation of the iGEMDOCK residue interaction predictions.

6.1.6.7.1 Mannitol

Docking for mannitol identified a single bound pose covering a small set of GCSF amino acid residues. Peptides that included these residues were highlighted on the differential data in Figure 72. There was a match between the peptides with largest negative differential uptake (protected from deuterium exchange) and the peptides with predicted excipient binding residues. There was also an increase in protection across the majority of the GCSF peptides as well as decrease in protection for many N-terminal peptides covering residues 1-40, which were not picked up by the binding screening results. This is most likely due to the global stabilisation of the protein mediated through a specific ligand interaction with the protein at one key site.

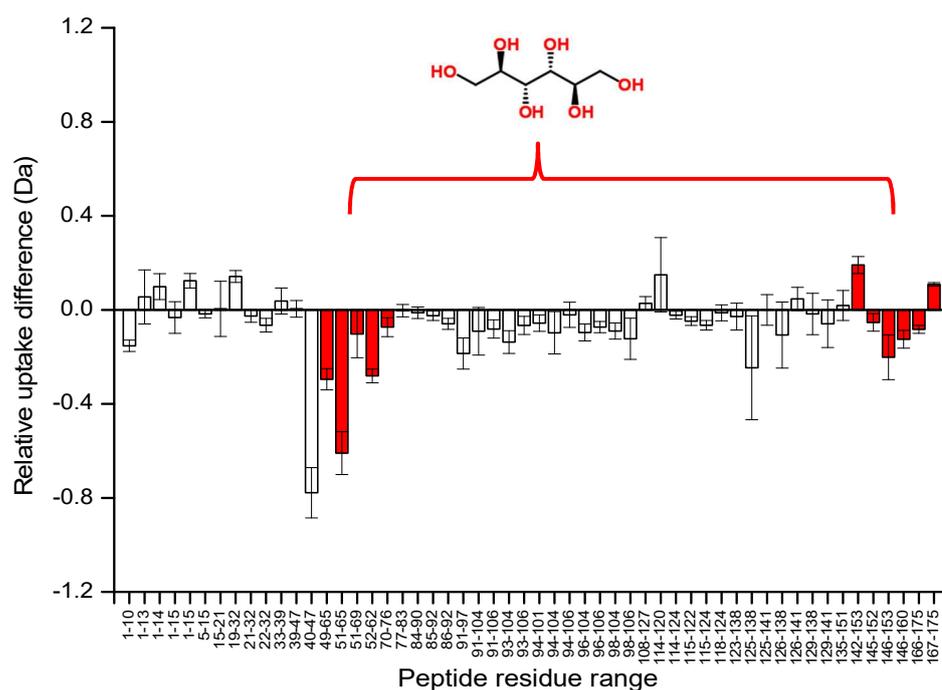


Figure 72. **Alignment of HDX-MS differential uptake with in silico docking data for GCSF and mannitol.** Differential uptake for GCSF peptides incubated with deuterium and deuterated mannitol for 8 hrs, peptides containing iGEMDOCK docked residues are highlighted in red.

6.1.6.7.2 Sucrose

The differential uptake of each peptide with sucrose is shown in Figure 73. The peptides that cover residues involved in docking were also highlighted. The data showed a match between the region of lowest differential uptake (protected from deuterium exchange) and excipients docking of sucrose. The mannitol peptides involved in docking showed an increase after 8 hrs labelling, whereas when sucrose was added the same peptides had a lower level of protection suggesting the docking data matches experimental data well. Most of the GCSF peptides showed a decrease in differential uptake with sucrose suggesting global protection occurred, rather than protection of a single docked position, however there was also a region of significant positive differential in peptides towards the C-terminal of the protein indicating that ligand interaction also destabilised local regions via localised shifts in structure.

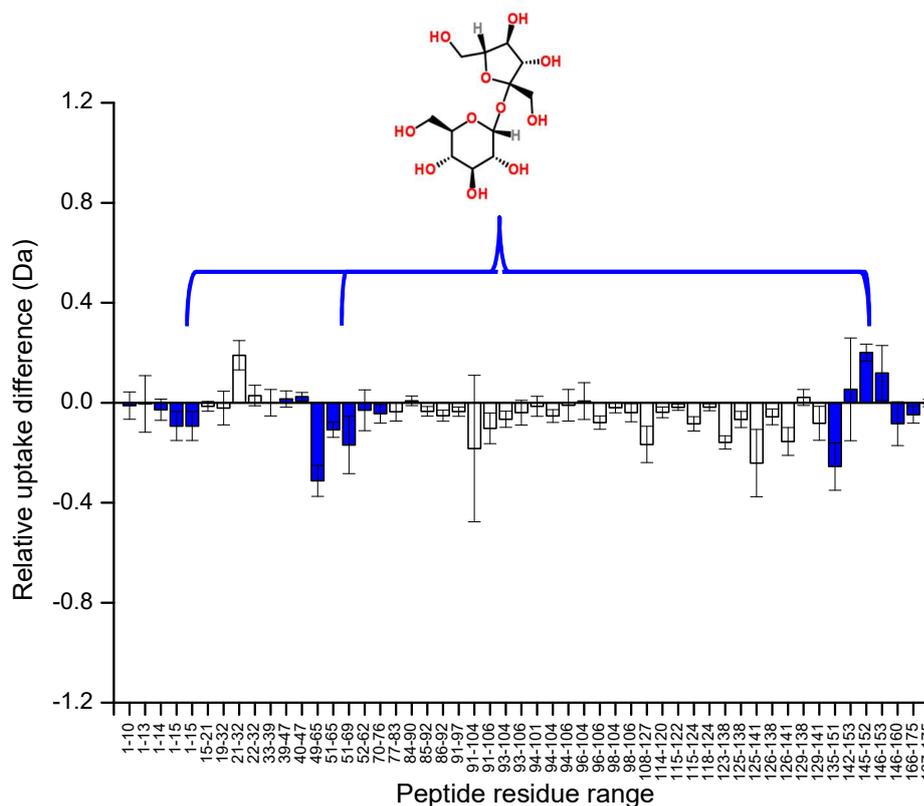


Figure 73. **Alignment of HDX-MS differential uptake with in silico docking data for GCSF and sucrose.** Differential uptake for GCSF peptides incubated with deuterium and deuterated sucrose for 8 hrs, peptides containing iGEMDOCK docked residues are highlighted in blue.

6.1.6.7.3 Phenylalanine

The differential uptake of each GCSF peptide with phenylalanine is shown in Figure 74. Docking for phenylalanine identified a single docked pose at the same region as mannitol and arginine. With mannitol there was a large decrease in exchange around the first docking region whereas with phenylalanine this region was not as protected. The lowest exchanging region (negative differential) for phenylalanine HDX was around the C-terminal of the protein, which was also identified as a region for phenylalanine docking. As such, the data again shows a match between the region of greatest negative differential uptake and excipient docking. As seen with sucrose, there was a region of peptides next to the C-terminal with increased uptake (positive differential) for phenylalanine suggesting perturbation of the structure upon interaction at the predicted docking region.

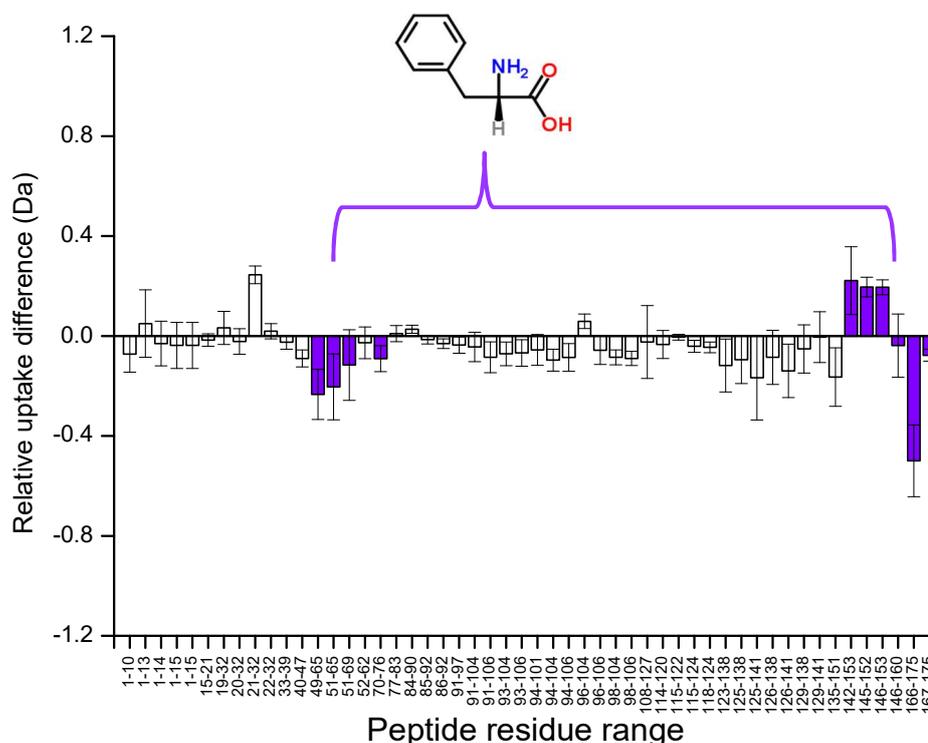


Figure 74. **Alignment of HDX-MS differential uptake with in silico docking data for GCSF and phenylalanine.** Differential uptake for GCSF peptides incubated with deuterium and deuterated phenylalanine for 8 hrs, peptides containing iGEMDOCK docked residues are highlighted in purple.

6.1.6.7.4 Arginine

The differential uptake of each peptide with arginine is shown in Figure 75. Docking for arginine identified one single docked pose also highlighted in Figure 3. From the deuterated excipient differential data, it appeared that inclusion of arginine during HDX experiments provided global protection from deuterium uptake. The data showed a weak match between peptides with strong negative differential uptake and peptides including residues involved with excipient docking; however, with arginine protection appeared to be global rather than one obvious region of protection.

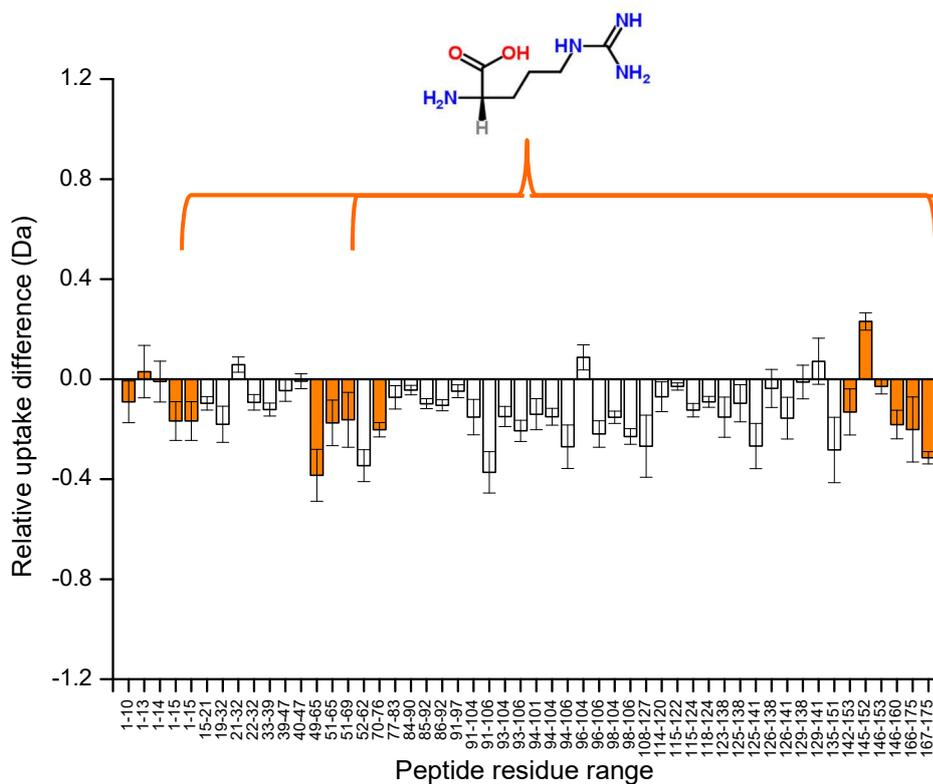


Figure 75. **Alignment of HDX-MS differential uptake with in silico docking data for GCSF and arginine.** Differential uptake for GCSF peptides incubated with deuterium and deuterated arginine for 8 hrs, peptides containing iGEMDOCK docked residues are highlighted in red.

6.1.7 Equilibration tests using intrinsic fluorescence and static light scattering

6.1.7.1 Equilibration analysis

Results from the HDX-MS deuterated excipient experiments suggested an equilibration time is required before analysis, as in the early time points (less than 1 hr) the protein structure appeared perturbed by the excipient compared to the control sample. To explore this further, T_m and T_{agg} measurements using the UNit with different GCSF and excipient mixing times was performed to investigate any changes caused by equilibration.

A faster than normal run time was designed by applying a temperature ramp from 30-90 °C at a rate of 2 °C/min, to ensure the main T_{agg}/T_m events would occur after 10-15 mins. Samples were prepared and equilibrated at RT to mimic the temperature used during HDX-MS labelling. Samples were prepared at the same concentrations used in the deuterated HDX-MS experiments, with the exception of the mannitol and sucrose as a 20% (w/v), 2X formulation, could not be achieved for mannitol due to solubility issues, therefore both sugars were prepared at a final concentration of 8% (w/v). A control sample was also included and prepared by mixing GCSF with 10 mM sodium acetate pH 4.25. The mixed samples were incubated at RT for 5 mins, 1 hr and 3hrs. All samples for the same formulation were analysed at the same time, therefore sample preparation was staggered. The 3 hr time point samples were prepared first, followed by the 1 hr and finally the 5 mins. The 3 hr and 1 hr samples were loaded prior to the 5 min sample preparation to reduce the time taken to prepare the last sample and run the equipment.

From the mannitol BCM and SLS data in Figure 76, there was a distinct difference between the 5 min incubated sample and the two longer incubation time points. This was true for all samples including the no excipient control (Appendix 13). The 5 min sample followed the same BCM profile as the 1 hr and 3 hr samples but had a higher value at each temperature point, which is usually indicative of being more stable. The same result was seen for the SLS profile, which at 5 min had a later aggregation onset seen by a shift in the curve to the right. The scatter intensity was also lower for the 5 min sample compared to the other two time points.

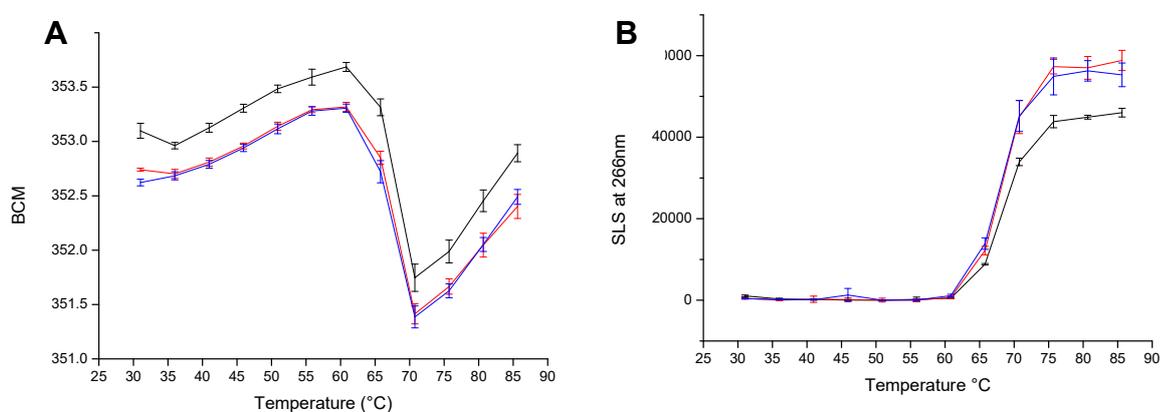


Figure 76 **Effect of sample equilibration time on GCSF SLS and fluorescence measurements.** Sample contained 0.5 mg/mL GCSF mixed with 8% (w/v) mannitol, 10mM sodium acetate pH 4.25 and incubated for different periods of time prior to thermal denaturation analysis. 5 min (black), 1hr (red), blue (3 hrs).

6.1.8 Effect of internal reference peptide on GCSF stability

The T_{onset} , T_m and T_{agg} values were determined for 0.3 mg/mL GCSF in 10 mM sodium acetate pH 4.25 with and without the IRP to assess its effect on protein stability. From the results in Figure 77 it can be seen that all values for the samples are the same. As such it can be concluded that IRP and the small amount of buffer it is in does not have an effect on the structural stability of GCSF samples during HDX-MS experiments.

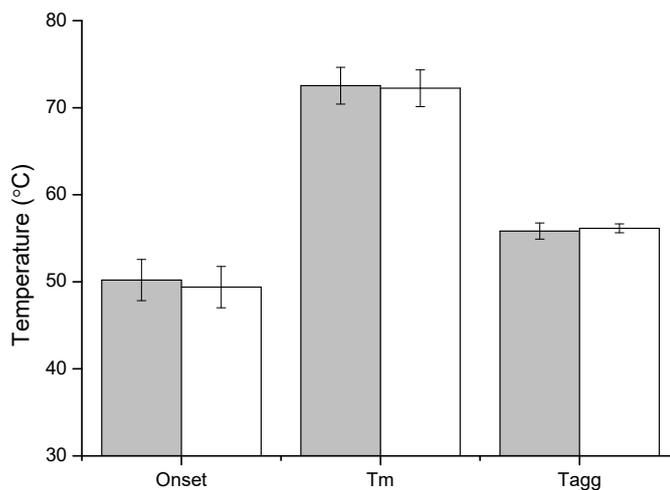


Figure 77. **Thermal stability of GCSF with and without an internal reference peptide.** Tandem intrinsic protein fluorescence (IPF; 266 nm excitation, 280 to 450 nm emission scan) and static light scattering (SLS) at 266 and 473 nm were measured for GCSF in 10 mM sodium acetate pH 4.25 with (white) and without (grey) 0.3 μ M internal reference peptide (IRP). The values T_{onset} , T_m and T_{agg} were calculated using analysis described in 2.4.6.

6.2 Discussion

The screening of excipients for formulations has traditionally taken an empirical approach with the majority of selections based on previous literature and indirect analysis. With the advancement of biophysical analysis, techniques such as HDX-MS, and NMR can measure the change in dynamics of proteins in different solutions, allowing for more in-depth insights into the mechanism of stabilisation. Protein-excipient interactions can also be modeled computationally using docking simulations of the 3D molecular structures, further reducing the amount of time and material required during early phase screenings. This chapter aimed to study WT GCSF protein-excipient relationships using both techniques in combination to further understand the effect of excipients and local flexibility on protein aggregation.

6.2.1 Excipient selection

The excipients used in this chapter were selected based on previous use in either GCSF formulation studies or common use in commercial parenteral formulations.

Arginine, a positively charged amino acid, is widely used as a solubilising agent in protein purification steps such as inclusion body recovery, as discussed in chapter 1. Arginine can also be used as a pharmaceutical excipient. Two patents by Michaelis et al. (1999) included arginine in aqueous and lyophilised preparations of GCSF, suggesting arginine can stabilise GCSF. The lyophilised preparation included arginine at twice the concentration of protein, whereas the aqueous formulation included arginine as the buffering agent at 5-80 mg/mL, pH 7.0 - 7.5, with 0.35 mg/mL GCSF.

Mannitol is a hexahydric sugar alcohol, commonly used in solid state formulations as a bulking agent and in aqueous formulations as a tonicity agent (Rowe et al. 2006). Mannitol is an isomer of sorbitol, an excipient included in the commercial formulation of GCSF (Neupogen®) at a concentration of 50 mM (Alebouyeh et al. 2016). Both mannitol and sorbitol were selected for study to determine any differences between the two.

Phenylalanine is a hydrophobic amino acid selected due to SLS data generated during preliminary GCSF formulation screenings, where phenylalanine reduced GCSF aggregation levels at elevated temperatures to the point where the scatter intensity at that temperature was significantly lower than any other excipient or control. Further to this, a patent filed by Sato et al. (2000) combined phenylalanine with another amino acid, such as arginine, in GCSF aqueous formulations to increase long-term storage stability of GCSF.

Sucrose was selected as it is one of the most commonly used sugars in pharmaceutical formulations. Trehalose was also included as an alternative sugar to sucrose (Carpenter et al. 1997). Both sucrose and trehalose are primarily used as cryoprotectants for solid state formulations, however, sucrose was also found to inhibit GCSF aggregation under physiological conditions in the liquid state (Krishnan et al. 2002; Zhang et al. 2015).

6.2.2 High throughput screening

GCSF-excipient short term stability measurements such as T_m/T_{agg} values have been previously shown to correlate with real-time shelf-life results (Maddux et al. 2014). As such, these measurements were determined for each protein-excipient formulation and used in this chapter as a way of determining the stabilising capability of each selected excipient. Two different concentrations of GCSF were included in the experiment, where it was found that T_{agg} was higher in all lower concentration samples compared to the identically formulated high concentration samples. This concentration dependent aggregation result follows the same trend shown in chapter 3 (Section 3.1.6). Increased T_{agg} of GCSF was observed for all excipients studied apart from arginine, and there was no clear difference between excipient T_{agg} values apart from arginine which reduced T_{agg} compared to the control. This is likely due to the fact the protein is in a stable buffering system, however, a study by Toth et al (2018) identified mAb formulation thermal measurements with clear differences between excipients such as the ones studied in this chapter using DSC. As the T_{agg} standard deviation data was so large perhaps any future work continuing from this could re-assess formulations using fluorescence data such as FLI or BCM. A study by Ablinger et al. (2012) also evaluated the effects of buffer conditions and excipients on the thermostability of GCSF where it confirmed mannitol, trehalose, sorbitol, and sucrose have significant positive impact on T_m , and trehalose and sorbitol were clearly the most stabilising.

6.2.3 Evaluating iGEMDOCK total binding energies

The use of computational predictions for stabilising excipients could alleviate formulation screening burden during early-phase development by reducing the number of potential candidates in a short space of time. The use of docking software has primarily been used in pharmacology to determine enzyme-substrate and protein-protein interactions (Yang et al. 2004), however it has recently been applied to biopharmaceutical formulation design (Li et al 2013; Barata et al. 2016). To determine the applicability of such software to predict GCSF-excipient interactions, iGEMDOCK was selected as a first trial. The total binding

energy of each excipient to GCSF was calculated from the sum of H-Bond, electrostatic and VDW. For the excipients screened it was found that no electrostatic binding occurred and as such the total binding energy was a total of H-bond and VDW.

When aligned with thermal stability T_{agg} measurements, it was found that the total binding energy positively correlated for all excipients apart from arginine where the T_{agg} was lower than the control and so this value was removed from the correlation. The high R^2 value suggests docking is consistent with previous results obtained by Barata et al. (2016) for Fab A33 formulations and could indeed be used as a pre-formulation screen to reduce excipient candidate numbers.

Arginine possessing destabilising qualities highlighted an issue with depending primarily on iGEMDOCK information to select excipient candidates during development as the software cannot predict stabilisation or destabilisation of a protein structure, only the location and energy of interaction. It is possible, however, that arginine is a uniquely destabilising excipient and all others, as seen in this body of work, would in fact stabilise the structure. Further work to build on these results could include more commonly used commercial excipients in the screening process such as other amino acids, sugars, polymers and surfactants, to determine if arginine is the only exception to the correlation between thermal denaturation measurements and total binding energies.

6.2.4 Evaluation of iGEMDOCK excipient docked poses

The iGEMDOCK docked pose analysis showed all excipients interacted with one or two regions, or “hotspots” on GCSF. It was expected that the excipients would interact at multiple sites on the protein due to non-specific binding and a higher concentration of excipient relative to protein. The identification of one to two sites most likely occurred because the software ranked and sorted all docked conformations into the top 10 most energetically favourable conformations. This kind of output is useful for identifying protein-protein or protein-ligand interaction – a task the software was originally designed for, however for excipient interactions all possible interactions and scorings would be a more useful output. The study by Barata et al. (2016) using iGEMDOCK for protein-excipient docking, also found *Drosophila* Su(dx) protein (WW34) docked with arginine at a single site, whereas using another docking software package, GLUE (Goodford et al. 1985), two preferential sites were identified suggesting the latter software may be more appropriate. Future protein-excipient interaction modelling work could perhaps screen small sections of a protein in series to identify all possible interactions and binding energies across the

protein, rather than the single most likely pose. Such strategy could be achieved using the software AutoDock (Morris et al. 2009) where a set of grids can be manually set before each docking run.

Another issue with using docking software is the assumption that the protein and excipients directly interact. The sugar alcohol and sugar excipients included in this study have well established mechanisms of stabilisation, namely preferential exclusion from the protein surface whereby interactions between protein and excipients do not occur. Therefore, it may not be scientifically accurate to apply computational docking to such molecules.

6.2.5 Exploring GCSF-excipient interactions using HDX-MS

To shed light on protein-excipient interaction experimentally, HDX-MS was utilised by adding deuterated excipients within deuterium solutions and measuring changes in peptide uptake rates compared to a control deuterium solution containing no excipients. Excipients selected for HDX-MS analysis included arginine, mannitol, phenylalanine and sucrose. Arginine was selected to determine if the negative effects observed during thermal analysis were caused by specific interactions and/or changes to the protein structure. Mannitol was selected over its isomer, sorbitol, as sorbitol collapses during the primary drying stage of freeze-drying (a drawback of the current freeze-drying deuteration technique), and as such could not be fully deuterated (Kadoya et al. 2010). To reduce the number of HDX-MS experiments required, sucrose was selected over trehalose because although they both had similar T_{agg} improvements, sucrose had previously been used in GCSF HDX-MS experiments allowing for direct comparison with published data. Finally, phenylalanine was selected as a stabilising amino acid because little formulation data has been published and as such, novel insights into its suppression of GCSF aggregation could be explored.

6.2.6 Internal reference peptide reports changes in intrinsic exchange

The internal reference peptide (PPPI) first described by Zhang et al. (2012) was spiked into GCSF HDX-MS experiment samples to monitor the differences of intrinsic HDX rates. The peptide was assumed to have no conformational protection and has only one backbone amide hydrogen (from the isoleucine) giving it a theoretical maximum uptake of 1 Da. The maximum uptake observed for the IRP, when included with GCSF over 8hrs labelling in D_2O , was 0.8 Da and occurred after 1 hr. This is the same level reported by Zhang et al. (2012) when looking at different concentrations of Gdn.HCl. However, the uptake of IRP at 30s

labelling time was lower than that reported by Zhang et al. (2012), which was probably due to the low pH of GCSF sample compared to the pH 7 sample they used.

6.2.7 Internal reference peptide does not alter GCSF stability

The effect on GCSF stability by including the IRP in the sample was assessed using thermal unfolding methods, it was shown that there were no differences between samples and as such no structural changes or interaction was considered to occur.

6.2.8 Mechanisms of GCSF stabilisation by excipients

Protein stabilisation is achieved by one of three mechanisms: strengthening of the protein-stabilising forces, destabilisation of the denatured state, or direct binding to the protein (Jorgensen et al. 2009). Interactions between protein and excipients has previously been identified using equilibrium dialysis experiments where excipients either “preferentially interact” with the protein and are present in excess at the protein surface compared to the concentration in the bulk phase, as shown in Figure 19A. The opposite case shown in Figure 19B, and is called “preferential hydration” or “preferential exclusion” where there is excess water at the protein surface.

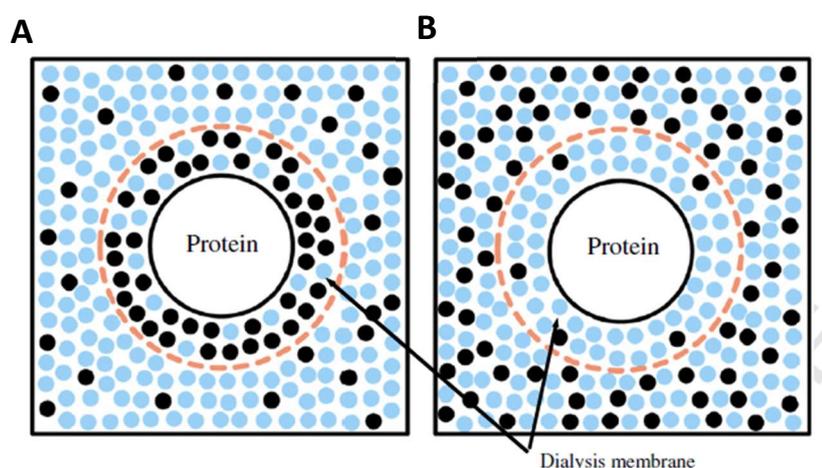


Figure 78. **Schematic presentation of A) preferential binding and B) preferential exclusion.** Performed in a dialysis equilibrium experiment where the protein is represented by a white circle, black circles represent excipient molecules and blue circles represent water molecules. . Taken from Ohtake et al. (2011)

The deuterated excipient HDX-MS experiments were performed to validate iGEMDOCK predicted protein-excipient regional interactions. By HDX-MS it was found that for mannitol, phenylalanine and sucrose, peptides with high level of protection by the excipient after labelling for 8 hrs matched up with peptides containing residues predicted to interact with the excipients by iGEMDOCK. Whilst this would provide evidence that the iGEMDOCK is successfully predicting the residues of interaction, most other peptides not containing residues with predicted interactions also showed differing levels of protection.

This result suggests both local and global protection may be occurring in solution; however, there is a large body of literature to show that most amino acids, sugars and polyols preferentially exclude themselves from the protein surface (Jorgensen et al. 2009). With this in mind, the data obtained during this chapter for each of the four excipients will now be considered along with previous protein-excipient interaction data to understand their potential mechanisms for stabilisation of GCSF.

6.2.8.1 Mannitol (and sorbitol)

Polyhydric compounds are among the most prevalent molecules used by nature to protect organisms against the stresses of high osmotic pressure (osmolytes) and freezing (cryoprotectants; Xie and Timasheff 1997). Consequently, these molecules have also been found to be excellent stabilisers of globular proteins in solution. Mannitol (and isomer sorbitol) molecules are hydrophilic and should have an affinity for the polar residues on the protein surface, however because these interactions are weaker than with protein-water molecules, the reaction favors preferential hydration, and exclusion of the excipients into the bulk solution (Xie and Timasheff 1997). From the GCSF-excipient T_{agg} measurements, addition of increasing concentrations of mannitol and sorbitol increased aggregation stability. This is consistent with previous work by Gekko and Morikawa (1981A) with chymotrypsinogen, where increasing either mannitol or sorbitol concentration increased the T_m . The increase in stabilisation with increasing concentration indicates the excipients are being preferentially excluded from the protein surface, as increasing excipient molecules in solution increases the compacting effects on protein species in the native state (Krishnan et al. 2002). Another study by Gekko and Morikawa (1981B) confirmed preferential hydration occurs with bovine serum albumin (BSA) and different polyhydric alcohols including sorbitol and mannitol. Density measurements were performed using a precision density meter where it was found that all preferential interaction parameters were negative, indicating a deficiency of the sugar alcohols at the surface of the protein. The deuterated HDX-MS experiments from this chapter also confirm preferential hydration is most likely occurring, as most GCSF peptides showed a level of decrease in deuterium uptake (global protection) in the presence of mannitol. The slower rate of deuterium exchange is most likely caused by compaction of the native state and reduction of protein flexibility (Carpenter et al. 2002). Mannitol's effect on decreasing peptide exchange was not the case for a 145 kDa mAb (IgG4) studied by Toth et al. (2018) using similar HDX-MS methods to those in this chapter. The group found deuterated mannitol at 0.8 M, pD 6.5 to 7.4 had stabilising effects on T_m as measured by DSC, but by

HDX-MS, mannitol caused a small increase in average global hydrogen exchange indicating destabilisation of the mAb. This difference in findings suggests mannitol and other polyhydric compounds may have varying levels of preferential exclusion depending on concentration and pH of solution, which may affect HDX differently.

By computational docking, mannitol was shown to interact at one specific position on GCSF and this matched with peptides showing higher levels of protection from deuterium uptake by HDX-MS. Whilst previous studies have shown the main mechanism of stabilisation for mannitol is via preferential exclusion, a study by Xie and Timasheff (1997) on RNase A and sorbitol interactions, found at 48 °C when increasing concentrations of sorbitol, preferential interaction parameter values decreased, indicating at concentrations (> 20% w/v) excipient molecules started to interact with RNase A. With this result, it is possible that both preferential exclusion and interaction are occurring between mannitol and GCSF in solution, as mannitol was included at a reasonably high concentration (10% w/v). Additional HDX-MS experiments could measure changes in differential uptake of docked regions by varying concentrations of mannitol (5%– 20% w/v) where no change in the differential with increased concentration would confirm site specific interactions.

6.2.8.2 *Phenylalanine*

The SLS profile for phenylalanine was different to the standard sigmoidal aggregation curve seen for the control and other excipient formulations. At the mid to high concentration range, the phenylalanine SLS signal remained very low throughout the run, even at the elevated temperatures of >70 °C where the protein usually melts and the scattering generally becomes unpredictable. The lack of SLS signal suggests the excipient is preventing protein-protein interactions.

Phenylalanine is a poorly studied excipient and as such its mechanism of protein stabilisation is currently unknown, however there have been many studies on L-phenylalanine in relation to phenylketonuria (PKU), a metabolic disorder where a person is unable to utilise phenylalanine, which may help elucidate its mechanism of action as an excipient. PKU studies have shown that increased blood phenylalanine concentration, results in self-assembly and fibril formation. A study by Adler-Abramovich et al. (2012) linked phenylalanine self-assembly with PKU and by using congo red dye binding, thioflavin T (ThT) binding and electron diffraction studies, found these self-assembled fibers are amyloid in nature. MD were also performed by Adler-Abramovich et al. (2012) where multiple microsecond-long simulations using a generalised a Born implicit solvent model,

aligned 27 monodispersed phenylalanine molecules, at different pH values and in the presence or absence of counter ions. The simulations predicted that at high pH the molecules in the presence of counter ions produced filamentous aggregates in high concentrations, at all temperature values (Figure 79A). Pairs of neighboring phenylalanines were found to be involved in direct H-Bonds or salt-bridged polar interactions. The self-assembly process of L-phenylalanine was further characterised by Singh et al. (2014) using a variety of analytical techniques including NMR, light scattering, particle size analysis, scanning electron microscopy (SEM) imaging, ThT and ANS binding. Self-assembly in water pH 5.8 was monitored by light scattering and ThT binding assays at different concentrations, where an increase in light scattering was observed at concentrations above 60 mM and an increase in fluorescence was observed at concentrations above 6 mM. By SEM, long fibrous structures were observed at low (300 μ M) and high concentrations (6 mM - 300 mM), as well as high and low pH (Figure 79), which suggests association also occurs at the low pH studied during this chapter. To probe the role of hydrophobic interactions in self-association, DLS was used to measure particle size with increasing ionic strength and temperature at 60 mM phenylalanine, where it was observed that high temperature and high ionic strength encouraged formation of higher order assemblies. As increased salt concentration would suppress charge interactions, this increase in higher order assembly suggests hydrophobic interactions are the main driving forces behind the associations.

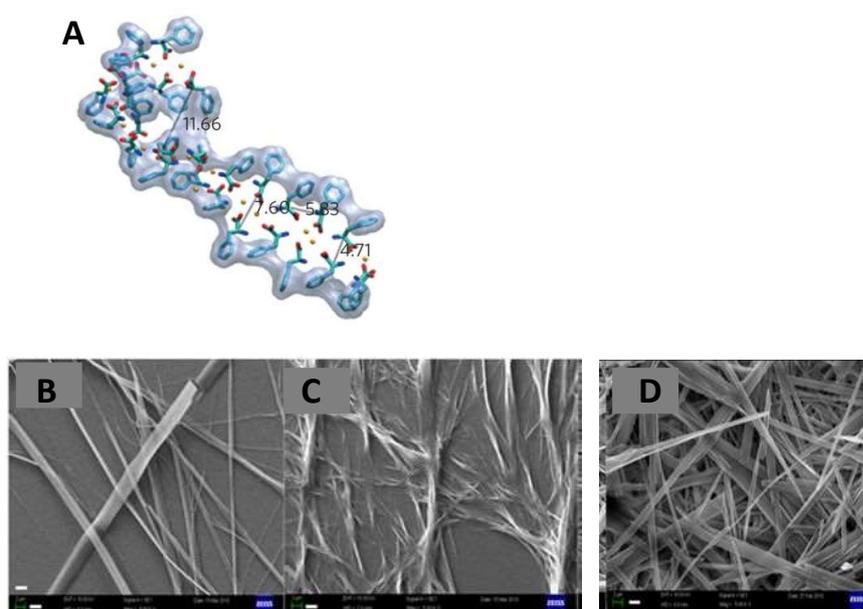


Figure 79. **Structure of phenylalanine in solution** A) Filamentous structure taken from Adler-Abramovich et al. (2012). Image obtained by molecular dynamics simulations with 27 monodisperse phenylalanine molecules (light blue) at high pH, in the presence of counterions (yellow spheres). The gray surface represents the van der Waals envelope causing tight packing of the aromatic rings. B) Scanning electron micrographs (SEM) of phenylalanine at 300 μ M, C) SEM of phenylalanine at 6mM and D) SEM of phenylalanine in 0.1M HCl. Scale bars shown in white for the SEM images are 2 μ m. Taken from Singh et al. (2014).

The PKU study data suggests that phenylalanine in solution has the potential to both interact with itself as well as with other hydrophobic amino acid residues. From the GCSF SLS data, increasing phenylalanine concentration drastically reduced aggregation onset and intensity, which suggests preferential exclusion is not occurring, as with the preferentially excluded sugars aggregation onset was delayed but changed only slightly with increasing excipient concentration. It is therefore, reasonable to assume interactions between the excipient and GCSF are occurring. Furthermore, intensity of the scatter lowered significantly at the mid and high concentrations of phenylalanine (50 – 100 mM) which is indicative of blocking of protein-protein interactions and likely occurred because of the hydrophobic interactions are occurring between phenylalanine molecules and temperature exposed hydrophobic regions of GCSF.

By HDX-MS, a large percentage of GCSF peptides showed a decrease in uptake indicating phenylalanine could be interacting with GCSF via preferential exclusion; however, there was a significant patch of peptides at the end of the protein (residues 142-175) that showed large changes in uptake relative to the control, both positively and negatively. This region also matched up with the region predicted by iGEMDOCK to be the most energetically favorable place for protein-excipient interactions. This suggests preferential interaction between GCSF and phenylalanine occurred as there is a non-uniform change in exchange. Future analysis to confirm this hypothesis could use vapour pressure osmometry as described by Schneider and Trout (2009) to determine phenylalanine's preferential parameter values with increasing concentration, as well as global conformation measurements such as CD and NMR to determine any structural perturbations.

6.2.8.3 *Sucrose (and Trehalose)*

In 1981, Lee and Timasheff (1981) established that thermodynamic stabilisation of proteins by sucrose is due to preferential exclusion of the sugar from the protein surface. By CD, the group found native conformations of α -chymotrypsin and chymotrypsinogen were not altered in the presence of preferentially excluded sucrose. Most likely sucrose did not alter the structure of these proteins, because the native conformation, in the absence of sucrose, is already representative of the most compact conformation. Kendrick et al. (1997) also showed sucrose was preferentially excluded from the surface of recombinant interleukin 1 receptor antagonist (rhIL-1ra), inhibiting global HDX and cysteine reactivity, which indicated reduction of the protein conformational flexibility and a compaction of the native state. By peptide level HDX-MS methods similar to those described in this chapter, Zhang et al. (2015) demonstrated a lack of regional interaction between 1 M sucrose and

GCSF, where in the presence of sucrose many peptides of GCSF showed small changes in deuterium exchange indicating non-specific stabilisation. Another excipient HDX-MS study by Manikwar et al. (2013) studied the effects of arginine and sucrose on local IgG1 mAb dynamics and correlated results with conformational and storage stability. They found sucrose at 0.5 M, pH 6, increased conformational stability (T_m), slowed the rate of monomer loss, reduced the formation of insoluble aggregates, and resulted in a global trend of small decreases in deuterium uptake in most peptides of the mAb. The recent Toth et al. (2018) deuterated excipient HDX-MS paper didn't include sucrose in their experiments, however they did cover another sugar, trehalose, with an IgG4 mAb where they found at 0.4 M, the sugar increased T_m and showed a global decrease in flexibility (protection) during HDX-MS at pH 6.5 – 7.4. Trehalose was shown to significantly rigidify the CH2 peptide, previously shown to be an aggregation hotspot region by Manikwar et al. (2013).

6.2.8.4 *Loop region analysis*

The data presented in this chapter further demonstrates the power of low pH HDX-MS for loop region exchange analysis. The HDX-MS excipient experiment by Zhang et al. (2015) studying sucrose and GCSF, found small conformational perturbations of GCSF occurred within the α -helices whereas there was a lack of any detectable effect on loop regions at 37 °C. Presumably, as discussed in chapter 3, at high temperature and physiological pH the loops exchange at a rate so fast they were not measurable. Only when reducing the exchange temperature to 4 °C was a measurable exchange possible for Zhang et al. (2015). This result was replicated by the GCSF and sucrose data in this chapter by lowering the solution pH rather than temperature, where not only was there was a distinct decrease in exchange in all peptides covering α -helices but peptides covering loop structures also showed a decrease in uptake. Interestingly peptides covering GCSF loopCD had the largest negative differential value of all loop peptides (Figure 80). As identified in chapter 3, loopCD is particularly dynamic and mutations to stabilise this region, as shown in chapter 4 and 5, were effective in reducing GCSF aggregation and increasing protein shelf-life. In combination the loopCD region appears to be an aggregation hotspot for GCSF due to its high flexibility, and therefore, methods to rigidify the area such as mutation and/or formulation can reduce GCSF aggregation propensity.

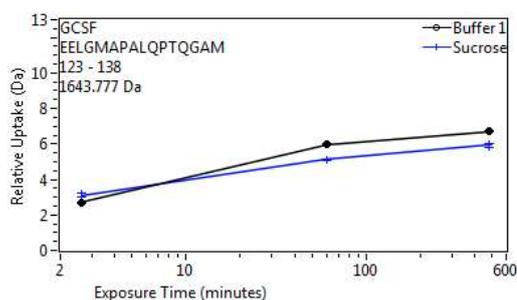


Figure 80. **Relative uptake of deuterium of GCSF loopCD peptide 123-138 over time.** The pH 4.25 uptake plot exported following DynamX 3.0 data processing, not corrected for differences in intrinsic uptake as measured by the internal reference peptide. Dark blue represents sucrose deuterium solution, and black represents control deuterium solution.

6.2.8.5 Arginine

Arginine was found to destabilise GCSF, as seen by the lowering of T_{agg} at all concentrations studied (50 - 200 mM). This was also seen in the work by Barata et al. (2016) who found inclusion of arginine had a mild detergent effect at 2% (w/v) with Fab A33 at 1 mg/mL, as well as Toth et al. (2018) who found arginine caused aggregation of an IgG4. Another study by Shal et al. (2011) compared the different in heat induced aggregation of three different model proteins: bovine serum albumin (BSA), lysozyme (LYZ), and b-lactoglobulin (BLG) when formulated with different concentrations of arginine. They found increased aggregation propensity occurred only for BSA and BLG, but not for LYZ, indicating that arginine's preferential interactions with certain residues over others determines the effect of the excipient on aggregation. The group performed density functional theory (DFT) calculations on the guanidinium group of arginine with the side chains of different residues and found the strongest interactions were with the acidic amino acids Asp and Glu. This coincided with BSA and BLG having high acidic residue content at 16.3 % and 14.6% respectively, compared to LYZ at 6.4%. GCSF contains 17 Glu and 4 Asp residues, distributed evenly across the entire protein sequence. This high level of acid residues, takes it to a similar range as the arginine aggregated proteins in the Shal et al. (2011) study, at 12.1%, therefore, it's possible the guanidinium part of arginine is forming interactions with these residues in GCSF and causing destabilisation.

Preferential interaction measurements have been performed on a number of compounds that are known to destabilise or denature proteins including urea (Prakash et al. 1981) and guanidine hydrochloride (Lee and Timasheff 1974). These compounds showed weak preferential exclusion or preferential binding, suggesting that they have a greater tendency

to bind to the proteins (Arakawa et al. 2001). It was expected that binding between GCSF and arginine would be observed by HDX-MS. From the data presented in this chapter it was found that the long labelling period of 8 hrs was necessary to observe excipient mediated changes in deuterium uptake for GCSF peptides for all excipients apart from arginine. By 8 hrs the majority of GCSF peptides showed a clear decrease in uptake with deuterated excipients, apart from with arginine, where a clear decrease in uptake for the majority of GCSF peptides with no distinct regional bias was seen at all labelling time points. This immediate protective effect of arginine with GCSF suggests that preferential binding does indeed occur and is in line with previously published data.

When comparing GCSF-arginine results with previous HDX studies using the excipient, the decrease in deuterium uptake in the majority of GCSF peptides by arginine after 8 hrs labelling was unexpected. Previous studies into the effects of arginine on an IgG1 and IgG4 mAb peptide HDX-MS, saw arginine at concentrations of 0.3 and 0.5M, cause substantial increases in uptake for specific regions of peptides (Manikwar et al. 2013; Toth et al. 2018). This increase in uptake indicated increases in backbone flexibility and destabilisation, which aligned with the mAb thermal destabilisation such as increased monomer loss and increased levels of soluble and insoluble aggregates. The discrepancy between this chapters GCSF-arginine HDX-MS data, and the mAb-arginine HDX-MS data could be due to differences in the protein molecule (small protein vs large multi-domain mAb), or the arginine concentration used (0.05 vs 0.3-0.5 M) and/or the pH used (4.25 vs 6 - 7.4). A mini review by Ishibashi et al. (2005) reported that arginine lowers the melt temperature of certain proteins, but to a low extent and is insufficient to cause denaturation of proteins at or below room temperature. This could explain why destabilisation of GCSF by 0.05 M arginine is seen by T_{agg} measurements but is not reflected in the HDX-MS data, where GCSF peptides show a reduction in deuterium uptake, indicative of stabilisation. Furthermore, investigations into protein-arginine preferential interaction coefficients performed by Schneider and Trout (2009) using both equilibrium dialysis and vapour pressure osmometry (VPO) found at concentrations < 0.5 M, arginine had a preferential interaction coefficient around zero indicating it was neither strongly bound nor excluded from the protein surface, whereas > at 0.5 M arginine became increasingly excluded. Such behavior might be indicative of the protein surface becoming saturated with arginine, thus causing any additional arginine added to the solution to be excluded from interacting with the surface (Schneider and Trout 2009). This result suggests the arginine concentration of 0.05 M used during this chapter's HDX-MS experiment may not have been sufficient to observe an

obvious change in uptake, however, it does not explain how the previous mAb studies observed such obvious destabilisation of structure with arginine where a concentration of 0.3 - 0.5 M would, according to VPO analysis, cause preferential exclusion of the excipient to occur. By SLS, at 25°C there was no difference in scatter intensity for GCSF with arginine even at the highest concentration (0.2 M) suggesting at this condition the protein is relatively stable and destabilisation by HDX-MS wouldn't be picked up, however further work is definitely required to determine why there is a discrepancy between protein-arginine HDX-MS results by different groups. Orthogonal advanced biophysical technique such as NMR has the potential to corroborate if local regions of GCSF higher order structure are perturbed by arginine in solution at 25 °C, pH 4.25. A study by Aubin et al. (2015) showed no differences in GCSF spectra (i.e. structure) when titrated with the surfactant polysorbate 80 and the excipient sorbitol. Perhaps a future study with ¹⁵N labelled GCSF and increasing concentrations of arginine could determine if low (and high) concentrations of arginine cause any changes in structure.

6.2.9 Formulation mixing equilibration

It was observed from both the SLS and HDX-MS experiments in this chapter that an equilibration window after protein and formulation mixing occurred, where the excipient altered the stability of GCSF and affected the biophysical analysis outcome. Preferential exclusion can potentially explain this equilibration effect. A two-step thermodynamic process between protein-excipient interactions in solution has been previously hypothesised by Timasheff (2002) and displayed in Figure 81. During step 1, following the mixing of protein and excipient solutions, a thermodynamically indifferent process occurs where the excipient is located in the bulk solvent as well as at the protein surface in equal concentrations. Following a certain amount of time (equilibration), step 2 occurs where protein-excipient interactions occur, and the protein's affinity for one solvent component than for the other dictates which is removed and replaced by the other solvent compound (water vs excipient). In the thermodynamically favourable process (negative free energy of binding) water is preferentially excluded and the excipient preferentially binds to the protein, whereas in the thermodynamically unfavourable (positive free energy of binding) process the protein is preferentially hydrated and the excipient is preferentially excluded.

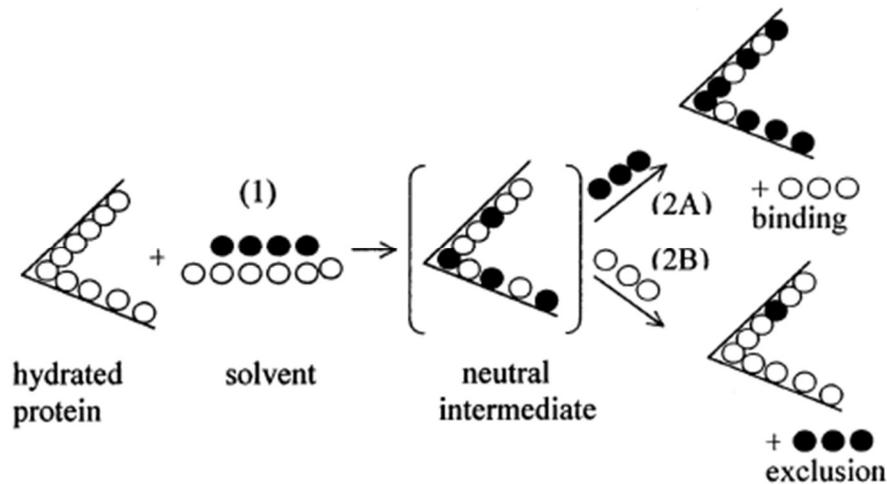


Figure 81. **Two-step preferential binding and exclusion model.** (1) Binding of cosolvent at the same composition as in the bulk solvent, thermodynamically indifferent, (2) replacement of solvent component molecules by each other A) thermodynamically favourable, B) thermodynamically unfavourable. Taken from Timasheff (2002).

During excipient HDX-MS, both stages of this thermodynamic process are observed and would explain the increase in exchange at the early labelling time points for mannitol, phenylalanine and sucrose where, upon mixing, there is a thermodynamically unfavourable interaction at the protein surface between the protein and excipient which may cause local destabilisation, but is reduced over time as the excipient is excluded from the protein surface. As it is suspected that arginine interacts with the protein surface immediately, such local destabilisation effects are not observed which could explain why arginine didn't exhibit an equilibration lag.

The issues surrounding mixing equilibration have not previously been reported by other excipient studies using biophysical analysis. Data from this chapter suggests experimental design should include post-mixing incubation time points of at least 1 hr for formulated samples prior to analysis to obtain a homogenous formulation and therefore, more exact results.

7 Mapping GCSF-excipient interactions in the solid state

With the development of a digestion and LC-MS method for HDX analysis of GCSF with high sequence coverage in chapter 3, combined with the learning of protein-excipient interactions in chapter 4, the aim of this chapter was to investigate the use of ssHDX-MS to study GCSF-excipient interactions and to link results to stability data and previous *in silico* docking.

The potential of ultra-scale down (USD) lyophilisation methods has been previously demonstrated by Grant et al. (2009) as a method of reducing material requirements and processing time, however residual moisture of USD cakes has never been accurately determined. This chapter used thermogravimetric analysis (TGA) to determine if USD cakes were suitable for formulation screenings and ultimately, ssHDX-MS experiments. Following this a screen of excipients, buffers and pH was performed with GCSF using USD methods and suitability determined by activity and monomer retention. This allowed for selection of formulations with varied levels of lyophilisation survival for ssHDX-MS characterisation.

As ssHDX-MS is currently performed using a custom built HDX manager, in the Topp lab at Purdue University (Moorthy et al. 2015), transfer of the method to the automated Waters HDX system was required. In addition, as GCSF has never been analysed using ssHDX-MS, the sample preparation and management during the ssHDX-MS process also required development.

In the previous chapter, the interaction of different excipients with GCSF in solution was probed by HDX-MS and results compared to *in silico* docking results. It was found that due to the preferential exclusion effects in solution no direct interactions between mannitol, sucrose, and phenylalanine occurred with GCSF in the native state. This meant comparisons between docking data, and HDX-MS were not possible. During lyophilisation, water is removed from samples and replaced by excipients to maintain the protein structure, directly interacting with residues on the protein surface. Therefore, computational docking has the potential to be applied for solid state formulation screening where H-bonding between protein and excipients occurs. In this chapter, excipients were co-lyophilised with GCSF and ssHDX-MS used to identify regions of interaction. Differential uptake data was compared between excipients as well as to docking results from chapter 4 to assess the predictability of interactions and stability in the solid state.

7.1 Results

7.1.1 Ultra-scale down lyophilisation validation

7.1.1.1 Residual moisture analysis by thermogravimetric analysis (TGA)

TGA was used to determine any differences in the residual moisture content of 15% (w/v) trehalose, lyophilised in glass vials and USD 96-well microtitre plate wells. A control sample of potassium tartrate was measured to ensure the system was working correctly and found to have average residual moisture content of 3.7% (w/w), which was within the specification of the control. Well samples were taken from the inner, corner and edge wells to determine differences in moisture with changing well position on the 96-well plate as shown in Table 25.

Table 25. **Categorisation for sample positions within a 96-well microtiter plate.** Black squares represent corner wells, white squares represent outer wells and grey squares represent inner wells.

	1	2	3	4	5	6	7	8	9	10	11	12
A	■											■
B		■	■	■	■	■	■	■	■	■	■	
C		■	■	■	■	■	■	■	■	■	■	
D		■	■	■	■	■	■	■	■	■	■	
E		■	■	■	■	■	■	■	■	■	■	
F		■	■	■	■	■	■	■	■	■	■	
G		■	■	■	■	■	■	■	■	■	■	
H	■											■

The average residual moisture % of glass vials and the different 96-well microtitre plate wells is shown in Figure 82. When comparing position within the 96-well microtitre plate, the residual moisture content was highest in the corner wells, and lowest in the inner wells, however all differences were within the standard deviation. The corner wells had a large standard deviation of the mean, indicating high variability, whereas the outer well residual moisture had the lowest standard deviation, indicating the least amount of variability. When comparing vials to 96-well microtitre plate wells, the average well residual moisture was $5.88 \% \pm 1.14$ whereas for the glass vial it was $2.85 \% \pm 0.41$. This was outside of the standard deviation, suggesting the two were significantly different.

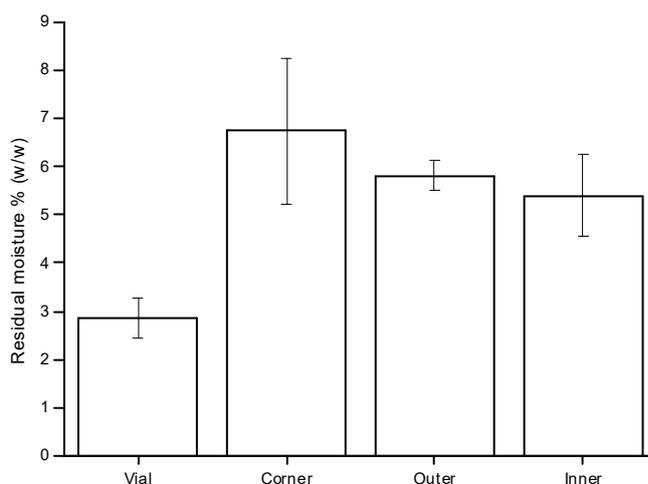


Figure 82. **Residual moisture of lyophilised cakes within glass vials and varied positions within a 96-well microtiter plates.** Measurements performed in triplicate using TGA. Positions on the 96-well microtiter plate included corners, outer wells and inner wells as depicted in Table 25

7.1.1.2 *Formulation screen*

To screen formulations for GCSF lyophilisation a 96-well microtitre plate USD screening method was used to study four different excipients in three different buffers. From the image in Figure 83 it can be determined that all samples containing phenylalanine and mannitol yielded cakes with a solid, white appearance, indicative of a successful lyophilised cake, whereas all samples containing sorbitol yielded small, collapsed cakes, where the rubber stopper above can be clearly seen. This is indicative of an unsuccessful lyophilised cake. Trehalose containing samples with sodium citrate and PBS buffers also produced robust, white cakes, whereas in sodium acetate buffer the cake collapsed. Finally, GCSF in sodium citrate and no excipients produced a solid, white cake, whereas GCSF with the two other buffers, sodium acetate and PBS, produced white, wispy cakes.



Figure 83. **Image of lyophilised GCSF formulations within a USD 96-well microtitre plate.** Samples of GCSF at 0.15 mg/mL in different formulations were lyophilised in duplicate. Columns represent the three different buffers were screened including 10 mM sodium acetate pH 4.25 (blue border) 50 mM sodium citrate pH 4.25 (orange border), and 10 mM phosphate buffered saline pH 7.4 (purple border). Rows represent excipients screened including 0.5% (w/v) trehalose (T), 3.5% (w/v) sorbitol (S), 0.1% (w/v) phenylalanine (P) and 3.5% (w/v) mannitol (M).

7.1.1.3 *Reconstitution scoring*

Samples within the 96-well microtitre plate were reconstituted with sterile H₂O to the same volume they were pre-lyophilisation. The ease of the reconstitution was assessed and scored in Table 26. For the sodium acetate samples, no formulation obtained a positive score: the trehalose, sorbitol and buffer control samples obtained neutral scores as they required pipetting to fully reconstitute, and the phenylalanine and mannitol samples did not reconstitute fully, even with pipetting, giving them a negative score. For the sodium citrate samples, the trehalose and buffer control samples reconstituted immediately upon contact with the H₂O and as such received positive scores. The sorbitol, mannitol and phenylalanine samples reconstituted after pipetting and as such received a neutral score. Finally, the PBS samples containing trehalose and buffer control reconstituted immediately, whereas the sorbitol, mannitol and phenylalanine samples required pipetting to fully reconstitute.

Table 26. **Reconstitution scores for GCSF in different lyophilised formulations.** Scoring includes immediate reconstitution (+1), reconstituted when mixed with pipette (0) and didn't reconstitute fully (-1). Letters represent buffers and excipients within the formulations: 50 mM sodium acetate pH 4.25 (A), 50 mM sodium citrate pH 4.25 (C), 10 mM PBS pH 7.4 (P), 0.5% (w/v) trehalose (T), 3.5% (w/v) sorbitol (S), 0.1 % (w/v) Phenylalanine (P) and 3.5% (w/v) Mannitol (M).

Formulation	Score	Formulation	Score	Formulation	Score
AT	0	CT	+1	PT	+1
AS	0	CS	0	PS	0
AP	-1	CP	0	PP	0
AM	-1	CM	0	PM	0
A	0	C	+1	P	+1

7.1.1.4 Monomer retention by SEC-HPLC

To assess the retention of GCSF monomer during the lyophilisation process, pre- and post-lyophilisation samples were analysed by SEC-HPLC and the % difference in monomer content calculated. The recovery of each formulation is displayed in Figure 84. Of the buffer control samples on their own, sodium citrate was the most suitable for the lyophilisation of GCSF with monomer retention of 84%. PBS recovered 44% and sodium acetate, 4%. For sodium acetate excipient formulations, sorbitol recovered the highest % of monomer at 78%. For sodium citrate, trehalose provided additional protective effects and increased recovery to 96%, whereas the other excipients did not improve on the buffer only recovery. For PBS, all excipients increased recovery. Sorbitol granted the highest recovery at 82%, followed by trehalose and then mannitol. Notably, all sorbitol formulations collapsed during lyophilisation but retained of a high level of monomer upon reconstitution.

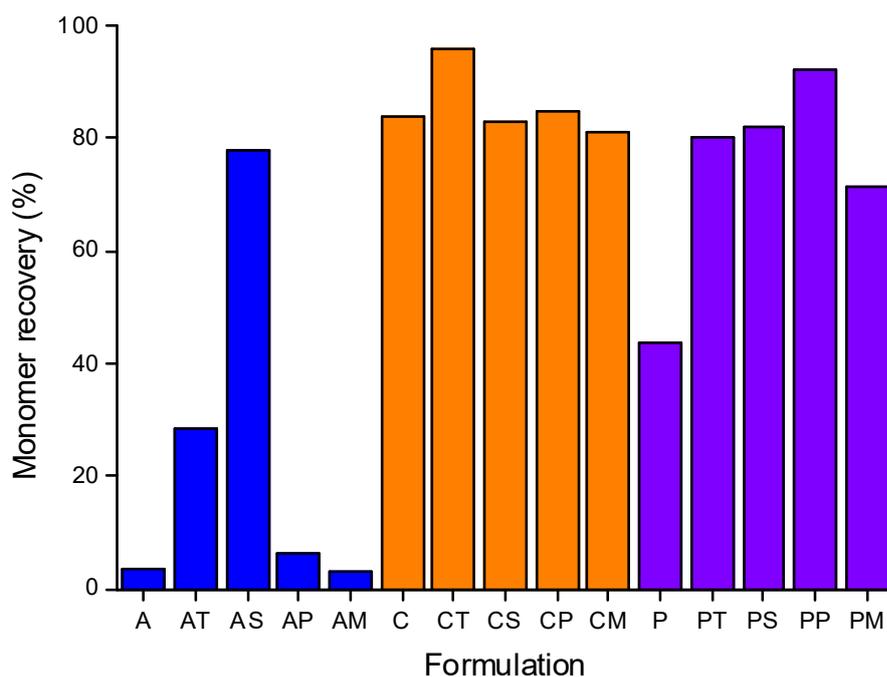


Figure 84. **SEC-HPLC monomer recovery % of GCSF lyophilised with different formulations.** Measured using SEC-HPLC peak areas of sample pre- lyophilisation and post reconstitution. Letters represent buffers and excipients within the formulations: 50mM sodium acetate pH 4.2 (A), 50 mM sodium citrate pH 4.2 (C), 10 mM PBS pH 7.4 (P), 0.5% (w/v) trehalose (T), 3.5% (w/v) sorbitol (S), 0.1 % (w/v) phenylalanine (P) and 3.5% (w/v) mannitol (M).

7.1.1.5 Activity retention by GNFS-60 cell proliferation bioassay

The activity of samples was measured using a cell proliferation bioassay. The response to serial dilution of samples pre- and post-lyophilisation is shown for each buffer and excipient combination in Figures 85-87.

For sodium acetate samples (Figure 85), a large loss of activity occurred for the control sample, as well as, with phenylalanine and mannitol following lyophilisation. A small loss of activity was also shown for the trehalose sample, whereas the inclusion of sorbitol allowed full retention of GCSF activity.

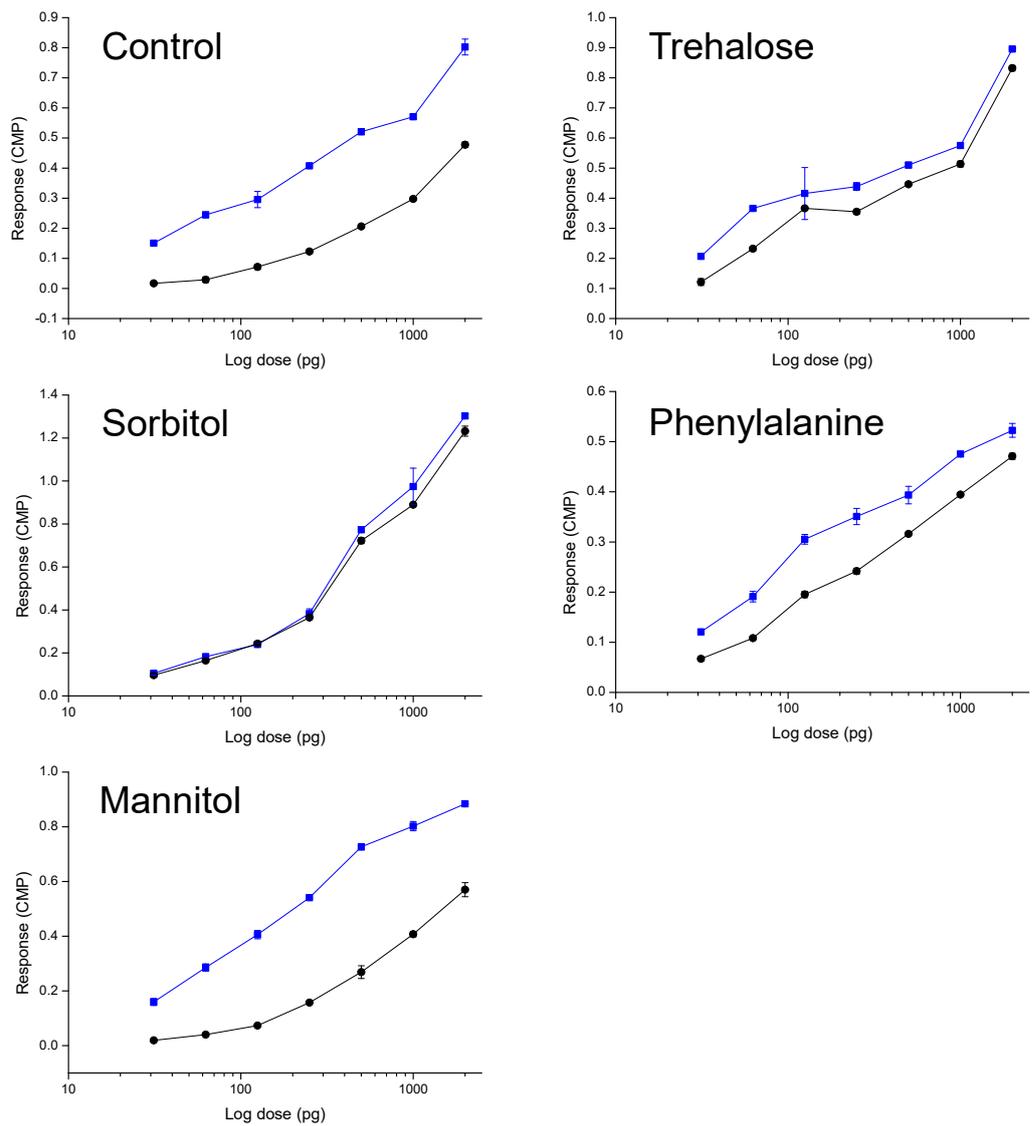


Figure 85. **Sodium acetate GNFS-60 cell proliferation bioassay.** Figures show comparisons between the same sample pre (blue) and post (black) lyophilisation. Samples contained 0.15 mg/mL GCSF formulated in 50 mM sodium acetate pH 4.25 with different excipients (as labelled).

With sodium citrate samples, no loss of activity occurred for the control sample, as such all samples containing excipients also showed no significant loss of activity following lyophilisation (Figure 86). The response of the GNFS-60 cells increased with increasing concentrations of GCSF as seen in the control sample, however, when trehalose was included there was a dip in response around the middle of the concentration range which then continued to increase with increasing dose concentration. The response dip was present in all trehalose containing samples (Figures 85-87) and was suspected to be caused by a component of the trehalose raw material, as it did not occur for control samples.

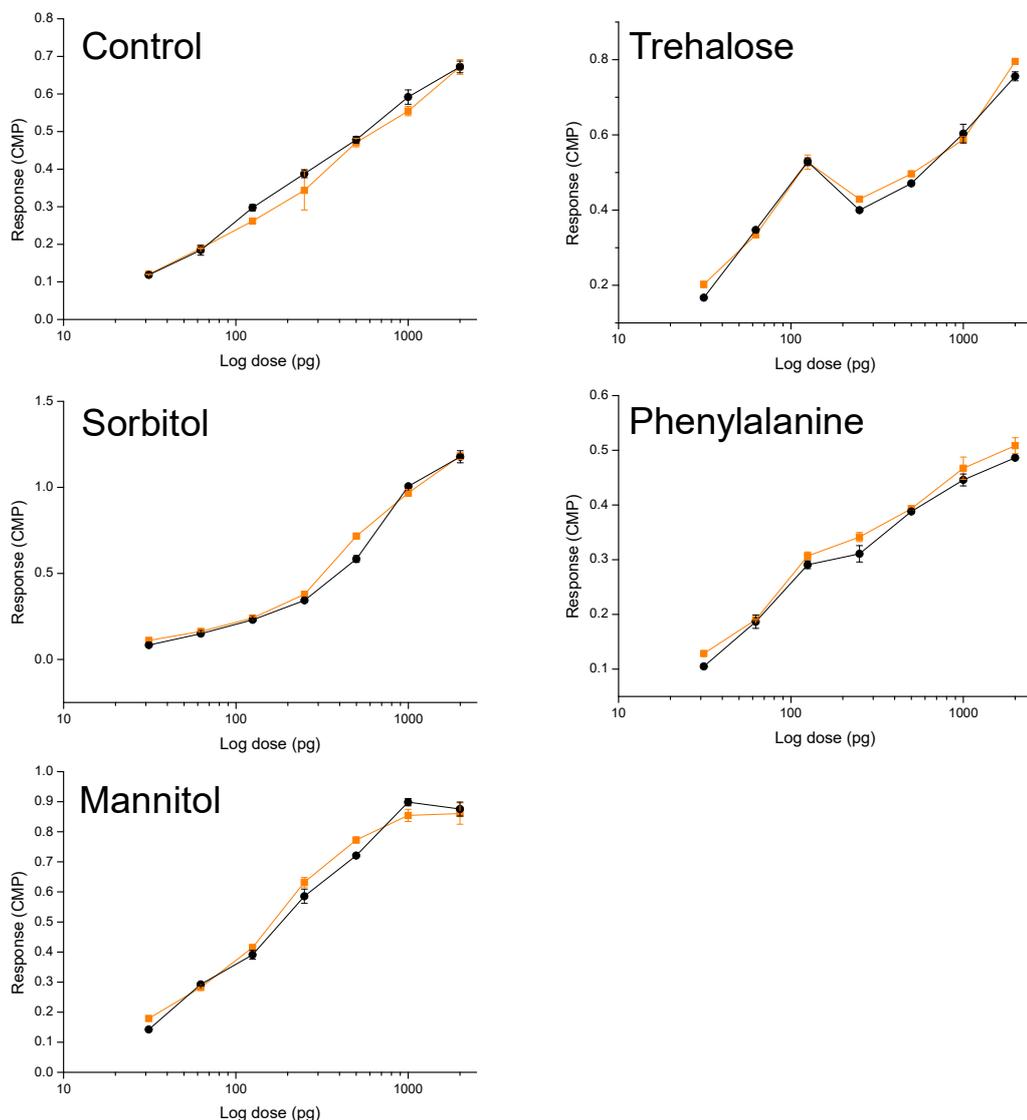


Figure 86. **Sodium citrate GNFS-60 cell proliferation bioassay.** Figures show comparisons between the same sample pre (blue) and post (black) lyophilisation. Samples contained 0.15 mg/mL GCSF formulated in 50 mM sodium citrate pH 4.25 with different excipients (as labelled).

With PBS, a loss of activity occurred for the control sample following lyophilisation (Figure 87); however, this was less prominent than that of the sodium acetate control sample (Figure 85). The inclusion of mannitol in the formulation gave the same level of activity loss as the control suggesting mannitol provided no protection during lyophilisation. The addition of phenylalanine and trehalose showed a slight loss of activity post lyophilisation, whereas the inclusion of sorbitol retained 100% of GCSF pre-lyophilisation activity. As observed with the sodium citrate samples, a dip in response around the middle of the concentration range can be seen when trehalose was included.

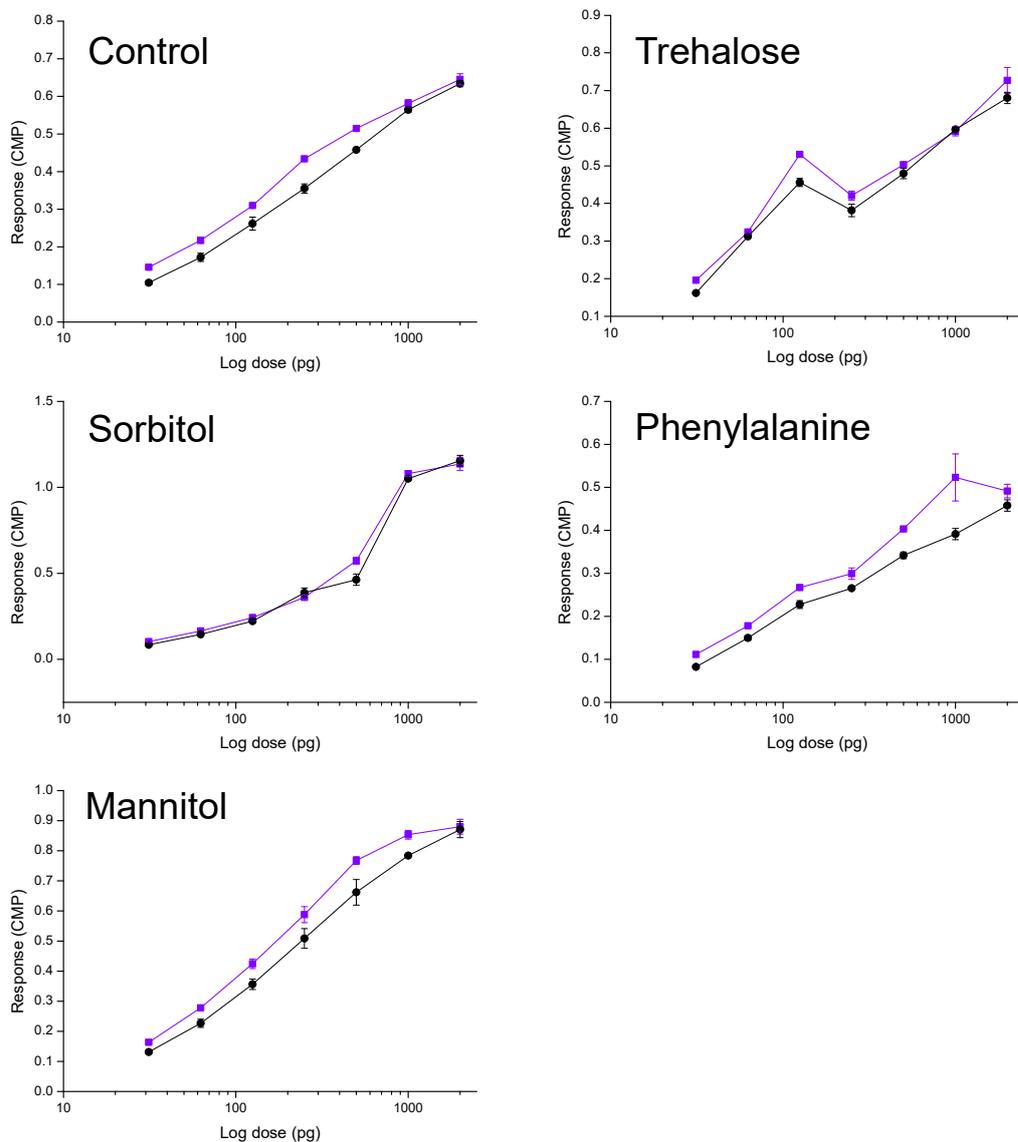


Figure 87. **PBS GNFS-60 cell proliferation bioassay.** Figures show comparisons between the same sample pre (blue) and post (black) lyophilisation. Samples contained 0.15 mg/mL GCSF formulated in 10 mM PBS pH 7.4 with different excipients (as labelled).

7.1.1.6 Determination of collapse (T_c) by freeze-drying microscopy (FDM)

Due to the difference in cake structure and retention of monomeric, active GCSF within buffers of the same pH, FDM was used to measure the collapse temperature (T_c) of GCSF in sodium acetate and citrate pH 4.25. From the images in Figure 88, the sodium acetate sample collapse occurred at an earlier temperature to that of the sodium citrate sample. The T_c of the sodium acetate sample was determined to be around $-48.2\text{ }^\circ\text{C}$, whereas for sodium citrate it was around $-37.2\text{ }^\circ\text{C}$.

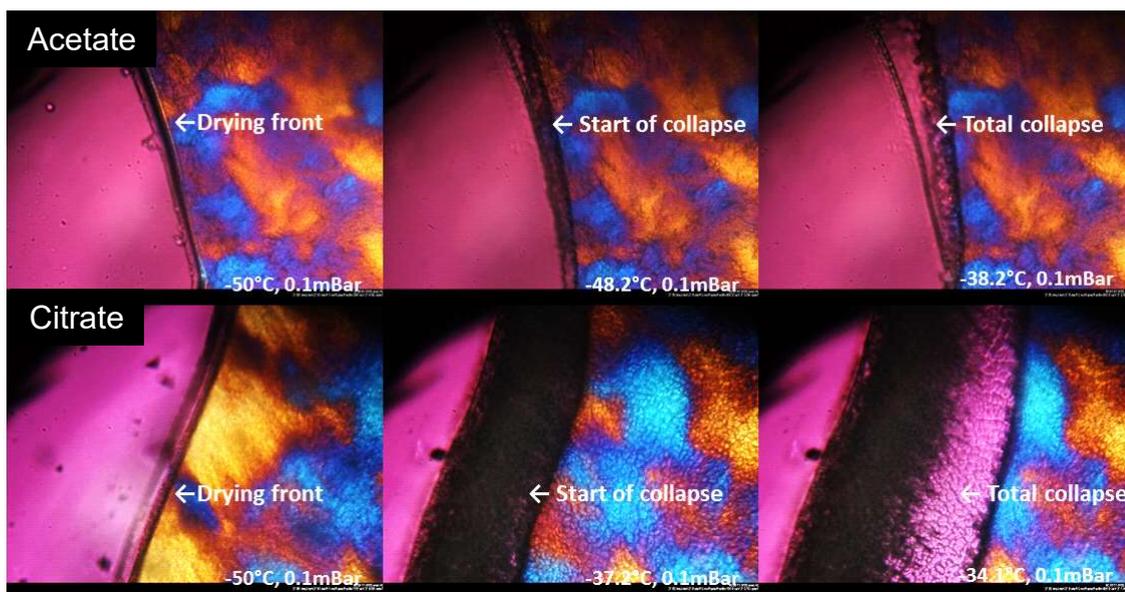


Figure 88. Freeze-drying microscope (FDM) images of GCSF in acetate and 50 mM sodium citrate. 0.3 mg/mL GCSF was formulated in either 10 mM sodium acetate pH 4.25 (acetate) or 50 mM sodium citrate pH 4.25 (citrate) and subjected to a freeze step to $-50\text{ }^\circ\text{C}$ at a ramp rate of $10\text{ }^\circ\text{C}/\text{min}$, followed by a hold for 5 mins at $-50\text{ }^\circ\text{C}$. A vacuum of 0.1 mBar was pulled for 5 mins at $-50\text{ }^\circ\text{C}$ followed by a temperature ramp of $5\text{ }^\circ\text{C}/\text{min}$ up to $25\text{ }^\circ\text{C}$. Images are shown in order of drying (left), collapse (middle) and total collapse (right).

7.1.2 GCSF-sucrose ssHDX-MS

A comparison between GCSF with and without the addition of sucrose and the IRP was performed. For this work, citric acid was selected as the buffer of choice for GCSF ssHDX-MS because as shown in the scale-down formulation screen, this buffer produced a solid lyophilised cake.

Four different samples were prepared, the first two were the control sample containing 0.3 mg/mL GCSF in 50 mM citric acid, pH 4.25 with and without the IRP, and the second two were the experimental sample containing 0.3 mg/mL GCSF in 50 mM citric acid, 1% sucrose (w/v) with and without the IRP. During preparation, samples were incubated on the bench at RT for 1 hr to ensure full equilibration following sample and formulation mixing. Samples were lyophilised in 2 mL glass vials, where all sample cakes formed during the freeze-drying

process were white, structurally sound, with slight shrink back from the vial side. Vials were removed from the freeze-dryer and placed into sealed desiccators containing D₂O and a RH of 43% for five different labelling time points: 30 min, 102 min, 240 min, 1440 min and 5760 min. Eight vials per formulation were kept un-labelled, where five were stored at -70 °C for the undeuterated HDX-MS samples and three stored at -20 °C for SEC-HPLC analysis. During the deuterium labelling process, collapse of the cakes was observed at the later labelling time points > 1440 min.

Upon removal of labelled samples, the vials were stoppered and snap frozen in liquid nitrogen to quench exchange. During snap freezing of the first-time point samples (30 min) the vials were stoppered and dropped into liquid nitrogen causing liquid nitrogen to leak into the vials and disrupt the sample. As such the 30 min samples were discarded. All other labelled vials were subsequently snap frozen by dipping the base of the stoppered vial into liquid nitrogen using large metal tongs for 5 s, where the top of the vial was not submerged and subsequently liquid nitrogen could not enter the vial.

7.1.2.1 Monomer retention by SEC-HPLC

The monomer retention of GCSF during lyophilisation between the samples with and without sucrose (not including the IRP) was determined using SEC-HPLC. Samples pre- and post-lyophilisation were analysed in triplicate and the average peak areas are displayed in Figure 89. The addition of 1 % (w/v) sucrose increased monomer recovery by around 1%.

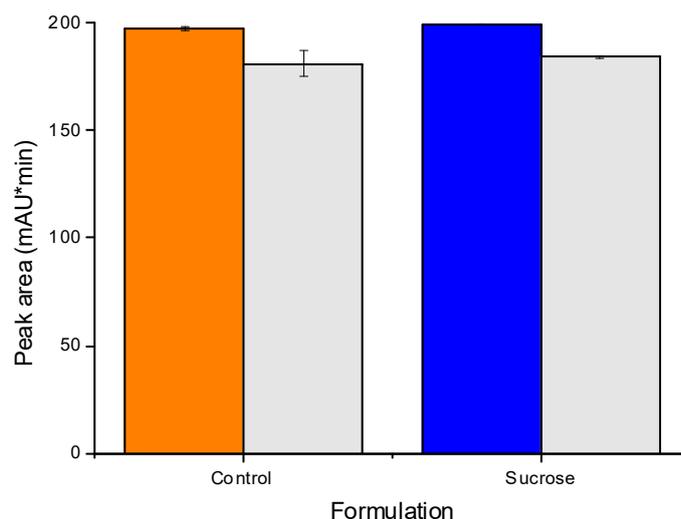


Figure 89. **Monomer content of GCSF samples with sucrose pre- and post-lyophilisation.** Measurements made of pre (coloured) and post (light grey) lyophilisation samples using SEC-HPLC. Control sample contained GCSF in 50 mM citric acid pH 4.25, Sucrose sample contained GCSF in 50 mM citric acid pH 4.25 and 1% sucrose (w/v). Measurements made from three vials per sample and ± 1 standard deviation illustrated.

7.1.2.2 Peptide map coverage

Three injections of reconstituted non-deuterated control sample were used for the peptide map, where the sequence coverage was 97.7% (Figure 90).

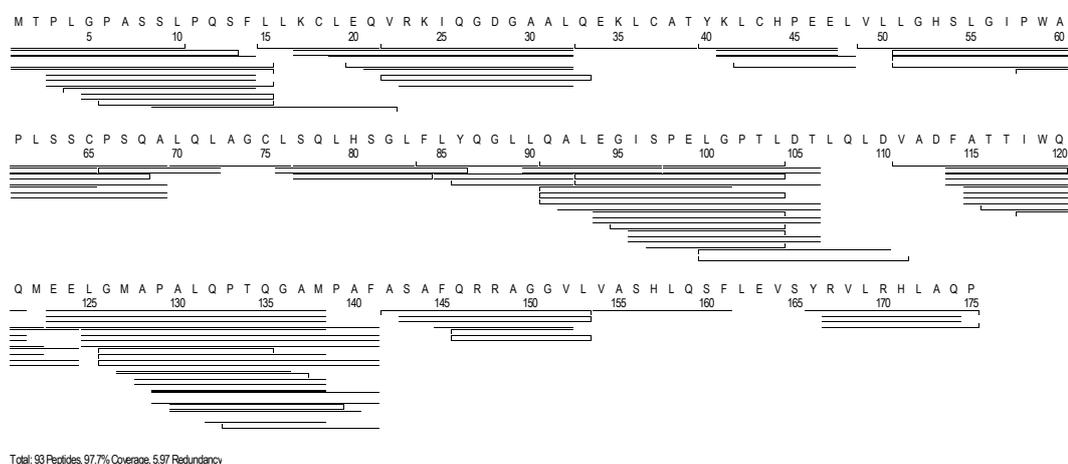


Figure 90. **Peptide map of reconstituted GCSF in 50 mM citric acid pH 4.25.** GCSF at 0.3 mg/mL was freeze-dried and reconstituted in 0.2% (w/w) formic acid at a 10x dilution. The peptide map was generated from 5 injections of the sample into the HDX-MS system following 1:1 mixing in quench solution. Peptides were identified using Waters PLGS excluding pepsin, with a minimum product per amino acid of 0.03, a maximum MH⁺ Error (ppm) of 20 and a file threshold of 3.

7.1.2.3 Sample reconstitution

All labelled samples, 2 hr, 4 hr and 24 hrs, were analysed using HDX-MS in triplicate. Samples containing the IRP were also labelled for 96 hrs in duplicate. A sample was removed from dry ice, hand thawed for 10 seconds, and reconstituted in ice cold 0.1 % (v/v) formic acid, followed by vortexing to ensure complete reconstitution. Vortexing for 10 s was found to be optimal for full sample reconstitution. The sample was mixed with ice cold quench solution within the LEAP PAL™ sampler and the HDX manager instructed to inject the sample into the LC-MS. During this study, the average time taken from removal of sample from dry ice to inject was 1 min 57 s ± 0.004.

7.1.2.4 Internal reference peptide (IRP)

The IRP was included to monitor any changes in intrinsic deuterium uptake with the inclusion of sucrose in the lyophilised cake (Table 27). For all labelling time points there was a lower level of relative deuterium uptake with the addition of sucrose. At the shorter labelling time points, 120 min and 240 min, the difference between the control and sucrose sample uptake was more prominent. The uptake of both samples plateaued around 1440 min, which coincided with the visual collapse of the lyophilised cakes.

Table 27. **Relative deuterium uptake for an internal reference peptide with sucrose solid state formulations.** Differential calculated as excipient uptake minus control uptake.

State	Exposure (min)	Uptake (Da)	Uptake SD	Differential (Da)	Differential SD
Control	120	0.261	0.004		
	240	0.648	0.003		
	1440	0.817	0.015		
	5760	0.813	0.007		
Sucrose	120	0.202	0.010	-0.059	0.014
	240	0.578	0.009	-0.070	0.012
	1440	0.782	0.011	-0.034	0.026
	5760	0.792	0.006	-0.021	0.013

7.1.2.5 HDX-MS

The differential relative uptake for GCSF peptides with and without 1% sucrose is plotted in Figure 91 for the three labelling time points 2 hrs (120 min), 4 hrs (240 min) and 24 hrs (1440 min). The 96 hr (5760 min) sample was discounted from data analysis as it showed similar values to the 24 hr sample due to the plateauing of uptake as seen with the reference peptide. As the labelling time increased, the variability of peptide uptake increased and the clarity of results decreased. By 1440 min there was minimal difference in uptake between the control and sucrose samples, which coincided with the visual collapse of the samples.

From the earliest labelling time point, 120 min, the inclusion of sucrose in the solid state greatly reduced uptake rate of specific regions of GCSF peptides, and slightly increased the uptake rate in others. The peptides with small increases in uptake include mostly α -helix regions, namely α A and α B, whereas the peptides with large negative differentials spanned the majority of loop regions apart from LoopD. When comparing changes in uptake for α -helix regions, α A and α C contained a larger number of peptides with a negative differential at 120 min than α B and α D. The short helix (sh) in the middle of LoopAB was not displayed in the figure due to the short sequence length; however peptides covering LoopAB (and the short helix) showed the second lowest level of negative differential values after LoopCD. LoopA and LoopBC showed moderate uptake protection and LoopD showed little protection.

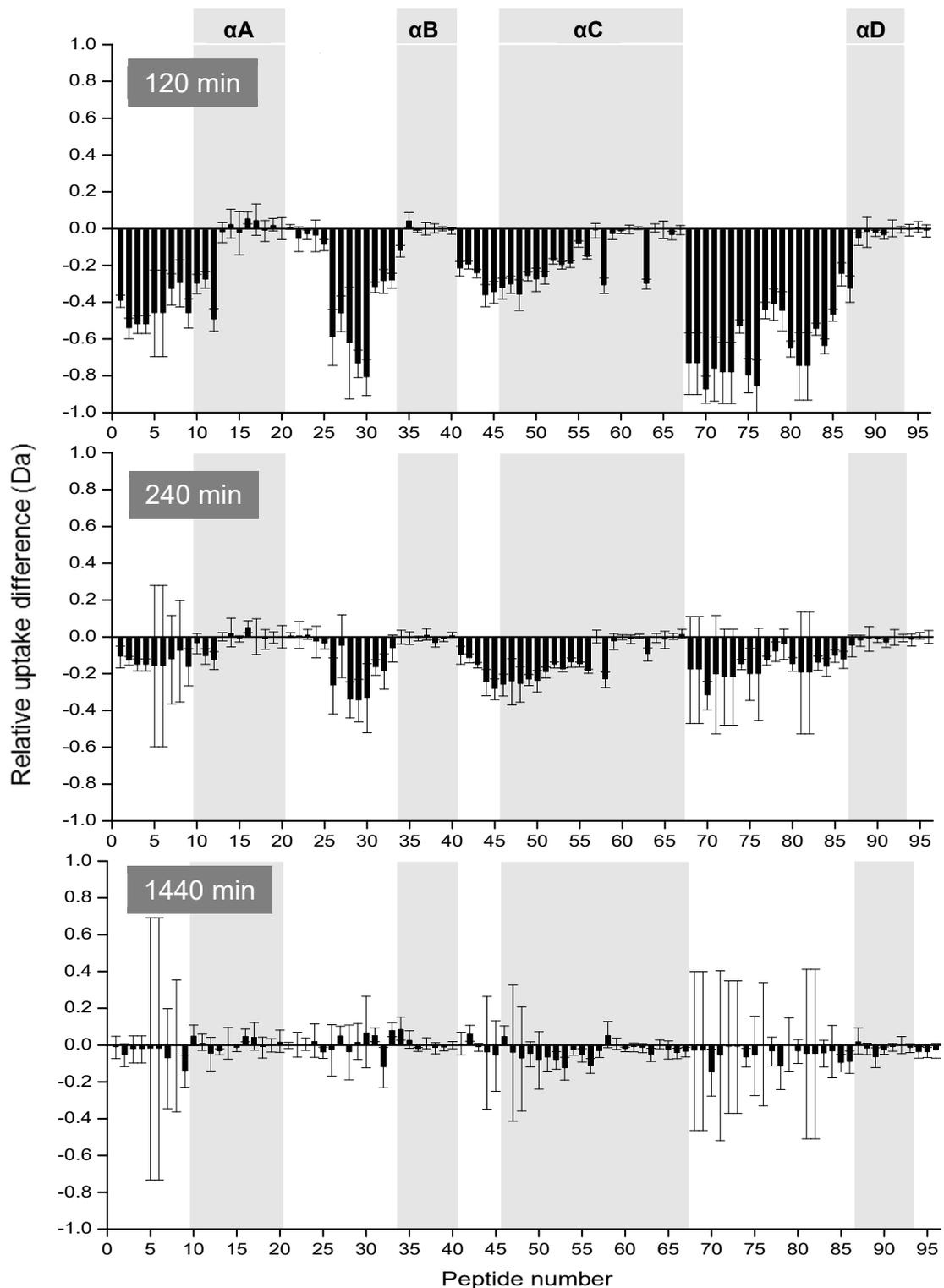


Figure 91. **Differential plots of lyophilised GCSF uptake with and without sucrose, as measured by ssHDX-MS.** Labelling time points are displayed in a grey box at the top left corner of each figure. The y -axis denotes the relative uptake calculated from the change in mass between the undeuterated and deuterated peptides. The x -axis denotes the ordinal peptide number, a sequential arrangement of 96 peptides of GCSF by the midpoints in the sequence. The different helical regions of GCSF are coloured in the background. The non-coloured regions represent the connecting loop regions. Appendix 14 contains the GCSF peptide sequence, residue numbers and locations.

7.1.2.6 *Effect of IRP on ssHDX-MS*

The differential uptake for control sample GCSF peptides with and without the IRP was plotted in Figure 92 for the three labelling time points. The addition of the peptide caused a general increase in uptake of deuterium, as seen by a positive differential, in specific regions of GCSF. The regions affected appeared to be similar to the regions that had a negative differential with sucrose. As labelling time increased the effect of the IRP on uptake were reduced, but the same general pattern was present. Increased variability between values at the later labelling time points also occurred, as seen with the sucrose ssHDX-MS data in Figure 91.

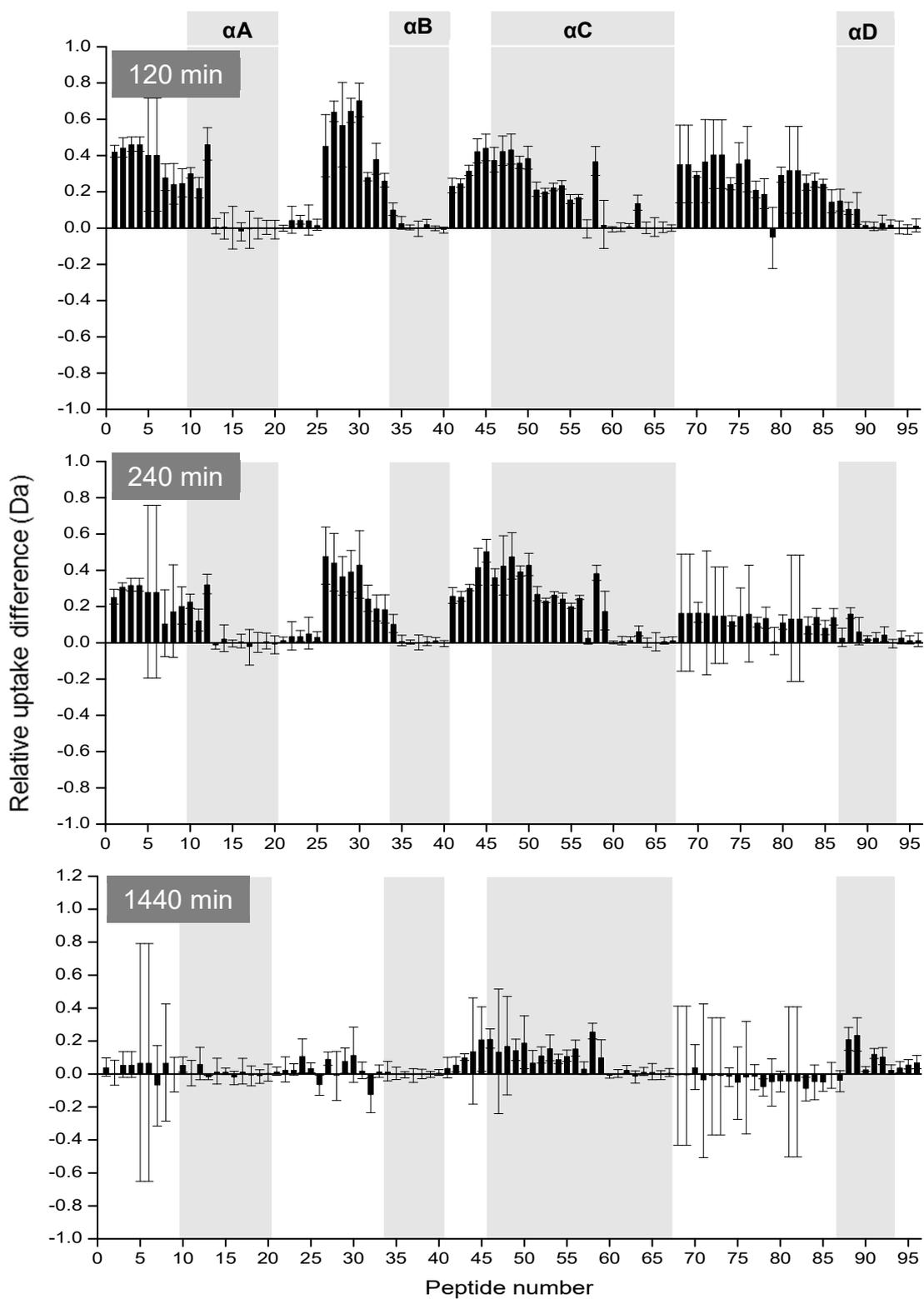


Figure 92. **Differential plots of lyophilised GCSF peptide uptake with and without the internal reference peptide, as measured by ssHDX-MS** Labelling time points are displayed in a grey box at the top left corner of each figure. The *y*-axis denotes the relative uptake calculated from the change in mass between the undeuterated and deuterated peptides. The *x*-axis denotes the ordinal peptide number, a sequential arrangement of 96 peptides of GCSF by the midpoints in the sequence. The different helical regions of GCSF are coloured in the background. The non-coloured regions represent the connecting loop regions. Appendix 14 contains the GCSF peptide sequence, residue numbers and locations..

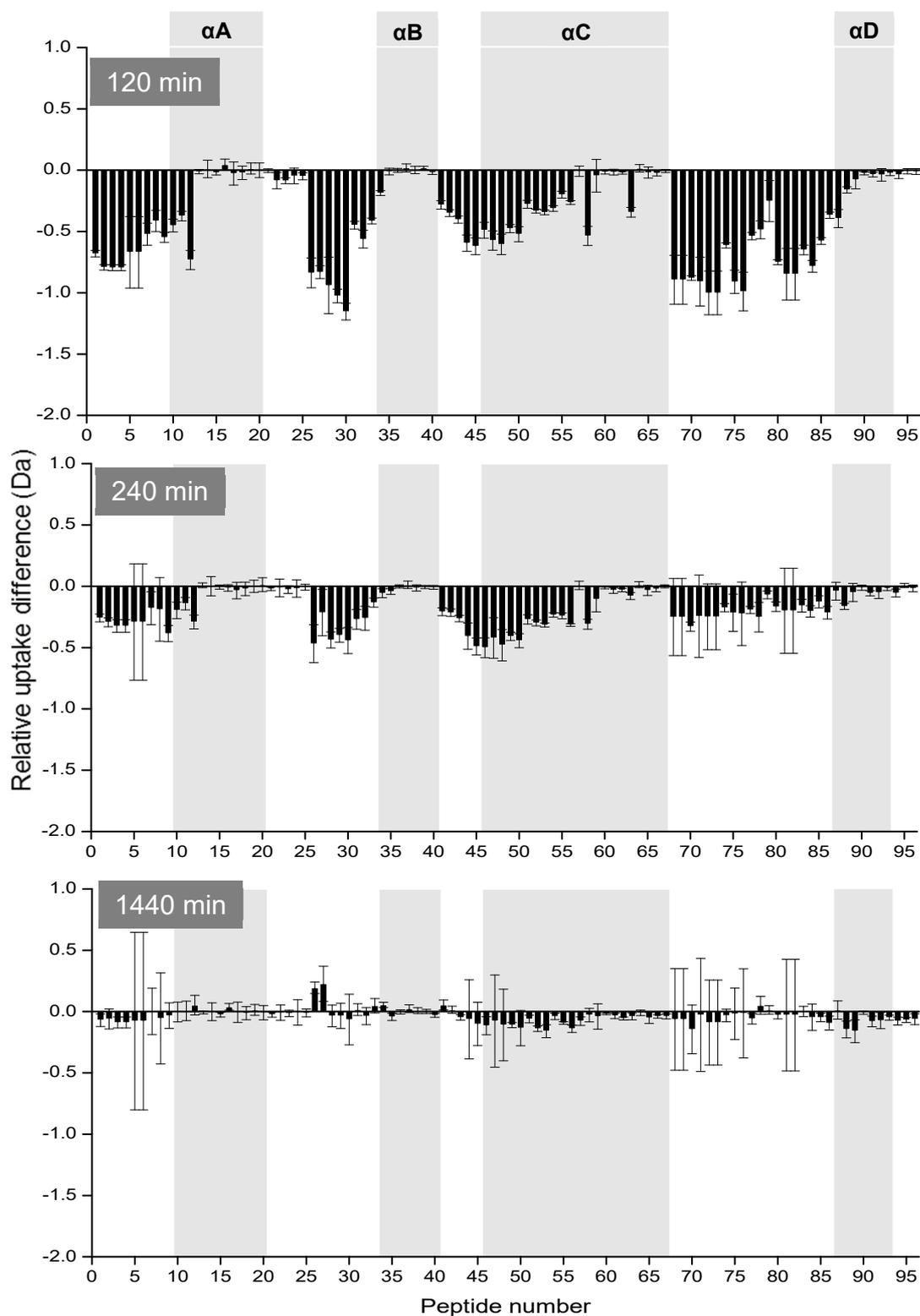


Figure 93. **Differential plots of lyophilised GCSF and uptake with and without sucrose and including the internal reference peptide in both samples, as measured by ssHDX-MS.** Labelling time points are displayed in a grey box at the top left corner of each figure. The y-axis denotes the relative uptake calculated from the change in mass between the undeuterated and deuterated peptides. The x-axis denotes the ordinal peptide number, a sequential arrangement of 96 peptides of GCSF by the midpoints in the sequence. The different helical regions of GCSF are coloured in the background. The non-coloured regions represent the connecting loop regions. Appendix 14 contains the GCSF peptide sequence, residue numbers and locations.

To determine the extent of changes to uptake when including the IRP in samples, the differential uptake for GCSF peptides between the control sample and the sucrose sample, both with IRP included was plotted in Figure 93 for the three labelling time points. It was found that the results were very similar to the differential plots comparing the effects of sucrose without the IRP (Figure 91). To determine how similar the results were between Figure 91 and Figure 93, a linear correlation was fit between the 120 min labelling data sets (Figure 94), where an adjusted R^2 value (Adj. R-Square) of 0.969 was obtained. This showed values strongly correlated, but as seen from the different axis ranges, values started fractionally lower for the IRP containing peptides. This is interesting as the IRP was shown to increase uptake with the control sample, so when including sucrose, it was expected to reduce effects of sucrose on uptake protection, whereas it provided greater potential for sucrose to show protective effects. There was also a peptide outlier, highlighted in Figure 94 (number 79, residues 128 to 138), which had a lower differential value with the peptide relative to the differential value without the peptide.

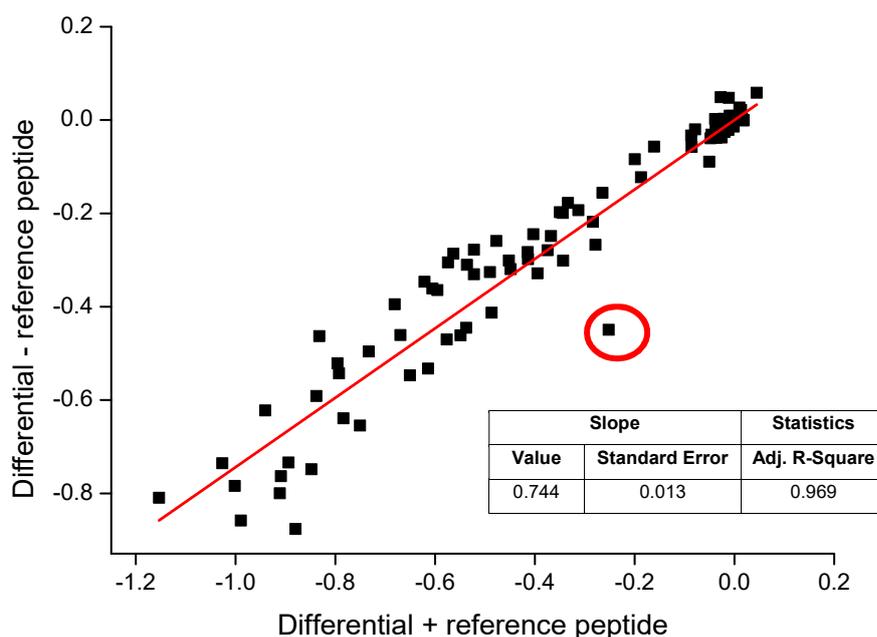


Figure 94 **Correlation between differential sucrose uptake values, with and without the internal reference peptide.** Incubation time of 120 mins used. Linear trend line coloured in red with Origin Pro calculated trend line statistics displayed with the adjoining table. Outlier circled in red denotes peptide number 79, residues 128-138.

7.1.3 USD moisture induced cake collapse study

Samples of 0.3 mg/ mL GCSF, 50 mM citric acid pH 4.25, containing four different concentrations of sucrose ranging from 0 % to 10 % (w/v) were prepared to determine if increasing concentration could reduce moisture induced collapse of solid state samples. The USD method was applied for the study where 50 µL of samples were lyophilised in 1.4 mL Micronic tubes in a 96-well format. The Micronic tubes were placed in 43 % RH desiccators for 96 hrs and the structural integrity of the lyophilised cakes monitored. All formulations formed solid white cakes that filled the base of the tubes, but by 96 hrs all samples had collapsed into small white lumps at the bottom of the tubes. This indicated that concentration of sucrose is not the cause of the collapse of lyophilised cakes during labelling.

7.1.4 USD excipient screen with internal reference peptide

To test the influence of different excipients on cake collapse, a USD cake collapse study was performed using different lyophilised excipients in place of sucrose and compared to a no-excipient control. Excipients selected included mannitol, arginine, glycine, and phenylalanine and all were included at 1% (w/v) in 50 mM citric acid pH 4.25. The effect on the uptake rate of deuterium with these excipients was also measured by including the IRP. Following a 1 hr, RT equilibration post-mixing, 50 µL of samples were added to 1.4 mL Micronic tubes with plug style TPE caps placed on top and lyophilised. Although mannitol was included in formulations, an anneal step was not included in the cycle. At the end of the cycle the chamber was backfilled with nitrogen, however for ease of labelling set up the tubes were not stoppered. All excipients and the buffer control formulations all formed neat white cakes at the bottom of the tubes. To reduce moisture sorption, the tubes were immediately placed around the edges of three empty 200 µL pipette tip holders. Samples were labelled at 43% RH, for three time points: 30 min, 2 hr and 4 hrs. The cakes were still present after 30 mins of 43% RH D₂O labelling; however, by 2 hrs and 4 hrs only the phenylalanine containing sample had retained its structure. The control, mannitol and arginine cakes had shrunk down to small white lumps at the bottom of the tubes and the glycine cake was not visible. During reconstitution in 0.2% FA, the mannitol and arginine samples were difficult to dissolve and required 15-20 s of vortexing rather than the sucrose optimised 10 s. The average time from removal of a sample from dry ice to injection into the LC-MS was 1 min 53 s ± 0.006.

7.1.4.1 Effect of excipients on relative deuterium uptake rate

The relative deuterium uptake of the IRP with different excipients is shown in Figure 95 alongside the no-excipient control. At the earliest labelling time point (30 min), the uptake of D₂O was low for all samples apart from glycine. At the second labelling time point (2 hrs) the mannitol and phenylalanine samples were within the error of the control, whereas the arginine and glycine samples had significantly lower exchange than the control. Finally, at the longest labelling time point (4 hrs) the mannitol and phenylalanine samples again followed the same uptake rate as the control, whereas the arginine sample remained lower than the control but increased in uptake at the same rate. The glycine sample did not increase in uptake and remained at the same level as the 2 hr time point. From cake appearance it suggested glycine took up moisture quicker than the others hence the increase in exchange and early plateau. All other excipients, although their cakes collapsed, still increased in uptake over time.

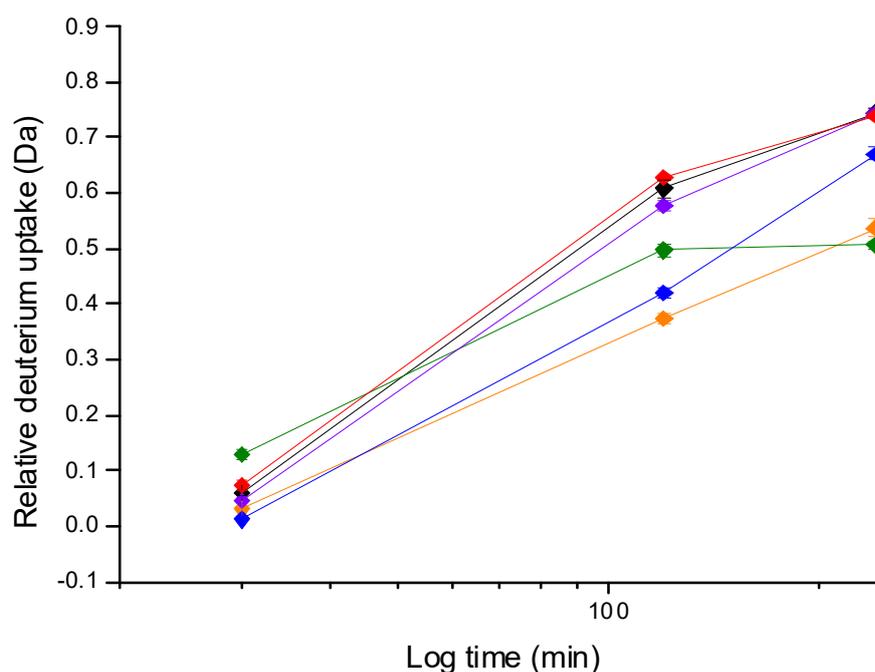


Figure 95. **Relative deuterium uptake for internal reference peptide lyophilised with different excipients.** Samples labelled for three different time points: 2 hours, 4 hours and 24 hrs. All samples were formulated in 50 mM citric acid pH 4.25. Excipients studied included arginine (orange), glycine (green), mannitol (red), phenylalanine (purple), sucrose (blue) and no-excipient control (black).

7.1.4.2 Differential scanning calorimetry

To further characterize the different excipient formulations, DSC was used to determine the transition temperatures. The T_g' and crystallisation of samples was determined by three different measurements: heat flow, reversed heat flow, non-reversed heat flow (Figure 96). All samples apart from phenylalanine had a measurable T_g' . The control sample had a T_g' value of 30-35 °C, and with the addition of 1% (w/v) arginine and glycine the T_g' value decreased by 10 °C. It was expected that mannitol would show a crystallisation dip in the reversed heat flow which did not occur; instead mannitol had a T_g' reading around 5 °C lower than the control. Phenylalanine showed a single crystallisation peak in the reverse heat flow at -12.05 °C.

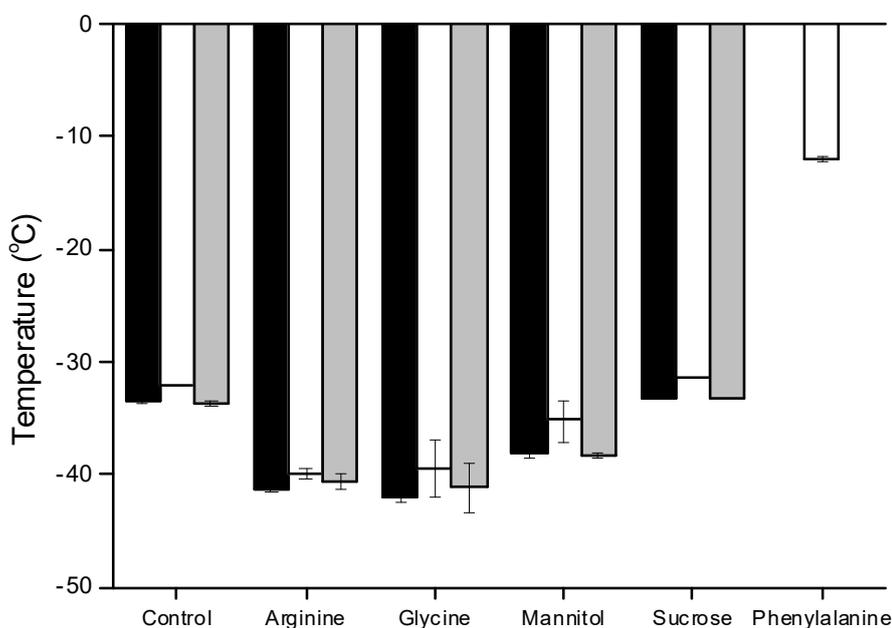


Figure 96. **DSC of 50 mM citric acid with different excipients.** Samples were analysed in triplicate and thermal events were determined using three different measurements: heat flow (black), reversed heatflow (white) and non-reversed heatflow (grey).

7.1.4.3 FDM to determine sample collapse temperatures (T_c)

FDM was used to visualize drying front progression and eventual collapse of selected excipient formulations compared to the no-excipient control. During each FDM run 10 s of images were captured, however only three main images are displayed for each in Figure 97. The first image shows the frozen sample held for 10 mins at $-50\text{ }^\circ\text{C}$, with a 0.1 mBar vacuum applied after 5 mins, the second shows the start of collapse (T_c) and the third shows total collapse. The drying front seen in the first image of each sample as a dark grey band that starts at the edge and moves across the sample, varies in size and rate of progression. For the citric acid control, the drying front was wispy with large gaps between frozen crystals, as seen by large spots of pink (Figure 97). The front progressed until collapse occurred around $-33\text{ }^\circ\text{C}$, which coincided with the DSC T_g' at -30 to $-35\text{ }^\circ\text{C}$. The inclusion of 1% (w/v) mannitol produced a thicker and darker drying front, which progressed at a slower rate than the control sample. The T_c was also lower at $-35.6\text{ }^\circ\text{C}$. This temperature also coincided with the reverse heat flow value from the DSC however; the heat flow and non-reversed heat flow T_g' had a lower temperature around $-40\text{ }^\circ\text{C}$. For 1% (w/v) phenylalanine the drying front was the thickest, with no visible pink gaps. The drying front progressed the longest with phenylalanine and eventually collapsed at $-19\text{ }^\circ\text{C}$. This low temperature collapse coincided with the DSC crystallisation event around $-12\text{ }^\circ\text{C}$. The addition of 1% sucrose (w/v) decreased the T_c slightly compared to the control, with $+2\text{ }^\circ\text{C}$, whereas by DSC the sucrose sample had the same T_g' as the control, suggesting the excipient has little effect on thermal characteristics of the formulation.

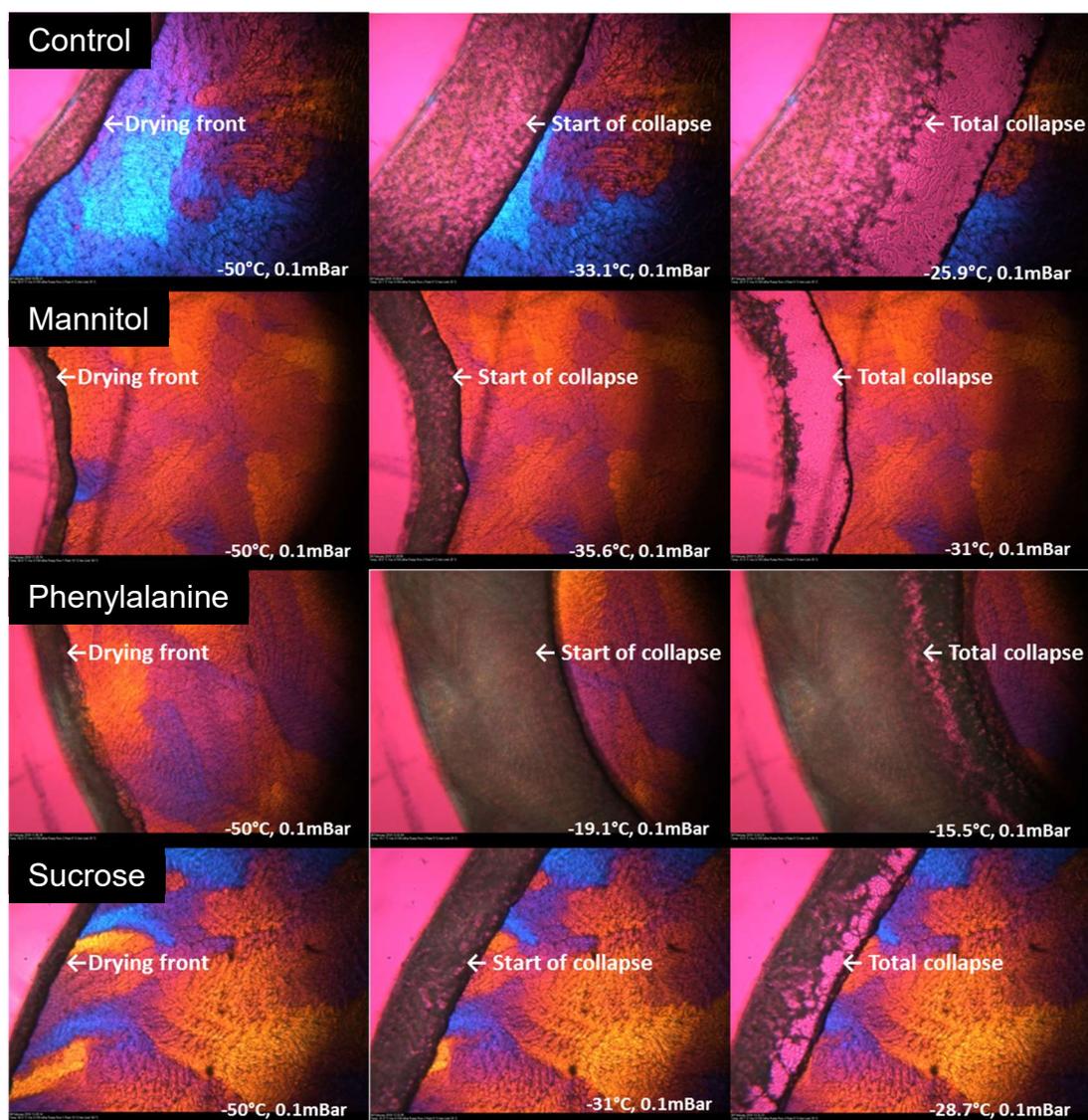


Figure 97. Freeze-drying microscope (FDM) images of GCSF with different excipients. 0.3 mg/mL GCSF was formulated in either 50 mM citric acid (control) along with either mannitol, phenylalanine or sucrose at 1% (w/v) and subjected to a freeze step to -50 °C at a ramp rate of 10 °C/ min, followed by a hold for 5 mins at -50 °C. A vacuum of 0.1 mBar was pulled for 5 mins at -50 °C followed by a temperature ramp of 5 °C/min up to 25 °C. Images are shown in order of drying (left), collapse (middle) and total collapse (right).

7.1.5 ssHDX-MS to explore the protective interactions of mannitol and phenylalanine

Excipients mannitol and phenylalanine were taken forward for ssHDX-MS analysis to determine the interactions of the molecules with GCSF in the solid state. Following the exact same methodology as the sucrose ssHDX-MS experiment in 7.1.2, three different formulations of GCSF were prepared. The first was the control sample containing 0.3 mg/mL GCSF in 50 mM citric acid, pH 4.25, and the second contained an additional 1% mannitol (w/v) and the third 1% phenylalanine (w/v). The samples were co-lyophilised with

GCSF alongside a 1 % (w/v) sucrose sample for visual comparison to the previous ssHDX-MS experiment. The cakes formed during lyophilisation are shown in Figure 98A and were solid and white. The appearance of the cakes was monitored at the different labelling time points as shown in Figure 98B-E. The control, mannitol and sucrose cakes shrank away from the edges of the vials meaning that the cake became loose (although still intact) during handling. The phenylalanine cake was cracked along the top but did not shrink away from the edges and remained fixed to the bottom of the vial during handling. The only change to the sucrose ssHDX-MS method was that the maximum labelling time in the 43% RH D₂O desiccators was 1400 min. After 30 and 60 mins of labelling in the 43% RH desiccators, all cakes were visually unchanged, whereas after 120 mins the mannitol sample had started to collapse. By 420 mins, the control and sucrose cakes also began to collapse, and the mannitol cake remained as it had been at 120 mins. By 1440 mins the control and sucrose cakes had completely disappeared and the mannitol cake was as it was at 420 mins hrs. The phenylalanine containing cake did not change throughout labelling.

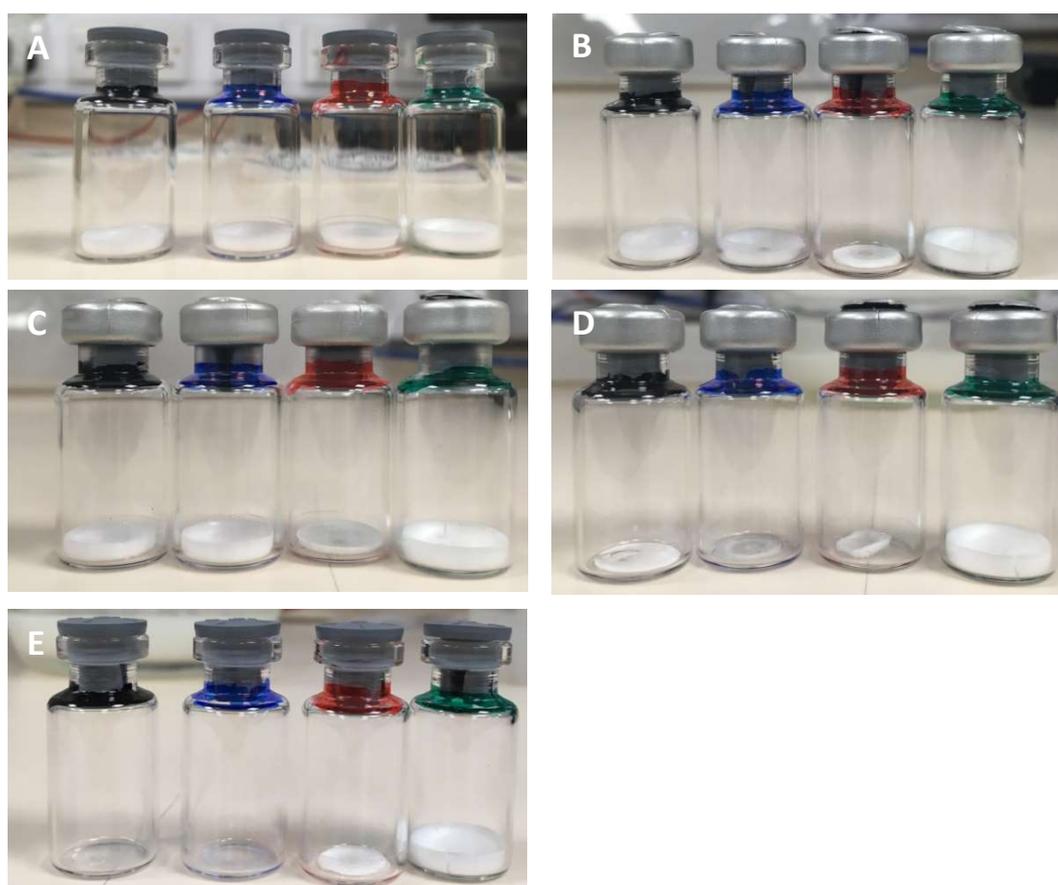


Figure 98. **Freeze-dried cake appearance during labelling with deuterium using a desiccator at 43% RH.** Neck colour of vials correspond to the excipient included within a sample of 0.3 mg/mL GCSF in 50mM citric acid buffer pH 4.25: black contains no excipient, blue contains 1% sucrose (w/v), red contains 1% mannitol (w/v) and green contains 1% phenylalanine (w/v). Letters represent labelling times: A) 0 s, B) 30 min, C) 1 hr, D) 2 hr, E) 4 hr

7.1.5.1 Monomer retention by SEC-HPLC

The retention of GCSF monomer during lyophilisation with both mannitol and phenylalanine compared to the no-excipient control was determined using SEC-HPLC. Samples pre- and post- lyophilisation were analysed in triplicate and the average peak area displayed in Figure 99. The addition of mannitol was found to increase recovery by 2%, indicating it provided increased protection during lyophilisation, whereas the addition of phenylalanine decreased monomer recovery compared to the control by 9%, suggesting phenylalanine destabilised GCSF during lyophilisation.

7.1.5.2 Reconstitution

Samples were defrosted, reconstituted and mixed with quench solution as developed in the sucrose ssHDX-MS experiment, where the average time taken from defrosting to LC-MS was 1 min 52 s \pm 0.007.

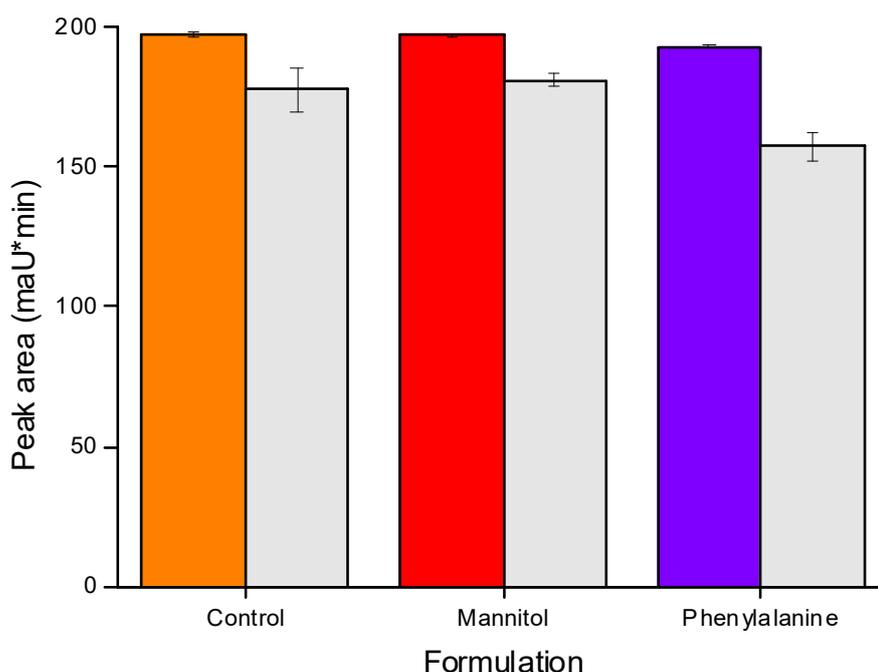


Figure 99. **Monomer content of GCSF with mannitol and phenylalanine pre- and post- lyophilisation.** Measurements made of pre (coloured) and post (light grey) lyophilisation samples using SEC-HPLC. Control sample contained GCSF in 50 mM citric acid pH 4.25, excipient samples contained GCSF in 50 mM citric acid pH 4.25 and either 1% (w/v) mannitol or phenylalanine (w/v). Measurements made from three vials per sample and standard deviation included.

7.1.6 GCSF-mannitol ssHDX-MS

The differential uptake values for GCSF peptides in the solid state with 1% (w/v) mannitol compared to the no excipient control are shown in Figure 100. During labelling, the control sample collapsed at 420 mins, therefore, only the time points prior to this event were used for analysis. From Figure 100 there was a clear reduction in deuterium uptake after 30 mins, in the regions seen previously to be protected by sucrose including all loop regions, apart from loopD and the start of most α -helices. As the labelling time increased the differential values decreased, and after 240 mins the uptake for the mannitol sample surpassed the control, as seen by positive differentials for most peptides (Figure 100). The level of protection imparted by mannitol was not as great as that shown by sucrose, especially in the loopA and loopAB regions.

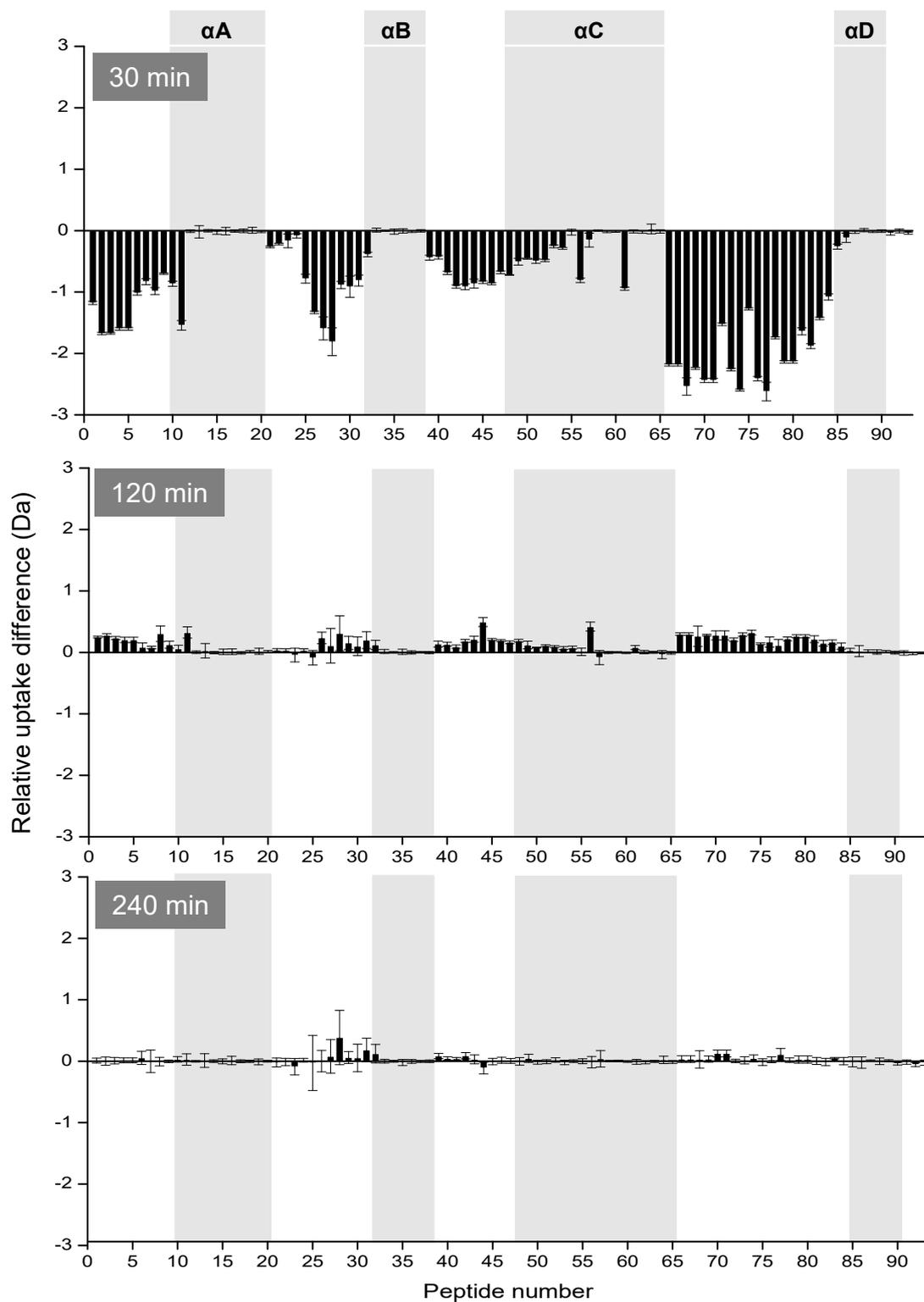


Figure 100. **Differential uptake plots for GCSF lyophilised with mannitol as measured by ssHDX-MS.** Negative values indicate a reduction in uptake in the presence of the excipient. The peptide number in x-axis represents the 93 overlapping peptides obtained from pepsin digestion. The helical locations are denoted at the top of the figure, and time points are displayed in a grey box at the top left corner of each figure. Amino acid sequence and the corresponding peptide numbers are provided in Appendix 17

7.1.7 GCSF-phenylalanine ssHDX-MS

The differential uptake values between GCSF peptides in the solid state, with and without 1% phenylalanine are shown in Figure 101. As seen with mannitol, there was a large decrease in uptake for phenylalanine peptides after 30 min labelling for loop and end of helix regions. At 420 mins, protective effects in most regions were still visible, but LoopCD (region between α C and α D) significantly increased in uptake as seen by the positive differential values. As with mannitol, the differential values of loop regions was not as low as sucrose suggesting that sucrose was the more protective excipient for these regions.

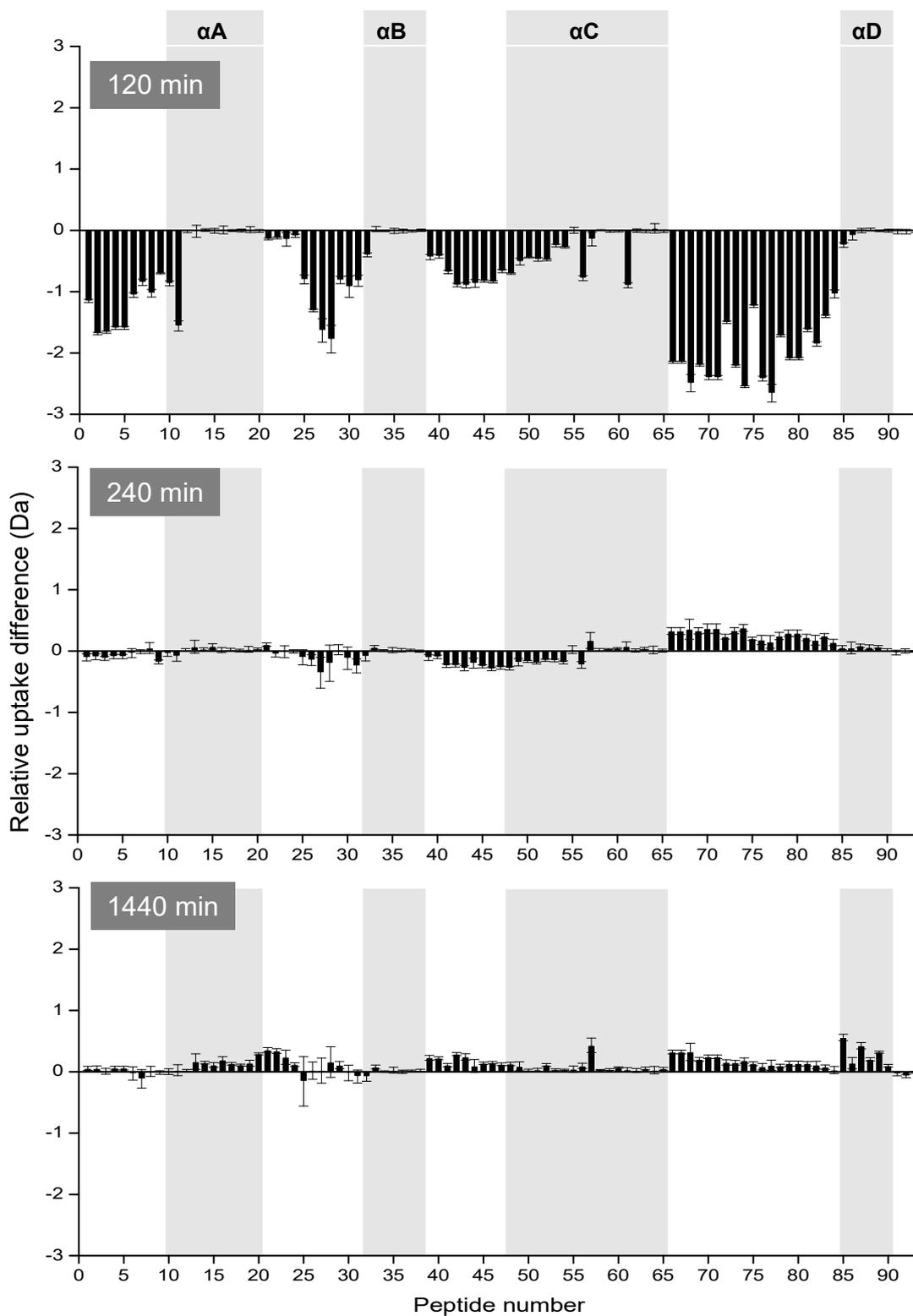


Figure 101. **Differential uptake plots for GCSF peptides lyophilised with phenylalanine as measured by ssHDX-MS.** Negative values indicate a reduction in uptake in the presence of the excipient. The peptide number in x-axis represents the 93 overlapping peptides obtained from pepsin digestion. The helical locations are denoted at the top of the figure, and time points are displayed in a grey box at the top left corner of each figure. Amino acid sequence and the corresponding peptide number range are provided in Appendix 14.

7.1.8 Comparing excipient protection factors

To assess the global effects of the three excipients, sucrose, mannitol and phenylalanine, on GCSF deuterium uptake in the solid state; the sum of differential values for all peptides at different labelling time points was calculated and displayed in Table 28. At 30 mins, where the most obvious protective effects were seen, it was found that mannitol and phenylalanine had very similar values at -74 Da and -73 Da, respectively. Sucrose data was not available at this time point due to issues with snap freezing the samples. At 120 mins, sucrose showed a negative differential sum whereas mannitol had a positive value and phenylalanine had a value close to neutral. This suggested that, of the three excipients, sucrose was the most protective at this time point. The same result was seen at the latest time point, 240 mins, where sucrose still maintained a negative differential; however phenylalanine had a positive differential value which, as described earlier, was due to the collapse of the control sample.

Table 28. Sum of all peptide differential values for each excipient at different labelling times

Excipient	Differential uptake (Da)		
	30 min	120 min	240 min
Sucrose	N/A	-26.809	-10.179
Mannitol	-74.168	10.727	1.579
Phenylalanine	-72.843	0.813	10.343

By SEC-HPLC, it was earlier determined that mannitol had the highest monomer retention during lyophilisation of 2 %, which was closely followed by sucrose with an increase of 1%. Phenylalanine destabilised GCSF during lyophilisation and retention was reduced by 9% compared to the no-excipient control. This did not coincide with the differential uptake data after 120 min of labelling (earliest time point where all samples had data available); however, at the 30 min labelling time point the difference in differential uptake total between mannitol and phenylalanine indicated mannitol was more protective. This is in line with the SEC-HPLC data, although the difference isn't as large as the loss in monomer indicates it should be. Due to the lack of obvious differences in ssHDX-MS total differential data versus SEC-HPLC monomer recovery, it was concluded that total differential uptake was not an effective predictor of lyophilisation stability.

7.1.9 Comparing ssHDX-MS and docking H-bond data

GCSF residues identified by iGEMDOCK in chapter 6 (Section 6.1.2) involved in H-bonds with sucrose, mannitol and phenylalanine at pH 4, were highlighted on the earliest labelling time point differential figures for each excipient (Figure 102). There was a match observed between regions with negative differential values identified by ssHDX-MS and the docked regions of all three excipients; however, the regions that matched up were not the most protected regions. The region with peptides showing the lowest differential was LoopCD, which showed no docking. *In silico*, sucrose docked at two different regions on the GCSF structure, whereas mannitol and phenylalanine only docked in a single region. The region not docked by mannitol and phenylalanine around LoopA, showed a larger negative differential value for sucrose compared to mannitol and phenylalanine during ssHDX-MS analysis (Figure 102), suggesting docking and ssHDX-MS protection data aligned at for this region.

Compared to the differential uptake plots for sucrose and mannitol, the positive differential values in peptides involved in the start of loop AB (number 21 and 22) in the differential plot for phenylalanine after 30 mins of labelling suggested increased flexibility around this region. The increased uptake at this region could be the cause of the 9% decrease in monomer after lyophilisation.

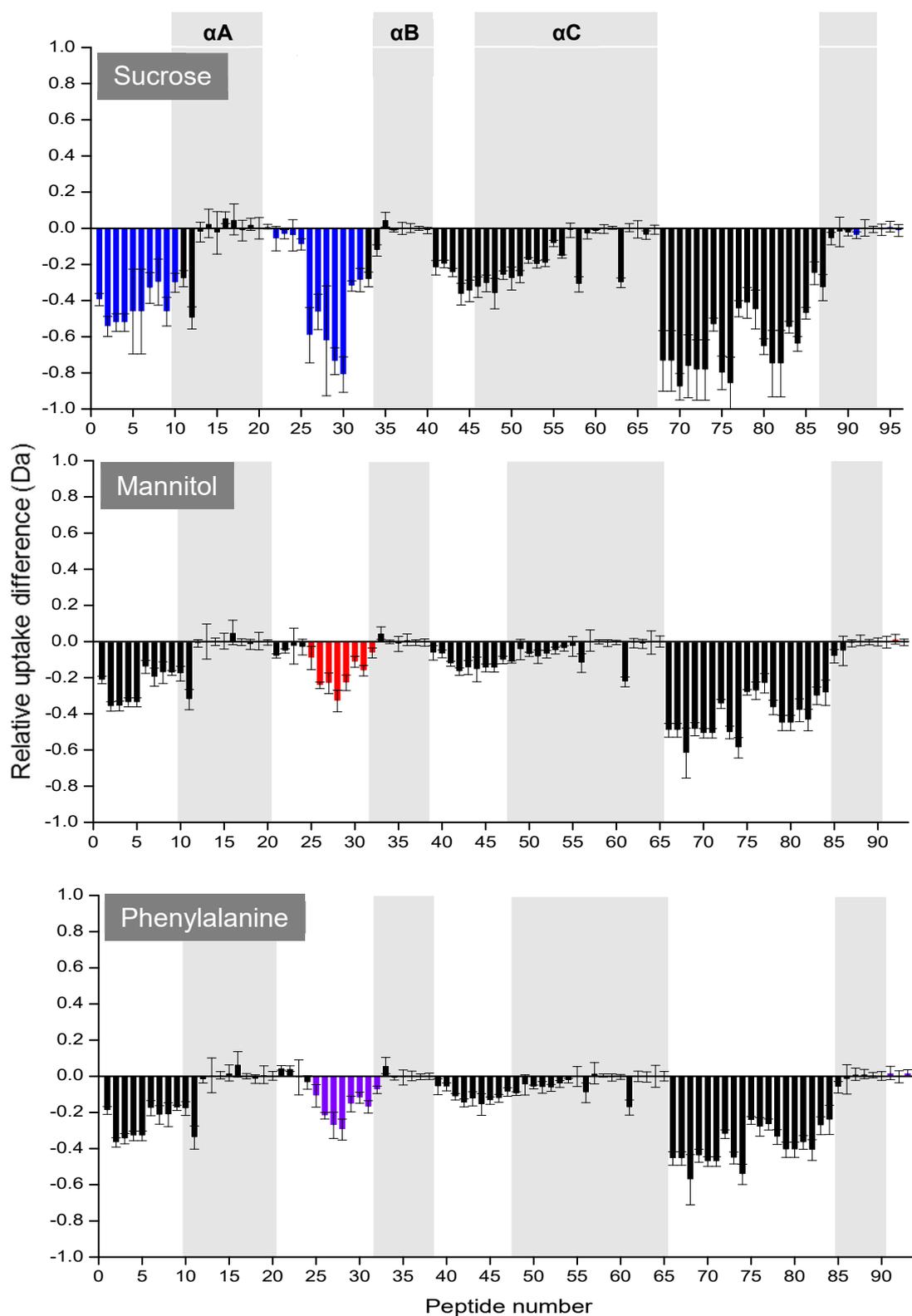


Figure 102. **ssHDX-MS protected regions aligned with docked regions for GCSF and excipients.** The excipient studied in each figure is displayed in the grey box. Differential values were obtained by ssHDX-MS experiment data with the shortest labelling time. For mannitol and phenylalanine, the labelling time was 30 min and for sucrose it was 2 hrs. Negative values indicate a reduction in uptake in the presence of the excipient. The peptide number in x-axis represents the 93 overlapping peptides obtained from pepsin digestion. The helical locations are denoted at the top of the figure, and time points are displayed in a grey box at the top left corner of each figure. Amino acid sequence and the corresponding peptide numbers are provided in Appendix 14. Highlighted on each figure are peptides which include residues identified to interact with the excipient by *in silico* docking

Discussion

7.1.10 USD lyophilisation

7.1.10.1 *USD wells show a higher level of residual moisture compared to vials*

Shrinkage of lyophilised cakes within 96-well microtitre plates has previously been observed during accelerated degradation studies at low storage temperatures, which does not occur in vials containing the same sample at the same temperature. In this chapter, TGA was used to measure the residual moisture of USD lyophilised cakes compared to the same cakes in glass vials. During TGA analysis, the addition of material from one USD well was found to be adequate for TGA pan filling, confirming the TGA as a method for measure USD residual moisture analysis. Upon comparison between containers, it was found that residual moisture differed by 3.03% (w/w). The increased moisture within the USD well may be due to the difference in container material. The plastic of the 96-well plate may not be as efficient at thermal conducting compared to glass vials, and as such samples do not reach shelf temperatures as quickly. Future experiments are required to determine if this difference in moisture is consistent when comparing wells and vials with different formulations, as if it is a scale factor could be established. Alternatively, the use of glass 96-well plates could be explored.

7.1.10.2 *Well position affects moisture content but not significantly*

It has been previously noted that during accelerated degradation storage, cake shrinkage occurs over time. TGA analysis also showed there wasn't a significant difference in residual moisture between different positions in the same 96-well microtiter plate, however the values were higher in the inner wells of the plate. This result was expected as higher sublimation rates have been shown to occur in outer wells, which would result in a drier lyophilised cake (Grant et al. 2009). It has been suggested by Grant et al. (2009) and later implemented by Zhang (2017), to avoid using the outer wells and corners of the UDS plates to minimise differences in residual moisture, which may occur due to the different sublimation rates. Another cause of increased moisture over time could be due to a weak seal between the USD well and stopper. During sample preparation in this chapter, stoppering of the 96 individual stoppers into the plates using the freeze-dryer hydraulic shelves proved to be mostly incomplete and most had to be pushed into the wells by hand once the plate was removed from the freeze-dryer.

7.1.10.3 *USD method useful for initial formulation screening*

Appearance and reconstitution analysis of the USD lyophilisation plate was relatively easy to perform due to the large number of samples within a small area, compared to the 2 mL glass vials which had to be imaged and reconstituted one at a time which increased processing time. During early-stage solid state formulation screening the number of samples can be large; therefore, the use of USD methods can alleviate some of the material requirements and speed of analysis for such screens. As the occurrence of moisture in the wells was roughly twice that of the glass vials, and no correction factor has been determined, it is recommended that USD methods only be used for cake appearance and lyophilisation survival, where immediate reconstitution of samples and analysis is performed. As increased moisture can cause increased degradation, the USD method in its current state does not translate well for accelerated degradation studies, and instead glass vials should be used where stoppers can be crimped in place and the relative moisture % of the sample maintained during studies.

7.1.10.4 *Sodium citrate is a stable GCSF lyophilisation buffer*

From the USD formulation screen, sodium citrate at pH 4.25 was found to be the most suitable buffer for GCSF lyophilisation compared to sodium acetate, also at pH 4.25, and PBS at pH 7.4. In fact, sodium citrate was such a stable buffer for lyophilised GCSF survival, that the addition of other cryoprotectant excipients only marginally improved monomer recovery. Stable lyophilisation of GCSF in citric acid/ sodium citrate has also been previously described in a patent by Zobel and Arndt (2005), and an arginine citrate buffer described in a patent by Michaelis et al. (1999).

As GCSF is stable at acidic pH, the added stress of lyophilisation whilst formulated in PBS at physiological pH explains the loss of monomer for this buffering system. For a low pH formulation, sodium acetate is the choice buffer for GCSF in solution, however, during lyophilisation screening it was detrimental to stability, causing greater than 95 % monomer loss. This possibly occurred because during lyophilisation the volatile acetic acid in the buffer is removed during the ice sublimation stage and causes a pH shift in the reconstituted solution (Franks 1998). The change in pH will have effects not only on GCSF monomer stability once reconstituted, but also affects excipient solubility where during reconstitution it was found that formulations containing mannitol and phenylalanine generated insoluble particulates. Sodium acetate cakes were difficult to reconstitute, and

the cakes were also fragile and easily broken up, forming highly static clumps of lyophilised cake coating the sides of the vials and cracks of the rubber stoppers. Presumably this was caused by the rapid removal of the acetic acid during sublimation. This breaking up and dispersion of material made reconstitution difficult and may have contributed in part to the loss of monomer.

7.1.10.5 Multiple excipients can stabilise GCSF in the solid state

The effect of excipient addition to the GCSF formulations during lyophilisation was measured by monomer and activity retention, where it was found that results varied depending on the buffer used. The activity recovery as measured by the GNFS-60 cell proliferation bioassay confirmed the results of the SEC-HPLC monomer retention, where the more monomer in the reconstitution sample, the closer the activity profiles of the pre- and post-lyophilisation samples.

For sodium acetate, trehalose and sorbitol increased monomer recovery, for sodium citrate, trehalose was the only excipient to slightly increase recovery, and for PBS all excipients increased recovery, with phenylalanine being the highest at ~ 90% recovery. Trehalose is a known cryoprotectant and a commonly used stabiliser in the lyophilisation of proteins (Pikal 1990) which explains its stabilisation of GCSF during lyophilisation. Mannitol is a bulking agent, and not a cryoprotectant, added to lyophilised formulations to provide mechanical strength, explaining the lack of effect on monomer and activity retention in acetate and citrate. Interestingly, mannitol and sorbitol are isomers and yet sorbitol was found to increase recovery of monomer in acetate and PBS buffers. As GCSF is unstable at physiological pH, the loss of monomer suggests the addition of lyophilisation stress causes the protein to aggregate in PBS. The stabilisation of phenylalanine in lyophilised PBS formulations indicates the amino acid reduced this aggregation, most likely via its interaction with exposed hydrophobic regions of GCSF, blocking protein-protein interactions.

7.1.10.6 Collapsed sorbitol sample still stabilised GCSF

Sorbitol, as discussed in chapter 6, is a notoriously difficult excipient to lyophilize. During cake appearance analysis, formulations containing sorbitol formed a small, moist, white droplet in the bottom of the USD microwell after lyophilisation, indicative of product collapse; however, sorbitol retained a high level of GCSF monomer and activity, especially in sodium acetate and PBS. In comparison with disaccharides, sorbitol has a poor glass

forming tendency, reflected by a low T_g' of $-43\text{ }^\circ\text{C}$, making it a poor choice for an efficient freeze-drying process (Foerst et al. 2010; Saffell-Clemmer 2015). The target product temperature during the primary drying stage of an optimised lyophilisation process is several degrees below a critical threshold value corresponding to the T_g' or T_c . The primary-drying of sorbitol in this chapter took place at $-35\text{ }^\circ\text{C}$, which is above the T_g' suggesting the formulation collapsed due to sorbitol being in a supercooled amorphous liquid state (Foerst et al. 2010).

Cake collapse, as seen in the sorbitol containing samples, can be detrimental to product quality. Collapse is generally caused by viscous flow of the glassy matrix, increasing the density of the cake structure and eventually resulting in blockage of pores. This leads to increased product resistance to water vapor flow, causing decreased sublimation rates and increased residual moisture levels (Schersh et al. 2010). However, the effect on stability of protein biopharmaceuticals during collapsed cake is still under debate. A number of studies suggest proteins do not necessarily require a porous cake structure. Wang et al. (2004) found no difference in the long-term stability and an increased stability of some collapsed samples at high storage temperatures ($40\text{ }^\circ\text{C}$) in a partially crystalline formulations consisting of sucrose and glycine. Another study by Schersh et al. (2010) investigated the effect of cake collapse during lyophilisation of a monoclonal IgG1 antibody and a biotherapeutic protein and found protein stability was not relevantly different between collapsed and non-collapsed cakes as measured by formation of soluble and insoluble aggregates, biological activity and conformational stability. On the other hand, Passot et al. (2007) investigated the effect of applying primary drying at temperatures above and below the formulation T_g' on the long-term stability of lyophilised proteins where they found a decreased long-term stability of proteins primary-dried at temperatures above (T_g') regardless of the onset of macroscopic cake collapse. The structural collapse of GCSF with sorbitol from this chapter lead to unattractive changes to the cake structure but improved GCSF stability short term, compared to other excipients suggesting collapse does not negatively affect this particular protein's stability, however the formation of a moist cake at the bottom of the wells also suggests a high residual moisture level, which could cause stability issues during long-term storage. Moreover, as the ssHDX-MS labelling the method involves D_2O vapor diffusion through a porous dried matrix to label the protein, sorbitol could not be studied using this technique, as the cake is not porous and the high residual

moisture may affect uptake rates. Sorbitol remains, however, a very interesting excipient in terms of preserving GCSF monomer and activity during lyophilised cake collapse.

7.1.10.7 *Taking formulations forward for biophysical analysis*

The formulation screen found GCSF with sodium citrate pH 4.25 was sufficient for lyophilisation survival with high monomer and activity levels, however, excipients are required in formulations not only to confer protection during processing, but also during subsequent storage. As demonstrated by Chang et al. (1993), elastase lyophilised without any excipients retained full activity immediately following freeze drying, however, it denatured upon storage at 40 °C and 75% RH, losing ~70% of the initial activity in 2 weeks. Furthermore, correlations have been made by Moorthy et al (2017) between shelf-life and level of protection in the solid state. The group studied four mAb formulations and measured ssHDX-MS at the initial time point prior to storage and found the level of deuterium uptake strongly correlated with the extent of aggregation and chemical degradation during storage. These results suggest ssHDX-MS can be valuable for characterizing different solid-state formulations, where the extent of deuterium incorporation and kinetic parameters directly reflect the degree of protein folding and structure (Moorthy et al. 2015). As such, work was continued with ssHDX-MS and GCSF lyophilised with different excipients with the intention of identifying those that could improve shelf-life of the sample.

7.2 ssHDX-MS with GCSF

7.2.1 ssHDX-MS implementation on Waters system

This chapter covered the first application of ssHDX-MS using GCSF, as well as implementation of the technique on a fully integrated and automated Waters HDX system. Due to the time critical steps of reconstitution, quench and injection into the LC-MS, changes to the automated Waters HDX-MS system involved removal of the LEAP PAL™ robotic sample handling and replacement with manual sample preparation. The average time from removal of the vials from dry ice to injection was around 2 mins and was highly reproducible when performed by the same operator. Furthermore, the standard deviation of relative uptake between three injections of different vials containing the same labelled sample was as low as those of standard HDX-MS samples when using the LEAP PAL™ automated sample preparation, which demonstrates the success and reproducibility of the ssHDX-MS sample preparation and reconstitution method to preserve labelling of samples.

7.2.2 Improvements observed for the GCSF peptide map coverage

The peptide map sequence coverage produced in the ssHDX-MS experiments was higher than seen in previous chapters 3, 4 and 5 using WT GCSF. This is thought to have occurred in the 0.1 % formic acid reconstitution step. Formic acid was selected due to its low pH providing ideal quench conditions minimising back exchange, however, it also has denaturing properties which could have enhanced protein digestion. Stacked spectra peaks were also notably well-defined during review and required less manual intervention. The clean spectral plots suggested a reduction in peptide carry over, a process by which a peptide is retained between sample injections and manifests as isotopic profile doublets consisting of both a deuterated peptide and a corresponding semi-deuterated peptide. A study by Manjumdar et al. (2012) found carry-over originates in the online-digestion stage (immobilised pepsin column) and could be substantially decreased by washing the online digestion flow-path and immobilised pepsin column with two wash steps containing A) 5% acetonitrile (v/v), 5% isopropanol (v/v) and 20% acetic acid (v/v) in water and B) 2 M guanidine hydrochloride, 100 mM phosphate buffer pH 2.5. It is proposed that the reconstitution of GCSF in 0.1 % formic acid enhanced GCSF digestion and this had a knock-on effect in reducing peptide carry over within the pepsin column, and resulted in well-defined extracted ion chromatograms.

7.2.3 Moisture induced collapse of samples caused issues during labelling

Moisture induced collapse of samples was a major issue during the ssHDX-MS labelling step. To explore whether the type of excipient was causing collapse, a second USD study was performed in this chapter to monitor the moisture induced collapse of different excipients co-formulated with the IRP. During this study all excipients were formulated at 1% (w/v), equivalent to 10 mg/mL. It was found that during labelling, all cakes collapsed apart from the formulation containing phenylalanine, which remained intact during all labeling time points. When reviewing previous ssHDX-MS experiments, there are no reports of moisture induced collapse of cakes. In fact, the RH of 43 % was selected based on a ssHDX-MS development paper recommending this RH to observe the greatest differences between sucrose and mannitol (Sophocleous et al. 2012; Sophocleous and Topp 2012). In a 2012 exchange kinetics paper, Sophocleous and Topp (2012) studied the vapour sorption for equine myoglobin (Mb) in lyophilised cakes containing mannitol and sucrose, at 5 °C and 43 % RH. The powder in each vial contained 10 wt % phosphate buffer, 45 wt % Mb, and 45 wt % sucrose or mannitol. The study used gravimetric sorption analysis (GSA) to

show sucrose cakes sorbed around 25 % more water than those containing mannitol, but the mannitol cake reached 50% of the water vapour sorption plateau quicker (0.06 h in mannitol and 0.09 h in sucrose). On the other hand, in more recently published studies, the % RH used by the ssHDX-MS group has decreased to 11% RH using a saturated D₂O solution with LiCl at 22°C (Moussa et al. 2018A; Moussa et al. 2018B) indicating that 43 % may be too high a RH for prolonged labelling. In a study by Moorthy et al. (2017) samples of mAb at 50 mg/mL were lyophilised with different concentrations of sucrose and mannitol ranging from 32-53 mg/mL and 8-80 mg/mL, respectively. They incubated samples in D₂O at a RH of 11% at 22 °C for 5 days and found no significant difference in cake morphology by SEM, indicating no moisture-induced cake collapse.

7.2.3.1 *Increasing excipient concentration does not prevent moisture induced collapse*

It's generally known that lyophilised sample stability increases as the content of both protein and excipient increases (Carpenter et al. 1997). From previous ssHDX-MS studies the concentration of protein included was an order of magnitude higher than those included in this chapter, which could imply that the lyophilised cakes produced were not concentrated enough to withstand the high RH labelling environment. Unfortunately, the concentration of GCSF could not be increased much more than that used during the ssHDX-MS studies, as too high a concentration (above 1 mg /mL) causes increased aggregation propensity (Raso et al. 2009; Treuheit et al.2002). To explore the effect of excipient concentration on moisture induced collapse, a USD study was performed in this chapter to study different concentrations of sucrose. Concentrations ranged up to 10 % (w/v), the equivalent to 0.1 g/mL, which is twice the highest concentration studied by Moorthy et al. (2017). The study found during 43 % RH labelling all sample cakes collapsed. This suggests that whilst increasing the concentration of sugars may decrease moisture sorption, the level of moisture taken up is still enough to collapse the cake. As the lower the RH % and temperature the lower the level of moisture sorption, it may be that future ssHDX-MS experiments use a salt solution yielding a lower RH % in a temperature controlled environment < 22 °C during labelling to prolong cake structural integrity.

The T_g of amorphous solids can be reduced by the presence of plasticising molecules dissolved in the amorphous solid such as water, so as the moisture level increases the T_g is reduced (Diuralliu et al. 2018). If moisture content is raised sufficiently, a glassy solid will revert to be a viscous fluid thereby facilitating complete cake collapse back to a liquid solution, as observed in this chapter (Roos 2010). An excipient moisture sorption study by

Costantino et al. (1998) employed GSA to measure moisture sorption of mannitol, sucrose and trehalose co-lyophilised with three different biotherapeutic proteins. The GSA technique placed lyophilised cakes on a microbalance within in a controlled-humidity environment where the relative humidity was increased in steps and the weight recorded. By FTIR, the lyophilised sucrose and trehalose spectra were relatively smooth indicating they were amorphous compared to mannitol which was found to be in the crystalline form. For pure lyophilised mannitol (no protein), no water uptake was seen with increasing RH by GSA, consistent with the formation of anhydrous crystals upon lyophilisation. At low protein-to-mannitol weight ratios water sorption was low and at high protein-to-mannitol weight ratios water sorption was high. For pure lyophilised sucrose and trehalose (no protein), water uptake was seen by GSA with increasing RH, up to 50 % RH where the weight plateaued. The moisture induced collapse of mannitol containing samples suggests the excipient was in the amorphous phase, and not crystallised, due to the lack of anneal step during freeze-drying. From DSC and FDM results it was confirmed that mannitol did not have a crystallisation event and had a T_c around $-35\text{ }^{\circ}\text{C}$, whereas phenylalanine had a clear crystallisation event and a T_c of $-19\text{ }^{\circ}\text{C}$. A study by Mattern et al. (1999) found recombinant human GCSF was stable in vacuum dried phenylalanine/maltose formulations at $40\text{ }^{\circ}\text{C}$ for at least 1 year. The inclusion of phenylalanine was shown to reduce lyophilised cake residual water to $\sim 1\%$ (w/w), and by X-ray diffraction and SEM it was shown to crystallise during vacuum-drying, forming networked structure. Consequently, their results combined with the results of this chapter suggest the hydrophobic, crystalline nature of solid state phenylalanine reduces its ability to adsorb moisture and prevented moisture induced collapse of the sample during labelling. It is therefore hypothesised that the collapse of the all samples apart from those containing phenylalanine was caused by a high residual moisture content inside the cakes following D_2O vapour sorption over time causing plasticisation of the excipients in the amorphous phase.

Furthermore, to prevent future collapse of mannitol containing cakes during labelling, an anneal step could be included within the freeze-dry cycle to ensure crystallisation.

7.2.3.2 *Internal reference peptide doesn't measure intrinsic exchange in ssHDX-MS*

This chapter demonstrated the first use of an internal reference peptide in both a full scale ssHDX-MS study looking at the protective effects of sucrose, and a USD screen of multiple excipients. Unstructured reference peptides in HDX-MS studies are included to monitor and correct for changes to the intrinsic exchange rate caused by differing solution conditions.

For ssHDX-MS, lyophilisation removes water from the sample causes the unstructured reference peptide to become captured within the lyophilised cake and as a result is not free for monitoring of intrinsic exchange rates. As such, the different rates of uptake seen for different excipients was most likely a measurement of their interaction with the peptide rather than differences in intrinsic exchange and therefore, did not need correcting for in the data.

Peptide differentials comparing the control sample with and without the IRP found that surface exposed regions of the GCSF showed an increase in exchange. Whilst direct interactions would most likely cause a decrease in exchange due to blocking of deuterium ions, the increase in exchange suggests destabilisation of GCSF. Destabilisation could be caused by the presence of organic solvents in the IRP solution including 6.25% (v/v) acetonitrile and 6.25 % (w/v) methanol, which would both decrease to 0.006 % (v/v) after spiking. Both acetonitrile and methanol have low freezing temperatures at $-42\text{ }^{\circ}\text{C}$ and $-97.8\text{ }^{\circ}\text{C}$, which would be increased by dilution in the sample but may still influence the overall freezing of the GCSF-PPPI samples (Sprung 2012). A change in ice crystal structure can affect primary drying as well as the structure of channels left by sublimed ice (Patapoff and Overcashier 2002). The presence of acetonitrile and/or methanol may have caused larger channels within the lyophilised cake, which may have increased deuterium access to GCSF molecules during labeling, and increased uptake levels.

7.2.3.3 Decreased uptake of deuterium by GCSF formulated with sucrose and the internal reference peptide

Peptide differentials comparing the sucrose/IRP, with the control/IRP were found to be larger than those without the IRP. Suggesting the peptide enhanced the effects of sucrose. This result was also confirmed during the USD excipient screening where sucrose decreased IRP exchange. As sucrose is not deuterated during labelling, the uptake data suggest sucrose interacts with the peptide and protects it, via competition, from deuterium uptake.

A strong linear correlation was made between the differential values comparing the control and sucrose samples with and without the IRP, after 120 mins of labelling. The strong correlation indicated that the difference with the addition of the IRP was non-specific (a global effect), and as such did not affect the ssHDX-MS result outcome. There was, however, one outlying peptide, number 79 (residues 128 to 138), which had a lower

differential value with the IRP relative to the differential value without the peptide. This suggested uptake was decreased in this region and could have been caused by residues within peptide 79 interacting with the reference peptide. Further work would need to be completed to confirm this, such as the use of computational docking as described in chapter 4 to identify regions of interactions between a protein and ligand.

7.2.3.4 Internal reference peptide can be used to screen excipient interactions

The IRP was used as a quick screening method for excipient solid state protection in a USD ssHDX-MS screen. However, there is some scepticism as to whether an unstructured IRP can be used in the place of a folded protein. The peptide was used instead of GCSF for ease of uptake data analysis, where a single peptide required identification rather than 90 with multiple charge states. Furthermore, as the IRP did not require digesting, the LC method was shortened reducing processing time and increasing throughput. The IRP excipient screen found sucrose, glycine and arginine reduced deuterium uptake over time, where sucrose and arginine had consistently lower rates of uptake whereas, glycine had the highest uptake rate after 30 mins but plateaued after 2 hours at a level lower than the control. The glycine uptake plateau coincided with collapse. Mannitol and phenylalanine had similar uptake levels to the control at all time points suggesting a lack of interaction.

7.2.4 Interactions of GCSF-excipients in the solid state

Three full scale ssHDX-MS experiments were performed in this chapter comparing lyophilised GCSF with three different excipients against a buffer control. Excipients selected were sucrose, mannitol and phenylalanine.

7.2.4.1 Moisture collapse caused issues with data analysis

The moisture induced collapse of lyophilised cakes caused significant changes to the uptake of deuterium by GCSF both in level of incorporation relative to the control sample and increased variability between samples. When comparing two collapsed samples such as the control and sucrose containing cakes, at 240 mins the exchange was comparable whereas when collapse hadn't occurred there was a clear increase in protection with sucrose. This result suggests upon collapse, the samples are both saturated with moisture, removing differences in protein-excipient interactions between the different samples.

When comparing the collapsed control sample and the non-collapsed phenylalanine sample, at 240 mins the differential showed an increase in exchange with phenylalanine

whereas at earlier time points there was a decrease in uptake with phenylalanine. The collapse of the control suggests the increase in uptake with phenylalanine is caused by the continued “normal” uptake of deuterium from the diffusion of D₂O vapour through the phenylalanine structured cake, whereas the collapsed control cake lacks any tracks for vapour to pass through causing retarded uptake, plateauing at a level lower than that of phenylalanine. Therefore, considering cake collapse with result analysis suggests early time point labelling where both the control and experimental cake is still structured is the only real indication of differences in interactions by different formulations.

7.2.4.2 Elucidating GCSF-excipient interactions in the solid state

During lyophilisation, the first step in the process is freezing. During this step water is still present in the sample and is gradually removed during ice crystallization. The freezing process involves a physical state that is described mainly by the interactions that are present in solution (Ohtake et al. 2011). In solution, the interactions between GCSF and sucrose, mannitol and phenylalanine were found by HDX-MS in chapter 6 to be due to stabilisation by preferential exclusion, however, during the two drying steps, water is removed from the protein, concentrating solutes in solution, which effectively forces protein-excipient interactions (Figure 103).

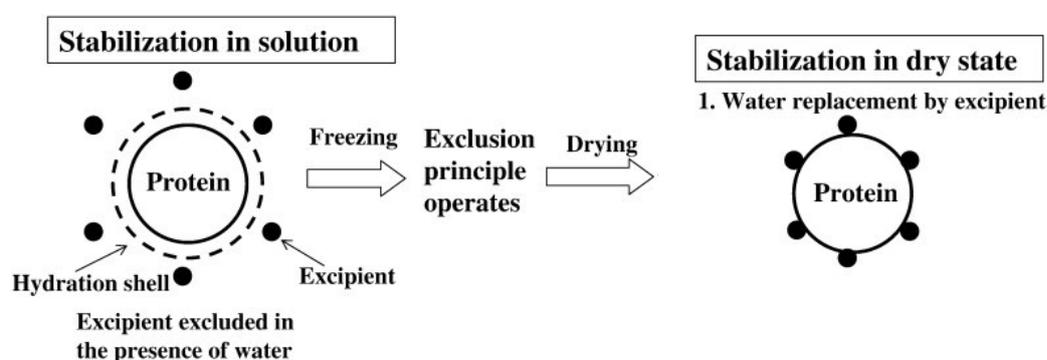


Figure 103. **Stabilising of proteins in solution and dry state.** Adapted from Ohtake et al. (2011)

There are two main mechanisms to stabilisation of proteins in the dried, solid state, by excipients: the water replacement hypothesis and vitrification. The common theme of both is the reduction of protein mobility. The early ssHDX-MS labelling time points showed a clear a reduction in uptake in GCSF peptides covering surface exposed regions of GCSF when formulated with all three excipients compared to a non-excipient control. This confirms the protection of GCSF in the solid state, it does not, however confirm the reduction of mobility as multiple time points could not be measured to obtain uptake rate

information. What can be compared from the early time point data is the amount of protection each excipient provided and how this varied for different regions of the GCSF. The sum of all peptide differential values was calculated for the three excipients where there were differences in the global levels of uptake protection. Mannitol and phenylalanine had similar differential values at 30 mins (30 mins data not available for sucrose) indicating that the excipients stabilise GCSF in a similar manner. Whereas at 120 mins, sucrose had the largest negative differential, phenylalanine was neutral and mannitol moved to a positive differential. This data, whilst not complete, is in keeping with previous ssHDX-MS study by Sophocleous et al. (2012) using ssHDX-MS to compare equine myoglobin (Mb) formulations containing either sucrose or mannitol, where they found ssHDX was lower with sucrose than mannitol both globally and at a peptide level. In addition, in recent studies by Moussa et al. (2018 A & B) comparing ssHDX in freeze-dried and spray dried IgG1 mAb it was found that mannitol formulations were the most structurally perturbed, compared to sucrose and had the fewest H-Bond interactions between the protein and the surrounding matrix. These results are consistent with sucrose being a cryoprotectant providing the most protection, whereas phenylalanine and mannitol are in a crystalline form that does not provide adequate stabilisation to the protein.

During local region analysis of excipient protection, it was found that the excipients studied generally reduced the uptake of deuterium in GCSF loops, and α C. LoopCD was the region with the greatest negative differential with all excipients. This region has previously been shown by HDX-MS in chapter 3, to be highly mobile in the aqueous phase, and in chapter 5 reductions in flexibility of this region via mutation stabilised GCSF. This confirms the stabilisation of highly mobile regions of proteins in the solid state with the addition of excipients, which has been shown to have direct consequences on long-term storage.

When comparing ssHDX-MS data to the excipient docking data from chapter 6 it was found that no docking was predicted to occur at LoopCD. This suggest that the docking data may once again not be accurate in predicting sites of interaction between protein and excipients. On the other hand, the docking data did correlate with sucrose interacting with LoopA, where a greater negative differential was observed with this excipient compared to mannitol and phenylalanine that were not predicted to interact at that site. As discussed at the end of chapter 6 perhaps the use of IGEMDOCK was not the best software to use for docking as the program has previously been shown to predict fewer sites of interaction

than other programs such as AutoDock and GLUE. Other studies have had success predicting the interaction of other protein-excipient combinations and linking the data to ssHDX-MS data, although none have been published this far. The potential is there for development of these tools and the interaction data from ssHDX-MS.

Additionally, ssHDX-MS identified GCSF peptides with a positive differential in loopAB with the inclusion of phenylalanine in the solid state formulation. Phenylalanine was found by SEC-HPLC to cause the loss of 9% of GCSF monomer during the lyophilisation and reconstitution process. As all other peptides showed either a negative differential or were similar to mannitol and sucrose peptides, which increased monomer retention during lyophilisation, it is thought that the peptides with increased uptake could be the cause of the monomer loss, however with phenylalanine the difference wasn't as obvious as the loss in monomer data indicated it could have been. The peptides, 20 and 21 contain the residues 33-39 and 40-47, respectfully. The peptide number 19 contained residues 23-32 and had no overlap with adjacent peptides, whereas peptides 22 and 23 contained residues 41-47 and 41-48, respectfully. The peptides 22 and 23 did not show the positive differential that peptide 21 did, therefore the residues different to these two, can be assumed to be the cause of the increase. Since the protection information is lost in the first residue (where amide is converted to amine) and typically compromised in the second residue (due to accelerated back-exchange under the slow exchange conditions (Bai et al. 1996), this means that the difference was caused by residues 40 + 2. Consequently, the region affected by phenylalanine destabilisation covered residues 33 to 42 within the end of α A and start of loopABI. Ultimately it is unknown if this region interacts with phenylalanine molecules during lyophilisation or becomes more exposed by allosteric effects from interactions with other regions. However, the loopAB region was previously shown to be implicated with GCSF aggregation from the mutants G51R and L71W which were shown to increase flexibility in the LoopAB region and increase GCSF aggregation rates during accelerated thermal stability studies. Additionally, in their work with WT GCSF, Raso et al. (2005) linked conformational changes to LoopAB to aggregation in low Gdn.HCl concentrations, at pH 7 as seen by increased fluorescence emission intensity of tryptophan at position 60. Moreover, GCSF mutant studies by Luo et al. (2002) and Buchanan et al. (2012) both identified mutants with mutations located in LoopAB (in combination with mutations in other regions) which increased stability. In combination with the results from this chapter, results suggest the LoopAB and short helix could be hotspots for GCSF aggregation. As such,

ssHDX-MS has been shown to be a useful technique to further characterise the cause of aggregation during lyophilisation.

7.2.5 Future work

7.2.5.1 *Preventing moisture induced collapse of lyophilised GCSF samples*

It is envisioned that a continuation of this work would include the development of solid state formulations of GCSF that do not collapse during labelling by either reducing the RH or by adding a crystalline excipient to each formulation to ensure cake stability. Alternatively, a different molecule could be selected such as a mAb, which can be formulated at higher concentrations. However, this also comes with its own issues such as complexity in digested peptide processing during HDX-MS. A recent study by Duralliu et al. (2018) described the use of real time video imaging combined with a GSA technique called dynamic Vapor Sorption (DVS) to measure the humidity induced collapse point, RH_{CP} . The technique is similar to that of FDM, whereby frame by frame image analysis (in combination with mass increase measurements) can pinpoint collapse of cakes at different RH % and temperatures. This technique lends itself well to future work with ssHDX-MS, where the deuterium labelling environment could be optimised for a given sample to prevent sample collapse.

7.2.5.2 *Next steps for ssHDX-MS GCSF analysis*

Long-term storage analysis of lyophilised samples could be performed indicate if the protection effects by ssHDX-MS translate into increased shelf life stability. Finally, the inclusion of stability mutants studied in the same formulations, such as Q120I and M131F produced in chapter 4, would allow for further understanding of the effects of different excipients on the lyophilisation survival of proteins with different levels of stability providing further understanding into protein-excipient interactions and shelf-life stability.

7.2.5.3 *Application of USD screening with ssHDX-MS*

During ssHDX-MS experiments GCSF was lyophilised with excipients in 2 mL glass vials with a fill volume of 200 μ L, however, after reconstitution and quench dilution steps (10X and 2X) it was observed that only 10 μ L of the lyophilised sample was injected onto the LC-MS and the bulk of reconstituted labelled GCSF sample went to waste. The fill volume of 200 μ L was selected as it is the minimal volume required to cover the base of the 2 mL glass vial, as such there was no scope to reduce the volume in this container. As discussed earlier, the use of USD 96-well microtiter plate lyophilisation method lends itself well to

formulation screening strategies where analysis is performed immediately. During ssHDX-MS sample preparation, samples are removed immediately from the freezer-drier and labelled meaning USD containers can be used and a reduction in sample volumes can be made. Furthermore, although poor stopper sealing means moisture uptake over time could be an issue, the storage of labelled samples at -80 °C means there is a lack of free moisture within the -80 °C storage environment.

This chapter demonstrated the first use of a USD ssHDX-MS method to successfully carry out two process development studies, where the sample fill volumes were reduced by 75 % (50 µL). USD lyophilisation of sucrose formulations was performed in standalone 1.4 mL Micron Tubes within a 96-tube format, rather than a fixed 96-well microtitre plate. This format was found to be best as samples required the same freeze-dry cycle but different labelling times, meaning samples had to be removed from the desiccator at different times. This would not have been feasible with a 96-well microtiter plate due to the large footprint multiple plates would have. The trays containing the tubes took up minimal space; so much so that a smaller desiccator could have been used.

As well as saving sample material, with the increased availability of liquid handling and automation instruments within laboratories, the automated preparation and reconstitution of formulations within 96-well format containers can increase throughput and reduce risks to repetitive strain injuries. Unfortunately, sample handling post labelling caused a process bottleneck and prevented this number of samples being feasible. The first bottleneck occurs during removal from the desiccators post-labelling where capping of tubes and snap freezing occurred. Attempts were made to stopper all 96-tubes using the individual rubber stoppers held in a 96-well format and snap-freeze en masse; however during submersion into liquid nitrogen, several tubes became loose. Future users of this technique may want to consider securing tubes into holders prior to submersion in liquid nitrogen. The second bottleneck occurred when samples required maintenance at -80 °C up until reconstitution. This extreme temperature requirement to minimise back exchange of labelling limits the use of automated sample reconstitution prior to LC-MS injection, however studies could be performed to measure the rate of back exchange of ssHDX-MS samples over time within attainable quenched solutions i.e. low pH, 0°C.

8 Conclusion

Structure and dynamics contribute significantly to the function of biotherapeutic proteins. (Engen 2009). It has been previously shown that the flexibility of the amide backbone and side-chain interactions can play a significant role in maintaining protein thermal stability (Leone et al. 2004; Feng et al. 2013; Manikwar et al. 2013); therefore, to fully understand stability, the interplay of structure, function, and dynamics must be investigated.

GCSF was used as a model protein for this body of work due to its small size and basic structure. Its inherent flexibility in a stable state was measured using the advanced biophysical technique, HDX-MS with pepsin digestion for peptide-level analysis. The method was optimised in chapter 3 to improve the protein sequence coverage and redundancy by enhancing the pepsin digestion via the inclusion of chaotropic agents and backpressure. The inclusion of both guanidine hydrochloride and TCEP were required in the quench solution to increase peptide numbers and improve coverage around disulphide bonds, respectively. Whilst backpressure has been shown to improve coverage for other proteins, it was found to decrease coverage for GCSF.

HDX-MS is typically performed with sample and labelling solutions at physiological pH, however GCSF is unstable at these conditions. Both high and low pH HDX-MS experiments were performed and the data compared through peptide differential uptake calculations to establish if changes in protein dynamics occur. As pH directly affects the intrinsic rate of HDX, the low pH labelling time points were converted to the standard pH 7.4 at 22°C, where it was found values with the same time differed significantly. Regions with increased dynamics at pH 7 included, LoopAB, LoopBC, α C and α D and LoopD, whereas the LoopCD decreased. Consequently, it was decided that GCSF HDX-MS experiments should be performed with low pH solutions to ensure measurements were being made on the native, folded structure. Additionally, through the use of snapshot thermal stability measurements, T_m and T_{agg} , it was determined that GCSF was most stable in 10 mM sodium acetate pH 4.25, and that both concentration of buffer and protein, decreased thermal stability. This result helped in modifying the GCSF purification process to alleviate observed precipitation during size exclusion chromatography. The result also molded the final HDX-MS analysis of GCSF to use lower concentration sodium acetate, where it was established that in its native folded structure, the regions, LoopABII, the short helix, α C, LoopCD and α D had relatively high flexibility compared to their respective structures.

The local dynamics of GCSF measured by HDX-MS in chapter 3 was found to weakly correlate with the B-factors of a number of solved GCSF crystal structures, and matched 2D9Q B-factors the closest. Consequently this structure was used for *in silico* predictions of GCSF stability through re-engineering and formulation. In chapter 3, biophysical data was combined to identify regions of high flexibility, where the regions loopCD and α C helix were selected as target regions for *in silico* engineering. Two *in silico* parallel approaches were used for stabilisation of GCSF. The first was the use of RosettaDesign, an online server; the second strategy was the Rosetta_ddg_monomer application ran on a High Performance Computing Facility. The online server was used to predict substitutions within the identified target regions, whereas the HPCF application was used to map the entire GCSF mutagenesis space. A total of 18 GCSF mutants were cloned, expressed and purified. Characterisation of the mutants relative to WT was performed for yield, purity, bioactivity and thermal stability. The majority of mutants produced from both design streams were high yielding and relatively pure with little effect on GCSF bioactivity. These results demonstrated the power of using targeted mutagenesis to decrease the number of candidates to screen and increase the probability of retained biological activity.

From the tandem thermal denaturation measurements of fluorescence and SLS, T_m and T_{agg} values for mutants were determined. It was found that the Rosetta_ddg_monomer had a higher accuracy of producing stabilising mutants; at 64% whilst the RosettaDesign accuracy was lower at 43%. The improved accuracy is most likely due to the larger number of iterations performed per mutation for Rosetta_ddg_monomer, as well as the minimisation of the structure pre-*in silico* screening which relaxed the structure and reduced collisions.

As well as determining predictive success, the two design streams also compared the mutagenesis success in regions with differing levels of flexibility. The majority of destabilising mutations occurred in the α A helix, a region shown by HDX-MS and B-factor values to have relatively low flexibility, whereas stabilising mutants were found to have prominent changes in LoopCD, α C and α D regions, also highlighted by B-Factor values and HDX-MS of WT as those of high flexibility. The preliminary data suggested rigidifying proteins could be an effective strategy to improve thermal stability and possibly shelf-life stability, however, further experiments were required to determine if mutations had an effect on the higher order structure flexibility to confirm if rigidification was the cause of improved thermal stability.

In chapter 4, the effect of mutation on GCSF shelf-life was further explored by accelerated thermal degradation and comparisons between the degradation rates at different time points for a select number of mutants. By this stability characterisation method it was found that only two mutants had increased degradation resistance relative to WT, and results were not in-keeping with T_m/T_{agg} values obtained in the previous chapter.

Peptide level HDX-MS was used for the first time to compare local changes in GCSF mutant mobility relative to WT and provide understandings for the discrepancy between accelerated degradation and T_m/T_{agg} measurements. It was found that mutants with high rates of thermal degradation also showed an increase in uptake for a large proportion of peptides during HDX-MS analysis, and the opposite was true for reduced degradation rate mutants. Consequently, a strong correlation was made between the overall positive or negative differential values and degradation rate indicating that protein flexibility is a valuable target for rational protein stabilisation. In addition, mutants with increased rates of aggregation relative to WT were found to have prominent changes in exchange around the loopABII and short helix regions of the protein, whereas aggregation stable mutants were found to stabilise LoopCD and α C regions highlighted by B-Factor values and HDX-MS of WT as regions of high flexibility. Therefore providing further mechanistic understanding of GCSF aggregation. Moreover, the work from the two GCSF mutant chapters demonstrated the efficiency combining *in silico* computational screening with HDX-MS and/or B-factor data to increase protein stability. More importantly, this work provided evidence that mutations that increase thermal stability cause rigidification of the molecule, which was only previously hypothesised.

The combination of HDX-MS and *in silico* predictions was also utilised in chapters 5 and 6 to study GCSF-excipient interactions in aqueous and solid state formulations. Both chapters included initial formulation screenings using high throughput techniques such as T_m/T_{agg} measurements and ultra-scale down (USD) methods to greatly reduce the time and amount of material required to identify excipient candidates for advanced biophysical analysis. Additionally, *in silico* interaction predictions between GCSF and excipients were performed using the software iGEMDOCK where it was found that docking energy data could predict the outcome of the SLS experiments to 76% of the experimental space, aligning well with previous results by Barata et al. (2016) with A33 Fab, and highlighting the potential of computational screening to reduce excipient candidate numbers prior to *in vitro* screenings. However, the computational screening could not detect the destabilising

qualities of arginine with GCSF, as such arginine would need to be removed from future computational docking screenings and assessed experimentally.

Predictions of the GCSF residues involved with excipient interactions by docking simulations were validated experimentally by HDX-MS. Excipients were included in the labelling solution which required them to be fully deuterated prior to dissolution. An unstructured internal reference peptide was included in the GCSF sample, monitored the intrinsic exchange of the different excipient labelling solutions and confirmed the success of excipient deuteration. During peptide differential uptake analysis the excipients mannitol, sucrose and phenylalanine saw the predicted docked residues have large changes in the level of deuterium uptake. This indicated that docking of protein-excipients was somewhat accurate to a regional level; however, arginine showed no evidence of protection for peptides containing predicted interacting residues. The majority of GCSF peptides showed low level protection with arginine and indicated that, as with the destabilising of GCSF, docking could not be used to predict the interactions of GCSF with arginine.

The deuterated excipient HDX-MS data also shed light on the mechanism of stabilisation for the excipients studied and contributed new information on stabilisation of flexible loop regions by excipients not seen by high pH HDX-MS experiments. A general decrease in deuterium uptake for most of protein peptides within the excipient solutions was observed. This indicated that stabilisation for the most part was via preferential exclusion, a well-studied phenomenon whereby the excipients exclude themselves from the protein surface and there is excess water at the protein surface. An increase in excluded excipients caused a compaction of the native state, reducing protein flexibility and causing the reduction of global uptake levels during HDX-MS analysis. Unfortunately, the confirmation of preferential exclusion for the excipients sucrose, mannitol and phenylalanine removed the relevance of docking data.

The deuterated excipient HDX-MS results also identified a minimum labelling time post-mixing between high concentration formulations and protein solutions to obtain a thermodynamically favourable system. This result was later applied to subsequent chapter sample preparations where an incubation of at least an hour was allowed for mixed samples prior to analysis. This can also be applied to future sample preparations not only

for biophysical analysis but also for formulation process design, ensuring product quality and patient safety.

Finally, in chapter 5 the stabilisation of GCSF with different excipients during lyophilisation was studied in the solid state. By USD lyophilisation methods, GCSF was shown to be stable in a sodium citrate buffer pH 4.25 as determined by monomer and activity retention upon reconstitution. Sorbitol collapsed the GCSF lyophilised cake in all buffer systems due to its low T_g' versus high primary drying temperature, forming a moist white cake at the bottom of the wells. Activity and monomer recovery with sorbitol was shown to be significantly improved in the acetate and PBS buffers compared to other excipients suggesting collapse of cakes is not detrimental to GCSF stability. Unfortunately, sorbitol could not be taken forward for ssHDX-MS analysis due to the collapse of the solid-state matrix, preventing D₂O vapour diffusion through the sample.

Coincidentally, collapse was found to be a major issue in the analysis of lyophilised GCSF samples using ssHDX-MS. The labelling of samples within a 43 % RH environment saw moisture induced collapse in the majority of samples studied within a few hours. Previous ssHDX-MS studies of sucrose and mannitol with mAbs have not reported observations of this occurring at the same RH; therefore USD studies were performed in the chapter in a bid to rectify this issue. The first included increasing the concentration of sucrose in the sample which did not prevent moisture induced collapse. The second included a range of different excipients within the GCSF formulation, where only phenylalanine containing samples did not collapse. By DSC, phenylalanine was shown to have a clear crystallisation event. All mannitol formulated samples collapsed during the work, which was due to the lack of anneal step during lyophilisation preventing crystallisation, as identified by large exothermic dips in heat flow by modulated DSC.

The moisture induced collapse of samples, including the control, over time was shown to drastically change the outcome of ssHDX-MS results. The differential uptake values of GCSF peptides were lessened with progression of sample collapse, removing observable stabilising and/or destabilising effects. Because of this, only the earliest labelling time points were of use for analysis and the dynamics of protein in the solid state could not be accurately measured.

The protective effects of sucrose, mannitol and phenylalanine with GCSF in the solid state were determined using early time point differential analysis of ssHDX-MS data, where it was found that sucrose had the highest level of protection, in keeping with its status as a

lyoprotectant, as well as with results of previous ssHDX-MS studies. Stabilised regions of GCSF in the solid state with excipients included all loop regions and α C. GCSF-excipient docking results from chapter 4 were compared to regions of protection identified by ssHDX-MS where it was found regions somewhat agreed, although the major region of reduced uptake was missed by the computational docking software.

Phenylalanine was identified by SEC-HPLC of pre and post-lyophilisation samples to be the only destabilising excipient of those studied, and decreased the amount of monomer in the sample by 9%. By ssHDX-MS, two peptides showed significant positive differential values relative to the control and other excipient differential uptake data, indicating increased exposure or flexibility. The peptides covered residues 33 to 42 within the end of α A and start of loopABI. Ultimately it is unknown if this region interacts with phenylalanine molecules during lyophilisation or if it becomes more exposed by allosteric effects from interactions with other regions. The second part of LoopAB (II) and short helix located in the middle of loopAB were among the regions with the highest flexibility as identified by HDX-MS in chapter 3. The loopAB region was also implicated in increased GCSF aggregation during analysis of the mutants G51R and L71W in chapter 4. Both mutants were shown to increase flexibility in the LoopAB region, and increase GCSF aggregation rates during accelerated thermal stability studies. In their work with WT GCSF, Raso et al. (2005) also identified conformational changes to LoopAB via increased fluorescence emission intensity of tryptophan at position 60 in low Gdn.HCl concentrations, at pH 7, which were linked to aggregation. Additionally, GCSF mutant studies by Luo et al. (2002) and Buchanan et al. (2012) both identified mutants with mutations located in LoopAB (along with mutations in other regions) which increased stability. In combination with the results from chapter 5, mutant data furthers the conclusion that the LoopAB and the short helix are hotspots for GCSF aggregation.

Finally, the use of USD Micronic tubes was demonstrated to drastically reduce sample volume and footprint during ssHDX-MS sample preparation and labelling. This moves the ssHDX-MS method forward towards automation of sample preparation, and potentially analysis; increasing the throughput of the relatively new technique. Due to lack of time, the chapter left the use of ssHDX-MS combined with USD sample preparation and *in silico* formulation screening at a place where the applicability of the techniques has been demonstrated, however further development work is required to reach the full potential of *in silico*, high-throughput pre-screening tools for solid state formulation development.

8.1 Future development of HDX-MS for lead candidate selection

In 2011, Yamamoto et al. (2011) developed a novel micro-reactor chip that made it possible to carry out accurate pulse labelling of deuterium and achieve millisecond order of time resolution for reliable mixing and quenching of reactions (Figure 104). The set-up consisted of a polydimethylsiloxane chip with 100 μM deep and wide microchannels. The chip was designed with two Y-shaped flow channels to allow mixing of three different solutions: protein sample, deuterium labelling solution and quench solution. The group demonstrated the use of the microfluidic device to measure the global HDX of 70S ribosome after 20 Ms of labelling.

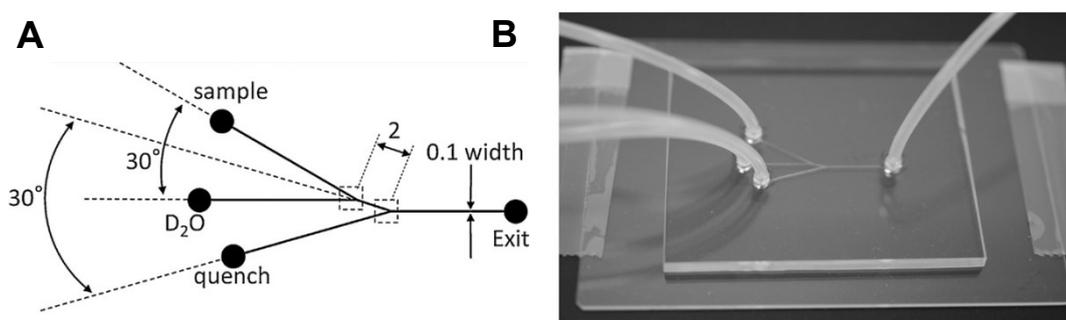


Figure 104. **HDX micro-reactor chip.** Schematic of the two Y-shaped channel microreactor chip, where the width unit is in mm. B) Image of the micro-reactor chip. Taken from Yamamoto et al. (2011).

In chapter 5, select GCSF mutants were shown to be distinguishable by the sum of their differential uptake values, which was essentially the global uptake. During early phase candidate development, multiple biotherapeutic protein structures are characterised in parallel. Due to the time constraints placed on development, high-throughput rapid analysis methods are required to meet deadlines. Future work with HDX-MS could combine the use of the micro-reactor chip HDX-MS technique to measure the global stability of GCSF mutants to determine if the strategy can also distinguish between stable and unstable structures when compared to long-term and accelerated stability studies. The removal of digestion for global analysis would also aid in the stability analysis of industrially relevant mAbs (> 150 kDa), as the high molecular weight increases the complexity of peptide mass spectra analysis due to increased number of fragment peptides with overlapping isotopic distributions. If successful, this high-throughput, 10 μL sample volume, strategy could be a powerful addition in the stability characterisation tool kit.

9 References

- Abdul-Fattah, A. M., Truong-Le, V., Yee, L., Pan, E., Ao, Y., Kalonia, D. S., & Pikal, M. J. 2007. Drying-induced variations in physico-chemical properties of amorphous pharmaceuticals and their impact on stability II: Stability of a vaccine. *Pharmaceutical research*, 24(4), pp.715-727.
- Ablinger, E., Hellweger, M., Leitgeb, S., & Zimmer, A. (2012). Evaluating the effects of buffer conditions and extremolytes on thermostability of granulocyte colony-stimulating factor using high-throughput screening combined with design of experiments. *International journal of pharmaceutics*, 436(1-2), 744-752.
- Adler-Abramovich, L., Vaks, L., Carny, O., Trudler, D., Magno, A., Caflish, A., Frenkel, D., & Gazit, E., (2012). Phenylalanine assembly into toxic fibrils suggests amyloid etiology in phenylketonuria. *Nature chemical biology*, 8(8), p.701.
- Ahn, J., Cao, M. J., Yu, Y. Q., & Engen, J. R. (2013). Accessing the reproducibility and specificity of pepsin and other aspartic proteases. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1834(6), 1222-1229.
- Ahn, J., Jung, M. C., Wyndham, K., Yu, Y. Q., & Engen, J. R. (2012). Pepsin immobilized on high-strength hybrid particles for continuous flow online digestion at 10 000 psi. *Analytical chemistry*, 84(16), 7256-7262.
- Alber, T., Sun, D. P., Nye, J. A., Muchmore, D. C., & Matthews, B. W. (1987). Temperature-sensitive mutations of bacteriophage T4 lysozyme occur at sites with low mobility and low solvent accessibility in the folded protein. *Biochemistry*, 26(13), 3754-3758.
- Alebouyeh, M., Tahzibi, A., Yaghoobzadeh, S., Zahedy, E.T., Kiumarsi, S., Soltanabad, M.H., Shahbazi, S., & Amini, H. (2016). Rapid formulation assessment of filgrastim therapeutics by a thermal stress test. *Biologicals*, 44(3), pp.150-156.
- Allison, T. M., Reading, E., Liko, I., Baldwin, A. J., Laganowsky, A., & Robinson, C. V. (2015). Quantifying the stabilizing effects of protein–ligand interactions in the gas phase. *Nature communications*, 6, 8551.
- Arakawa, T., Prestrelski, S. J., Kenney, W. C., & Carpenter, J. F. (2001). Factors affecting short-term and long-term stabilities of proteins. *Advanced drug delivery reviews*, 46(1-3), 307-326.
- Aritomi, M., Kunishima, N., Okamoto, T., Kuroki, R., Ota, Y., & Morikawa, K. (1999). Atomic structure of the G-CSF–receptor complex showing a new cytokine–receptor recognition scheme. *Nature*, 401(6754), 713-717.
- Arvedson, T. L., & Giffin, M. J. (2012). Structural biology of G-CSF and its receptor. In *Twenty Years of G-CSF* (pp. 61-82). Springer, Basel.
- Aubin, Y., Hodgson, D. J., Thach, W. B., Gingras, G., & Sauvé, S. (2015). Monitoring effects of excipients, formulation parameters and mutations on the high order structure of filgrastim by NMR. *Pharmaceutical research*, 32(10), 3365-3375.
- Babaeipour, V., Khanchezar, S., Mofid, M. R., & Abbas, M. P. H. (2015). Efficient process development of recombinant human granulocyte colony-stimulating factor (rh-G-CSF) production in *Escherichia coli*. *Iranian biomedical journal*, 19(2), 102.
- Bai, Y., Milne, J. S., Mayne, L., & Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins: Structure, Function, and Bioinformatics*, 17(1), 75-86.

- Barata, T. S., Zhang, C., Dalby, P. A., Brocchini, S., & Zloh, M. (2016). Identification of Protein–Excipient Interaction Hotspots Using Computational Approaches. *International journal of molecular sciences*, 17(6), 853.
- Bendall, L. J., & Bradstock, K. F. (2014). G-CSF: From granulopoietic stimulant to bone marrow stem cell mobilizing agent. *Cytokine & growth factor reviews*, 25(4), 355-367.
- Berkowitz, S. A., & Houde, D. J. (2015). The Complexity of Protein Structure and the Challenges it Poses in Developing Biopharmaceuticals. In *Biophysical Characterization of Proteins in Developing Biopharmaceuticals* (pp. 1-21).
- Betts, M. J., & Russell, R. B. (2003). Amino acid properties and consequences of substitutions. *Bioinformatics for geneticists*, 289-316.
- Bhambure, R., & Rathore, A. S. (2013). Chromatography process development in the quality by design paradigm I: Establishing a high-throughput process development platform as a tool for estimating “characterization space” for an ion exchange chromatography step. *Biotechnology progress*, 29(2), 403-414.
- Bishop, B., Koay, D. C., Sartorelli, A. C., & Regan, L. (2001). Reengineering granulocyte-colony stimulating factor (G-CSF) for enhanced stability. *Journal of Biological Chemistry*.
- Bou-Assaf, G. M., & Marshall, A. G. (2015). Biophysical mass spectrometry for biopharmaceutical process development: Focus on hydrogen/deuterium exchange. In *Biophysical Characterization of Proteins in Developing Biopharmaceuticals* (pp. 307-339).
- Brader, M.L., Estey, T., Bai, S., Alston, R.W., Lucas, K.K., Lantz, S., Landsman, P., & Maloney, K.M. (2015). Examination of thermal unfolding and aggregation profiles of a series of developable therapeutic monoclonal antibodies. *Molecular pharmaceutics*, 12(4), pp.1005-1017.
- Brems, D. N. (2002). The kinetics of G-CSF folding. *Protein Science*, 11(10)pp.2504-2511.
- Brier, S., Lemaire, D., DeBonis, S., Kozielski, F., & Forest, E. (2006). Use of hydrogen/deuterium exchange mass spectrometry and mutagenesis as a tool to identify the binding region of inhibitors targeting the human mitotic kinesin Eg5. *Rapid communications in mass spectrometry*, 20(3), 456-462.
- Bristow, A. F., Bird, C., Bolgiano, B., & Thorpe, R. (2012). Regulatory requirements for therapeutic proteins: the relationship between the conformation and biological activity of filgrastim. *Pharmeuropa bio & scientific notes*, 2012, 103-117
- Brokx, S., Scrocchi, L., Shah, N., & Dowd, J. (2017). A demonstration of analytical similarity comparing a proposed biosimilar pegfilgrastim and reference pegfilgrastim. *Biologicals*, 48, 28-38.
- Buchanan, A., Ferraro, F., Rust, S., Sridharan, S., Franks, R., Dean, G., McCourt, M., Jermutus, L., & Minter, R. (2012). Improved drug-like properties of therapeutic proteins by directed evolution. *Protein Engineering, Design & Selection*, 25(10), pp.631-638.
- Carpenter, J. F., & Crowe, J. H. (1989). An infrared spectroscopic study of the interactions of carbohydrates with dried proteins. *Biochemistry*, 28(9), 3916-3922.
- Carpenter, J. F., Chang, B. S., Garzon-Rodriguez, W., & Randolph, T. W. (2002). Rational design of stable lyophilized protein formulations: theory and practice. In *Rational design of stable protein formulations* (pp. 109-133). Springer, Boston, MA.
- Carpenter, J. F., Pikal, M. J., Chang, B. S., & Randolph, T. W. (1997). Rational design of stable lyophilized protein formulations: some practical advice. *Pharmaceutical research*, 14(8), 969-975.

- Chakroun, N., Hilton, D., Ahmad, S. S., Platt, G. W., & Dalby, P. A. (2016). Mapping the aggregation kinetics of a therapeutic antibody fragment. *Molecular pharmaceutics*, 13(2), 307-319.
- Chang, B. S., Randall, C. S., & Lee, Y. S. (1993). Stabilization of lyophilized porcine pancreatic elastase. *Pharmaceutical research*, 10(10), 1478-1483.
- Chi, EY., Krishnan, S., Kendrick, BS., Chang, BS., Carpenter, JF., Randolph, TW. (2003). Roles of conformational stability and colloidal stability in the aggregation of recombinant human granulocyte colony-stimulating factor. *Protein Science*, 12(5)pp.903-913.
- Clogston, C. L., Hu, S., Boone, T. C., & Lu, H. S. (1993). Glycosidase digestion, electrophoresis and chromatographic analysis of recombinant human granulocyte colony-stimulating factor glycoforms produced in Chinese hamster ovary cells. *Journal of Chromatography A*, 637(1), 55-62.
- Coales, S. J., Lee, J. E., Ma, A., Morrow, J. A., & Hamuro, Y. (2010). Expansion of time window for mass spectrometric measurement of amide hydrogen/deuterium exchange reactions. *Rapid Communications in Mass Spectrometry*, 24(24), 3585-3592.
- Codevilla, C. F., Brum Jr, L., de Oliveira, P. R., Dolman, C., Rafferty, B., & Dalmora, S. L. (2004). Validation of an SEC-HPLC Method for the Analysis of rhG-CSF in Pharmaceutical Formulations. *Journal of liquid chromatography & related technologies*, 27(17), 2689-2698.
- Colandene, J. (2007). Lyophilization cycle development for a high-concentration monoclonal antibody formulation lacking a crystalline bulking agent. *Journal of pharmaceutical sciences*, 96(6), pp.1598–1608.
- Costantino, H. R., Curley, J. G., Wu, S., & Hsu, C. C. (1998). Water sorption behavior of lyophilized protein–sugar systems and implications for solid-state interactions. *International Journal of Pharmaceutics*, 166(2), 211-221.
- Crommelin, D. J. A. (2013). Formulation of Biotech Products, Including Biopharmaceutical Considerations. Pp. 69–99 in *Pharmaceutical biotechnology*, edited by Daan J. A. Crommelin, Robert D. Sindelar, and Bernd Meibohm. New York, London: Springer.
- Cryar, A., Groves, K., & Quaglia, M. (2017). Online hydrogen-deuterium exchange traveling wave ion mobility mass spectrometry (HDX-IM-MS): a systematic evaluation. *Journal of The American Society for Mass Spectrometry*, 28(6), 1192-1202.
- Das, P., King, J. A., & Zhou, R. (2011). Aggregation of γ -crystallins associated with human cataracts via domain swapping at the C-terminal β -strands. *Proceedings of the National Academy of Sciences*, 108(26), 10514-10519.
- Dayhoff, M. O., Schwartz, R. M., & Orcutt, B. C. (1978). 22 a model of evolutionary change in proteins. *Atlas of protein sequence and structure*, 345-352.
- DbBrowser (2018) The single-letter amino acid code. <http://130.88.97.239/bioactivity/aacodefrm.html>. Accessed 07 September 2018
- Duralliu, A., Matejtschuk, P., & Williams, D. R. (2018). Humidity induced collapse in freeze dried cakes: A direct visualization study using DVS. *European Journal of Pharmaceutics and Biopharmaceutics*, 127, 29-36.
- Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry*, 6(7), 1948-1954.
- Engen, J. R. (2009). Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS.

Eschweiler, J. D., Rabuck-Gibbons, J. N., Tian, Y., & Ruotolo, B. T. (2015). CIUSuite: a quantitative analysis package for collision induced unfolding measurements of gas-phase protein ions. *Analytical chemistry*, 87(22), 11516-11522.

European Pharmacopeia 6.3 (1999) Concentrated Filgrastim Solution <http://varban.airbase.ru/old-airbase/EPh/2206e.pdf>. Accessed 09 March 2018

Ewing, T. J., Makino, S., Skillman, A. G., & Kuntz, I. D. (2001). DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *Journal of computer-aided molecular design*, 15(5), 411-428.

Fang, J., Yu, Y. Q., Lawler, R., & Berger, S. (2015) Pressurized online pepsin digestion of proteins for hydrogen/deuterium exchange mass spectrometry. http://www.waters.com/webassets/cms/library/docs/hdx_high_pressure_hos_2015.pdf. Last accessed 13 September 2018.

Fauchere, J. L., & Pliska, V. (1983). Hydrophobic parameters π of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides. *Eur. J. Med. Chem*, 18(3), 369-375.

Fekete, S., & Guillarme, D. (2015). Estimation of pressure-, temperature- and frictional heating-related effects on proteins' retention under ultra-high-pressure liquid chromatographic conditions. *Journal of Chromatography A*, 1393, 73-80.

Feng, X. L., Zhao, X., Yu, H., Sun, T. D., & Huang, X. R. (2013). Molecular dynamics simulations of the thermal stability of tteRBP and ecRBP. *Journal of Biomolecular Structure and Dynamics*, 31(10), 1086-1100.

Fersht, A. (1999). *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*. Macmillan.

Fersht, A. R., & Serrano, L. (1993). Principles of protein stability derived from protein engineering experiments. *Current Opinion in Structural Biology*, 3(1), 75-83.

Fissore, D., Pisano, R., & Barresi, A. A. (2018). Process analytical technology for monitoring pharmaceuticals freeze-drying—A comprehensive review. *Drying Technology*, 1-27.

Foerst, P., Reitmaier, J., & Kulozik, U. (2010). ¹H NMR investigation on the role of sorbitol for the survival of *Lactobacillus paracasei* ssp. *paracasei* in vacuum-dried preparations. *Journal of applied microbiology*, 108(3), 841-850.

Frans, F. (1998). Freeze-drying of bioproducts: putting principles into practice. *European journal of pharmaceuticals and biopharmaceutics: official journal of Arbeitsgemeinschaft fur Pharmazeutische Verfahrenstechnik eV*, 45(3), 221-229.

Frauenfelder, H., Sligar, S. G., & Wolynes, P. G. (1991). The energy landscapes and motions of proteins. *Science*, 254(5038), 1598-1603.

Gao, K., Rao, C., Tao, L., Han, C., Shi, X., Wang, L., Fan, W., Yu, L., & Wang, J. (2012). Development and calibration of a standard for the protein content of granulocyte colony-stimulating factor products. *Biologicals*, 40(2), 151-157.

Gekko, K., & Morikawa, T. (1981A). Thermodynamics of polyol-induced thermal stabilization of chymotrypsinogen. *The Journal of Biochemistry*, 90(1), 51-60.

Gekko, K., & Morikawa, T. (1981B). Preferential hydration of bovine serum albumin in polyhydric alcohol-water mixtures. *The Journal of Biochemistry*, 90(1), 39-50.

Gervais, V., Zertal, A., & Oschkinat, H. (1997). NMR Investigations of the Role of the Sugar Moiety in Glycosylated Recombinant Human Granulocyte-Colony-Stimulating Factor. *The FEBS Journal*, 247(1), 386-395.

Gill, S. C., & Von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Analytical biochemistry*, 182(2), 319-326.

Goodford, P. J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *Journal of medicinal chemistry*, 28(7), 849-857.

Goodford, P. J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *Journal of medicinal chemistry*, 28(7), 849-857.

Goodsell, D. S., Morris, G. M., & Olson, A. J. (1996). Automated docking of flexible ligands: applications of AutoDock. *Journal of Molecular Recognition*, 9(1), 1-5.

Grant, Y. G. (2011). Engineering the rational design and optimisation of lyophilization processes for biological materials (Doctoral dissertation, UCL (University College London)).

Grant, Y., Matejtschuk, P., & Dalby, P. A. (2009). Rapid optimization of protein freeze-drying formulations using ultra scale-down and factorial design of experiment in microplates. *Biotechnology and bioengineering*, 104(5), 957-964.

Grant, Y., Matejtschuk, P., Bird, C., Wadhwa, M., & Dalby, P. A. (2012). Freeze drying formulation using microscale and design of experiment approaches: a case study using granulocyte colony-stimulating factor. *Biotechnology letters*, 34(4), 641-648.

Henzler-Wildman, K., & Kern, D. (2007). Dynamic personalities of proteins. *Nature*, 450(7172), 964.

Herman, A. C., Boone, T. C., & Pearlman, H. L. R and YJ. (1996). Characterization, Formulation, and Stability of Neupogen (Filgrastim), a Recombinant Human Granulocyte-Colony Stimulating Factor. *Formulation, Characterization, and Stability of Protein Drugs: Case Histories*.

Herman, A. C., Boone, T. C., & Lu, H. S. (2002). Characterization, formulation, and stability of Neupogen®(Filgrastim), a recombinant human granulocyte-colony stimulating factor. In *Formulation, characterization, and stability of protein drugs: case histories* (pp. 303-328). Springer, Boston, MA.

Highsmith, J. (2012). *Biologic therapeutic drugs: technologies and global markets*. BCC Research.

Hill, C. P., Osslund, T. D., & Eisenberg, D. (1993). The structure of granulocyte-colony-stimulating factor and its relationship to other growth factors. *Proceedings of the National Academy of Sciences*, 90(11), 5167-5171.

Houde, D. J., & Berkowitz, S. A. (2015A). Biopharmaceutical Industry's Biophysical Toolbox. In *Biophysical Characterization of Proteins in Developing Biopharmaceuticals* (pp. 49-78).

Houde, D. J., & Berkowitz, S. A. (2015B). An Introduction and Hierarchical Organization of the Biophysical Tool in Section II. In *Biophysical Characterization of Proteins in Developing Biopharmaceuticals* (pp. 79-85).

Houde, D., & Engen, J. R. (2013). Conformational analysis of recombinant monoclonal antibodies with hydrogen/deuterium exchange mass spectrometry. In *Glycosylation Engineering of Biopharmaceuticals* (pp. 269-289). Humana Press, Totowa, NJ.

Ishibashi, M., Tsumoto, K., Tokunaga, M., Ejima, D., Kita, Y., & Arakawa, T. (2005). Is arginine a protein-denaturant?. *Protein expression and purification*, 42(1), 1-6.

- Ishikawa, M., Iijima, H., Satake-Ishikawa, R., Tsumura, H., Iwamatsu, A., Kadoya, T., Shimada, Y., Fukamachi, H., Kobayashi, K., Matsuki, S., & Asano, K. (1992). The Substitution of Cysteine 17 of Recombinant HumanG-CSF with Alanine Greatly Enhanced its Stability. *Cell structure and function*, 17(1), 61-65
- Jiang, Y., Jiang, W., Qiu, Y., & Dai, W. (2011A). Effect of a structurally modified human granulocyte colony stimulating factor, G-CSFa, on leukopenia in mice and monkeys. *Journal of hematology & oncology*, 4(1), 28.
- Jiang, W., Kim, S., Zhang, X., Lionberger, R. A., Davit, B. M., Conner, D. P., & Lawrence, X. Y. (2011B). The role of predictive biopharmaceutical modeling and simulation in drug development and regulatory evaluation. *International journal of pharmaceutics*, 418(2), 151-160.
- Jorgensen, L., Hostrup, S., Moeller, E. H., & Grohgan, H. (2009). Recent trends in stabilising peptides and proteins in pharmaceutical formulation—considerations in the choice of excipients. *Expert opinion on drug delivery*, 6(11), 1219-1230.
- Kadoya, S., Fujii, K., Izutsu, K. I., Yonemochi, E., Terada, K., Yomota, C., & Kawanishi, T. (2010). Freeze-drying of proteins with glass-forming oligosaccharide-derived sugar alcohols. *International journal of pharmaceutics*, 389(1-2), 107-113.
- Kaltashov, I. A., Bobst, C. E., Nguyen, S. N., & Wang, S. (2013). Emerging mass spectrometry-based approaches to probe protein–receptor interactions: Focus on overcoming physiological barriers. *Advanced drug delivery reviews*, 65(8), 1020-1030.
- Kamerzell, T. J., & Middaugh, C. R. (2008). The complex inter-relationships between protein flexibility and stability. *Journal of pharmaceutical sciences*, 97(9), 3494-3517.
- Kamerzell, T. J., Esfandiary, R., Joshi, S. B., Middaugh, C. R., & Volkin, D. B. (2011). Protein–excipient interactions: Mechanisms and biophysical characterization applied to protein formulation development. *Advanced drug delivery reviews*, 63(13), 1118-1159.
- Kang, S. H., Na, K. H., Park, J. H., Park, C. I., Lee, S. Y., & Lee, Y. I. (1995). High level expression and simple purification of recombinant human granulocyte colony-stimulating factor in *E. coli*. *Biotechnology letters*, 17(7), 687-692.
- Kaushik, J. K., & Bhat, R. (2003). Why is trehalose an exceptional protein stabilizer?: An analysis of the thermal stability of proteins in the presence of compatible osmolyte trehalose. *Journal of Biological Chemistry*.
- Kellogg, E. H., Leaver-Fay, A., & Baker, D. (2011). Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins: Structure, Function, and Bioinformatics*, 79(3), 830-838.
- Kendrick, B. S., Chang, B. S., Arakawa, T., Peterson, B., Randolph, T. W., Manning, M. C., & Carpenter, J. F. (1997). Preferential exclusion of sucrose from recombinant interleukin-1 receptor antagonist: role in restricted conformational mobility and compaction of native state. *Proceedings of the National Academy of Sciences*, 94(22), 11917-11922.
- Kessler, M., Goldsmith, D., & Schellekens, H. (2006). Immunogenicity of biopharmaceuticals. *Nephrology Dialysis Transplantation*, 21(suppl_5), v9-v12.
- Kim, H. S., Le, Q. A. T., & Kim, Y. H. (2010). Development of thermostable lipase B from *Candida antarctica* (CaLB) through in silico design employing B-factor and RosettaDesign. *Enzyme and Microbial Technology*, 47(1-2), 1-5.
- Koide, H., Muto, Y., Kasai, H., Kohri, K., Hoshi, K., Takahashi, S., Tsukumo, K.I., Sasaki, T., Oka, T., Miyake, T., & Fuwa, T. (1992). A site-directed mutagenesis study on the role of isoleucine-23 of

human epidermal growth factor in the receptor binding. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, 1120(3), 257-261.

Kolvenbach, C. G., Narhi, L. O., Philo, J. S., Li, T., Zhang, M., & Arakawa, T. (1997). Granulocyte-colony stimulating factor maintains a thermally stable, compact, partially folded structure at pH 2. *The Journal of peptide research*, 50(4), 310-318.

Konermann, L., Pan, J., & Liu, Y. H. (2011). Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chemical Society Reviews*, 40(3), 1224-1234.

Krasucka, D. M., Kos, K. A. T. A. R. Z. Y. N. A., Cybulski, W. A., Mitura, A., Lysiak, E., & Pietron, W. J. (2012). Karl Fisher determination of residual moisture in veterinary vaccines—practical implementation in market monitoring. *Acta Pol. Pharm*, 69(6), 1364-7.

Krishnan, S., Chi, E. Y., Webb, J. N., Chang, B. S., Shan, D., Goldenberg, M., Manning, M.C., Randolph, T.W., & Carpenter, J. F. (2002). Aggregation of granulocyte colony stimulating factor under physiological conditions: characterization and thermodynamic inhibition. *Biochemistry*, 41(20), 6422-6431.

Kuga, T., Komatsu, Y., Yamasaki, M., Sekine, S., Miyaji, H., Nishi, T., Sato, M., Yokoo, Y., Asano, M., Okabe, M., & Morimoto, M. (1989). Mutagenesis of human granulocyte colony stimulating factor. *Biochemical and biophysical research communications*, 159(1), pp.103-111.

Lawrence, X. Y. (2008). Pharmaceutical quality by design: product and process development, understanding, and control. *Pharmaceutical research*, 25(4), 781-791.

Lee, J. C., & Timasheff, S. N. (1974). Partial specific volumes and interactions with solvent components of proteins in guanidine hydrochloride. *Biochemistry*, 13(2), 257-265.

Lee, J. C., & Timasheff, S. N. (1981). The stabilization of proteins by sucrose. *Journal of Biological Chemistry*, 256(14), 7193-7201.

Leone, M., Di Lello, P., Ohlenschläger, O., Pedone, E.M., Bartolucci, S., Rossi, M., Di Blasio, B., Pedone, C., Saviano, M., Isernia, C., & Fattorusso, R. (2004). Solution structure and backbone dynamics of the K18G/R82E *Alicyclobacillus acidocaldarius* thioredoxin mutant: a molecular analysis of its reduced thermal stability. *Biochemistry*, 43(20), pp.6043-6058.

Leopold, P. E., Montal, M., & Onuchic, J. N. (1992). Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proceedings of the National Academy of Sciences*, 89(18), 8721-8725.

Levy, M. J., Gucinski, A. C., Sommers, C. D., Ghasriani, H., Wang, B., Keire, D. A., & Boyne, M. T. (2014). Analytical techniques and bioactivity assays to compare the structure and function of filgrastim (granulocyte-colony stimulating factor) therapeutics from different manufacturers. *Analytical and bioanalytical chemistry*, 406(26), 6559-6567.

Li, C., Wang, J. X., Le, Y., & Chen, J. F. (2013). Studies of Bicalutamide–Excipients Interaction by Combination of Molecular Docking and Molecular Dynamics Simulation. *Molecular pharmaceutics*, 10(6), 2362-2369.

Li, Y., Williams, T. D., & Topp, E. M. (2008). Effects of excipients on protein conformation in lyophilized solids by hydrogen/deuterium exchange mass spectrometry. *Pharmaceutical research*, 25(2), 259-267.

Li, Y., Williams, T. D., Schowen, R. L., & Topp, E. M. (2007). Characterizing protein structure in amorphous solids using hydrogen/deuterium exchange with mass spectrometry. *Analytical biochemistry*, 366(1), 18-28.

Linderstrøm-Lang, K. (1958). Deuterium exchange and protein structure. In *Symposium on protein structure* (pp. 23-34). Methuen London.

Lipscomb, L. A., Gassner, N. C., Snow, S. D., Eldridge, A. M., Baase, W. A., Drew, D. L., & Matthews, B. W. (1998). Context-dependent protein stabilization by methionine-to-leucine substitution shown in T4 lysozyme. *Protein science*, 7(3), 765-773.

Liu, Y., & Kuhlman, B. (2006). RosettaDesign server for protein design. *Nucleic acids research*, 34(suppl_2), W235-W238.

Lohitha (2017) Hydrogen Deuterium exchange mass spectrometry (HDX-MS). <https://www.slideshare.net/KarumazziLohitha/hydrogen-deuterium-exchange-mass-spectrometry-hdxms>. Accessed 16 September 2018

Lovejoy, B., Cascio, D., & Eisenberg, D. (1993). Crystal structure of canine and bovine granulocyte-colony stimulating factor (G-CSF). *Journal of molecular biology*, 234(3), 640-653.

Luo, P., Hayes, R.J., Chan, C., Stark, D.M., Hwang, M.Y., Jacinto, J.M., Juvvadi, P., Chung, H.S., Kundu, A., Ary, M.L. & Dahiyat, B.I. (2002). Development of a cytokine analog with enhanced stability using computational ultrahigh throughput screening. *Protein science*, 11(5), pp.1218-1226.

Maddux, N.R., Iyer, V., Cheng, W., Youssef, A.M., Joshi, S.B., Volkin, D.B., Ralston, J.P., Winter, G., & Middaugh, C.R. (2014). High throughput prediction of the long-term stability of pharmaceutical macromolecules from short-term multi-instrument spectroscopic data. *Journal of pharmaceutical sciences*, 103(3), pp.828-839.

Maggio, E. T. (2016). Polysorbates, peroxides, protein aggregation, and immunogenicity—a growing concern. *Journal of Excipients and Food Chemicals*, 3(2).

Mamonova, T. B., Glyakina, A. V., Galzitskaya, O. V., & Kurnikova, M. G. (2013). Stability and rigidity/flexibility—Two sides of the same coin?. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1834(5), 854-866.

Manikwar, P., Majumdar, R., Hickey, J.M., Thakkar, S.V., Samra, H.S., Sathish, H.A., Bishop, S.M., Middaugh, C.R., Weis, D.D., & Volkin, D.B. (2013). Correlating excipient effects on conformational and storage stability of an IgG1 monoclonal antibody with local dynamics as measured by hydrogen/deuterium-exchange mass spectrometry. *Journal of pharmaceutical sciences*, 102(7), pp.2136-2151.

Manning, M. C., Chou, D. K., Murphy, B. M., Payne, R. W., & Katayama, D. S. (2010). Stability of protein pharmaceuticals: an update. *Pharmaceutical research*, 27(4), 544-575.

Marshall, S. A., Lazar, G. A., Chirino, A. J., & Desjarlais, J. R. (2003). Rational design and engineering of therapeutic proteins. *Drug discovery today*, 8(5), 212-221.

Matejtschuk, P. (2007). Lyophilization of proteins. In *Cryopreservation and Freeze-Drying Protocols* (pp. 59-72). Humana Press.

Mattern, M., Winter, G., Kohnert, U., & Lee, G. (1999). Formulation of proteins in vacuum-dried glasses. II. Process and storage stability in sugar-free amino acid systems. *Pharmaceutical development and technology*, 4(2), 199-208.

Michaelis, U., Rudolph, R., Winter, G., & Woog, H. (1999). U.S. Patent No. 5,919,443. Washington, DC: U.S. Patent and Trademark Office

- Mo, J., Tymiak, A. A., & Chen, G. (2013). Characterization of disulfide linkages in recombinant human granulocyte-colony stimulating factor. *Rapid Communications in Mass Spectrometry*, 27(9), 940-946.
- Moorthy, B. S., Iyer, L. K., & Topp, E. M. (2015). Mass spectrometric approaches to study protein structure and interactions in lyophilized powders. *Journal of visualized experiments: JoVE*, (98).
- Moorthy, B. S., Schultz, S. G., Kim, S. G., & Topp, E. M. (2014). Predicting protein aggregation during storage in lyophilized solids using solid state amide hydrogen/deuterium exchange with mass spectrometric analysis (ssHDX-MS). *Molecular pharmaceutics*, 11(6), 1869-1879.
- Moorthy, B. S., Zarraga, I. E., Kumar, L., Walters, B. T., Goldbach, P., Topp, E. M., & Allmendinger, A. (2017). Solid-state hydrogen–deuterium exchange mass spectrometry: correlation of deuterium uptake and long-term stability of lyophilized monoclonal antibody formulations. *Molecular pharmaceutics*, 15(1), 1-11.
- Moroco, J. A., & Engen, J. R. (2015). Replication in bioanalytical studies with HDX MS: aim as high as possible. *Bioanalysis*, 7(9), 1065-1067.
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. and Olson, A. J. (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Computational Chemistry* 2009, 16: 2785-91.
- Morris, P., Rios-Solis, L., García-Arrazola, R., Lye, G. J., & Dalby, P. A. (2016). Impact of cofactor-binding loop mutations on thermotolerance and activity of *E. coli* transketolase. *Enzyme and microbial technology*, 89, 85-91.
- Mott, H. R., & Campbell, I. D. (1995). Four-helix bundle growth factors and their receptors: protein-protein interactions. *Current opinion in structural biology*, 5(1), 114-121.
- Moussa, E. M., Singh, S. K., Kimmel, M., Nema, S., & Topp, E. M. (2018). Probing the Conformation of an IgG1 Monoclonal Antibody in Lyophilized Solids Using Solid-State Hydrogen–Deuterium Exchange with Mass Spectrometric Analysis (ssHDX-MS). *Molecular pharmaceutics*, 15(2), 356-368.
- Moussa, E. M., Wilson, N. E., Zhou, Q. T., Singh, S. K., Nema, S., & Topp, E. M. (2018). Effects of Drying Process on an IgG1 Monoclonal Antibody Using Solid-State Hydrogen Deuterium Exchange with Mass Spectrometric Analysis (ssHDX-MS). *Pharmaceutical research*, 35(1), 12.
- NIBSC (2013) WHO International Standard 2nd International Standard for Granulocyte Colony Stimulating Factor (Human rDNA derived) <http://www.nibsc.org/documents/ifu/09-136.pdf>. Last accessed 26 August 2018
- Nölting, B. (2006). Physical interactions that determine the properties of proteins (pp. 17-25). Springer Berlin Heidelberg.
- Novatia (2017) ESI Adduct Ions. <http://www.enovatia.com/services/ms/ms-resources/esi-adduct-ions/>. Last accessed 11 December 2017
- Novikov, F. N., & Chilov, G. G. (2009). Molecular docking: theoretical background, practical applications and perspectives. *Mendeleev Communications*, 5(19), 237-242.
- Oheda, M., Hase, S., Ono, M., & Ikenaka, T. (1988). Structures of the sugar chains of recombinant human granulocyte-colony-stimulating factor produced by Chinese hamster ovary cells. *The Journal of Biochemistry*, 103(3), 544-546.
- Ohtake, S., Kita, Y., & Arakawa, T. (2011). Interactions of formulation excipients with proteins in solution and in the dried state. *Advanced drug delivery reviews*, 63(13), 1053-1073.

- Ono, M. (1994). Physicochemical and biochemical characteristics of glycosylated recombinant human granulocyte colony stimulating factor (lenograstim). *European journal of cancer (Oxford, England: 1990)*, 30, S7-11.
- Palmer III, A. G. (2004). NMR characterization of the dynamics of biomacromolecules. *Chemical reviews*, 104(8), 3623-3640.
- Passot, S., Fonseca, F., Barbouche, N., Marin, M., Alarcon-Lorca, M., Rolland, D., & Rapaud, M. (2007). Effect of product temperature during primary drying on the long-term stability of lyophilized proteins. *Pharmaceutical development and technology*, 12(6), 543-553.
- Patapoff, T. W., & Overcashier, D. E. (2002). The importance of freezing on lyophilization cycle development. *BIOPHARM-EUGENE-*, 15(3), 16-21.
- Patel, J., Nadine Ritter, M., Ruchi Kothari, R. T., & Binita Tunga, S. (2011). Stability considerations for biopharmaceuticals: overview of protein and peptide degradation pathways. *BioProcess International*, 9, 2-11.
- Perchiacca, J. M., & Tessier, P. M. (2012). Engineering aggregation-resistant antibodies. *Annual review of chemical and biomolecular engineering*, 3, 263-286.
- Pikal, M. J. (1990). Freeze-drying of proteins. Part II: Formulation selection. *BioPharm*, 3(9), 26-30.
- Prakash, V., Loucheux, C., Scheufele, S., Gorbunoff, M. J., & Timasheff, S. N. (1981). Interactions of proteins with solvent components in 8 M urea. *Archives of Biochemistry and Biophysics*, 210(2), 455-464.
- Prestrelski, S. J., Tedeschi, N., Arakawa, T., & Carpenter, J. F. (1993). Dehydration-induced conformational transitions in proteins and their inhibition by stabilizers. *Biophysical journal*, 65(2), 661-671.
- Privalov, P. L. (1979). Stability of proteins small globular proteins. In *Advances in protein chemistry* (Vol. 33, pp. 167-241). Academic Press.
- Querol, E., Perez-Pons, J. A., & Mozo-Villarias, A. (1996). Analysis of protein conformational characteristics related to thermostability. *Protein Engineering, Design and Selection*, 9(3), 265-271.
- Quint, S., Widmaier, S., Minde, D., Hornburg, D., Langosch, D., & Scharnagl, C. (2010). Residue-specific side-chain packing determines the backbone dynamics of transmembrane model helices. *Biophysical journal*, 99(8), 2541-2549.
- Raso, S.W., Abel, J., Barnes, J.M., Maloney, K.M., Pipes, G., Treuheit, M.J., King, J., & Brems, D.N. (2005). Aggregation of granulocyte-colony stimulating factor in vitro involves a conformationally altered monomeric state. *Protein science*, 14(9), pp.2246-2257.
- Razvi, A., & Scholtz, J. M. (2006). Lessons in stability from thermophilic proteins. *Protein Science*, 15(7), 1569-1578.
- Reetz, M. T., Carballeira, J. D., & Vogel, A. (2006). Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angewandte Chemie*, 118(46), 7909-7915.
- Ribarska, J., Jolevska, S., Panovska, A., & Dimitrovska, A. (2008). Studying the formation of aggregates in recombinant human granulocyte-colony stimulating factor (rHuG-CSF), lenograstim, using size-exclusion chromatography and SDS-PAGE. *Acta Pharmaceutica*, 58(2), 199-206.

Robinson, M. J., Matejtschuk, P., Bristow, A. F., & Dalby, P. A. (2017). T_m-Values and Unfolded Fraction Can Predict Aggregation Rates for Granulocyte Colony Stimulating Factor Variant Formulations but Not under Predominantly Native Conditions. *Molecular pharmaceuticals*, 15(1), 256-267.

Robinson, M. J. (2017) unpublished (Doctoral dissertation, UCL (University College London)).

Roessl, U., Wiesbauer, J., Leitgeb, S., Birner-Gruenberger, R., & Nidetzky, B. (2012). Non-native aggregation of recombinant human granulocyte-colony stimulating factor under simulated process stress conditions. *Biotechnology journal*, 7(8), 1014-1024.

Roos, Y. H. (2010). 25 Crystallization, Collapse, and Glass Transition in Low-Water Food Systems. *Water Properties in Food, Health, Pharmaceutical and Biological Systems: ISOPOW 10*, 335.

Rose, P. W., Bi, C., Bluhm, W. F., Christie, C. H., Dimitropoulos, D., Dutta, S., Green, R.K., Goodsell, D.S., Prlić, A., Quesada, M., & Quinn, G. B. (2012). The RCSB Protein Data Bank: new resources for research and education. *Nucleic acids research*, 41(D1), D475-D482.

Rowe, R. C., Sheskey, P. J., & Owen, S. C. (Eds.). (2006). *Handbook of pharmaceutical excipients* (Vol. 6). London: Pharmaceutical press.

Rudolph, R. (1996). Successful protein folding on an industrial scale. *Protein engineering: principles and practice*, 283-298.

Saffell-Clemmer, W (2015) Formulation and lyophilisation cycle development of an ADC. https://www.baxterbiopharmasolutions.com/pdf/920955-00_Formulation_ADC_Case_Study_FINAL.pdf. Last Accessed 23 August 2018

Sandoz. (2015). FDA Advisory Committee briefing document. <http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/OncologicDrugsAdvisoryCommittee/UCM428782.pdf>. Last accessed 16 September 2015

Sarkar, C. A., Lowenhaupt, K., Horan, T., Boone, T. C., Tidor, B., & Lauffenburger, D. A. (2002). Rational cytokine design for increased lifetime and enhanced potency using pH-activated "histidine switching". *Nature biotechnology*, 20(9), 908.

Sato, Y. (2005). U.S. Patent No. 6,908,610. Washington, DC: U.S. Patent and Trademark Office.

Schneider, C. P., & Trout, B. L. (2009). Investigation of cosolute– protein preferential interaction coefficients: New insight into the mechanism by which arginine inhibits aggregation. *The Journal of Physical Chemistry B*, 113(7), 2050-2058.

Shin, H., & Cho, B. K. (2015). Rational protein engineering guided by deep mutational scanning. *International journal of molecular sciences*, 16(9), 23094-23110.

Shukla, D., & Trout, B. L. (2011). Understanding the synergistic effect of arginine and glutamic acid mixtures on protein solubility. *The Journal of Physical Chemistry B*, 115(41), 11831-11839.

Singh, V., Rai, R. K., Arora, A., Sinha, N., & Thakur, A. K. (2014). Therapeutic implication of L-phenylalanine aggregation mechanism and its modulation by D-phenylalanine in phenylketonuria. *Scientific reports*, 4, 3875

Sinha, S., Li, Y., Williams, T. D., & Topp, E. M. (2008). Protein conformation in amorphous solids by FTIR and by hydrogen/deuterium exchange with mass spectrometry. *Biophysical journal*, 95(12), 5951-5961.

Skrlin, A., Radic, I., Vuletic, M., Schwinke, D., Runac, D., Kusalic, T., Paskvan, I., Krsic, M., Bratos, M., & Marinc, S. (2010). Comparison of the physicochemical properties of a biosimilar filgrastim with those of reference filgrastim. *Biologicals*, 38(5), pp.557-566.

Sophocleous, A. M., & Topp, E. M. (2012). Localized Hydration in Lyophilized Myoglobin by Hydrogen–Deuterium Exchange Mass Spectrometry. 2. Exchange Kinetics. *Molecular pharmaceuticals*, 9(4), 727-733.

Sophocleous, A. M., Zhang, J., & Topp, E. M. (2012). Localized Hydration in Lyophilized Myoglobin by Hydrogen–Deuterium Exchange Mass Spectrometry. 1. Exchange Mapping. *Molecular pharmaceuticals*, 9(4), 718-726.

Sopkova, J., Vincent, M., Takahashi, M., Lewit-Bentley, A., & Gallay, J. (1998). Conformational flexibility of domain III of annexin V studied by fluorescence of tryptophan 187 and circular dichroism: the effect of pH. *Biochemistry*, 37(34), 11962-11970.

Sprung, J. (2012) Need a lyophilizer read this before you buy. <https://www.labconco.com/articles/need-a-lyophilizer-read-this-before-you-buy>. Accessed 26 August 2018

Suzuki, T., Imamura, K., Fujimoto, H., & Okazaki, M. (1998). Relation between thermal stabilizing effect of sucrose on LDH and sucrose-LDH hydrogen bond. *Journal of chemical engineering of Japan*, 31(4), 565-570.

Tamada, T., Honjo, E., Maeda, Y., Okamoto, T., Ishibashi, M., Tokunaga, M., & Kuroki, R. (2006). Homodimeric cross-over structure of the human granulocyte colony-stimulating factor (G-CSF) receptor signaling complex. *Proceedings of the National Academy of Sciences*, 103(9), 3135-3140.

Tang, K. E., & Dill, K. A. (1998). Native protein fluctuations: the conformational-motion temperature and the inverse correlation of protein flexibility with protein stability. *Journal of Biomolecular Structure and Dynamics*, 16(2), 397-411.

Tang, X. C., & Pikal, M. J. (2004). Design of freeze-drying processes for pharmaceuticals: practical advice. *Pharmaceutical research*, 21(2), 191-200.

Taylor, W. R. (1986). The classification of amino acid conservation. *Journal of theoretical Biology*, 119(2), 205-218.

Teilum, K., Olsen, J. G., & Kragelund, B. B. (2009). Functional aspects of protein flexibility. *Cellular and Molecular Life Sciences*, 66(14), 2231.

Timasheff, S. N. (2002). Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components. *Proceedings of the National Academy of Sciences*, 99(15), 9721-9726.

Timasheff, S. N., & Arakawa, T. (1988). Mechanism of protein precipitation and stabilization by co-solvents. *Journal of Crystal Growth*, 90(1-3), 39-46.

Treuheit, M. J., Kosky, A. A., & Brems, D. N. (2002). Inverse relationship of protein concentration and aggregation. *Pharmaceutical research*, 19(4), 511-516.

Tsuchida, D., Yamazaki, K., & Akashi, S. (2014). Characterization of stress-exposed granulocyte colony stimulating factor using ELISA and hydrogen/deuterium exchange mass spectrometry. *Journal of The American Society for Mass Spectrometry*, 25(10), 1747-1754.

Ullrich, S., Seyferth, S., & Lee, G. (2015). Measurement of Shrinkage and Cracking in Lyophilized Amorphous Cakes. Part I: Final-Product Assessment. *Journal of pharmaceutical sciences*, 104(1), 155-164.

Vemula, S., Thunuguntla, R., Dedaniya, A., Kokkiligadda, S., Palle, C., & Ronda, S. R. (2015). Improved production and characterization of recombinant human granulocyte colony stimulating factor from *E. coli* under optimized downstream processes. *Protein expression and purification*, 108, 62-72.

Wadhwa, M., Bird, C., Hamill, M., Heath, A. B., Matejtschuk, P., & Thorpe, R. (2011). The 2nd International Standard for human granulocyte colony stimulating factor. *Journal of immunological methods*, 367(1-2), 63-69.

Wales, T. E., Fadgen, K. E., Gerhardt, G. C., & Engen, J. R. (2008). High-speed and high-resolution UPLC separation at zero degrees Celsius. *Analytical chemistry*, 80(17), 6815-6820.

Wang, D. Q., Hey, J. M., & Nail, S. L. (2004). Effect of collapse on the stability of freeze-dried recombinant factor VIII and α -amylase. *Journal of pharmaceutical sciences*, 93(5), 1253-1263.

Wang, W. (2005). Protein aggregation and its inhibition in biopharmaceutics. *International journal of pharmaceutics*, 289(1-2), 1-30.

Waters (2018) <http://www.waters.com/webassets/cms/support/docs/715004032ra.pdf> ACQUITY UPLC M-Class with HDX Technology System Guide. Last accessed 13 September 2018

Wei, H., Ahn, J., Yu, Y. Q., Tymiak, A., Engen, J. R., & Chen, G. (2012). Using hydrogen/deuterium exchange mass spectrometry to study conformational changes in granulocyte colony stimulating factor upon PEGylation. *Journal of the American Society for Mass Spectrometry*, 23(3), 498-504.

Wei, H., Mo, J., Tao, L., Russell, R.J., Tymiak, A.A., Chen, G., Iacob, R.E., and Engen, J.R. (2014). Hydrogen/deuterium exchange mass spectrometry for probing higher order structure of protein therapeutics: methodology and applications. *Drug discovery today*, 19(1), pp.95-102.

Weiss IV, W. F., Young, T. M., & Roberts, C. J. (2009). Principles, approaches, and challenges for predicting protein aggregation rates and shelf life. *Journal of pharmaceutical sciences*, 98(4), 1246-1277.

Welte, K. (2012). Discovery of G-CSF and early clinical studies. In *Twenty years of G-CSF* (pp. 15-24). Springer, Basel.

WHO (2018) <http://www.who.int/biologicals/biotherapeutics/biotherapeutic-products/en/>. Accessed 15 September 2018.

Wunderlich, M., Martin, A., Staab, C. A., & Schmid, F. X. (2005). Evolutionary protein stabilization in comparison with computational design. *Journal of molecular biology*, 351(5), 1160-1168.

Xie, G., & Timasheff, S. N. (1997). Mechanism of the stabilization of ribonuclease A by sorbitol: preferential hydration is greater for the denatured than for the native protein. *Protein Science*, 6(1), 211-221.

Yamamoto, T., Shimizu, Y., Ueda, T., Shiro, Y., & Suematsu, M. (2011). Application of micro-reactor chip technique for millisecond quenching of deuterium incorporation into 70S ribosomal protein complex. *International Journal of Mass Spectrometry*, 302(1-3), 132-138.

Yang, J. M., & Chen, C. C. (2004). GEMDOCK: a generic evolutionary method for molecular docking. *Proteins: Structure, Function, and Bioinformatics*, 55(2), 288-304.

- Yu, H., Yan, Y., Zhang, C., & Dalby, P. A. (2017). Two strategies to engineer flexible loops for improved enzyme thermostability. *Scientific reports*, 7, 41212.
- Yu, L. (2001). Amorphous pharmaceutical solids: preparation, characterization and stabilization. *Advanced drug delivery reviews*, 48(1), pp.27–42.
- Zhang, C., Samad, M., Yu, H., Chakroun, N., Hilton, D., & Dalby, P. A. (2018). Computational-design to reduce conformational flexibility and aggregation rates of an antibody Fab fragment. *Molecular pharmaceuticals*.
- Zhang, C. (2017). Freeze-drying of engineered proteins using protein modelling tools and experimental validation (Doctoral dissertation, UCL (University College London)).
- Zhang, J., Banks, D. D., He, F., Treuheit, M. J., & Becker, G. W. (2015). Effects of sucrose and benzyl alcohol on GCSF conformational dynamics revealed by hydrogen deuterium exchange mass spectrometry. *Journal of pharmaceutical sciences*, 104(5), 1592-1600.
- Zhang, Z., Zhang, A., & Xiao, G. (2012). Improved protein hydrogen/deuterium exchange mass spectrometry platform with fully automated data processing. *Analytical chemistry*, 84(11), 4942-4949.
- Zhou, D., Zhang, G. G., Law, D., Grant, D. J., & Schmitt, E. A. (2002). Physical stability of amorphous pharmaceuticals: Importance of configurational thermodynamic quantities and molecular mobility. *Journal of pharmaceutical sciences*, 91(8), 1863-1872.
- Zink, T., Ross, A., Lueers, K., Cieslar, C., Rudolph, R., & Holak, T. A. (1994). Structure and dynamics of the human granulocyte colony-stimulating factor determined by NMR spectroscopy. Loop mobility in a four-helix-bundle protein. *Biochemistry*, 33(28), 8453-8463.
- Zobel, H.P., & Arndt, S. (2005). Lyophilised preparation comprising immunocytokines. U.S. Patent Application 10/503,615.

10 Appendix

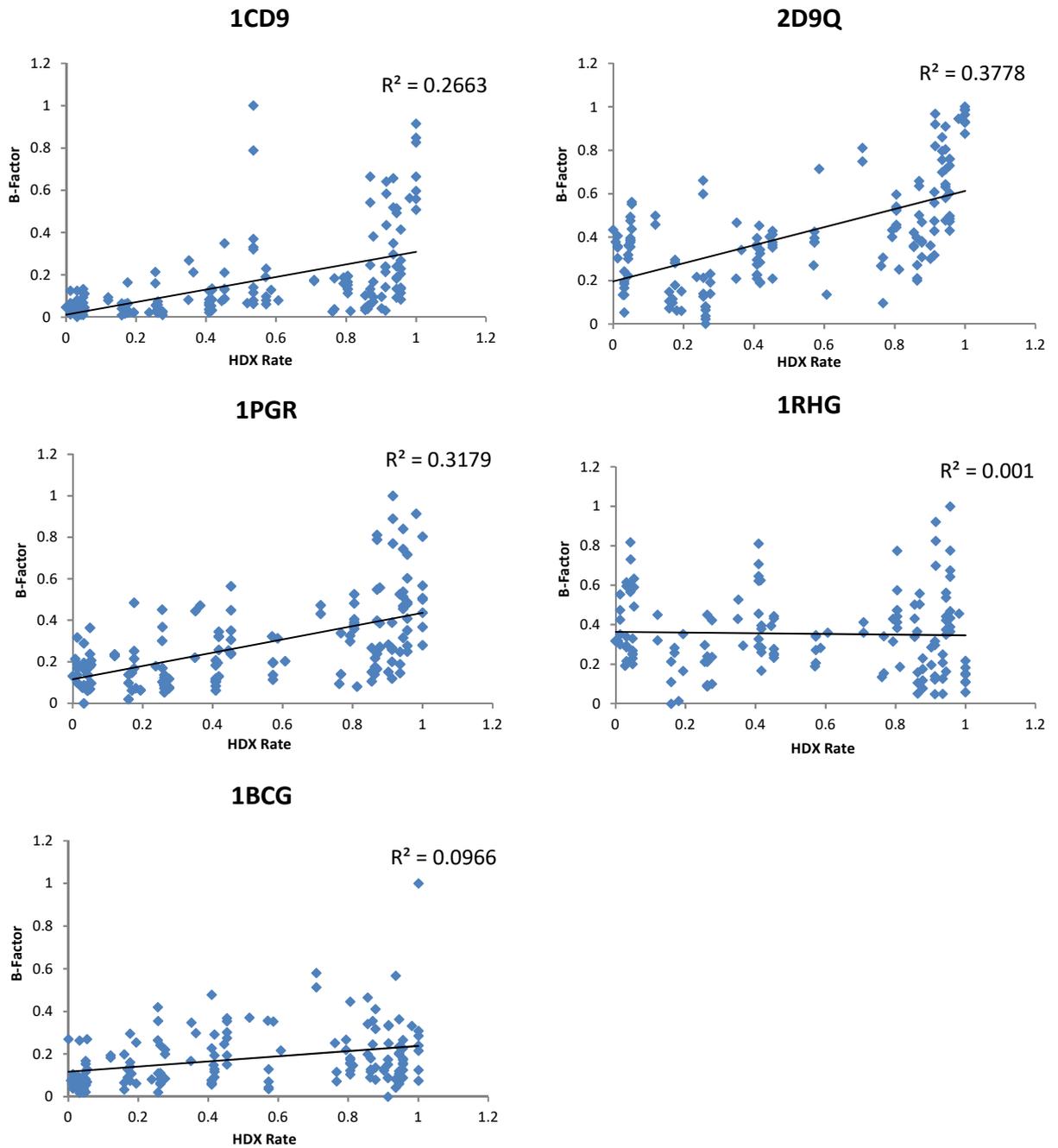
Appendix 1. GCSF DNA sequence

ACACCCCTAGGCCCTGCCAGCTCCCTGCCCCAGAGCTTCTGCTCAAGTGCTTAGAGCAAGTGA
 GGAAGATCCAGGGCGATGGCGCAGCGCTCCAGGAGAAGCTGTGTGCCACCTACAAGCTGTGCC
 ACCCCGAGGAGCTGGTGTCTCGGACACTCTCTGGGCATCCCCTGGGCTCCCCTGAGCAGCT
 GCCCCAGCCAGGCCCTGCAGCTGGCAGGCTGCTTGAGCCAACCTCCATAGCGGCCTTTCTCTA
 CCAGGGGCTCCTGCAGGCCCTGGAAGGGATCTCCCCGAGTTGGGTCCCACCTTGACACACT
 GCAGCTGGACGTCGCCGACTTTGCCACCACCATCTGGCAGCAGATGGAAGAAGTGGGAATGGCC
 CTGCCCCTGCAGCCCACCCAGGGTGCCATGCCGGCCTTCGCCTCTGCTTTCCAGCGCCGGGCA
 GGAGGGGTCTGGTTGCCTCCCATCTGCAGAGCTTCTGGAGGTGTCGTACCGCGTTCTACGCC
 ACCTTGCCCAGCCC

Appendix 2. Amino acid sequences for all available PDB structures for GCSF as of 2016.

PDB	Amino Acid Sequence
2D9Q	TPLGPASSLPQSFLKCLEQVRKIQGDGAALQEKLKCATYKLCHEPELVLLGHSLGIPWAPLSS CPSQALQLAGCLSQLHSGLFLYQGLLQALEGISPELGPTLDTLQLDVADFATTIWQQMEEL GMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVSYRVLRLHQAQP
1CD9	TPLGPASSLPQSFLKCLEQVRKIQGDGAALQEKLKCATYKLCHEPELVLLGHSLGIPWAPLSS CPSQALQLAGCLSQLHSGLFLYQGLLQALEGISPELGPTLDTLQLDVADFATTIWQQMEEL GMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVSYRVLRLHQAQP
1PGR	TPLGPASSLPQSFLKCLEQVRKIQGDGAALQEKLKCATYKLCHEPELVLLGHSLGIPWAPLSS CPSQALQLAGCLSQLHSGLFLYQGLLQALEGISPELGPTLDTLQLDVADFATTIWQQMEEL GMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVSYRVLRLHQAQP
1GNC	TPLGPASSLPQSFLKCLEQVRKIQGDGAALQEKLKCATYKLCHEPELVLLGHSLGIPWAPLSS LSSCPSQALQLAGCLSQLHSGLFLYQGLLQALEGISPELGPTLDTLQLDVADFATTIWQQMEEL ELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVSYRVLRLHQAQP
1RHG	TPLGPASSLPQSFLKCLEQVRKIQGDGAALQEKLKCATYKLCHEPELVLLGHSLGIPWAPLSS CPSQALQLAGCLSQLHSGLFLYQGLLQALEGISPELGPTLDTLQLDVADFATTIWQQMEEL GMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVSYRVLRLHQAQP
1BGC	TPLGPASSLPQSFLKCLEQVRKIQADGAELQERLCAAHKLCHEPELMLLRHSLGIPQAPLSS CSSQSLQLRGCLNQLHGGLFLYQGLLQALAGISPELAPTLDLQLDVTDFATNIWLQMEDL GAAPAVQPTQGAMPTFTSAFQRRAGGVLVASHLQSFLEVSYRVLRLHQAQP
1BGD	MAPLGPTGPLPQSFLKCLEQMRKVQADGTALQETLQATHQLCHPEPELVLLGHSLGIPQPP LSSCSSLQALQMLGCLRQLHSGLFLYQGLLQALAGISPELAPTLDLQLDVTDFATNIWLQMEDL DLGMAPAVPPTQGTMPAFTSAFQRRAGGVLVASHLQSFLEVSYRVLRLHQAQP
1BGE	MAPLGPTGPLPQSFLKCLEQMRKVQADGTALQETLQATHQLCHPEPELVLLGHSLGIPQPP LSSCSSLQALQMLGCLRQLHSGLFLYQGLLQALAGISPELAPTLDLQLDVTDFATNIWLQMEDL DLGMAPAVPPTQGTMPAFTSAFQRRAGGVLVASHLQSFLEVSYRVLRLHQAQP

Appendix 3. R² calculation of GCSF PDB structure B-factors plotted against HDX-MS uptake rate



Appendix 4. Peptide information for **Repeatability of GCSF peptide relative uptake as measured by HDX-MS**

Peptide number	Peptide residue range	Sequence	Region
1	1-10	MTPLGPASSL	LoopA
2	1-13	MTPLGPASSLPQS	
3	1-14	MTPLGPASSLPQSF	
4	1-15	MTPLGPASSLPQSFL	
5	1-15	MTPLGPASSLPQSFL	
6	5-15	PLGPASSLPQSF	α A
7	5-33	PLGPASSLPQSFL	
8	19-32	LGPASSLPQSF	
9	19-34	GPASSLPQSFL	
10	21-32	PASSLPQSFL	
11	22-32	SLPQSFLKCLEQV	
12	24-39	LLKCLEQ	
13	27-41	KCLEQVRKIQGDGAAL	
14	27-47	LEQVRKIQGDGAAL	LoopAB
15	33-39	EQVRKIQGDGAAL	
16	40-47	QVRKIQGDGAAL	
17	40-48	VRKIQGDGAAL	
18	41-47	VRKIQGDGAALQ	
19	42-48	RKIQGDGAAL	
20	49-65	QEKLCAAT	
21	51-65	YKLCHPEE	
22	51-68	KLCHPEE	
23	51-69	KLCHPEEL	
24	58-65	LCHPEEL	
25	58-69	VLLGHSLGIPWAPLSSCPSQA	
26	70-76	LGHSLGIPWAPLSSC	α B
27	77-83	LGHSLGIPWAPLSSCPSQ	
28	84-90	LGHSLGIPWAPLSSCPSQA	
29	84-91	PWAPLSSC	
30	84-92	PLSSCPSQA	
31	85-92	PSQALQL	
32	86-92	LQLAGCL	
33	91-97	LSQLHSGLFLY	LoopBC
34	91-104	SQLHSGL	
35	91-106	SQLHSGLF	
36	93-104	FLYQGLL	
37	93-106	LYQGLLQA	
38	94-101	YQGLLQA	α C
39	94-104	LQALEGIS	
40	94-106	QALEGIS	

41	96-104	QALEGISPELG	
42	96-106	QALEGISPELGPTL	
43	98-104	QALEGISPELGPTLDT	
44	98-106	ALEGISPELGPTLDT	
45	98-127	LEGISPELGPTL	
46	108-116	LEGISPELGPTLDT	
47	113-127	EGISPELGPTL	
48	114-120	EGISPELGPTLDT	
49	114-124	GISPELGPTL	
50	115-122	ISPELGPTL	
51	115-124	ISPELGPTLDT	
52	118-124	SPELGPTL	
53	121-138	PELGPTL	
54	123-138	PELGPTLDT	
55	125-138	LGPTLDTLQLD	
56	125-141	LGPTLDTLQLDV	
57	126-138	VADFATTIWQQ	
58	126-141	FATTIWQ	
59	129-138	FATTIWQQM	
60	129-141	FATTIWQQMEE	
61	136-156	ATTIWQQ	
62	142-153	ATTIWQQM	
63	146-153	ATTIWQQMEE	
64	146-160	TTIWQQM	
65	154-161	IWQQMEE	
66	166-175	EELGMAPALQPTQGAM	
67	167-175	EELGMAPALQPTQGAM	

LoopCD

αD

LoopD

Appendix 5. Peptide information for Relative deuterium uptake of GCSF in 10mM sodium acetate pH 4.25 as measured by HDX-MS.

Peptide number	Peptide residue range	Sequence	Region
1	1-10	MTPLGPASSLPQSF	LoopA
2	1-13	MTPLGPASSLPQSFL	
3	1-14	MTPLGPASSLPQSFL	
4	1-15	MTPLGPASSLPQSFLKLC	
5	1-15	GPASSLPQSFL	α A
6	5-15	PASSLPQSFL	
7	5-33	SLPQSFLKCLEQV	
8	19-32	LLKCLEQ	
9	19-34	LKCLEQVRKIQGDGAAL	
10	21-32	LEQVRKIQGDGAAL	
11	22-32	EQVRKIQGDGAAL	
12	24-39	QVRKIQGDGAAL	
13	27-41	VRKIQGDGAAL	
14	27-47	QEKLCAT	
15	33-39	YKLCHPEE	
16	40-47	YKLCHPEEL	
17	40-48	VLLGHSLGIPWAPLSSC	
18	41-47	VLLGHSLGIPWAPLSSCPSQA	
19	42-48	LGHSLGIPWAPLSSC	
20	49-65	LGHSLGIPWAPLSSCPSQ	
21	51-65	LGHSLGIPWAPLSSCPSQA	
22	51-68	PSQALQL	
23	51-69	ALQLAGCLSQLH	α B
24	58-65	LQLAGCL	
25	58-69	SQLHSGL	
26	70-76	FLYQGLL	
27	77-83	FLYQGLLQA	
28	84-90	YQGLLQA	
29	84-91	YQGLLQAL	
30	84-92	QALEGISPELGPTL	LoopBC
31	85-92	QALEGISPELGPTLDT	
32	86-92	LEGISPELGPTL	
33	91-97	LEGISPELGPTLDT	α C
34	91-104	EGISPELGPTL	
35	91-106	EGISPELGPTLDT	
36	93-104	ISPELGPT	
37	93-106	ISPELGPTL	
38	94-101	SPELGPTLDT	
39	94-104	PELGPTLDT	
40	94-106	FATTIWQQMEE	
41	96-104	ATTIWQQMEE	

42	96-106	EELGMAPALQPTQGAM	LoopCD
43	98-104	EELGMAPALQPTQGAMPAF	
44	98-106	EELGMAPALQPTQGAMPAFASA	
45	98-127	LGMAPALQPTQGAM	
46	108-116	LGMAPALQPTQGAMPAF	
47	113-127	LGMAPALQPTQGAMPAFASA	
48	114-120	GMAPALQPTQGAM	
49	114-124	GMAPALQPTQGAMPAF	
50	115-122	GMAPALQPTQGAMPAFASA	
51	115-124	PALQPTQGAMPAF	
52	118-124	PTQGAMPAF	
53	121-138	ASAFQRRAGGVL	
54	123-138	QRRAGGVL	
55	125-138	VASHLQSF	
56	125-141	VASHLQSF	
57	126-138	ASHLQSF	
58	126-141	YRVLRLHAQP	LoopD
59	129-138	RVLRLHAQP	

Appendix 6. Forward and reverse primer DNA sequences for site-directed mutagenesis

Mutant	Forward primer (5'-3')	Reverse primer (5'-3')
S12W	ctccctgccccagtggttctctgctcaagtg	cacttgagcaggaaccactggggcagggag
F13A	agcacttgagcagggcgctctggggcaggg	ccctgccccagagcgccttctcaagtct
C17A	ttctcacttgctctaaggccttgagcaggaagctctg	cagagcttctgctcaaggccttagagcaagtggaa
I24A	gcgccatcgccctgggccttctcacttgctc	gagcaagtgaggaaggcccagggcgatggcgc
G28I	gaagatccagggcgatatcgacgctccag	ctggagcgcctgcgatatcgccctggatcttc
T38W	ggagaagctgtgtcctggtacaagctgtgccacc	ggtggcacagctgtaccaggcacacagcttctcc
G51R	gctggtgctgctcagacactctctggg	cccagagagtgtctgagcagcaccagc
L71W	gttgggtcccacctgtatacactgcagctggacg	caagcagcctcccactgcagggcctgg
S80W	ctgcttgagccaactccattggggcctttcttc	gaggaaaaggcccaatggagttggctcaagcag
D104Y	gttgggtcccacctgtatacactgcagctggacg	cgtccagctgcagtgatacaaggtgggacccaac
Q107Y	accttggacacactgtatctggacgtcggcagc	gtcggcgacgtccagatacagtggtccaaggt
Q120I	ccattcccagttcttccattatctccagatggtggtggca	tgccaccaccatctggcagataatggaagaactgggaatgg
M126S	agggcagggcgcttcccagttcttccatctgc	gcagatggaagaactgggaagcggccttgcct
I31F	ggcccctgcccgttccccacccaggggtg	caccctgggtggggaacagggcaggggccc
F132E	catggcaccctgggttctctgcagggcaggggc	gcccctgcccctgaggaaccaggggtgcatg
M137L	gaaggccggcaaggcaccctggg	cccaggggtccttggcggccttc
I55Y	ggggtcctggtgcctatcatctgcagagcttcc	ggaagctctcagatgataggcaaccagggacccc
I64L	gcttctggaggtgtgtaccggttctacg	cgtagaacgcggtacaacacctccaggaagc

Appendix 7. BLAST® blastp alignment of WT and mutant amino acid sequences

F13A						
Score	Expect	Method	Identities	Positives	Gaps	
341 bits(875)	4e-127	Compositional matrix adjust.	173/174(99%)	173/174(99%)	0/174(0%)	
Query 1		TPLGPASSLPQSFLKLCLEQVRKIQGDGAALQEKL	CATYKLC	HP	PEELVLLGHSLGIPWAP	60
Sbjct 1		TPLGPASSLPQSFLKLCLEQVRKIQGDGAALQEKL	CATYKLC	HP	PEELVLLGHSLGIPWAP	60
Query 61		LSSCPSQALQLAGCLS	QLHSGFLFYQGLLQALEGISPEL	GPTLDTLQ	LDVADFATTIWQQ	120
Sbjct 61		LSSCPSQALQLAGCLS	QLHSGFLFYQGLLQALEGISPEL	GPTLDTLQ	LDVADFATTIWQQ	120
Query 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS	YRVL	RHLAQP		174
Sbjct 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS	YRVL	RHLAQP		174

C17A						
Score	Expect	Method	Identities	Positives	Gaps	
341 bits(875)	4e-127	Compositional matrix adjust.	173/174(99%)	173/174(99%)	0/174(0%)	
Query 1		TPLGPASSLPQSFLKLCLEQVRKIQGDGAALQEKL	CATYKLC	HP	PEELVLLGHSLGIPWAP	60
Sbjct 1		TPLGPASSLPQSFLKLCLEQVRKIQGDGAALQEKL	CATYKLC	HP	PEELVLLGHSLGIPWAP	60
Query 61		LSSCPSQALQLAGCLS	QLHSGFLFYQGLLQALEGISPEL	GPTLDTLQ	LDVADFATTIWQQ	120
Sbjct 61		LSSCPSQALQLAGCLS	QLHSGFLFYQGLLQALEGISPEL	GPTLDTLQ	LDVADFATTIWQQ	120
Query 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS	YRVL	RHLAQP		174
Sbjct 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS	YRVL	RHLAQP		174

I24A						
Score	Expect	Method	Identities	Positives	Gaps	
342 bits(878)	2e-127	Compositional matrix adjust.	173/174(99%)	173/174(99%)	0/174(0%)	
Query 1		TPLGPASSLPQSFLKLCLEQVRKIQGDGAALQEKL	CATYKLC	HP	PEELVLLGHSLGIPWAP	60
Sbjct 1		TPLGPASSLPQSFLKLCLEQVRKIQGDGAALQEKL	CATYKLC	HP	PEELVLLGHSLGIPWAP	60
Query 61		LSSCPSQALQLAGCLS	QLHSGFLFYQGLLQALEGISPEL	GPTLDTLQ	LDVADFATTIWQQ	120
Sbjct 61		LSSCPSQALQLAGCLS	QLHSGFLFYQGLLQALEGISPEL	GPTLDTLQ	LDVADFATTIWQQ	120
Query 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS	YRVL	RHLAQP		174
Sbjct 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS	YRVL	RHLAQP		174

Q120I						
Score	Expect	Method	Identities	Positives	Gaps	
342 bits(876)	3e-127	Compositional matrix adjust.	173/174(99%)	173/174(99%)	0/174(0%)	
Query 1		TPLGPASSLPQSFLKLCLEQVRKIQGDGAALQEKL	CATYKLC	HP	PEELVLLGHSLGIPWAP	60
Sbjct 1		TPLGPASSLPQSFLKLCLEQVRKIQGDGAALQEKL	CATYKLC	HP	PEELVLLGHSLGIPWAP	60
Query 61		LSSCPSQALQLAGCLS	QLHSGFLFYQGLLQALEGISPEL	GPTLDTLQ	LDVADFATTIWQQ	120
Sbjct 61		LSSCPSQALQLAGCLS	QLHSGFLFYQGLLQALEGISPEL	GPTLDTLQ	LDVADFATTIWQQ	120
Query 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS	YRVL	RHLAQP		174
Sbjct 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS	YRVL	RHLAQP		174

M126S

Score	Expect	Method	Identities	Positives	Gaps
341 bits(875)	4e-127	Compositional matrix adjust.	173/174(99%)	173/174(99%)	0/174(0%)
Query 1		TPLGPASSLPQSFLKCLEQVRKIQGDGAALQEKLKATYKLCHEPELVLLGHSLGIPWAP			60
Sbjct 1		TPLGPASSLPQSFLKCLEQVRKIQGDGAALQEKLKATYKLCHEPELVLLGHSLGIPWAP			60
Query 61		LSSCPSQALQLAGCLSQLHSGFLFYQGLLQALEGISPELGPTLDTLQLDVADFATTIWQQ			120
Sbjct 61		LSSCPSQALQLAGCLSQLHSGFLFYQGLLQALEGISPELGPTLDTLQLDVADFATTIWQQ			120
Query 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS YRVLRLH LAQP			174
Sbjct 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS YRVLRLH LAQP			174

P132E

Score	Expect	Method	Identities	Positives	Gaps
342 bits(876)	3e-127	Compositional matrix adjust.	173/174(99%)	173/174(99%)	0/174(0%)
Query 1		TPLGPASSLPQSFLKCLEQVRKIQGDGAALQEKLKATYKLCHEPELVLLGHSLGIPWAP			60
Sbjct 1		TPLGPASSLPQSFLKCLEQVRKIQGDGAALQEKLKATYKLCHEPELVLLGHSLGIPWAP			60
Query 61		LSSCPSQALQLAGCLSQLHSGFLFYQGLLQALEGISPELGPTLDTLQLDVADFATTIWQQ			120
Sbjct 61		LSSCPSQALQLAGCLSQLHSGFLFYQGLLQALEGISPELGPTLDTLQLDVADFATTIWQQ			120
Query 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS YRVLRLH LAQP			174
Sbjct 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS YRVLRLH LAQP			174

M137L

Score	Expect	Method	Identities	Positives	Gaps
342 bits(878)	1e-127	Compositional matrix adjust.	173/174(99%)	174/174(100%)	0/174(0%)
Query 1		TPLGPASSLPQSFLKCLEQVRKIQGDGAALQEKLKATYKLCHEPELVLLGHSLGIPWAP			60
Sbjct 1		TPLGPASSLPQSFLKCLEQVRKIQGDGAALQEKLKATYKLCHEPELVLLGHSLGIPWAP			60
Query 61		LSSCPSQALQLAGCLSQLHSGFLFYQGLLQALEGISPELGPTLDTLQLDVADFATTIWQQ			120
Sbjct 61		LSSCPSQALQLAGCLSQLHSGFLFYQGLLQALEGISPELGPTLDTLQLDVADFATTIWQQ			120
Query 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS YRVLRLH LAQP			174
Sbjct 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS YRVLRLH LAQP			174

S12W

Score	Expect	Method	Identities	Positives	Gaps
341 bits(875)	4e-127	Compositional matrix adjust.	173/174(99%)	173/174(99%)	0/174(0%)
Query 1		TPLGPASSLPQSFLKLCLEQVRKIQGDGAALQEKLCA			60
Sbjct 2		TPLGPASSLPQSFLKLCLEQVRKIQGDGAALQEKLCA			61
Query 61		LSSCPSQALQLAGCLS			120
Sbjct 62		LSSCPSQALQLAGCLS			121
Query 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS			174
Sbjct 122		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS			175

G28I

Score	Expect	Method	Identities	Positives	Gaps
338 bits(867)	7e-126	Compositional matrix adjust.	172/174(99%)	172/174(98%)	0/174(0%)
Query 1		TPLGPASSLPQSFLKLCLEQVRKIQGDGAALQEKLCA			60
Sbjct 2		TPLGPASSLPQSFLKLCLEQVRKIQGDGAALQEKLCA			61
Query 61		LSSCPSQALQLAGCLS			120
Sbjct 62		LSSCPSQALQLAGCLS			121
Query 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS			174
Sbjct 122		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS			175

T38W

Score	Expect	Method	Identities	Positives	Gaps
341 bits(874)	5e-127	Compositional matrix adjust.	173/174(99%)	173/174(99%)	0/174(0%)
Query 1		TPLGPASSLPQSFLKLCLEQVRKIQGDGAALQEKLCA			60
Sbjct 2		TPLGPASSLPQSFLKLCLEQVRKIQGDGAALQEKLCA			61
Query 61		LSSCPSQALQLAGCLS			120
Sbjct 62		LSSCPSQALQLAGCLS			121
Query 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS			174
Sbjct 122		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS			175

G51R

Score	Expect	Method	Identities	Positives	Gaps
342 bits(877)	2e-127	Compositional matrix adjust.	173/174(99%)	173/174(99%)	0/174(0%)
Query 1		TPLGPASSLPQSFLKLCLEQVRKIQGDGAALQEKLCA			60
Sbjct 2		TPLGPASSLPQSFLKLCLEQVRKIQGDGAALQEKLCA			61
Query 61		LSSCPSQALQLAGCLS			120
Sbjct 62		LSSCPSQALQLAGCLS			121
Query 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS			174
Sbjct 122		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS			175

L71W

Score	Expect	Method	Identities	Positives	Gaps
342 bits(878)	2e-127	Compositional matrix adjust.	173/174(99%)	173/174(99%)	0/174(0%)
Query 1		TPLGPASSLPQSFLKCLEQVRKIQDGAALQEKLKATYKLCHEPELVLLGHSLGIPWAP			60
Sbjct 2		TPLGPASSLPQSFLKCLEQVRKIQDGAALQEKLKATYKLCHEPELVLLGHSLGIPWAP			61
Query 61		LSSCPSQALQLAGCLS QLH SGLFLYQGLLQALEGISPELGPTLDLQLQDVADFATTIWQQ			120
Sbjct 62		LSSCPSQALQLAGCLS QLH SGLFLYQGLLQALEGISPELGPTLDLQLQDVADFATTIWQQ			121
Query 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS YR VLRLHQAQP			174
Sbjct 122		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS YR VLRLHQAQP			175

S80W

Score	Expect	Method	Identities	Positives	Gaps
341 bits(875)	4e-127	Compositional matrix adjust.	173/174(99%)	173/174(99%)	0/174(0%)
Query 1		TPLGPASSLPQSFLKCLEQVRKIQDGAALQEKLKATYKLCHEPELVLLGHSLGIPWAP			60
Sbjct 1		TPLGPASSLPQSFLKCLEQVRKIQDGAALQEKLKATYKLCHEPELVLLGHSLGIPWAP			60
Query 61		LSSCPSQALQLAGCLS QLH SGLFLYQGLLQALEGISPELGPTLDLQLQDVADFATTIWQQ			120
Sbjct 61		LSSCPSQALQLAGCLS QLH SGLFLYQGLLQALEGISPELGPTLDLQLQDVADFATTIWQQ			120
Query 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS YR VLRLHQAQP			174
Sbjct 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS YR VLRLHQAQP			174

D104Y

Score	Expect	Method	Identities	Positives	Gaps
341 bits(875)	5e-127	Compositional matrix adjust.	173/174(99%)	173/174(99%)	0/174(0%)
Query 1		TPLGPASSLPQSFLKCLEQVRKIQDGAALQEKLKATYKLCHEPELVLLGHSLGIPWAP			60
Sbjct 2		TPLGPASSLPQSFLKCLEQVRKIQDGAALQEKLKATYKLCHEPELVLLGHSLGIPWAP			61
Query 61		LSSCPSQALQLAGCLS QLH SGLFLYQGLLQALEGISPELGPTLDLQLQDVADFATTIWQQ			120
Sbjct 62		LSSCPSQALQLAGCLS QLH SGLFLYQGLLQALEGISPELGPTLDLQLQDVADFATTIWQQ			121
Query 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS YR VLRLHQAQP			174
Sbjct 122		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS YR VLRLHQAQP			175

Q107Y

Score	Expect	Method	Identities	Positives	Gaps
342 bits(878)	2e-127	Compositional matrix adjust.	173/174(99%)	173/174(99%)	0/174(0%)
Query 1		TPLGPASSLPQSFLKCLEQVRKIQDGAALQEKLKATYKLCHEPELVLLGHSLGIPWAP			60
Sbjct 2		TPLGPASSLPQSFLKCLEQVRKIQDGAALQEKLKATYKLCHEPELVLLGHSLGIPWAP			61
Query 61		LSSCPSQALQLAGCLS QLH SGLFLYQGLLQALEGISPELGPTLDLQLQDVADFATTIWQQ			120
Sbjct 62		LSSCPSQALQLAGCLS QLH SGLFLYQGLLQALEGISPELGPTLDLQLQDVADFATTIWQQ			121
Query 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS YR VLRLHQAQP			174
Sbjct 122		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS YR VLRLHQAQP			175

Q131F

Score	Expect	Method	Identities	Positives	Gaps
341 bits(875)	5e-127	Compositional matrix adjust.	173/174(99%)	173/174(99%)	0/174(0%)
Query 1		TPLGPASSLPQSFLKCLEQVRKIQGDGAALQEKLKATYKLCHEPEELVLLGHSLGIPWAP			60
Sbjct 2		TPLGPASSLPQSFLKCLEQVRKIQGDGAALQEKLKATYKLCHEPEELVLLGHSLGIPWAP			61
Query 61		LSSCPSQALQLAGCLSQLHSGFLFYQGLLQALEGISPELGPTLDTLQLDQVADFATTIWQQ			120
Sbjct 62		LSSCPSQALQLAGCLSQLHSGFLFYQGLLQALEGISPELGPTLDTLQLDQVADFATTIWQQ			121
Query 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS YRVLRRHLAQP			174
Sbjct 122		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS YRVLRRHLAQP			175

S155Y

Score	Expect	Method	Identities	Positives	Gaps
342 bits(877)	2e-127	Compositional matrix adjust.	173/174(99%)	173/174(99%)	0/174(0%)
Query 1		TPLGPASSLPQSFLKCLEQVRKIQGDGAALQEKLKATYKLCHEPEELVLLGHSLGIPWAP			60
Sbjct 2		TPLGPASSLPQSFLKCLEQVRKIQGDGAALQEKLKATYKLCHEPEELVLLGHSLGIPWAP			61
Query 61		LSSCPSQALQLAGCLSQLHSGFLFYQGLLQALEGISPELGPTLDTLQLDQVADFATTIWQQ			120
Sbjct 62		LSSCPSQALQLAGCLSQLHSGFLFYQGLLQALEGISPELGPTLDTLQLDQVADFATTIWQQ			121
Query 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS YRVLRRHLAQP			174
Sbjct 122		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS YRVLRRHLAQP			175

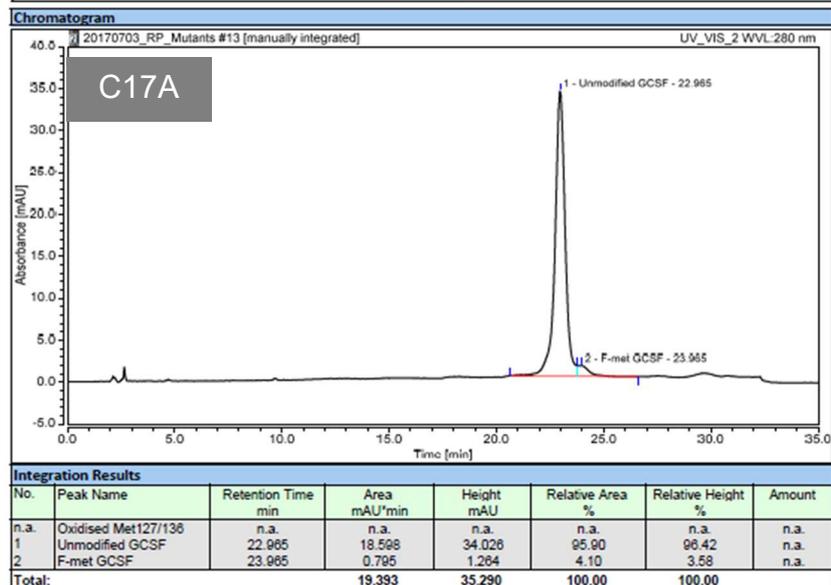
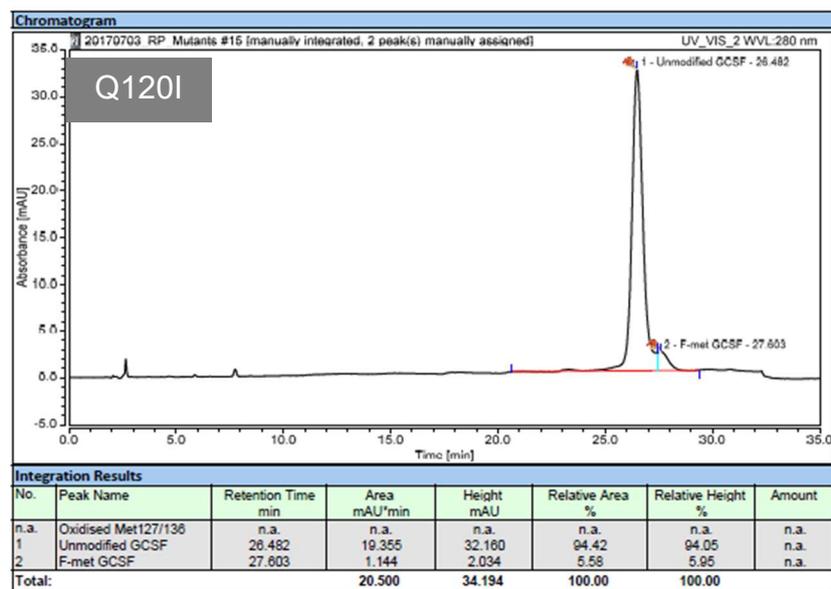
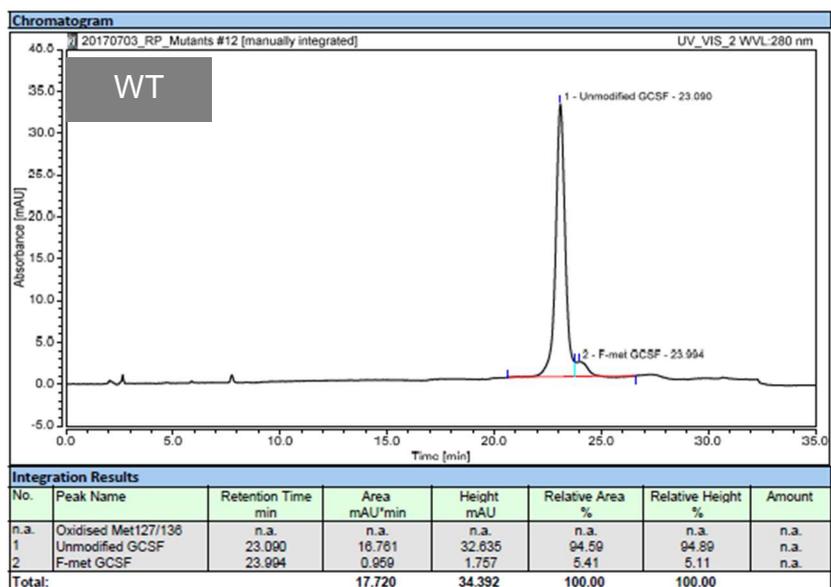
S164L

Score	Expect	Method	Identities	Positives	Gaps
342 bits(876)	3e-127	Compositional matrix adjust.	173/174(99%)	173/174(99%)	0/174(0%)
Query 1		TPLGPASSLPQSFLKCLEQVRKIQGDGAALQEKLKATYKLCHEPEELVLLGHSLGIPWAP			60
Sbjct 2		TPLGPASSLPQSFLKCLEQVRKIQGDGAALQEKLKATYKLCHEPEELVLLGHSLGIPWAP			61
Query 61		LSSCPSQALQLAGCLSQLHSGFLFYQGLLQALEGISPELGPTLDTLQLDQVADFATTIWQQ			120
Sbjct 62		LSSCPSQALQLAGCLSQLHSGFLFYQGLLQALEGISPELGPTLDTLQLDQVADFATTIWQQ			121
Query 121		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS YRVLRRHLAQP			174
Sbjct 122		MEELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVS YRVLRRHLAQP			175

Appendix 8. Extinction coefficients (EC) for GCSF mutants

Sample	EC	Δ EC
WT	15820	1.00
F13A	15820	1.00
C17A	15700	0.99
I24A	15820	1.00
Q120I	15820	1.00
M126S	15820	1.00
P132E	15820	1.00
M137L	15820	1.00
S12W	21510	1.4
G28I	15820	1.0
T38W	21510	1.4
G51R	15820	1.0
L71W	21510	1.4
S80W	21510	1.4
D104Y	17100	1.1
Q107Y	17100	1.1
Q131F	15820	1.0
S155Y	17100	1.1
S164L	15820	1.0

Appendix 9. RP-HPLC chromatograms with integration



Appendix 10. Peptide information for G51R and Q120I peptide differential uptake plots.

Peptide number	Peptide residue range	Sequence	Region
1	1-10	MTPLGPASSL	LoopA
2	1-13	MTPLGPASSLPQS	
3	1-14	MTPLGPASSLPQSF	
4	1-15	MTPLGPASSLPQSFL	
5	1-15	MTPLGPASSLPQSFL	
6	3-14	PLGPASSLPQSF	
7	3-15	PLGPASSLPQSFL	
8	5-15	GPASSLPQSFL	α A
9	6-14	PASSLPQSF	
10	6-15	PASSLPQSFL	
11	15-21	LLKCLEQ	
12	16-32	LKCLEQVRKIQGDGAAL	
13	17-32	KCLEQVRKIQGDGAAL	
14	19-32	LEQVRKIQGDGAAL	
15	22-32	VRKIQGDGAAL	
16	22-34	VRKIQGDGAALQE	
17	33-39	QEKLCAT	LoopAB
18	40-47	YKLCHPEE	
19	40-48	YKLCHPEEL	
20	41-47	KLCHPEE	
21	41-48	KLCHPEEL	
22	42-48	LCHPEEL	
23	51-65	LGHSLGIPWAPLSSC	
24	51-68	LGHSLGIPWAPLSSCPSQ	
25	51-69	LGHSLGIPWAPLSSCPSQA	
26	66-72	PSQALQL	
27	70-76	LQLAGCL	α B

28	76-83	LSQLHSGL	
29	76-86	LSQLHSGFLY	
30	77-83	SQLHSGL	
31	77-84	SQLHSGLF	
32	84-90	FLYQGLL	
33	84-91	FLYQGLLQ	
34	84-92	FLYQGLLQA	
35	85-92	LYQGLLQA	
36	86-92	YQGLLQA	
37	90-97	LQALEGIS	
38	91-97	QALEGIS	
39	91-104	QALEGISPELGPTL	
40	93-101	LEGISPELG	LoopBC
41	93-104	LEGISPELGPTL	
42	93-106	LEGISPELGPTLDT	
43	94-100	EGISPEL	
44	94-101	EGISPELG	
45	94-104	EGISPELGPTL	
46	94-106	EGISPELGPTLDT	
47	96-104	ISPELGPTL	
48	96-106	ISPELGPTLDT	
49	97-104	SPELGPTL	
50	98-104	PELGPTL	
51	98-106	PELGPTLDT	
52	100-110	LGPTLDTLQLD	
53	100-111	LGPTLDTLQLDV	
54	107-116	LQLDVADFAT	
55	114-120	FATTIWQ	
56	114-124	FATTIWQQMEE	
57	115-121	ATTIWQQ	αC

58	115-122	ATTIWQQM		
59	115-124	ATTIWQQMEE		
60	118-124	IWQQMEE		
61	121-137	QMEELGMAPALQPTQGA	LoopCD	
62	123-138	EELGMAPALQPTQGAM		
63	125-138	LGMAPALQPTQGAM		
64	125-141	LGMAPALQPTQGAMPAF		
65	126-135	GMAPALQPTQ		
66	126-138	GMAPALQPTQGAM		
67	126-141	GMAPALQPTQGAMPAF		
68	129-138	PALQPTQGAM		
69	129-141	PALQPTQGAMPAF		
70	130-140	ALQPTQGAMPA		
71	132-138	QPTQGAM		
72	133-141	PTQGAMPAF		
73	139-153	PAFASAFQRRAGGVL		α D
74	141-153	FASAFQRRAGGVL		
75	143-153	SAFQRRAGGVL		
76	146-153	QRRAGGVL		
77	154-161	VASHLQSF		
78	167-174	RVLRHLAQ	LoopD	
79	167-175	RVLRHLAQP		
80	168-175	VLRHLAQP		

Appendix 11. Peptide information for C17A differential uptake plot.

Peptide number	Peptide residue range	Sequence	Region
1	1-10	MTPLGPASSL	LoopA

2	1-13	MTPLGPASSLPQS	
3	1-14	MTPLGPASSLPQSF	
4	1-15	MTPLGPASSLPQSFL	
5	1-15	MTPLGPASSLPQSFL	
6	3-14	PLGPASSLPQSF	
7	5-15	GPASSLPQSFL	
8	6-15	PASSLPQSFL	
9	15-21	LLKCLEQ	
10	16-32	LKCLEQVRKIQGDGAAL	
11	17-32	KCLEQVRKIQGDGAAL	
12	19-32	LEQVRKIQGDGAAL	
13	20-32	EQVRKIQGDGAAL	
14	22-32	VRKIQGDGAAL	
15	22-34	VRKIQGDGAALQE	α A
16	33-39	QEKLCAT	
17	40-47	YKLCHPEE	
18	40-48	YKLCHPEEL	
19	41-47	KLCHPEE	
20	41-48	KLCHPEEL	
21	42-48	LCHPEEL	
22	51-65	LGHSLGIPWAPLSSC	
23	51-68	LGHSLGIPWAPLSSCPSQ	
24	51-69	LGHSLGIPWAPLSSCPSQA	
25	66-72	PSQALQL	LoopAB
26	70-76	LQLAGCL	
27	76-83	LSQLHSGL	
28	76-86	LSQLHSGLFLY	
29	77-83	SQLHSGL	
30	77-84	SQLHSGLF	
31	84-90	FLYQGLL	α B

32	84-91	FLYQGLLQ	
33	84-92	FLYQGLLQA	
34	85-92	LYQGLLQA	
35	86-92	YQGLLQA	
36	90-97	LQALEGIS	
37	91-97	QALEGIS	
38	91-104	QALEGISPELGPTL	
39	91-106	QALEGISPELGPTLDT	
40	93-101	LEGISPELG	
41	93-104	LEGISPELGPTL	LoopBC
42	93-106	LEGISPELGPTLDT	
43	94-101	EGISPELG	
44	94-104	EGISPELGPTL	
45	94-106	EGISPELGPTLDT	
46	96-104	ISPELGPTL	
47	96-106	ISPELGPTLDT	
48	98-104	PELGPTL	
49	98-106	PELGPTLDT	
50	100-110	LGPTLDTLQLD	
51	109-141	LDVADFATTIWQQMEELGMAPALQPTQGAMPAF	
52	114-120	FATTIWQ	
53	114-124	FATTIWQQMEE	
54	115-121	ATTIWQQ	
55	115-122	ATTIWQQM	
56	115-124	ATTIWQQMEE	
57	116-122	TTIWQQM	
58	118-124	IWQQMEE	α C
59	121-137	QMEELGMAPALQPTQGA	
60	125-138	LGMAPALQPTQGAM	
61	126-135	GMAPALQPTQ	LoopCD

62	126-138	GMAPALQPTQGAM	
63	126-141	GMAPALQPTQGAMPAF	
64	129-138	PALQPTQGAM	
65	129-141	PALQPTQGAMPAF	
66	130-140	ALQPTQGAMPA	
67	132-138	QPTQGAM	
68	139-153	PAFASAFQRRAGGVL	αD
69	146-153	QRRAGGVL	
70	154-161	VASHLQSF	
71	166-175	YRVLRLHAQP	LoopD
72	167-174	RVLRLHAQ	
73	167-175	RVLRLHAQP	
74	168-175	VLRHLAQP	

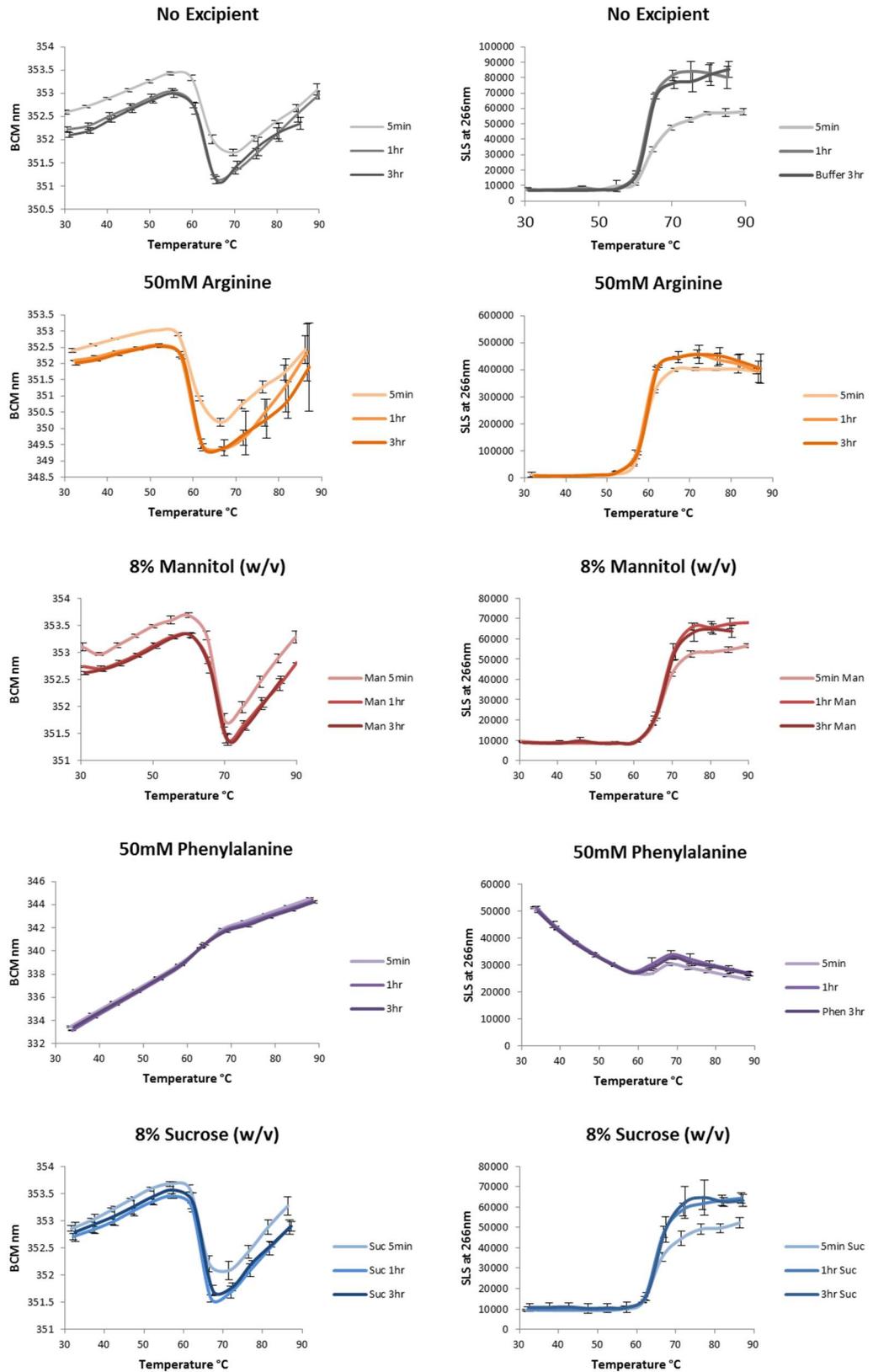
Appendix 12. Peptide information for L71W, Q107Y and Q131F differential uptake plots.

Peptide number	Peptide residue range	Sequence	Region
1	1-10	MTPLGPASSL	LoopA
2	1-13	MTPLGPASSLPQS	
3	1-14	MTPLGPASSLPQSF	
4	1-15	MTPLGPASSLPQSFL	
5	1-15	MTPLGPASSLPQSFL	
6	3-14	PLGPASSLPQSF	
7	5-15	GPASSLPQSFL	αA
8	6-15	PASSLPQSFL	
9	15-21	LLKCLEQ	
10	16-32	LKCLEQVRKIQGDGAAL	
11	17-32	KCLEQVRKIQGDGAAL	
12	19-32	LEQVRKIQGDGAAL	

13	20-32	EQVRKIQGDGAAL		
14	22-32	VRKIQGDGAAL		
15	22-34	VRKIQGDGAALQE		
16	33-39	QEKLCAT	LoopAB	
17	40-47	YKLCHPEE		
18	40-48	YKLCHPEEL		
19	41-47	KLCHPEE		
20	41-48	KLCHPEEL		
21	42-48	LCHPEEL		
22	51-65	LGHSLGIPWAPLSSC		
23	51-68	LGHSLGIPWAPLSSCPSQ		
24	51-69	LGHSLGIPWAPLSSCPSQA		
25	66-72	PSQALQL		
26	70-76	LQLAGCL		α B
27	76-86	LSQLHSGFLY		
28	77-83	SQLHSGL		
29	77-84	SQLHSGLF		
30	84-90	FLYQGLL		
31	84-91	FLYQGLLQ		
32	84-92	FLYQGLLQA		
33	85-92	LYQGLLQA		
34	86-92	YQGLLQA		
35	90-97	LQALEGIS	LoopBC	
36	91-97	QALEGIS		
37	91-104	QALEGISPELGPTL		
38	91-106	QALEGISPELGPTLDT		
39	93-101	LEGISPELG		
40	93-104	LEGISPELGPTL		
41	93-106	LEGISPELGPTLDT	α C	
42	94-101	EGISPELG		

43	94-104	EGISPELGPTL	
44	94-106	EGISPELGPTLDT	
45	96-104	ISPELGPTL	
46	96-106	ISPELGPTLDT	
47	98-104	PELGPTL	
48	98-106	PELGPTLDT	
49	100-110	LGPTLDTLQLD	
50	114-120	FATTIWQ	
51	114-124	FATTIWQMEE	
52	115-121	ATTIWQQ	
53	115-122	ATTIWQQM	
54	115-124	ATTIWQMEE	
55	116-122	TTIWQQM	
56	118-124	IWQQMEE	
57	121-137	QMEELGMAPALQPTQGA	LoopCD
58	125-138	LGMAPALQPTQGAM	
59	126-135	GMAPALQPTQ	
60	126-138	GMAPALQPTQGAM	
61	126-141	GMAPALQPTQGAMPAF	
62	129-138	PALQPTQGAM	
63	129-141	PALQPTQGAMPAF	
64	130-140	ALQPTQGAMPA	
65	132-138	QPTQGAM	
66	139-153	PAFASAFQRRAGGVL	
67	146-153	QRRAGGVL	
68	154-161	VASHLQSF	
69	166-175	YRVLRLHAQP	LoopD
70	167-174	RVLRLHAQ	
71	167-175	RVLRLHAQP	
72	168-175	VLRHLAQP	

Appendix 13. Effect of sample equilibration time on GCSF SLS and fluorescence measurements.



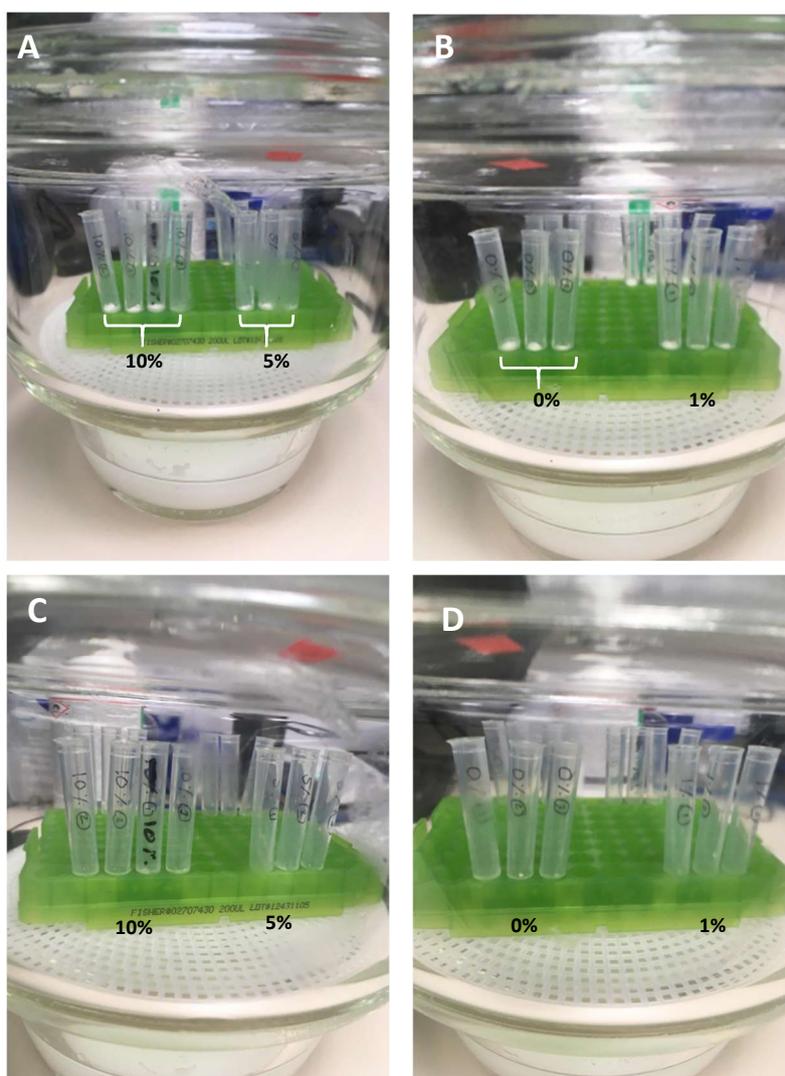
Appendix 14. Peptide information for Differential plots of lyophilised GCSF peptide uptake with and without sucrose, as measured by ssHDX-MS.

Peptide number	Peptide residue range	Sequence	Region
1	1-10	MTPLGPASSL	LoopA
2	1-13	MTPLGPASSLPQS	
3	1-14	MTPLGPASSLPQSF	
4	1-14	MTPLGPASSLPQSF	
5	1-15	MTPLGPASSLPQSFL	
6	1-15	MTPLGPASSLPQSFL	
7	3-14	PLGPASSLPQSF	
8	3-15	PLGPASSLPQSFL	
9	4-14	LGPASSLPQSF	
10	5-15	GPASSLPQSFL	
11	6-15	PASSLPQSFL	
12	9-22	SLPQSFLKCLEQV	
13	15-21	LLKCLEQ	
14	17-32	KCLEQVRKIQGDGAAL	
15	19-32	LEQVRKIQGDGAAL	
16	20-32	EQVRKIQGDGAAL	
17	21-32	QVRKIQGDGAAL	
18	22-32	VRKIQGDGAAL	
19	22-33	VRKIQGDGAALQ	
20	23-32	RKIQGDGAAL	LoopAB
21	33-39	QEKLCAT	
22	40-47	YKLCHPEE	
23	41-47	KLCHPEE	
24	41-48	KLCHPEEL	
25	42-48	LCHPEEL	
26	49-64	VLLGHSLGIPWAPLSS	
27	49-69	VLLGHSLGIPWAPLSSCPSQA	
28	51-65	LGHSLGIPWAPLSSC	
29	51-68	LGHSLGIPWAPLSSCPSQ	
30	51-69	LGHSLGIPWAPLSSCPSQA	
31	58-65	PWAPLSSC	
32	61-69	PLSSCPSQA	α B
33	66-72	PSQALQL	
34	70-76	LQLAGCL	
35	76-86	LSQLHSGLFLY	
36	77-83	SQLHSGL	
37	77-84	SQLHSGLF	
38	84-90	FLYQGLL	
39	85-92	LYQGLLQA	

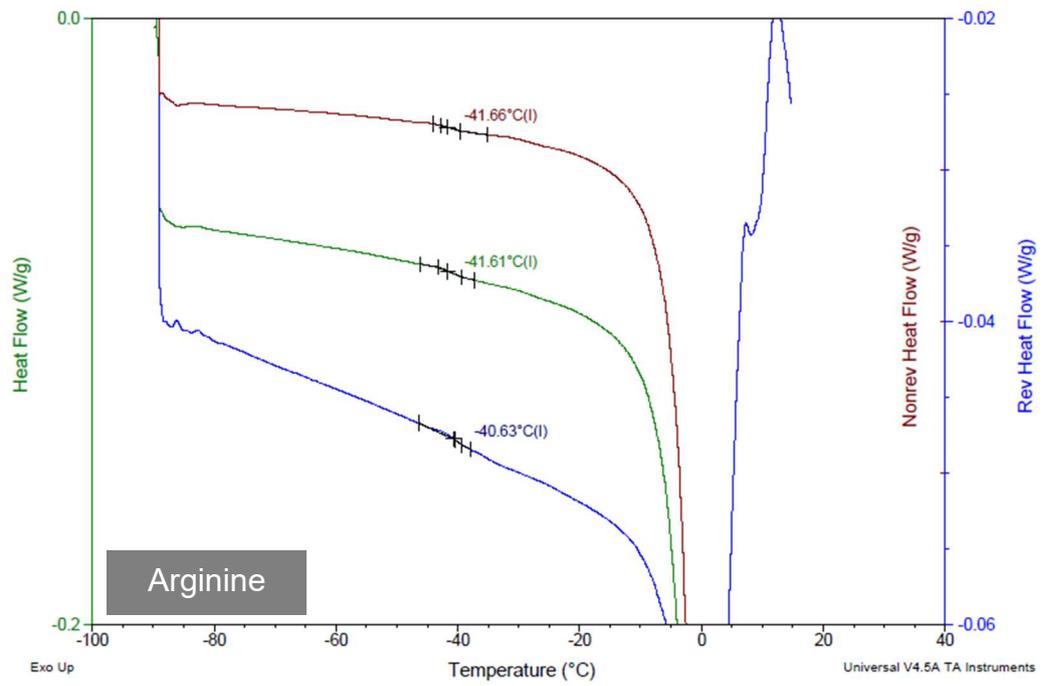
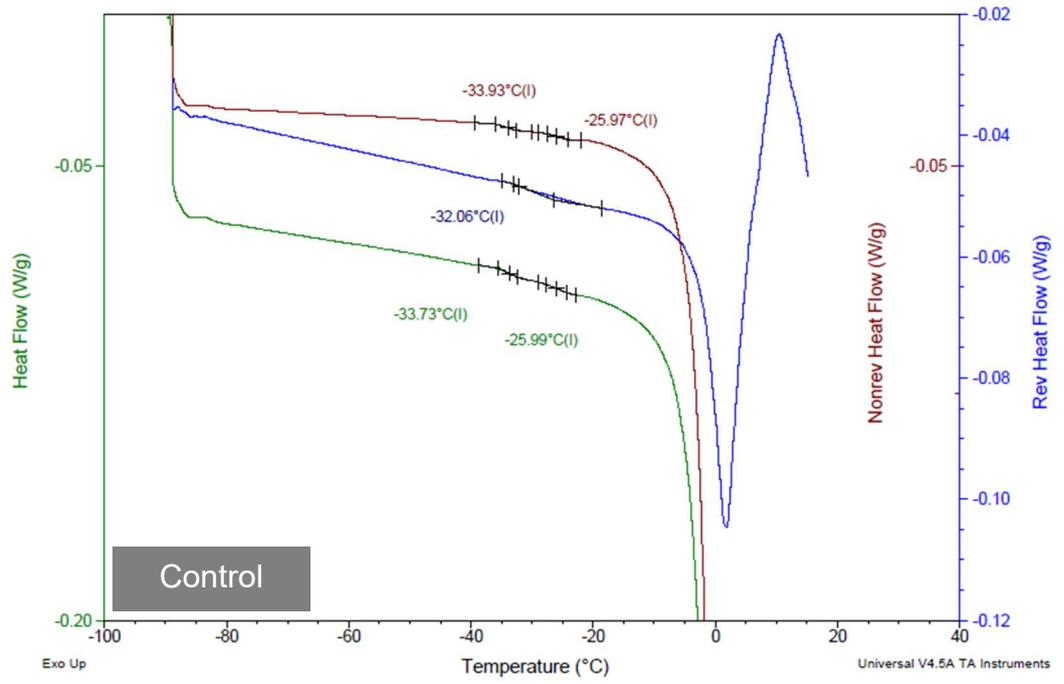
40	86-92	YQGLLQA	
41	90-97	LQALEGIS	LoopBC
42	91-97	QALEGIS	
43	91-101	QALEGISPELG	
44	91-104	QALEGISPELGPTL	
45	91-106	QALEGISPELGPTLDT	
46	92-106	ALEGISPELGPTLDT	α C
47	93-104	LEGISPELGPTL	
48	93-106	LEGISPELGPTLDT	
49	94-104	EGISPELGPTL	
50	94-106	EGISPELGPTLDT	
51	95-104	GISPELGPTL	
52	96-104	ISPELGPTL	
53	96-106	ISPELGPTLDT	
54	97-104	SPELGPTL	
55	98-104	PELGPTL	
56	98-106	PELGPTLDT	
57	100-110	LGPTLDTLQLD	
58	100-111	LGPTLDTLQLDV	
59	111-121	VADFATTIWQQ	
60	114-120	FATTIWQ	
61	114-122	FATTIWQQM	
62	114-124	FATTIWQQMEE	
63	115-121	ATTIWQQ	
64	115-122	ATTIWQQM	
65	115-124	ATTIWQQMEE	
66	116-122	TTIWQQM	
67	118-124	IWQQMEE	
68	123-138	EELGMAPALQPTQGAM	LoopCD
69	123-138	EELGMAPALQPTQGAM	
70	123-141	EELGMAPALQPTQGAMPAF	
71	125-138	LGMAPALQPTQGAM	
72	125-141	LGMAPALQPTQGAMPAF	
73	125-141	LGMAPALQPTQGAMPAF	
74	126-135	GMAPALQPTQ	
75	126-138	GMAPALQPTQGAM	
76	126-141	GMAPALQPTQGAMPAF	
77	127-136	MAPALQPTQG	
78	127-137	MAPALQPTQGA	
79	128-138	APALQPTQGAM	
80	129-138	PALQPTQGAM	
81	129-141	PALQPTQGAMPAF	
82	129-141	PALQPTQGAMPAF	
83	130-139	ALQPTQGAMP	
84	130-140	ALQPTQGAMPA	

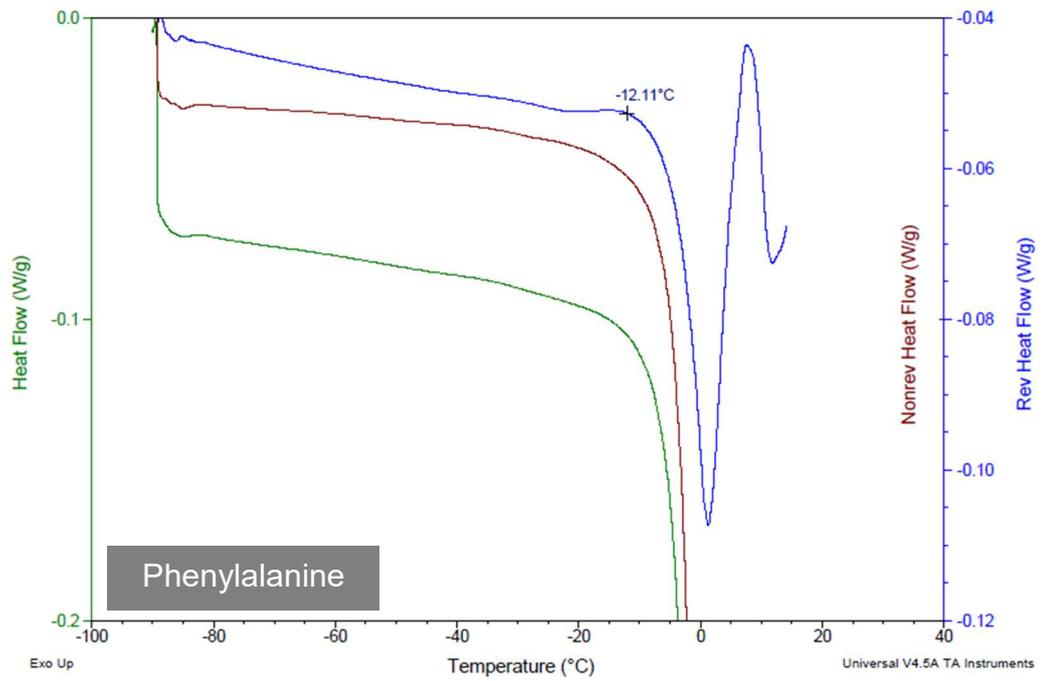
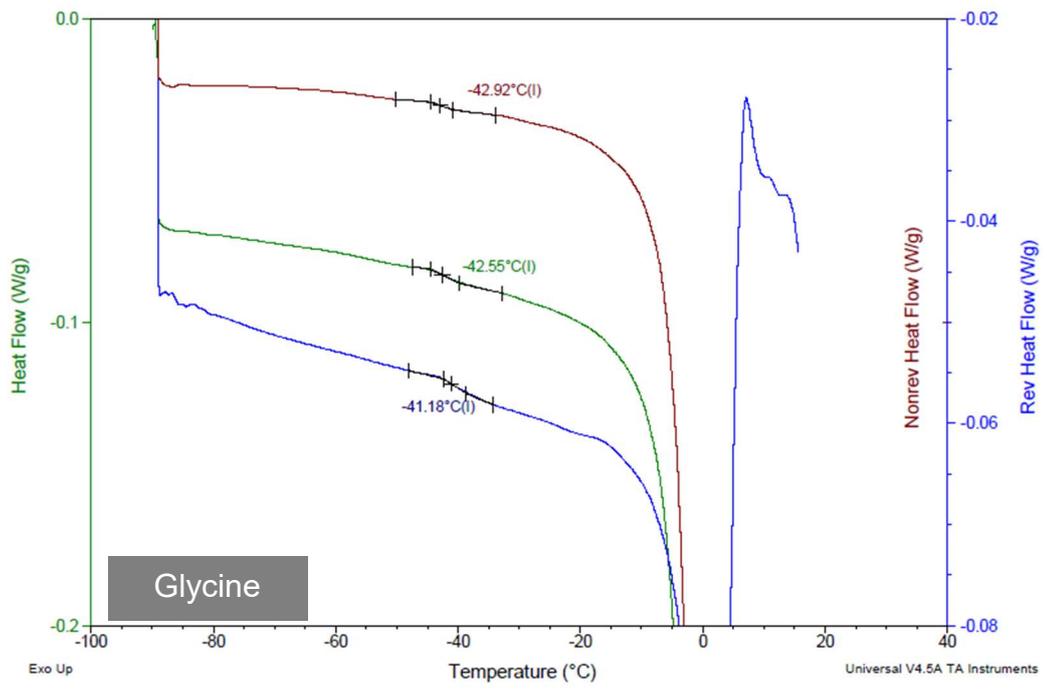
85	132-138	QPTQGAM	αD
86	132-141	QPTQGAMPAF	
87	133-141	PTQGAMPAF	
88	142-153	ASAFQRRAGGVL	
89	143-153	SAFQRRAGGVL	
90	145-152	FQRRAGGV	
91	146-152	QRRAGGV	
92	146-153	QRRAGGVL	
93	154-161	VASHLQSF	LoopD
94	166-175	YRVLRLHLAQP	
95	167-174	RVLRLHLAQP	
96	167-175	RVLRLHLAQP	

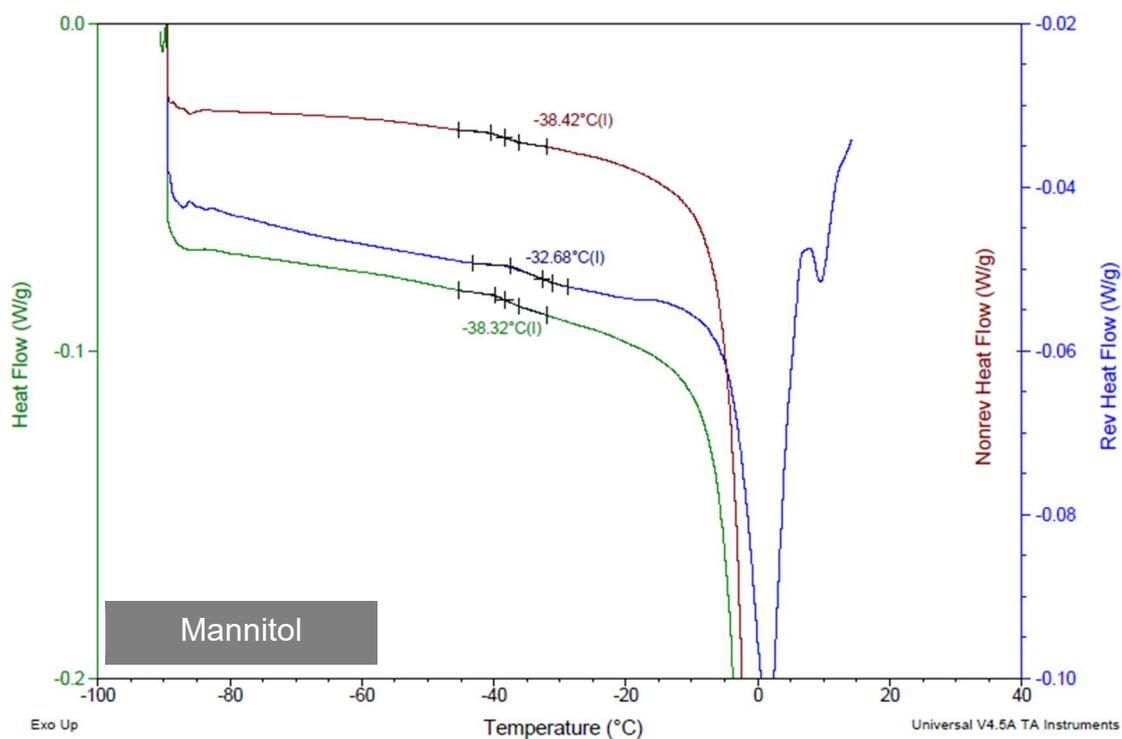
Appendix 15. **Images of moisture-induced USD cake collapse.** Different concentrations of sucrose (black writing; % w/v), formulated in 50 mM citric acid pH 4.25 were lyophilised in 1.4 mL Micronic tubes and placed within sealed desiccators at 43% RH. Images were taken at the start of incubation (A and B), and after 3 days (C and D).



Appendix 16. Modulated DSC figures







Appendix 17. Peptide information for Differential uptake plots for GCSF lyophilised with mannitol as measured by ssHDX-MS & Differential uptake plots for GCSF lyophilised with phenylalanine as measured by ssHDX-MS

Peptide number	Peptide residue range	Sequence	Region
1	1-10	MTPLGPASSL	LoopA
2	1-13	MTPLGPASSLPQS	
3	1-14	MTPLGPASSLPQSF	
4	1-15	MTPLGPASSLPQSFL	
5	1-15	MTPLGPASSLPQSFL	
6	3-14	PLGPASSLPQSF	
7	3-15	PLGPASSLPQSFL	
8	4-14	LGPASSLPQSF	
9	5-15	GPASSLPQSFL	αA
10	6-15	PASSLPQSFL	
11	9-22	SLPQSFLKCLEQV	
12	15-21	LLKCLEQ	
13	17-32	KCLEQVRKIQGDGAAL	
14	19-32	LEQVRKIQGDGAAL	
15	20-32	EQVRKIQGDGAAL	
16	21-32	QVRKIQGDGAAL	
17	22-32	VRKIQGDGAAL	
18	22-33	VRKIQGDGAALQ	

19	23-32	RKIQGDGAAL	
20	33-39	QEKLCAAT	LoopAB
21	40-47	YKLCHPEE	
22	41-47	KLCHPEE	
23	41-48	KLCHPEEL	
24	42-48	LCHPEEL	
25	49-69	VLLGHSLGIPWAPLSSCPSQA	
26	51-65	LGHSLGIPWAPLSSC	
27	51-68	LGHSLGIPWAPLSSCPSQ	
28	51-69	LGHSLGIPWAPLSSCPSQA	
29	58-65	PWAPLSSC	
30	61-69	PLSSCPSQA	
31	66-72	PSQALQL	α B
32	70-76	LQLAGCL	
33	76-86	LSQLHSGLFLY	
34	77-83	SQLHSGL	
35	77-84	SQLHSGLF	
36	84-90	FLYQGLL	
37	85-92	LYQGLLQA	
38	86-92	YQGLLQA	LoopBC
39	90-97	LQALEGIS	
40	91-97	QALEGIS	
41	91-101	QALEGISPELG	
42	91-104	QALEGISPELGPTL	
43	91-106	QALEGISPELGPTLDT	
44	92-106	ALEGISPELGPTLDT	
45	93-104	LEGISPELGPTL	
46	93-106	LEGISPELGPTLDT	α C
47	94-104	EGISPELGPTL	
48	94-106	EGISPELGPTLDT	
49	95-104	GISPELGPTL	
50	96-104	ISPELGPTL	
51	96-106	ISPELGPTLDT	
52	97-104	SPELGPTL	
53	98-104	PELGPTL	
54	98-106	PELGPTLDT	
55	100-110	LGPTLDTLQLD	
56	100-111	LGPTLDTLQLDV	
57	111-121	VADFATTIWQQ	
58	114-120	FATTIWQ	
59	114-122	FATTIWQQM	
60	114-124	FATTIWQQMEE	
61	115-121	ATTIWQQ	
62	115-122	ATTIWQQM	
63	115-124	ATTIWQQMEE	

64	116-122	TTIWQQM	
65	118-124	IWQQMEE	LoopCD
66	123-138	EELGMAPALQPTQGAM	
67	123-138	EELGMAPALQPTQGAM	
68	123-141	EELGMAPALQPTQGAMPAF	
69	125-138	LGMAPALQPTQGAM	
70	125-141	LGMAPALQPTQGAMPAF	
71	125-141	LGMAPALQPTQGAMPAF	
72	126-135	GMAPALQPTQ	
73	126-138	GMAPALQPTQGAM	
74	126-141	GMAPALQPTQGAMPAF	
75	127-136	MAPALQPTQG	
76	127-137	MAPALQPTQGA	
77	128-138	APALQPTQGAM	
78	129-138	PALQPTQGAM	
79	129-141	PALQPTQGAMPAF	
80	129-141	PALQPTQGAMPAF	
81	130-139	ALQPTQGAMP	
82	130-140	ALQPTQGAMPA	
83	132-138	QPTQGAM	
84	133-141	PTQGAMPAF	
85	142-153	ASAFQRRAGGVL	
86	143-153	SAFQRRAGGVL	
87	145-152	FQRRAGGV	
88	146-152	QRRAGGV	
89	146-153	QRRAGGVL	
90	154-161	VASHLQSF	LoopD
91	166-175	YRVLRLHAQP	
92	167-174	RVLRLHAQ	
93	167-175	RVLRLHAQP	