



UCL

---

# **Freeze-drying of engineered proteins using protein modelling tools and experimental validation**

A thesis submitted to University College London

for the degree of

DOCTOR OF PHILOSOPHY

**Cheng Zhang**

Department of Biochemical Engineering

University College London

2016

Primary supervisor: Professor Paul Dalby (UCL)

Secondary supervisor: Professor Steve Brocchini (UCL)

I, Cheng Zhang, 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.

Signature: \_\_\_\_\_

Date: \_\_\_\_\_

**This thesis is dedicated to my parents**

For their endless love, support and encouragement

# Acknowledgements

I would like to thank my supervisors Professor Paul Dalby and Professor Steve Brocchini for their great help throughout my PhD studies. Thanks to their extensive scientific knowledge and deep insight into current development trend of bioprocessing and formulation, I have been given this long-standing but still vivid project - freeze-drying of proteins. Following their instructions, I have gradually mastered basic ways to do research and familiarised myself with the project background. They help me to build up my theoretical system and indispensable confidence of undertaking research independently.

I would like to thank Dr Paul Matejtschuk from NIBSC. He is very generous to provide me much useful knowledge about freeze-drying. From the engineers' perspective, he has helped me so much to solve practical problems, which may take ages to find a solution for a beginner like me. He has guided me into the laboratory of NIBSC to use freeze dryer, DSC, FDM and made me familiarise with the industrial working environment.

I would like to thank Dr Nesrine Chakroun, Dr Teresa Silva Barata and Dr Keying Li. They are very helpful and always provide constructive advice to carry out my research. Their individual specialties have helped me design experiment more reasonable and rational. I would like to thank David Hilton, Garima Sharma and all the other colleagues from UCL Department of Biochemical Engineering and Department of Pharmaceutics at UCL School of Pharmacy. It is their consistent help and support that enables me to tackle a great deal of difficulties and progress step by step. I would like to thank Dr Shahina Ahmad, who has done the pioneer work for the creating of protein mutants, which has facilitated us with valuable experimental materials. In addition, I would like to acknowledge the use of the UCL Legion High Performance Computing Facility (Legion@UCL), and associated support services, in the completion of this work.

I would like to pay great thanks for my family, especially my parents Hejie Zhang and Xin Jiang and cousin Da Ma. Being a self-funded overseas student, it is their meticulous care and caring that makes me to have the progress so far.

Last but not least, I would like to thank all the other people that have helped me during these years from UCL, my accommodation ILSC and any other places. The excellent study environment and support they have given to me is a great encouragement for me to move my project onwards and upwards.

# Abstract

The development of therapeutic proteins is a driving force in the current manufacture of biopharmaceuticals. Freeze drying is widely used in the fabrication of final dosage forms of therapeutic proteins. Using a series of A33 Fab mutants, this thesis aimed to correlate their physicochemical properties to the outcomes of freeze-drying.

Preliminary studies employed a homogeneous freeze-drying process on 96-well plates. It was found that K65M and K133M surface mutations, the use of acetate buffer, low pH, increased ionic strength, and the use of NaCl, caused the most monomer loss; whereas S75K, C226S, and L50K mutations, high pH, and the use of Na<sub>2</sub>SO<sub>4</sub> caused the least monomer loss.

Several *in-silico* modelling tools were used to design mutants for studying the impact of protein conformational stability. Rosetta software, RMSF and B-factor analyses were used to evaluate the mutant candidates and restrict the mutations mainly located in the flexible regions. Unstable mutants were prepared as controls to validate the prediction accuracy.

In freeze-drying, most of the stabilising mutants had 20% less monomer loss than C226S, while the destabilising ones had 14-46% more monomer loss.  $T_m$  and  $\Delta\Delta G$  estimated the monomer loss in freeze-drying with low degree of accuracy. Compared to freeze-drying, a more distinct difference was observed in the aqueous phase as all the destabilising mutants aggregated more than 5 times faster than C226S and the stabilising mutants did.  $T_m$  correlated well with the aggregation in aqueous phase, indicating conformational stability was more important in aqueous phase than that in freeze-drying.

In addition, excipients barely exerted influence on the stable mutants but provided sufficient protection for the unstable ones, which was reflected by their correlations to  $T_m$  values. The rank-order of excipient effects for individual mutants, relative to that of wild type, became less similar as the mutant  $\Delta T_m$  magnitude increased.

# Table of Contents

Freeze-drying of engineered proteins using protein modelling tools and experimental validation .....	1
Acknowledgements .....	4
Abstract.....	6
Nomenclature .....	14
Units.....	17
List of Figures .....	18
List of Tables .....	24
1 Introduction .....	26
1.1 The significance of freeze-drying for proteins .....	26
1.2 The freeze-drying application to overcome the limitation in aqueous formulation.....	27
1.2.1 Protein aggregation .....	27
1.2.2 Improved storage condition .....	28
1.2.3 High concentration dosage form upon administration.....	28
1.3 Basic introduction of freeze-drying .....	29
1.3.1 Freeze dryer equipment .....	29
1.3.2 Three steps in freeze-drying.....	31
1.4 Interplay of freeze-drying parameters.....	42
1.4.1 Formulation and freezing.....	42
1.4.2 Primary drying.....	50
1.4.3 Secondary drying.....	57
1.4.4 Real-time monitoring for product properties.....	60

1.5	Analytical methods to characterise protein properties.....	62
1.5.1	Size/Conformation/Structure .....	62
1.5.2	Surface area .....	65
1.5.3	Water content and sorption .....	66
1.5.4	Thermal analysis.....	68
1.5.5	(Thermo) Imaging analysis .....	71
1.5.6	Storage conditions.....	73
1.5.7	Reconstitution .....	74
1.5.8	Robustness verification studies for process development.....	75
1.6	Summary of protein freeze-drying research .....	76
1.7	Guidelines for systematic development of freeze-drying processes for proteins	77
1.8	The influence of Gibbs free energy on protein stability in freeze-drying and liquid formulation.....	80
1.9	Aims and objectives .....	82
1.9.1	The gap in current freeze-drying research .....	82
1.9.2	Aims and objectives of the project.....	83
2	Materials and Methods.....	86
2.1	Materials.....	86
2.1.1	Buffer .....	86
2.1.2	Growth media for <i>E. coli</i> .....	87
2.1.3	Model proteins .....	87
2.2	Methods .....	88
2.2.1	Plasmid Mutagenesis .....	88

2.2.2	Production of cell and plasmid stocks .....	89
2.2.3	Pilot scale Fab production .....	92
2.2.4	Protein quantification .....	97
2.2.5	SDS-PAGE .....	98
2.2.6	DNA gel electrophoresis.....	98
2.2.7	Freeze-drying.....	98
2.2.8	Size-exclusion chromatography HPLC (SEC-HPLC).....	100
3	Impact of surface-charge mutations on the freeze-drying of Fab.....	102
3.1	Introduction .....	102
3.1.1	Mutation impact by A/G substitution.....	104
3.1.2	Mutation impact by proline.....	105
3.1.3	Mutation impact on protein charge and hydrophobicity .....	105
3.1.4	Mutation impact on secondary structure .....	107
3.1.5	Aims of the chapter.....	108
3.2	Materials and Methods.....	108
3.2.1	Homogeneity of drying across the freeze-dryer shelf .....	108
3.2.2	Differential scanning calorimetry (DSC) .....	110
3.2.3	Freeze drying microscopy (FDM) .....	111
3.2.4	X-ray diffraction (XRD) .....	111
3.2.5	Fab sample preparation.....	111
3.2.6	Liquid formulations.....	112
3.2.7	Size-exclusion chromatography HPLC (SEC-HPLC).....	112
3.2.8	Freeze-drying.....	112
3.2.9	Cake morphology scoring and reconstitution .....	113

3.2.10	Net charge calculation .....	113
3.2.11	Sweet plot .....	114
3.3	Results and discussion .....	114
3.3.1	Homogeneity of drying across the freeze-dryer shelf .....	114
3.3.2	DSC and FDM to validate the freeze-drying cycle parameters.....	117
3.3.3	Formulation and mutant effects on monomer loss .....	120
3.3.4	Formulation and mutant effects on cake morphology .....	138
3.3.5	Monomer loss vs cake score .....	144
3.3.6	Sweet plot for “monomer loss” and “cake morphology” .....	145
3.4	Conclusion .....	145
4	Hybrid mutagenesis design and pilot scale production .....	148
4.1	Introduction .....	148
4.1.1	Bioinformatics tools to indicate the disorder of residues.....	148
4.1.2	Design of stable proteins based on consensus tools.....	150
4.1.3	Homology modelling .....	151
4.1.4	Computational prediction of protein $\Delta\Delta G$ upon mutation .....	152
4.1.5	Rosetta.....	154
4.1.6	Aims of the chapter.....	156
4.2	Materials and Methods.....	156
4.2.1	<i>In-silico</i> mutagenesis .....	156
4.2.2	Analysis of residue flexibility .....	158
4.2.3	Design of stable and unstable mutant candidates .....	160
4.2.4	Primer sequence design.....	162
4.2.5	Laboratory production.....	163

4.3	Results and Discussion.....	164
4.3.1	Analysis of residue flexibility.....	164
4.3.2	<i>In-silico</i> mutagenesis.....	168
4.3.3	Designed mutants and corresponding primer sequences.....	183
4.3.4	Laboratory production.....	184
4.4	Conclusions.....	186
5	Freeze-drying tolerance and thermal stability measurement for mutant Fabs	188
5.1	Introduction .....	188
5.1.1	Experimental methods to measure the $\Delta G$ and melting temperature	188
5.1.2	Freeze-drying formulation of antibodies.....	189
5.1.3	Aims of the chapter.....	190
5.2	Materials and Methods.....	191
5.2.1	Freeze-drying of the new mutants.....	191
5.2.2	Melting temperature ( $T_m$ ) measurement.....	191
5.3	Results and discussion .....	192
5.3.1	Freeze-drying for the designed mutants.....	192
5.3.2	Thermal stability of designed mutants and their relations to freeze-drying	194
5.3.3	Rosetta's $\Delta\Delta G$ prediction upon $T_m$ and freeze-drying monomer loss	198
5.4	Conclusion .....	201
6	Liquid aggregation kinetics for designed mutant formulations .....	203
6.1	Introduction .....	203
6.1.1	Antibody aggregation in liquid formulations .....	203

6.1.2	Formulation with excipients to mitigate antibody aggregation in the liquid state	204
6.1.3	Aims of the chapter.....	205
6.2	Materials and methods.....	206
6.2.1	Liquid aggregation kinetics .....	206
6.2.2	Melting temperature ( $T_m$ ) of C226S formulated with excipients .....	207
6.3	Results and discussion .....	207
6.3.1	Liquid solution condition scouting to study liquid aggregation kinetics	207
6.3.2	Effect of mutations upon liquid aggregation kinetics .....	209
6.3.3	Effect of excipients upon liquid aggregation kinetics .....	217
6.3.4	The $T_m$ with excipients and its correlation with liquid aggregation kinetics	222
6.4	Conclusion .....	224
7	Conclusion .....	226
8	Future work .....	228
8.1	Short-term future work .....	228
8.1.1	Improve the liquid kinetics operations .....	228
8.1.2	Compare aggregation driven by global unfolding and native states ..	229
8.1.3	Examine the excipients' stabilising effect .....	229
8.1.4	Analyse the degradation stresses in freeze-drying .....	230
8.2	Long-term future work.....	231
8.2.1	Improve the mutagenesis strategy .....	231
8.2.2	Standardised formulation studies .....	231
8.2.3	Study other antibody structures.....	232

8.2.4	Activity.....	232
9	Reference.....	233
10	Appendix .....	257
10.1	Pymol visualisation for the PDB of C226S .....	257
10.2	Rosetta script.....	257
10.2.1	Mutfile .....	257
10.2.2	Option file.....	258
10.3	The $\Delta\Delta G$ of 8398 mutants .....	258
10.4	The $\Delta\Delta G$ of all the stabilising mutants and their relation with CDR regions, RMSF and Depth .....	264
10.5	The sequence alignment for C226S from human Fab PDB.....	269
10.6	The correlation between $\Delta T_m$ , $\Delta\Delta G$ .....	275
10.7	GROMACS code.....	276
10.7.1	Code for “job.sh” file .....	276

# Nomenclature

API	Active Pharmaceutical Ingredients
ACN	Acetonitrile
BET	Brunauer, Emmett and Teller
BSA	Bovine Serum Albumin
CD	Circular Dichroism
CESEM	Cryoenvironmental Scanning Electronic Microscopy
CV	Column Volume
DEA	Dielectric Analysis
DLS	Dynamic Laser Scattering
DMTA, DMA	Dynamic Mechanical Thermal Analysis
DoE	Design of Experiment
DOT	Dissolved Oxygen Tension
DSC	Differential Scanning Calorimetry
DTA	Differential Thermal Analysis
DVS	Dynamic Vapour Sorption
EDTA	Ethylene Diamine Tetraacetic Acid
Fab	Fragment Antigen-Binding
FDM	Freeze-Drying Microscopy
FPLC	Fast Protein Liquid Chromatography
FTIR (IR)	Fourier Transform Infrared Spectroscopy
G-CSF	Granulocyte Colony-Stimulating Factor
HIC	Hydrophobic Interaction Chromatography
HCP	Host Cell Protein
HPLC	High Performance Liquid Chromatography
IEF	Isoelectric Focusing
IMC	Isothermal Microcalorimetry

IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
KF	Karl Fischer
LB	Luria Bertani
LDH	Lactate Dehydrogenase
LOD	Loss on Drying
MTDSC	Modulated Temperature DSC
NMR	Nuclear Magnetic Resonance
OD	Optical Density
PBS	Phosphate Buffer Saline
PEG	Poly(ethylene glycol)
PHC	Polyhydroxide Compound
PPG	Poly(propylene glycol)
PSD	Particle Size Distribution
RP	Reverse Phase (Chromatography)
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SEM	Scanning Electronic Microscopy
SLS	Static Light Scattering
SOP	Standard Operation Protocol
SVP	Saturation Vapour Pressure
$T_g, T_g$	Glass Transition Temperature
$T_c$	Collapse Temperature
$T_e$	Eutectic Point
TEA	Thermoelectric Analysis
TFA	Trifluoroacetic Acid
TGA	Thermogravimetric Analysis
TMA	Thermal Mechanical Analysis
TSC	Thermally Stimulated Current

UV	Ultraviolet
WT	Wild Type
XRPD, XRD	X-Ray Powder Diffractometry

# Units

°C	degrees Celsius
min	minute
h	hour
Å	angstrom
bp	base pair
rpm	revolutions per minute
Da	Daltons
μ	micro
m	milli
M	molar
g	gram
L	litre
mAU	milli absorbance units
OD	optical density
pH	$-\log_{10}[\text{H}^+]$
pI	isoelectric point
v/v	volume per volume
w/v	weight per volume

# List of Figures

Figure 1.1 A typical freeze dryer.....	30
Figure 1.2 The description of crystalline and amorphous states.....	32
Figure 1.3 A phase diagram for sucrose-water system.....	35
Figure 1.4 Undercooling was found to preserve LDH activity .....	36
Figure 1.5 The phase diagram for water .....	40
Figure 1.6 Solid/liquid phase diagram for the binary water-NaCl system .....	48
Figure 1.7 Diagram of coupled heat transfer and mass transfer.....	51
Figure 1.8 Effect of chamber pressure on sublimation rate .....	53
Figure 1.9 Contributions of radiation (left, diagonals), conduction (centre, cross-hatched) and convection (right, blank) to the total heat transfer for three vial types and two pressures .....	53
Figure 1.10 An ideal operation of adjusting temperature between $T_s$ and $T_g$ .....	59
Figure 1.11 A more practical way of adjusting the product temperature between $T_g$ and $T_s$ .....	59
Figure 1.12 The CD features of pure secondary structure elements at far-UV region .....	65
Figure 1.13 A typical loop approach to optimise the freeze-drying process .....	78
Figure 1.14 An example of a rational way to systematically undertake freeze-drying process development .....	79
Figure 1.15 Schematic diagram for the strategy to conduct the study of freeze-drying of proteins .....	84
Figure 2.1 A typical SEC-HPLC chromatogram. ....	100
Figure 3.1 The layout of the homogeneity test with 96-well plates .....	110
Figure 3.2 The evaluation criteria for cake morphology .....	113
Figure 3.3 The mapping of water remaining in the four 96-well plates for the homogeneity test of shelf .....	116
Figure 3.4 A proposed layout for freeze-drying in 96-well microplates.....	117
Figure 3.5 The DSC profile for Fab. ....	118

Figure 3.6 Freeze-drying microscopy for 1 mg/ml Fab at pH 7, and NaCl to 200 mM ionic strength.....	119
Figure 3.7 The monomer loss of Fab mutants determined by SEC-HPLC.....	121
Figure 3.8 Average monomer loss for formulations with different pH.....	123
Figure 3.9 The impact of pH for each mutant upon monomer loss.....	125
Figure 3.10 Average monomer loss for S75K formulations with/without acetate at pH 4 and 5.....	126
Figure 3.11 Average monomer loss for formulations at different salt types and levels of ionic strength .....	127
Figure 3.12 Average monomer loss for formulations at different salt types, levels of ionic strength and mutants.....	130
Figure 3.13 The X-ray diffraction patterns for NaCl and Na <sub>2</sub> SO <sub>4</sub> conditions with and without Fab .....	131
Figure 3.14 Average monomer loss for the different mutants.....	133
Figure 3.15 The overlay of the predicted structure before and after the K65M mutation.....	135
Figure 3.16 The overlay of the predicted structure before and after the K133M mutation .....	135
Figure 3.17 The overlay of the predicted structure before and after the S75K mutation.....	136
Figure 3.18 The overlay of the predicted structure before and after the L50K mutation.....	136
Figure 3.19 The cake rating of freeze-dried Fab mutants averaged from two freeze-drying batches, except for K133M with one batch. ....	138
Figure 3.20 The impact of pH for each mutant upon cake morphology.....	139
Figure 3.21 The correlations between monomer loss and cake mass at different pH....	140
Figure 3.22 Overall cake morphology score for formulations adjusted with NaCl or Na <sub>2</sub> SO <sub>4</sub> .....	141
Figure 3.23 Average cake morphology score for formulations at different ionic strength adjusted with NaCl or Na <sub>2</sub> SO <sub>4</sub> .....	141
Figure 3.24 The theoretical mass formulated in each well by considering buffer salt, neutral salt (NaCl or Na <sub>2</sub> SO <sub>4</sub> ) and protein .....	143

Figure 3.25 The correlation between cake score and monomer loss .....	144
Figure 3.26 The sweet plot for freezing drying performance with cake score > 0.6 and monomer loss < 0.15 .....	145
Figure 4.1 The schematic diagram for processing the B-factor .....	160
Figure 4.2 The schematic diagram to design stable mutants .....	161
Figure 4.3 The overlay of C226S secondary structure and average RMSF at pH 4, 200 mM ionic strength .....	164
Figure 4.4 normalised B-factors of light chain and heavy chain .....	165
Figure 4.5 The normalised B-factor after window averaging with 5 neighbouring residues .....	166
Figure 4.6 The correlation between RMSF and B-factor.....	167
Figure 4.7 A breakdown of secondary structures within the correlation between RMSF and B-factor derived from Figure 4.6 .....	167
Figure 4.8 $\Delta\Delta G$ for 8398 (19 mutation/residue x 442 residue) candidates.....	169
Figure 4.9 Histogram for the mutant frequency distribution based on $\Delta\Delta G$ .....	170
Figure 4.10 A blue-white-red plot to represent the locations of stable mutants.....	171
Figure 4.11 A blue-white-red plot to represent the residue depth of C226S Fab. ....	172
Figure 4.12 The correlation between lowest $\Delta\Delta G$ of 19 mutants and the depth from bulk solvent for each residue. ....	172
Figure 4.13 The occurring times for the 20 amino acids among the 442 residues in the C226S.....	176
Figure 4.14 The average RMSF of individual residues of C226S derived from Figure 4.3 .....	177
Figure 4.15 The average depth of individual residues of C226S derived from Figure 4.11 .....	177
Figure 4.16 The correlation between RMSF (Figure 4.3) and depth (Figure 4.11) of C226S.....	178
Figure 4.17 The relation between RMSF, CDR and sum of the stabilising $\Delta\Delta G$ of each residue.....	178

Figure 4.18 The total $\Delta\Delta G$ for stabilising mutants grouped by the type of amino acids derived from Figure 10.2. ....	179
Figure 4.19 The weighted RMSF of individual amino acid substitutions based on their $\Delta\Delta G$ stabilising effect .....	179
Figure 4.20 The weighted depth of individual amino acid substitutions based on their $\Delta\Delta G$ stabilising effect .....	180
Figure 4.21 An illustration for the calculation for the weighted RMSF and weighted depth in Figure 4.19 and Figure 4.20 derived from Figure 10.2. ....	180
Figure 4.22 A summation for the stabilising $\Delta\Delta G$ based on the type of residues.....	181
Figure 4.23 The positions of the top 25 destabilising mutations .....	182
Figure 5.1 The averaged monomer loss for the stabilising mutants during freeze-drying .....	193
Figure 5.2 The averaged monomer loss for the destabilising mutants during freeze-drying .....	193
Figure 5.3 The normalised monomer loss for the stabilising and destabilising mutants during freeze-drying.....	194
Figure 5.4 The melting temperature ( $T_m$ ) for the stabilising mutants measured by Optim1000 at pH 7 .....	196
Figure 5.5 The melting temperature ( $T_m$ ) for the destabilising mutants measured by Optim1000 at pH 7 .....	196
Figure 5.6 The temperature difference of $T_m$ ( $\Delta T_m$ ) for the stabilising and destabilising mutants against pseudo wild-type C226S at pH 7 .....	197
Figure 5.7 The correlation between normalised monomer loss in freeze-drying and $\Delta T_m$ against C226S .....	198
Figure 5.8 The $\Delta\Delta G$ calculated by Rosetta for stabilising (green) and destabilising (red) mutants. ....	199
Figure 5.9 The correlation between $\Delta T_m$ and $\Delta\Delta G$ .....	200
Figure 5.10 The correlation between normalised monomer loss in freeze-drying and $\Delta\Delta G$ .....	201

Figure 5.11 A summary of correlations between freeze-drying aggregation rate, $T_m$ and $\Delta\Delta G$ .	201
Figure 6.1 Monomer retention kinetics at 65°C for 1 mg/ml C226S, 20 mM acetate, pH 5, with NaCl to bring the total ionic strength to 200 mM	208
Figure 6.2 Monomer retention kinetics at 45°C for 1 mg/ml C226S, 20 mM citrate, pH 4, with NaCl to bring the total ionic strength to 200 mM	208
Figure 6.3 Monomer retention of kinetics at 65°C for 1 mg/ml C226S, 20 mM citrate, pH 4, with NaCl to bring the total ionic strength to 200 mM	208
Figure 6.4 The monomer retention of liquid kinetics for 1 mg/ml Fab mutants at 20 mM citrate, pH 4 with NaCl to bring the total ionic strength to 200 mM at 65°C	212
Figure 6.5 The normalised aggregation rates derived from Figure 6.4.	213
Figure 6.6 Correlations between $\Delta T_m$ , $\Delta\Delta G$ , normalised freeze-drying monomer loss and normalised monomer loss rate.	214
Figure 6.7 A summary of correlations between liquid aggregation rate, freeze-drying aggregation rate, $T_m$ and $\Delta\Delta G$ .	216
Figure 6.8 The impact of excipients on the liquid monomer retention of mutants.	218
Figure 6.9 The absolute aggregation rate correlations between C226S and designed mutants, and their correlation accuracy as a function of $\Delta T_m$ .	221
Figure 6.10 The $T_m$ of 1 mg/ml C226S added with various excipients at 10 mM sodium phosphate, pH 7.	223
Figure 6.11 The correlations between monomer retention and $T_m$ as impacted by the excipients.	224
Figure 10.1 The C226S structure predicted by Rosetta.	257
Figure 10.2 The $\Delta\Delta G$ of all the stabilising mutants and their relation with CDR regions, RMSF and Depth.	269
Figure 10.3 The correlation between $\Delta T_m$ and $\Delta\Delta G$ without destabilising mutants.	275
Figure 10.4 The correlation between normalised monomer loss in freeze-drying and $\Delta\Delta G$ without destabilising mutants	275



# List of Tables

Table 1.1 Essential points of annealing.....	37
Table 1.2 Collapse phenomena at different temperature.....	39
Table 1.3 Impact of cooling rate on ice crystal dimensions and ice sublimation time .....	42
Table 1.4 Three freezing methods.....	44
Table 1.5 Choices of excipients to protect APIs from various denaturation process during lyophilisation .....	50
Table 1.6 The parameter definitions of Equation 1.1 for heat and mass transfer.....	50
Table 1.7 Conduction of heat .....	52
Table 1.8 The operation margin temperature between product temperature and $T_c$ .....	55
Table 1.9 The forms of residual water after primary drying.....	57
Table 1.10 Temperature monitoring methods .....	60
Table 1.11 Chamber pressure monitoring methods.....	61
Table 1.12 The characterisation of denatured status by RP-HPLC .....	63
Table 1.13 Infrared spectroscopy characterisation for protein denaturation during lyophilisation .....	64
Table 1.14 The analysis of moisture content for freeze-dried solids.....	66
Table 1.15 Interpretation of (Modulated) DSC sensorgram for typical phase transitions in freeze-drying.....	69
Table 1.16 The impact of crystallisation of amorphous solids during storage .....	73
Table 1.17 Undesirable consequences for proteins during reconstitution process .....	74
Table 1.18 A brief summary for the process control and analytical methods during freeze-drying.....	76
Table 2.1 A list of buffer compositions.....	86
Table 2.2 The composition of growth media .....	87
Table 2.3 PCR operating parameters.....	88
Table 2.4 An example for the process cycle parameters of freeze-drying.....	99
Table 3.1 The net charge of mutants at different pH as calculated by PropKa .....	132

Table 4.1 The Complementarity Determining Regions (CDR) of C226S.....	162
Table 4.2 The equations for calculating the $T_m$ of primers from Eurofins .....	163
Table 4.3 The top 25 destabilising mutants.....	181
Table 4.4 The designed stable mutants and corresponding primer sequences .....	183
Table 4.5 The designed unstable mutants and corresponding primer sequences .....	183
Table 4.6 A list of Fab mutants that had been successfully expressed and purified in reasonable amount.....	184
Table 4.7 The pipelines for commercial drug development of therapeutic recombinant antibodies .....	186
Table 6.1 The formulation conditions for liquid aggregation study.....	206
Table 6.2 The formulation conditions to perform the $T_m$ measurement .....	207
Table 10.1 The sequence alignment for heavy chain .....	269
Table 10.2 The sequence alignment for light chain .....	272

# 1 Introduction

## 1.1 The significance of freeze-drying for proteins

After downstream processing, purified proteins need to be formulated into their final dosage forms to preserve activity and efficacy. A therapeutic protein is deemed to be the active pharmaceutical ingredient (API) while the auxiliary inert ingredients are called excipients. The selection of proper dosage forms depends largely on protein characteristics, patient requirements, and the mode by which it is administered into patients.

Aqueous, ready to use formulations of protein therapeutics are desired as they avoid the need for reconstitution at the point of use. However, a significant number of protein pharmaceuticals (biopharmaceuticals) formulated as aqueous solutions, are prone to degrade, denature or aggregate (Edwards & Hrkach 2000). One solution is to formulate such proteins into solid forms with long-term retention of integrity at lower moisture levels (Abdelwahed et al. 2006).

There are various ways to fabricate and to formulate dry powders, including freeze-drying (lyophilisation), spray-drying and spray freeze-drying. Among them, freeze-drying is one of the most common ways to obtain dry powders from protein solutions. During a freeze-drying process, the protein solution is frozen, followed by sublimation of water at reduced pressure and low temperature. Upon removal of the frozen water, a dry cake can be obtained in different solid forms (mostly as amorphous). The solid proteins, together with the inert ingredients, will mostly be reconstituted into a liquid dosage form immediately before injecting into patients (i.e. parenteral administration).

Freeze-drying can be easily carried out under sterile conditions. Sub-zero environments and low moisture conditions can preserve the biological activity of a product. Well-structured cake morphologies, presented as a porous plug, have won great popularity among end-users as these improve the final activity recovery and time required to reconstitute into the final solid dosage of product by rehydration. In addition,

compared to the manufacturing cost of purified active proteins, the cost of freeze-drying accounts for a relatively small proportion (Sharma et al. 2005).

## 1.2 The freeze-drying application to overcome the limitation in aqueous formulation

The instability of proteins in aqueous phase has been reviewed extensively (Wang 1999). The benefit of freeze-drying mainly derived from its reduction in the deleterious chemical reactions as the diffusion of protein molecules is greatly inhibited in the vitrified solid state (Tonnis et al. 2015). This section will focus on the limitation and instabilities of liquid formulation that could potentially be minimised or improved by freeze-drying.

### 1.2.1 Protein aggregation

Protein aggregation is one of the major challenges in formulation studies. Unfolded or partially denatured proteins would associate and form oligomers under processing conditions like expression and purification (Wang et al. 2010). Excipients were found to mitigate the aggregation. For example, sucrose increased the conformation stability and reduced insoluble aggregates of an IgG1 monoclonal antibody (Manikwar & Majumdar 2013). The excipients stabilise protein in aqueous phase through direct binding or indirect interactions (e.g. preferential interaction); while in freeze-drying, excipients play roles as cryoprotectants or lyoprotectants and provide high  $T_g$  in the dried state (Ohtake et al. 2011). When excipients could not provide sufficient stabilising effect in aqueous phase, alternative to freeze-drying was considered. It was found that a lyophilized form of recombinant human epidermal growth factor (rhEGF) decreased degradation rate of 70 times at 50°C compared to aqueous formulation (Santana et al. 2014). Therefore, it would be worthwhile to alter protein degradation pathway and use appropriate excipients in the solid formulation.

## 1.2.2 Improved storage condition

Protein pharmaceuticals are desired to be stored under dried condition and cold temperature to extend their shelf life. The dried state provided reduced hydrolysis and deamidation reactions induced by inappropriate pH or temperature (Ohtake et al. 2011). The low temperature could reduce the protein motion rate, minimise the interaction and aggregation of protein molecules in aqueous phase. However, the refrigeration and transportation cost remains as a concern. Upon fabricating the liquid form into solid state, the motion of proteins along with excipients are highly restricted, therefore providing opportunities for ambient or higher temperature storage conditions.

It was reported that the denaturation temperature ( $T_d$ ) of lysozyme substantially increased from 70°C to more than 120°C as the water content was decreased from more than 25% (w/w) to less than 5% (Martínez et al. 2016). Another study showed that lyophilized IgG1 mAb samples with 6.8% moisture resulted more 2-5  $\mu\text{m}$  particles than samples with 0.6% moisture after accelerated shipping-like stress (Telikepalli et al. 2015). Therefore, it is obvious that a high moisture content is detrimental to preserve the lyophilized protein samples though a minimal moisture content (less than 1%) is necessary (Colandene 2007). In addition, it was often suggested that the storage temperature should be below the  $T_g$  so as to maintain the physical stability of dried product (Breen et al. 2001).

## 1.2.3 High concentration dosage form upon administration

Subcutaneous and intramuscular routes are the common delivery methods to administer the mAb drugs (Neergaard et al. 2014). Due to the limited volume restricted for a single administration dosage, the liquid solution needs to be formulated in an adequately high concentration (>100 mg/ml) (Yearley et al. 2014). As a result, problems include aggregation, solubility and high viscosity rise in such highly concentrated solution (Singh et al. 2014).

Converting the aqueous protein into solid form could mitigate the association of protein molecules in the liquid. A desired high concentration solution could also be achieved upon reconstitution of dried powders without ultra-filtration to concentrate the protein in a liquid state (Harris et al. 2004). Sucrose was used to stabilise the high protein formulation, which caused prolonged reconstitution time (Beech et al. 2015). A systematic study was done to tackle the long reconstitution time issue and it was found that incorporating wetting agents, a high temperature annealing step and reconstitution under vacuum conditions could reduce the time for reconstitution (Cao et al. 2013).

### 1.3 Basic introduction of freeze-drying

Freeze-drying is a widely used technique to obtain dry solids and has been applied in the pharmaceutical and biopharmaceutical industry since 1930s (Constantino & Pikal 2004). The operation of freeze-drying is straightforward. However, the physical chemical theories involved in this technique are fairly complex and some remain to be understood. The product properties and processing parameters are always interdependent, and one needs to have a deep understanding so as to design an appropriate freeze-drying process.

The basic operation procedures are normally similar and there are some terminologies that need to be introduced before elaborating the theories in detail. Therefore, this section aims to introduce the basic theories while the detailed kinetics and thermodynamics would be discussed in the next section.

#### 1.3.1 Freeze dryer equipment

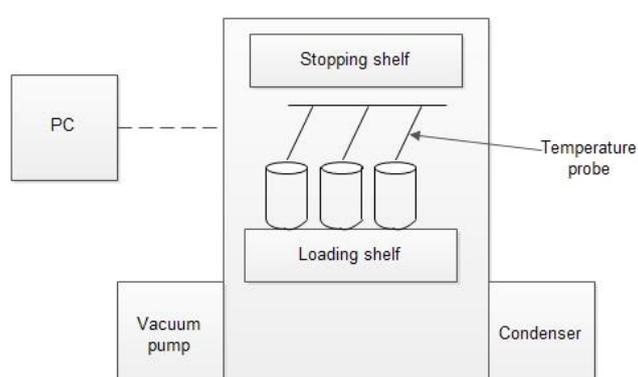


Figure 1.1 A typical freeze dryer

Before conducting a freeze-drying process, APIs together with proper additives are mixed in aqueous buffer solution. The solution is then filled into containers such as vials, which are then put onto the shelf of a freeze-dryer to be freeze-dried.

A typical freeze dryer is illustrated in Figure 1.1. Freeze-drying is conducted within a confined space. The shelf temperature can be controlled with cooling fluid inside the shelf. The heat conduction between the shelf surface and the bottom of the containers ensures that the desired temperature of product is achieved. In order to achieve a good heat transfer, one needs to make sure that the whole bottom of the container contacts well with the metal surface of shelf. This is simple to achieve for some glass vials with flat bottoms. However, for a standard 96-well plate, there is a gap between the bottom of the wells and the shelf surface. One can either trim off the bottom lips (Grant et al. 2009) or place an aluminium plate in between the gap to achieve full contact.

The condenser, which is usually underneath the shelf, is designed to collect water vapour in the chamber. The condenser is the coldest part of a freeze-dryer unit. Water vapour accumulates on the condenser surface and deposits as ice. The condenser ensures that a saturated water vapour is avoided at a certain pressure and temperature, so that water can be continuously sublimed from product.

The vacuum pump reduces the pressure within the chamber and ensures that the ice sublimates into the gaseous state directly, bypassing the liquid state. It also facilitates the movement of water vapour from areas of higher pressure, to areas of lower pressure, thus accelerating the drying.

When sterile conditions are required, stoppers are partially inserted into vials before loading onto the shelves. At the end of the freeze-drying, the stoppers are fully inserted to seal the vials by either raising the chamber pressure (e.g. pumping in nitrogen) (Geigert 2004) or by pushing them down with the shelf above them. After taking the samples out, the condenser is heated to melt the ice, and the liquid water is collected in the tray underneath it.

A modern freeze-dryer can be monitored using a computer with real-time measurement of the temperatures of the product, shelf and condenser, as well as the pressure of the chamber. Step-by-step procedures of freeze-drying can be programmed in advance and adjustments can be made during the process.

### 1.3.2 Three steps in freeze-drying

Freeze-drying consists of three consecutive operation procedures, which are:

- Freezing, in which the protein solution is cooled down and frozen;
- Primary drying, in which most of the ice water is sublimed; and
- Secondary drying, in which the residual water is further removed until the final solid form is constructed.

Although these three steps are relatively distinctive, they are inherently interdependent. A previous operation may exert dramatic influence on a subsequent one. For example, one needs to carefully manipulate the temperature and rate of change of temperature, as these parameters will influence the number and size of ice crystals formed, and thus exert impact on the drying rate (Section 1.4).

#### 1.3.2.1 Freezing

There are three states to describe the morphology of the frozen solids, which are crystalline, polycrystalline and amorphous states as shown in Figure 1.2. 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 an amorphous state, the molecules are stochastically arranged. 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.

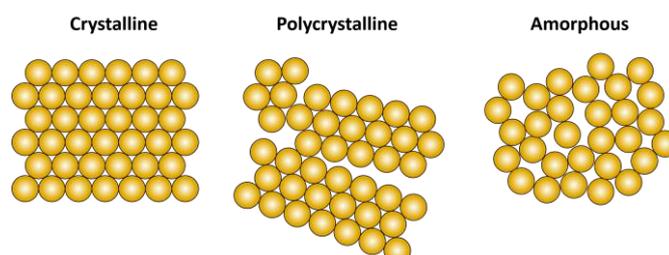


Figure 1.2 The description of crystalline and amorphous states.  
(Redrawn from <http://en.wikipedia.org/wiki/Crystal>)

There are two types of classifications for crystalline states. One is a variety of polymorphs; another is different solvates (e.g. hydrates). For the polymorphs, molecules may arrange into different morphologies at certain temperatures and pressures. For example, needle, feather, and spherulite forms of ice can be generated at various environmental conditions, while the familiar hexagonal stellar dendrites of snowflakes are the most stable form (Libbrecht 2001). In terms of solvates, the solutes can interact with solvent and form complexes. Hydrates are a common form of solvate if water is used as solvent, while some salts or other compounds will form crystals together with water molecules. The importance of different crystalline morphologies lies in that it renders the composition of the freeze-drying solution more complex, as phase transitions between them may occur, which each increasing the uncertainty and unpredictability of the process.

Theoretically, the crystalline state is more thermally stable, whereas the amorphous states are prone to experience a series of relaxation stages, and can also rearrange into crystalline states at specific temperatures and pressures (Craig et al. 1999). However, the kinetics of relaxation processes are often beyond the experimental time-scale, and may even last for years and centuries. For this reason, amorphous states can exist for a long time. Most of the water (typically >99%) is usually frozen into ice crystals formed via non-covalent hydrogen bonds. The solutes, and the rest of the water, can either become crystalline or amorphous, depending on the properties of the solutes, and also the cooling rate.

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). A comprehensive understanding of the properties of the two solid forms, and a robust and well-understood manufacturing procedure is indispensable to ensure products with an acceptable period of validity and a consistent

efficacy. For example, the glass transition temperature  $T_g$  and the eutectic point  $T_e$  are characteristic temperatures for amorphous and crystalline states, respectively. At the end of freezing, the temperature of the drug phase should be well below the  $T_g$  or  $T_e$  (Tang & Pikal 2004).

In a typical formulation for freeze-drying, the solutes are not saturated. As the temperature is decreased, the water typically crystallises first, while the solutes remain in solution. In fact, the solute concentration increases, in a process called freeze-concentration.

The formation of ice is initialised by the generation of nuclei, followed by the growth of the water crystal. As water molecules diffuse in the liquid solution, there are certain domains in which the local density becomes higher. If the size and lifetime of these random high-density domains can be reached to a critical nucleation point, the nuclei will generate and surrounding water molecules would accumulate to form a cluster and start the growth of a crystal. This theory is the basis of the quantitative "homogeneous nucleation model", in which the water is assumed to be without any impurities (Oxtoby 1992; Anisimov 2003). In practice, however, nucleation can also be induced by a substance in the protein solution, or by other external factors. For example, a temperature probe immersed in solution can unavoidably provide the nuclei source.

Once the nuclei of ice have been formed, the growth of ice crystals will immediately take place. There are many forms of ice and most the stable one is hexagonal stellar dendrites, which is obtained when water is slowly cooled to sub-zero Celsius (Kenneth 2005; Franks & Auffret 2007). The occupied volume of water is enlarged due to the density drop and the molar heat capacity is halved after freezing.

Freeze-concentration occurs when water is extracted from solution due to ice formation. With the formation of ice, the solutes become more concentrated at the interface between the ice front and the bulk liquid solution, and would diffuse into the bulk solution. If the diffusion cannot keep up with the ice formation, adverse impacts can be induced such as instability of the solutes and protein denaturation at the interfaces of protein-air and protein-ice (Bhatnagar et al. 2008). If the solute continues to be

concentrated beyond its saturation concentration, the solution would become either super-saturated, or the solute would crystallise together with the ice formation.

Crystallisation of solutes requires an initial nucleation and takes time to reach an equilibrium crystallisation stage at a certain temperature. It is influenced by the cooling rate, the solute type and concentration, and the crystal morphology. Although crystals are the most stable solid state, the actual states formed depend largely on the interactions among proteins, excipients and water, together with the processing conditions. In practice, polymorphic forms (including hydrates) are more common for a multi-component solution, in which the metastable forms can experience solid-solid transitions to more stable states during the subsequent process and even during storage (Franks & Auffret 2007).

A super-saturated solution will form if nuclei do not appear in time or, if the crystallisation equilibrium is not fully achieved before the product is cooled further and equilibrated. A super-saturated solution induced by cooling is also known as an “undercooled” solution. The temperature at which ice, the crystalline solute and a super-saturated solution begin to coexist is known as the eutectic point ( $T_e$ ).

Theoretically, there is no liquid solution below the  $T_e$  if equilibrium has been reached. In practice, an undercooled liquid solution still exists when the cooling rate is too fast. This partially frozen stage can remain until the temperature is further lowered to a point where the super-saturated solution is converted into a solid-like glassy state.

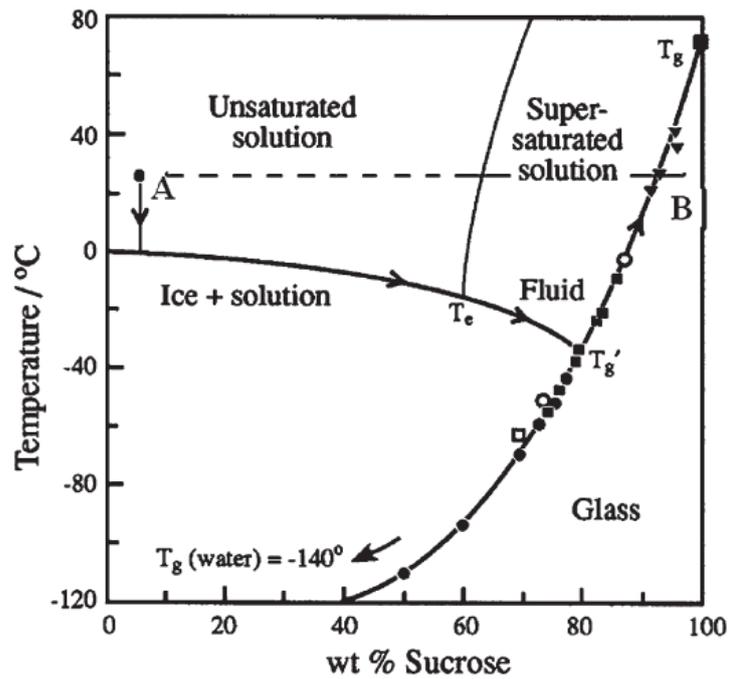


Figure 1.3 A phase diagram for sucrose-water system

The solution is in fluid phase and unsaturated at Point A. As the temperature drops, the solution often experiences a super-saturated state before being frozen into a solid state. As increasing amount of water is frozen, the solute (i.e. sucrose) concentration increases in the rest of fluid water along the equilibrium freezing curve. When the  $T_e$  is reached, the solution either crystallises or keeps increasing the concentration until reaching the  $T_g'$ , depending on the complexity of solution. This diagram illustrates that one needs to operate the freezing under the  $T_g'$  (Heljo 2013; Franks & Auffret 2007).

The term “chilled” is often used to describe a supersaturated protein solution induced by cooling, but which is still not frozen. It is stated that even though an undercooled solution could cause a protein to denature by unfolding that leads to inactivation, this cold denaturation can be completely reversible (Singh & Nema 2010).

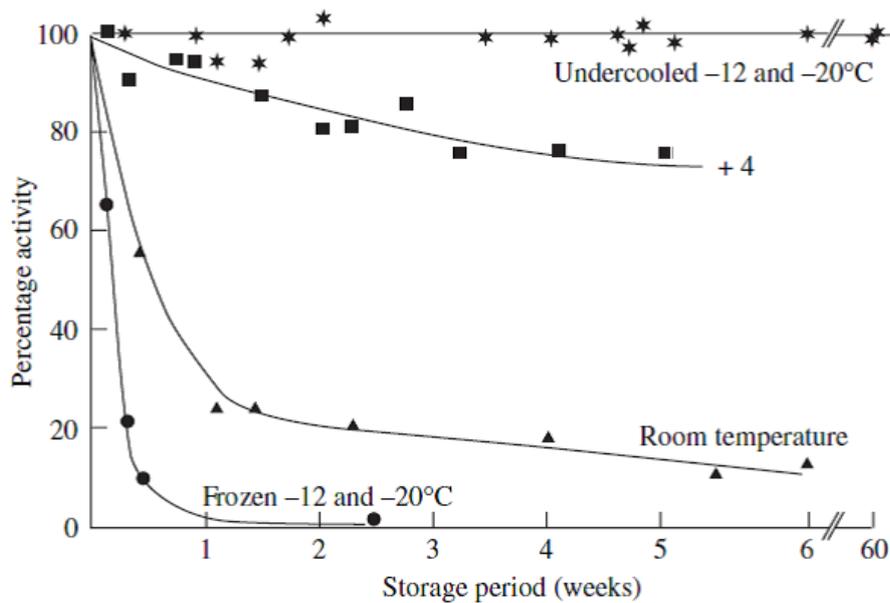


Figure 1.4 Undercooling was found to preserve LDH activity

Hexagrams, squares, triangles and circles represent storage conditions at undercooled at -12 and -20°C, 4°C, room temperature and frozen at -12 and -20°C, respectively. No activity loss was found for the undercooled conditions over 60 weeks while other conditions all reduced the LDH activities in 6 weeks (Franks & Auffret 2007).

It was found previously that undercooled conditions (i.e. -12 & -20°C) could almost fully preserve the activity of LDH for more than one year (Figure 1.4). By contrast, the activity was lost within two weeks when LDH was kept frozen at -12 & -20°C. The reason can be that in the frozen solution, the solute concentration is dramatically increased, which may accelerate competitive degradation reactions such as aggregation at elevated protein concentration. In the undercooled solution, the solute concentration is not changed significantly, and any degradation reaction rates are decreased by the low temperature.

As the product temperature is further reduced from  $T_e$  to a certain point, super-saturated solutions, in which the solutes do not crystallise, are maximally concentrated and converted into amorphous solids. This solid state is regarded as a “glass” and the temperature at which it forms is known as the glass transition temperature ( $T_g$ ). This phase transition is known as “vitrification”, “glassification” or “solidification” (Liu 2006).

$T_g$  is an important parameter used to develop a freeze-drying process. If the amorphous state is the desired form of the product, a necessary condition is that the shelf

temperature should be lower than the  $T_g$  so that all of the substance can be converted to a glass, and the product not damaged by exposure to a phase shift (Tang & Pikal 2004).

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. Inadequate crystallisation of bulking agents may lead to a decreased  $T_g$ , and also vial breakage during primary drying (Milton et al. 2007). Annealing can also increase the ice particle size, which helps to reduce the primary drying time, but both increasing and decreasing reconstitution rate have been reported (Franks & Auffret 2007; Searles et al. 2001).

Some of the features of annealing are listed in Table 1.1 (Tang & Pikal 2004; Franks & Auffret 2007; Liu 2006). Differential Scanning Calorimetry (DSC) is often used to design a proper annealing process, as discussed in Section 1.5.4.1 (Liu 2006).

Table 1.1 Essential points of annealing

Time point	At the end of freezing
Duration	2 h or longer (fill depth 1 cm or more)
Temperature	$T_e > T (10-20^\circ\text{C}) > T_g$
Crystallisation rate	Higher mass ratio of bulking agent > lower mass ratio bulking agent
	Lower temperature results in higher crystallinity but lower crystallisation rate
Crystallinity monitoring	<ul style="list-style-type: none"> <li>● Frozen solution X-ray diffraction</li> <li>● DSC</li> </ul>
Features	<ul style="list-style-type: none"> <li>● Full crystallisation of bulking agent</li> <li>● Increased ice particle size</li> <li>● Reduced specific surface area</li> </ul>
Advantages	<ul style="list-style-type: none"> <li>● Prevents decreasing <math>T_g</math> &amp; also vial breakage during drying</li> <li>● Shortens primary drying</li> <li>● Improves cake appearance</li> </ul>
Disadvantages	<ul style="list-style-type: none"> <li>● Increased residual moisture &amp; prolonged secondary drying</li> <li>● Decreased reconstitution rate</li> <li>● Phase separation, pH shifts, etc.</li> </ul>

Often the protein solution to be freeze-dried is made up of several components and the phase changes are very complex. Both crystalline and amorphous states can

coexist under certain conditions. As water accounts for the majority of protein solution, the formation of ice is generally the first thing that is induced by freezing. The crystallisation of other substances, including hydrates, and the formation of amorphous states, can both take place if stable nucleation exists and  $T_e$  and  $T_g$  have been reached, respectively.

Partial crystallisation can also commonly occur during the freeze-drying process. This situation can happen when the crystallisation of solutes has not been fully completed due to a quick temperature drop to the  $T_g$  point.

### *1.3.2.2 Primary drying*

After the solution has been frozen, the temperature of the condenser is set to a very low point (e.g.  $-60^{\circ}\text{C}$ ) to initialise the primary drying phase. A vacuum is established and more than 90% of the water during the initial freezing process will sublime.

Ice sublimation is an endothermic process. During primary drying, the shelf provides the majority of the heat for the ice to sublime. Increasing the temperature difference between the shelf and product makes heat transfer more efficient. In fact, the shelf temperature can be raised to a certain point to accelerate the heat transfer while maintaining the product structure integrity. However, it is advisable to retain the shelf temperature below the  $T_g$  to avoid the ice melting back into the supersaturated solution.

The collapse temperature ( $T_c$ ) indicates the temperature above which an amorphous product loses its macroscopic structure (Tang & Pikal 2004) due to a decreased viscosity (Liu 2006). It is typically at approximately  $2^{\circ}\text{C}$  above  $T_g$  (Colandene 2007). A collapsed cake results in high residue moisture, which may cause sample degradation, and loss of an acceptable product appearance (Liu 2006).

$T_c$  is equivalent to  $T_e$  when the product is in the crystalline state (Tang & Pikal 2004). 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  $T_e$ . Operating between  $T_g$  and  $T_e$  is sometimes used to shorten the primary drying time in cases where

no significant degradation takes place. Cake collapse phenomena are listed in Table 1.2 (Tang & Pikal 2004; Liu 2006).

Table 1.2 Collapse phenomena at different temperature

State status	Collapse phenomena ("T" is the product temperature)	
Amorphous state ( $T_g < T_c$ )	$T_c < T$	
	Amorphous phase collapse & macroscopic collapse with decreasing viscosity	
A mixture of amorphous and excess crystalline phases	$T_g < (T_c \text{ of microcollapse}) < T < T_e$	$T_e (T_c \text{ of macrocollapse}) < T$
	Amorphous phase collapses onto the surface of crystalline phase (microcollapse without macroscopic collapse)	Macroscopic collapse (melt down)
Crystalline state ( $T_c = T_e$ )	$T_c < T$	
	Macroscopic collapse (melt down)	

During primary drying, sublimation occurs and the water vapour is condensed as ice on the condenser. The concept of water vapour pressure and how it is influenced by temperature is important to understand as the drying force during sublimation.

The "saturation vapour pressure (SVP) of water/ice" is the pressure at which water vapour is saturated at a given temperature. It occurs as a dynamic equilibrium where the rate of evaporation equals that of condensation into water droplets. The SVP increases exponentially with temperature as shown in Figure 1.5.

During the sublimation process, the temperature of the condenser is always maintained below that of the shelf. As water vapour sublimates from the ice front, the vapour pressure of water is unsaturated due to the relatively higher temperature with respect to the equilibrium temperature. The water vapour will be generated continuously from the ice as long as the vapour pressure does not reach the SVP of water. As water vapour accumulates above the ice, its vapour pressure increases and would migrate to zones of lower water-vapour pressure. As it approaches an area of lower temperature, such as at the condenser, the SVP of water decreases. If the vapour pressure of water is higher than its SVP at that temperature, the water vapour would form ice droplets. In fact, the temperature difference between the ice front area and the condenser ensures

that ice continues to sublime to water vapour, travels to the condenser and then deposits into ice. Therefore, it is preferable to set the lowest achievable temperature of the condenser so as to ensure a sufficient mass transfer rate into water vapour. The condenser also plays a role in minimising the movement of water vapour into the vacuum pump.

The drying process must be undertaken in a vacuum with pressure below the solid-gas phase-boundary for sublimation to occur as shown in Figure 1.5. The reason is that ice will only sublime at a low pressure when maintained at a low temperature. Only when the pressure is sufficiently low, the solid water (i.e. ice) can undergo a phase transition directly to the vapour state.

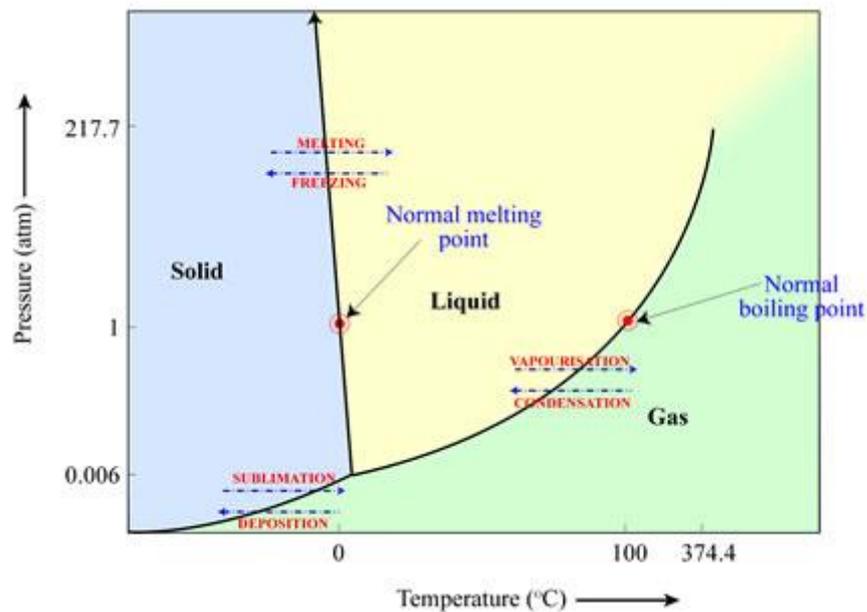


Figure 1.5 The phase diagram for water

Figure was extracted from <http://astronomy.swin.edu.au/cosmos/s/sublimation>. The pressure-temperature relations for water is shown in the figure. The boundary lines of Solid/Gas and Liquid/Gas represent the "Saturation vapour pressure (SVP)". As pressure is reduced, the ice will be converted to vapour if the pressure is lower than the SVP. The water vapour will also be converted to ice if the temperature is reduced for the pressure to be lower than the SVP.

### 1.3.2.3 Secondary drying

Secondary drying aims to remove residual, often more tightly bound water that is trapped within the solid product mass. The residual water can be either in the form of hydrates or simply free from the substance.

During secondary drying, the shelf temperature is further increased so that residual water within the product can diffuse to the surface and evaporate. However, one needs to be sure that the product temperature is not so high as to induce the glass transition and the subsequent collapse of amorphous solid. The temperature that the amorphous state starts to experience glass transition during second drying process is the  $T_g$ , and this phase transition is known as devitrification. In most cases, a higher  $T_g$  is achieved where there is less residual water.

The diffusion of residual water plays a major role during secondary drying. At this point, little water is in the form of ice and so cannot be removed simply by sublimation. Instead, it needs to migrate from within the solid matrix to the outer surface and then evaporate.

Conventionally, the content of residual water is expressed as a mass ratio relative to the product. This is easy to determine by established techniques including Karl Fischer titration (Section 1.5.3.1). However, as the molecular weights of different products vary, it cannot actually represent an absolute measure of the dryness of final solid product. It is obvious that for two products with the same mass ratio of residual water, the lower molecular-weight product would be surrounded by fewer water molecules. In fact, it is usually the molar ratio of product and water that truly affect the subsequent stability/activity of product (Franks & Auffret 2007).

As the drying process approaches the end, one needs to consider the proper dryness of product so that maximum quality can be obtained. Even though the freeze-drying process will remove most of the water and provide a low moisture storage condition for the product, this does not mean simply that lowering the water content leads necessarily to a better product. For example, a study (Chang et al. 2005) showed that the chemical degradation rate of a IgG1 antibody was minimal at 2-3% residual moisture within the examined 1-5% residual moisture conditions; another study (Breen et al. 2001) showed that a humanized monoclonal antibody formulated at 2-3% moisture resulted in less aggregates than formulation at 1% moisture when samples were stored below their  $T_g$  values.

Containers are left open during the freeze-drying process and can be sealed when taking them out from the chamber after the process cycle. If sterile conditions are necessary, stoppers can be partially inserted into the containers when loading and then fully inserted after the cycle by bringing down the shelf above them. This also avoids re-absorption of moisture from the atmosphere.

## 1.4 Interplay of freeze-drying parameters

In this section, the formulation and various freeze-drying process parameters will be discussed and correlated. A major aim is to develop a deeper understanding of operation parameters in order to optimise the freeze-drying process.

### 1.4.1 Formulation and freezing

#### 1.4.1.1 *Product property*

The formulation and freezing steps are used to build the structure of products, which influence both heat transfer and mass transfer. Several aspects of the product attributes are of particular importance and they are described below.

Thermal conductivity refers to the efficiency of heat transfer to, and through, the product. As ice accounts for the majority of the solid, the ice crystal size distribution strongly affects the thermal conductivity.

Table 1.3 Impact of cooling rate on ice crystal dimensions and ice sublimation time

Cooling rate	Nucleation rate	Number of ice crystals	Size of ice crystals	Ice sublimation time
Low	Low	Small	Large	Shorter
High	High	Large	Small	Longer

(Extracted from (Franks & Auffret 2007))

The size distribution can be influenced by the cooling rate during freezing (Franks & Auffret 2007; Tang & Pikal 2004). As shown in Table 1.3, a high cooling rate results in small ice crystals and decreases the sublimation rate due to less ice connectivity. By contrast, a low cooling rate results in large ice crystals with better ice connectivity that,

therefore, reduces the time for ice sublimation. The formation of an amorphous product can reduce the connectivity compared to crystal states. Also, if a glassy product forms around crystals, it would act as an insulating skin that inhibits the heat/mass transfer. In addition to the cooling rate, these aspects can also be influenced by formulation (e.g. solid content).

The specific surface area of the ice front is another significant property that greatly impacts the subsequent drying efficiency. Clearly, larger specific surface areas lead to more efficient sublimation of the ice into gaseous state. Excipient choice, solid porosity, solid content, fill volume and cooling rate have all been found to strongly influence the specific surface area.

- Excipient modification

Excipients can be designed to increase the ice front surface area. It was found that increased ratio of mannitol/sucrose resulted in larger specific surface area and less extent of collapsed cake for a IgG1 formulation (Schersch et al. 2010); another study showed that increased sucrose concentration resulted in an increased ice surface area (Bhatnagar et al. 2008).

- Porosity/solid content

A porous product would present a larger ice surface area, and the product would become more porous as drying proceeds. Usually, a low solid content leads to a more porous product. However, too little of the solid would result in poor mechanical strength and break up of the solid plug.

- Fill volume

Deeper filling of a vial by the frozen product, leads to a lower specific surface area. However, in practice, increasing the surface area by using vials with larger diameters, can lead to the suboptimal utilisation of the freeze-dryer shelf space.

The fill volume can also affect the choice of containers. Usually between one third and one half of the vial volume is filled with solution, in order to achieve good heat diffusion. A fill depth of more than 2 cm should be avoided, while the final freezing

temperature should be held for 1-2 hours before primary drying, to ensure sufficient time for complete freezing (Tang & Pikal 2004).

The fill volume (or more precisely, fill depth) contributes differently to the specific surface area throughout the primary and secondary drying processes. In the beginning of primary drying, there is little void space within the frozen product, and so the available surface area is only made up of the top surface of the frozen solid. As drying proceeds, the solid should become more porous with more available void space, and thus increases the specific surface area.

- Cooling rate

Three different cooling methods have been discussed previously (Tang & Pikal 2004) and these are summarised in Table 1.4 with their influence on supercooling. It is also found that supercooling heterogeneity between vials is caused when samples are placed onto a precooled shelf. A recommended cooling rate of 1°C/min is suggested which provides uniform ice structure and minimises phase separation between proteins and stabilisers.

Table 1.4 Three freezing methods

<b>Freezing methods</b>	<b>Supercooling effect</b>
Liquid nitrogen freezing	High
Loading vials onto precooled shelves	Low
Ramped cooling	Medium

As illustrated before, a high cooling rate would create more ice crystals and thus larger specific surface area. Even though thermal conductivity outweighs specific surface area in ice sublimation during primary drying, the specific surface area will play a more important role in the diffusive process in secondary drying. In addition, as proteins are likely to denature at the ice-water interface, it has been found that the adverse impact of fast cooling is greater or equal to slow cooling of 11 cycles (Hang & Endrick 1996).

### ***1.4.1.2 Excipients***

Excipients are defined in the *Handbook of Pharmaceutical Excipients* (Raymond C Rowe 2009) as "inert or inactive ingredients" which "aid the formulation and manufacture

of the subsequent dosage form for administration to patients". They can be used to protect active drug-substance proteins from denaturing under the processing stresses and storage conditions, and thus preserve and prolong their stability and activity. To be more specific, they can act as buffer salts to maintain the solution within a proper physiological pH; they can physico-chemically alter the  $T_g$  and  $T_g$  of product-excipient complexes, so that the active ingredients can resist more extreme conditions; they can also play a role in providing a robust cake structure or elegant morphology that meet the market demand. Actually, it is the excipients, instead of active ingredients, that usually account for the majority of the composition of the final solid dosage. Therefore, the selection of suitable excipients and their composition within the initial protein solutions are of equal importance as the adjustment of operating parameters (e.g. temperature, pressure) during freeze-drying.

In practice, certain excipients have gained particular popularities, and their specific functions have been investigated. For example, mannitol is a crystalline excipient that is added to support the formation of the physical structure of the product, whereas trehalose and sucrose work as amorphous excipients to stabilise proteins with higher  $T_g$  and reduced motion (Ohtake et al. 2011). PEG, sorbitol and many other excipients are also useful additives that are widely applied in the biopharmaceutical industry to reduce temperature and dehydration stress or/and inhibit protein adsorption (Kamerzell et al. 2011). However, a rational selection or design strategy, of suitable excipients and their composition, is still not well developed, and it is found that some excipients can serve in multi-functional roles to stabilise proteins.

A straightforward way of choosing excipients is by undertaking so-called "Design of Experiment (DoE)". In DoE, a range of potential excipients are formulated with active proteins, in a specific set of compositions. The impacts of different excipients in stabilising proteins are then compared from their corresponding protein activities after freeze-drying. As a result, the combination of excipients that can achieve optimal activity is determined and an equation indicating how individual excipients can contribute to the stabilisation is proposed by statistical calculation and regression. This approach

was used previously (Grant et al. 2009; Grant, Matejtschuk, et al. 2012) to determine the most suitable excipients to preserve model proteins of lactose dehydrogenase (LDH) and granulocyte colony-stimulating factor (G-CSF). In their research, 96-well plates were used to reduce the sample material requirements, and to rapidly screen the possible excipient alternatives and to analyse their interactions in the protection of active proteins.

The DoE method has presented a generic way to choose the suitable excipients. Its application, however, provides only very limited, if any, understanding of the mechanism of excipients. Meanwhile, as one can only apply one operation condition at a time in a single freeze dryer, the screen of suitable processing parameters (e.g. temperature, pressure) is not as straightforward as that of suitable excipients. To achieve this, the recognition of molecular structure of both products and additives, and the understanding of thermodynamics and kinetics are crucially important.

Theoretically, all inert compounds can be used as excipients. Generally, excipients can be categorised as buffer salts, bulking agents and lyoprotectants (Wang 2000; Kamerzell et al. 2011). Their functions are explained as follows.

- Buffer salts: maintain the required pH and salinity (ionic strength) during freezing and upon reconstitution.
- Bulking agents: support the appearance and physical strength of the product, e.g. crystalline and amorphous structures.
- Cryoprotectants and lyoprotectants: alter the physicochemical properties (e.g.  $T_g$ ,  $T_e$ ,  $T_f$ ) of product by interacting with products or forming product-excipient complexes.

It is emphasised that in many cases, one excipient can provide many attributes to the product.

- Excipients as buffer salts

Buffer salts are used to maintain the pH and salinity (ionic strength) of the protein solution, not only during freeze-drying, but also for the reconstitution of dried solid at the point of use. There are two issues that need to be considered when choosing suitable buffer salts:

- (1) Whether the buffer salts are likely to precipitate; and
- (2) What is the form of precipitation?

The first issue involves the potential to induce a pH shift if one of the pair of the buffer salts precipitates prior to the other. For example, in a sodium phosphate buffer system, the mole ratio of  $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$  to obtain a pH 7 buffer is 0.72. The individual eutectic points for  $\text{NaH}_2\text{PO}_4$  and  $\text{Na}_2\text{HPO}_4$  are  $-9.7^\circ\text{C}$  and  $-0.5^\circ\text{C}$ ; while at a ternary eutectic point, in which  $\text{NaH}_2\text{PO}_4$ ,  $\text{Na}_2\text{HPO}_4$  and water co-crystallise, the mole ratio is 57. This means that if the initial pH of the buffer is 7 with a mole ratio of 0.72 for  $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$ , which is far from 57,  $\text{Na}_2\text{HPO}_4$  would crystallise prior to the co-crystallisation of  $\text{NaH}_2\text{PO}_4$  and  $\text{Na}_2\text{HPO}_4$ . As the basic buffer salt being subtracted from the buffer system, an acidic pH-shift would be induced (Murase & Franks 1989). It was found that a decreased initial buffer concentration at 8 mM and slightly lower initial pH at 5.7 could mitigate the pH shift of sodium phosphate buffer compared to 100 mM buffer at pH 7.4 on freezing (Gó Mez et al. 2001), while potassium phosphate had stronger resistance towards pH shift compared to sodium phosphate buffer system (Pikal-Cleland et al. 2000). Other studies show that citrate buffer is better than phosphate buffer for pH 6.5, and glycocholate buffer is better than succinate for pH 3 to 5 (Wang 2000).

The second issue that needs to be considered when selecting a proper buffer is the form of precipitation. Instead of an anhydrous crystal, the buffer salts may form hydrates. Some of the hydrates are of a metastable state and would decompose into water molecules and the anhydrous form. Take NaCl for example (Figure 1.6), it would spontaneously precipitate with two water molecules and form the hydrate of  $\text{NaCl}\cdot 2\text{H}_2\text{O}$  as the temperature is dropped below  $0^\circ\text{C}$ . At its eutectic point of  $-21^\circ\text{C}$ , almost all the  $\text{Na}^+$  and  $\text{Cl}^-$  ions become crystallised. This dihydrate, which is not stable, decomposes into NaCl and water when temperature is raised to nearly  $1^\circ\text{C}$  at its peritectic point (Franks & Auffret 2007). The undesired water cannot be dried during the sublimation of ice and would prolong the secondary drying. If the water exists in the final solid dosage of the product, it will promote the destabilisation and inactivation of the protein.

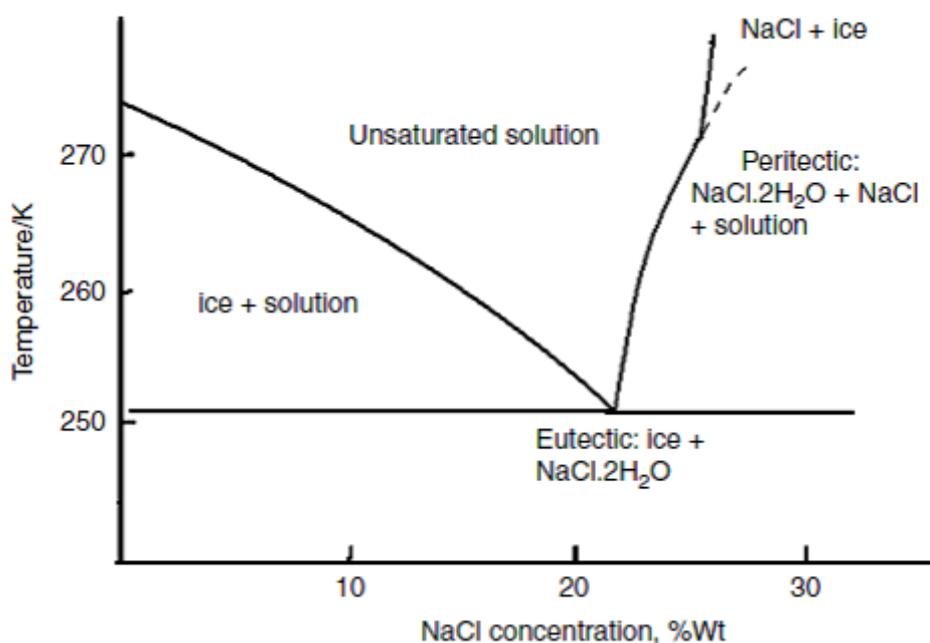


Figure 1.6 Solid/liquid phase diagram for the binary water-NaCl system

The dihydrate NaCl.2H<sub>2</sub>O is crystallised from the solution at the eutectic point. As temperature increases, it decomposes into anhydrous crystalline NaCl and water at the peritectic point (Franks & Auffret 2007).

- Excipients as bulk agents

Bulk agents can be utilised to confer a robust physical structure for the final solid dosage. In other words, they provide the desired mechanical strength for solid states. Crystal and glass are the two typical solid states that formulation scientists would aim for. The choice between amorphous and crystalline formulation depends on the market demands and the property of proteins. Crystalline solids give rise to an attractive appearance but offer little chemical protection for the unstable active ingredients. Amorphous states provide good reconstitution performance and their storage conditions can be adjusted by suitable excipients. For example, insulin exhibited better stability in an amorphous state than a crystalline one (Pikal & Rigsbee 1997).

In order to achieve the crystal form of the product, excipients that can spontaneously crystallise are favoured. Mannitol is one of the common crystalline agents and it can help form a rigid homogeneous cake for the lyophilised plug (Raymond C Rowe 2009; Kaialy et al. 2016).

Unlike the spontaneous crystallization of compounds like mannitol, some organic compounds, especially poly hydroxide compounds (PHCs), do not form crystals at their  $T_e$  even at low cooling rates. As the temperature further decreases, these compounds vitrify into glass directly. Excipients with high  $T_g$  are suggested as they can increase the  $T_g$  of the complex, and thus provide a higher operation temperature for freezing and primary drying. For instance, sucrose is a commonly used amorphous agent and it could hinder the crystallisation of mannitol formulation (Park et al. 2013). Due to these favourable features of PHCs, they can be mixed with products if an amorphous state is the desired form for end-users. As amorphous excipients reduce the ice formed, they can shorten the primary drying time while prolonging that for secondary drying.

In cases where an amorphous state is to be achieved, excipients that can raise the  $T_g$  give rise to a higher operation temperature for secondary drying, and the condition to preserve the solid form during storage. Sucrose, dextran and many other excipients (e.g. polymers) had been found to stabilise the amorphous solid state of proteins (Ohtake et al. 2011). However, excipients that offer amorphous structures may devitrify into crystalline state. Therefore, storage temperature should be well below their  $T_g$  (Bianco et al. 2013). For example, lactose crystallises at 360 K to form monohydrates from an amorphous solid, leaving the rest of the glass with less water content. This transition renders the protein further dried and out of the protection of lactose. Interestingly, however, by removing water from the amorphous state, the  $T_g$  of the remaining product will increase therefore leading to improved storage stability. This is deemed as “self-stabilisation” (Franks & Auffret 2007).

- Excipients as stabilisers

Stabilisers are the excipients that can help active protein resist the undesired chilling/freezing/drying conditions during freeze-drying. Theoretically, bulk agents that can assist to provide a stable morphology (e.g. glass) can be also regarded as stabilisers. The stabilisers can be further classified into cryoprotectant for freezing protection, and lyoprotectant for drying protection (Wang 2000).

Various excipients corresponding to different denaturation processes during freeze-drying have been summarised in the literature (Wang 2000) and are listed here in Table 1.5.

Table 1.5 Choices of excipients to protect APIs from various denaturation process during lyophilisation

Processing stresses		Type of excipients		
Freezing stress	Concentration effect with accelerated chemical reactions	<ul style="list-style-type: none"> <li>● Polymers to increase viscosity</li> </ul>	<ul style="list-style-type: none"> <li>● Polyhydric alcohols (PEG, DMSO, DMF)</li> <li>● Salts (potassium phosphate)</li> <li>● Amines</li> </ul>	<ul style="list-style-type: none"> <li>● Sugars/polyols (sucrose, trehalose)</li> <li>● Polymers (serum albumin, dextran)</li> <li>● Amino acids</li> <li>● Metal ions</li> <li>● Amphiphilic excipients</li> </ul>
	Ice-water interface with high surface tension	<ul style="list-style-type: none"> <li>● Surfactants (tween 80)</li> </ul>		
	pH changes	<ul style="list-style-type: none"> <li>● Amino acids (glycine)</li> <li>● Polymers (BSA)</li> </ul>		
Drying stress (Dehydration)		<ul style="list-style-type: none"> <li>● Excipients to form glass (e.g. Sucrose)</li> <li>● Excipients of high <math>T_g</math> (trehalose)</li> </ul>		

(Extracted from (Wang 2000). Note: dehydration is the removal of hydration shell around proteins, which would induce aggregation/inactivation of proteins.)

## 1.4.2 Primary drying

The step of primary drying accounts for the majority of sublimation of ice water.

This process is a combination of heat transfer and mass transfer. Specifically, the heat provided by the freeze dryer (e.g. shelf) gives rise to the driving force for the sublimation of ice from product. The relation between heat transfer and mass transfer can be mathematically described (Franks & Auffret 2007) as

$$K_v(T_s - T_p) = \Delta H_s(dm/dt) \quad \text{Equation 1.1}$$

In Equation 1.1, the definitions of each parameter are listed in Table 1.6.

Table 1.6 The parameter definitions of Equation 1.1 for heat and mass transfer

Parameters	Definitions
$K_v$	Heat transfer coefficient
$T_s$	Shelf temperature
$T_p$	Ice front temperature
$\Delta H_s$	Latent heat of ice sublimation required at $T_p$
$dm/dt$	Rate of mass transfer of water vapour from ice front to the condenser

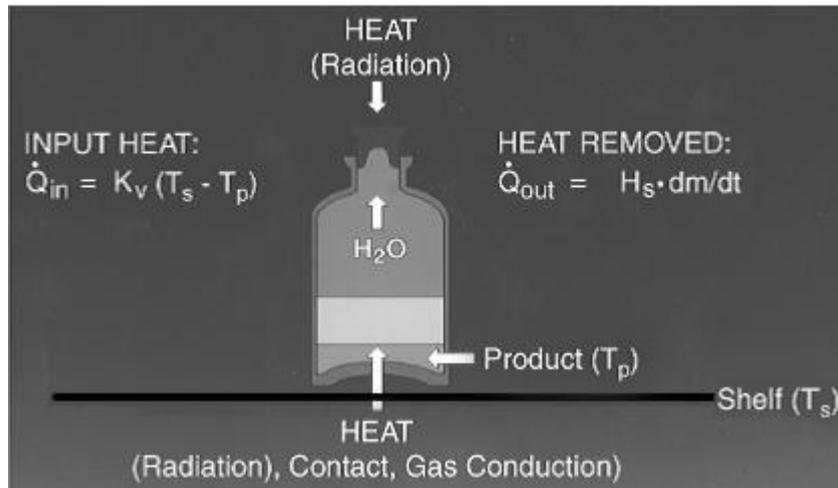


Figure 1.7 Diagram of coupled heat transfer and mass transfer

The shelf supplies the energy ( $Q_{in}$ ) for the water to sublime with energy ( $Q_{out}$ ) at a mass transfer rate  $dm/dt$ . At product temperature  $T_p$ ,  $H_s$  indicates the sublimation latent heat (Franks & Auffret 2007).

Figure 1.7 describes how the heat input is transferred to the removal of ice. It should be noted that Equation 1.1 represents an instantaneous situation, and so some values would change according to different parameters. For example, the latent heat of ice sublimation varies with ice temperature. Moreover, the heat transfer coefficient depends largely on the structure of the product. In the following section, a variety of factors that influence heat and mass transfers will be discussed, and their contributions weighted.

#### 1.4.2.1 Heat transfer

Heat transfer is made up of three mechanisms, i.e. radiation, conduction and convection.

##### (1) Radiation

In most cases, the exterior of the freeze dryer is warmer than the interior and heat can radiate from the door and walls of the freeze dryer.

##### (2) Conduction

Conduction of heat is caused by direct contact between two materials of different temperatures. For example, conduction takes place between shelf and vial, vial and product, and between products of various temperatures.

### (3) Convection

Heat convection is mainly a result of gas collision and flow. As gas molecules randomly travel through the space (Brownian diffusion), they interact with each other and exchange heat from higher temperature to lower temperature.

Compared to conduction and convection of heat, radiation accounts for only a minor part of the entire heat transfer.

The thermal conduction is mainly affected by two factors, i.e. temperature difference and conductivity. They can be illustrated as follows.

Table 1.7 Conduction of heat

Conduction of heat	
Temperature difference	Thermal conductivity
Shelf/vial	Contact area with shelf
Vial/product	Thermal conductivity of container
Temperature gradient within product	Thermal conductivity of product

As the container is heated by the shelf, there would be a temperature gradient within the product from the highest at the bottom to the lowest at the ice front, which ensures continuous heat conduction. The thermal conductivity of the product thus critically affects the heat transfer efficiency. In general, ice crystals have better thermal conductivity compared to amorphous solids. In addition, some aspects described in Table 1.7 undergo continuous changes as well. For instance, the thermal conductivity of the product will decrease as the product becomes more porous.

As convection of heat is caused by the motion and collision of gas molecules, the pressure, which indicates the density of gas molecules, is used to reflect the extent of convection. The gas not only transfers heat from the shelf to the ice front, but also transfers heat generated by condensation of water vapour. Figure 1.8 shows that the sublimation rate is proportional to the chamber pressure but will reach a plateau when the pressure goes above a certain point. It should be noted that the pressure within the chamber does not exceed the SVP of ice at the operating temperature (Franks & Auffret 2007).

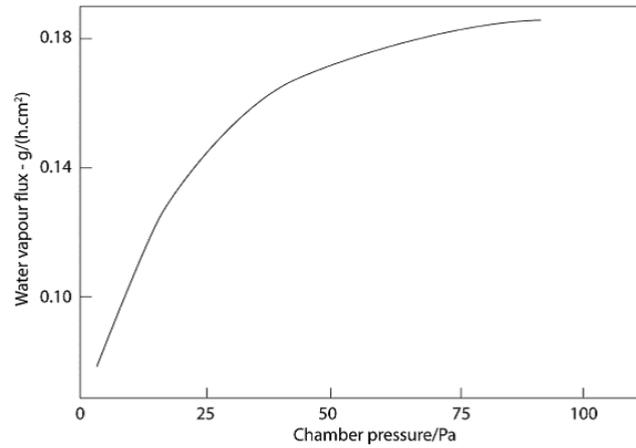


Figure 1.8 Effect of chamber pressure on sublimation rate

The water sublimation rate increases with the chamber pressure but will reach a plateau when the pressure goes above a certain point (Franks & Auffret 2007)

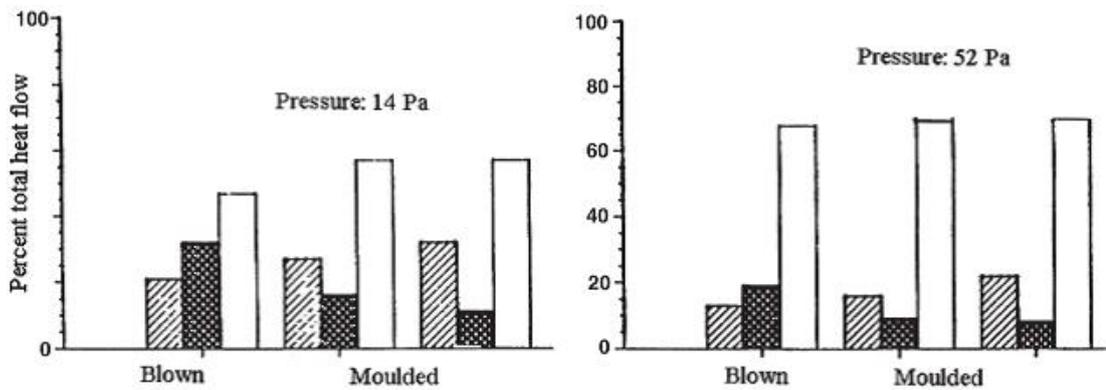


Figure 1.9 Contributions of radiation (left, diagonals), conduction (centre, cross-hatched) and convection (right, blank) to the total heat transfer for three vial types and two pressures (Franks & Auffret 2007).

The weights of the three heat transfer mechanisms have been compared in Figure 1.9 (Franks & Auffret 2007). It is shown that convection is a major influence among the three mechanisms, especially at higher pressure. The contributions between radiation and conduction vary from the containers, i.e. conduction account for more if a container of better contact with shelf is used.

### 1.4.2.2 Mass transfer

Mass transfer refers to the water migrating away from the product. Clearly, sublimation of ice is a crucial part and the surface area of the ice front where sublimation takes place is a key factor. In addition, water can also migrate inside the product before

sublimation, and therefore the morphology of the product affects the rate of mass transfer to a certain extent as well. It is summarised that fill depth, total solid content, and cake porosity would influence the product morphology, and therefore exerts an impact on the specific surface area. In the meantime, after ice becomes water vapour, it would migrate to the condenser and condense onto its surface. The rate for ice migration depends on the pressure difference between the area near the ice front and the condenser.

In addition to the pressure difference providing the driving force for the migration of water vapour to the condenser, the pressure within the chamber itself also plays a role in accelerating the mass transfer. Obviously, a higher pressure would lead to more possibilities for gas molecules to move around more intensively and reach the condenser.

#### *1.4.2.3 Monitoring the end-point of primary drying*

It is of critical importance to determine the end-point of primary drying so as to minimise the drying period and lessen product denaturation.

- Temperature

As the drying process approaches the end, the temperature of the product approaches the shelf temperature. However, it might not be accurate to measure the temperature of a dry product without any interference with the product (Corbellini et al. 2010). In addition, the vials with thermocouples dry faster than vials without thermocouples (Patel et al. 2010), so it is suggested that one needs to extend the primary drying time to at least 10 - 20% more after the end-point determined by temperature meters (Tang & Pikal 2004). A series of product temperature measurements are listed in Section 1.4.4.1.

- Gas/pressure

Dew point sensors can detect the relative humidity of gas to indicate the change in gas composition. It is said that there is a significant dew point decrease when it comes to the end of primary drying as gas composition changes from ~100% water vapour to mostly air or nitrogen (Tang & Pikal 2004).

A "pressure rise" test is a common indicator of the sublimation endpoint. In the test, the chamber is isolated from any gas flow and its pressure is measured. The completion of ice removal is determined by an absence of a pressure rise within one second (Chouvenc et al. 2004). However, the limitation of this test lies in that the isolation of gas flow is not always practically feasible. For example, there is sometimes no particular valve between chamber and condenser to cut off the gas flow.

According to Table 1.11, the completion of sublimation can also be indicated when the thermal conductivity of the gas (i.e. Pirani gauge) decreases and levels off to total pressure measurement (i.e. capacitance pressure probe) (Franks & Auffret 2007; Colandene 2007). Anemometers can measure the gas flow but they cannot differentiate different gas types (Franks & Auffret 2007). When mass spectrometers are incorporated into the detection of gas, more sensitive monitoring is provided not only for the gas elements but also for the product profiles, which offers great insight into protein denaturation (Barresi et al. 2009).

#### *1.4.2.4 Practical recommendations for operation parameters*

- Target product temperature

The product temperature in primary drying is 5 to 40°C below the shelf temperature (Tang & Pikal 2004), due to the heat removal by sublimation. It would be ideal to hold the product at an optimum target product temperature throughout the primary drying. Theoretically, the product temperature should be lower than  $T_c$  so as to avoid collapse. Practically, however, a slight increase in product temperature would decrease the primary time (e.g. 13% time is reduced by 1°C elevated temperature) (Tang & Pikal 2004). The operational safety margin temperature between product temperature and  $T_c$  is suggested, as below, in terms of primary drying duration (Tang & Pikal 2004). Basically, the longer the primary drying, the smaller the safety margin used, so as to minimise the cycle period.

Table 1.8 The operation margin temperature between product temperature and  $T_c$

Safety margin temperature (lower than $T_c$ )	Primary drying duration
2°C	t > 2 days
3°C	10 h < t < 2 days
5°C	t < 10 h

Interestingly, it is also found that the product temperature can be operated above  $T_g$  with negligible denaturation. This is because the viscosity of the formulation is high enough at low temperature (e.g. -15°C) to inhibit protein unfolding (Tang & Pikal 2004).

Addition of crystalline bulking agents with high eutectic melting temperatures can also aid avoiding collapse when operating above  $T_g$  (Colandene 2007).

- Chamber pressure

In primary drying, the partial pressure of water is the same as the chamber pressure except at the endpoint of primary drying. Low partial pressure of water facilitates high sublimation rate. But very low chamber pressures would produce larger heterogeneity in heat transfer (Tang & Pikal 2004). It is recommended that 50-200 mTorr (typically 100-150 mTorr) provides optimal homogeneity (Tang & Pikal 2004).

#### 1.4.2.5 Conclusion

In the primary drying process, a perfect situation lies in that the heat provided equals the heat needed for mass transfer. If heat is over-provided, the product temperature will increase and may exceed  $T_g$ . If heat provided is insufficient, the drying process will be prolonged and may induce protein denaturation.

Basically, raising of the shelf temperature is a common way to accelerate the primary drying process, but the temperature of the ice front should not be higher than the glass transition temperature  $T_g$ . If the product temperature is likely to surpass its  $T_g$ , increasing the pressure can be used to enhance the mass transfer, and thus cool down the product temperature.

## 1.4.3 Secondary drying

### 1.4.3.1 Two existence forms of water

After the primary drying, some residual water still remains in the product. This water can be classified into two existence forms, as listed in Table 1.9. The majority of it is trapped within the glass as free mobile water while the rest of it is within the crystalline system, either as a thin adsorbed layer on the crystal surface, or as an integral part of crystalline hydrates with stoichiometric ratios to the crystalline solutes (Yu et al. 1999; Wahl et al. 2015).

Table 1.9 The forms of residual water after primary drying

<b>In amorphous system (majority)</b>	<b>In crystalline system</b>	
Trapped in amorphous solid & free to diffusive	Thin adsorbed layer	Crystalline hydrates

### 1.4.3.2 Diffusion as the rate-limiting step in secondary drying

Water removal during secondary drying involves two steps, i.e. diffusion and desorption. The diffusion refers to the diffusive process of water from the bulk product to the surface. It is then followed by desorption, in which water evaporates rapidly and migrates to the condenser.

Though it is not clear if water diffusion is the rate-limiting step compared to water evaporation, many models assumed diffusion of water as the rate-limiting step for water desorption during secondary drying (Kodama et al. 2014). Diffusion is significantly influenced by temperature and specific surface area while pressure and water content does not significantly influence the secondary drying rate (Pikal et al. 1990).

- Specific surface area

The specific surface area is a key aspect that affects the drying rate in secondary drying. As illustrated in Section 1.4.1, it can be affected by formulation, porosity, fill depth of the product and the cooling rate during freezing. Basically, high porosity, low fill

depth and high cooling rate would induce a large specific surface area. However, these aspects contribute differently for the drying rate in secondary drying. It has been found that cake thickness has little effect on the drying kinetics (Franks & Auffret 2007). This may be due to that in the secondary drying there is already a high degree of void space within the cake, so the fill depth will not affect the specific surface area significantly. On the other hand, a high cooling rate during freezing generates a large specific surface area and accelerates secondary drying, but the primary-drying rate would be compromised (Franks & Auffret 2007).

- Temperature

Adjusting the shelf temperature is one of the most common ways for process control in secondary drying. The aim is to accelerate the drying rate by raising the product temperature but maintaining it below  $T_g$  so as to minimise cake shrinkage (Rambhatla et al. 2005).

As water is gradually removed from the product, the drying rate will decrease to a plateau if a constant shelf temperature is set. In the meantime, the  $T_g$  of the product is supposed to increase with less water content. Therefore, an ideal operation would gradually increase the shelf temperature in order to efficiently reduce water content to a desired composition (Pisano et al. 2012). Specifically, it is recommended to increase the temperature at the rate of 0.1 or 0.15°C/min for amorphous formulations, and 0.3 or 0.4°C/min for crystalline products (Tang & Pikal 2004).

Softening point ( $T_s$ ) is the temperature at which the glass softens into a liquid. In some cases where the residual water can only be removed by raising the product temperature above  $T_g$ , it is suggested that the temperature should remain between  $T_g$  and  $T_s$  (Figure 1.10) (Franks & Auffret 2007).

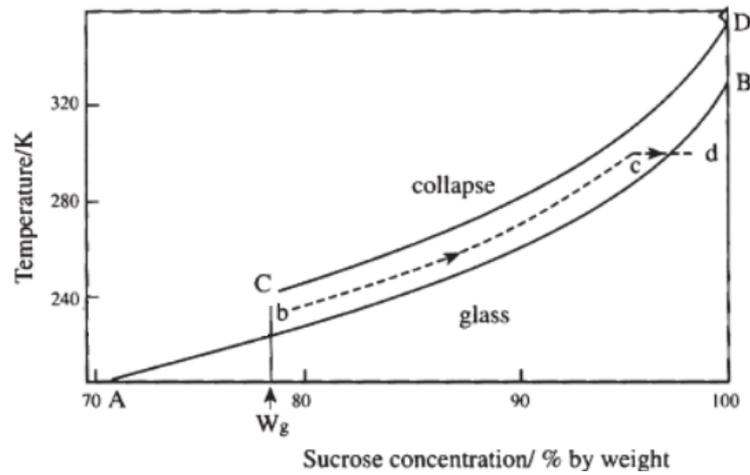


Figure 1.10 An ideal operation of adjusting temperature between  $T_s$  and  $T_g$ . Line AB and CD are glass transition and softening temperature profiles derived from DSC.  $W_g$  is the sucrose concentration of maximally freeze-concentrated solution. Line bcd is the ideal secondary drying pathway. (Franks & Auffret 2007)

A more practical way of adjusting the product temperature is shown in Figure 1.11, as it is not straightforward to control the temperature exactly as shown in Figure 1.10.

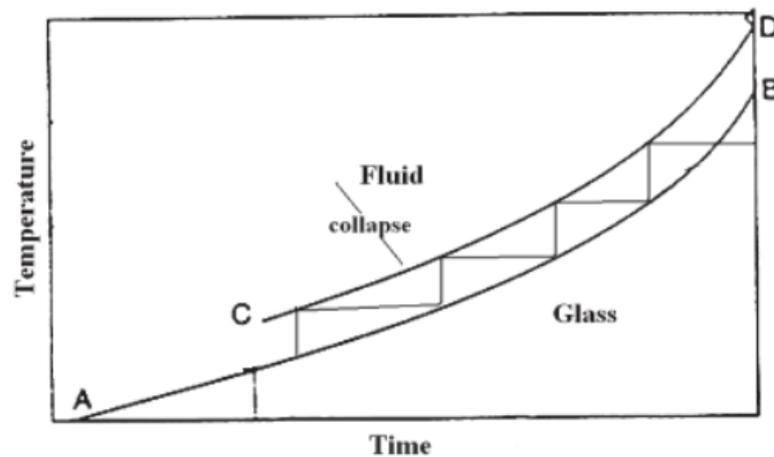


Figure 1.11 A more practical way of adjusting the product temperature between  $T_g$  and  $T_s$ . The step line in between line AB and CD is a more practical secondary drying pathway.

### 1.4.3.3 Monitoring the end-point of secondary drying

When the desired moisture of products is achieved, the secondary drying should be terminated otherwise products would experience unnecessary denaturing processes at high temperatures. It is a delightful fact (Tang & Pikal 2004) that the freeze-dried samples can be extracted by the use of a “sample thief” without interrupting the freeze-

drying cycle, and the moisture can be determined by TGA, near IR, Karl Fisher titration, or a modified MTM method.

#### 1.4.3.4 Conclusion

In conclusion, during secondary drying, the diffusion of water relies largely on the specific area whereas water content has little effect to the diffusive process. Temperature control is a more efficient way to control the product quality compared to pressure, and the product temperature is advised to be kept below  $T_g$  so as to avoid collapse from the amorphous state to a crystalline or solution state. A gradual increase of shelf temperature is proposed so as to reduce the drying period while retaining the product integrity.

### 1.4.4 Real-time monitoring for product properties

#### 1.4.4.1 Temperature monitoring

There are different ways to monitor the temperature of products. Each of them involves strengths and weakness. The following table is a summary for the various methods discussed in the literature (Franks & Auffret 2007; Tang & Pikal 2004).

Table 1.10 Temperature monitoring methods

General aspects			
<b>Limitation</b>	Close contact with product	Probe liberate heat	
		Increase the ice nucleation probability	
"Thin wire thermocouples" is preferred to "resistance thermometers (RTD sensors)"			
Various measurements			
	Application	Mechanism	Limitation
<b>Thermocouple</b>	The temperature at the bottom of vials	The junction of two different metals is heated/cooled and a voltage	Cause less supercooling & larger ice crystal

		is produced to correlate temperature	
<b>RTD sensors</b>	The temperature at the bottom of vials	a resistor (e.g. platinum) that changes resistance value as its temperature changes	
<b>Barometric temperature measurement (Manometric temperature measurement, MTM)</b>	<ul style="list-style-type: none"> <li>● The temperature of ice sublimation interface (0.5 to 2°C lower than vial bottom)</li> <li>● The completion of ice sublimation</li> </ul>	A pressure rise test (PRT) with a water vapour–temperature diagram	An interruption of the cycle & lacks the continuous output
<b>Conductimetry</b>	<ul style="list-style-type: none"> <li>● Conductance</li> <li>● Eutectic temperature</li> <li>● Warning of abnormal conditions</li> </ul>	Conductimeter	Eutectic temperatures are unrelated to the physical state of the drug substance

#### 1.4.4.2 Chamber pressure monitoring

Various chamber pressure measurements are summarised as below (Franks & Auffret 2007; Tang & Pikal 2004).

Table 1.11 Chamber pressure monitoring methods

	<i>Feature</i>	<i>Mechanism</i>
<b>Heat conductivity vacuum gauge (Pirani pressure gauge)</b>	Calibrated against air <ul style="list-style-type: none"> <li>● correction for pure water vapour</li> <li>● <math>\frac{\text{water vapour}}{\text{air or nitrogen}} = 1.5</math></li> </ul>	Measure the electrical energy needed to maintain a constant temperature
<b>Membrane differential gauge (capacity gauge or MKS Baratron gauge)</b>	Independent of the gas type	Capacitance changes caused by a deflection of the membrane sealed against a fixed low pressure

## 1.5 Analytical methods to characterise protein properties

### 1.5.1 Size/Conformation/Structure

#### 1.5.1.1 SEC-HPLC

SEC-HPLC is one of the most widely used analytical methods to characterise protein monomer/aggregate profiles in aqueous phase. Based on the differences in molecular weight, pure protein monomers would exhibit a single peak in the chromatogram while aggregates (e.g. dimers, trimers, etc.) would present peaks before the monomer peak. A typical SEC-HPLC process can be completed within 15 min. The disadvantage of SEC-HPLC is that it cannot distinguish aggregates of different molecular weight, and becomes less sensitive for large aggregates. SDS-PAGE is needed if the exact molecular weights are required.

#### 1.5.1.2 DLS

Dynamic Light Scattering (DLS) can be a complementary method to SEC-HPLC that can also indicate the molecular weight of proteins, especially for large molecules and particles. However, it provides more of a qualitative indication of the molecular size profiles. Plots of intensity versus particle size obtained by DLS, often only show 1-3 peaks, and these are heavily biased towards the larger particle sizes present, because intensity is proportional to the sixth-power of the particle radius. The PSD of individual species therefore cannot be directly compared with the profiles obtained from other analyses (e.g. SEC-HPLC) that are more quantitative for the number of particles.

#### 1.5.1.3 RP-HPLC

RP-HPLC is used to differentiate the hydrophobicity of different protein species. During a freeze-drying process, intact protein monomers can degrade into fragments or interact with each other to form aggregates. Proteins of different states would present

different hydrophobic patches and interact differently with the carbon chains immobilised on the resin.

Under good control of ACN (Acetonitrile) composition and its gradient, RP-HPLC can be used to differentiate between aggregates formed by covalent and non-covalent bonding (Wang 2005), which is a great strength compared to other aggregation characterisation techniques based on size (e.g. SEC-HPLC, DLS, SDS-PAGE). Theoretically, covalently bonded aggregates are more hydrophobic compared to monomers, especially when they are unfolded, which makes them reside longer on the resin. As for non-covalently bonded aggregates, they would dissociate when interacting with ACN and display more than one peak. Other types of denatured status can also be characterised, which is summarised in Table 1.12. In order to obtain more detailed molecular profiles (e.g. intrinsic heterogeneity such as glycosylation) of protein, RP-HPLC can be coupled with mass spectroscopy (Dillon et al. 2006; Carr 2002).

Table 1.12 The characterisation of denatured status by RP-HPLC

Denatured status		Differences from intact monomers	Reference
Aggregation	Covalent binding	Longer residence time	(Wang 2005)
	Non-covalent binding	More than one peak	
Deamidation		Neutral pH, shorter residence time	(Carr 2002)
Oxidation		Shorter residence time	
Structure heterogeneity (LC/MS)	Differences in disulfide bond	Multiple peaks for whole antibody but merely two peaks (heavy & light chains) after reduction	(Dillon et al. 2006)
	Fragments by hydrolysis cleavage	Shorter residence time	
Partial unfolding		Broad peaks	

It is known that the mobile phase used in RP-HPLC (typically acetonitrile, ACN) would also unfold proteins to a certain extent upon binding or elution (Lau et al. 1984). The hydrophobic peptides inside the protein would then become exposed to the outside and proteins would become more hydrophobic, leading to a longer residence time on an RP-HPLC column. The unfolded proteins can potentially further re-fold or form

aggregates with other proteins. Due to this uncertainty, one needs to first ensure that the composition of ACN that can elute proteins does not render the proteins to re-fold and form aggregates.

#### 1.5.1.4 FTIR

Fourier transform infrared spectroscopy (FTIR or IR) is a commonly used in-line or off-line method to characterise protein secondary structure. The principle is that the amide I, II, III and A vibrational modes of proteins absorb IR distinctly and exhibit unique band shapes (Hayashi & Mukamel 2007). Both liquid and solid states of proteins can be analysed (Colandene 2007). Various denaturation processes, together with corresponding changes in IR spectra, have been discussed in the literature (Wang 2000) and these are summarised in Table 1.13.

Table 1.13 Infrared spectroscopy characterisation for protein denaturation during lyophilisation

<b>Common denaturation</b>	<b>Reflection in IR spectra</b>
Disruption of hydrogen bonds	An increase in frequency and a decrease in intensity of hydroxyl stretching bands
Unfolding of proteins	Broadening and shifting (to higher wave numbers) of amide I component peaks
Aggregation and/or increased intermolecular interaction	Conversion of $\alpha$ -helix to $\beta$ -sheet

#### 1.5.1.5 CD

Circular Dichroism (CD) is a valuable analytical method to evaluate the protein structure preservation. During a measurement, left and right circularly polarised lights (LCP light and RCP light) are passed through a protein sample, and the difference in absorption detected over a range of wavelengths. For instance, secondary structures such as  $\alpha$ -helix,  $\beta$ -sheet and random coil present distinct circular dichroism in the far-UV range (Figure 1.12) while near-UV CD can be an indicator of tertiary structure, which includes dipole orientation and interactions of side chain aromatic amino acids (Kelly et al. 2005). As environmental factors exert significant impact for the protein structure, CD

can be used to evaluate the formulation efficacy for the stabilising of proteins in terms of pH, salinity, temperature and excipient types.

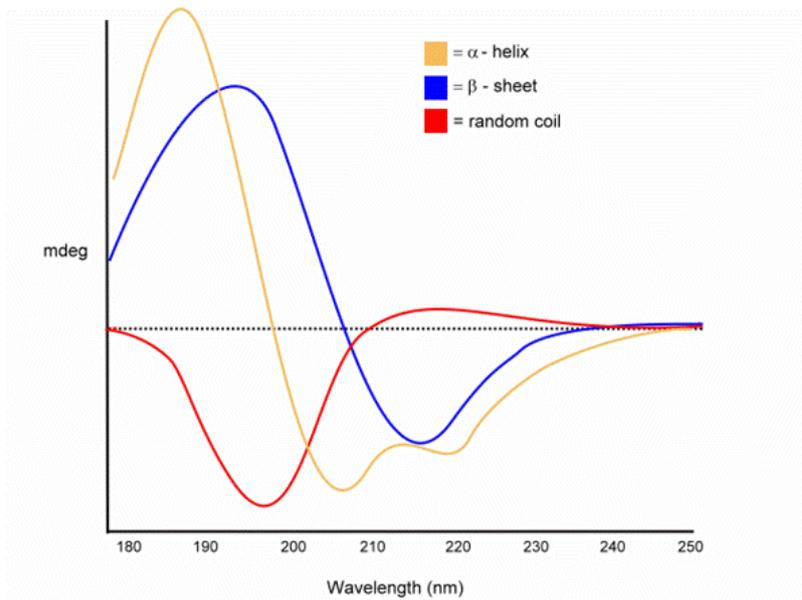


Figure 1.12 The CD features of pure secondary structure elements at far-UV region (Extracted from <http://www.proteinchemist.com/cd/cdspec.html>)

## 1.5.2 Surface area

Surface area is an important feature of the freeze-dried product. The BET (Brunauer, Emmett, and Teller) surface area testing method measures the specific surface of finely divided and porous solids. It is essentially based on the BET theory of multilayer adsorption assumptions, and measurement is carried out by obtaining equilibrium pressure and the amount of gas adsorbed onto a sample surface (Brunauer et al. 1938). Multipoint BET testing is a useful method to determine the surface area and porosity of samples (Colandene 2007). Attention should be paid to cakes with collapsed features, as the BET data in these samples do not fully reflect the previous ice trace (Liu 2006).

## 1.5.3 Water content and sorption

### 1.5.3.1 Residual moisture

Because water is often involved in protein degradation (Section 1.2), its presence in the final product can be deleterious in preserving the potency and stability of product as well as its appearance (Krasucka et al. 2012). For example, water serves as a plasticizer to decrease the  $T_g$  of amorphous solid (Towns 1995). Therefore, it is important to characterize the final product for its water content.

A variety of analyses can be used for the determination of water content, such as loss on drying, thermogravimetry, near-infrared spectroscopy, gas chromatography and Karl Fischer Titration, which have been discussed in literature (Krasucka et al. 2012; Towns 1995; Mary 1967; Zhou et al. 1998) and summarised in Table 1.14. Among these, the most widely used is Karl Fischer Titration. Karl Fischer Titration (KF Titration) is a standard and versatile method to measure the moisture of samples and is commonly used in the biopharmaceutical industry (Crescent 2004). Other moisture measurement alternatives can be used if the analytes are not compatible with the KF titrant. A major downside of most moisture analyses, is that they measure only the total amount of water, and do not provide information on the distribution of the water within the sample (Franks & Auffret 2007). As water interacts differently with amorphous and crystalline solids, more advanced moisture analysis is required if the degradation mechanism by moisture needs to be investigated. Near-infrared (NIR) imaging provides an efficient analysis for the water content across the shelf so the water distribution could be visualised (Trnka, Palou, et al. 2014).

Table 1.14 The analysis of moisture content for freeze-dried solids

Measurement classification		Name	Mechanism	Type of moisture	Feature
Destructive technique	Thermal method	Gravimetric method (loss on drying, LOD)	Measurements before and after the removal of water by heating (20 - 30°C) under vacuum	Surface moisture & loosely bound water of hydration	Under estimation of total moisture General criteria: not over 1.0%

	<b>Thermogravimetric analysis (TGA)</b>	Record loss of mass by heating with linearly increasing temperature to over 1200°C		Sensitive to 2 mg Need to differentiate between moisture & volatile Volatiles verified by MS
<b>Non-thermal method</b>	<b>Gas chromatography using a thermal conductivity detector</b>	Calculate based on water density and calibration (e.g. Peak height/area) of internal standard solution (e.g. N-propanol)	Generally total water content	Overestimation of moisture Micro scale High throughput
	<b>Karl Fischer titration (KF titration)</b>	Titants stoichiometrically react with water		10 mg sample Complex sample handling to prevent ambient water
<b>Non-destructive technique</b>	<b>Near-infrared spectroscopy</b>	Fiber-optic diffuse-reflectance probe measures reflectance through intact glass vials		Rapid (20 s/per analysis) In-situ measurement Little sample preparation Prevent atmospheric moisture Need reference moisture (e.g. KF titration)

### 1.5.3.2 Water desorption

Dynamic Vapour Sorption (DVS) is used to measure the sorption/desorption isotherm of product samples. Samples are exposed and equilibrated at individual levels of gas moisture, in which isotherms are plotted by recording sample mass against the corresponding temperature.

DVS aims to evaluate how formulation, processing, and storage conditions would be affected by water interaction. Sorption degree and kinetic data can be obtained for a range of humidities (e.g. 5-80%) and correlated to the product stability, and water-related reactions such as crystallisation, hydrate formation. A Symmetrical Gravimetric Analyzer (SGA) is a similar water sorption instrument, which can be operated at high vacuum conditions (Liu 2006).

## 1.5.4 Thermal analysis

Thermal analysis plays an important role in determining the stability of the drug in the final dosage form. An explicit classification and discussions for a variety of thermal analyses have been carried out previously (Kett et al. 2004; Liu 2006).

### 1.5.4.1 Calorimetric analyses

- DSC

Differential scanning calorimetry (DSC) is a thermo analytical method that measures the heat absorbed by a sample as the temperature is steadily changed. A reference sample or empty metal crucible (“pan”) with a well-defined heat capacity, and a sample (in an identical pan) are placed within the device. The temperatures of both reference and sample are raised at a controlled rate, and the instrument then measures the difference in heat-input (current) required to retain an identical temperature change between reference and sample, as measured by thermocouples.

DSC is typically used to determine the temperature (or temperature range) of a phase transition by detecting its heat capacity change. Compared to the sample maintained at a fixed phase, sample adsorbs or liberates heat during an endothermic or exothermic phase transition process, respectively. Therefore, the corresponding temperature that would induce a phase transition can be characterised by a peak or trough in the heat input to the sample (Coleman & Craig 1996). Modulated temperature DSC (MTDSC) differs in that a sinusoidal wave modulation is applied to the linear temperature programme used in conventional DSC. This technique was developed to resolve reversible and irreversible processes (Coleman & Craig 1996). The characterisation of glass transition temperatures in frozen state (i.e.  $T_g$ ) and lyophilised state (i.e.  $T_g$ ) is crucial for the understanding and design of a particular formulation and freeze-drying cycle. The interpretation of a DSC thermogram for typical phase transitions in freeze-drying is briefly summarised in Table 1.15 from the literature (Liu 2006; Craig & Reading 2007).

Table 1.15 Interpretation of (Modulated) DSC sensorgram for typical phase transitions in freeze-drying

		Features in thermogram	Examples of events
Thermal events	Endothermic	Peaks	Glass transition, melting (fusion), dehydration
	Exothermic	Reverse peaks	(Re)crystallisation
	Enthalpy change	Step change	Glass transition
Thermodynamic reversibility of processes	Reversible processes	Equilibrium with its surroundings at each stage during that process	Glass transition
	Irreversible processes	Kinetically controlled processes which are dependent on absolute temperature	Crystallisation, enthalpy recovery
Applications	$T_g$ , $T_e$ , crystallisation, melting, $T_g$ , annealing, heat capacity change		

- DTA

Differential Thermal Analysis (DTA) has similar features to that of DSC. It was invented earlier than DSC but is not as popular. The major difference is, in DTA, the temperature of the reference sample and the unknown sample is recorded while applying the same controlled heat flow to both reference and sample. The phase transition is indicated by the difference in temperature change observed between the reference sample and the unknown sample.

- IMC

Isothermal microcalorimetry (IMC) detects heat flow under isothermal and humidity controlled conditions, which can reflect the real-time molecular mobility and reactions within samples. Enthalpy relaxation and recovery processes have been quantified by IMC (Kawakami & Ida 2003) and the relaxation times enables the prediction of stability of freeze dried solid at particular storage condition (Liu 2006). Further development can be made on the characterisation of phase transitions in the frozen state as IMC offers higher sensitivity as well as analysis at conditions more similar to those of freeze-drying, compared to DSC (Liu 2006).

#### *1.5.4.2 Dielectric and electrical analysis*

- TEA

Thermoelectric analysis (TEA) (or Electrical Resistance Analysis, ERA, ER, Electrical Thermal Analysis, ETA, Freezing Resistance Analysis, FRA) measures the electric resistance while the temperature is ramped. The principle is that the sample exhibits high electric resistance (low electric conductance) at low temperature (e.g. frozen state) compared to that at high temperature (e.g. liquid state). Although works have shown that TEA can measure  $T_g$ ,  $T_e$  and ice melting temperature, this analysis is not widely employed due to inconsistency of outputs from different instruments due to their non-standard feature of the frequency applied (Liu 2006).

- DEA

Dielectric Analysis (DEA) measures the dielectric properties of samples. The permittivity of sample is plotted against temperature, which can be used to reflect certain thermal properties (e.g.  $T_c$ ) of formulations for freeze-drying. Applications of DEA have been found in the characterisation of collapse temperature, the difference between frequency-independent first-order transitions and frequency-dependent higher-order transitions for frozen solutions, and in correlating product degradation with dielectric relaxation kinetics and activation energy (Liu 2006).

- TSC

Thermally Stimulated Current (TSC) Spectrometry is another dielectric thermal analysis, which correlates molecular mobility with temperature. TSC has been recommended as a sensitive tool (even better than DSC) to detect the properties of freeze-dried solid, such as  $T_g$ , relaxation in glass state, fragility, crystallisation and melting temperature while its application in the frozen state has been rarely reported (Liu 2006).

#### *1.5.4.3 Mechanical analysis*

Thermal Mechanical Analysis (TMA) measures the macro dimensional change under a constant stress with regard to temperature for both frozen and freeze-dried solid states,

which is specifically employed to prevent vial breakage during freezing (Liu 2006).

Dynamic Mechanical Thermal Analysis (DMTA, DMA), in which periodic stress under alterable frequency is applied, can provide more sensitive detection of state transitions (e.g.  $T_g$ ) than DSC (Kararli et al. 1990).

## **1.5.5 (Thermo) Imaging analysis**

Naked-eye observation would provide a direct insight into the appearance of freeze-drying formulation. In the meantime, observation with instrumentation offers greater insight for the microstructure or other physical properties (e.g. mobility) of frozen and freeze-dried samples. Moreover, a cooling system and vacuum pump can be attached to mimic the conditions applied in freeze-drying.

### *1.5.5.1 Naked-eye observation*

Determination of the frozen state, freeze-drying extent and cake morphology based on naked-eye observation is a simple way to analyse the freeze-drying state. Although naked-eye observation largely depends on individual subjectivity, it is straightforward to operate without the limitations of advanced instruments. It can be at least used to screen the formulation for aesthetic appearance during process development, and is regarded as one of the criteria for product market-acceptance. Freezing and sublimation rates, as well as the observation of cake-collapse events, can be qualitatively indicated before more advanced instruments are employed for quantitative characterisation.

### *1.5.5.2 FDM*

Freeze-Drying Microscopy (FDM) is a valuable tool to simulate freeze-drying conditions while visualising the sample with a microscopic view. Equipped with an adjustable cooling and a vacuum system, freezing and freeze-drying processes can be mimicked and observed on the microscope. In addition, the crystallinity of the sample

can be detected if a polarised light is used such that crystalline-state solids would be brightly or colourfully reflected.

FDM has been successfully applied to determine the collapse temperature (Meister & Gieseler 2008). Other applications, such as the ice morphology, and water diffusion coefficient, have also been undertaken but their transferability to a real vial needs to be evaluated (Liu 2006).

#### ***1.5.5.3 SEM & CESEM***

Scanning Electronic Microscopy (SEM) utilises a focused electron beam, which enables higher resolution for the microstructure morphology (10-20 nm). Therefore, this resolution allows the characterisation of macromolecules (Millqvist-Fureby et al. 1999) and traces left by the sublimation of ice (Doillon et al. 1986).

Similarly to FDM, a cooling system can also be affiliated to SEM to achieve Cryo-environmental Scanning Electronic Microscopy (CESEM). Changes in microstructures can be monitored as a function of temperature change, although it mainly provides an observation of sample surface texture (Liu 2006).

#### ***1.5.5.4 XRPD & FDXRPD***

X-ray Powder Diffractometry (XRPD, XRD) is a standard method to qualitatively indicate the crystallinity of samples. When the X-ray chamber is temperature and humidity controlled, the impact of storage conditions on the physical structure can be studied (Liu 2006). A recent study has shown that XRPD can differentiate bulk amorphous phase separation with only one  $T_g$  detected by DSC (Newman et al. 2008).

Similar to FDM and CESEM, freeze-drying accessories (i.e. a vacuum pump and cooling system) have been mounted to the XRPD instrument to enable the *in situ* monitoring for the crystallisation during a freeze-drying cycle (Cavatur & Suryanarayanan 1998), which is called FDXRPD. However, it is not practical to investigate the amorphous phase transitions and needs to be utilised together with DSC (Liu 2006).

### 1.5.5.5 NMR

Solid-state Nuclear Magnetic Resonance (NMR) can measure the molecular mobility of drug and excipient molecules. By correlating the molecular mobility with physical structure, many bio-physical/chemical properties (e.g. aggregation, activity, stability) can be studied to assess the formulation and storage conditions (Liu 2006). NMR characterise fast molecular dynamics of solid-state formulation with timescales much shorter than structural relaxation (Yoshioka et al. 2011). It has been found that this fast dynamics correlates with formulation instability better than structural relaxation does (Cicerone & Soles 2004).

## 1.5.6 Storage conditions

Essentially, the stability of active protein ingredients only has a practical meaning when particular storage conditions (e.g. pH, temperature) and period are also defined. The storage period, under certain storage conditions, that can preserve half of the active drug molecules, is called the half shelf-life. A variety of structure changes would take place during storage, which gradually leads to a certain degree of irreversible destabilisation and deactivation. For instance, crystallisation of amorphous solids during storage is a common phenomenon. Its advantages and disadvantages have been discussed in the literature (Liu 2006) and summarised in Table 1.16.

Table 1.16 The impact of crystallisation of amorphous solids during storage

	<b>Crystallization of amorphous drug molecules</b>	<b>Crystallization of stabilizers</b>	
<i>Relative possibility</i>	Less likely	More common	
<i>Consequence</i>	Increase the storage stability, reduced solubility	Crystallise to an anhydrate	Crystallise to a hydrate
		Lose its function as a stabilizer, confer water to the drug-containing amorphous phase	Remove plasticising water from amorphous phase

Therefore, a robust freeze-drying formulation and cycle does not only depend on the sample preparation and freeze-drying operation but also largely relies on subsequent preservation during storage, which inevitably should target the end-use point of the drug. Due to the scope of this project of freeze-drying process development, the impact of storage conditions to the product quality would be less of a focus. The reconstitution of lyophilised products was conducted on the same day upon completion of freeze-drying, which minimised the instability occurred in the solid state.

### 1.5.7 Reconstitution

Freeze-dried products need to be dissolved into a liquid solution at the point of use. This is known as reconstitution. It is important to ensure that a solid cake can be fully dissolved within a certain period, and that end-users do not encounter problems caused by un-dissolved powders.

Protein stability and efficacy may be impacted by several factors during reconstitution. This is summarised in Table 1.17 from the literature (Wang 2000; Liu 2006). In order to prevent any undesirable situations, surfactants, stabilisers, buffers are recommended to be used instead of pure water for reconstitution (Wang 2000). This is a feasible strategy but it renders the drug administration more complex and requires patients to be fully educated.

Table 1.17 Undesirable consequences for proteins during reconstitution process

<b>Factors</b>	<b>Mechanisms</b>	<b>Consequences</b>
High residual moisture of product	N/A	Prolonged reconstitution time
Too rapid rehydration	Not refold to native form	Denaturation, aggregation, reduced activity
Loss of formulation elements during lyophilisation	pH shift	
Undesirable reconstitution temperature for APIs with temperature sensitivity		

## **1.5.8 Robustness verification studies for process development**

A robust process should be able to withstand various environmental changes. In the following, a range of factors are discussed that can be used to define the desired criteria of an optimised freeze-drying cycle.

### ***1.5.8.1 Reproducibility, feasibility & transferability***

A good process should firstly be reproducible with only slight and acceptable deviations in the product quality occurring with the expected degree of process parameter variability. Secondly, it should be feasible to operate, which means it should not be labour-intensive and time-consuming. More specifically, operational procedures and expense should be minimised. The experience of practitioners and resources required (e.g. computer source) should be sufficient to operate the required tasks throughout the freeze-drying process and sample analysis. Thirdly, transferability is also important. For example, an optimised process should also be well operated in another instrument or plant.

### ***1.5.8.2 Tolerance for operating failures***

A good process cycle should withstand slight deviations in the operating parameters and maintain the majority of product stability/activity. Even though it is crucial to undertake experiments under Standard Operation Protocols (SOP), and to regularly maintain the devices (e.g. calibration of the meters), the aging of instrumentation is somewhat unpredictable and would lead to a small deviation of parameters. Therefore, testing the cycle performance below and above the desired parameters is a good way to assess the tolerance of withstanding operating failures. For example, a more conservative condition (lower temperature for drying) and a more aggressive condition (higher temperature for drying) has been utilised to test the cycle robustness of freeze-drying for an antibody (Colandene 2007).

### 1.5.8.3 Scale-up study

An optimised cycle would lose its market value if it cannot be applied to the manufacture at an industrial scale. For instance, 96-well plates with 200 µl fill volume can be used in the initial screening step for the selection of formulation alternatives, followed by a verification test in 22 mm external diameter vials with 1 ml fill volume (Grant et al. 2009). However, scale-up to larger volumes of bottles and freeze dryers may also be required in some cases.

Scale-up studies also involve the use of mathematical equations to describe mass/heat transfer (e.g. Equation 1.1) and thus define the factors that change the process upon scale-up. The essence is to ensure that at least one of the parameters (e.g. sublimation rate) remains constant when scaling up.

## 1.6 Summary of protein freeze-drying research

Based on the previous discussion, it is concluded that freeze-drying is a complex bioprocess operation to fabricate therapeutic proteins as stable solids for reconstitution. It involves not only the selection of excipients type and concentration but also the optimisation of processing parameters. Thus far, many theories have been described to understand processes leading to undesirable reactions. Advanced analytical methods have also been developed to maximise the characterisation of product properties (Ohtake et al. 2011; Kamerzell et al. 2011).

Significant considerations for process control of freeze-drying at each step has been briefly summarised in Table 1.18. It is noted that the influence of some controlling methods are not limited in certain steps. Instead, they may impact some properties at a subsequent step. For example, the cooling is conducted during the freezing step but the cooling rate will impact the morphology of the ice and thus affect the sublimation rate and the morphology of the final dried product.

Table 1.18 A brief summary for the process control and analytical methods during freeze-drying

<b>Process steps</b>	<b>Essential properties/events</b>	<b>Process control &amp; analytical methods</b>
----------------------	------------------------------------	---

<i>Formulation</i>	General product properties		Excipients, container, solid content, fill volume, etc.
<i>Characterisation of liquid/frozen product</i>	$T_g$ , $T_e$ , $T_c$ , annealing, etc.		DSC, DTA, FDM, TEA, DEA, TMA & DMTA
<i>Freezing</i>	Thermal conductivity, specific surface area of ice front, porosity		Cooling rate
<i>Primary drying (ice sublimation)</i>	Heat transfer	Radiation	(not significant)
		Conduction	Temperature difference
		Convection	Shelf temperature, pressure
	Mass transfer		Condenser temperature, pressure
<i>Secondary drying (water diffusion)</i>	Drying rate	Diffusion (rate-limiting)	Temperature
		Desorption (rapid)	(Not significant)
<i>Characterisation of freeze dried product</i>	Size/Conformation/Structure		HPLC, DLS, FTIR, CD, etc.
	Surface area		BET
	Water content and sorption	Residual moisture	LOD, TGA, gas chromatography, KF titration, near-infrared spectroscopy
		Water desorption	DVS, SGA
	Thermal property ( $T_g$ , crystallisation propensity, etc.)	Calorimetric analysis	DSC, DTA, IMC
		Dielectric and electrical analysis	TEA, DEA, TSC
		Mechanical analysis	TMA & DMTA
	(Thermo) Imaging analysis (collapse, mobility, surface texture, crystallinity)		Naked-eye, SEM, XRPD, NMR
Storage, reconstitution, process robustness		N/A	

## 1.7 Guidelines for systematic development of freeze-drying processes for proteins

Thus far, there is not a generic way of conducting a freeze-drying process for a new product. Most process development is based on empirical knowledge and a range of key factors are selected to optimise the process in a limited range. As can be seen from Table 1.18, there are a number of parameters to be considered and many of them interact with each other. Moreover, numerous physical and chemical reactions (e.g. aggregation, unfolding) take place at different stages of freeze-drying, which make the deactivation and destabilisation of API unpredictable and difficult to interpret, and it may

take decades to fully understand their correlations. To date, a general methodology of undertaking the development of freeze-drying has not been systematically reported though extensive work has been done for specific aspects (Hang & Endrick 1996; Pyne et al. 2003; Taylor et al. 2010b) or particular protein products (Chang & Fischer 1995; Abdelwahed et al. 2006).

A loop approach is usually undertaken to formulate, freeze-dry and ensure preservation of freeze-dried product is achieved. This will be accomplished in parallel with the characterisation of liquid/frozen/freeze-dried state and cycle design as proposed (Liu 2006). This approach is a typical approach for the development of a freeze-drying process, and it features the data input from analytical methods to directly optimise the formulation (excipients, concentration, etc.) and process cycle (temperature, time, pressure, etc.). Obviously, this loop approach largely relies on experience and the protocol developed for a particular product may not be applicable to another active ingredient.

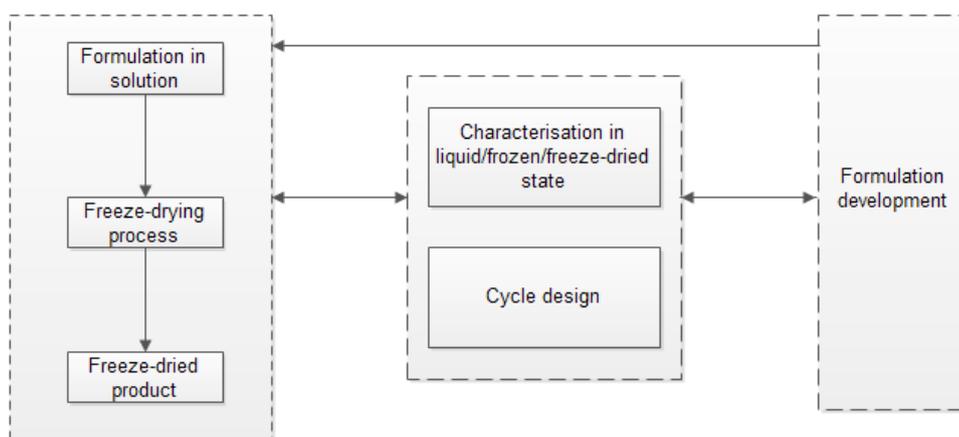


Figure 1.13 A typical loop approach to optimise the freeze-drying process (Adjust and redraw from (Liu 2006))

A more rational and systematic approach has been proposed in Figure 1.14. In this approach, the initial development focuses on the formulation and characterization of the thermal properties. The cycle development is then conducted after the optimised formulation has been determined, which is accompanied by a cycle robustness test and process analytical technology (PAT) tools. In the end, the scale-up study and PAT

implementation will be carried out before the optimised freeze-drying process can be licensed.

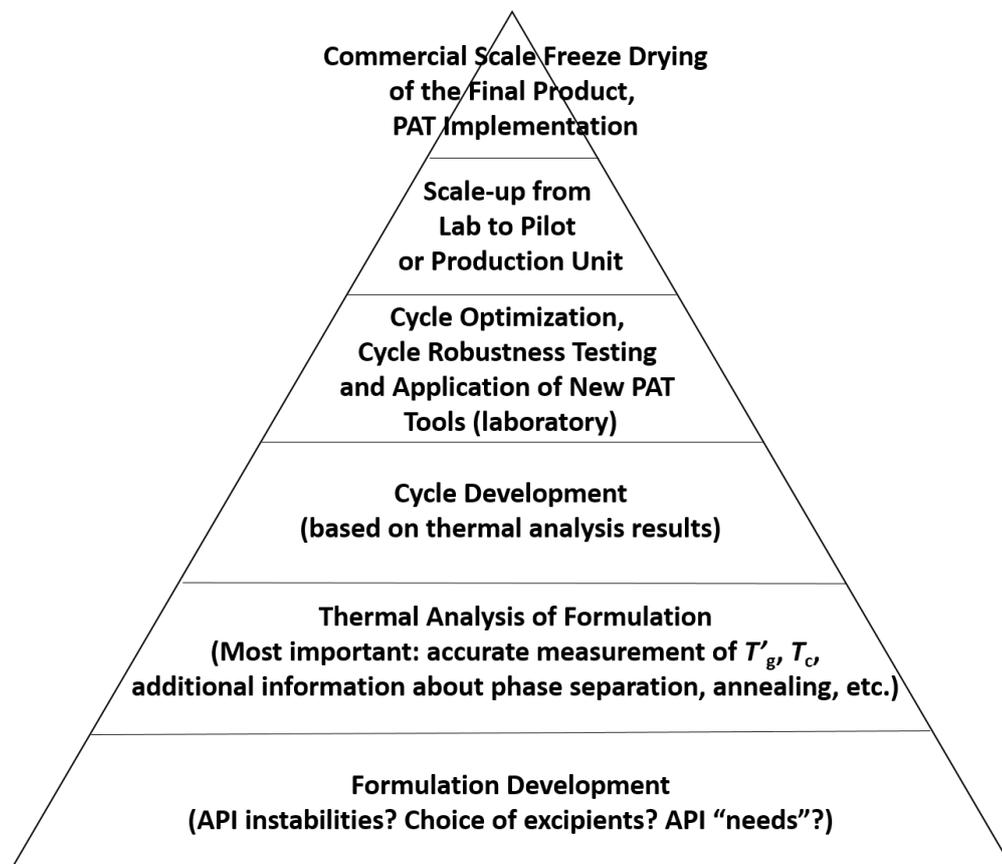


Figure 1.14 An example of a rational way to systematically undertake freeze-drying process development

(Redrawn from <http://freeze-drying.eu/html/research.html>)

Based on the above schematic diagram and knowledge about freeze-drying, the features of the methodology of a rational freeze-drying process development can be summarised as follows.

- From formulation development to cycle development

Due to the feature of the freeze-drying unit operation, a range of formulations can be tested in 96-well plates, vials, ampoules within a single batch while the process parameters (e.g. shelf temperature, time, etc.) can only be tested sequentially. Thus, it would be wise to screen the formulation alternatives in the initial stage with a set of “generically good” process cycle parameters.

- From high-throughput scale-down test to low-throughput scale-up validation

This thumb of rule is also generally applied to other development for the optimisation of individual unit operations. Obviously, it is an effective way to test the potential operating conditions as many as possible and minimise the cost and risk of failures in large scale manufacture.

- From single, more general objective to more specific, multi-objective optimisation

The formulation of pharmaceuticals involves the transition from thinking in an engineer's perspective to a pharmacist's one. From an engineer's point of view, the target to optimise a process focuses more on the protein itself, which includes aggregation, yield of product, etc. From a pharmacist's view, the market value of the product is more concerned, which includes type of formulation, shelf-life, route of administration, etc. Certainly, these two points of views are correlated and they both cover the stability and activity of the drug molecules.

Nevertheless, it is important to notice that it is impractical to satisfy all the criteria in the very beginning. Thus, a few more generic targets (e.g. aggregation) should always come first.

To sum up, a rational and systematic methodology for the development of freeze-drying of proteins is required. It should be straightforward to carry out with detailed guidelines and should be applicable for most proteins.

## 1.8 The influence of Gibbs free energy on protein stability in freeze-drying and liquid formulation

Formulating protein in a desired form to prevent environmental stress is a long standing topic. Numerous studies have been undertaken to examine the influence of various conditions that would affect protein stability. It has been found (Wang et al. 2007) that antibodies would denature or aggregate under physical stresses (temperature change, shear, etc.) and chemical stresses (disulfide formation, isomerization, oxidation, etc.). Liquid (Wang 1999) and lyophilised formulations (Wang 2000) are common methods to preserve the proteins and adequate considerations should be made on the

choice of concentration, pH, excipients and process equipment/containers. However, due to the diversity of proteins and research groups, the formulation recipes are not standardised and are mostly species-dependent for various proteins, which greatly undermines the cross-comparison and usually one formulation technique could not directly be applied to another one. As a result, it takes great effort to screen potential formulation conditions for different proteins.

Therefore, a more fundamental understanding is required of protein instability so as to assist more quantitative, standardised and efficient optimisation of formulation conditions. Several theories and mathematical terms have been developed to describe the protein-protein interaction and stability. For example, the osmotic second virial coefficient,  $B_{22}$ , describes how strong the proteins interact with each other in solution (Gabrielsen et al. 2010). The interaction depends not only on the proteins themselves but also on their surrounding environment including pH, temperature and ionic strength (Guo et al. 1999; Neal et al. 1999; Haas et al. 1999).

Gibbs free energy could be used as a term to describe the stability of proteins. If a process is advantageous, the change in Gibbs free energy is negative and the product is more stable ( $\Delta G < 0$ ); if a process is disadvantageous, the change in Gibbs free energy is positive and the product is less stable ( $\Delta G > 0$ ). This theory has been applied to describe protein folding and unfolding processes. The spontaneous folding or unfolding is regarded as the change of Gibbs free energy towards a more stable state. Due to the lack of consensus,  $\Delta G$  has been used to indicate either folding or unfolding free energy in different research groups (Kaufmann et al. 2010)(Kumar et al. 2006). If  $\Delta G$  indicates the folding energy, the protein with lower  $\Delta G$  at certain condition is more stable. In the subsequent discussion,  $\Delta G$  will be used to indicate the folding energy.

In order to lower the protein  $\Delta G$  so as to preserve the protein in a native folded state, one can either change the solution condition (e.g. pH, ionic strength, temperature) or modify the protein structure like mutation or chemical modification (e.g. PEGylation). Using mutation to stabilise the protein has been widely used in pharmaceutical industry. If a mutation is more stable than the wild type, the change in  $\Delta G$  (i.e.  $\Delta\Delta G$ ) is negative.

It has been found that mutations on residues would alter the charge and hydrophobicity so as to impact the non-covalent interactions including electrostatic interaction, hydrophobic interaction and salt bridge (Lehermayr & Mahler 2011; Chiti, Taddei, et al. 2002; Bosshard et al. 2004). It may also alter the secondary structure as more  $\beta$ -sheet and less  $\alpha$ -helix structure would raise the aggregation propensity (Chiti & Dobson 2006; Pawar et al. 2005). As a result, the change in protein conformation would influence its stability and result in unfolding, misfolding and irreversible aggregation (Fu et al. 2010; Bloom et al. 2006; Chiti & Dobson 2006). In order to generalise and quantify the conformation change impact on the protein stability,  $\Delta\Delta G$  has been used to predict the structural alteration influence under certain forcefield by calculating the various contributions from different interactions (Benedix et al. 2009; Schymkowitz et al. 2005; Das & Baker 2008). In the later chapters, the mutation impact on protein aggregation and stability will be reviewed (Chapter 3); the use of various *in silico* prediction methods for  $\Delta\Delta G$  will be discussed and their performance will be evaluated (Chapter 4); the *in vitro* measurement technique for  $\Delta G$  and  $T_m$  will be assessed (Chapter 5).

## 1.9 Aims and objectives

### 1.9.1 The gap in current freeze-drying research

Based on the discussion above, we can see that it is not straightforward to select the best composition of formulation, optimised process development and rapid and reliable measurement. Many literatures focus on formulation or operating parameters to develop freeze-drying process (Wang 2000; Kasper et al. 2013). Thermal properties (e.g.  $T_g$ ,  $T_c$ ), moisture content and morphology (e.g. surface area) have been adequately characterised and great efforts have been made to ensure the quality of product using excipients and adjusted processing parameters (Chang et al. 2005; Meister & Gieseler 2008).

However, limited reports are found on development of the lyophilized products by modifying the active pharmaceutical ingredient (API) and improve overall Gibbs free

energy. On the other hand, reports have been found that it is practical and useful to mutate one or a few residues for proteins in order to raise the stability (Shoichet & Baase 1995), the production efficiency in fermentation (Kabir & Shimizu 2003) and stability in aqueous solution (Teilum et al. 2011). Therefore, it would be a promising strategy if the lyophilised protein products can be stabilised by mutagenesis so as to ease the subsequent effort in formulation and process parameters. As protein stability directly relates to the efficacy of the drug of interest, the activity and aggregation of proteins should always be considered in the first instance to filter out suboptimal candidates.

With respect to the regulatory chain for clinical trials, the application of mutagenesis needs to be performed in the beginning of clinical and process development. Once the mutational work has been conducted, the intrinsic amino acid sequence will need to be maintained unchanged throughout the clinical trials.

## **1.9.2 Aims and objectives of the project**

This research project aims to investigate how to apply protein engineering principles to enhance the quality of freeze-dried biological macromolecules such as antibodies. A33 Fab and a range of mutants derived from it will be used as model proteins to study freeze-drying.

Figure 1.15 illustrates the step-by-step strategies used to develop the freeze-drying of engineered proteins.

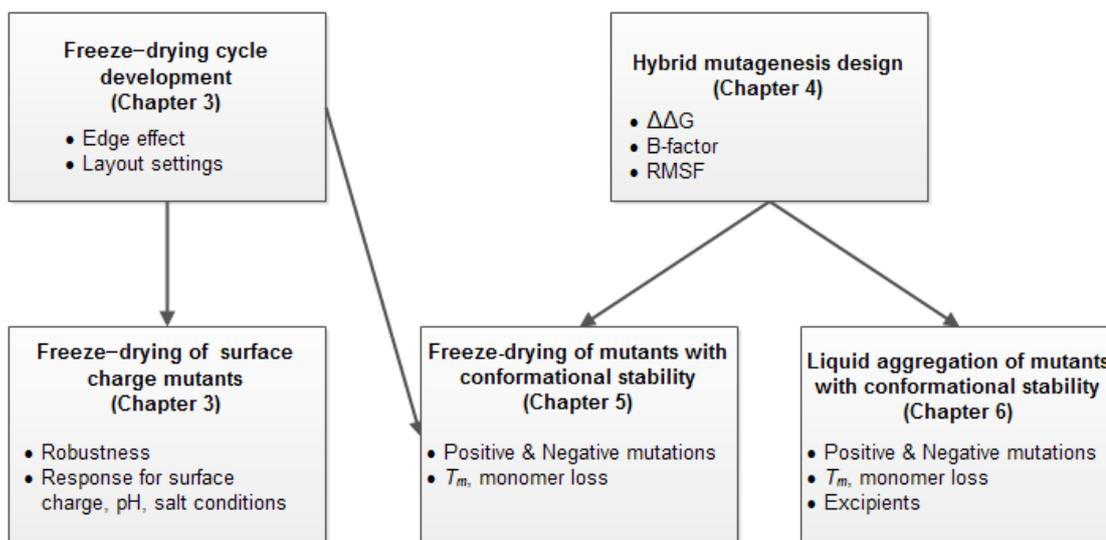


Figure 1.15 Schematic diagram for the strategy to conduct the study of freeze-drying of proteins

As can be seen from Figure 1.15, the first study (Chapter 3) aimed to characterise the edge effect of the freeze-drying, and propose a method to minimise that. This method was then used to run freeze-drying on 96-well plates in the subsequent studies. Afterwards, the impact of surface-charge on Fab was studied for freeze-drying across a range of pH, ionic strength and salt types. Monomer loss and cake morphology were used to indicate the Fab stability and cake macro appearance.

The next step is “Hybrid mutagenesis design” (Chapter 4). It aimed to develop stabilising and destabilising mutants for both freeze-drying and liquid aggregation kinetics work in the subsequent two chapters. Both B-factor and RMSF would be used to identify the flexible sites, and with Rosetta to propose potential mutations that could rigidify those flexible sites. In addition, several destabilising mutants would also be developed as a reference to validate the mutagenesis strategy.

In the section of “Freeze-drying of mutants with conformational stability” (Chapter 5), the aim was to assess the impact of protein conformational stability changes ( $\Delta T_m$ ) due to mutations designed from the previous chapter, upon the aggregation tolerance of proteins to freeze-drying. The results implied that both  $T_m$  and  $\Delta\Delta G$  could not elucidate adequately the monomer loss after freeze-drying.

Due to the more complex range of factors involved in freeze-drying than that in liquid phase, it was decided to conduct a series of liquid aggregation kinetics studies (Chapter 6) upon the designed mutants from Chapter 4. It aimed to validate the usefulness of  $T_m$  and  $\Delta\Delta G$  more generally for protein formulation, and provide a simpler denaturing pathway whereby the aggregation rate might be expected to correlate better. The stabilising effect offered by excipients would also be evaluated, such that propose a more efficient formulation screening strategy with good  $T_m$  correlations.

## 2 Materials and Methods

### 2.1 Materials

#### 2.1.1 Buffer

Various solvents and buffers were used for this project. Most chemicals were purchased from Sigma, which included ammonium hydroxide, isopropanol, sodium acetate, monohydrate citric acid, sodium citrate tribasic dihydrate, sodium monobasic phosphate, sodium dibasic phosphate, potassium monobasic phosphate, phosphoric acid, sodium chloride, potassium chloride, and ethanol were purchased from Sigma Chemical Co. Ltd. (Dorset, UK). Chemicals used to prepare HPLC solution were HPLC analysis grade. All of the other chemicals were reagent grade (or above). After preparation, buffers were filtered through a Stericup Filter Unit with pore size 0.22  $\mu\text{m}$  (Millipore, Watford, UK).

A series of compositions for the buffer used is listed in Table 2.1.

Table 2.1 A list of buffer compositions

<b>Buffer</b>	<b>Composition</b>	<b>Final pH</b>
Phosphate Buffer Saline (PBS)	$\text{KH}_2\text{PO}_4$ 0.2 g/L, $\text{Na}_2\text{HPO}_4$ 1.15 g/L, NaCl 8 g/L, KCl 0.2 g/L	7.4
Extraction buffer	12.11 g/L Tris base, 2.92 g/L EDTA	7.4
Protein G equilibration buffer	$\text{Na}_2\text{HPO}_4$ 2.655 g/L, $\text{NaH}_2\text{PO}_4$ 0.757 g/L	7.4
Protein G intermediate buffer	$\text{Na}_2\text{HPO}_4$ 2.655 g/L, $\text{NaH}_2\text{PO}_4$ 0.757 g/L, 99.8% isopropanol 50 ml/L	7.4
Protein G elution buffer	Monohydrate citric acid 9.638 g/L, sodium citrate tribasic dihydrate 4.155 g/L	3.4

## 2.1.2 Growth media for *E. coli*

The growth media were prepared according to the formulations indicated in Table 2.2. Deionised H<sub>2</sub>O was used to dissolve the media ingredients, which was then autoclaved at 121°C for 20 minutes.

Agar plates were prepared under aseptic conditions by pouring 15 ml of warm liquid agar solution into a petri dish and cooled at ambient temperature. Afterwards, these plates were placed up-side-down, dried at 37°C and incubated at 4°C.

Table 2.2 The composition of growth media

Growth Media	Composition	Final pH
Luria Bertani (LB) agar	10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl + agar	7
Luria Bertani (LB) media	10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl	7
2XPY media	16 g/L phytone, 10 g/L yeast, 5 g/L NaCl	7
2XPY agar	16 g/L phytone, 10 g/L yeast, 5 g/L NaCl, 20 g/L select agar	7
SM6GC media	5.2 g/L NaH <sub>2</sub> PO <sub>4</sub> , 3.3 g/L Na <sub>2</sub> HPO <sub>4</sub> , 4.4 g/L KCl, 1.04 g/L MgSO <sub>4</sub> , 4.16 g/L citric acid, 0.25 g/L CaCl <sub>2</sub> , 112 g/L glycerol, 10 ml/L SM6 elements	6.8, adjust with w/50% NH <sub>4</sub> OH

(The pH adjustment for SM6GC media should be achieved by a slow addition of NH<sub>4</sub>OH as an excess of pH 7.15 in the local pH would induce the precipitation of phosphates and calcium within the solution.)

## 2.1.3 Model proteins

A33 Fab and a range of mutants derived from it were used as model proteins to study freeze-drying. The A33 Fab targets a cell surface marker in cancer cells (King et al. 1995; Welt et al. 2003) but its exact antigen has not been recognised yet. The wild type (WT) A33 Fab was provided by UCB Celltech, UK and the various mutants were engineered at UCL by Shahina Ahmad during her PhD project (Ahmad 2011). The WT A33 Fab molecule, which exhibits a free cysteine in position 226, tends to dimerise with another Fab molecules through disulfide bridge formation. To avoid this dimerization, the free cysteine was mutated into a serine in the C226S mutant (pseudo-wild type). Additional point mutants (e.g. K65M, K133M, S75K, L50K) were designed by varying the surface charge of the molecule and thus impacting on the aggregation propensity and

stability of Fab. For example, S75K was less prone to form aggregates than the WT Fab (Ahmad 2011).

The production, purification and assays involved to characterise the Fab properties are illustrated throughout Section 2.2 and will be used in the subsequent chapters.

## 2.2 Methods

### 2.2.1 Plasmid Mutagenesis

The mutagenesis aimed to change the genes in wild type plasmid so that different Fab could be expressed during cell culture. It is mainly made up of “site directed mutagenesis” and “TOP 10 transformation”.

#### 2.2.1.1 Site directed mutagenesis

In this step, reagents and protocol was used by referring to “QuikChange II XL Site-Directed Mutagenesis Kit”. Specifically, a mixture of 5 µl 10x reaction buffer, 2 µl dsDNA template (100 ng/µl), 2 µl forward primer, 2 µl reverse primer, 1 µl dNTP mix, 3 µl Quik Solution were prepared, which was then topped up to 50 µl by adding ddH<sub>2</sub>O. After that, 1 µl Pfu Turbo DNA polymerase was added to the 50 µl mixture and the whole mixture was used for PCR. The entire PCR consisted of 25 cycles and lasted for around 7.3 hours with cycling parameters as shown in Table 2.3. After PCR completed, the PCR reaction solution was digested with 1µl Dpn I, which was followed by gentle mixing, spinning down for 1 min and incubated for 1 hour at 37 C. A DNA agarose gel (Section 2.2.6) was conducted to confirm the reaction products were present and detectable. The details of designed primers were listed in the subsequent chapters.

Table 2.3 PCR operating parameters

	Temp (°C)	Time (min)
Heated lid	105	
Initial denature	95	2
25 cycles	95	1
	60	1
	68	15

Final extension	72	10
Final hold	4	

### 2.2.1.2 TOP 10 transformation

In this step, reagents and protocol were used by referring to “One Shot TOP10 Chemically Competent *E. coli*”. Specifically, competent cells were thawed on ice after being taken out from -80°C freezer. Then 5 µl of PCR synthesis solution was added into the competent cell vial with gentle mixing. The vial was then incubated on ice for 30 min. After that, heat-shock was executed by putting the cells into 42°C water bath for 30 seconds without shaking, which was followed by incubation on ice for 2 min. Afterwards, 250 µl S.O.C. solution was added to the competent cell vial and the vial was shaken at 225 rpm in 37°C incubator for 1 hour. Then the cells were spread onto Tet<sup>r</sup> LB agar and incubated at 37°C overnight. If colonies could be observed the next day, single colonies would be picked and grown overnight for plasmid extraction and sequencing (Section 2.2.2.3, 2.2.2.4) to verify the success of transformation.

The plasmid purified from TOP10 cells was further transformed into W3110 *E. coli* (Section 2.2.2.5) to prepare glycerol stock (Section 2.2.2.6) for subsequent fermentation use.

## 2.2.2 Production of cell and plasmid stocks

The production practise of cell and plasmid stocks was undertaken under aseptic conditions with the use of laminar flow cabinet and necessary disinfectants (e.g. 75% ethanol). After being taken out from storage, cells and plasmid stocks were preserved in ice.

### 2.2.2.1 Streak Cultures

Streak plates were made to obtain the strict condition to select plasmid before culturing. Cells kept in glycerol stock were scraped and spread onto LB agar following a quadrant streak technique, which offers the growth condition for isolated bacterial

colonies. The streak plates were wrapped and sealed with parafilm followed by incubation at 37°C for 12-14 hours.

### *2.2.2.2 Overnight Cultures*

5 ml LB Media was placed in a 50 ml Falcon tube. An inoculation loop was used to pick cells from a single cell colony in the streak plate and inoculate the Falcon tube with LB Media. The Falcon tube was then incubated attached onto a thermo shaker at 37°C, 250 rpm for 12-14 hours.

### *2.2.2.3 Plasmid extraction*

Protocols adjusted from a QIAprep Spin Miniprep Kit were applied for the purification of plasmid DNA from the 5 ml overnight culture in a Falcon tube. The following procedures were conducted with slight adjustment from the kit instructions.

1. The overnight culture was centrifuged at 5400 g, 4°C for 10 minutes.
2. After removing the supernatant, the pelleted cells were resuspended in resuspension buffer containing RNase A and LyseBlue. The resuspension of cell/buffer mixture was then transferred to a microcentrifuge tube.
3. Lysis buffer was added with an equal amount of resuspension buffer. The previously added LyseBlue is a colour indicator to visually indicate the extent of mixing during the alkaline lysis stage. A homogeneous coloured suspension was eventually obtained after gentle mixing.
4. A neutralization buffer with the 1.4 fold volume of lysis buffer was added. The solution was drastically mixed until became colourless, which indicated an effective SDS precipitation.
5. The solution complex was then centrifuged at 17900 g for 10 minutes to achieve a tight white pellet.
6. The supernatant was transferred to a 1 ml QIAprep spin column and the DNA would bind to the silica-gel membrane within a brief period. The QIAprep spin column was centrifuged for 30-60 seconds and the flow through was disposed.

7. To remove the remaining nucleases and carbohydrates bound on the column, 0.5 ml binding buffer was added and flow through was discarded again after centrifugation for 30-60 seconds.
8. To wash the column, 0.75 ml wash buffer was added. The column was centrifuged for 30-60 seconds and flow through was discarded. A further centrifuge of 60 seconds was used to remove residual wash buffer within the column.
9. The plasmid was eluted by adding elution buffer or water of 50  $\mu$ l to the column. The column was stood for 1 minute, centrifuged for another minute, and flow through was collected for the recovery of plasmid.

Plasmids were stored in water at -20°C for short term storage and elution buffer for long term storage.

#### *2.2.2.4 Plasmid Sequencing*

A Thermo Scientific Nanodrop (Wilmington, USA) was used to verify the plasmid solution purity before sequencing. The sensor was initially washed and cleaned with ddH<sub>2</sub>O. A baseline was set by the use of water or elution buffer according to the plasmid storage condition, followed by the scanning of a 2  $\mu$ l plasmid solution sample.  $A_{260/280}$  of 1.8 is desired to indicate the low contaminant of protein and  $A_{260/230}$  of less than 2.0-2.2 shows minimal contaminant of lysis/wash buffer or carbohydrate.

After dilution with ddH<sub>2</sub>O to 100 ng/ $\mu$ l, the plasmid solution was send to Wolfson Institute for Biomedical Research for sequencing together with custom primers at 5 pmols/ $\mu$ l requested from Operon Biotechnologies. A BioEdit sequence alignment editor (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>) was used to compare the sequence analysis results with the master sequence to recognise any errors within the plasmid sequence.

### ***2.2.2.5 Bacterial Transformation***

The transformation of *E. coli* with the plasmids was conducted by using a Bio-Rad Micropulser electroporation unit (Hertfordshire, UK). After allowing the cells to thaw on ice, 1 µl plasmid solution (~100 ng plasmid) was transferred to a tube with 50 µl cells followed by gentle tapping to mix. The tube was incubated for 2 minutes and the mixture was then transferred to a pre-chilled electroporation cuvette (1 mm gap width). The cuvette was inserted into the pre-chilled electroporation chamber without any condensation in the unit, and a 1.8 kV electrical pulse was instantly applied to the sample. 250 µl SOC media was added to the mixture and the contents of the cuvette were transferred to a 15 ml falcon tube and incubated at 37°C, 250 rpm for 1 hour. After incubation, different volumes of pre-culture were dispersed onto 2XPY agar by a lawn spreader. Finally, the plates were sealed with parafilm and incubated at 37°C for 14-16 hours.

### ***2.2.2.6 Glycerol stocks***

Glycerol stocks were used for long term storage of *E. coli* with or without the presence of plasmids. On the condition that the growth is within the log phase, 50 µl of 40% v/v glycerol/dH<sub>2</sub>O was filtered with a sterilised 0.2 µm PVDF filter and added to an equal volume of overnight cell culture. The stocks were stored at -80°C and would be thawed only for immediate use.

## **2.2.3 Pilot scale Fab production**

### ***2.2.3.1 Inoculum preparation***

A pre-culture was undertaken to provide a vigorous cell growth condition with a log phase production. The cell suspension of 250 µl was taken from an overnight culture and added to 200 ml 2XPY media in a 2 L baffled shake flask. OD<sub>600</sub> was monitored periodically by a Cecil Aquarius Spectrophotometer. After incubating for 4 hours at

37°C, 250 rpm and OD<sub>600</sub> reaching 1-2, 30 ml pre-culture aliquots were transferred into four 2 L baffled flasks each containing 30 ml SM6GC media. These seed cultures were incubated for 12-16 hours until a final OD<sub>600</sub> of 4-5 was achieved. All of the seed culture was the inoculum for the subsequent fermentation.

### 2.2.3.2 Fermentation (30 L Sartorius)

The pilot scale fermentation was performed in a 30 L Sartorius Stedim Biostat C Plus bioreactor for the production of Fab proteins. The sterilisation was carried out *in situ* by decanting 18 L SM6GC media into the bioreactor before inoculation. The feed bottles, tubing, inoculum flasks and all other ancillary components were sealed and externally autoclaved before using. After sterilisation, the bioreactor vessel was cooled and maintained at 30°C throughout the fermentation process. After that, the inoculation was achieved by a gravity feed of 2 L log phase seed culture into the fermenter vessel via the sterilised inoculum flask, yielding an entire working volume of 20 L.

The fermentation was monitored and controlled on-line. The overall homogeneity of the internal environment was conserved by agitation using a 3-tiered impeller. A pH of 6.95 was maintained by the addition of diluted acid (i.e. 15-20% v/v H<sub>3</sub>PO<sub>4</sub>) and base (i.e. 15% v/v NH<sub>4</sub>OH) solutions through a Biostat control unit. Adequate oxygen level was controlled by sparging sterilised air at 20 L/min. The mixing of the feed gas with pure oxygen (6:4 volume ratio of oxygen:nitrogen) was conducted if the dissolved oxygen tension (DOT) was lower than ~40% (assuming standard air as 100%). An anti-foaming agent, namely PPG 2000, was added by Biostat control system to mitigate the cell lysis rate.

The vent gas composition was monitored online during the fermentation. Culture aliquots were taken every 2 hours for OD<sub>600</sub> measurement to follow the cell growth. In order to enhance the structural integrity of cell walls so as to reduce the leakage of desired intracellular protein, 150 ml of magnesium “shot” (1 M MgSO<sub>4</sub>·7H<sub>2</sub>O) was added to the fermenter and the temperature was lowered to 25°C to slow down growth rate once the OD<sub>600</sub> reached 40.

A DOT spike and a decrease of carbon dioxide level indicated the advent of a stationary phase in which carbon source within the media had been used up and cell growth was limited. At this stage, the bioreactor was switched to fed-batch operation with a constant feed flow of 0.7 ml/min 80% w/w glycerol/dH<sub>2</sub>O, which held the cell culture in a stationary phase. The protein expression was induced by adding 50 ml of IPTG (64 mM). Based on previous fermentation runs, an optimal intracellular target protein amount was achieved by commencing the harvest of the cell culture after 22 hours of the induction.

### *2.2.3.3 200 ml fermentation*

DASbox Mini Bioreactor was used to perform fermentation for multiple mutants in the same time. A maximum of four bioreactors were used with working volume around 170 to 180 ml. Most of the operating parameters were determined by scaling down from the 30 L Sartorius fermenter protocol (Section 2.2.3.2). Due to the miniature property of the reactors, several parameters were optimised after conducting a few cycles.

In the pre-culture stage, 1 ml glycerol stock was taken from -80°C freezer and inoculated into 20 ml 2XPY define in a 250 ml shake flask. The cell culture was conducted at 37°C, 250 rpm for 3-4 hours until the OD 600 reached 1-2. Afterwards, 2 ml of the cell culture was transferred into 20 ml define media, which was cultured at 30°C, 250 rpm for around 16 hours until the OD 600 reached 3-4.

The entire cell culture in define media was then inoculated into the DASbox bioreactor, which had already been filled with 150 ml define media. A minimum DOT of 30% was controlled by a combination of agitation, gas flowrate and oxygen proportion in the gas. The screw at the end of the gas sparger was removed to prevent blockage due to high cell density. The pH value was maintained at 6.95 by 15-20% v/v H<sub>3</sub>PO<sub>4</sub> and 15% v/v NH<sub>4</sub>OH. Temperature was kept at 30°C in the beginning.

About 20 hours after inoculation, magnesium shot was conducted when OD reached 40 and temperature was decreased to 25°C. Then after around 10 hours, a DOT spike was observed when the nutrition had been used up and growth entered a stationary

phase. The OD usually reached 150 to 200. A fed-batch was conducted immediately after DOT spike by adding 0.5 ml 64 mM IPTG and continues 80% w/w glycerol feeding at 0.7 ml/h. It is noted that different mutants would exert various growth profiles so one may customise some parameters including agitation rate, Mg shot/IPTG amount and glycerol feeding rate. The fermentation broth was harvested 16-24 hours after induction.

#### ***2.2.3.4 Tubular bowl centrifugation***

Cell harvest was achieved by using a Carr Powerfuge P6 Centrifuge (Sittingbourne, UK) with a feedstock flow rate of 500 ml/min. The slurry temperature was cooled at 10°C. The sedimentary solid in the centrifuge were discharged periodically when the capacity of centrifuge bowl was reached. The dewatered cell paste was then split into aliquots, packed in sample bags and kept at -80°C.

#### ***2.2.3.5 200 ml centrifugation***

Cell harvest with low volume (100 – 300 ml) was clarified by BECKMAN COULTER J2-MC centrifuge (High Wycombe, UK) at 10,000 rpm and 4°C for 90 min. The supernatant was removed afterwards and the sediment cell paste was grabbed out, stored into sample bags and kept at -80°C freezer.

#### ***2.2.3.6 Protein extraction***

In order to obtain the target protein whilst minimising the contamination by the host cell proteins (HCP), EDTA, which is a chelating agent, was used to disrupt the cell wall by removing stabilising agents such as magnesium and release the target proteins in the periplasm of *E. coli*.

The cell paste was resuspended in extraction buffer to form a 15% w/v slurry. 400 ml aliquots of the slurry were individually transferred into 2 L baffled shake flasks and incubated for 14-16 hours at 50°C, 250 rpm. The extraction solution was cooled to ambient temperature, centrifuged at 10,000 rpm, 4°C for 1.5 hours in a Sorvall Super T21 centrifuge (Basingstoke, UK) and the supernatant collected and stored at 4°C.

### *2.2.3.7 Protein filtration*

Near solid-free solution from the last protein extraction step was filtered to remove any large particles and unwanted proteins of larger size than the target proteins. This was done by using a vacuum filtration system with filtration membrane filters of decreasing membrane pore size down to 0.2  $\mu\text{m}$ .

### *2.2.3.8 Protein G chromatography*

Protein G chromatography was conducted on an AKTA Purifier FPLC system installed with a XK50 column. The column was packed with Sepharose Fast Flow Protein G resin. The purification was initialised by equilibrating the column with 3 column volumes (CVs) of 25 mM sodium phosphate, pH 7.4. Afterwards, the protein filtrate was loaded through a sample tube by a sample pump without reaching the maximum dynamic binding capacity of the Protein G resin. A first washing step was carried out with three CVs of equilibration buffer to remove any unbound substance, followed by two CVs of equilibration buffer plus 5% v/v isopropanol to remove any hydrophobically bound impurities. Finally, the target proteins were eluted by 3 CVs of 60 mM sodium citrate, pH 3.5, and neutralised by 1 M Tris base, pH 8.5. The neutralised samples were aliquoted and stored at  $-80^{\circ}\text{C}$ .

### *2.2.3.9 Size-exclusion chromatography*

The product solution at this point contained some protein fragments and aggregates of target proteins. A preparative size-exclusion chromatography (gel filtration) was applied to further purify the product. It was done using an AKTA Prime FPLC system with a Superdex 200 chromatography column. The protein solution was injected into a 5 ml sample loop and connected to the FPLC system. Phosphate buffer saline (PBS) solution was used during the gel filtration step. It was conducted by equilibrating the column with 2 CV, injecting the sample from the sample loop and recovery for the fractions of protein monomers.

### 2.2.3.10 0.02 $\mu\text{m}$ filtering

When multiple Fab mutants were needed, it took excessively long period to prepare individual mutants through size-exclusion chromatography. Therefore, Fab samples were filtered through Anotop 25 0.02  $\mu\text{m}$  syringe filters (GE Healthcare, Buckinghamshire, UK) to remove any aggregates.

### 2.2.3.11 Buffer exchange

Buffer exchange was conducted by using Slide-A-Lyzer Dialysis cassettes with a 10 kDa cut-off (Fisher Scientific, Leicestershire). A syringe was used to inject protein sample into the dialysis membrane chamber. A small amount of air was left in the syringe before filling the syringe with the sample in order to minimise any sample loss in the syringe's dead volume during injection. After that, the dialysis cassette was kept within a bucket filled with the dialysis buffer, which was more than 200 times the volume of the sample. The bucket was incubated at 4°C with a stirrer rotating at the bottom to facilitate the dialysis. The dialysis buffer was replaced after 2 h before dialysis overnight. The dialysis cassette was taken out from the dialysis buffer the next morning and the sample inside the cassette was transferred to a falcon tube and store at 4°C.

## 2.2.4 Protein quantification

The protein concentration was calculated using the Beer-Lambert law:

$$c = \frac{A}{\epsilon l} \quad \text{Equation 2.1}$$

where  $A$  is the absorbance value,  $\epsilon$  is the extinction coefficient,  $l$  is the path length of cuvette and  $c$  is the unknown protein concentration. The extinction coefficient of Fab is  $1.4 \text{ (mg/ml)}^{-1}\text{cm}^{-1}$  (Ng & Osawa 1997).

## **2.2.5 SDS-PAGE**

SDS-PAGE was operated to qualitatively determine the Fab amount from fermentation and purification sampling. For each sample, 15  $\mu$ l sample, 7.5  $\mu$ l 0.2 M DTT and 7.5  $\mu$ l 4x Laemmli sample buffer (BIO-RAD) were mixed. Then the mixture was incubated at 95°C for 10 min to denature the protein. After heat block, a Mini-PROTEAN TGX Gel (BIO-RAD) was installed on a gel tank. Protein samples together with marker and standard sample were loaded onto the gel wells, which was then run at 200 V for around 40 min.

After running the gel, the gel was stained by InstantBlue (Expedeon) for 30 min. The staining buffer was then replaced with RO water to de-stain. The washing water was repeated for 2-3 times until clear protein bands could be visualised. Images were taken to record the sample characterisation by SDS-PAGE under white light.

## **2.2.6 DNA gel electrophoresis**

DNA gel was conducted to verify if the plasmid of interest existed in the sample. A gel was prepared by adding 0.5 g agarose into 50 ml TBE. The solution was then heated by microwave and shaken to fully dissolve the agarose. Then 2.5  $\mu$ l ethidium bromide was added and sufficiently mixed. After that, the entire agarose melted solution was poured into a plastic tray and a comb was inserted on its top. After 10-15 min, the agarose cooled down and solidified. Wells were created by vertically remove the comb.

The plastic tray containing solidified agarose gel was then put into a gel tank. The electrophoresis was operated at 100 V for around 30 min. When finish, the gel was visualised under UV lamp.

## **2.2.7 Freeze-drying**

Freeze-drying was conducted on the freeze dryer Virtis Genesis 25EL. Micro-titre plates filled with solutions were loaded onto the freeze-drying shelves. Freeze-drying cycle was conducted after programming the method on PC. If not specified, the method

parameters were shown in Table 2.4, but it would be adjusted according to practical requirement.

Table 2.4 An example for the process cycle parameters of freeze-drying

		Temp (°C)	Time (min)	Ramp/Hold	
Freezing	Step # 1	20	30	H	
	Step # 2	-40	120	R	
	Step # 3	-40	60	H	
Endpoint of freezing	Freeze	-40	(°C)		
	Extra Freeze	0	(Minutes)		
	Condenser	-40	(°C)		
	Vacuum	500	(Milli Torr)		
Primary & secondary drying		Temp (°C)	Time (min)	Ramp/Hold	Vacuum (mTorr)
	Step # 1	-40	60	H	100
	Step # 2	-20	30	R	100
	Step # 3	-20	600	H	100
	Step # 4	25	300	R	20
	Post Heat	25	60	H	20
	Secondary drying setpoint	27	N/A	N/A	N/A

After a freeze-drying cycle had finished, samples were taken out from the chamber. Precautionary measures (e.g. stoppers) were used to prevent water vapour from coming into the containers. An eye observation by practitioner was applied to indicate the morphology of freeze dried samples in individual containers. Cake with plumpy appearance would be determined as good morphology while those of collapse structures would be regarded as bad ones. The freeze dried solids were then reconstituted with reconstitution buffer and shaken onto a thermomixer to achieve reasonable solid dissolution at 300-500 rpm for 5 min. OD<sub>600</sub> was carried out again to check the turbidity and compare to the previous values of liquid formulation. The spinning down of sample was taken if the reconstitution solution was of high turbidity and solution was transferred to another container if necessary. In the end, the reconstitution solution was analysed

by assays as required. Each type of the assays for a same batch was completed within 24 h.

## 2.2.8 Size-exclusion chromatography HPLC (SEC-HPLC)

Fab fragments, monomer and soluble aggregates within the sample, were analysed by size-exclusion chromatography HPLC (SEC-HPLC) on an Agilent Zorbax Bio Series GF-250 column (Agilent, Berkshire, UK) with an Agilent 1200 HPLC system, and a mobile phase of 200 mM sodium phosphate, pH 7, at a flowrate of 1 ml/min, for 4.5 min in each cycle. Calibration curves were established prior to each batch of analyses. Protein samples at 1 mg/ml were loaded in 10  $\mu$ l onto the column, and the Fab monomers eluted at 2.6 min. A typical SEC-HPLC profile is shown in Figure 2.1. Analysis was performed on liquid formulations prior to freeze-drying to determine the initial monomer concentration in each sample, and then again for fully reconstituted freeze-dried products to determine the final monomer concentrations. Relative monomer loss was calculated as shown in Equation 2.2.

$$\text{relative monomer loss} = \frac{\text{initial monomer} - \text{final monomer}}{\text{initial monomer}} \quad \text{Equation 2.2}$$

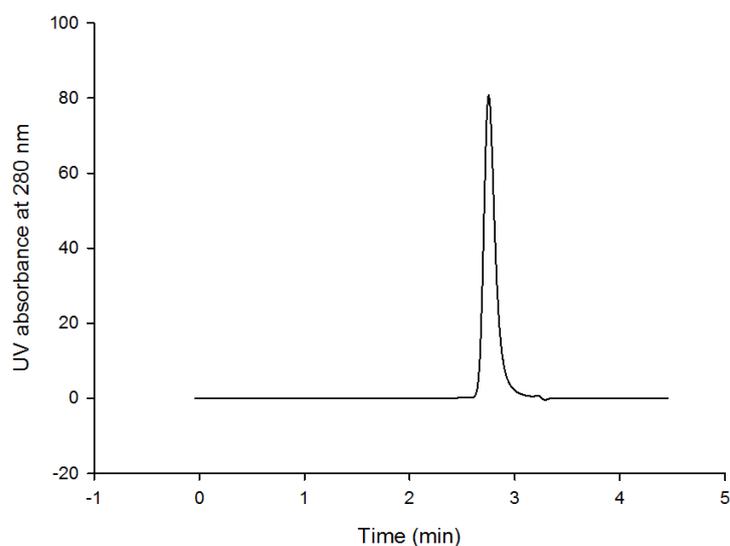


Figure 2.1 A typical SEC-HPLC chromatogram.

10 µg of Fab was injected to the Agilent Zorbax Bio Series GF-250 column (Agilent, Berkshire, UK) with an Agilent 1200 HPLC system, and a mobile phase of 200 mM sodium phosphate, pH 7, at a flowrate of 1 ml/min. Fab monomers were observed at 2.6-2.8 min.

# 3 Impact of surface-charge mutations on the freeze-drying of Fab

## 3.1 Introduction

Formulating proteins in a desired form, while preventing degradation during manufacturing and storage, is a long-standing challenge. Considerable research has increased our understanding of the conditions that influence protein stability to denaturation, chemical modification, or aggregation, such as physical (e.g. temperature, shear) and chemical stresses (e.g. pH, redox potential, ionic strength) (Wang et al. 2007). Both liquid and lyophilised formulations are commonly used to preserve proteins, and typically require the optimisation of protein concentration, pH, buffer type, excipient composition, vial/syringe enclosure materials, and process equipment/containers (Wang 2000; Daugherty & Mrsny 2006; Uchiyama 2014). Freeze-dried dosage forms offer longer shelf-life and ease of storage and transportation (Manning et al. 2010). The influence of process development and formulation upon freeze-drying performance has been well characterised (Kasper et al. 2013), and a range of properties including thermal transitions ( $T'_g$ ,  $T_{eu}$ ,  $T_c$ ) (Kett et al. 2004; Meister & Gieseler 2008), porosity (Fonte et al. 2012), residual moisture (Schersch et al. 2010), reconstitution rate (Telikepalli et al. 2015) have been found to be impacted by the freeze-drying process.

Freeze-drying of proteins is usually conducted with the addition of excipients (e.g. disaccharide) to inhibit aggregate (Shukla et al. 2011) and reform the hydrogen bond as water is depleted during drying (Mensink et al. 2015). Lyophilised protein formulations with higher concentration has often been suggested as it reduces the fraction of proteins accumulated to the ice-liquid interface during freezing (Kuelto & Wang 2008). It was found that freeze-drying of high protein concentration (115 mg/mL) in buffer-free formulations resulted in comparable stabilization to buffer-based formulations (Garidel et al. 2015).

Empirical testing of various formulation compositions is typically required in pre-formulation stage, and usually carried out in vials. The lower sample volume, and shelf-space requirements of microscale freeze-drying, now offers the potential to screen many more formulations for stability to the freeze-drying process, and simultaneously an acceptable final cake morphology. We previously developed a freeze-drying framework using 96-well plates with trimmed bottom lips, in which the microscale activity data was found to scale up well into stoppered vials (Grant et al. 2009; Grant, Matejtschuk, et al. 2012; Grant, Dalby, et al. 2012). Recently, the freeze-drying of lysozyme formulations was explored on a microscale heating stage, with samples under a glass cover slip, or inside a hollow circular polypropylene holder (Peters et al. 2014). This freeze-drying platform was similar to that used in freeze-drying microscopy, and the sample holders further enabled the pore size and microcollapse to be analysed by scanning electron microscopy and micro-X-ray computed tomography, without additional transfer or manipulation of samples. Others have found the freeze-dried solids formed within custom fabricated brass well plates, commercial 96-well plates, and conventional vials to be comparable in terms of polymorphic form, residue moisture, cake collapse and reconstitution time (Trnka, Rantanen, et al. 2014).

Beyond freeze-drying process and formulation optimisation, proteins can also potentially be mutated, or chemically modified (e.g. PEGylation) to improve their stability. Mutations have been explored widely for stabilising proteins in the aqueous phase, and *in vivo* (Chiti, Calamai, et al. 2002; Calloni et al. 2005; Teilum et al. 2011). However, few studies have examined the impact of mutations upon freeze-drying performance, and these have been limited to only a few specific mutations. For example, the S80R mutation of RNase A was found to increase the population of domain-swapped oligomers after freeze-drying in 40% acetic acid solutions (Vottariello et al. 2011).

In the following sections, the mutation impact on protein aggregation and stability will be reviewed. The impact will be classified into substitution of alanine/glycine and proline, charge, hydrophobicity, and secondary structure.

### 3.1.1 Mutation impact by A/G substitution

The mutation between alanine and glycine has been studied by many researchers. Due to the additional methyl in the alanine side chain, the replacement of A/G is an ideal mutational example to investigate the impact of  $\beta$ -carbon on protein stability. The mechanism lies in the  $\alpha$ -helix stabilisation by alanine (Rezácová et al. 2008), which would impact on conformation entropy (Rezácová et al. 2008; Scott et al. 2007).

In the secondary structure, glycine is not good to form  $\alpha$ -helix structure while alanine contributes to its formation and occurs more frequently in  $\alpha$ -helix structure (Pace & Scholtz 1998). As glycine has no  $\beta$ -carbon, this amino acid can exert more backbone conformation flexibility than alanine (López-Llano et al. 2006). Therefore, it requires more free energy to fold the regions with glycine due to the greater loss of entropy required. It was found that mutations of glycines to alanines in amyloid beta-peptides would increase the helix content and reduce the beta-sheet formation, which could inhibit the amyloid fibril formation (Xu et al. 2005). Scott (Scott et al. 2007) measured that the Ala to Gly mutation typically causes a 0.4 kcal/mol increase in conformational entropy in the unfolded state. After studying 22 A/G mutations in  $\alpha$ -helices, Rezacova (Rezácová et al. 2008) found that alanine persistently stabilised the conformation compared to glycine as more non-polar regions could be buried on folding so as to reduce the backbone entropy. This technique was also used to reengineer G-CSF to increase its  $\alpha$ -helical propensity in the anti-parallel 4-helix bundle (Bishop et al. 2001).

Alternative findings suggest that glycine acts in a more conserved way in evolution and inhibits aggregation (Parrini et al. 2005). The reason is that glycine exhibits low propensity to adopt  $\beta$ -sheet structure, which is a factor that promotes the forming of amyloid fibrils. It is also observed that in elastomeric proteins, a high glycine proportion prohibits the formation of aggregate into a stable  $\beta$ -sheet structure (Monsellier & Chiti 2007). Therefore, there is a trade-off between using alanine to stabilise the  $\alpha$ -helix structure and glycine to prevent aggregation.

### 3.1.2 Mutation impact by proline

The substitution of proline is another strategy to improve protein stability. It performs in the way that its unique side chain pyrrolidine confers to proline a strong conformational rigidity, which greatly reduce its entropic instability (Eijsink et al. 2004). It is found that the pyrrolidine makes proline difficult to constitute a  $\beta$ -sheet structure, which renders it more conserved in fibronectin type III domains and less prone to aggregate (Steward et al. 2002). The introduction of proline substitution in amyloidogenic polypeptide can substantially increase their solubility so as to lessen amyloid formation (Williams et al. 2004).

Proline may also play a destabilising role. Morimoto (Morimoto et al. 2002) conducted several mutations into prolines on amyloid  $\beta$  peptides. Though most of the substitution showed less aggregation propensity, one proline mutant showed much higher aggregation propensity, which was thought to form a  $\beta$ -turn structure. Moreover, due to pyrrolidine ring hindrance, proline would disrupt the secondary structure like  $\alpha$ -helix and  $\beta$ -sheet (Savage & Gosline 2008). Thus, the position of proline in the protein secondary structure needs to be considered so as to assist favourable folding free energy. Wedemeyer (Wedemeyer et al. 2002) investigated the protein folding mechanism by incorporating non-native proline isomers. They found that prolines in loosely ordered loop positions are not essential; prolines that are completely buried or in tight turns should be in their native cis-trans forms so as to properly fold into functional structures.

### 3.1.3 Mutation impact on protein charge and hydrophobicity

Naturally, the polypeptides of proteins fold *in vivo* to enable the burial of non-polar, hydrophobic residues and exposure of polar, chargeable ones (Dill & MacCallum 2012). The compact protein tertiary structure is stabilised by hydrophobic interaction, Van der Waals force, hydrogen bond, disulphide bond and salt bridge within the protein, hydrogen

bond interaction with solute molecules as well as long-range electrostatic interaction between proteins considered as colloidal stability (King et al. 2002; Dobson 2004; Baneyx & Mujacic 2004; Stefani & Dobson 2003; Uversky 2003). The aggregation of proteins upon solution condition change (pH, ionic stress) is mostly initiated by unfolding or misfolding of proteins, which expose their hydrophobic core to the outer surface and enhance the intermolecular interactions including hydrophobic interaction, electrostatic attraction and hydrogen bonding (Stefani & Dobson 2003; Uversky 2003). In order to strengthen the intrinsic protein stability, mutagenesis strategies could be applied to better pack the hydrophobic core, increase surface net charge and alter secondary structure to favour lower aggregation propensity (Lehmann et al. 2000; Eijssink et al. 2004)

Improving the net surface charge so as to enhance the electrostatic repulsion has been utilised by many researchers (Chiti, Calamai, et al. 2002; Lehermayr & Mahler 2011; Sheinerman et al. 2000). Chiti (Chiti, Calamai, et al. 2002) studied the mutation impact on muscle acylphosphatase (AcP) and found that the aggregation was favoured when the charge of mutants was close to neutrality. Strickler (Strickler et al. 2006) engineered more surface charge on five different proteins and observed a notable increase in stability over the wild type.

The extent of charge mutation on protein stability, however, should not be over-emphasised and should be carefully selected. As hydrogen bonds and van der Waals mainly contribute to the folding stability, charge mutations only give rise to small effect on structural stability (Dill et al. 2008). The free energy change is more accurately predicted if the charge mutation locates at the surface (Seeliger & de Groot 2010). It is suggested that the charge-charge interactions stabilise the protein in both folded and unfolded states, which may cause the net contribution to be destabilising (Pace et al. 2000). An introduction of charge mutation within the hydrophobic core was found to destabilise the folded protein structure (Wang & Moult 2001).

Increasing the hydrophobicity on the protein surface would likely promote aggregation. Munch (Munch & Bertolotti 2010) used a sensitive fluorescent-based assay to quantify the hydrophobicity of mutants from copper-zinc superoxide dismutase-1

(SOD1). It was found that the aggregation was initiated by exposing hydrophobic surfaces of the mutations. Wu (Wu et al. 2010) conducted mutagenesis on an anti-IL-13 monoclonal antibody and found that mutants with lower hydrophobic surface area demonstrated improved solubility without affecting the antigen binding affinity.

As hydrophobic interaction is the essential element that stabilises proteins (Nick Pace et al. 2014), mutations that alter the hydrophobic core will destabilise the proteins. It has been found that mutations of isoleucine to valine in the buried regions of *Borrelia burgdorferi* protein all decreased the stability compared to the wild type with average of 1.6 kcal/mol (Pace et al. 2011). Wang and Moulton (Wang & Moulton 2001) found that several single nucleotide polymorphisms (SNPs) related mutants reduce hydrophobic interaction with a loss of non-polar burial area of more than 50 Å<sup>2</sup> on folding.

However, the substitution of an amino acid would often affect both the charge and hydrophobicity simultaneously and so it is not always straightforward to separate their individual contribution to protein stability. An (DuBay et al. 2004) equation has been developed to illustrate the aggregation rate by incorporating the intrinsic factors (hydrophobicity, hydrophobic patterns, charge) and extrinsic factors (pH, ionic strength, protein concentration).

### **3.1.4 Mutation impact on secondary structure**

The modification on secondary structure could change the aggregation propensity (Chiti & Dobson 2006; Chiti, Taddei, et al. 2002; Fernandez-Escamilla et al. 2004). Chiti (Chiti, Taddei, et al. 2002) found that mutants of human muscle acylphosphatase were of higher aggregation sensitivity for regions with higher  $\beta$ -sheet propensity. They also found that (Chiti & Dobson 2006) more  $\alpha$ -helical and less  $\beta$ -sheet structure would inhibit the aggregation. Studies on prion protein (Lee et al. 2010; Tahiri-Alaoui et al. 2004) revealed that the normal cellular prion protein (PrP<sup>c</sup>) is predominantly  $\alpha$ -helix while pathogenic form (PrP<sup>Sc</sup>) is rich in  $\beta$ -sheet. The conversion from PrP<sup>c</sup> to PrP<sup>Sc</sup> is influenced by the common methionine/valine polymorphism at 129<sup>th</sup> residue in PrP, in which harmful mutants form the intermolecular  $\beta$ -sheet conformation. A study on human

Islet Amyloid Polypeptide (hIAPP) (Dupuis et al. 2011) showed that the early oligomers of hIAPP fibrils are formed by self-interaction of  $\beta$ -hairpin monomers.

### **3.1.5 Aims of the chapter**

Increasing the net surface charge of a protein has been used increasingly to stabilise proteins against aggregation in the liquid state (Chiti, Calamai, et al. 2002; Lehermayr & Mahler 2011; Sheinerman et al. 2000; Strickler et al. 2006), and is most likely due to an enhanced colloidal stability through electrostatic repulsion, as surface charge mutations typically only give rise to small effects on structural stability (Dill et al. 2008), compared to the structurally destabilising effects of introducing buried charges (Wang & Moulton 2001). The impact of surface charge mutations upon stability to freeze-drying has not been explored previously to our knowledge.

Here aimed firstly to more accurately define the freeze-drying process that would minimise variability in drying between microwells, and established a process in which  $T_c$  is not the primary driver for destabilization of the therapeutic Fab fragment under study. We then varied the pH, salt type and ionic strength, to study for the first time, the impact of protein surface-charge mutations on the robustness of a protein to freeze-drying, and also on the cake morphology achieved. While excipients can be added to stabilise against potential charge effects in either liquid or freeze-drying formulations, we aimed to examine whether charge effects on the protein actually exerted any influence in freeze-drying.

## **3.2 Materials and Methods**

### **3.2.1 Homogeneity of drying across the freeze-dryer shelf**

A test of shelf homogeneity was carried out using polystyrene, flat-bottom 96-well microplates (Greiner Bio-one Ltd, Gloucestershire, UK). The layout (Figure 3.1) of

plates on the freeze-dryer shelf enabled a comparison between the presence and absence of aluminium plates placed underneath microplates. It also enabled an assessment of the impact of the presence or absence of neighbouring water-filled microplates upon the water removal from wells of central four microplates. Four microplates were aligned in the centre of the shelf, alternated with and without aluminium plates. Five microplates placed along the left side of the central four plates had their bottom lip trimmed so that each well rested directly on the shelf. All of the 96 wells in the nine microplates were filled with 300  $\mu$ L water. No microplates were placed on the right of the central four plates.

In the freeze-drying cycle, the temperature was initially kept at 0°C for 60 min, then ramped to -30°C over 60 min and held at -30°C for 600 min. Drying was initialised by reducing the pressure to 100 mTorr and held at -30°C for 60 min, then ramped to -25°C over 30 min, held at -25°C for 60 min, and then the process stopped, which achieved approximately one third of the frozen solution removed by sublimation. Microplates were warmed to room temperature with all the ice thawed, then removed from the chamber and immediately covered with an impermeable sticky coversheet (Sealplate 146 x 79 mm 50  $\text{\AA}$ m, VWR International, UK) to prevent water evaporation. Before measuring the water remaining in the wells, any droplets adhered to the inner coversheet surface were added to the corresponding wells by pipette. The water from each well of the central four microplates, was transferred by pipette to a balance to determine the masses remaining in each well. All measurements were conducted on the same day of the freeze-drying.



Figure 3.1 The layout of the homogeneity test with 96-well plates

The four sample plates (light blue and dark blue), filled with 300 µL water, were placed in the centre of the shelf. The 1<sup>st</sup> and 3<sup>rd</sup> sample plates (light blue) sat directly onto the shelf without aluminium plates underneath; the 2<sup>nd</sup> and 4<sup>th</sup> sample plates (dark blue) were in contact with aluminium plates underneath. Trimmed plates filling with water (brown) were placed on the left of the sample plates to minimise the edge effect during sublimation.

### 3.2.2 Differential scanning calorimetry (DSC)

Aqueous protein samples of 1-2 mg/ml were loaded to 80 µL into pre-weighed steel pans, with lids and O-rings, and the pan and lid then crimped together. Sealed pans were weighed again to calculate the net sample weight. Differential scanning calorimetry (DSC) was performed on a Q2000 DSC (TA instrument, Crawley, Surrey, UK) using a sample pan together with a reference pan. Samples were held isothermally for 2 minutes, and then cooling ramped at 10°C/min to -90°C. Modulation was then applied at +/- 1.5°C every 60 seconds with a sampling interval of 1 second. Heating was then ramped at 3°C/min to 25°C. The glass transition values were determined from the transition midpoints using Universal Analysis 2000 software (TA Instruments, New Castle, US).

### 3.2.3 Freeze drying microscopy (FDM)

Freeze drying microscopy was performed on a Lyostat4 (Biopharma Technology Limited, Winchester, UK). Samples were dispensed as 2  $\mu$ L onto the centre of the FDCS SP 70  $\mu$ m spacer, and a W13G 13 mm diameter glass slide placed on top using a vacuum pen. The spacer did not contact the sample, and sample edge was moved over the microscope aperture via stage manipulators. Samples were analysed throughout a freeze-drying cycle in which the stage was cooled to  $-40^{\circ}\text{C}$  at  $20^{\circ}\text{C}/\text{min}$ , held at  $-40^{\circ}\text{C}$  for 30 minutes, and then heated to  $20^{\circ}\text{C}$  at  $1^{\circ}\text{C}/\text{min}$ . Images were recorded every 5 seconds to follow the morphology change, which was used to indicate  $T_c$  or  $T_{eu}$ .

### 3.2.4 X-ray diffraction (XRD)

X-ray diffraction (XRD) was performed on Rigaku MiniFlex 600 (Ettlingen, Germany). Freeze-dried cake samples from a single well or combined from multiple wells, were loaded onto the centre of the circular sample holders, and the sample surface flattened using a glass slide. Incident X-rays were fired onto the material and diffraction recorded for angles at  $3 - 90^{\circ}$ , at  $5^{\circ}/\text{min}$  with a step size of  $0.02^{\circ}$ . The recorded XRD data was firstly converted to .xrdml format by PowDLL (Kourkoumelis 2013) and then analysed by PANalytical X'Pert Data Viewer (Almelo, The Netherlands).

### 3.2.5 Fab sample preparation

Fab was expressed and purified as described previously (Chakroun et al. 2016). Aliquots were stored at  $-80^{\circ}\text{C}$  until required. Fab samples were thawed at room temperature and dialysed overnight at  $4^{\circ}\text{C}$  against MilliQ water using 20 kDa cut-off dialysis cassettes (Slide-A-Lyzer Dialysis Cassette, Thermo Scientific, UK). Samples were then filtered through Anotop 25  $0.02 \mu\text{m}$  syringe filters (GE Healthcare, Buckinghamshire, UK) to remove any aggregates, and then concentrated to 2 mg/ml with 30 kDa cut-off Vivaspins (Generon Ltd, Berkshire, UK), and stored at  $4^{\circ}\text{C}$  to be used within 1 day.

### **3.2.6 Liquid formulations**

Each final sample for freeze-drying contained 1 mg/ml Fab, 20 mM buffer, pH 4-9, and either NaCl or Na<sub>2</sub>SO<sub>4</sub> to bring the total ionic strength to 50, 100 or 200 mM. To achieve this, 100 µl 2 mg/ml Fab in water was added to 100 µl double-concentrated stock buffer in each well of a microplate. The buffers used were sodium citrate for pH 4, sodium acetate for pH 5, HCl-histidine for pH 6, sodium phosphate for pH 7, HCl-Tris for pH 8 and TAPS for pH 9. For each freeze-drying batch, one mutant was formulated in 36 different conditions (6 pH x 2 salts x 3 ionic strength), in triplicate, for a total of 108 samples. Samples were randomly assigned to the 60 central wells of a 96-well microplate, to minimise variations arising from edge effects and plate locations. Water was placed in the outer wells. Two freeze-drying batch repeats were obtained for each mutant (except K133M due to limited sample) to independently examine the process reproducibility.

### **3.2.7 Size-exclusion chromatography HPLC**

#### **(SEC-HPLC)**

SEC-HPLC was performed based on Section 2.2.8.

### **3.2.8 Freeze-drying**

Freeze drying was conducted on a Virtis Genesis 25EL freeze dryer (SP Scientific, US). Microplates filled with 200 µl samples were loaded onto the freeze-drying shelves. The shelf was initially held at 20°C for 30 min, then ramped to -40°C over 120 minutes and held at -40°C for 60 minutes. Primary drying was initiated by reducing the pressure to 500 mTorr, then holding at -40°C for 60 min with pressure further decreased to 100 mTorr once holding was initialised. The temperature was then ramped to -20°C over 30 minutes and held at -20°C for 600 minutes. For secondary drying, the temperature was ramped to 25°C over 300 minutes with pressure further decreased to 20 mTorr as secondary drying started, and held at 25°C for 60 min before stopping. Samples were

stoppered in the freeze-dryer, and then removed from the chamber for immediate cake morphology scoring.

### 3.2.9 Cake morphology scoring and reconstitution

An eye observation was used to score the morphology of freeze-dried samples in individual wells. Cake morphology was scored against a reference image (Figure 3.2), where a plump cake occupying the well was scored 0.9-1; a cake with edge defects or non-homogeneous covering of the well was scored 0.5-0.8; and a cake that occupied less than half of the well was scored below 0.4, with a score of 0 representing no visible cake. Scores were obtained by two independent practitioners, and averaged to provide a semi-quantitative judgement of relative cake morphology. Freeze-dried solids were then reconstituted with 190  $\mu$ l water, and mixed gently by pipette aspiration 10 to 20 times to fully dissolve the protein.  $OD_{340}$  and  $OD_{600}$  nm measurements in a platereader were used to verify the turbidity, and mixing repeated if high turbidity was found in any wells. Reconstituted samples were held at 4°C and analysed as above by SEC-HPLC within 24 h.

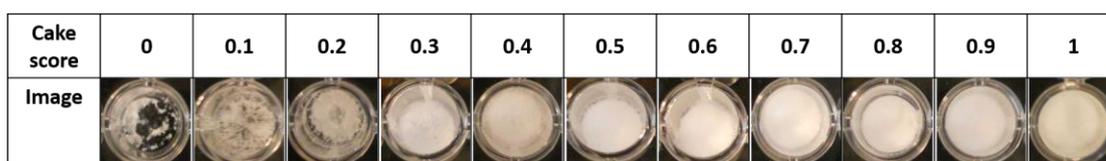


Figure 3.2 The evaluation criteria for cake morphology

### 3.2.10 Net charge calculation

The net charge of each mutant (in the absence of salts), was calculated by loading PDB files into Propka (Dolinsky et al. 2004).

### **3.2.11 Sweet plot**

The sweet plot was generated by Design-Expert 8.0.6 from the monomer loss and cake score response surfaces. The software automatically searched for the best fitting models that resulted in lowest p-value, and suggested models were then used to predict the sweet zones that met the inputted user requirement.

## **3.3 Results and discussion**

### **3.3.1 Homogeneity of drying across the freeze-dryer shelf**

A homogeneous removal of water from samples is a pre-requisite for the quantitative comparison of formulation stabilities in a freeze-drying process cycle. The freeze dryer is specified to provide homogeneous cooling to within 1°C difference across the shelf. However, due to the structure of a freeze dryer, certain areas of the shelf can experience more or less efficient drying rates due to the impact of the condenser, vacuum pump and the chamber shell. Edge effects, in which samples sitting on the edge experience a higher drying rate than those in the centre, is fairly common for freeze drying, and needs to be minimised. The water vapour pressure above those samples at the edge is lower than that in the centre, leading to a higher mass-transfer rate for samples at the edge. Moreover, the drying rate also varies for different container geometries and materials within. As a result, any factor that can induce heterogeneity should be minimised.

The aim of the homogeneity test was to map water removal during freeze-drying in different areas of the shelf, to determine the extent of edge effects when using 96-well microplates, and to trial the use of aluminium plates for establishing thermal contact between wells and the shelf. A previous pioneering study facilitated the use of 96-well microplates for freeze-drying, with hundreds of samples assessable in one process cycle (Grant et al. 2009). That work also revealed a decreased drying rate from the central wells within a single 96-well microplate, compared to those on the outer edges. To

achieve relatively homogeneous water drying, only the central 60 wells were used for samples, while the outer-edge wells were filled with water. As an alternative approach we tested the placement of water-filled microplates around the sample-containing plates, to provide necessary water vapour pressure. In the previous study, the bottom lips of microplates were trimmed off so that the well bottoms were in direct contact with the freeze-drier shelf. To avoid trimming every microplate, a simpler approach would be to place an aluminium plate in the air gap between the well bottoms and the shelf.

Figure 3.1 shows the layout of the homogeneity test using 96-well microplates. The centre region of the shelf was chosen to load plates because it could offer relatively homogeneous drying compared to the shelf edge. Trimmed plates filled with water were placed on the left of the four sample plates to investigate whether this setting could minimise edge effects during freeze drying of 96-well microplates. Figure 3.3 shows the mass of water remaining in each well of the four microplates, after partial drying in the homogeneity test. As can be seen, edge effects within each microplate were observed with higher drying rates in the outer wells of all four microplates, compared to those in the centre. The edge wells on the left side, adjacent to the water-filled trimmed plates, had a slightly lower drying rate than those on the right.

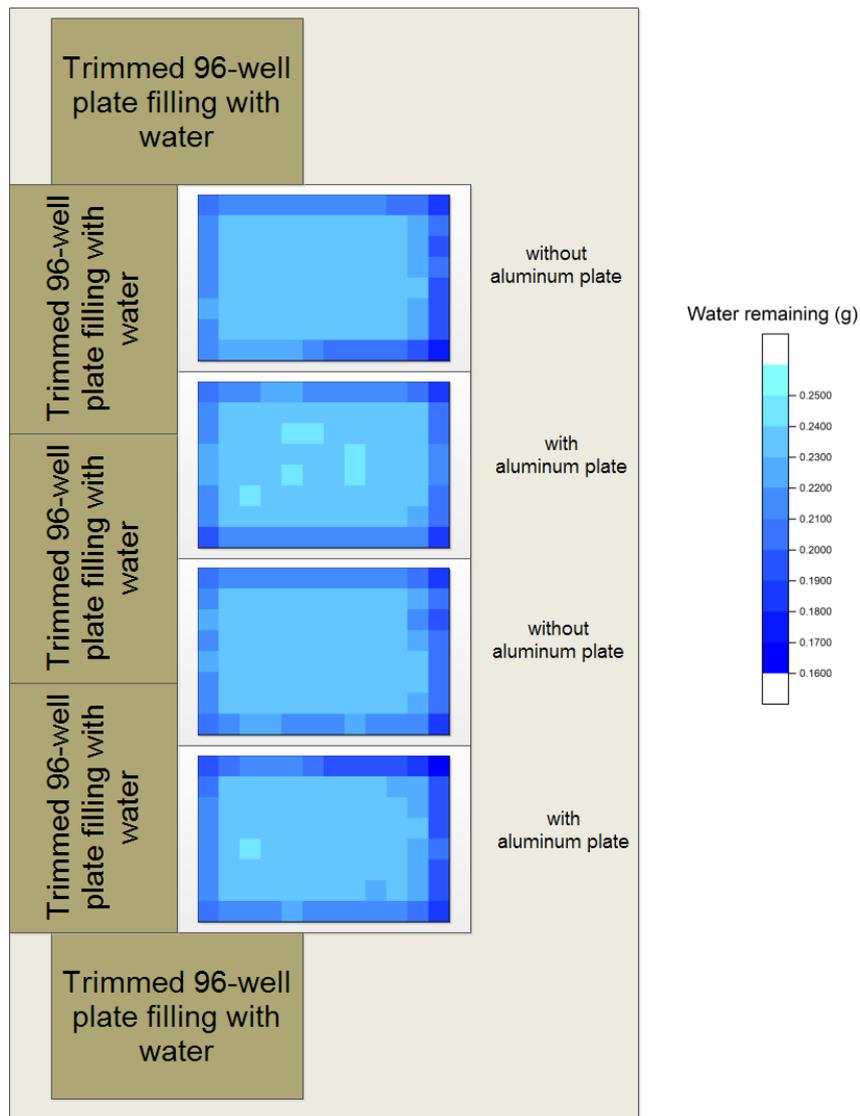


Figure 3.3 The mapping of water remaining in the four 96-well plates for the homogeneity test of shelf

The layout details were shown in the Figure 3.1 legend. A gradient of light blue to dark blue was used to indicate the remaining water from more than 0.25 g to 0.16 g of each well after the homogeneity test.

Very little significant difference in drying rates was observed between the drying of microplates with and without aluminium plates underneath. This suggests that the aluminium plate does not improve the heat-flow considerably between the shelf and well bottoms as expected, and may even have introduced some slight heterogeneity as shown in the second and fourth plates.

The final setting in Figure 3.4 was selected to achieve homogeneous drying for samples in the central 60 wells highlighted, with water in the outer wells, for each of the

four central 96-well microplates. The surrounding trimmed plates are also filled with water. Aluminium plates were not used beneath the plates. Using SEC-HPLC for the characterisation of freeze-dried solutions with a 5 minute cycle time, a total of 240 samples required 20 hours to process.

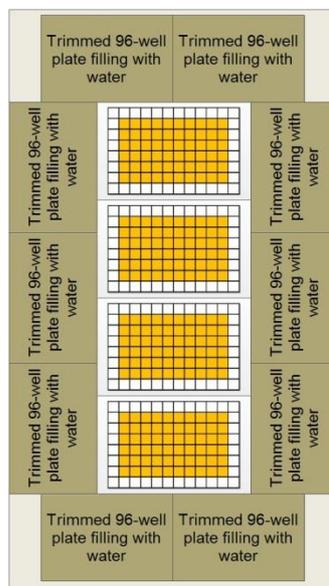


Figure 3.4 A proposed layout for freeze-drying in 96-well microplates

### 3.3.2 DSC and FDM to validate the freeze-drying cycle parameters

#### 3.3.2.1 DSC to measure the glass transition temperature

The glass transition temperature ( $T_g$ ) is a key characteristic for a freeze-drying formulation. The freezing temperature should ideally be lower than the  $T_g$ , to ensure that the frozen sample is vitrified into an amorphous state. Modulated-temperature differential scanning calorimetry (DSC) enables the deconvolution of reversing and non-reversing signal (Gill et al. 2010), to allow a more accurate determination of  $T_g$  from the reverse heat-flow component (Kett 2001). Figure 3.5 shows the DSC profile of Fab in pure water for the reverse heat-flow (green line). A  $T_g$  could not be detected ascribed to the low concentration and low protein molecular weight of Fab, with no other thermal events observed such as endothermic relaxation. The inflection point ( $-4.58^\circ\text{C}$ ) was the

initial melting of ice. The profile for Fab at pH 4, with NaCl added to an ionic strength of 200 mM, was dominated by the eutectic crystallisation of NaCl at above  $-24^{\circ}\text{C}$ , still the  $T_g$  of Fab could not be clearly identified (Figure 3.5, blue line). Therefore, FDM was used to characterise the  $T_c$  as the DSC could not detect the  $T_g$  due to the low concentration of Fab.

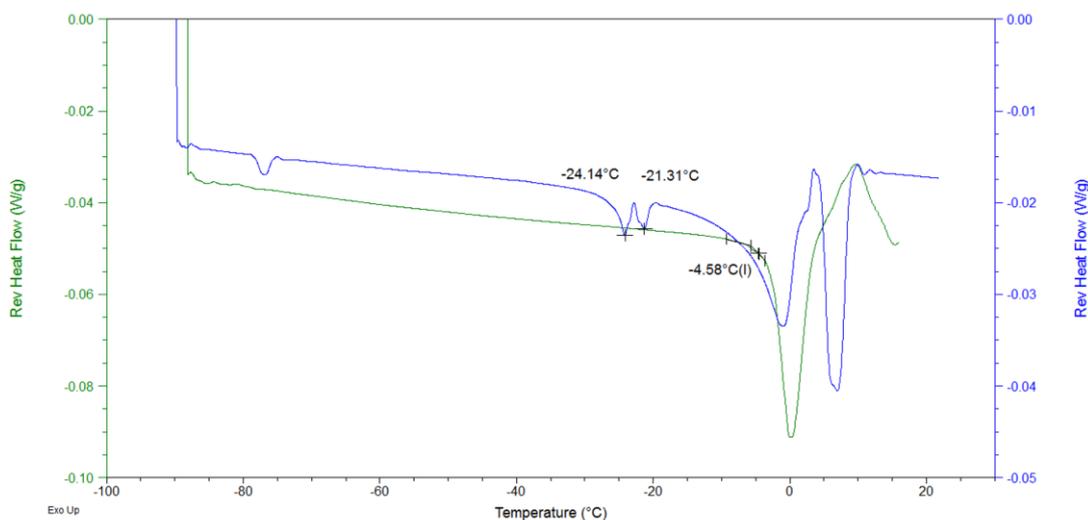


Figure 3.5 The DSC profile for Fab.

Green line: Fab at 2 mg/ml in water; blue line: 1 mg/ml Fab at pH 4 with NaCl to make the total ionic strength at 200 mM

### 3.3.2.2 FDM to measure the collapse temperature

FDM was used to determine the critical collapse temperature,  $T_c$ , which directly represents the collapse phenomenon under vacuum conditions (Meister & Gieseler 2008). It has been observed that collapse during freeze-drying had no effect on aggregation or the integrity of freeze-dried proteins in a saccharide system (Sarciaux et al. 1999; Taylor et al. 2010c), which might be different from the salt system applied in this study. Therefore, we selected our highest salt condition at pH 7 for  $T_c$  measurement by FDM, as a worst-case formulation.

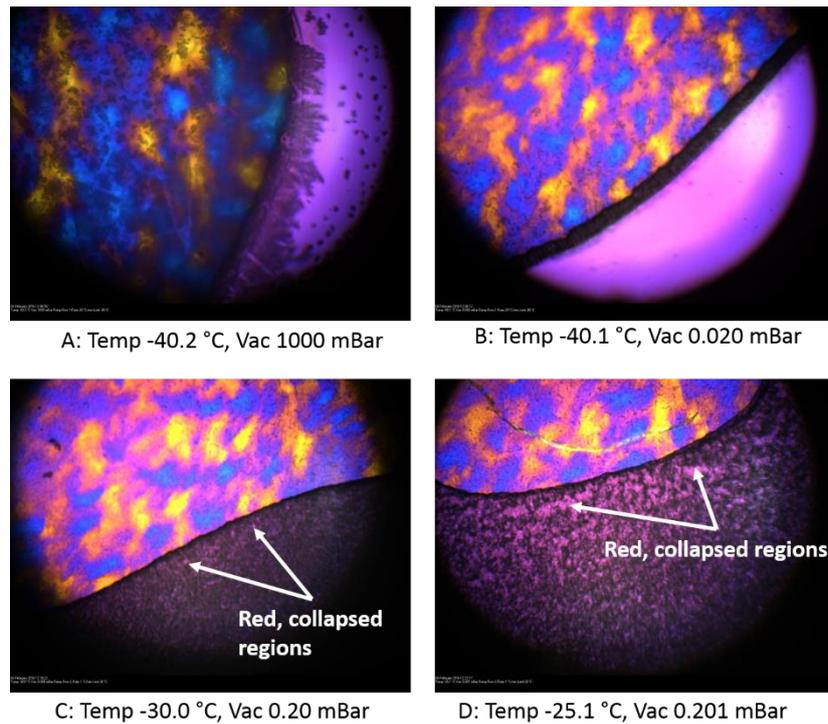


Figure 3.6 Freeze-drying microscopy for 1 mg/ml Fab at pH 7, and NaCl to 200 mM ionic strength

As shown in Figure 3.6, the vacuum was applied at  $-40^{\circ}\text{C}$  and the drying frontline, observed as a black thin line, migrated to the centre of the sample as the temperature was increased progressively, and as the ice sublimed. No structural change was observed within the drying frontline until around  $-30^{\circ}\text{C}$ , where some pink regions appeared behind the sublimation front in the dried region. These indicated degeneration in the dried structure, and led afterwards to complete structure loss near the drying front. Therefore, the  $T_c$  was close to  $-30^{\circ}\text{C}$ , and so carrying out the primary drying at  $-40^{\circ}\text{C}$  was sufficiently low to prevent the cake structure from collapsing in most formulations.

Having established a microscale freeze-drying process in which the well-to-well variability was not significant, where the freezing and primary drying temperatures were well below the  $T_c$ , we then investigated the influence of pH, ionic strength, salt type, and protein surface charge mutations, on the retention of the protein monomer, and the cake morphology.

### **3.3.3 Formulation and mutant effects on monomer loss**

The scope of this research focused on the effect of buffer conditions and mutations upon freeze-drying. Therefore, the conventional preservation by disaccharide was not applied. The monomer loss for each mutant during freeze-drying is shown in Figure 3.7, where it can be seen that the freeze-drying process led to 10-30% monomer loss in most cases. The data varied greatly for different pH, mutant types, salt types and ionic strengths. Despite the clear general trends arising from formulation factors, we observed batch-to-batch variations, and hence we averaged the batch data together to determine the underlying trends in the following discussion.

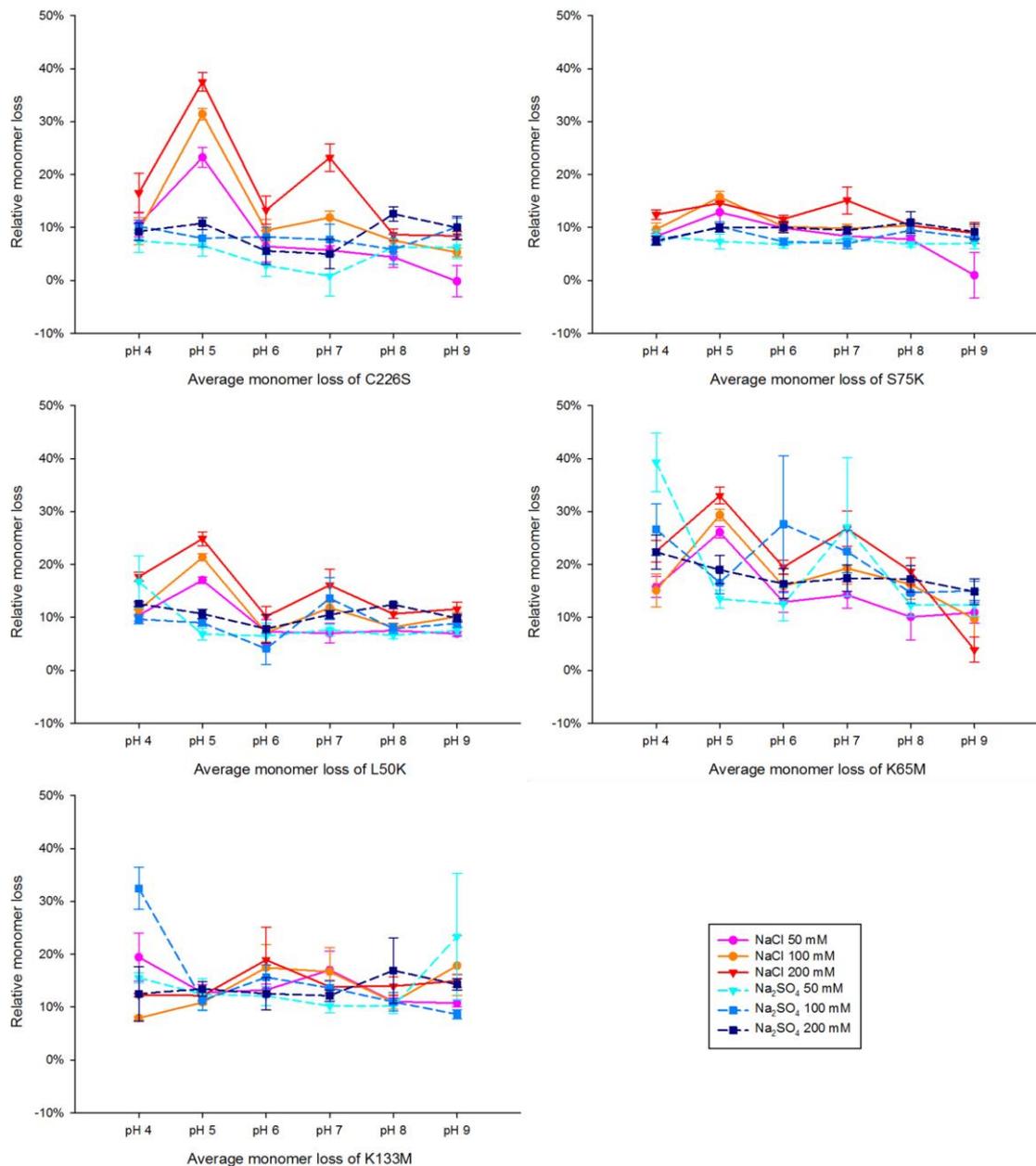


Figure 3.7 The monomer loss of Fab mutants determined by SEC-HPLC

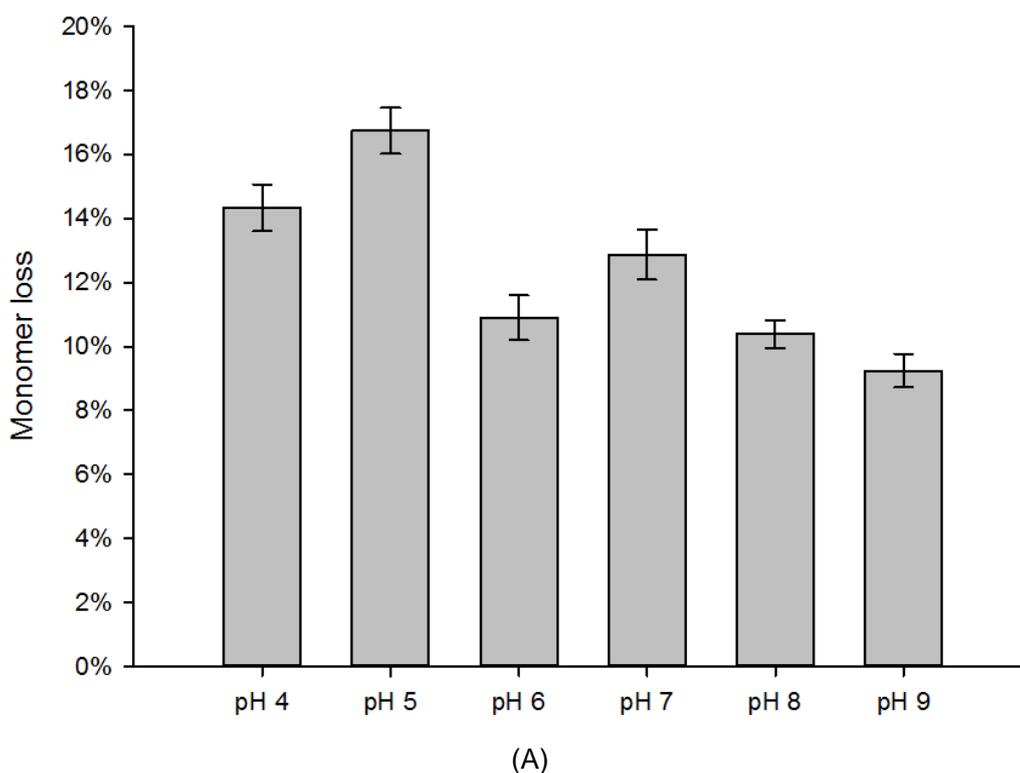
Data were averaged from two freeze-drying batches, except K133M with one batch. Triplicates were used within each batch. Error bars are Standard Error of the Mean (SEM). The Y-axis scales are matched for ease of cross comparison. One data point of K133M at pH 8, Na<sub>2</sub>SO<sub>4</sub>, 100 mM was ignored where the cake floated out from the well.

### 3.3.3.1 Influence of pH on monomer loss

The influence of pH on monomer loss is shown in Figure 3.8, averaged from the 9 freeze-drying batches carried out (two repeats for each of the five mutants, except K133M). The monomer loss of all the mutants in general decreased as the pH was

increased from 4 to 9. The trend was mostly monotonic, except for a spike at pH 5, which gave an unexpectedly higher monomer loss, and a dip at pH 6, such that interestingly, pH 6, 8 and 9 were more favourable than the physiological pH 7.

The *t*-test (Figure 3.8B) showed that monomer loss at pH 4 was significantly higher than any other higher pH except pH 7, and significantly lower than at pH 5; monomer loss at pH 5 was significantly higher than any other pH conditions; monomer loss at pH 7 was significantly higher than Ph 8 and 9. Formulation at pH 6 achieved comparable monomer loss with respect to pH 7, 8 and 9; while monomer loss was also comparable between pH 8 and 9.



	pH 4	pH 5	pH 6	pH 7	pH 8	pH 9
pH 4	N/A	0.019	8E-04	0.174	5E-06	3E-08
pH 5	0.019	N/A	2E-08	3E-04	6E-13	1E-15
pH 6	8E-04	2E-08	N/A	0.064	0.533	0.059
pH 7	0.174	3E-04	0.064	N/A	0.006	2E-04
pH 8	5E-06	6E-13	0.533	0.006	N/A	0.091
pH 9	3E-08	1E-15	0.059	2E-04	0.091	N/A

(B)

Figure 3.8 Average monomer loss for formulations with different pH

(A) Each data point was averaged from all the 3 ionic strengths, 2 salt types and repeats of in total 9 batches (n=162) for the 5 mutants. The error bars were standard error of the mean (SEM).

(B) The p-values of two-sample *t*-test assuming unequal variances for the monomer loss in (A). The p-values less than 0.05 were highlighted in green.

All mutants and also C226S tended to aggregate the most upon freeze-drying at low pH. While the increased net positive charge at low pH can exert greater electrostatic repulsion between protein molecules, the increased ionic strength can conversely lessen intermolecular long-range electrostatic repulsions, allowing hydrophobic interactions to dominate. The same conditions also lead to decreased conformational stability of A33 Fab (Chakroun et al. 2016), due to increased short-range intramolecular electrostatic repulsion, and also the weakening of salt bridges. The combined effect would be a greater population of both partially and globally unfolded protein states, and faster aggregation from those states during freeze-drying.

It is interesting that pH 7 was not the optimal pH condition for the freeze-drying of Fab, and that the pH-dependence was not simply monotonic. It is well known that the pH in the solution phase during freeze-drying can undergo dramatic changes due to the freeze-concentration effect as liquid water freezes, in a manner that depends on the buffer species and also the ionic strength. In addition, some buffer species can become frozen prior to the others. For example, Na<sub>2</sub>HPO<sub>4</sub> can crystallise prior to the co-crystallisation of NaH<sub>2</sub>PO<sub>4</sub> and Na<sub>2</sub>HPO<sub>4</sub> if the initial pH is 7, which induces a shift to lower pH (Franks & Auffret 2007). Therefore, the lower performance at pH 7 for Fab

freeze-drying may also indicate that the phosphate buffer did not maintain pH 7 during freezing, whereas Tris and TAPS buffers are not greatly affected by freezing.

The stabilising effect at pH 8 and 9 was potentially also caused by attractive electrostatic interactions at close to the pI of 9, which can promote precipitation in the native folded state. The protein was freeze-concentrated during freezing, and its solubility dropped at subzero temperature, which caused it to precipitate when its concentration was over-saturated. Such precipitation may be reversible and hence serve to protect Fab from aggregation or unfolding during the freezing and drying processes. Owing to the precipitation, the protein is less likely to interact with Cl<sup>-</sup> or SO<sub>4</sub><sup>2-</sup> at higher pH, and this would also potentially result in less conformational change.

### *3.3.3.2 Influence of pH and mutants on monomer loss*

Figure 3.9 shows the pH impact on individual mutants for the freeze-drying monomer loss. The rank order of the mutants was broadly similar from pH 6-9, and also at pH 4, but changed considerably at pH 5, at which condition, the salting in and destabilising effect of acetate was likely to have had an adverse impact. We separately compared citrate and acetate buffers at pH 4 and found that acetate led to considerably more monomer loss than citrate (Section 3.3.3.3). Pseudo wild-type C226S and the two positively modified mutants, S75K and L50K, each witnessed a spike at pH 5, while two negatively modified mutants, K65M and K133M, showed a decrease in monomer loss at pH 5 compared to pH 4. In general, except for K65M, pH 6 to 9 provided similar degradation stress in freeze-drying, where the monomer loss varied within 3% for each mutant. The K65M mutant was most sensitive to pH change, and dipped 5% at pH 6, rose 4% at pH 7, and decreased again by 7% and 3% at pH 8 and 9 respectively, and influenced most of the overall trend in Figure 3.8, as it also resulted in the highest monomer loss.

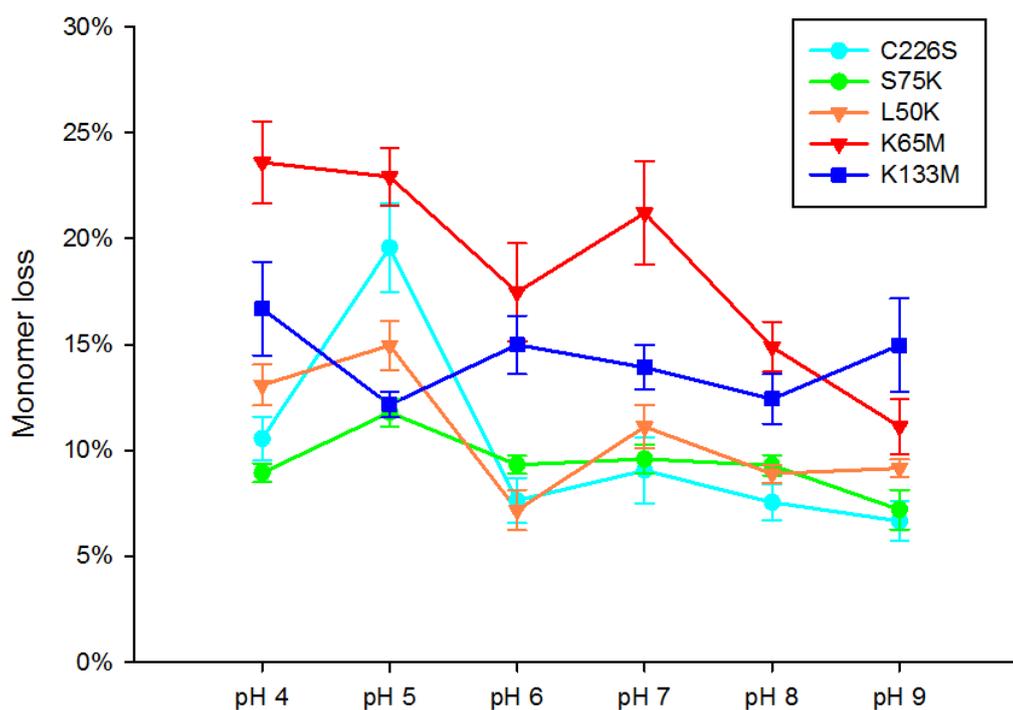


Figure 3.9 The impact of pH for each mutant upon monomer loss

Each data point was averaged from all the three ionic strengths, two salt types and repeats of two batches (n=36). The error bars were standard error of the mean (SEM).

The sensitivity of the mutants to pH, during freeze-drying was greatest at pH 5 overall. The average monomer loss (across all pH) followed K65M > K133M > C226S > L50K/S75K. This sensitivity difference was caused probably by mutants' surface charge. S75K and L50K increased the net charge by one, which made them generally more colloidally stable, and less sensitive to positive hydrogen ions at low pH compared to C226S. K65M and K133M decreased the net charge by one so they were less colloidally stable and caused more monomer loss almost across all the conditions. The charge decreases on K65M and K133M also results in increased hydrophobicity, which made them less sensitive to pH change. The significant drop for K133M at pH 5 is very interesting, as it was the pH opposite to all other mutants. This may have resulted from an increased rate of protective precipitation with this mutant of higher surface hydrophobicity, when the salting-in destabilising effect by acetate buffer was reduced.

### 3.3.3.3 Influence of acetate and citrate pH buffer on monomer loss

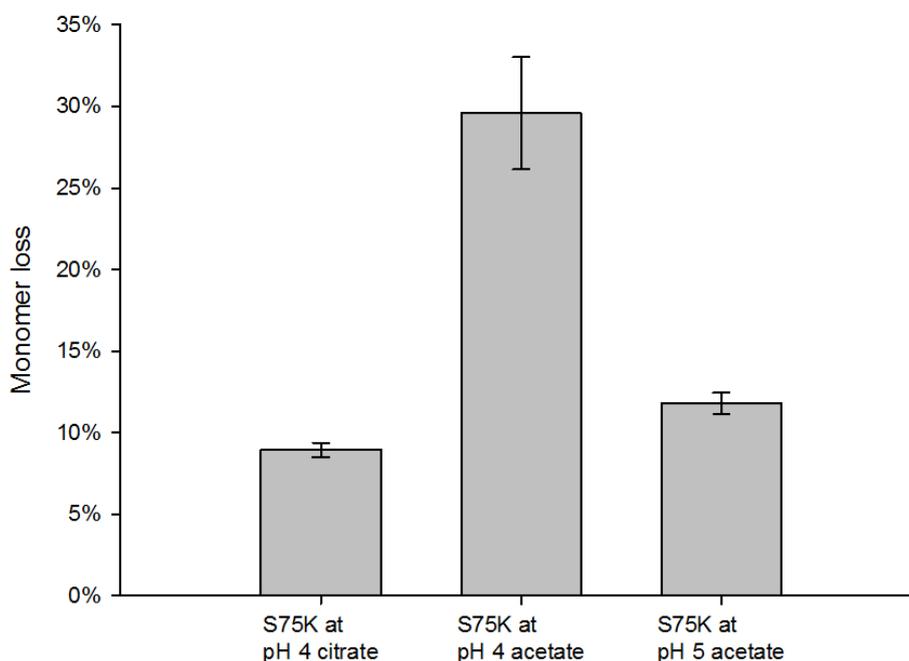


Figure 3.10 Average monomer loss for S75K formulations with/without acetate at pH 4 and 5. Error bar: standard error of the mean.

A stable pH and proper ionic strength are the key elements that affects the protein aggregation in freeze-drying. A good selection of buffer type and salt concentration plays an important role to stabilise protein in a native state. Figure 3.10 shows that the use of acetate buffer at pH 5 resulted in more monomer loss than those with citrate buffer at pH 4. This suggests that even though pH 4 is more extreme for protein, probably the salting-in effect of acetate destabilised proteins. This may also be attributed to sodium acetate's low  $T_g$  ( $-80^{\circ}\text{C}$ ) that would cause collapse (Chang & Randall 1992). In order to make sure acetate buffer is detrimental to Fab stability, the most stable mutant S75K was formulated at pH 4 with acetate buffer and went through freeze drying. As shown in Figure 3.10, it turned out that when S75K was used, pH 4 acetate condition resulted in three times the monomer loss compared to pH 4 citrate and pH 5 acetate. Therefore, it is very obvious that the acetate buffer is not a good option for freeze-drying.

### 3.3.3.4 Influence of salt types and ionic strength to monomer loss

The influence of salt types and ionic strength (IS) on monomer loss is shown in Figure 3.11, averaged from the 9 freeze-drying batches. The monomer loss was comparable between formulation with NaCl and Na<sub>2</sub>SO<sub>4</sub> at 50 and 100 mM IS. At 200 mM IS, samples containing NaCl had significantly 5% more monomer loss than samples containing Na<sub>2</sub>SO<sub>4</sub>.

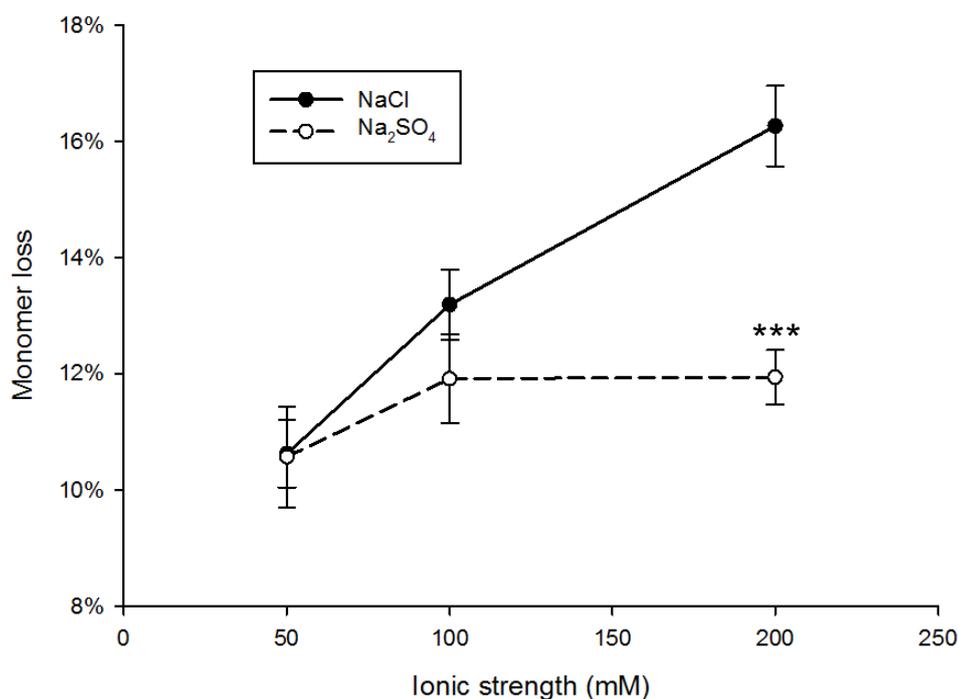


Figure 3.11 Average monomer loss for formulations at different salt types and levels of ionic strength

Each data point was averaged from all the 6 pH and 3 repeats in 9 batches for the 5 mutants (n=162). Error bars were standard error of the mean. Two-sample *t*-test assuming unequal variances were performed between formulation with Na<sub>2</sub>SO<sub>4</sub> and NaCl at 50, 100 and 200 mM ionic strength, respectively (\*\*\*)  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ).

The Hofmeister series ranks the ability of anions and cations to salt-in and salt-out protein (Baldwin 1996). For anions, Cl<sup>-</sup> is of higher ion chaotropicity than SO<sub>4</sub><sup>2-</sup>, and exerts more salting-in effect, in which anions preferably interact to polar protein residues, increase the protein solubility and change its conformation (Roberts et al. 2015). Anions with lower chaotropicity (higher kosmotropicity) salt-out more strongly, where preferable

hydration of anions induces the dehydration of protein, increased protein-protein attraction, and ultimately precipitation.

The effect of the Hofmeister series, and particularly NaCl and Na<sub>2</sub>SO<sub>4</sub>, on freeze drying has been studied previously. It was noted that kosmotropic salts could preserve enzyme activity through preferential hydration (Ru et al. 2000), while freeze-drying with NaCl resulted in remarkable aggregation, oxidation and deamidation (Pikal et al. 1991; Sarciaux et al. 1999). It was also found (Izutsu & Kojima 2000) that NaCl lowered  $T'_g$  of protein-nonionic polymers mixture 5 to 20 °C more than Na<sub>2</sub>SO<sub>4</sub> when salt molarity ranged from 50 to 200 mM. It was suggested that the salting-in NaCl largely removed protein from the polymers' protection and formed polymer-rich and protein-rich phase separately; the salting-out Na<sub>2</sub>SO<sub>4</sub>, however, was separated from the protein-polymer mixture, remained the mixture in a single phase and did not alter its  $T'_g$ .

As shown in Figure 3.11, formulations with Na<sub>2</sub>SO<sub>4</sub> performed better than their NaCl counterparts at 200 mM ionic strengths. Increasing ionic strength resulted in greater monomer loss, with an average monomer loss of 10.63%, 13.20% and 16.27% for 50, 100 and 200 mM NaCl respectively. By contrast, Na<sub>2</sub>SO<sub>4</sub> led to an initial increase in monomer loss from 10.57% to 11.92% for 50 and 100 mM, respectively, but then did not increase further at 200 mM Na<sub>2</sub>SO<sub>4</sub>.

Protein-protein and protein-salt interactions in the aqueous phase have been studied extensively (Roberts et al. 2015; Bye & Falconer 2014; Roberts et al. 2014). The difference between NaCl and Na<sub>2</sub>SO<sub>4</sub> at 50 and 100 mM IS was comparable, as it is generally thought that favourable hydration of protein molecules occurs at low ionic strength, where the net charge of a protein dominates the overall repulsive protein-protein interaction (Lehermayr & Mahler 2011). Therefore, comparable monomer loss was detected at low IS. By contrast, the monomer loss difference was significant at 200 mM IS, as the hydration layer is removed from proteins as ions compete for water at higher ionic strengths (Bye & Falconer 2014), while the repulsive electrostatic interaction between proteins becomes more shielded. During freezing, the solute concentration rapidly increases as liquid water forms ice. For Na<sub>2</sub>SO<sub>4</sub> samples, the SO<sub>4</sub><sup>2-</sup> ions would

exclude Fab from its hydration shell and promote protein precipitation. This salting-out effect could effectively maintain the native state of Fab if precipitation occurs rapidly. By contrast, Cl<sup>-</sup> ions would more likely bind the Fab surface, and neutralise the protein net positive charge, but also destabilize and partially unfold the protein by weakening salt-bridges, that leads to unfolding and also aggregation as the Fab becomes concentrated in the unfrozen phase. During reconstitution, most of the Fab precipitated in the Na<sub>2</sub>SO<sub>4</sub> sample would dissolve back to a native state, while those in NaCl may misfold or remain aggregated.

### 3.3.3.5 Influence of salt types, ionic strength and mutants to monomer loss

Figure 3.12 analyses the influence of salt for individual mutants. Fab mutants all responded more to an increase in ionic strength with NaCl than with Na<sub>2</sub>SO<sub>4</sub>, except K133M which was insensitive to ionic strength with NaCl. The response of monomer loss across 50-200 mM ionic strength with NaCl was similar for all mutants (except K133M). With Na<sub>2</sub>SO<sub>4</sub>, the monomer loss for pseudo wild-type C226S increased initially with IS from 50 mM to 100 mM, and then plateaued to 200 mM. K65M and K133M also potentially had minor increases with IS from 50 mM to 100 mM, and dropped to a comparable level at 200 mM compared to 50 mM. By contrast, the two mutants with one additional positive charge, S75K and L50K, were each relatively insensitive to IS from 50-200 mM.

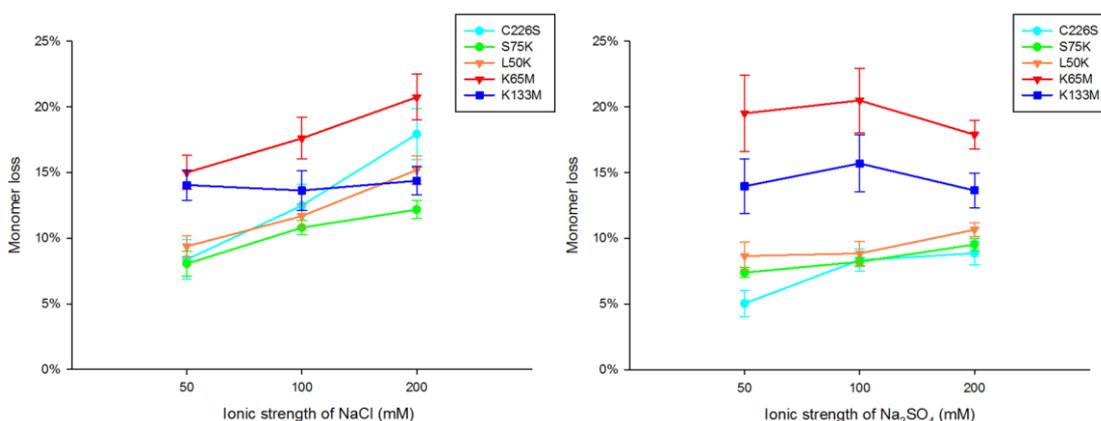


Figure 3.12 Average monomer loss for formulations at different salt types, levels of ionic strength and mutants

Error bars are standard error of the mean. Each data point was averaged from all the 6 pH and repeats of 2 batches for each mutant (n=36, except K133M with one batch and n=18).

Ionic strength clearly differentiated the monomer loss of mutants based on their charge at 50-100 mM, more than it did at 200 mM. This is consistent with the net charge on the proteins having an influence on colloidal stability that becomes more effectively weakened at higher ionic strength, particularly as freeze concentration raises the local ionic strength. At 50 mM in particular, pseudo wild-type C226S, the mutants S75K and L50K with one additional positive charge, were more colloiddally stable than those with one less positive charge (K65M and K133M). The insensitivity of K133M to ionic strength in NaCl appears to be derived from an increase in monomer loss at low ionic strength, given that K133M was more stable than K65M under all other conditions. This indicates that the K133M variant was more sensitive to unfolding at lower NaCl concentration, possibly due to removal of a potential salt bridge interaction that it forms with E216 in the heavy chain. This further suggests that this salt bridge interaction is particularly critical to stability of the Fab.

### *3.3.3.6 XRD to examine the crystallinity of the sample*

As formulation with Na<sub>2</sub>SO<sub>4</sub> generally resulted in decreased monomer loss and a higher cake morphology score than with NaCl at the same ionic strength, from 50 mM to 200 mM, X-ray diffraction (XRD) was used to measure the crystallinity of freeze-dried samples, so as to determine the influence of crystallinity in freeze-drying. It has been found previously that the crystal pattern depends on material composition, drying process, storage condition, and relative humidity (Haque & Roos 2005). To avoid significant interference by signal and noise from the sample holder, samples from a total of 12 replicate wells were combined onto the centre of the sample holder. This was sufficient to obtain an XRD pattern of predominantly the freeze-dried material.

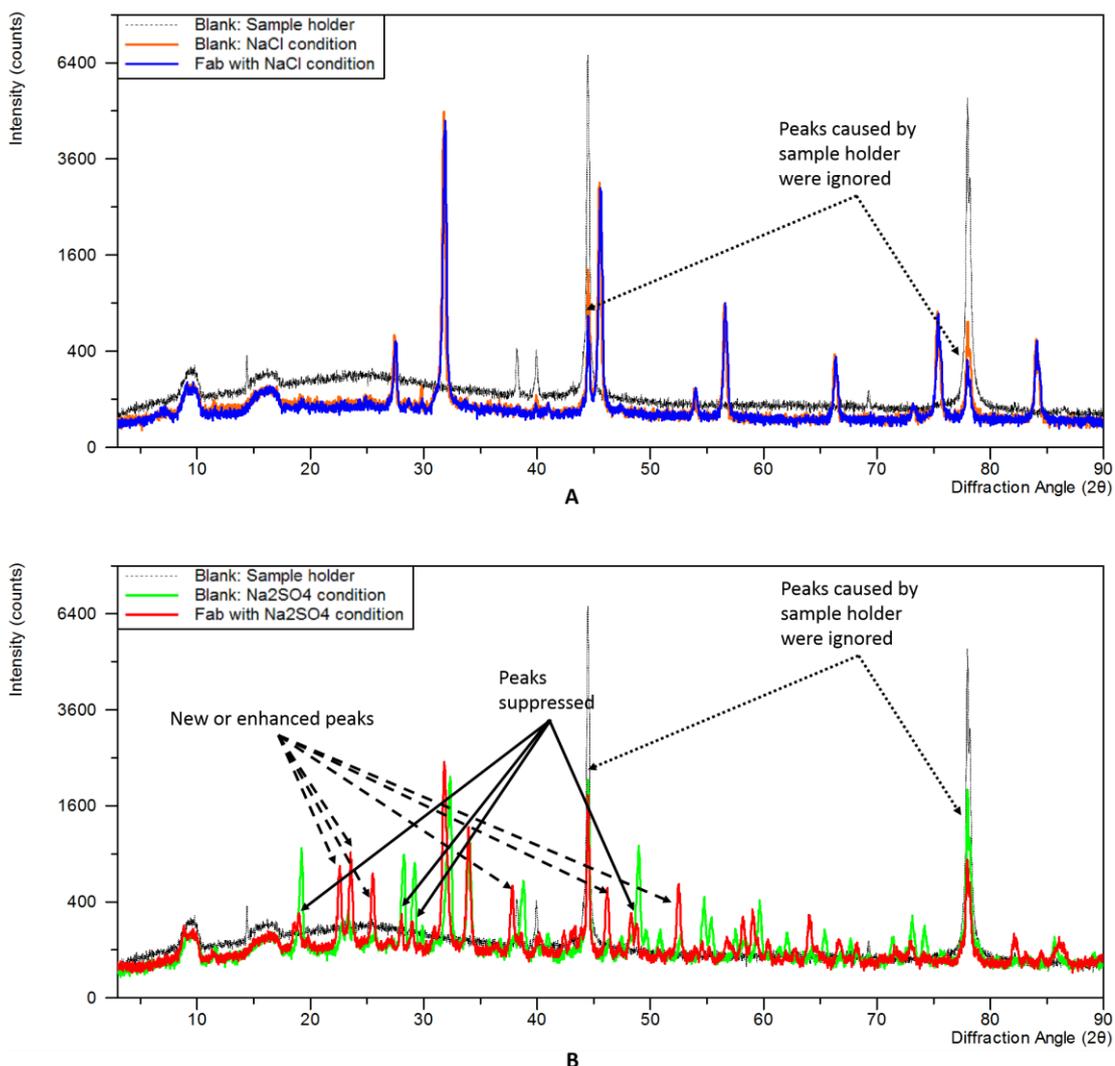


Figure 3.13 The X-ray diffraction patterns for NaCl and Na<sub>2</sub>SO<sub>4</sub> conditions with and without Fab. The signal from a blank sample holder is shown in dash, thin, black line for both Figure A and Figure B. All the samples were combined from 12 wells of freeze-dried cake in identical conditions. Samples were in pH 5, 20 mM acetate buffer, and either NaCl or Na<sub>2</sub>SO<sub>4</sub> to bring the total ionic strength to 200 mM, with or without 1 mg/ml C226S Fab. In Figure A, NaCl condition with no Fab is coloured in orange; NaCl condition with Fab is coloured in blue. In Figure B, Na<sub>2</sub>SO<sub>4</sub> condition with no Fab is coloured in green; Na<sub>2</sub>SO<sub>4</sub> condition with Fab is coloured in red.

The samples freeze-dried at pH 5, 200 mM were analysed by XRD as Fab stability was most significantly different between the Na<sub>2</sub>SO<sub>4</sub> and NaCl formulations in these conditions (Figure 3.7). Figure 3.13 shows the XRD intensities for a blank sample holder, and also NaCl and Na<sub>2</sub>SO<sub>4</sub> samples, both with and without Fab. The two peaks that appeared at diffraction angles of 44.5° and 78.0° with a blank sample holder were ignored when analysing the freeze-dried samples. The NaCl formulations gave higher crystallinity at diffraction angles larger than 50°, whereas Na<sub>2</sub>SO<sub>4</sub> samples resulted in

higher crystallinity at diffraction angles smaller than 50°. The addition of Fab in the NaCl formulation partially suppressed three of the peak intensities, with no peak enhancements, or new peaks formed. By contrast, with Na<sub>2</sub>SO<sub>4</sub>, the presence of Fab resulted in new or enhanced peaks at 22°, 23°, 26°, 38°, 46° and 52°, and suppression of peaks at 19°, 28°, 29° and 48°.

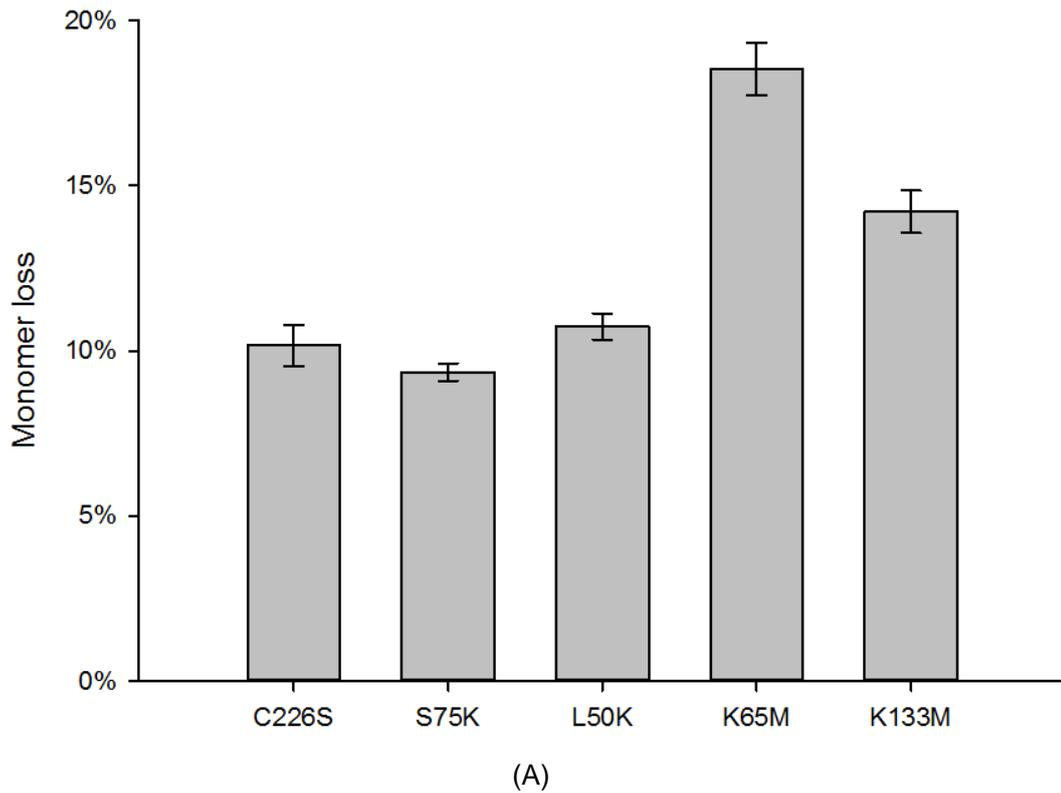
These results suggest that a slightly more amorphous phase, or proportion of the material, was achieved in the presence of Fab, for the NaCl formulation. By contrast, the presence of Fab in the Na<sub>2</sub>SO<sub>4</sub> formulation led to a slightly altered but still crystalline state. In general, the differences caused by Fab were small and the XRD patterns indicated that the freeze-drying with Fab retained most of the crystallinity attributable to the salts. Figure 3.7 shows that the C226S Fab with NaCl resulted in nearly 30% greater monomer loss than with Na<sub>2</sub>SO<sub>4</sub> at pH 5 acetate, 200 mM ionic strength. While previous studies found that proteins in an amorphous state correlated to stabilisation by excipients that could interact with them (Section 1.3.2.1) (Izutsu & Kojima 2002), the present study indicated the opposite effect. Given the small differences in crystallinity, it is unlikely that relative effects of NaCl and Na<sub>2</sub>SO<sub>4</sub> on monomer loss are due to their impact on crystallinity.

### *3.3.3.7 Molecular analysis for the influence of mutations on monomer loss*

Mutational charge is linked strongly to aggregation and related amyloid diseases (Chiti, Calamai, et al. 2002; Calloni et al. 2005), with increased aggregation when the protein net charge approaches its isoelectric point (pI) (Chiti, Calamai, et al. 2002; Chi, Krishnan, Kendrick, et al. 2003).

Table 3.1 The net charge of mutants at different pH as calculated by PropKa

<b>pH</b>	<b>C226S</b>	<b>S75K</b>	<b>L50K</b>	<b>K65M</b>	<b>K133M</b>
<b>4</b>	26.68	27.83	27.77	25.62	25.79
<b>5</b>	16.43	17.47	17.65	15.28	15.54
<b>6</b>	11.1	12.1	12.2	10	10.12
<b>7</b>	8.46	9.49	9.35	7.33	7.46
<b>8</b>	6.84	7.85	7.81	5.81	5.85



	<b>C226S</b>	<b>S75K</b>	<b>L50K</b>	<b>K65M</b>	<b>K133M</b>
<b>C226S</b>	N/A	0.225	0.443	2E-15	8E-06
<b>S75K</b>	0.225	N/A	0.004	5E-23	9E-11
<b>L50K</b>	0.443	0.004	N/A	1E-16	7E-06
<b>K65M</b>	2E-15	5E-23	1E-16	N/A	3E-05
<b>K133M</b>	8E-06	9E-11	7E-06	3E-05	N/A

(B)

Figure 3.14 Average monomer loss for the different mutants

(A) Each data point was averaged from all the three ionic strengths, two salt types, 6 pH and repeats of two batches (n=216). Error bars were standard error of the mean.

(B) The p-values of two-sample *t*-test assuming unequal variances for the monomer loss in (A). The p-values less than 0.05 were highlighted in green.

The net charge of each mutant (in the absence of salts) are listed in Table 3.1. The average monomer loss of the mutants during freeze-drying were ranked in the order S75K < C226S < L50K < K133M < K65M as shown in Figure 3.14A, although the differences between S75K, C226S, L50K were statistically insignificant (Figure 3.14B). The average monomer loss increased from 9-11% for S75K, C226S, and L50K, and then to 14% for K133M and 18.5% for K65M. This shows that while adding one positive charge (L50K and S75K) did not significantly stabilise C226S colloiddally, the removal of one positive charge (K133M and K65M) was colloiddally destabilising. This suggests that although C226S already possessed sufficient net charge to provide colloiddal stability, this colloiddal stability was easily lost. As the pI of C226S Fab is calculated to be between 8.41 (Chakroun et al. 2016) and 9.44 (Ahmad 2011), Fab was positively charged across essentially all of the pH conditions tested. Removal of just one positive charge would significantly increase the number of conditions at which the net charge is close to zero. By contrast, the addition of one more positive charge has less potential for impact where the net charge is already more than 10 at pH 4-6 as calculated from PropKa.

Mutational charge has been recognised as one of the reasons that contribute to aggregation and related amyloid diseases (Chiti, Calamai, et al. 2002; Calloni et al. 2005). Protein tends to associate with each other when they do not possess net charge as an electrostatic repulsion for long-range interaction. It has been found that aggregation is favoured when the protein net charge approaches its isoelectric point (pI) (Chiti, Calamai, et al. 2002; Chi, Krishnan, Kendrick, et al. 2003). Studies also found the removal of charged residues would result in destabilisation (Sheinerman et al. 2000). In this work, one positive charge is removed on the surface of K65M and K133M. Therefore, compared to the wild type C226S, the instability results of K65M and K133M is in accordance with previous work as their net charge is one unit closer to neutrality when the working pH condition is below their pI.

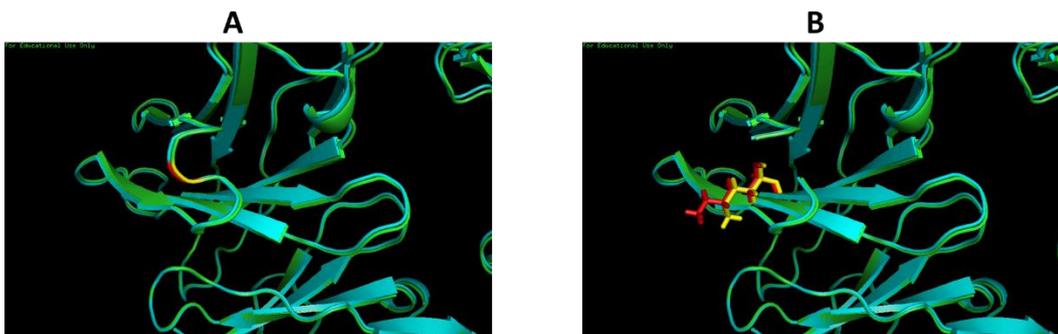


Figure 3.15 The overlay of the predicted structure before and after the K65M mutation. Mutation is shown in cartoon in Picture A and stick in Picture B. C226S is coloured in cyan with residue 65 in red; K65M is coloured in green with residue 65 in yellow.

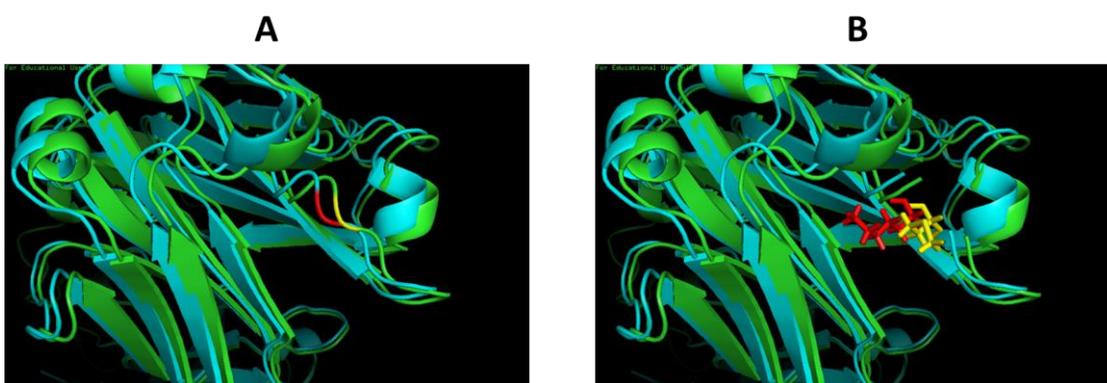


Figure 3.16 The overlay of the predicted structure before and after the K133M mutation. Mutation is shown in cartoon in Picture A and stick in Picture B. C226S is coloured in cyan with residue 133 in red; K133M is coloured in green with residue 133 in yellow.

The monomer loss in K65M is 4% more than that in K133M. As both of the mutations are on the loop regions of heavy chain, the difference in monomer loss may be due to their mutational positions that lead to different biophysical properties. As calculated by POPS (Fraternali & Cavallo 2002; Cavallo et al. 2003), the K65M and K133M positions are of SASA fraction 68.81% and 54.16% before mutating, and SASA fraction 38.08% and 63.97% after mutating. Shahina also found that K65M is of lower solvent accessible surface area (SASA) ( $21,208.3 \text{ \AA}^2$ ) than that of K133M ( $21,241 \text{ \AA}^2$ ) (Ahmad 2011). This liquid kinetic work also showed that K65M was significantly less stable than any of other mutants. Therefore, nearly half of the SASA is reduced by the K65M mutation while K133M mutation even yields more SASA. As a result, the electrostatic repulsion is greatly weakened by K65M.

Compared to C226S, adding one more positive charge almost has no effect to the monomer loss in S75K and L50K. This phenomenon suggests that even though the electrostatic repulsion could inhibit the association of Fab but its effect is limited. Moreover, long-range electrostatic interaction is a weaker force compared to hydrophobic interaction and hydrogen bonding. As a result, its benefit for stabilising may be shielded and even acts in a destabilising role if short-range non-covalent interaction is compromised. Studies (Pace et al. 2000) have found that charge-charge interactions actually stabilise the protein in a native state, but they also reduce the free energy in the unfolded states.

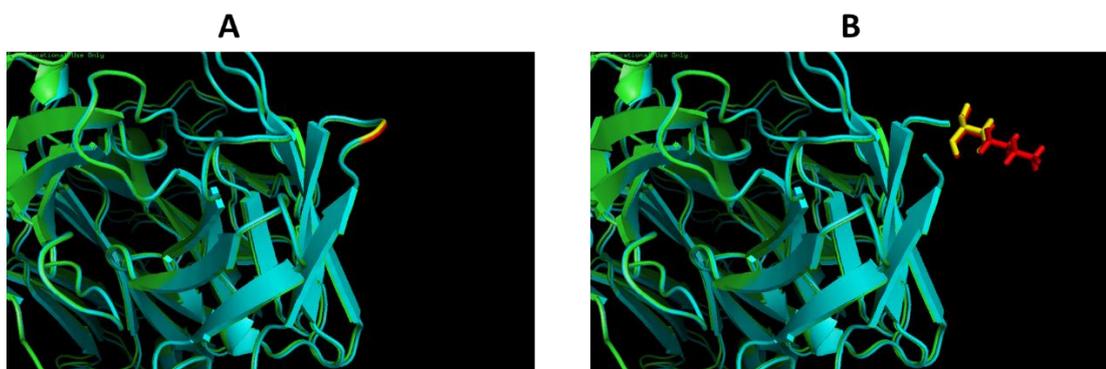


Figure 3.17 The overlay of the predicted structure before and after the S75K mutation. Mutation is shown in cartoon in Picture A and stick in Picture B. C226S is coloured in cyan with residue 75 in red; S75K is coloured in green with residue 75 in yellow.

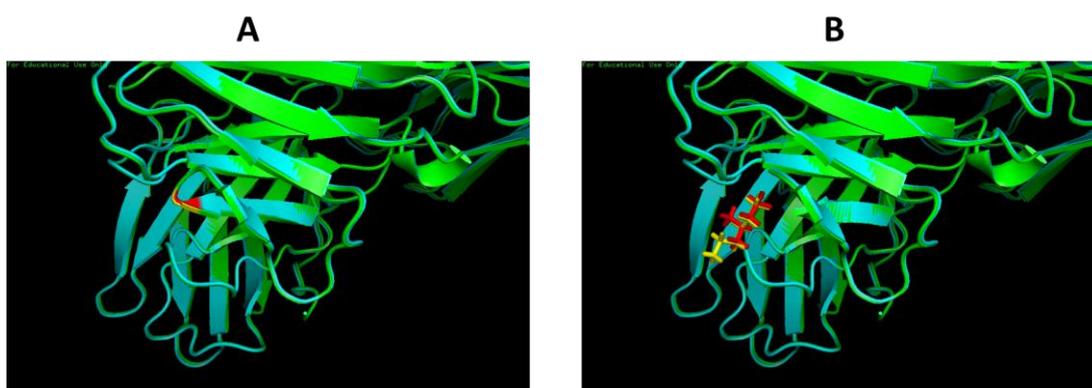


Figure 3.18 The overlay of the predicted structure before and after the L50K mutation. Mutation is shown in cartoon in Picture A and stick in Picture B. C226S is coloured in cyan with residue 50 in red; L50K is coloured in green with residue 50 in yellow.

In the present work, the mutational charge changes in S75K and L50K led to less than 1% less and less than 1% more monomer loss, respectively. As shown in Figure

3.17 and Figure 3.18, even though both of the mutations are not predicted to change the secondary structure, the relative exposure of the two positions to the outer surface is different. As calculated by POPS (Fraternali & Cavallo 2002; Cavallo et al. 2003), the S75K and L50K positions are of SASA fraction 52.59% and 25.51% before mutating, and SASA fraction 91.22% and 41.03% after mutating. Therefore, the S75K position locates relatively outward compared to L50K. The mutated lysine in S75K could play an important role in suppressing protein-protein attraction by repulsive electrostatic interaction. The mutation to lysine in L50K, on the other hand, may create more steric disturbance and less intra-chain hydrophobic stability.

### 3.3.4 Formulation and mutant effects on cake morphology

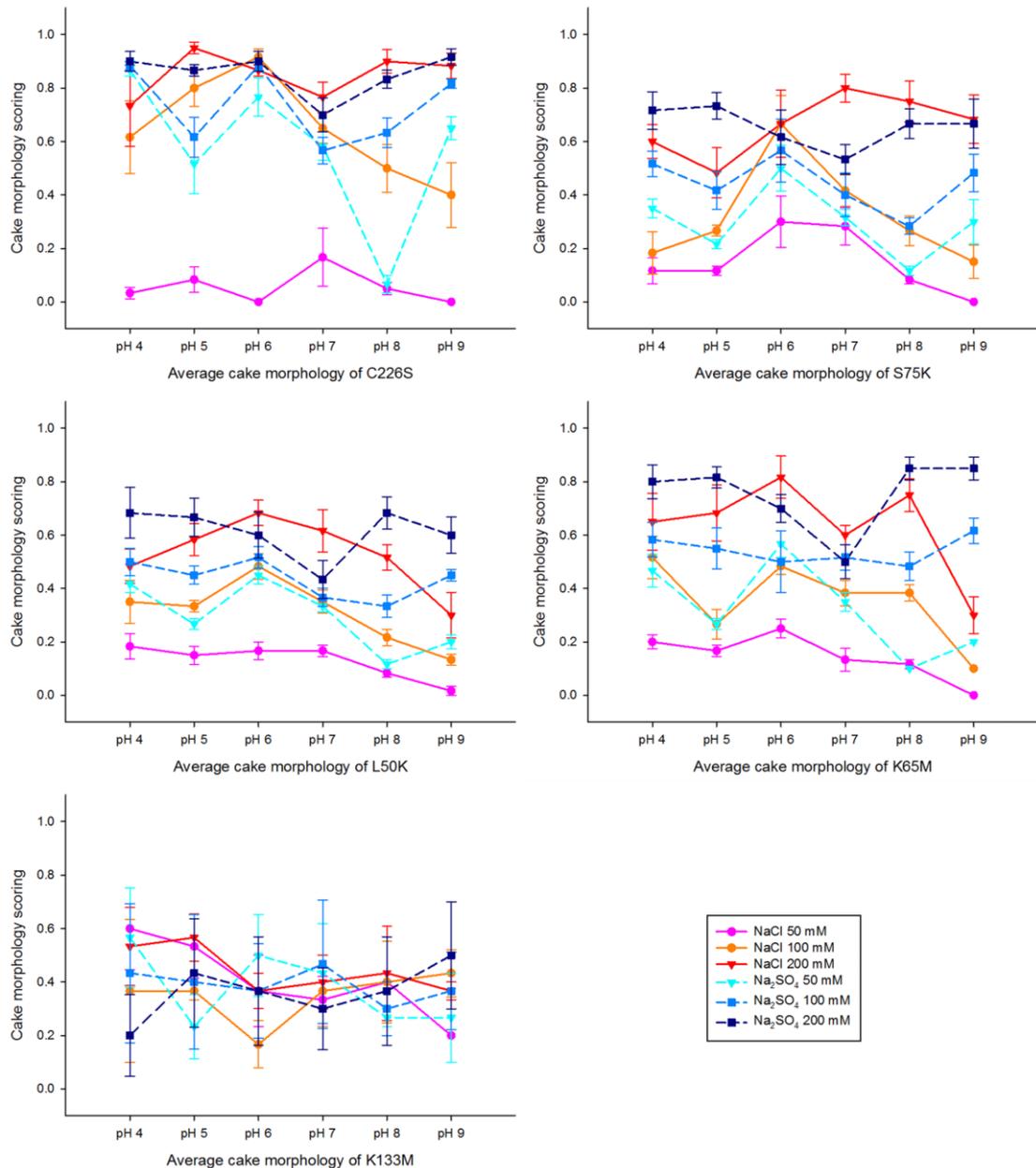


Figure 3.19 The cake rating of freeze-dried Fab mutants averaged from two freeze-drying batches, except for K133M with one batch.

Triplicates were used in each batch. Error bars were standard error of the mean. The scales of all the vertical axes were set at the same level for the ease of cross comparison. One data point of K133M at pH 8, Na<sub>2</sub>SO<sub>4</sub>, 100 mM was ignored due to the cake floating out from the well.

The cake morphology scores obtained by visual inspection of all freeze-dried Fab samples are shown in Figure 3.19. Scores ranged from 0 to 1 with more acceptable

cakes formed at higher ionic strengths. The cake scores showed a complex dependence on pH and mutants.

### 3.3.4.1 Influence of pH and mutants on cake morphology

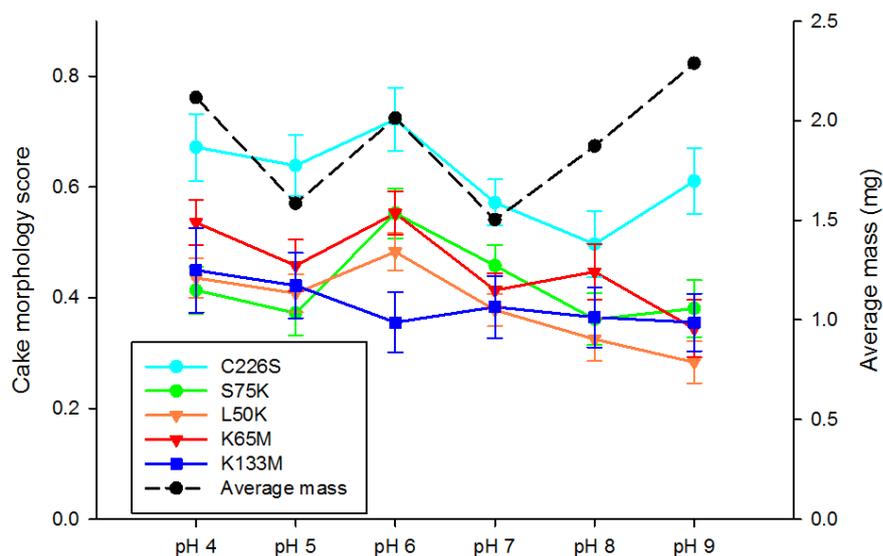


Figure 3.20 The impact of pH for each mutant upon cake morphology. Each data point of cake morphology was averaged from all the 3 ionic strengths, 2 salt types, and repeats of 2 batches for each mutant (n=36).

Figure 3.20 shows the mutants' cake morphology at different pH after freeze-drying. The mutants in general behave similarly at different pH. The cake morphology dipped at pH 5 and increased at pH 6 (except K133M), then it fluctuated between score 0.3 to 0.45 at pH 7-9. C226S resulted in 0.1-0.2 higher than the average of the four other mutants, but this was due to the observers' criteria change over the later freeze-drying batches.

In fact, the cake scores largely depended on the mass of the cake. Therefore, a green dash line was plotted against the five mutants line in Figure 3.20. This is to determine if the variations among different pH were caused by cake mass or the pH conditions. At pH 4-7, the cake of all mutants, except K133M, closely related to the trend of their mass change. At pH 8-9, most mutants, except C226S, tended to remain plateau instead of rising to a highest score. The C226S mutant was most sensitive to the change of pH; while K133M was least sensitive across all the pH.

The pattern of average mass change (black dash line in Figure 3.20) across the pH looked relatively the opposite of the trend of monomer loss as observed in Figure 3.8. Therefore, a correlation was plotted in Figure 3.21. The plot showed that approximately 30% of negative correlation could be seen between the monomer loss and cake mass. This is very interesting as an increasing cake mass would reduce the monomer loss on a pH level. This may be explained as a higher cake mass was more likely to retain the cake within the well, instead of floating out during freeze-drying.

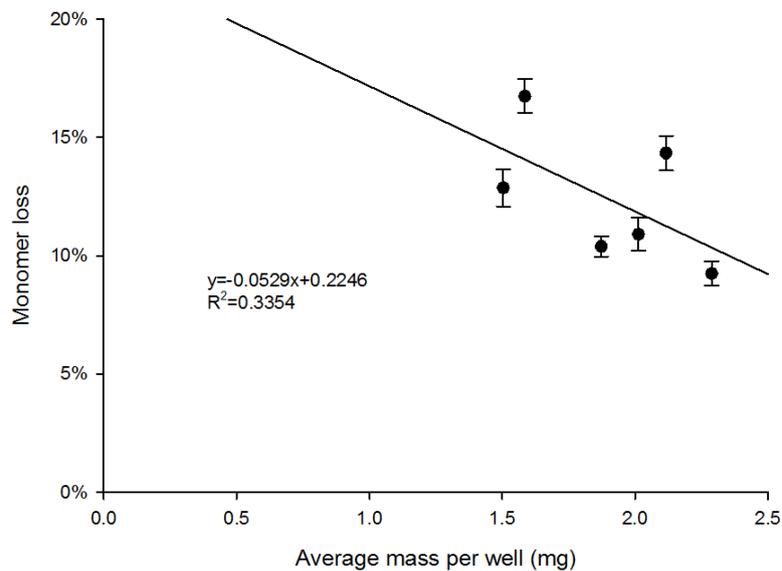


Figure 3.21 The correlations between monomer loss and cake mass at different pH

### 3.3.4.2 Influence of salt types, ionic strength and mutants on cake morphology

#### morphology

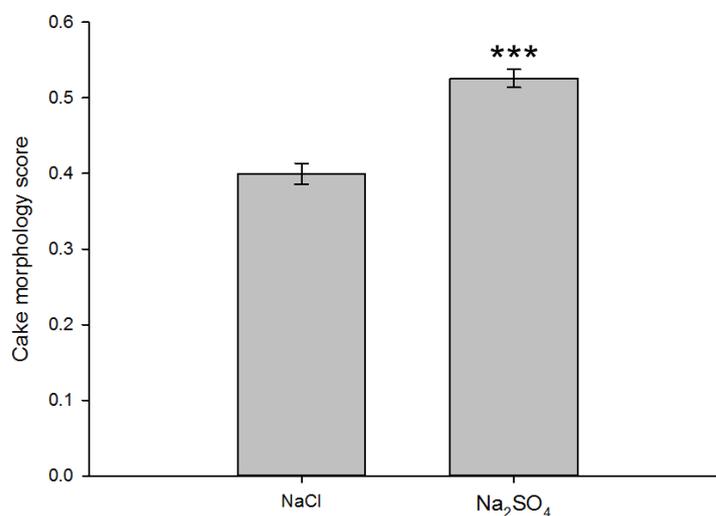


Figure 3.22 Overall cake morphology score for formulations adjusted with NaCl or Na<sub>2</sub>SO<sub>4</sub>

Each data point was averaged from all the three ionic strengths, 6 pH and repeats of 9 batches for the 5 mutants (n=486). Error bars were standard error of the mean. Two-sample *t*-test assuming unequal variances were performed between formulation with NaCl and Na<sub>2</sub>SO<sub>4</sub> (\*\*\*) *p*<0.001, \*\* *p*<0.01, \* *p*<0.05).

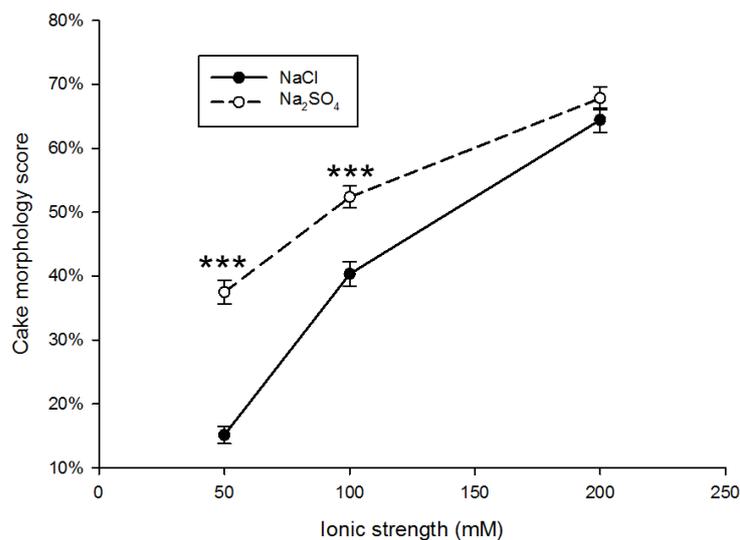


Figure 3.23 Average cake morphology score for formulations at different ionic strength adjusted with NaCl or Na<sub>2</sub>SO<sub>4</sub>

Each data point was averaged from all the 6 pH and repeats of 9 batches for the 5 mutants (n=162). Error bars were standard error of the mean. Two-sample *t*-test assuming unequal variances were performed between formulation with Na<sub>2</sub>SO<sub>4</sub> and NaCl at 50, 100 and 200 mM ionic strength, respectively (\*\*\*) *p*<0.001, \*\* *p*<0.01, \* *p*<0.05).

The impact of ionic strength on the average cake morphology scores for NaCl and Na<sub>2</sub>SO<sub>4</sub> formulations are shown in Figure 3.23. Cake morphology was improved at higher ionic strength, and the scores converged at 200 mM Na<sub>2</sub>SO<sub>4</sub> and NaCl. The significantly higher cake morphology scores for Na<sub>2</sub>SO<sub>4</sub> relative to NaCl at 50 and 100 mM ionic strength were not due to the higher cake mass obtained for Na<sub>2</sub>SO<sub>4</sub> (Figure 3.24). Alternatively, XRD analysis indicated that Na<sub>2</sub>SO<sub>4</sub> formulations had the higher crystallinity, which may also explain the difference in cake morphology. SO<sub>4</sub><sup>2-</sup> is also known to be able to raise the  $T'_g$  of L-arginine while Cl<sup>-</sup> had minor effect or decreased the  $T'_g$  (Izutsu et al. 2005), while the use of salts with higher eutectic melting temperatures (Na<sub>2</sub>SO<sub>4</sub> -2°C, NaCl -23°C) can also avert cake collapse (Chang & Randall 1992). This suggests that in the current work, higher  $T'_g$  or crystallinity of Na<sub>2</sub>SO<sub>4</sub> may reduce the chances of cake collapse and give rise to a more aesthetic cake form.

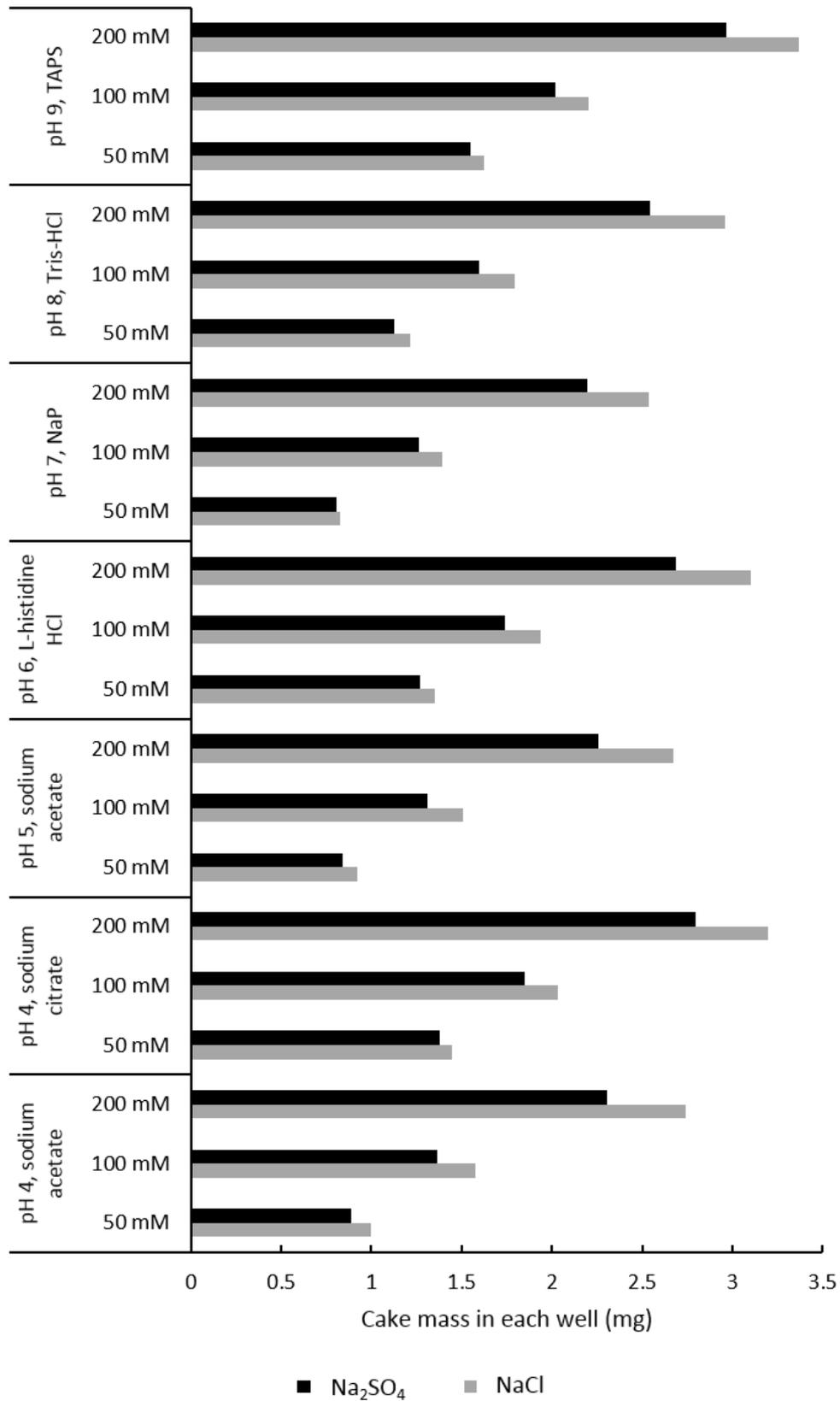


Figure 3.24 The theoretical mass formulated in each well by considering buffer salt, neutral salt (NaCl or Na<sub>2</sub>SO<sub>4</sub>) and protein

### 3.3.5 Monomer loss vs cake score

Though the cake appearance is an important factor for the drug final dosage form, its relevance to protein stability, however, is not validated. A poor looking cake is usually accompanied with collapse. It has been found (Schersch et al. 2010) that compared to non-collapsed one, even though the collapsed cake had lower specific surface area and more residual moisture, the protein stability was preserved and with similar reconstitution times. In this work, the cake morphology was not correlated with monomer loss ( $R^2 = 0.0131$ ) as shown in Figure 3.25, which is consistent with previous work.

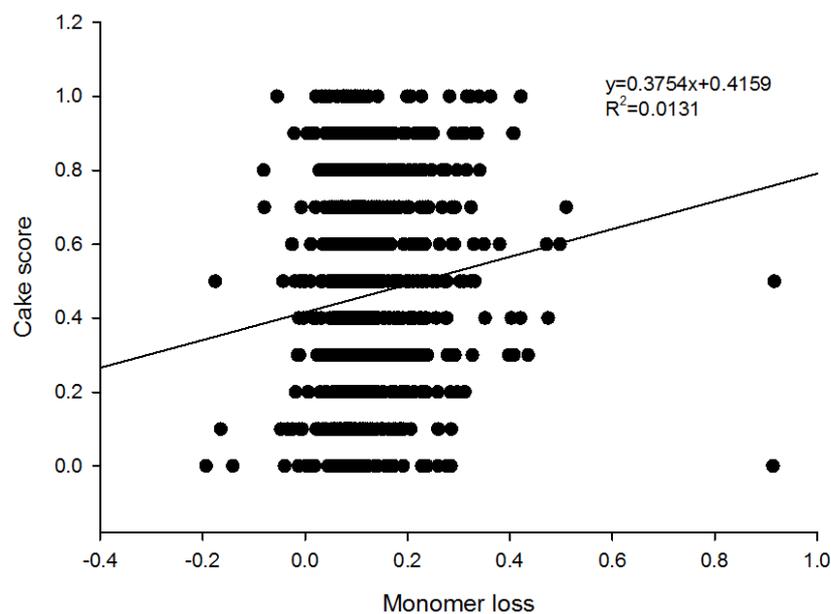


Figure 3.25 The correlation between cake score and monomer loss

### 3.3.6 Sweet plot for “monomer loss” and “cake morphology”

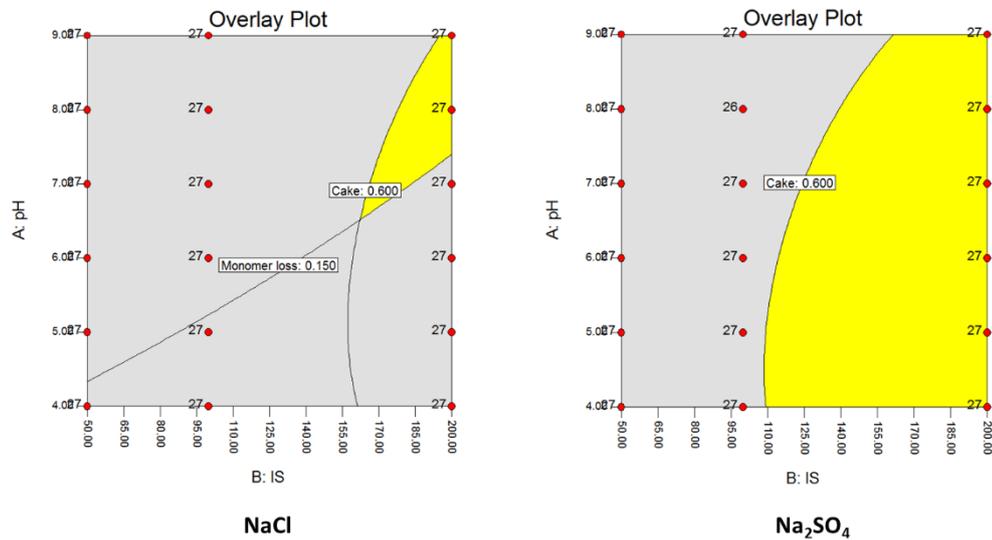


Figure 3.26 The sweet plot for freezing drying performance with cake score > 0.6 and monomer loss < 0.15

The cake morphology and monomer loss were not correlated, and thus the conditions that are acceptable for both objectives can be investigated using a “sweet plot”, in which several threshold boundary lines are overlaid (Grant et al. 2009). Figure 3.26 shows the sweet zones (yellow areas) when setting a cake score > 0.6 and monomer loss < 0.15. Clearly, the formulation with Na<sub>2</sub>SO<sub>4</sub> yields a much larger acceptable area than with NaCl. For NaCl, the sweet zone is at an ionic strength above 170 mM and at pH > 7. For Na<sub>2</sub>SO<sub>4</sub>, the larger sweet zone is at an ionic strength above 110 mM, and across all pH tested. Therefore, if monomer loss and cake appearance are the main factors to consider, it is preferable to select Na<sub>2</sub>SO<sub>4</sub> as the additive with which to control ionic strength.

## 3.4 Conclusion

This experiment demonstrated a rapid way to screen key formulation and protein property factors that influence the freeze-drying performance. This was applied to investigate for the first time, the impact of protein surface-charge mutations on the

tolerance to freeze-drying. It is interesting that even only single amino acid substitutions that alter the surface charge of the Fab molecule, could greatly influence the aggregation during freeze-drying.

With no additional excipients added, the solution composition was relative simple to enable a clearer interpretation of the impact of mutants, pH and ionic strength and neutral salt type. Mutants with one net surface charge removed led to increased monomer loss, whereas addition of one net charge had relatively little impact. This indicated that the pseudo-wild-type Fab was already just above a critical point in colloidal stability whereby other factors had more dominant influences on monomer loss. High ionic strength, low pH and use of NaCl led to greater monomer loss. The ionic strength or the mass of the cake contribute mostly in improving the cake appearance while the other factors showed less impact. Compared to NaCl, the use of Na<sub>2</sub>SO<sub>4</sub> is a better option to minimise the monomer loss while retaining the cake morphology.

The monomer loss data suggests that the freeze-drying system is complicated even in a relative simple solution composition without any excipients. As batch-to-batch noise and replicates deviations may contribute to the low variation, several measures are recommended to provide a more robust, reproducible dataset. These measures include 1) use mat stoppers to cover the 96-well plates to prevent cake floating out; 2) use single channel pipette if multi-channel pipette is not accurate enough; 3) prepare stock working buffer at a time and aliquot them in freezer; 4) consider more repeats if triplicates deviate considerably; 5) it is better to run more mutants in each cycle with fewer conditions if one aims to differentiate the mutants' response on freeze drying stress.

The monomer loss observed in this experiment is not large enough (less than 50%) to differentiate the mutational impact on protein stability. A suboptimal freeze-drying process could be developed in order to cause more monomer loss and larger improvement space for protein stability. Double or triple mutants that encompass advantageous substitutions could be considered to reduce the aggregation propensity. To better elucidate the stresses, freeze-thawing could be used to separate the stresses

caused by freezing and drying individually. In addition, the activity would also need to be assessed.

The present study proposed a strategy of altering the active pharmaceutical ingredients (API) to improve their stability in freeze-drying. Though clinical formulation would not contain high salt, the salt conditions used here help us probe the impact of charge modification. Further study could focus on excipient screening on the API and buffer condition with improved cryoprotection and lyoprotection (e.g. higher charge, low salt and pH closing to pI). We hope this API-based modification could be a complementary method for the traditional excipient-based formulation development.

# 4 Hybrid mutagenesis design and pilot scale production

## 4.1 Introduction

### 4.1.1 Bioinformatics tools to indicate the disorder of residues

#### 4.1.1.1 RMSF

The advance of computer technology has enabled researchers to model complex biomolecules, their dynamics and interactions, as well as to calculate their free-energy (Pronk et al. 2013). Root mean square fluctuation, or RMSF, measures the deviation over time between a particle position and its reference position. By modelling the solvent environment around a protein, as a virtual box of water molecules, and then simulating the effects of thermal energy on molecular dynamics over time, the disturbance of atomic coordinates can be determined as an RMSF and used to indicate the relative flexibility of protein residues. This information would offer insights that can guide potential mutations with the aim of replacing the flexible residues with amino acids that lead to more stability through improved interactions or packing.

Many software packages and forcefields are currently available to carry out molecular dynamics (MD) simulations for macromolecules, including CHARMM (Brooks & Brooks 2009), Gromacs (Pronk et al. 2013; Abraham et al. 2015) and AMBER (Case et al. 2005; Salomon-Ferrer et al. 2013). MD simulations can be carried out on the molecular equivalent of the ps to ns time scale, which could cover local flexibility of methyl rotation, loop motion and side-chain rotamers (Henzler-Wildman & Kern 2007). Therefore, MD simulation is a valuable way to further analyse the structures generated by

X-ray crystallography or NMR, as deposited in the protein databank (PDB). Ohmura (Ohmura et al. 2001) conducted a simulation for mutants of hen egg white lysozyme and successfully identified stabilising mutants. RMSF values calculated from MD were found to be very similar to the B-factor values obtained by X-ray crystallography. Together with dynamic Van der Waals energy, the stable mutants could fill the cavity in the core and strengthen the hydrophobic interaction without introducing unfavourable Van der Waals energy. Another study (van der Kamp & Daggett 2010) on pathogenic mutations discovered that conformational flexibility was altered by mutations. Two mutants exhibited notable C $\alpha$  RMSF increases, which was most significant for residues around the points of mutation.

#### *4.1.1.2 B-factor*

B-factors (or temperature factors) are determined by X-ray crystallography, and are normally annotated within the standard PDB file format in the column after atom's Cartesian coordinates and occupancy value. It represents the extent of thermal motion of an atom such that a high B-factor value indicates a high fluctuation for that atom (Yuan et al. 2005). Similar to RMSF, the B-factor is another index that can be used to determine if a specific residue is highly flexible.

If the PDB file is not available, the B-factor can be predicted based on its sequence, and comparison to a large dataset of known PDBs (Yuan et al. 2005). The B-factor is closely correlated with the protein secondary structure, as it reflects the disorder and flexibility of the folded conformation. It has been found (Linding et al. 2003) that protein disorder is mostly found in loop or random coil regions of protein structures, as these exhibit the highest B-factors for C $\alpha$  atoms. To predict a more accurate loop structure for a sequence with unknown structure, available structures with homologous sequences can be cross-compared. Those with lower B-factors in the target loop, and also similar sequence identities can be used to model the unknown loop structure (Choi & Deane 2010).

Due to the different settings in X-ray crystallography and the heterogeneity of various proteins, the B-factor values are often determined at different scales. Therefore, normalisation is required to compare multiple B-factor values on a similar scale and to remove outliers (Smith et al. 2003). After normalising, B-factors have been used to guide iterative saturation mutagenesis and increase the thermostability of a lipase, by modifying only sequence regions with the highest B-factors (Reetz & Carballeira 2007). In that approach, a program was developed to calculate the amino acid B-factor by averaging all of its atom B-factor values except hydrogen. It was found that two hyperthermophilic mutants displayed melting temperature 40°C more than the wild type after a 60 min heat treatment.

### **4.1.2 Design of stable proteins based on consensus tools**

Functional proteins evolve so as to adjust to their neighbouring environment. Deleterious mutations become extinct and helpful ones remain. As a result, the regions that keep their consistency usually play an essential role to maintain the stability or functionality of proteins (Miller & Kumar 2001). For the structures of antibodies, most of the residues retain high sequence similarities in order to form a “Y” shaped conformation and target the various antigens through their CDR (Complementarity Determining Region). The consensus theory can be used for two purposes, 1) it would be better to apply mutagenesis on non-consensus regions or, mutate the residues back to conserve ones originally exist in the ancestors so as to inherit its stability from ancestor species; 2) if the atomic structure is unknown for the protein of interest, one can use homology modelling to predict its structure by using a resolved, homologous structure. However, the non-consensus regions may locate at the CDR. Therefore, mutations upon the CDR would affect the binding and activity, which needs to be considered before mutagenesis.

The influence of stability upon protein evolvability has been well studied (Bloom et al. 2005; DePristo et al. 2005). It is found that more than half of the single mutants

introduced into most proteins, would retain their native functions (Bloom et al. 2006). However, most of the mutations are neutral or deleterious to the level of that function. As missense mutations accumulate, the stability of a protein is exponentially decreased, which greatly affects the ability of a protein to adapt new functions (Tokuriki & Tawfik 2009). In order to improve the stability without compromising the functionality (or *vice versa*) of proteins, one can align the sequences of homologous species, identify the consensus sequence and design amino acid substitutions towards the consensus. It has been found that (Steipe et al. 1994) the most frequently occurring amino acids exhibit a greater stabilising effect than less frequent ones. To test the “consensus approach” the consensus sequence of a phytase was first calculated from a total of 19 homologous sequences (Lehmann et al. 2000). Then, they constructed the Consensus phytases were then constructed, expressed, purified and their  $T_m$ -values measured. The results showed that over 20°C improvements in the  $T_m$  could be obtained without compensating the catalytic activity. In later research, many others also confirmed the “consensus approach” with improved protein stabilities (Lehmann & Wyss 2001; Jackel et al. 2010).

### 4.1.3 Homology modelling

Homology modelling (or comparative modelling) aims to predict the 3D structure of a protein based on its amino acid sequence and available structures with homologous sequences. It is used when the experimental structures (e.g. x-ray, NMR) are not available in the database (Ginalski 2006). Homology modelling can be classified into two categories, which are template-based modelling and *ab initio* methods. The template-based method is used when the unknown structure has a high amino acid sequence similarity with one or more experimental structures. The *ab initio* (or “*de novo*”) method is used when the sequence similarity to known structures is low. So far, most templated-based methods have focused on cases when the sequence identity is greater than 35% (França 2015).

The homology modelling building procedure is carried out step-by-step. A general protocol involves template selection, sequential alignment, backbone/loop/sidechain

modelling, model optimization and validation (Venselaar et al. 2010). In order to produce an accurate estimation for the native-like structure, the prediction work requires a comprehensive understanding for the evolutionary mechanism, sufficient sequence similarity searches from protein database and reasonable estimation for primary and secondary structures according to statistical distribution (Pavlopoulou & Michalopoulos 2011).

Various computational tools are available to achieve this purpose, which include Modeller (Webb & Sali 2014), SWISS-MODEL (Biasini et al. 2014; Arnold et al. 2006), and Phyre2 (Kelley et al. 2015). A contest, Critical Assessment of Structure Prediction (CASP), for evaluating the performance of these predictions tools is held every two years (Moult et al. 2014). Enormous progress has been achieved over the past 20 years for prediction accuracy. The Z-score is used as a statistical measurement to describe how likely the energy of a predicted structure deviates from its misfolded ensemble conformations (Zhang & Skolnick 1998). A good Z-score value should be within the characteristic range for proteins with a similar size in the same group (Wiederstein & Sippl 2007). In the CASP10 competition, “BAKER-ROSETTASERVER” achieved the top position in the Z score analysis in the “Assessment of template-based protein structure predictions” (Huang et al. 2014). Therefore, it is advantageous to use Rosetta to build the homology model based on an existing homologous structure.

#### **4.1.4 Computational prediction of protein $\Delta\Delta G$ upon mutation**

A range of computational methods have been developed to predict the overall change in protein stability ( $\Delta\Delta G$ ) due to mutations. Due to the complexity of macromolecules resolved in atomic scale, a computational method needs to compromise between the “search problem” and the “scoring problem” (Potapov et al. 2009). With limited computational resources, the “search problem” aims to search potential conformations in greater number and detail; whereas the “scoring problem” aims to

describe as many structural interactions as possible. Two types of forcefield are generally used. One is the physical-based potentials (PBP), which considers the fundamental physical forces within atoms; and the other is the knowledge-based potentials (KBP), or statistical potentials, which depend more on a statistical summation derived from protein datasets (Potapov et al. 2009; Khan & Vihinen 2010). Some methods are an integration of PBP and KBP (e.g. Rosetta), which can thus be classified as “empirical potentials” (Potapov et al. 2009; Khan & Vihinen 2010). In addition, some prediction methods have been developed based on machine learning algorithms (e.g. neural network, support vector machine) (Capriotti et al. 2004; Cheng et al. 2006) and a combination of PBP, KBP and machine learning has also been developed (Khan & Vihinen 2010).

Potapov (Potapov et al. 2009) has evaluated six computational tools to predict the protein  $\Delta\Delta G$ . It has been found that the correlation coefficients range from 0.26 to 0.59 for Rosetta, Hunter, FoldX, I-Mutant2.0, CC/PBSA and EGAD, respectively. However, those methods perform differently on different structure locations. For example, CC/PBSA and Rosetta work better for exposed residues than buried ones while the other four methods work in the opposite way; FoldX and Rosetta predicted best for unstructured regions while the other four methods behave best for  $\beta$ -sheet mutations.

Khan & Vihinen (Khan & Vihinen 2010) conducted a performance test for 11 online stability predictors. It has been found that I-Mutant3.0 (structure version), Dmutant and FoldX achieved the best accuracies from 0.54 to 0.64. They also compared the predictors performance on the mutation location properties (secondary structure, accessible surface area) and the overall protein structure categories (e.g. mainly  $\alpha$ -helical, mainly  $\beta$ -stranded, etc.). They found that most programs behave with similar accuracy on secondary structural elements. I-Mutant2.0, FoldX, MUpro, MultiMutate, and CUPSAT resulted in the best in sensitivity for structure categories. Most programs except MultiMutate gave better prediction accuracies on exposed mutations than buried ones.

Niroula (Niroula & Vihinen 2016) has conducted a comprehensive review of the protein stability prediction by bioinformatics approaches from the human medical perspective. Various aspects have been discussed, which include mutation databases, the principles implemented to identify variations-function correlations (machine learning, evolutionary conservation, energy functions) and a range of prediction tools. Moreover, the extant problems are illustrated, which include improper transplanting of methods, overfitting and imbalance of training data.

As discussed above, recent years have witnessed a tremendous progress in predicting protein stability. Besides the problems mentioned above, there are still some other issues that impede the accuracy of the algorithm. The most important one is the theoretical solution condition is usually not explicitly defined. It is known that the  $\Delta G$  not only relates to the protein itself but also depends on its surrounding environment. However, the aqueous solution conditions that are used to generate the experimental data differ across different labs. As a result, the impacts of pH, temperature, ionic strength have not been sufficiently considered, or an “average” condition is reflected based on the sampling data. Similarly, the processing parameters are also not fully addressed. Though many studies are undertaken in aqueous solution, some stability tests are carried out in some other forms like freeze-thawing, freeze-drying and spray drying. Some stresses involved in these treatments are not the same as those in aqueous phase or exhibit at different degrees. In conclusion, the trained protein stability predictors are not adequately optimised if the contribution of solution interaction ingredients and processing parameters are not taken into account.

#### **4.1.5 Rosetta**

Rosetta is a flexible, multi-purpose protein modelling software. It has been developed since 1998 and has achieved considerable progress in structure prediction (Chivian & Baker 2006), design (Kuhlman et al. 2003), stability improvement, and protein molecule docking (Leaver-fay et al. 2014).

In general, Rosetta uses both physical-based and knowledge-based potentials to capture hydrophobic effects, van der Waals interactions and hydrogen bonding (Das & Baker 2008). To describe covalent bonding, Rosetta does not use physical-based potentials to describe bond torsion, but instead applies empirical torsion angle distributions (preferred rotamers) derived from the database of PDB structures. To describe the protein folding or packing, the hydrophobic effect and electrostatic desolvation cost are computed by a Lennard-Jones potential with an implicit solvation model. Hydrogen bonding is computed by an explicit hydrogen-bonding potential. In order to score a protein structure so as to indicate its stability, the free energy is broken down into various elements (e.g. Van der Waals, solvation, electrostatics) and each element is assigned a weighted contribution. As a result, the total free energy is computed as the weighted sum of the component energies (Leaver-fay et al. 2014).

However, there are still some problems not optimally addressed. The first one is that the long-range electrostatic interaction is ignored (Das & Baker 2008). Therefore, some colloid stability caused by protein-protein electrostatic repulsion is not considered, which may lead to some bias when linking unfolding with aggregation. The second one is that the entropy of a structure is not sufficiently considered. It is approximated that the conformational entropies of various properly folded protein are similar (Das & Baker 2008). As a result, the stability scoring of a structure is based on a single PDB structure instead of ensembles of potential structures. The third one is that due to the empirical nature of the scoring function, some energy terms are considered more than once. For example, due to the complexity of quantum mechanical effects, the torsional potentials could not be decomposed into individual classical contributions, which may be over-emphasised in non-bonded interaction energy terms.

Potapov assessed the performance of Rosetta together with five other available tools for prediction of protein  $\Delta\Delta G$  upon mutation (Potapov et al. 2009). It was found that Rosetta performed less well than the others as scored by correlation coefficient. However, the operation parameters were not listed in the paper and we do not know the exact application executable file used. As a response to the poor performance of

Rosetta, Kellogg conducted a thorough search to identify the optimal relaxation approach upon point mutation (Kellogg 2011). It was found that the minimisation method involving limited backbone minimisation after repacking all sidechains achieved the highest experiment-prediction correlation coefficient of 0.69. In spite of the improved prediction accuracy, it was noted that the structure samples used in Kellogg's testing were all no greater than 350 residues. Therefore, there remains a question as to whether the protocol is also applicable for larger protein molecules like antibodies including the antibody fragment Fab used in this project with 442 residues.

### **4.1.6 Aims of the chapter**

The aim of this chapter was to develop stabilising and destabilising mutants for both freeze-drying and liquid aggregation kinetics work in the subsequent two chapters. As flexibility is an important mechanism in defining overall protein stability (Fields 2001), both RMSF and B-factor would be used to determine the flexible sites, and then Rosetta used to identify potential mutations that could rigidify those flexible sites (Yu & Huang 2014). Several destabilising mutants would also be developed as a reference to validate the strategy.

## **4.2 Materials and Methods**

### **4.2.1 *In-silico* mutagenesis**

The design of new mutants was conducted using the Rosetta software (Das & Baker 2008). Rosetta is a protein modelling software suite that uses algorithms for *in-silico* modelling and structure analysis. The entire mutagenesis procedure encompasses the following three elements:

- Structure cleaning
- Homology modelling
- Mutating and scoring

In the “Structure cleaning” step, the crystal structure of human germline antibody 5-51/O12 (PDB ID 4KMT) was downloaded from the PDB website (<http://www.rcsb.org/pdb/explore/explore.do?structureId=4KMT>). The raw PDB file was cleaned and only the “ATOM” section was retained. The residue number was renumbered from 1 to 442, taking the original light chain numbering of 1 to 214, and heavy chain from 215 to 442. Then hydrogen atoms were added to make a full-atom structure.

In the “Homology modelling” step, Rosetta method “minirosetta” (Chivian & Baker 2006; Raman & Vernon 2009) was applied to replace the residues in the 4KMT model with those from A33 C226S. There are five disulfide bonds within the C226S variant, which are

- Interchain disulfide bond: LC214 – HC220
- Intrachain disulfide of light chain (LC): LC23 – LC88, LC134 – LC194
- Intrachain disulfide of heavy chain (HC): HC144 – HC200, HC96 – HC22

It was found that disulfide bonds randomly existed after residue replacement by “minirosetta”. Therefore, in order to obtain a relatively promising model with all five disulfide bonds intact, 20,000 PDB were generated, clustered and ranked according to Rosetta Energy Unit (REU). After residue replacement, 6811 structures contained all five intact disulfide bonds. Among them, 1000 structures of the lowest score were selected. A clustering step was then used to group multiple structures based on their similarities. It aims to identify which category of structures is mostly likely to represent the real model. During the clustering step, the largest category consisted of 573 structures and the structure with the lowest score among that category was selected as the model of C226S to use in the subsequent “Mutating and scoring” step.

In the “Mutating and scoring” step, a single point mutation procedure was conducted by Rosetta method “ddg\_monomer” (Kellogg 2011). Each of the 442 residues in the PDB structure was mutated into the other 19 amino acid residues. As a result, a total of 8398 structures with single mutations were created. The files associated with the structure preparation were generated by using several python scripts, and the

corresponding jobs were submitted to UCL Legion High Performance Computing Facility (Legion@UCL) with Rosetta Version 2015.31.58019. After mutating, the change in stability ( $\Delta\Delta G$ ) induced by each point mutation, was calculated with reference to the original C226S model.

## 4.2.2 Analysis of residue flexibility

After determining the  $\Delta\Delta G$  for mutation candidates, several *in-silico* protein analysis methods were used to validate the potential mutants from different perspectives.

### 4.2.2.1 RMSF simulation in Gromacs

In the preparation step, the C226S and mutant PDB structures were obtained from Rosetta. The protonation states of chargeable residues at pH 4 were determined by uploading PDB files to [http://nbc-222.ucsd.edu/pdb2pqr\\_1.9.0/](http://nbc-222.ucsd.edu/pdb2pqr_1.9.0/). In Gromacs, the Fab PDB file was initially converted to a topology file with its five inter/intra-disulfide bonds retained. A OPLS-AA/L all-atom force field was selected and protonation status was manually adjusted. The Fab was then centred into a cubic box with 1 nm away from the edge of the box. After that, the box was filled with water molecules as solvent. Then the entire solution box was neutralised and adjusted to an ionic strength of 200 mM by adding Na<sup>+</sup> and Cl<sup>-</sup>. In the end, the structure was subjected to an energy minimisation step, and equilibrated at 300 K and atmospheric pressure. For equilibration, each mutant was simulated for more than 40 ns. Jobs were submitted to UCL Legion High Performance Computing Facility (Legion@UCL) to facilitate parallel simulations. At least three repeats were conducted to validate if data was reproducible. The RMSF data was exported based on trajectories beginning from 20 ns to allow for relaxation at the beginning of the simulation.

### 4.2.2.2 B-factor

In this work, multiple homologous structures were aligned to C226S in order to infer the flexibility across all residues. The raw PDB files of human Fab were downloaded

from <http://www.rcsb.org/>. All the PDB files were modified so that only one set of light chain and heavy chain remained. Due to the crystallisation uncertainties for excessive thermal motion of certain residues, it is noted that some residues were not displayed in their PDB files but shown in their sequence files. As a result, only the residues that exist in PDB files could be used for sequence alignment with C226S and B-factor cross comparison. Therefore, instead of using the FASTA sequence file, the actual sequence information was extracted from the PDB for heavy chain and light chain separately. After that, the sequence outputs of all the human Fab PDB were aligned with that of C226S using BioEdit (Hall 1999).

In the PDB profiles, each atom has its own B-factor. In order to have an overall inference for the whole residue, all the atomic B-factors of a same residue were averaged and assigned to their corresponding residues. These residual B-factors were then tabulated into the sequence alignment file so that they were aligned according to the C226S residues. In addition, the residual B-factors within each protein were normalised into a distribution with average 0 and standard deviation 1 (Reetz & Carballeira 2007). In the end, only the B-factors that accounted for existing C226S residues were retained. The average B-factors of C226S residues were calculated by averaging the B-factors from individual homologous structures. In order to reduce the scattering, the B-factors were further processed by window-averaging across 5 residues. The entire procedure to process the B-factor is shown in Figure 4.1.

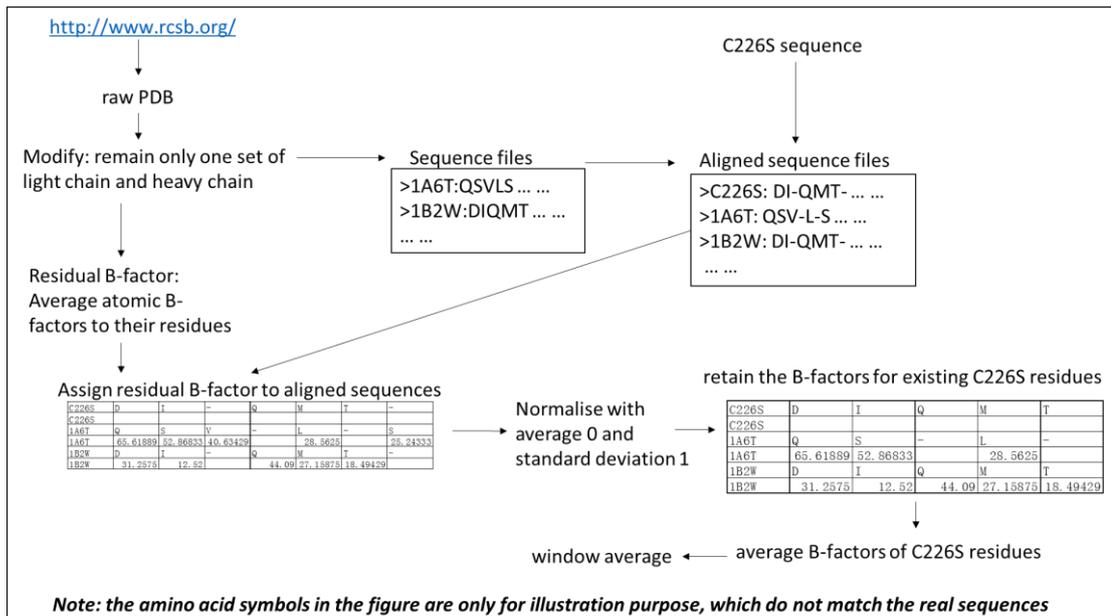


Figure 4.1 The schematic diagram for processing the B-factor

Raw PDB profiles were downloaded from the website, and only one set of light chain and heavy chain was retained for each profile. Amino acid sequence files were extracted from these modified PDB, and were then aligned against the C226S sequence file. In addition, the residual B-factors were obtained by averaging their atomic B-factors, and tabulated along the amino acid sequence alignment. Normalisation was conducted for the tabulated "B-factor/sequence alignment" and the averaged B-factors accounting for C226S residues were retained, and were further processed by window-averaging across 5 residues.

## 4.2.3 Design of stable and unstable mutant candidates

After the *in-silico* mutagenesis by Rosetta, RMSF and B-factor analysis, mutant candidates were selected for generation in the wet lab. Mutants were categorised into theoretical stable and unstable classes.

### 4.2.3.1 Stable mutant candidates

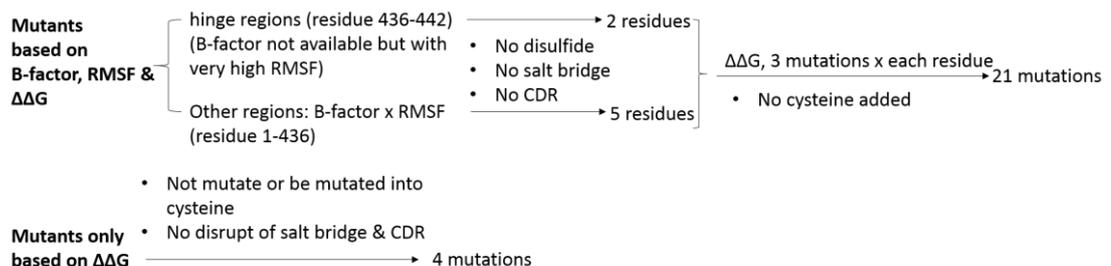


Figure 4.2 The schematic diagram to design stable mutants

21 mutants were designed based on a combined analysis of B-factor, RMSF and  $\Delta\Delta G$ . For hinge regions, HC-A227 and HC-A228 were selected for mutation. For other regions, the top 5 residues with highest B-factor and RMSF were selected. For a total of 7 selected residues, each was mutated into the three amino acids predicted by Rosetta to have the lowest  $\Delta\Delta G$  values from across all 19 candidates. There were also 4 mutations that only based on lowest  $\Delta\Delta G$ . All the mutations had no influence on disulfide bond, salt bridge, CDR and cysteine addition.

For stable mutants, Figure 4.2 shows the schematic procedure to select those for construction. 21 mutants were designed based on a combined analysis of B-factor, RMSF and  $\Delta\Delta G$ . The hinge regions in the heavy chains had no available B-factor values, but their RMSF values were very high. It was suggested that the hinge regions accounted for the flexibility and instability of an IgG, and switching it to another subclass could potentially improve the formulation stability while maintaining its binding affinity (Neergaard et al. 2014). Thus, the last two hinge residues, HC-A227 and HC-A228, were selected for mutation. For the other regions (residues 1 to 436), both B-factor and RMSF were considered. Because the scales of B-factor and RMSF were not equal, the B-factor and RMSF values were firstly normalised to between 0 and 1 according to Equation 4.1. Individual residues were then ranked based on the product of normalised B-factor and RMSF, and the top 5 residues were selected. For a total of 7 selected residues, each was mutated into the three amino acids predicted by Rosetta to have the lowest  $\Delta\Delta G$  values from across all 19 candidates. Three mutations were selected because Rosetta is expected to fit with 69% prediction accuracy (Kellogg 2011), so theoretically, three mutations should yield at least one stable mutant. An additional 4 mutations were selected based only on the lowest  $\Delta\Delta G$  values predicted by Rosetta, regardless of the flexibility of the target site.

$$\text{Normalised value} = \frac{\text{Raw value} - \text{Min}}{\text{Max} - \text{Min}} \quad \text{Equation 4.1}$$

In order to design the mutants only based on B-factor, RMSF and  $\Delta\Delta G$  values without losing or introducing new features, several additional filter criteria were applied. Mutations were avoided that would introduce a cysteine, remove a disulphide bond, or disrupt salt bridges. In addition, mutations were avoided in the Complementarity Determining Regions (CDR) of the Fab as this would affect Fab function, and guide

mutations to regions that could be potentially useful in any Fab generated by industry.

The CDR regions are listed in Table 4.1.

Table 4.1 The Complementarity Determining Regions (CDR) of C226S

Light chain	Heavy chain
CDR-L1: 24-KASQNVRTVVA	CDR-H1: 26-GFAFSTYDMS
CDR-L2: 50-LASNRHT	CDR-H2: 50-TISSGGSYTYYLDSVKG
CDR-L3: 89-LQHWSYPLT	CDR-H3: 99-TTVVPFAY

#### 4.2.3.2 Unstable mutant candidates

The unstable mutants were designed from only those candidates with the highest  $\Delta\Delta G$  values as predicted by Rosetta. These would increase the range of protein stabilities for further study. Moreover, if all the stable mutants resulted in higher stability, the unstable mutants would provide a “negative control” role. As above, the designed mutants avoided mutations from and to cysteine, salt bridge modifications, and CDR regions of Fab.

### 4.2.4 Primer sequence design

Primers were designed for site-directed mutagenesis of codons at the DNA level. Though one type of amino acid could be translated by different codons, certain species have their own preference towards specific codons. It is a good strategy to mutate towards codons preferred by the host cell species. In this research, the codons used mostly by *E. coli* were adopted from a previous codon usage work (Sharp et al. 1988).

In addition to the codon optimisation, the primers for individual mutants were also designed to meet certain commonly used rules, which include:

- Locate the mutation in the middle of the primer
- A GC content of more than 40%
- Include a GC clamp (G/C at the 3' end)
- Avoid more than 3 G/C bases in the final 5 at the 3' end
- Use a primer length of 25-45 bases
- Ensure a melting temperature ( $T_m$ ) of less than 75°C

The calculation of the  $T_m$  is based on Eurofins's equations as shown in Table 4.2.

Table 4.2 The equations for calculating the  $T_m$  of primers from Eurofins

Sequences with 15 or less bases	$T_m [^{\circ}C] = 2(n_A + n_T) + 4(n_G + n_C)$
Sequences with more than 15 bases	$T_m [^{\circ}C] = 69.3 + 41 \frac{(n_G + n_C)}{L} - \frac{650}{L}$

(The L indicates the length of a primer.  $n_A$ ,  $n_T$ ,  $n_G$ ,  $n_C$  are the base number of adenine, thymine, guanine and cytosine, respectively.)

The primers were ordered from Eurofins Genomics (Ebersberg, Germany) at HPSF purity and with a concentration at 50 pmol/ $\mu$ l. Their sequences were listed in Table 4.4 and Table 4.5.

## 4.2.5 Laboratory production

After obtaining the primers, Fab mutants were produced in the lab, which included:

- Laboratory mutagenesis
- 200 ml fermentation
- Purification
- Buffer exchange to water

For fermentation, one cycle usually lasts for one week. It was hoped that each Fab could be produced to a reasonable quantity (e.g. 30 mg) to allow initial screening for stability. Therefore, the DASbox® Mini Bioreactor (working volume 60-250 ml) (Eppendorf, Germany) was used as four different mutants could be run in parallel reactors.

For purification, a 1 ml HiTrap Protein G column was tested initially due to the small amount of Fab produced in a 200 ml fermentation. However, the binding capacity was not high enough and some Fab was lost. As a result, a 90 ml Protein G column was used, which could bind all the Fab but resulted in more diluted protein upon elution. In order to minimise aggregation during freezing and thawing, the eluted proteins were concentrated to 1-2 mg/ml first before storing in a -80°C freezer. A detailed procedure for the above can be found in Sections 2.2.1, 2.2.2 and 2.2.3 with scaled-down operating parameters.

## 4.3 Results and Discussion

### 4.3.1 Analysis of residue flexibility

#### 4.3.1.1 RMSF simulation in Gromacs

The RMSF data from molecular dynamics simulation of the pseudo wild-type C226S Fab are shown in Figure 4.3. It can be seen that the data deviate dramatically from different batches. The deviation is especially significant when the RMSF is of high value (e.g. the last tail region). Despite the large deviation, a general trend of fluctuation associated with secondary structures could be observed as shown in Figure 4.3 and it is relatively consistent at each repeat. In particular, loop regions are always of high RMSF while strand and helix regions are of relatively low RMSF values.

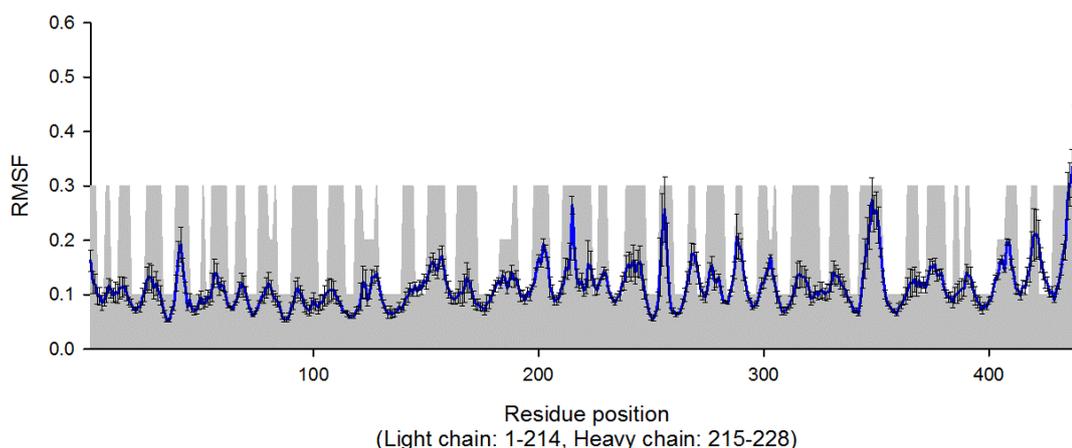


Figure 4.3 The overlay of C226S secondary structure and average RMSF at pH 4, 200 mM ionic strength

The standard error of the mean is used as the error bar of RMSF from three repeats of simulation. The secondary structure is determined by PyMol. The height of loop, helix, and strand regions are of height 0.3, 0.2 and 0.1, respectively.

#### 4.3.1.2 B-factor and its correlation with RMSF

The PDB IDs of 26 accessed human PDB files were: 1A6T, 1B2W, 1C5D, 1DFB, 1DN0, 1DQD, 1FGN, 1IT9, 1L7I, 1OPG, 1T3F, 2Z4Q, 2ZKH, 3D69, 3G6A, 3HC0, 3HI5, 3VG0, 4GSD, 4HBC, 4HH9, 4HIE, 4LKX, 4OCY, 4OSU, 7FAB.

The alignment results for C226S and those human Fab are shown in Table 10.1 and Table 10.2 (Appendix, Section 10.4) for heavy chain and light chain, respectively.

The normalised B-factors are plotted in Figure 4.4.

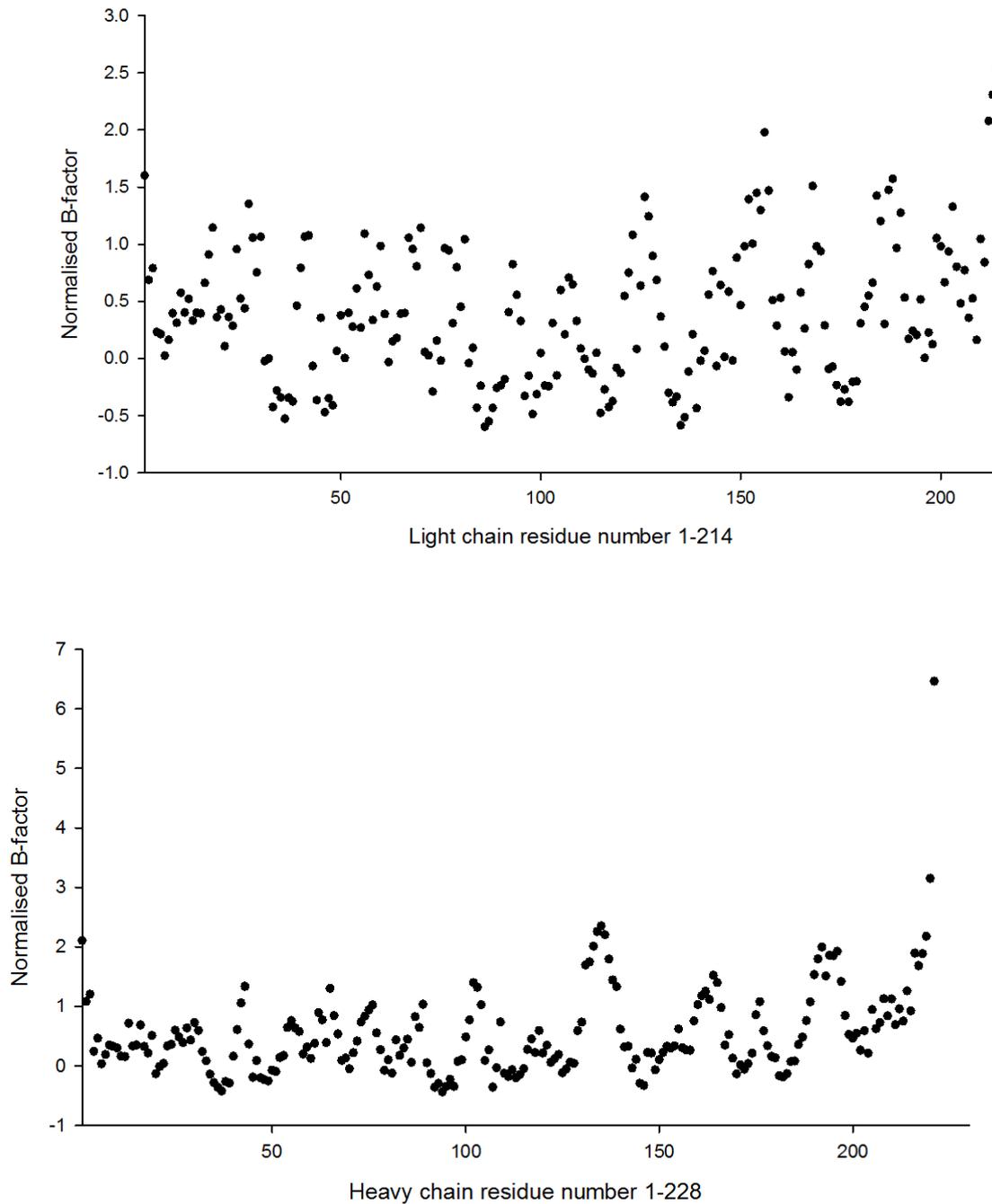


Figure 4.4 normalised B-factors of light chain and heavy chain

As shown in Figure 4.1, the residual B-factors were obtained by averaging their atomic B-factors, and tabulated along the amino acid sequence alignment. Normalisation was conducted for the tabulated "B-factor/sequence alignment" and the averaged B-factors accounting for C226S residues were retained.

The normalised B-factor after window averaging is shown in Figure 4.5.

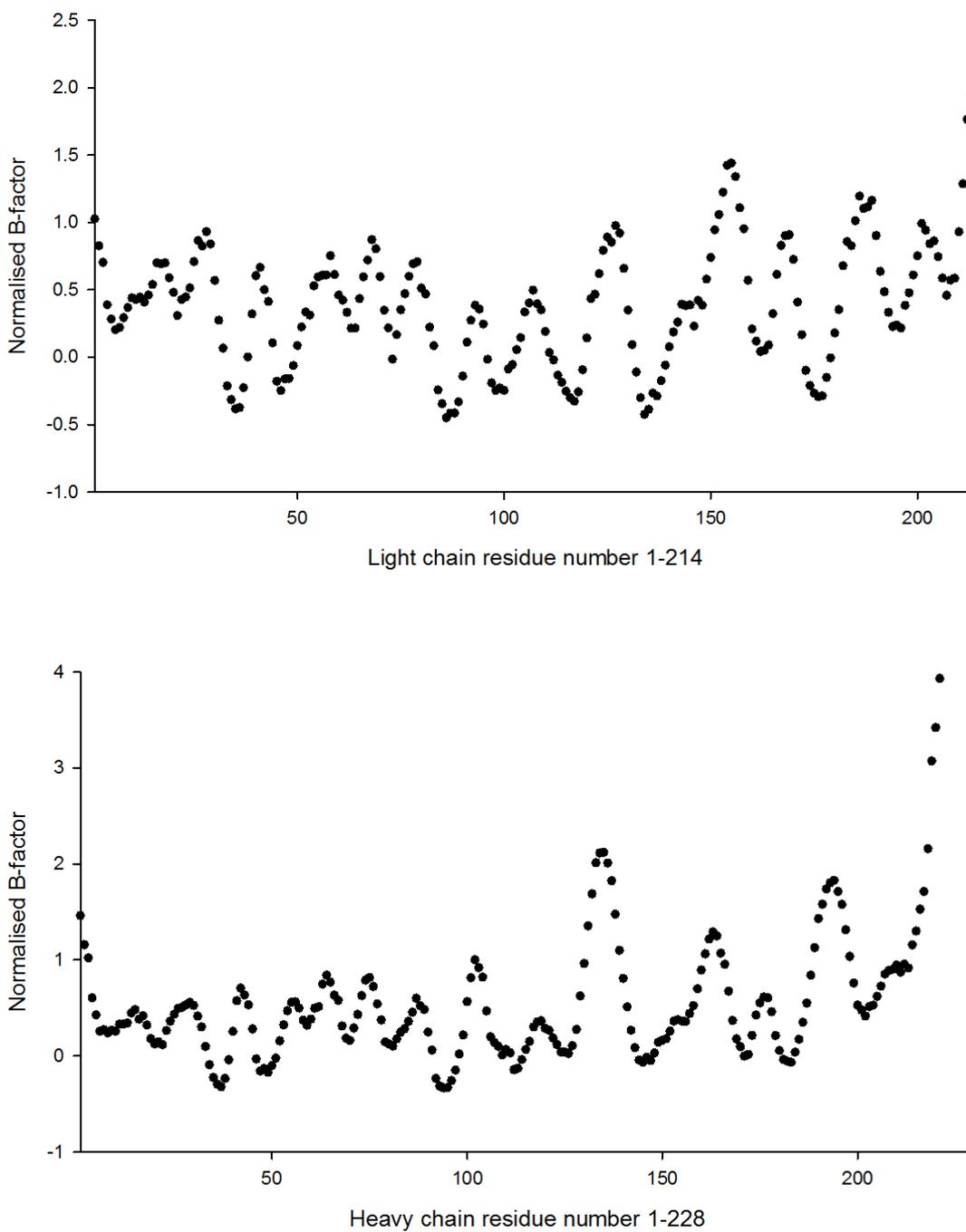


Figure 4.5 The normalised B-factor after window averaging with 5 neighbouring residues  
 Following the B-factor normalisation step shown in Figure 4.4, the B-factors were further window-averaged across 5 residues. The last 7 residues of heavy chain have no B-factor values as this part of sequence has no consensus with any of the other human Fab used for sequence alignment.

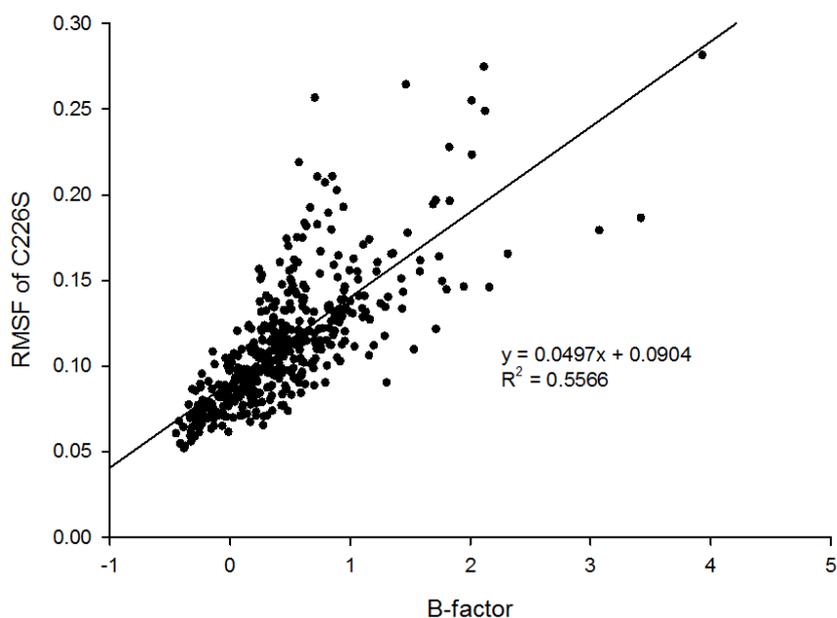


Figure 4.6 The correlation between RMSF and B-factor

The RMSF is derived from Figure 4.3; the B-factor is derived from Figure 4.5. The last 7 residues of heavy chain are not included as they have no B-factor values.

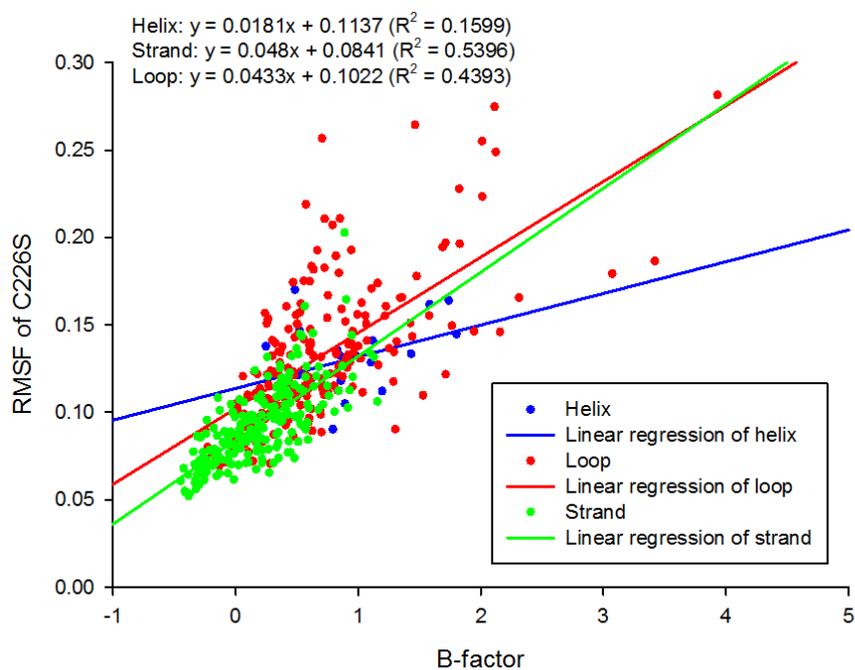


Figure 4.7 A breakdown of secondary structures within the correlation between RMSF and B-factor derived from Figure 4.6

As shown in Figure 4.5, the B-factor values from the human Fab structures had a similar sequence dependence as the RMSF. This implies that B-factors from the crystal

structures of Fab reflect a disorder *in vitro* that can be simulated by Gromacs *in silico*. In order to verify that both independent measurements achieve similar flexibility results for the protein residues, it was useful to find out if they correlated with each other to some extent. Figure 4.6 shows the correlation between RMSF and B-factor. It is shown that a 55.66% variation could be explained by the linear regression model. Therefore, the RMSF and B-factor values were statistically correlated.

Figure 4.7 further breaks the correlations down by secondary structure type. This indicates that the  $\beta$ -strand structures are the most rigid, whereas the significantly less common helical structure is relatively flexible. The loop regions cover almost the entire magnitude for RMSF and B-factor, but also show the highest flexibility in some locations. As discussed in Section 3.1.4, proteins with more  $\beta$ -sheet regions were more prone to form aggregates. However, strand structures are relatively less flexible than helical ones. Therefore, in addition to flexibility, the formation of aggregates also depends on the structural conformation. For the loop regions, it is interesting to see that only a small proportion of them exhibit significantly high flexibility, while the majority of them still retain low flexibility compared to strand and helix regions. This result indicated that a mutagenesis strategy could be applied to convert those residues with high flexibility in the loop regions into ones with low flexibility, based on a combination of RMSF and B-factor analyses.

## **4.3.2 *In-silico* mutagenesis**

### ***4.3.2.1 Analysis for the overall distribution and the most stabilising mutants for individual residues***

After *in-silico* mutagenesis, 8398 mutations were generated. Their  $\Delta\Delta G$  values (in arbitrary Rosetta Energy Units) are plotted in Figure 4.8 and their frequency distribution is plotted in Figure 4.9. It can be seen that most of the mutations had  $\Delta\Delta G$  values of close to 0, which implies that most single mutations exert only a limited impact on the protein stability. Moreover, negative  $\Delta\Delta G$  values were all greater than -9.4, whereas positive

$\Delta\Delta G$  values extended much further with a maximum value of 235. This result implies that the A33 C226S sequence already had a relatively stable form, which is not surprising given that it is already the result of significant selection and engineering as a potential therapeutic. However, there remained some room to further stabilise the protein even though the extent was very limited. As the  $\Delta G$  is around -1100, a -10  $\Delta\Delta G$  improvement would only provide less than 1% stabilising impact on the free energy of the global structure. On the contrary, the destabilising mutants may exhibit great detrimental effect with the maximal exerting more than 20% loss in  $\Delta G$ , which include polar or chargeable mutations in the protein core, hydrophobic mutations on the outer surface, hindrance caused by large amino acid substitutions like aromatic ones and disruption of salt bridge and hydrogen bond.

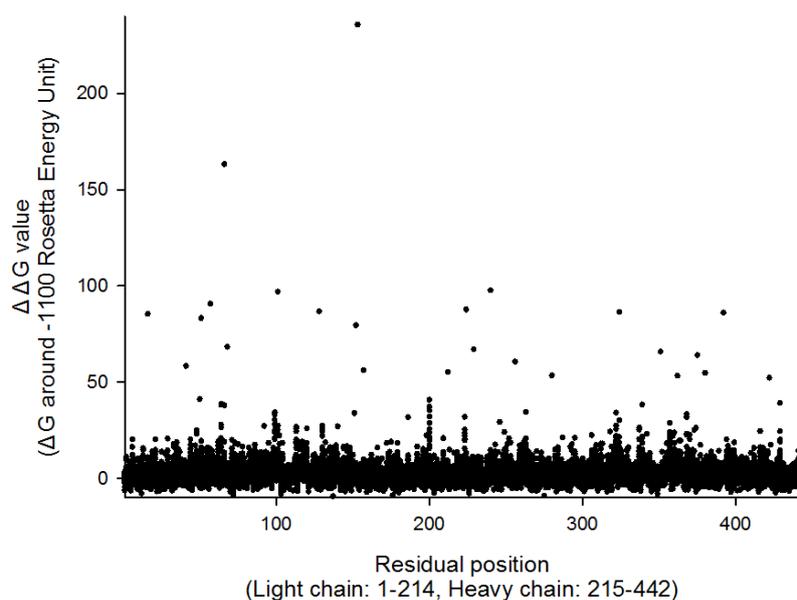


Figure 4.8  $\Delta\Delta G$  for 8398 (19 mutation/residue x 442 residue) candidates

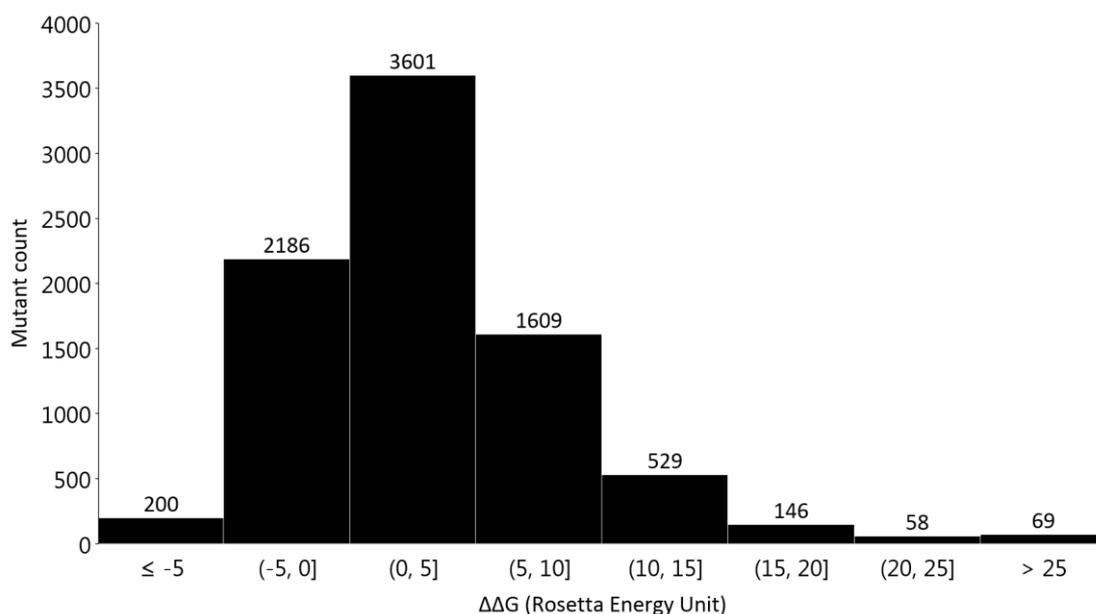


Figure 4.9 Histogram for the mutant frequency distribution based on  $\Delta\Delta G$

There were 2386 potentially stabilising mutants (i.e.  $\Delta\Delta G$  values lower than 0) as shown in Figure 4.9. To investigate whether they were biased towards particular secondary structure types, the  $\Delta\Delta G$  values were plotted by colour on the Fab structure in Figure 4.10. The blue regions in the figure represent stabilising effects, and the red ones indicate where no improvement could be made by any of the 19 mutations. As shown in Figure 4.10, most of the blue regions are located within  $\beta$ -sheet structures (e.g. LC-N137, LC-S176), while some were in the turn or random coil regions (e.g. HC-L61, HC-T135). As  $\beta$ -sheet is the most popular secondary structure pattern in this Fab, the stabilising locations do not show any statistical bias towards a particular secondary structure type.



Figure 4.10 A blue-white-red plot to represent the locations of stable mutants.

For each residual position, there are 19 mutations. The mutation with lowest  $\Delta\Delta G$  is retained. If the lowest  $\Delta\Delta G$  of that position is still greater than 0, value of 0 is used to indicate no improvement. Gradient colours from blue to white to red is used to demonstrate the  $\Delta\Delta G$  from lowest negative value to 0.

It was also interesting to investigate whether the stabilising mutants located more towards the outer surface regions, or within inner more buried regions. During the folding of a protein, the inner core is heavily influenced by the hydrophobic interactions and Van der Waals forces to reach a conformational stability, while the outer surface regions are relatively flexible, and surface charges offer additional colloidal stability. Figure 4.11 represents the depth of each residue in the C226S, with blue regions for lower depth and red regions for greater depth ([http://mspc.bii.a-star.edu.sg/tankp/run\\_depth.html](http://mspc.bii.a-star.edu.sg/tankp/run_depth.html)) (Tan et al. 2013). It is very clear that most of the residues are close to the bulk solvent while the inner  $\beta$ -sheet regions are well protected by their surrounding outer structures. Figure 4.12 correlates the lowest  $\Delta\Delta G$  of 19 mutants calculated by Rosetta with their depth for individual residues. As no correlation is observed, it implied that there was no significant bias in the location of stabilising mutants relative to the surface, and the stabilising mutants were spread comparably across the Fab structure. The poor correlation could come from the very rigid

conformation of Fab such that residues are already very stable in both the inner and outer regions. This also implied that there were many elements (e.g. electrostatics, hydrogen bonding) accounting for the free energy and they did not exert an overall bias upon the locations for this Fab.

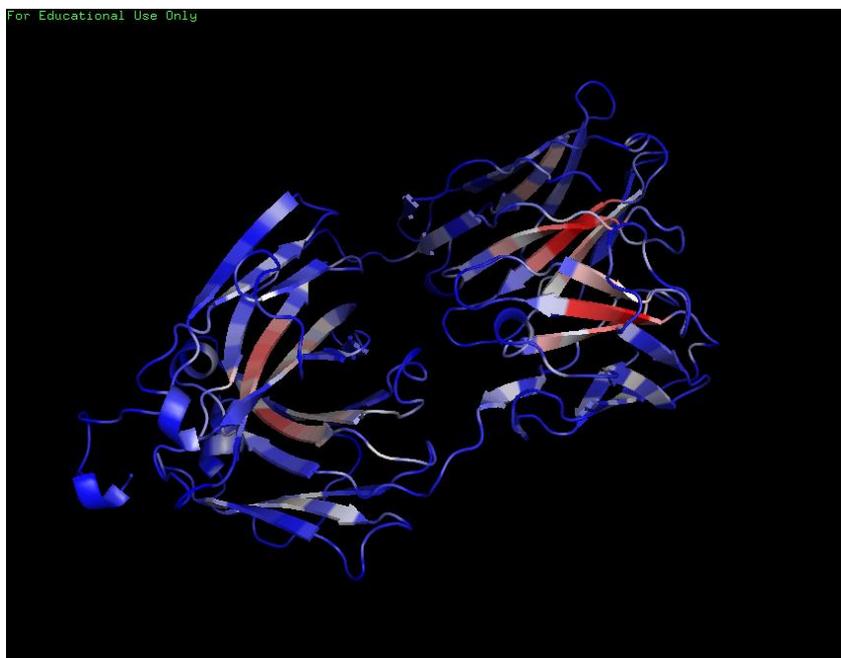


Figure 4.11 A blue-white-red plot to represent the residue depth of C226S Fab. Gradient colours from blue to white to red is used to demonstrate the residue depth from low to high as calculated from [http://mspc.bii.a-star.edu.sg/tankp/run\\_depth.html](http://mspc.bii.a-star.edu.sg/tankp/run_depth.html) with default settings.

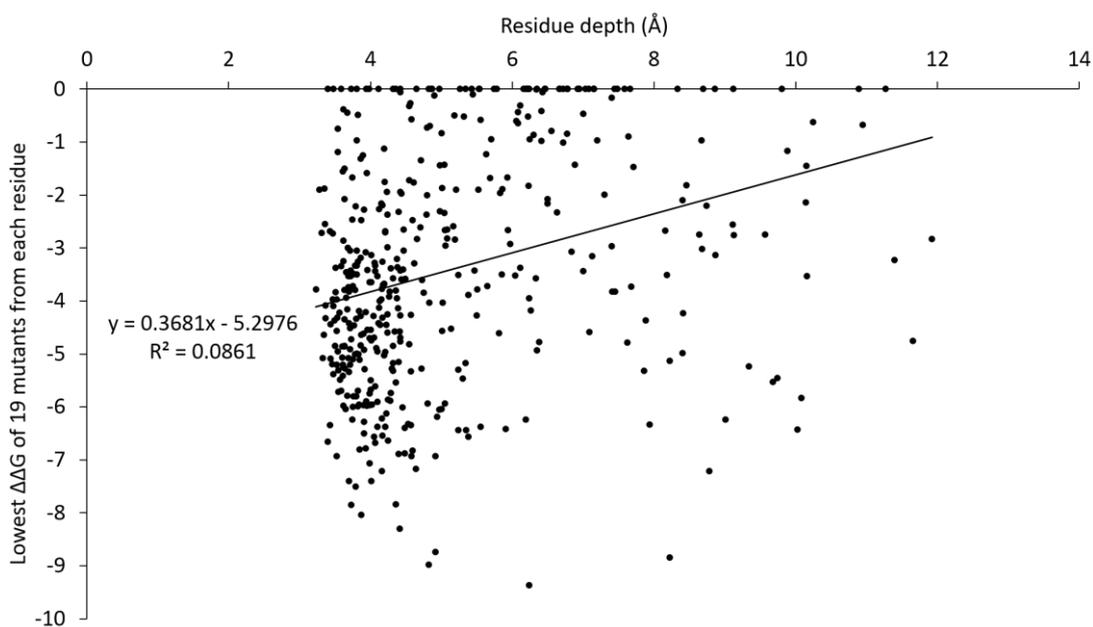


Figure 4.12 The correlation between lowest  $\Delta\Delta G$  of 19 mutants and the depth from bulk solvent for each residue.

The lowest  $\Delta\Delta G$  and residue depth were extracted from Figure 4.10 and Figure 4.11, respectively.

#### ***4.3.2.2 Analysis for all of the stabilising mutants ( $\Delta\Delta G < 0$ )***

Firstly, the distribution of the 20 amino acids in the C226S is shown from Figure 4.13 to Figure 4.15. Figure 4.13 summarises the occurrence of the 20 amino acids among the 442 residues. Figure 4.14 and Figure 4.15 list the average RMSF and depth for the C226S amino acids, respectively. Afterwards, all the stabilising mutants were populated in Figure 10.2, in which red to yellow gradient was used to indicate the strong to weak stabilising effect. The CDR, RMSF and depth information were also listed alongside with the residue so as to show their influence upon the stabilising effect proposed by Rosetta.

A clear opposite trend was observed between RMSF and depth. Figure 4.16 shows that this trend could be approximately fitted by a power equation with  $R^2$  of 0.41. This implies that the Fab residues on the outer regions generally have higher flexibility compared to the ones buried inside. Most of the residues locate at depths between 3-5 Å and with RMSF lower than 0.2. The residues become less flexible as they locate closer to the protein inner core, but still retain at least 0.05 RMSF flexibility. The residues on the tail of the heavy chain, which directly contact with bulk solvent, are the most flexible with RMSF values of more than 0.25.

The opposite relation between RMSF and depth was also reflected on the distribution of amino acids in C226S. As shown in Figure 4.14 and Figure 4.15, polar and chargeable amino acids, histidine and glycine have very high RMSF values, and they thus locate relatively close to the surface. This implies that residues with increased polarity are more prone to be near the surface region so as to provide colloidal stability, while residues with decreased polarity are more prone to be inside the protein to provide conformational stability through Van der Waals's force and hydrophobic interaction. Alanine also has a high RMSF but it distributes relatively deeper, which may due to its non-polar property.

RMSF values closely relate with the stabilising mutants proposed by Rosetta. As shown in Figure 4.17, all of the stabilising  $\Delta\Delta G$  were added up for individual residues with a window average of 9 neighbouring residues, and were superimposed with RMSF and CDR region. It can be seen that for some locations, regions with increased RMSF have decreased total  $\Delta\Delta G$  values (e.g. residues 39-44, 167-169, 198-206, 254-258, 287-289, 306-307, 346-351), while regions with decreased RMSF have increased total  $\Delta\Delta G$  values (e.g. residues 115-121, 293-296, 340-343, 359-361). This implies that for some particularly flexible locations, it is a promising strategy to mutate towards residues with less flexibility. However, not all highly flexible regions would benefit from further mutation. For example, there were few stabilising mutational recommendations for regions of increased flexibility such as at residues 79-81, 92-94, 214-216, 276-277, 420-422. As checked in the PDB, those locations are all loop regions except for residues 79-81 which are in a  $\beta$ -turn structure. This implies that the Fab actually requires several flexible regions to properly connect rigid structures like  $\beta$ -sheets and to easily adapt to any environmental disturbance without altering the overall conformation.

Similarly, it is also interesting to observe in Figure 4.17 that although all of the CDR regions (except CDR-L2) exhibit increased flexibility, only very few stabilising mutants were suggested for CDR-L2, L3, H1. This implies that the flexibility of CDR regions does not impact the global conformational stability to a large extent. This is beneficial as those regions could be more extensively engineered based on the target antigen, and the Fab could appropriately adjust their conformation upon binding without losing the overall structural rigidity of the protein.

It is surprising to see that a few low flexibility regions with relatively high depth could potentially be stabilised further, such as at residues 70-72, 103 and 235-237. The PDB structure shows that these are all  $\beta$ -sheet structures. Phenylalanine and tyrosine are predicted to be most stabilising for residues 70, 72 and 103, which suggests that they might be better able to fill the cavity, or offer stronger hydrophobic interactions. Residue 236 is cysteine, and yet other non-polar substitutions, glycine, proline and methionine, were preferred. Since cysteine was strongly suggested at neighbouring positions 235

and 237, it implies that these two positions may offer a more stable intra-chain disulfide bond.

Figure 4.18 to Figure 4.22 subdivide the stabilising effect based on the type of amino acid mutations. Figure 4.18 sums up all the stabilising  $\Delta\Delta G$  from the 8398 mutants for each residue mutation. For Figure 4.19 and Figure 4.20, weighted values of RMSF and depth are used to reflect the locations of different stabilising amino acid substitution. The weighted values were calculated as shown in Figure 4.21. It aimed to capture the individual  $\Delta\Delta G$  contribution at various RMSF or depth. For example, if the particular amino acid always stabilises the Fab at a high RMSF location, its weighted RMSF would also be high. Figure 4.22 summarises the extent of the residues in the C226S stabilised by point mutations based on their  $\Delta\Delta G$ .

Figure 4.18 shows that aromatic amino acids accounted for most of the stabilising substitutions, followed by methionine, aliphatic amino acids, cysteine, and polar amino acids, while glycine, proline and ionisable ones were the least stabilising. Figure 4.22 shows that aromatic and aliphatic amino acids stabilise wild-type serine and threonine residues to a large extent. The large sidechains of aromatic residues greatly limit their mobility in solution (Gunasekaran & Nussinov 2007) and thus are preferable to stabilise the protein structure. Though aliphatic ones are not as rigid as aromatic ones, aliphatic residues are still beneficial to replace flexible residues serine and threonine near the surface due to their non-polar property.

The ionisable residues contribute relatively less compared to the others with aspartic acid the least favourable (Figure 4.18). The reason might be due to the positive or negative charge on the amido or carboxyl side chain. Those charges may exert some electrostatic repulsion within the structure, increase the residue distance and weaken the hydrophobic interaction. Lysine and arginine, the two positively charged amino acids, are preferable near the surface while aspartic acid and glutamic acid, the two negatively charged ones, are more likely to locate deeper (Figure 4.20). This suggests that positive surface charge could provide some colloid stability as the pI of the Fab is around 9, while the negatively charged substitutions would neutralise the net charge if placed at

surface regions. It shows that all of the four charged substitutions locate at highly flexible regions (Figure 4.19), and could reasonably stabilise serine and threonine (Figure 4.22). This implies that salt-bridges might be good for lowering the flexibility (Yu & Huang 2014; Amini-Bayat et al. 2012), in which buried negatively charged residues non-covalently interact the positively charged ones near the surface.

Glycine and proline are two of the three least preferred substitutions. As  $\beta$ -sheet is the main secondary structure in the Fab, it is not surprising that glycine and proline, which disrupt  $\beta$ -sheets (discussed in Section 3.1.1 and 3.1.2), are less favoured. As confirmed in the PDB, all of the glycine and proline mutations with  $\Delta\Delta G$  less than -4, locate at loop or  $\beta$ -turn regions while the exceptions are cysteines at the  $\beta$ -sheet. In addition, Figure 4.20 shows that glycine tends to locate at the inner core while proline tends to be near the surface. This depth difference mainly comes from their structure difference. Glycine is the only amino acid without a side chain, so it is more likely to adapt to the inner compact hydrophobic core of protein without causing structural disruption. By contrast, proline is very rigid so it has to be near the surface so as to have minimal disruptive influence on its surrounding residues.

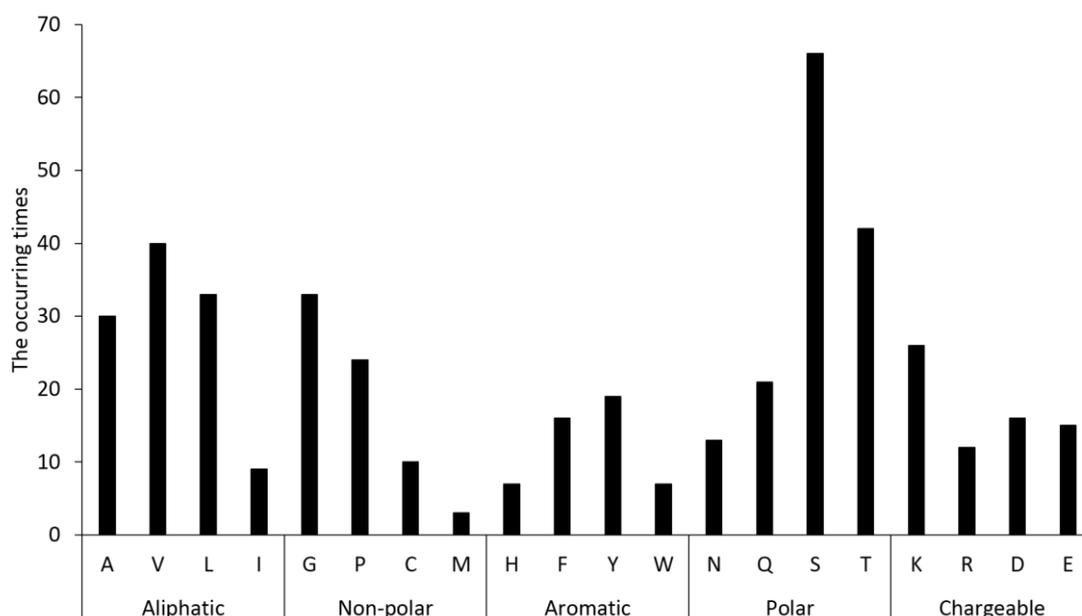


Figure 4.13 The occurring times for the 20 amino acids among the 442 residues in the C226S. The 20 natural amino acids were categorised into 5 groups. The occurring time of each type of amino acid within the 442 residues was shown in the figure.

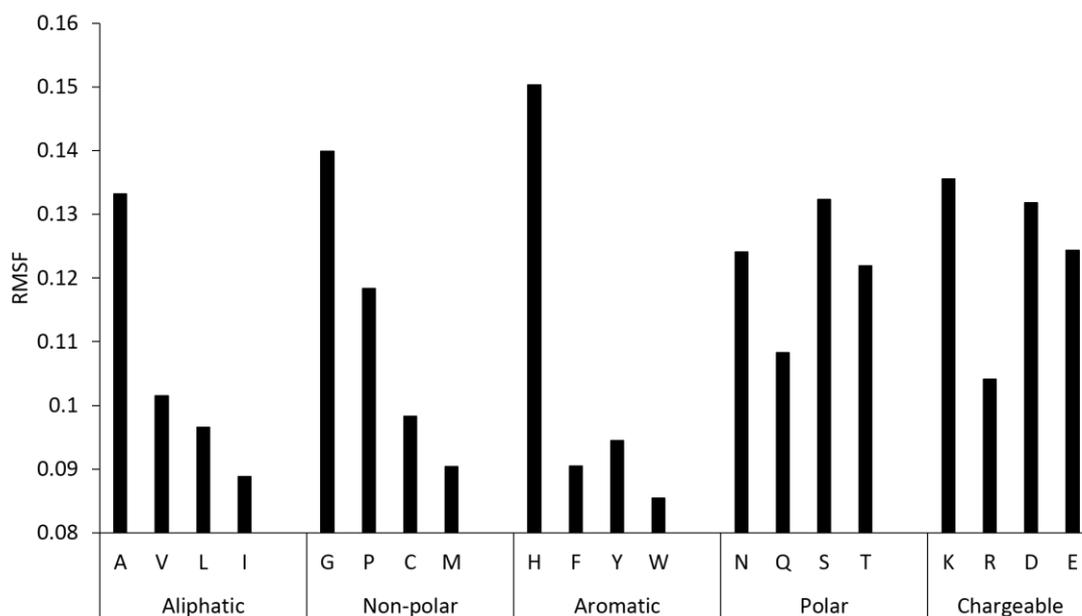


Figure 4.14 The average RMSF of individual residues of C226S derived from Figure 4.3. The 20 natural amino acids were categorised into 5 groups. Their average RMSF were calculated from Figure 4.3.

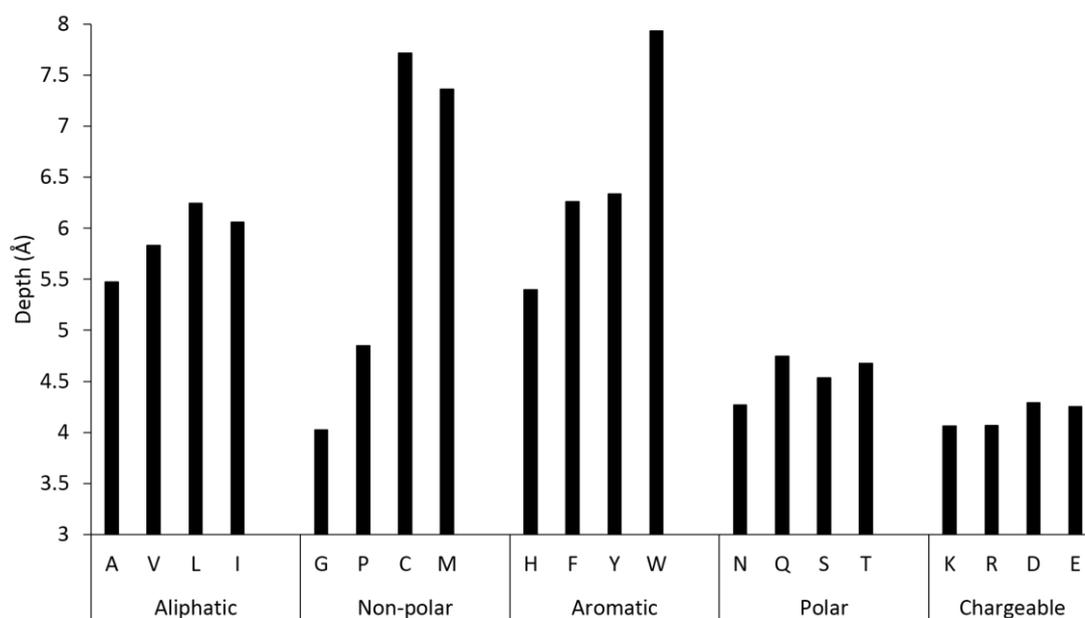


Figure 4.15 The average depth of individual residues of C226S derived from Figure 4.11. The 20 natural amino acids were categorised into 5 groups. Their average depth was calculated from Figure 4.11.

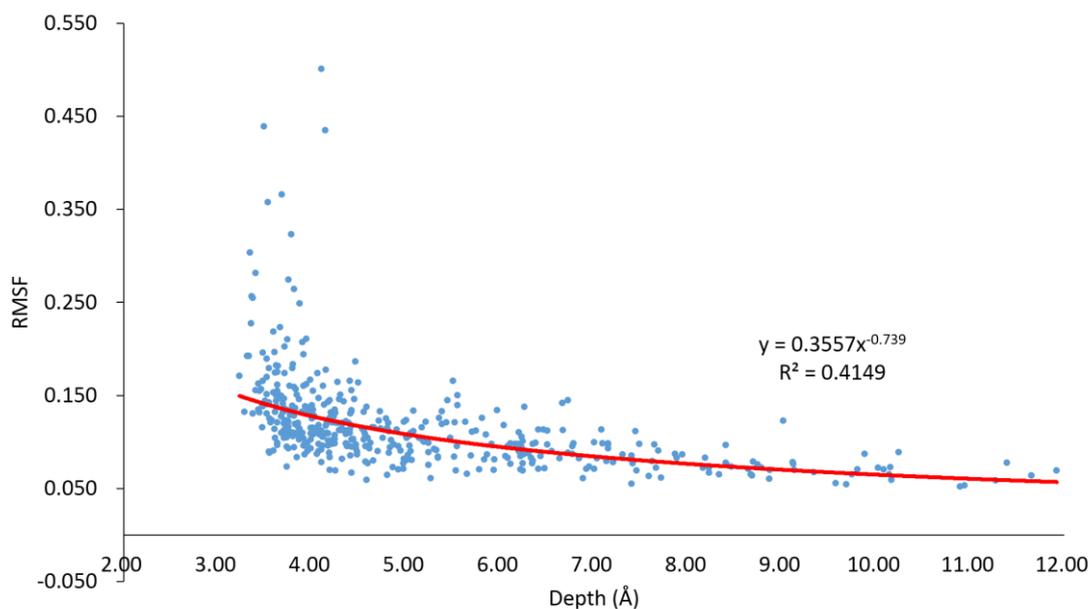


Figure 4.16 The correlation between RMSF (Figure 4.3) and depth (Figure 4.11) of C226S

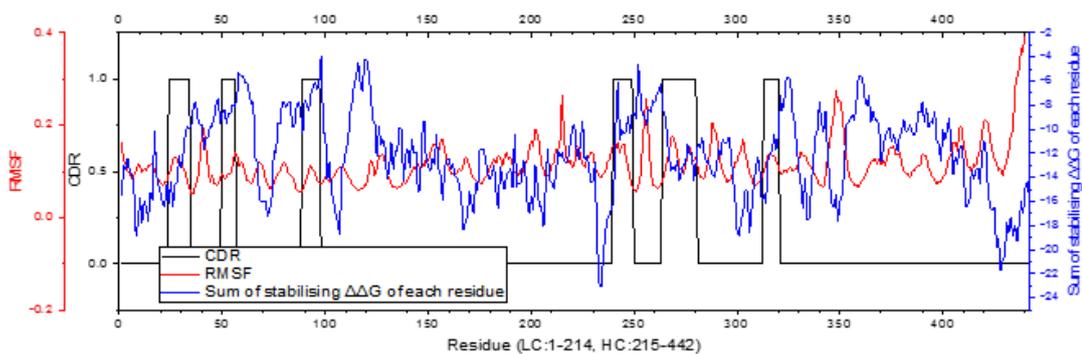


Figure 4.17 The relation between RMSF, CDR and sum of the stabilising  $\Delta\Delta G$  of each residue. Data was derived from Figure 10.2. A window average of 9 neighbouring residues was used for the sum of the stabilising  $\Delta\Delta G$  of each residue. The CDR regions were assigned with value 1 while non-CDR regions were of value 0.

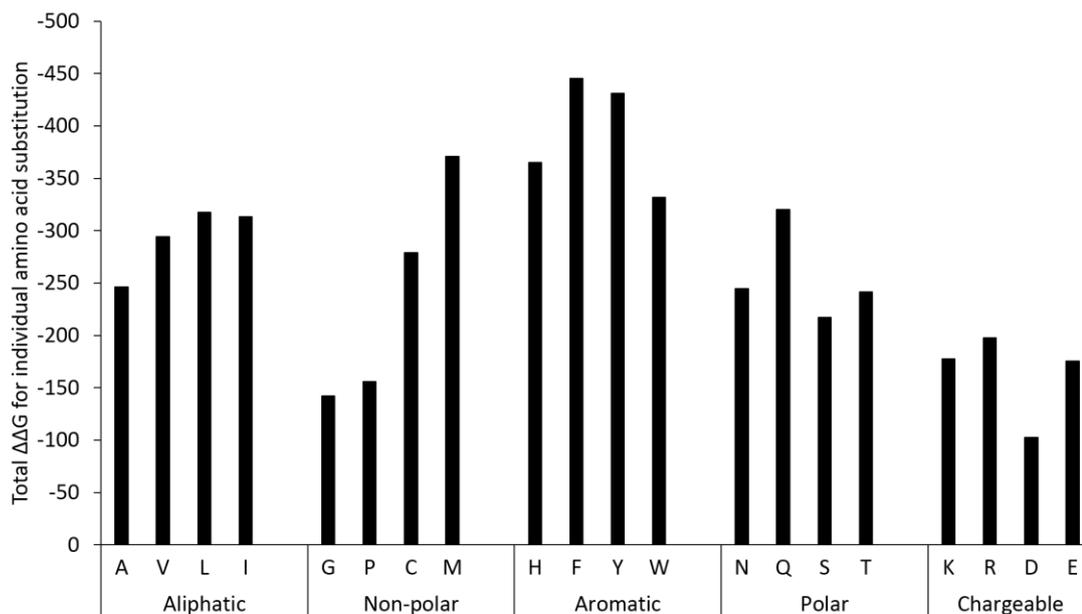


Figure 4.18 The total  $\Delta\Delta G$  for stabilising mutants grouped by the type of amino acids derived from Figure 10.2.

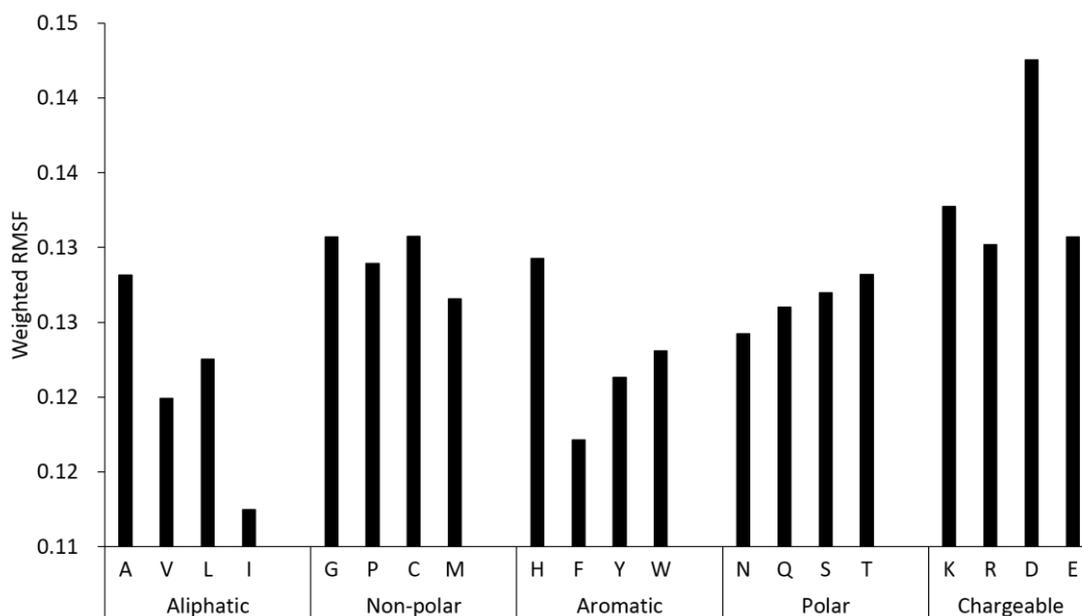


Figure 4.19 The weighted RMSF of individual amino acid substitutions based on their  $\Delta\Delta G$  stabilising effect

Data was derived from Figure 10.2, calculated as shown in Figure 4.21.

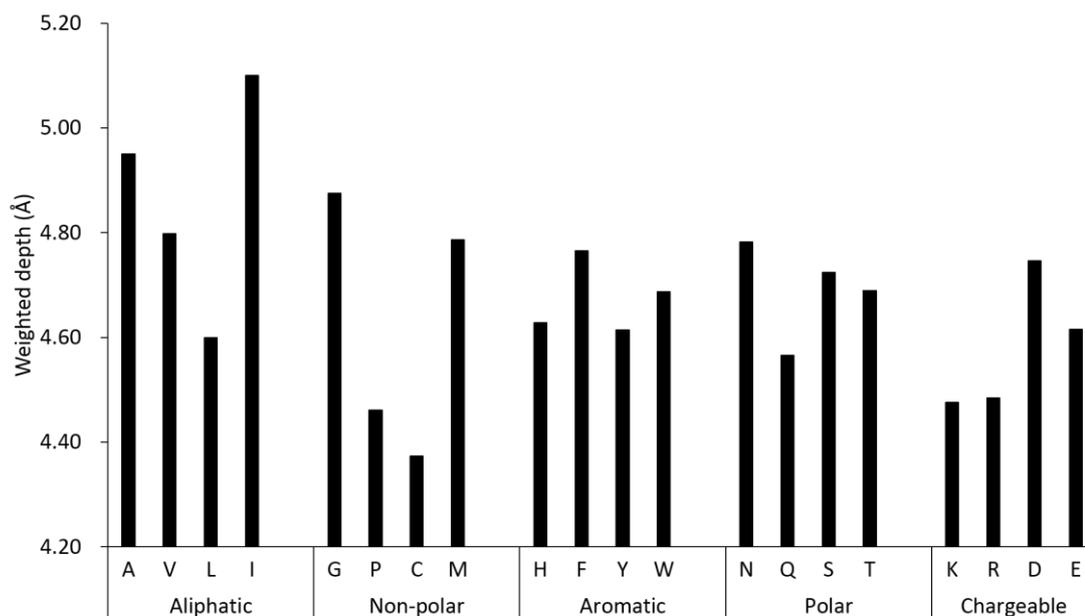


Figure 4.20 The weighted depth of individual amino acid substitutions based on their  $\Delta\Delta G$  stabilising effect

Data was derived from Figure 10.2, calculated as shown in Figure 4.21.

Residue	RMSF or Depth	$\Delta\Delta G$ from a amino acid substitution	Weight
1	$A_1$	$B_1$	$C_1$
2	$A_2$	$B_2$	$C_2$
3	$A_3$	$B_3$	$C_3$
...	...	...	...
n	$A_n$	$B_n$	$C_n$

n: the total number of residues (i.e. 442)

$$\text{Sum} = \sum_{i=1}^n B_i$$

$$C_i = B_i / \text{Sum}$$

$$\text{Weighted value} = \sum_{i=1}^n A_i \times C_i$$

Figure 4.21 An illustration for the calculation for the weighted RMSF and weighted depth in Figure 4.19 and Figure 4.20 derived from Figure 10.2.

		Target mutation																				Sum	
		Aliphatic				Non-polar				Aromatic				Polar				Chargeable					
		A	V	L	I	G	P	C	M	H	F	Y	W	N	Q	S	T	K	R	D	E		
Wild type residue	Aliphatic	A	0.00%	0.52%	0.41%	0.18%	0.31%	0.23%	0.45%	0.56%	0.60%	0.53%	0.40%	0.21%	0.26%	0.62%	0.34%	0.51%	0.11%	0.08%	0.30%	0.29%	6.94%
		V	0.19%	0.00%	0.35%	0.57%	0.03%	0.09%	0.28%	0.38%	0.35%	0.59%	0.54%	0.44%	0.15%	0.24%	0.26%	0.35%	0.22%	0.17%	0.10%	0.08%	5.38%
		L	0.16%	0.36%	0.00%	0.22%	0.00%	0.10%	0.17%	0.37%	0.28%	0.60%	0.35%	0.12%	0.15%	0.16%	0.07%	0.12%	0.00%	0.09%	0.00%	0.02%	3.36%
		I	0.01%	0.00%	0.06%	0.00%	0.00%	0.00%	0.11%	0.00%	0.02%	0.08%	0.10%	0.00%	0.00%	0.05%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.42%
	Non-polar	G	0.25%	0.22%	0.30%	0.03%	0.00%	0.06%	0.18%	0.26%	0.46%	0.32%	0.22%	0.18%	0.40%	0.24%	0.22%	0.17%	0.17%	0.26%	0.06%	0.19%	4.19%
		P	0.12%	0.14%	0.09%	0.05%	0.04%	0.00%	0.00%	0.15%	0.20%	0.04%	0.22%	0.02%	0.02%	0.09%	0.09%	0.31%	0.02%	0.04%	0.03%	0.08%	1.73%
		C	0.39%	0.20%	0.22%	0.35%	0.34%	0.26%	0.00%	0.32%	0.33%	0.21%	0.37%	0.22%	0.48%	0.19%	0.47%	0.30%	0.29%	0.30%	0.24%	0.32%	5.80%
		M	0.00%	0.07%	0.09%	0.15%	0.00%	0.00%	0.00%	0.00%	0.03%	0.01%	0.00%	0.00%	0.01%	0.00%	0.07%	0.04%	0.00%	0.00%	0.00%	0.00%	0.47%
	Aromatic	H	0.03%	0.01%	0.00%	0.03%	0.00%	0.00%	0.00%	0.02%	0.00%	0.08%	0.10%	0.03%	0.00%	0.08%	0.06%	0.05%	0.02%	0.00%	0.00%	0.03%	0.55%
		F	0.00%	0.04%	0.04%	0.00%	0.00%	0.00%	0.01%	0.01%	0.05%	0.00%	0.08%	0.03%	0.00%	0.00%	0.03%	0.00%	0.00%	0.00%	0.07%	0.00%	0.36%
		Y	0.00%	0.04%	0.09%	0.07%	0.00%	0.00%	0.00%	0.00%	0.06%	0.38%	0.00%	0.15%	0.01%	0.05%	0.00%	0.02%	0.00%	0.00%	0.00%	0.01%	0.89%
		W	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.03%	0.00%	0.00%	0.04%	0.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.11%
	Polar	N	0.10%	0.40%	0.31%	0.26%	0.06%	0.08%	0.07%	0.32%	0.23%	0.09%	0.41%	0.25%	0.00%	0.53%	0.05%	0.09%	0.12%	0.15%	0.13%	0.31%	3.94%
		Q	0.14%	0.21%	0.28%	0.25%	0.11%	0.26%	0.25%	0.29%	0.25%	0.20%	0.24%	0.22%	0.19%	0.00%	0.15%	0.19%	0.09%	0.13%	0.07%	0.15%	3.67%
		S	1.05%	0.79%	0.86%	0.97%	0.69%	0.82%	1.37%	1.78%	1.49%	1.67%	1.93%	1.70%	1.02%	1.27%	0.00%	0.91%	0.96%	0.88%	0.31%	1.00%	21.48%
		T	0.77%	0.98%	1.07%	1.17%	0.25%	0.22%	0.83%	0.72%	0.79%	1.09%	0.92%	1.20%	0.60%	0.64%	0.70%	0.00%	0.78%	0.61%	0.16%	0.27%	13.78%
	Chargeable	K	0.57%	0.70%	0.45%	0.53%	0.19%	0.15%	0.53%	0.46%	0.58%	1.06%	1.12%	0.58%	0.29%	0.62%	0.63%	0.42%	0.00%	0.44%	0.33%	0.18%	9.84%
		R	0.19%	0.32%	0.53%	0.30%	0.13%	0.26%	0.18%	0.38%	0.27%	0.14%	0.21%	0.11%	0.18%	0.28%	0.31%	0.28%	0.14%	0.00%	0.00%	0.22%	4.46%
		D	0.29%	0.26%	0.39%	0.29%	0.27%	0.23%	0.36%	0.55%	0.33%	0.64%	0.32%	0.35%	0.49%	0.59%	0.39%	0.43%	0.30%	0.25%	0.00%	0.11%	6.86%
		E	0.33%	0.22%	0.37%	0.40%	0.22%	0.14%	0.36%	0.35%	0.45%	0.49%	0.48%	0.35%	0.31%	0.30%	0.22%	0.32%	0.10%	0.28%	0.11%	0.00%	5.79%
Sum		4.59%	5.48%	5.91%	5.83%	2.65%	2.90%	5.20%	6.91%	6.80%	8.30%	8.03%	6.18%	4.56%	5.96%	4.05%	4.50%	3.31%	3.68%	1.91%	3.26%	100.00%	

Figure 4.22 A summation for the stabilising  $\Delta\Delta G$  based on the type of residues

Data was derived from Figure 10.2. Each row represents the total  $\Delta\Delta G$  when a particular amino acid in the C226S is mutated to the other 19 residues. Each column represents the total stabilising  $\Delta\Delta G$  of the 19 residues in the C226S that are mutated to another amino acid. Gradient colours are used to indicate the magnitude of the  $\Delta\Delta G$ .

#### 4.3.2.3 Analysis for the destabilising mutants ( $\Delta\Delta G > 0$ )

There were 6012 potentially destabilising mutants as shown in Figure 4.9. The 25 mutants with the highest  $\Delta\Delta G$  values are displayed in Table 4.3. As shown in the table, all of them are mutated into proline and 19 of them are glycine to proline mutations. As glycine has no  $\beta$ -carbon this residue would exert more backbone conformation flexibility. Proline, on the contrary, is rather rigid due to its unique side chain pyrrolidine. The stability of X to P mutations depends largely on the secondary structures. If proline is in the middle of an  $\alpha$ -helix or  $\beta$ -sheet, it tends to break down the secondary structure and destabilise the protein. As shown in Figure 4.23, most of the mutations were in the loop regions. So the destabilising effect is not mainly due to proline's structural disruption tendency. As loop regions require relatively high flexibility to connect two ordered structures, the instability caused by glycine to proline might arise from proline's rigidity that disables the flexibility for folding offered by glycine.

Table 4.3 The top 25 destabilising mutants

Mutant_ID	$\Delta\Delta G$
LC-A153P	235.72
LC-G66P	163.28
HC-G26P	97.66
LC-G101P	96.961

LC-G57P	90.729
HC-G10P	87.695
LC-G128P	86.686
HC-G110P	86.35
HC-G178P	86.001
LC-G16P	85.343
LC-A51P	83.233
LC-N152P	79.539
LC-G68P	68.295
HC-G15P	66.995
HC-G137P	65.807
HC-G161P	63.958
HC-G42P	60.577
LC-G41P	58.336
LC-G157P	56.228
LC-G212P	55.195
HC-G166P	54.732
HC-G66P	53.383
HC-D148P	53.251
HC-N208P	52.194
LC-L50P	41.124

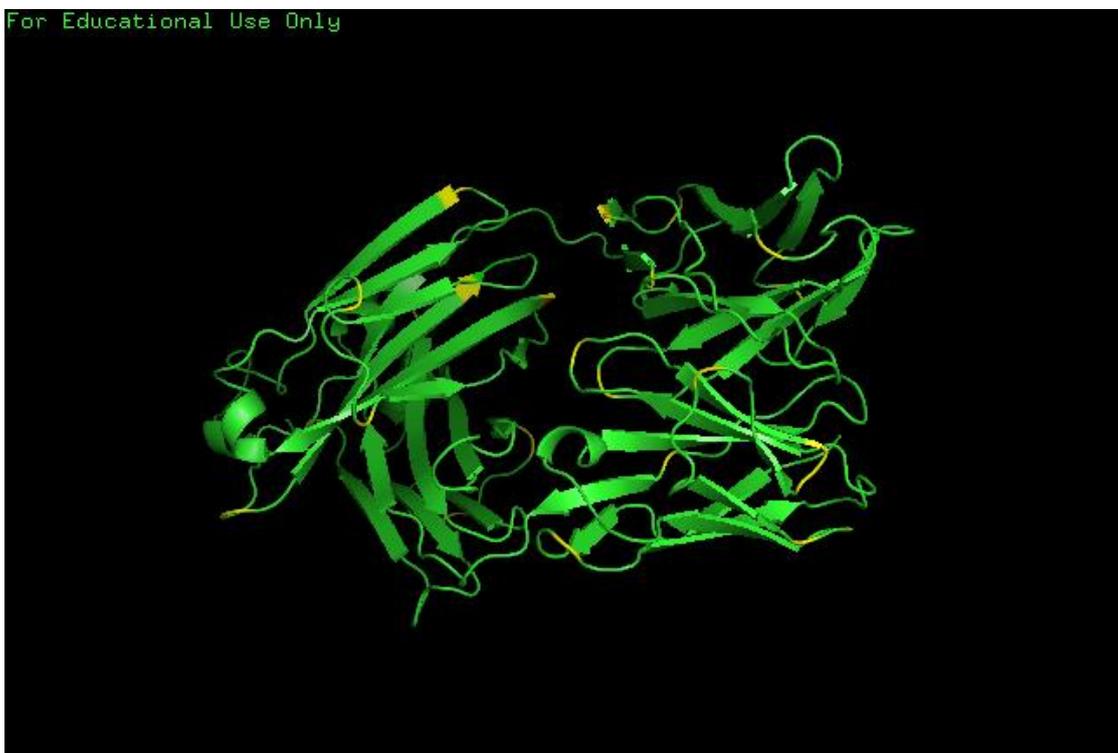


Figure 4.23 The positions of the top 25 destabilising mutations  
The highest  $\Delta\Delta G$  values are achieved by mutating the yellow residues into prolines

### 4.3.3 Designed mutants and corresponding primer sequences

#### 4.3.3.1 Designed stable mutants

The designed stable mutants and their primer sequences are shown in Table 4.4.

The three oligonucleotides responsible for mutations are highlighted in red colour.

Table 4.4 The designed stable mutants and corresponding primer sequences

		Primer Sequence (5' to 3')		Complementary primer (5' to 3')
Mutants based on B-factor, RMSF & $\Delta\Delta G$	Hinge regions	HC-A227	HC-A227T CAAAACCTCACACAAGC <b>ACC</b> GCCTGATGAGGATCC	GGATCCTCATCACCGCGTGTGTGAGTTTTG
		HC-A227E	GACAAAACCTCACACAAGC <b>GAA</b> GCCTGATGAGGATCCAAGC	GCTTGGATCCTCATCACGCTTCGCTTGTGTGAGTTTTGTC
		HC-A227W	CAAAAACCTCACACAAGC <b>TGG</b> GCCTGATGAGGATCC	GGATCCTCATCACGCCAGCTTGTGTGAGTTTTG
		HC-A228H	CTCACACAAGCGCC <b>ATT</b> GATGAGGATCCAAG	CTTGGATCCTCATCAATGGCGCTTGTGTGAG
		HC-A228N	CTCACACAAGCGCC <b>AAT</b> GATGAGGATCCAAG	CTTGGATCCTCATCAGTTGGCGCTTGTGTGAG
		HC-A228M	CTCACACAAGCGCC <b>ATG</b> TATGAGGATCCAAG	CTTGGATCCTCATCACATGGCGCTTGTGTGAG
	Other regions	HC-T135W	CCTCCAAGAGC <b>TGG</b> TCTGGGGGCAC	GTGCCCCAGACCAGCTCTTGGAGG
		HC-T135I	CCTCCTCCAAGAGC <b>ATT</b> TCTGGGGGCACAGC	GCTGTGCCCCCAGAATGCTCTTGGAGGAGG
		HC-T135Y	CCTCCTCCAAGAGC <b>TAT</b> TCTGGGGGCACAGC	GCTGTGCCCCCAGAATAGCTCTTGGAGGAGG
		HC-S134Y	CACCCTCCTCCAAG <b>TAT</b> ACCTCTGGGGGCAC	GTGCCCCAGAGGTATACCTTGGAGGAGGGTG
		HC-S134M	CCTCCTCCAAG <b>ATG</b> ACCTCTGGGGGCAC	GTGCCCCAGAGGTATCTTGGAGGAGG
		HC-S134P	CACCCTCCTCCAAG <b>CCG</b> ACCTCTGG	CCAGAGGTCGGCTTGGAGGAGGGTG
		HC-S136G	CAAGAGCACCC <b>GGC</b> GGGGGCACAGC	GCTGTGCCCCCAGCGGTGCTCTTG
		HC-S136Y	CTCCTCCAAGAGCAC <b>TAT</b> GGGGGCACAGC	GCTGTGCCCCCATAGGTGCTCTTGGAGGAG
		HC-S136W	CCAAGAGCAC <b>TGG</b> GGGGGCACAGC	GCTGTGCCCCCAGGTGCTCTTGG
		HC-S219Y	GACAAGAAAGTTGAGCCCA <b>AAATAT</b> TGTGACAAAACCTCACACAAGC	GCTTGTGTGAGTTTTGTACAAATATTGGGGCTCAACTTTCTTGTG
		HC-S219R	GACAAGAAAGTTGAGCCCA <b>AAACCG</b> TGTGACAAAACCTCACACAAGC	GCTTGTGTGAGTTTTGTACAGCGTTTGGGGCTCAACTTTCTTGTG
		HC-S219L	GACAAGAAAGTTGAGCCCA <b>AACTGT</b> TGTGACAAAACCTCACACAAGC	GCTTGTGTGAGTTTTGTACACAGATTTGGGGCTCAACTTTCTTGTG
		LC-L154A	GAAGGTGGATAACGC <b>CCG</b> CAATCGGGTAACTC	GAGTTACCCGATTGCGCGCGGTTATCCACCTTC
		LC-L154M	GGAAGGTGGATAACGC <b>ATG</b> CAATCGGGTAACTCC	GGAGTTACCCGATTGCATGGCGTTATCCACCTTCC
LC-L154P	GAAGGTGGATAACGC <b>CCG</b> CAATCGGGTAACTC	GAGTTACCCGATTGCGGGCGGTTATCCACCTTC		
Mutants only based on $\Delta\Delta G$	LC-N137L	LC-N137L GTTGTGTGCGCTGCT <b>GCTGA</b> ACTTCTATCCCAGAG	CTCTGGGATAGAAGTTACAGCAGCAGGCACACAAC	
	LC-S176W	LC-S176W GCACCTACAGCCTC <b>TGG</b> AGCACCCCTGACG	CGTCAGGGTCTCCAGAGGCTGATAGTGTC	
	LC-T72Y	LC-T72Y GGTAGTGGTACTGATTT <b>CTAT</b> CTGACTATCAGTAGTCTCCAGC	GCTGGAGACTACTGATAGTCAGATAGAAATCAGTACCACTACC	
	LC-S12F	LC-S12F CTCAGAGTCCAAGTAGTCT <b>TTT</b> GCTAGTGTAGGTGATAGG	CCTATCACCTACACTAGCAAAGAGACTACTTGGACTCTGAG	

#### 4.3.3.2 Designed unstable mutants

The designed unstable mutants and their primer sequences are shown in Table 4.5.

The three oligonucleotides responsible for mutations are highlighted in red colour. It was found that the top 25 mutants with highest  $\Delta\Delta G$  were all mutated into proline, and 14 of them were glycine to proline mutants. In order to cover other types of mutations, the mutants ranked after the first 25 were selected as well as shown in Table 4.5.

Table 4.5 The designed unstable mutants and corresponding primer sequences

$\Delta\Delta G$	Primer Sequence 5' to 3'	Complement 5' to 3'
LC-A153P 236	GGAAGGTGGATAAC <b>CCG</b> CTCCAATCGGGTAAC	GTTACCCGATTGGAGCGGGTTATCCACCTTCC
LC-G66P 163	CTAGATTCAGTGGTAGC <b>CCG</b> AGTGGTACTGATTTACAC	GTGAAATCAGTACCACCTCGGGCTACCACTGAATCTAG
LC-G101P 97	CCTCTCACGTTCCGGT <b>CCG</b> ACTAAAGTAGAAATCAAAC	GTTTGTATTCTACTTTAGTCCGGCTGACCGAACCTGAGAGG
LC-G200W 40.8	GTCACCCATCAG <b>TGG</b> CTGAGCTCACCAG	CTGGTGTGAGCTCAGCCACTGATGGGTGAC
HC-V215W 39.1	CCAAGGTGCACAAGAA <b>TGG</b> GAGCCCAAATCTTGTG	CACAAGATTTGGGCTCCCAATTTCTTGTCCGACCTTGG
LC-G64F 38.7	GGAGTACCATCTAGATT <b>CTT</b> AGCGGTAGTGGTACTG	CAGTACCACTACCCTAAAACCTGAATCTAGATGGTACTCC
LC-G64Y 38.6	GGAGTACCATCTAGATT <b>TAT</b> AGCGGTAGTGGTACTG	CAGTACCACTACCCTATAACTGAATCTAGATGGTACTCC
HC-V125W 38.3	GAAGGGCCATCG <b>TGG</b> TTCCCCCTGG	CCAGGGGAACCACGATGGGCCCTTC

### 4.3.4 Laboratory production

The primers were designed as shown in Table 4.4 and Table 4.5. Some primers (e.g. HC-S219L, LC-N137L) were not able to provide successful site direct mutagenesis and so the corresponding mutants were not produced. Some mutants (e.g. HC-A227T) were successfully mutated at the plasmid level, but could not grow sufficiently well in cell culture. Due to the availability of bioreactors, some mutants were not expressed even though they could yield optical density more than 10 in shake flasks. Table 4.6 shows the various Fab variants finally successfully expressed and purified for more than 5 mg, which were used for subsequent Fab stability studies.

Table 4.6 A list of Fab mutants that had been successfully expressed and purified in reasonable amount

	Mutant*	Date*	DOT spike*	Abnormal event*	Data at Harvest			Fab mass (mg)
					OD <sub>600</sub> at harvest*	Wet cell weight (g)*	Fab concentration (mg/ml)*	
Stable mutants	HC-A227E	02.09.2015 - 04.09.2015	No	pH not adjusted since inoculation for 1.5 h	N/A	23.6	0.06	9.9
	HC-A227W	02.09.2015 - 04.09.2015	No	pH not adjusted since inoculation for 1.5 h	N/A	20.9	0.03	5.8
	HC-A228H	23.09.2015 - 25.09.2015	Yes	N/A	N/A	44.9	0.36	61.0
	HC-A228N	23.09.2015 - 25.09.2015	No	N/A	N/A	27.6	0.06	9.8
	HC-A228M	23.09.2015 - 25.09.2015	No	Broth overflowed	N/A	28.0	0.05	8.4
	HC-T135W	09.11.2015 - 11.11.2015	No	Gas outlet blocked, unable to pump gas	55	24.6	0.04	6.5
	HC-T135Y	09.11.2015 - 11.11.2015	Yes	N/A	124	39.1	0.23	39.6
	HC-S134Y	09.11.2015 - 11.11.2015	Yes	N/A	122	41.4	0.12	20.0
	HC-S134M	09.12.2015 - 11.12.2015	Yes	10-20 overflowed	176	48.3	0.11	18.0
	HC-S134P	09.12.2015 - 11.12.2015	No	foam to the headplate	161	56.2	0.13	21.4

	HC-S136G	09.12.2015 - 11.12.2015	No	foam to the headplate	181	40.2	0.14	23.3
	HC-S219Y	10.01.2016 - 12.01.2016	Yes	N/A	281	38.1	0.25	42.8
	LC-L154A	10.01.2016 - 12.01.2016	Yes	N/A	288	36.0	0.27	45.4
	LC-S176W	10.01.2016 - 12.01.2016	Yes	very viscous broth	313	43.8	0.22	37.8
Unstable mutants	LC-A153P	13.06.2016 - 15.06.2016	Yes	N/A	219	38.0	0.08	14.3
	LC-G66P	13.06.2016 - 15.06.2016	Yes	impeller failed to rotate on the last day	256	31.6	0.13	21.4
	LC-G200W	13.06.2016 - 15.06.2016	Yes	N/A	241.0	38.0	0.19	33.0
	HC-V215W	13.06.2016 - 15.06.2016	Yes	viscous broth	390.0	40.4	0.11	19.3

\* Mutant: The mutants listed here were used for this research project. There were other fermentations failed and the purification was not done.

\* Date: The dates ranged from inoculation to the DASbox bioreactor until the harvest. The pre-culture date was one day prior to the inoculation date.

\* DOT spike: This was a good indication of normal cell growth, which usually resulted in high cell density and rich Fab expression.

\* Abnormal event: Due to the lack of operating experience and unforeseen reasons, the fermentation was not always operated in a standard manner. The resulting cell density and Fab expression might be low, which should not be considered as the detrimental effect caused by the plasmid transformed.

\* Wet cell weight: The harvest broth was centrifuged at 10,000 rpm for 90 min. After decanting the supernatant, the weight of the sediment was recorded as the "wet cell weight". The "wet" comment indicates the cell sediment was still partially fluidic.

\* OD<sub>600</sub> at harvest: "N/A" indicates the OD<sub>600</sub> was not measured for the harvested sample.

\* Fab concentration: The exact volume at the harvest point was not recorded. So an estimated volume of 170 ml was used. The concentration, which denoted the Fab in the harvest broth, was calculated based on the estimated 170 ml harvest broth and the final Fab mass (mg) obtained after purification.

\* Fab mass: This was measured after purification.

It took 3-4 runs to optimise the fermentation with the small-scale DASbox bioreactors. As a result, only a small amount of Fab protein was produced in the beginning (e.g. for HC-A227E, HC-A227W). Therefore, it was not sensible to infer the Fab stability from their expression yields, as conditions were different.

Admittedly, the engineering of proteins would potentially impact the host cell expression machinery, which would result in varied cell viability, density, expression level and folding state (Kabir & Shimizu 2003; Angov 2011). The varied expression profile and altered host cell proteins (HCPs) would also impact the downstream stage and the

final purified product obtained (Tscheliessnig et al. 2013). In addition, the integrity and efficacy of therapeutic proteins include the minimisation of chemical instability (e.g. deamidation, oxidation), physical instability (e.g. denaturation, aggregation) and preservation of biological functions (e.g. activity) (Manning et al. 2010). However, due to the focus of this work, the impact on protein expression was not examined, and mainly monomer loss and melting temperature would be assessed for the stability in the next two chapters.

## 4.4 Conclusions

This chapter illustrated several *in-silico* methods to identify potential mutational sites and to design specific mutants with the aim of improving and diminishing Fab stability. Rosetta software was used to build a homology model and screen for the predicted impact of all the possible mutant candidates. RMSF and B-factor analyses were used to restrict the mutants mainly towards flexible sites, as the increased entropy associated with flexibility is mainly thought to result from fewer interactions and hence less enthalpy available to offset the entropy change upon folding. In the meantime, unstable mutants were also prepared as negative control to validate the Rosetta prediction accuracy.

Due to some technical reasons, not all the designed mutants were finally expressed. For example, some regions with very low G/C content in the plasmid DNA are not appropriate to design suitable primers. Other variants were not able to grow in the bioreactors, while there is also a limited availability for accessing the bioreactors. In the end, 14 stable and 4 unstable mutants were prepared for Fab freeze-drying, liquid kinetics and biophysical stability characterisation.

Table 4.7 The pipelines for commercial drug development of therapeutic recombinant antibodies

Strategy, Management and Health Policy				
Enabling Technology, Genomics, <b><u>Proteomics</u></b>	Preclinical Research	Preclinical Development Toxicology, <b><u>Formulation</u></b> Drug Delivery, Pharmacokinetics	Clinical Development Phases I-III Regulatory, Quality, Manufacturing	Postmarketing Phase IV

The table illustrates the pipelines for commercial drug development of therapeutic recombinant antibodies (Harris et al. 2004). The words highlighted in bold and underline, “Proteomics” and “Formulation”, are the stages that might be assisted by the protein engineering work proposed here.

The application of protein engineering should take place in the early stage of drug development pipeline. The genomics and proteomics research takes place prior to the preclinical research, which was followed by studies on formulation, drug delivery, toxicology and pharmacokinetics (Table 4.7) (Harris et al. 2004; Li et al. 2010). Therefore, the protein engineering could 1) help the early proteomics work to screen out potential drug molecules with desired conformational and colloidal stability; 2) assist the formulation development when proper excipients or dosage form cannot meet the requirement on protein stability. As changing the amino acid composition of a protein molecule constitutes a new molecule, the protein engineering requires the recommencing of all the corresponding regulatory implications. Here we adopted a hybrid mutagenesis strategy to propose several mutants with *in silico* stability and would examine their *in vitro* stability in freeze-dried and aqueous forms in the next two chapters.

# 5 Freeze-drying tolerance and thermal stability measurement for mutant Fabs

## 5.1 Introduction

### 5.1.1 Experimental methods to measure the $\Delta G$ and melting temperature

Up to now, ProTherm has been the most widely used database to store experimental data about thermal stability, including the Gibbs free energy change ( $\Delta G$ ), enthalpy change ( $\Delta H$ ) and transition temperature for both wild-type and mutant proteins (Kumar et al. 2006; Prabakaran et al. 2001; Bava 2004). In the dataset, fluorescence, DSC, and CD have been the commonly used methods to measure the  $\Delta G$  and melting transition temperature ( $T_m$ ).

One popular fluorescence method is the ThermoFluor assay (Ericsson et al. 2006), whereby the protein is mixed with SYPRO Orange dye, which fluoresces upon binding to hydrophobic residues of proteins. As the temperature is increased, the native protein unfolds and exposes the hydrophobic core residues to the solvent, which bind more of the dye. The resulting increase in fluorescence is monitored typically using a qPCR instrument. The temperature that results in 50% of the total change in the fluorescence signal is used to indicate the protein stability. Differential scanning fluorimetry (DSF) characterises the difference in  $T_m$  ( $\Delta T_m$ ) from different samples and indicate the stabilising effect induced by mutation or ligand binding (Niesen et al. 2007; Rosa et al. 2015). In addition, a dye-free technique has also been developed based on the tryptophan's intrinsic fluorescence property. Tryptophan fluoresces maximally at  $\lambda_m \approx 331$

nm in low polar, hydrophobic environment and at  $\lambda_m \approx 350$  nm in high polar aqueous environment (Burstein et al. 1973). As the protein unfolds, the fluorescence shifts from 331 to 350 nm and the transition temperature would indicate the stability.

Differential Scanning Calorimetry (DSC) can also be used to investigate the thermal stability of protein samples. By heating the protein samples together with a reference sample, both endothermic and exothermic thermal events (e.g. crystallisation, melting) are recorded (Coleman & Craig 1996). A change in thermal property (e.g. heat capacity) is reflected in the thermal event shift on the endotherm (D'Amico et al. 2003).

Circular dichroism spectroscopy (CD) is another type of measurement used to study protein structure and stability. The mechanism and operating guidelines have been well reviewed (Kelly & Price 2000; Kelly et al. 2005). CD is often used to determine the types and relative proportions of secondary structure existing in the protein (Mehl et al. 2009). The  $\alpha$ -helical structure exhibits strong CD ellipticity minimum at 208 nm and 222 nm. By plotting the ellipticity at 222 nm versus temperature, the  $T_m$  can be determined from the temperature at which 50% of the native protein unfolds when the sample is step-wise heated.

### **5.1.2 Freeze-drying formulation of antibodies**

Most approved monoclonal antibodies are formulated as liquids or in freeze-dried forms, to be administered through the intravenous (IV) route to maximise systemic distribution. By contrast, oral, pulmonary and transdermal administrations are challenged by the instability and the low systemic distribution of antibodies (Harris et al. 2004). The freeze-drying formulation of antibodies has been well reviewed and studied (Daugherty & Mersny 2006; Awotwe-Otoo et al. 2012; Harris et al. 2004). Compared to liquid formulation, freeze-drying provides an alternative approach that reduces the physical and chemical degradation, aggregation and fragmentation of the protein during storage (Daugherty & Mersny 2006; Chang & Hershenson 2002). Freeze-drying is often used in those cases where stable liquid formulations are difficult to identify. The

techniques utilised for protein freeze-drying development are generally applicable to antibodies as discussed in Chapter 1.

Antibody formulations are featured by their high dosage forms, which can be achieved by reconstituting the freeze-dried cake with a lower volume of buffer (Shire et al. 2004). However, aggregation is one of the major challenges that greatly impacts the drug efficacy and potentially causes unwanted and adverse immune responses. Non-reducing sugars, typically disaccharides like sucrose and trehalose, are excipients commonly used to minimise both aggregation and denaturation, arising from freezing and drying stresses. It has also been found that freeze-dried samples with high residual moisture content (5-8% w/w) were more likely to undergo aspartate isomerization, including deamidation, than those with only 2% moisture; but that aggregation was reduced at higher residual moisture for storage temperatures under  $T_g$  (Breen et al. 2001). Reconstitution of the freeze-dried material with water can also lead to protein degradation, but slow reconstitution rates were previously suggested to allow a fully recovery of the native antibody conformation (Daugherty & Mersny 2006). To assess the stability under storage in the solid state, accelerated studies at high temperatures should be performed at below the  $T_g$  so as to be more relevant for its formulation stability (Duddu & Dal Monte 1997).

### **5.1.3 Aims of the chapter**

This Chapter aims to assess the impact of protein conformational stability changes ( $\Delta T_m$ ) due to mutations designed from the previous chapter, upon the monomer loss tolerance of proteins to freeze-drying. Hence both aspects as well as  $\Delta\Delta G$  needed to be assessed to evaluate to what extent the *in silico* and *in vitro* conformational stability could reflect the aggregation in freeze-drying.

## 5.2 Materials and Methods

### 5.2.1 Freeze-drying of the new mutants

The freeze-drying protocol used was mostly the same as described in Section 2.2.7. Each final sample for freeze-drying contained 1 mg/ml Fab, 20 mM sodium phosphate buffer at pH 7, and NaCl to bring the total ionic strength to 200 mM. Five to six replicates were used to minimise the deviations within a running batch. The bottom lips were trimmed off for the flat-bottom 96-well microplates (Greiner Bio-one Ltd, UK) before filling the sample solutions. A pierceable TPE lycapcluster-96 cover mat (Kinesis, UK) was put on top of each trimmed microplate to prevent cake floating out during the cycle, but allow water sublimation through the vent on the mat. The Fab monomer was quantified by SEC-HPLC (Section 2.2.8) before freeze-drying and after reconstitution to determine the monomer loss caused by freeze-drying.

### 5.2.2 Melting temperature ( $T_m$ ) measurement

The thermal transition temperatures of Fab mutants were characterised by their intrinsic fluorescence using the Optim1000 (Unchained Laboratories, Wetherby, UK). Samples were prepared in triplicates at 1 mg/ml in 10 mM sodium phosphate pH 7 buffer if not specified. 9  $\mu$ l was pipetted onto the cuvette for each sample and loaded into the pre-warmed machine. Samples were ramp-heated from 20-90°C at 1°C/min with 30 s incubation period for each temperature. The barycentric mean (BCM) of the spectra ranging 280-460 nm was recorded against each temperature, and was then fitted to a sigmoid curve equation as shown in Equation 5.1. It was observed that the fluorescence value decreased when sample was heated at more than 80-85°C during  $T_m$  measurement. This was probably due to the aggregation of denatured proteins, which decreased the amount of light available for excitation of fluorophores. This range of decreased reading was removed when fitting the sigmoid curve.

$$\text{BCM} = y_0 + \frac{a}{1 + e^{-\frac{T-T_m}{b}}} \quad \text{Equation 5.1}$$

## 5.3 Results and discussion

### 5.3.1 Freeze-drying for the designed mutants

Due to the number and availability of mutants to examine, it was not practical to screen also a large range of formulation conditions for pH, salt type and ionic strength (IS). Therefore, only the physiological condition at pH 7 was used as this would represent an ideal target formulation. Based on the previous results, 200 mM NaCl resulted in the most monomer loss for C226S. Therefore, the Fab samples were brought to an IS of 200 mM by NaCl in order to most readily differentiate the different stabilities of mutants to freeze-drying stress. The pseudo wild-type C226S and initially only the 13 stabilising mutants were freeze-dried, and their monomer losses determined by SEC-HPLC as shown in Figure 5.1. It was promising that 6 out of 13 stabilising mutants improved the monomer loss during freeze-drying by up to 3.97% (HC-A228M) compared to C226S. Admittedly, as C226S Fab was known to be very stable, only small effects were expected from the designed stable mutants. The average monomer loss of most mutants ranged between 9-11%, which was not significant enough to differentiate the stability of the mutants, although it was consistent with the relatively low range of  $\Delta\Delta G$  improvements predicted for these mutants by Rosetta.

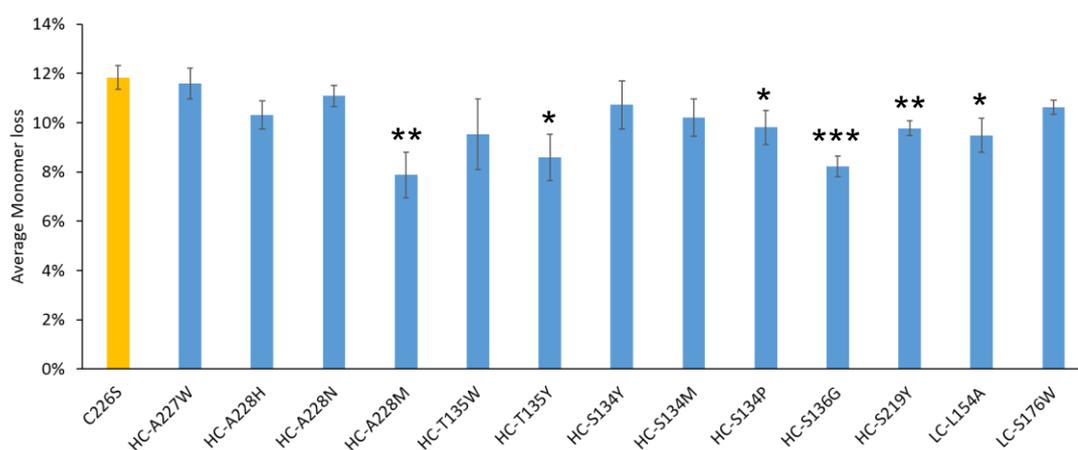


Figure 5.1 The averaged monomer loss for the stabilising mutants during freeze-drying. Error bar is SEM (standard error of the mean). Pseudo wild type C226S was coloured in yellow while the others were in blue. Two-sample *t*-test assuming unequal variances were performed between C226S and other mutants (\*\**p*<0.01, \*\**p*<0.001, \**p*<0.05).

Despite the overall improved monomer loss for the stabilising mutants, it still remained uncertain if the enhanced stability was caused coincidentally. Due to the limited amount of purified Fab, this freeze-drying process could not be performed again to validate its reproducibility between experimental batches. There might exist a possibility that the particular preparation of purified C226S was accidentally slightly unstable, which made the other mutants all appear to be improved. Therefore, to validate the design strategy proposed by Rosetta with a wider range of  $\Delta\Delta G$ , several unstable mutants were also developed, expressed and purified as described in Chapter 4.

C226S and 4 destabilising mutants were freeze-dried with monomer loss determined by SEC-HPLC as shown in Figure 5.2. It was shown that 3 out of 4 of the destabilising mutants designed exhibited significantly decreased monomer retention. It was expected that the difference would be extended if a more unstable condition like lower pH was used. This result was very promising and could exactly confirm the advantage of Rosetta in predicting the protein stability upon point mutations for the Fab.

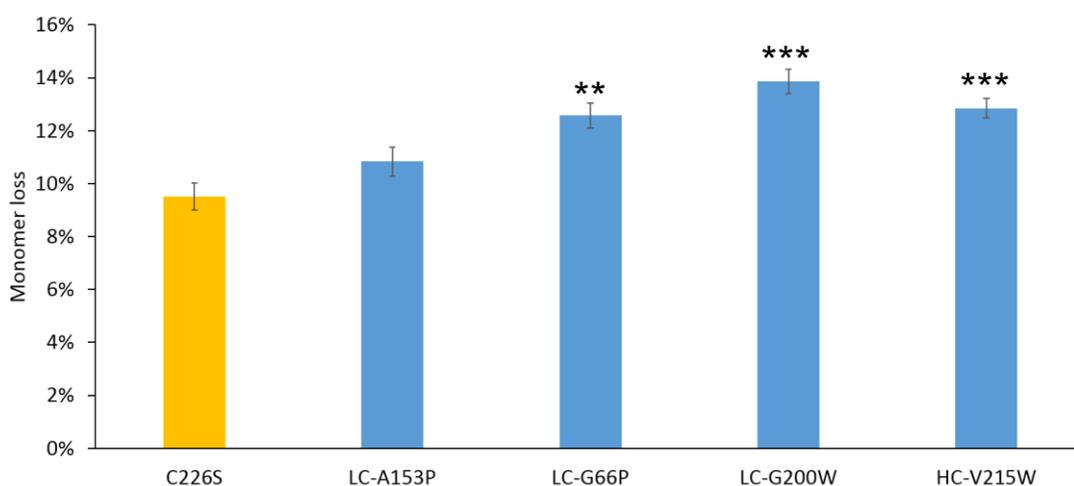


Figure 5.2 The averaged monomer loss for the destabilising mutants during freeze-drying. Error bar is SEM (standard error of the mean). Pseudo wild type C226S was coloured in yellow while the others were in blue. Two-sample *t*-test assuming unequal variances were performed between C226S and other mutants (\*\**p*<0.01, \*\**p*<0.001, \**p*<0.05).

It was found that C226S performed slightly differently between the two batches as shown in Figure 5.1 and Figure 5.2 with average monomer loss of 11.84% and 9.51%, respectively. This implies that a minimum of 2% monomer loss should be considered as the batch-to-batch variation, assuming a systematic effect that applies equally to the whole batch of mutants. In order to conduct an unbiased cross comparison to assess stability between stabilising and destabilising mutants, all the monomer loss of designed mutants was normalised against C226S in the corresponding batch with C226S set as 1, and plotted in Figure 5.3.

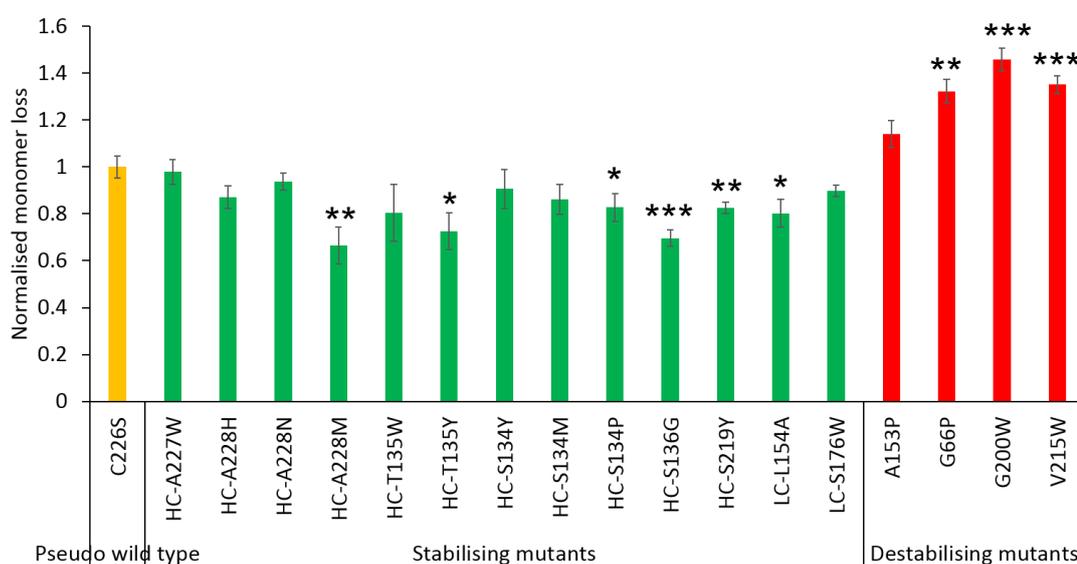


Figure 5.3 The normalised monomer loss for the stabilising and destabilising mutants during freeze-drying

Error bar is SEM (standard error of the mean). Pseudo wild-type C226S, stabilising mutants and destabilising mutants were coloured in yellow, green and red, respectively. Two-sample *t*-test assuming unequal variances were performed between C226S and other mutants (\*\**p*<0.001, \*\**p*<0.01, \**p*<0.05). The SEM of C226S was averaged from the two batches.

### 5.3.2 Thermal stability of designed mutants and their relations to freeze-drying

After determining the freeze-drying monomer loss of the mutants designed, it was useful to see if this instability could be related to the intrinsic properties of the mutant proteins. In that case, the performance of mutants could be predicted more efficiently and with less sample. Melting temperature ( $T_m$ ) is one of the most commonly used

descriptors to indicate the thermal stability of protein upon step-wised elevated thermal treatment.  $T_m$  measurement was carried out for both stabilising and destabilising mutants, and is shown in Figure 5.4 and Figure 5.5, respectively. All the stabilising mutants ranged in a narrow window with  $T_m$ -values from 77.5-80.4°C. For the four tested destabilising mutants, all of them were significantly lower than the wild type, with LC-A153P revealing a highest  $T_m$  of 78.8°C and LC-G66P being the most unstable with a  $T_m$  of 72.9°C.

Surprisingly, the pseudo wild-type C226S mutant achieved the highest  $T_m$  in both of the measurements. The highest  $T_m$  of C226S may come from its large-scale manufacture process. C226S was expressed in a 30 L fermenter with 2-3 g product yield, and therefore went through a shorter concentration process after purification. For the other mutants, as they were expressed in 200 ml bench-top bioreactors, they resulted in only 10-20 mg product and were very dilute into Protein G column eluate. As a result, they were concentrated multiple times before storing at -80°C. This excess concentration process may cause the Fab degrade to some extent, although all of the initial Fab samples were prepared to 100% monomer as confirmed by SEC-HPLC.

The C226S gave a slightly lower  $T_m$  when measured along with the stabilising mutants (80.7°C), than when repeated alongside the destabilising mutants (81.1°C), which demonstrated a good batch-to-batch reproducibility for the  $T_m$  measurement. The difference may be caused by the total number of samples performed at the same time in the Optim. When more samples are measured, it has been found that all the samples are incubated for longer at each step-wise increase in temperature. At temperatures closest to  $T_m$ , the proteins are more likely to unfold and aggregate during the longer incubations associated with using more samples. Therefore, the C226S would be found to have a slightly lower  $T_m$  as in Figure 5.4, compared to that in Figure 5.5.

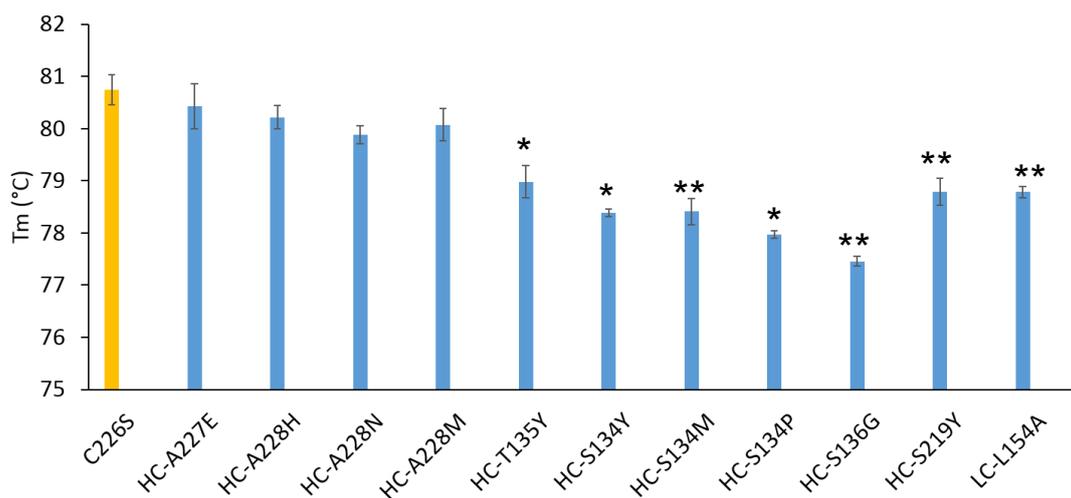


Figure 5.4 The melting temperature ( $T_m$ ) for the stabilising mutants measured by Optim1000 at pH 7

Error bar is SEM (standard error of the mean). Pseudo wild type C226S was coloured in yellow while the others were in blue. HC-A227W and HC-T135W were not available for this measurement due to limited sample amount. Two-sample  $t$ -test assuming unequal variances were performed between C226S and other mutants (\*\* $p < 0.01$ , \* $p < 0.05$ ).

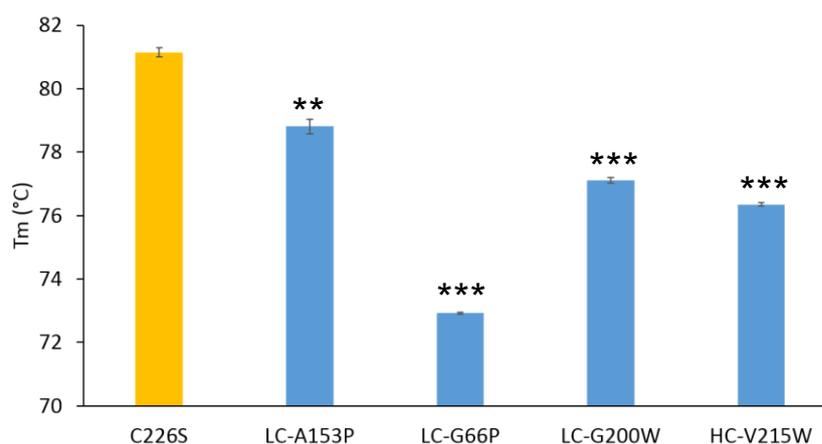


Figure 5.5 The melting temperature ( $T_m$ ) for the destabilising mutants measured by Optim1000 at pH 7

Error bar is SEM (standard error of the mean). Pseudo wild-type C226S was coloured in yellow while the others were in blue. Two-sample  $t$ -test assuming unequal variances were performed between C226S and other mutants (\*\* $p < 0.01$ , \* $p < 0.05$ ).

To enable cross-comparison while minimising the batch influence, the  $T_m$  of the designed mutants were subtracted from the same-batch C226S value as  $\Delta T_m$  so as to represent the positive or negative impact of point mutations, while setting C226S as 0 as shown in Figure 5.6 (Cheng et al. 2012). Figure 5.7 correlated the normalised monomer loss in freeze-drying (Figure 5.3) with the  $\Delta T_m$ . As expected, a negative correlation ( $R^2$

= 0.40) was found as a decrease in  $T_m$  resulted in more monomer loss due to freeze-drying. Therefore,  $T_m$  was a good intrinsic property, with approximately 40% accuracy, for predicting the stability improvement against freeze-drying, as caused by point mutations. This promising result indicated that  $T_m$  can be used as an efficient pre-screening tool, with 40% confidence to assess the developed mutations, and with a shorter analysis period (1 day vs 5 days) and less sample (9  $\mu$ l vs 200  $\mu$ l) than when running a freeze-drying screen. However, due to the imbalance in the relative number of stabilising and destabilising mutants, with most of the dots scattered around the right-bottom corner of Figure 5.7, the negative correlation was perhaps overly influenced by the four unstable mutants coloured in red. Therefore, it would be beneficial to include more destabilising mutants to equally distribute the data points across the graph.

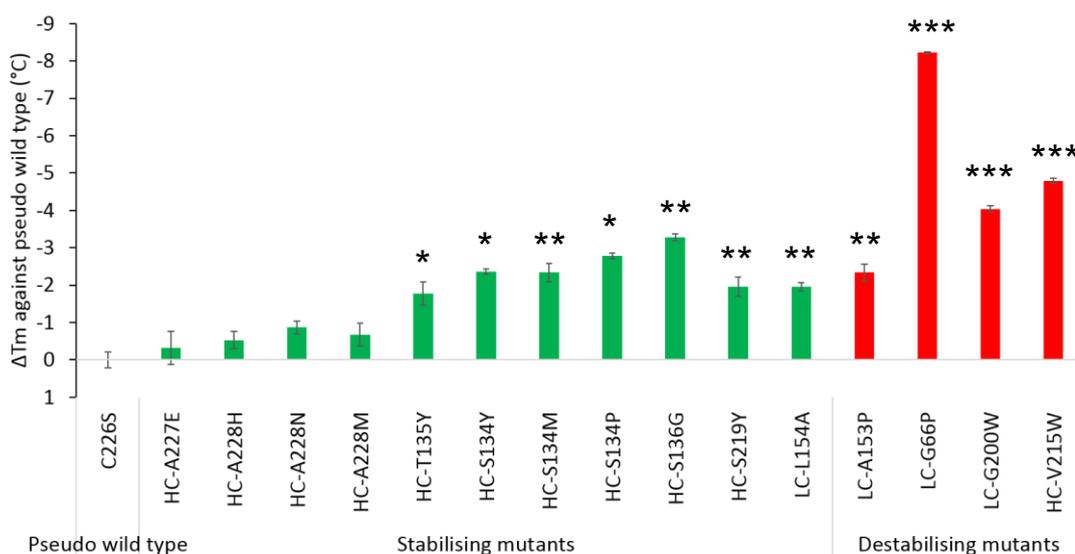


Figure 5.6 The temperature difference of  $T_m$  ( $\Delta T_m$ ) for the stabilising and destabilising mutants against pseudo wild-type C226S at pH 7

Error bar is SEM (standard error of the mean). Pseudo wild-type C226S was set at 0. Stabilising mutants and destabilising mutants were coloured in green and red, respectively. The SEM of C226S was averaged from the two batches. LC-S176W was excluded due to its sigmoid  $T_m$  curve could not be obtained. Two-sample  $t$ -test assuming unequal variances were performed between C226S and other mutants (\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ).



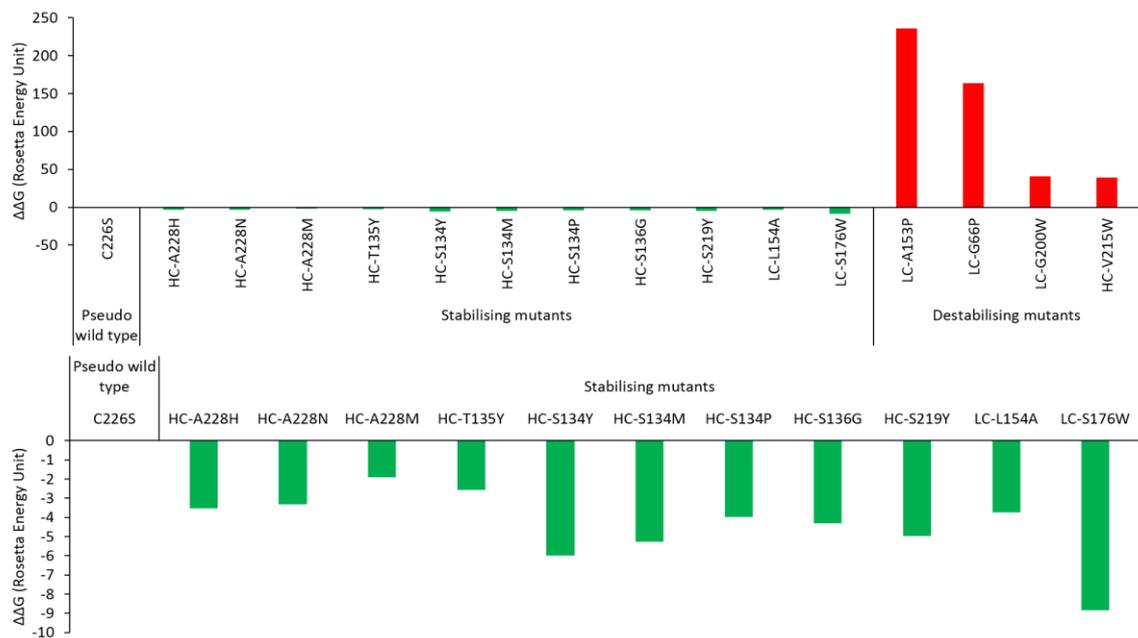


Figure 5.8 The  $\Delta\Delta G$  calculated by Rosetta for stabilising (green) and destabilising (red) mutants. The C226S had  $\Delta\Delta G$  of 0 as it served as the benchmark for point mutation. Due to the small magnitudes of  $\Delta\Delta G$  for stabilising mutants, they were also plotted separately.

Figure 5.9 and Figure 5.10 show the correlation of  $\Delta\Delta G$  with  $T_m$  and freeze-drying monomer loss, respectively. Figure 5.11 showed a summary for the correlations between freeze-drying monomer loss,  $T_m$ , and  $\Delta\Delta G$ . The results indicated that increases in  $\Delta\Delta G$  resulted in decreases in  $T_m$  ( $R^2 = 0.26$ , Figure 5.9) and increases in monomer loss ( $R^2 = 0.33$ , Figure 5.10), which corresponded to the theory although the correlations were poor, and perhaps only significant for destabilising mutants. In general, Rosetta was an acceptable prediction tool that could differentiate mutational effects with 30% confidence, but with much greater confidence for the destabilising mutants. When an *in vitro* study is limited in time and labour cost, this computational method could offer the efficiency to screen more than 8000 point mutational candidates for a 400-residue protein. It would certainly be useful for avoiding mutations that cause significant destabilisation, or when designing improved mutants in a protein that starts with a lower stability than the A33 C226S Fab.

Similarly to Figure 5.7, the correlation established in Figure 5.9 and Figure 5.10 relied heavily upon the four destabilising mutants as all the stabilising ones clustered very closely at  $\Delta\Delta G$  around 0. If the four unstable mutants were excluded, the correlation

would actually be in the opposite direction (Figure 10.3) or no correlation (Figure 10.4). This implied that when the mutants had reached sufficient stability, the  $\Delta\Delta G$ , which denoted the global stability, became less accurate or sensitive in estimating the thermal stability and level of aggregate formed due to freeze-drying stress. The mechanisms leading to aggregation (when measuring  $T_m$  or when freeze drying) could be different with a protein that is already at high conformational stability, and hence other properties could provide a better prediction under those conditions. For example, local instability in particular sequence regions may contribute more to aggregation, and other descriptors like surface charge, solvent accessible area, and cavity volume may play a more important role in predicting the stability.

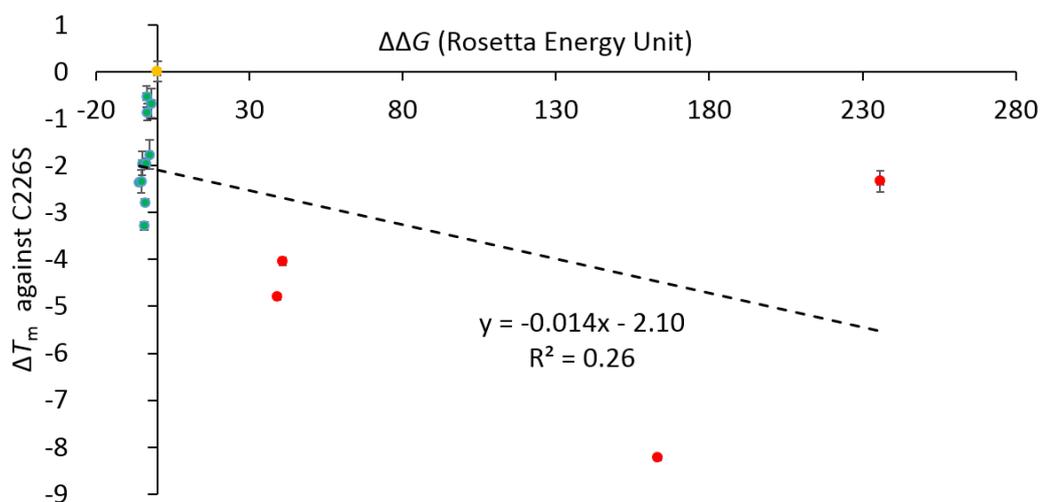


Figure 5.9 The correlation between  $\Delta T_m$  and  $\Delta\Delta G$

$\Delta T_m$  was calculated against C226S (Figure 5.6) and  $\Delta\Delta G$  is calculated by Rosetta (Figure 5.8). Error bar is SEM (standard error of the mean). Pseudo wild type C226S, stabilising mutants and destabilising mutants were coloured in yellow, green and red, respectively.

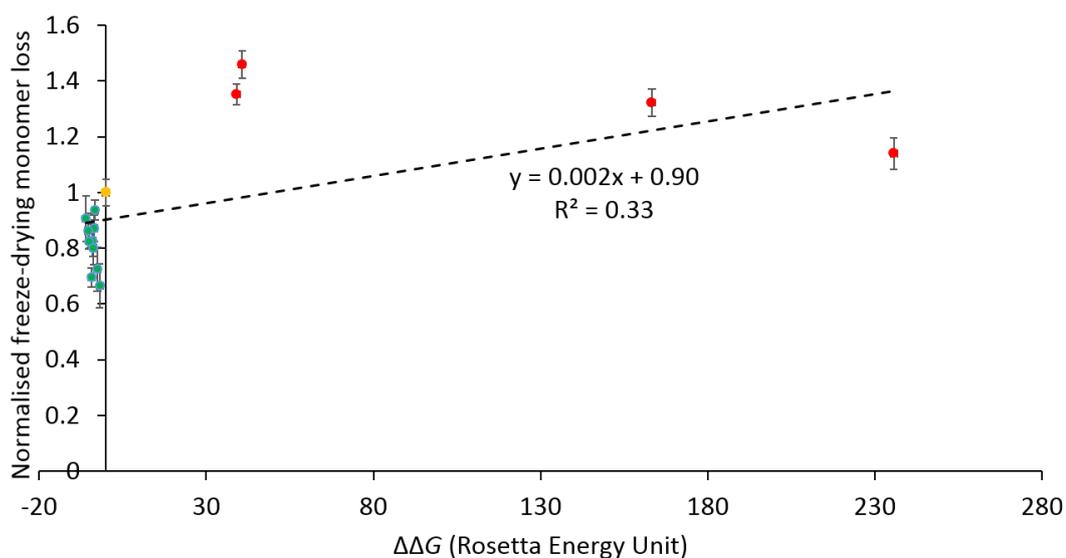


Figure 5.10 The correlation between normalised monomer loss in freeze-drying and  $\Delta\Delta G$ . Normalised monomer loss in freeze-drying was from Figure 5.3 and  $\Delta\Delta G$  calculated by Rosetta was from Figure 5.8. Error bar is SEM (standard error of the mean). Pseudo wild-type C226S, stabilising mutants and destabilising mutants were coloured in yellow, green and red, respectively.

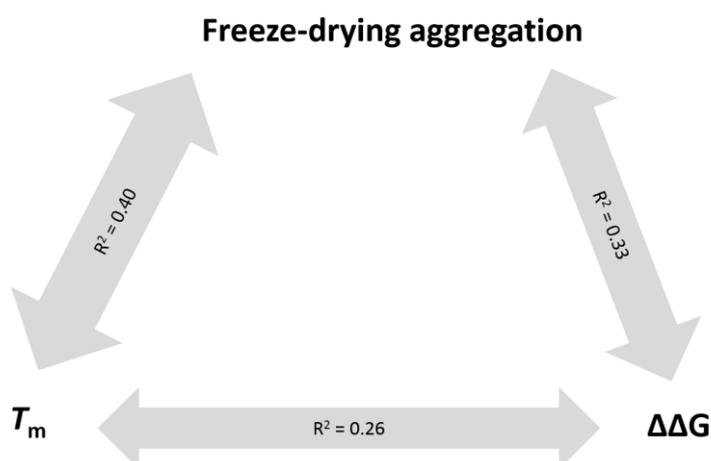


Figure 5.11 A summary of correlations between freeze-drying aggregation rate,  $T_m$  and  $\Delta\Delta G$ . The widths of the arrows were proportional to their coefficient of determination ( $R^2$ ). “Freeze-drying aggregation” denoted the normalised monomer loss (Figure 5.3); “ $T_m$ ” denoted the  $\Delta T_m$  against C226S (Figure 5.6); “ $\Delta\Delta G$ ” was derived from Figure 5.8.

## 5.4 Conclusion

This chapter analysed the freeze-drying performance for the designed stabilising and destabilising mutants. The mutational design strategy was shown to be successful for nearly half of the stabilising mutants, as 6 out of 13 of them had 20% less monomer loss

than C226S after freeze-drying, while 3 out of 4 of the destabilising ones had significantly 14-46% more monomer loss. The same design strategy was also successful in obtaining mutants with altered  $T_m$  values, although the relative  $T_m$  of C226S appeared to be higher than for all mutants, possibly due to the different scales of purification process. The different extent of monomer loss of the mutants could be described by their differences in  $\Delta\Delta G$ , as stabilising mutants had  $\Delta\Delta G$  less than 10 Rosetta Energy Unit (REU) while destabilising ones had  $\Delta\Delta G$  of 40-235 (REU). The mutational stability was also revealed from measurement of  $T_m$  values, whereby destabilising mutants had lower  $T_m$ -values than stabilising mutants. It was shown that  $T_m$  and  $\Delta\Delta G$  estimated the monomer loss in freeze-drying with some degree of accuracy ( $R^2 = 0.40$  and  $0.33$ ). The correlation between  $T_m$  and  $\Delta\Delta G$ , however, was poor with an  $R^2$  of  $0.26$ . Overall, the poor estimation for freeze-drying aggregation by  $T_m$  and  $\Delta\Delta G$  may come from the complex degradation during freeze-drying. As  $T_m$  and  $\Delta\Delta G$  mainly characterise the unfolding and enthalpy (by Rosetta), the freezing and drying stresses could not be effectively captured. Therefore, it was worthwhile to examine the aggregation in aqueous phase so as to validate the performance of  $T_m$  and  $\Delta\Delta G$  estimates in assessing the stability of Fab in the liquid form.

# 6 Liquid aggregation kinetics for designed mutant formulations

## 6.1 Introduction

### 6.1.1 Antibody aggregation in liquid formulations

Liquid formulation strategies to minimise the denaturation and aggregation of therapeutic Fab and full antibodies have been well reviewed previously (Uchiyama 2014; Lowe et al. 2011). It is known that colloidal stability and conformational stability are two key factors that influence the rate and extent of aggregation (Uchiyama 2014). Second virial coefficients ( $B_{22}$ ) and the concentration dependence of the apparent diffusion constant ( $k_{\text{Diff}}$ ) are two measurable parameters that can be used to describe the colloidal stability. The free energy difference between native and denatured state ( $\Delta G_{\text{ND}}$ ), and the temperatures where unfolding starts ( $T_{\text{onset}}$ ), where aggregates are first detected ( $T_{\text{agg}}$ ), or for the midpoint of the unfolding/aggregation transition ( $T_m$ ) are useful parameters that describe or indicate the conformational stability. Various measurement techniques have been developed to characterise the protein conformation and progress of aggregation, which include DSC, DLS, SEC, optical density and analytical ultracentrifugation (Lowe et al. 2011). To enable efficient screening of optimal formulation conditions, accelerated stability studies are also commonly conducted (Samra & He 2012; Taylor et al. 2010a), while *in silico* screening is also emerging as a potential method for predicting the behaviour of mutational candidates (Lauer & Agrawal 2012; Wijma et al. 2013).

## 6.1.2 Formulation with excipients to mitigate antibody aggregation in the liquid state

The use of excipients is a common way to preserve the stability and activity of proteins either in the aqueous phase or under freeze-drying conditions (Daugherty & Mrsny 2006; Cheng et al. 2012). Excipients exhibit their stabilising effect on proteins through multiple mechanisms, that are still poorly understood. Understanding the mechanism by which excipients stabilise proteins, and the interaction between the impacts of excipients and mutations, would provide valuable insights that could guide us in the rational engineering of the protein.

Sugars (sucrose, trehalose) and polyols (mannitol and sorbitol) interact with proteins through their hydroxyl groups, which replace the hydrogen bond formed with water molecules, and minimise protein denaturation during drying. They therefore improve the conformational stability of proteins, but have little effect on colloidal stability, as observed with a native antibody (Uchiyama 2014). Sugars and polyols are also thought to increase the  $T_m$  of protein formulations through preferential exclusion from protein cavities, which stabilises the protein native state with more hydration layers on the protein (Abbas et al. 2012; Timasheff 2002).

Tween 20 and Tween 80 are widely used excipients for antibody drug formulations. They are surfactants that can reduce the surface tension in the air-solvent and surface-solvent interfaces, and can coat the hydrophobic surfaces of containers and air-solvent boundaries. This then decreases the tendency of proteins to populate and unfold at the solvent interface (Kerwin 2008). However, Tween may also favourably bind to the hydrophobic surface of an unfolded antibody which can promote denaturation, but also suppress aggregation (Uchiyama 2014). Tween has also been found to degrade into unwanted products by heat and near-UV light (Uchiyama 2014).

Basic amino acids have been shown to stabilise antibodies by increasing their melting temperature (Arakawa et al. 2007; Falconer et al. 2011). It was found that the positive charge of histidine, arginine and lysine performed better in stabilising an IgG

compared to neutral amino acids serine and alanine, or acidic amino acids (Falconer et al. 2011). Arginine was found to behave differently to the other stabilising amino acids. It decreased the  $T_m$  when its concentration exceeded 0.5 M, but was found to suppress proteins interactions, and aggregation against heat-induced unfolding (Arakawa et al. 2007). An equal molarity (up to 200 mM) of arginine and glutamic acid together appeared to suppress mAb aggregation at elevated temperature and pH 7, compared to the optimal pH 5 formulation condition (Kheddo et al. 2014).

### 6.1.3 Aims of the chapter

The freeze-drying process described in Chapter 5 involves a more complex range of factors that could influence degradation and aggregation, than does liquid storage, which may have convoluted any attempts to correlate  $T_m$  and  $\Delta\Delta G$  to the monomer loss after freeze-drying. In order to validate the usefulness of  $T_m$  and  $\Delta\Delta G$  more generally for protein formulation, it is worthwhile to examine the aggregation in aqueous phase, which provides a simpler denaturing pathway whereby the aggregation rate might be expected to correlate better (Kamerzell et al. 2011; Chi, Krishnan, Randolph, et al. 2003). Therefore, this chapter aimed to perform a liquid aggregation study for the designed stabilising and destabilising mutants. The  $T_m$  and  $\Delta\Delta G$  were each correlated to the aggregation in the aqueous phase, and compared with those correlations found previously for freeze-drying. The stabilising effect provided by excipients was also assessed. This provided insights into the extent that the instability of destabilising mutations could be minimised by osmolyte or surfactant. The excipient effects on aggregation were also correlated with their  $T_m$  so as a potential route to more efficient screening of new formulations.

## 6.2 Materials and methods

### 6.2.1 Liquid aggregation kinetics

Thawed from -80°C stock, Fab was filtered, buffer exchanged to water, concentrated at 2 mg/ml and stored at 4°C (Section 2.2.3). Excipients and buffer were prepared in stock solutions before formulating with Fab. The final formulation conditions used were shown in Table 6.1. Each formulation contained Fab, one type of excipient (or without excipient) and one type of buffer.

Table 6.1 The formulation conditions for liquid aggregation study

<b>Fab</b>		1 mg/ml
<b>Excipient</b>	<b>Mannitol</b>	40 mg/ml (219.6 mM)
	<b>Sorbitol</b>	40 mg/ml (219.6 mM)
	<b>Tween 80</b>	4 mg/ml (3.1 mM)
	<b>Glycine</b>	20 mg/ml (266.4 mM)
<b>Buffer condition</b>	<b>20 mM acetate pH 5</b>	NaCl to bring the total ionic strength to 200 mM
	<b>20 mM citrate pH 4</b>	

Each sample was aliquoted into 20 µL in a safe-lock micro-centrifuge tube (Fisher Scientific, UK). During the aggregation study, the samples were firstly incubated at 45 or 65°C. Samples in triplicates were taken out every 15 or 30 min and centrifuged at 15,000 rpm, 4°C for 15 min, from which 15 µL supernatant was transferred into a HPLC vial insert. 5 µL sample was used to determine the monomer analysed by the SEC-HPLC (Section 2.2.8).

The monomer retention at each condition was plotted against incubation time. To fit a linear regression curve, the initial plateau stage (0.5-1 hour) and the monomer retention closing to 0 were excluded. Only data points that covering the same period in a batch for all the mutants were used to fit the linear regression. The slope of each regression curve was extracted and normalised against C226S to determine the aggregation rate.

## 6.2.2 Melting temperature ( $T_m$ ) of C226S

### formulated with excipients

The  $T_m$  measurement followed the same protocol as described in Section 5.2.2 except the different formulation conditions as shown in Table 6.2. Each formulation was run in triplicate and contained 1 mg/ml Fab, one type of excipient and 10 mM sodium phosphate buffer at pH 7,

Table 6.2 The formulation conditions to perform the  $T_m$  measurement

<b>Fab</b>		1 mg/ml
<b>Excipient</b>	<b>Trehalose</b>	50 mg/ml
	<b>Sucrose</b>	50 mg/ml
	<b>Mannitol</b>	40 mg/ml
	<b>Sorbitol</b>	40 mg/ml
	<b>Tween 20</b>	4 mg/ml
	<b>Tween 80</b>	4 mg/ml
	<b>Glycine</b>	20 mg/ml
	<b>Arginine</b>	20 mg/ml
<b>Buffer</b>	<b>Sodium phosphate</b>	10 mM, pH 7

## 6.3 Results and discussion

### 6.3.1 Liquid solution condition scouting to study liquid aggregation kinetics

In order to perform an efficient screening of the mutants and to minimise sample evaporation during long-term incubations, a scouting study was conducted to identify the conditions that could complete one full kinetics study in a single working day. It aimed to aggregate the majority of the Fab monomers in 5-10 hours. Based on previous work (Chakroun et al. 2016), low pH, high salt concentration and high temperature would cause rapid aggregation. In this work, C226S was firstly tried to provide a benchmark for the designed mutants.

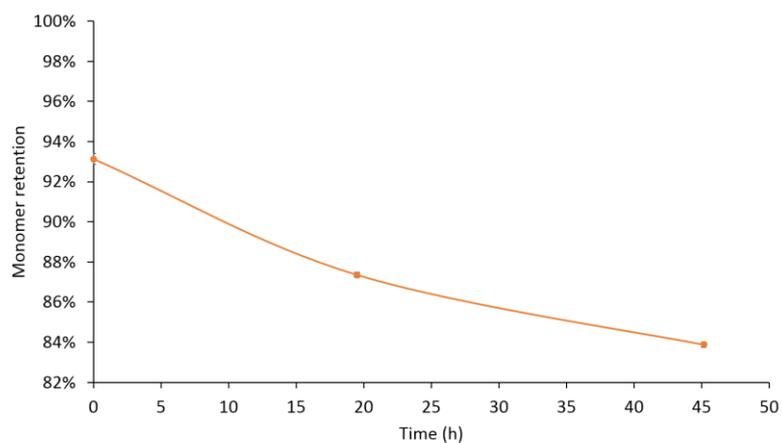


Figure 6.1 Monomer retention kinetics at 65°C for 1 mg/ml C226S, 20 mM acetate, pH 5, with NaCl to bring the total ionic strength to 200 mM  
Error bar: standard error of the mean.

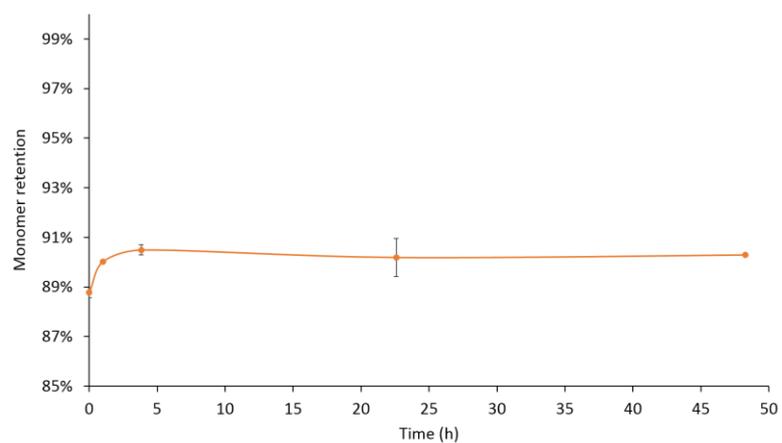


Figure 6.2 Monomer retention kinetics at 45°C for 1 mg/ml C226S, 20 mM citrate, pH 4, with NaCl to bring the total ionic strength to 200 mM  
Error bar: standard error of the mean.

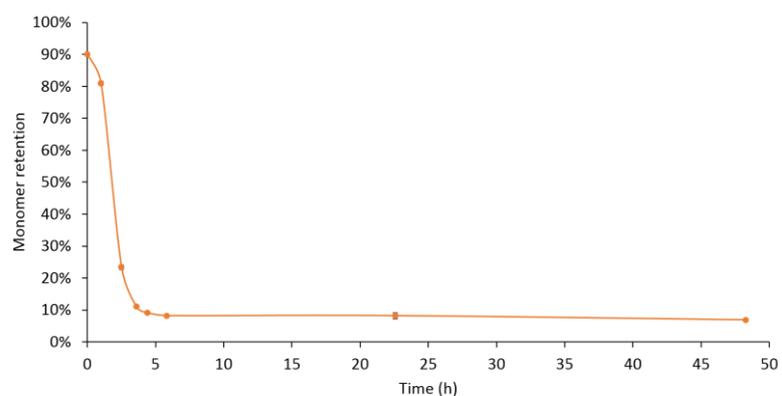


Figure 6.3 Monomer retention of kinetics at 65°C for 1 mg/ml C226S, 20 mM citrate, pH 4, with NaCl to bring the total ionic strength to 200 mM  
Error bar: standard error of the mean.

Figure 6.1 showed the C226S monomer retention at 20 mM acetate, pH 5 with NaCl to bring the total ionic strength to 200 mM at 65°C, which indicated that the monomer population decreased by less than 20% in two days. Based on this trend, it might take a week to fully aggregate the monomer at this condition. A more extreme condition of pH 4 (20 mM citrate) but with more moderate incubation temperature at 45°C was then tested as shown in Figure 6.2. However, this condition was even milder as no detectable monomer was observed for the first 2 days. After that, a low pH at 4 with 65°C incubation temperature was further tested as shown in Figure 6.3. The required curve for monomer decay was obtained such that more than 90% monomer was lost in 5 hours. Based on the results, this condition was selected to evaluate the stability of mutants in the aqueous phase. In order to obtain a more accurate record for the monomer loss, a sampling interval of at least every 30 min would also be used.

### **6.3.2 Effect of mutations upon liquid aggregation kinetics**

The liquid aggregation kinetics data for C226S and the mutants are shown in Figure 6.4, in which the aggregation rate was determined based on the slope from a linear regression curve as shown on the right. All of the stabilising mutants (LC-S176W, LC-L154A, HC-T135Y) performed better than the destabilising ones (LC-A153P, LC-G200W, HC-V215W, LC-G66P). The aggregation rate differences between stabilising mutants were much smaller than the differences between destabilising ones, which corresponded to the extent shown in the predicted  $\Delta\Delta G$  and the experimentally determined  $T_m$ . There was evidence of evaporation at later time points in some samples and so these data points were excluded from the analysis.

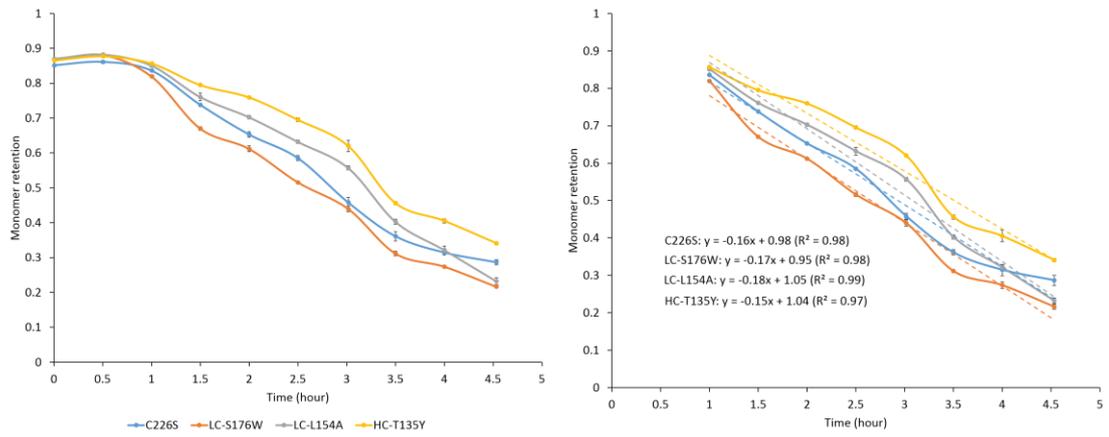


Figure A. C226S and stabilising mutants without excipient

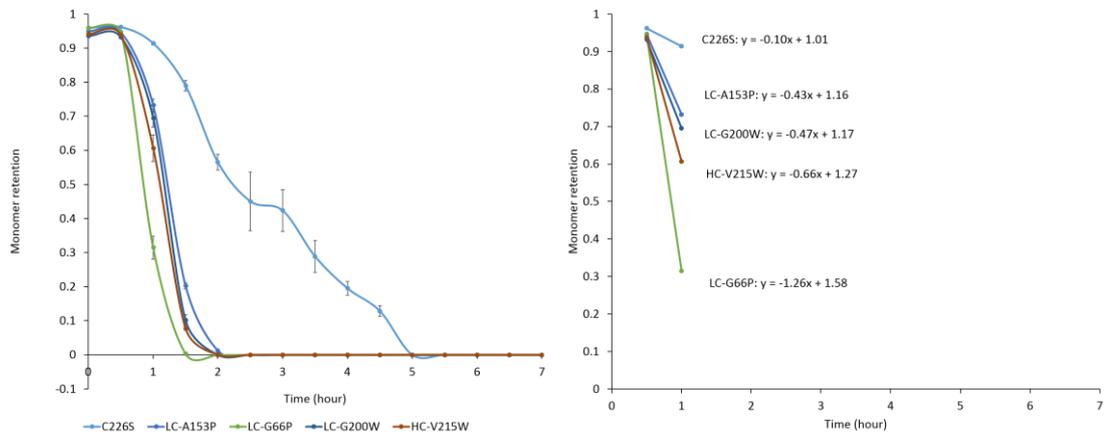


Figure B. C226S and destabilising mutants without excipient (The LC-G66P already achieved 0 monomer retention at 1.5 hour. So its aggregation rate was only fitted from 0.5 and 1-hour data points. In order to have an equivalent comparison, the aggregation rates of all the other mutants were also fitted from those two time points.)

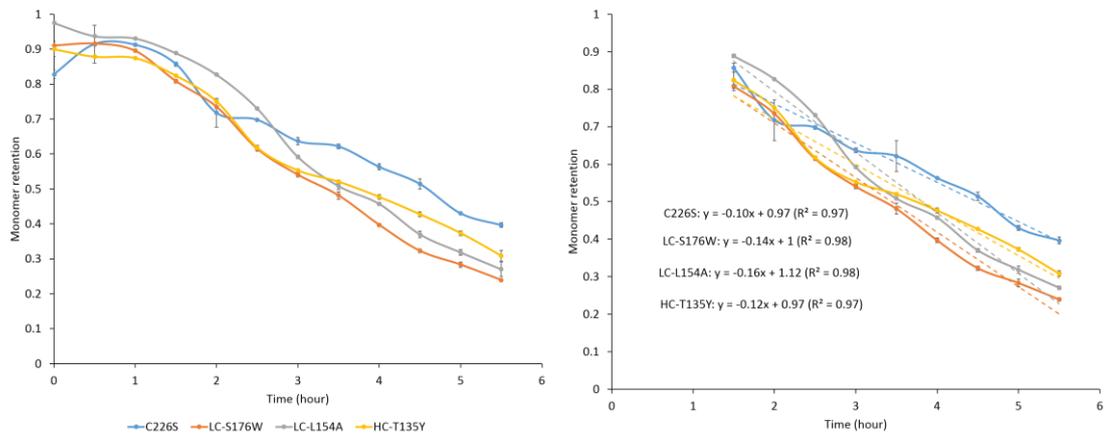


Figure C. C226S and stabilising mutants with 4 mg/ml Tween 80

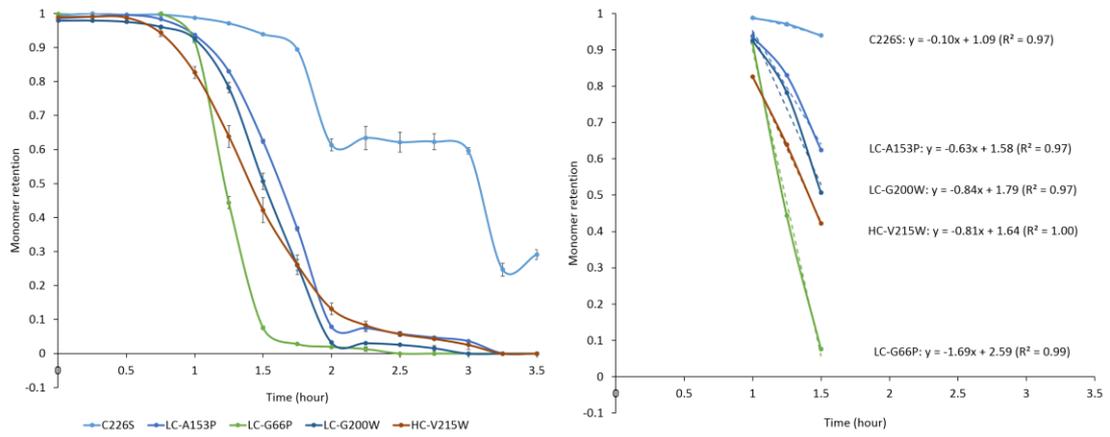


Figure D. C226S and destabilising mutants with 4 mg/ml Tween 80

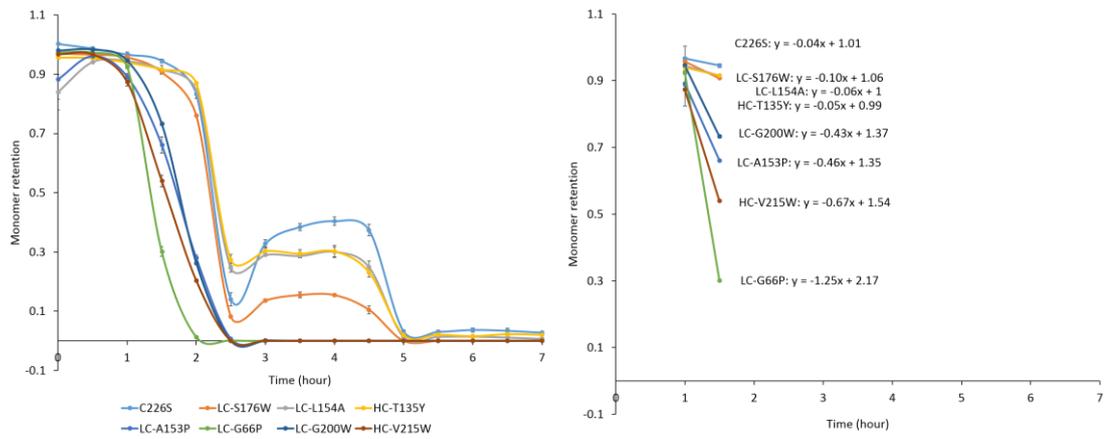


Figure E. C226S, stabilising and destabilising mutants with 40 mg/ml mannitol

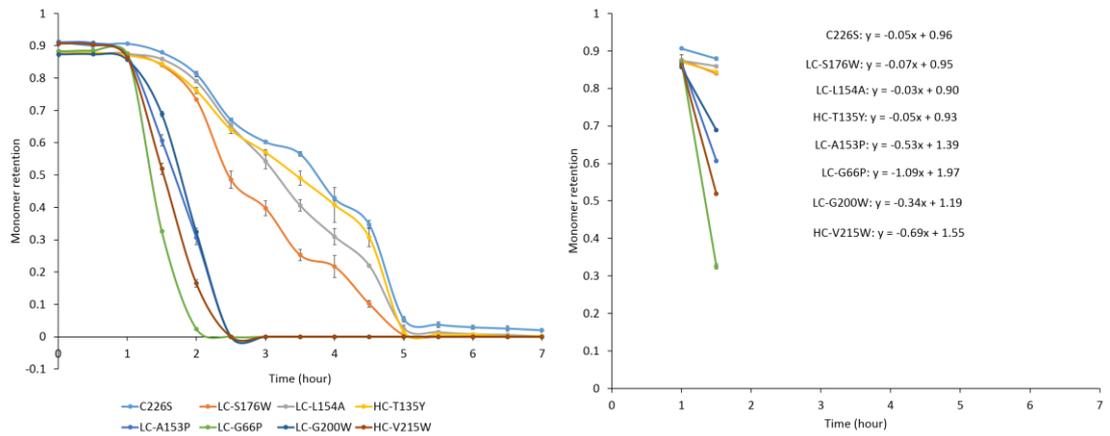


Figure F. C226S, stabilising and destabilising mutants with 40 mg/ml sorbitol

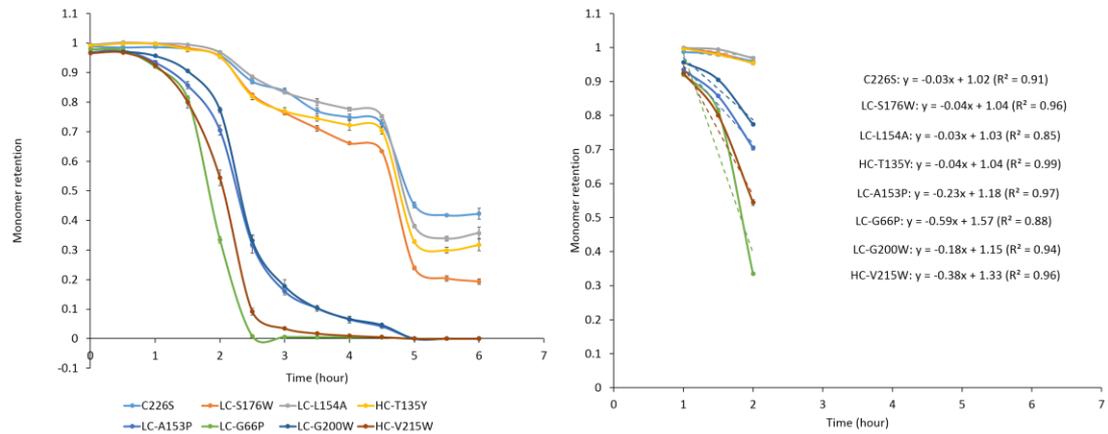


Figure G. C226S, stabilising and destabilising mutants with 20 mg/ml glycine

Figure 6.4 The monomer retention of liquid kinetics for 1 mg/ml Fab mutants at 20 mM citrate, pH 4 with NaCl to bring the total ionic strength to 200 mM at 65°C

Full triplicates were used for each data point with error bars indicating the standard error of the mean. One type of excipients was added as specified individually. Left: overall monomer retention kinetics; Right: the truncated data points to obtain aggregation rate.  $R^2$  values were not shown if only two data points were used to obtain the linear regression.

Considering the batch-to-batch variation, it would be better to normalise the aggregation rate before cross-comparing the mutants' behaviours for quantitative analysis. Figure 6.5 shows the aggregation rate relative to that of C226S for each excipient group. It can be seen that the impact of mutations on aggregation varied differently for each excipient type. The mutants behaved with the most sensitivity for mannitol, followed by Tween 80, sorbitol and glycine, and exhibited least sensitivity to without the excipient.

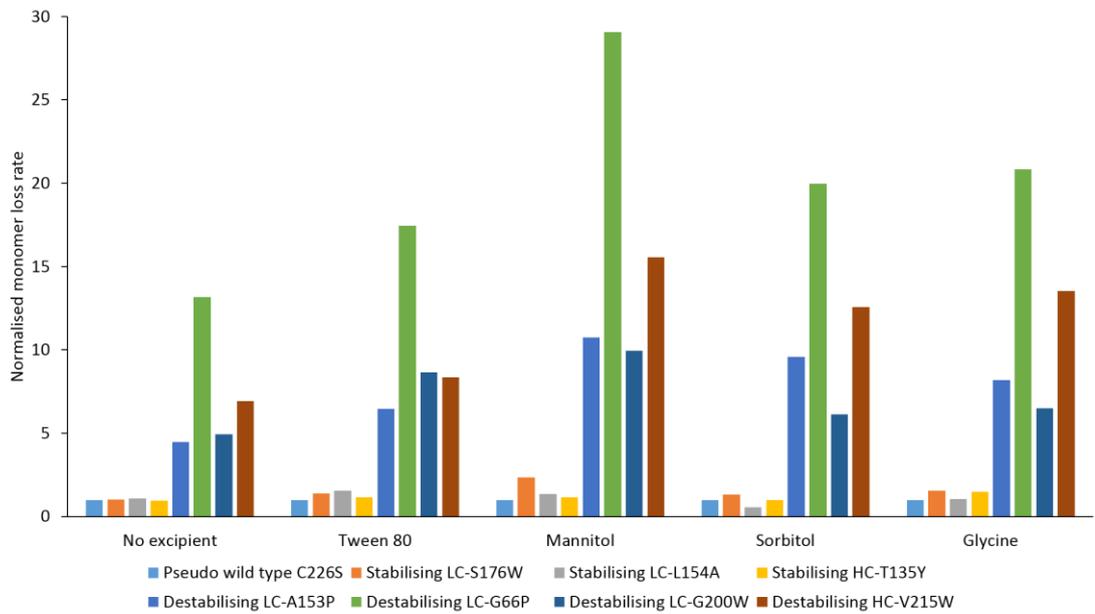
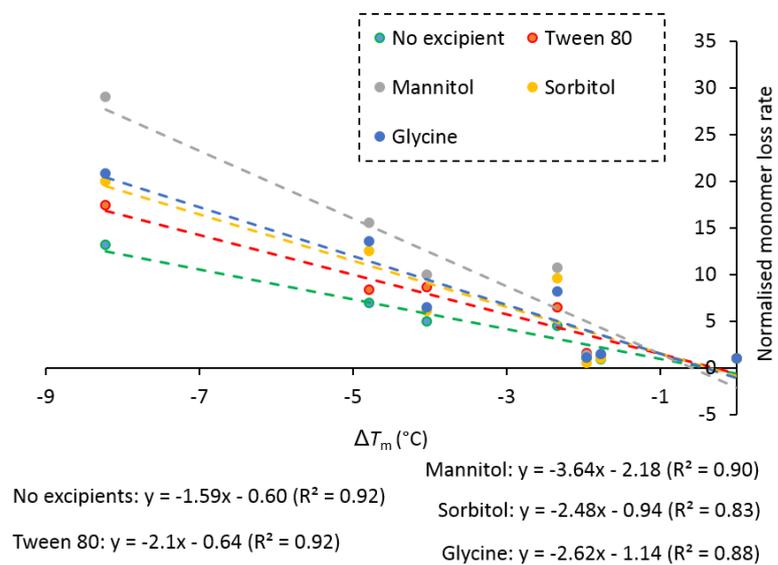


Figure 6.5 The normalised aggregation rates derived from Figure 6.4.

For each of the excipient group, the aggregation rate was normalised against the C226S in the same group. The error bars were not shown as the rates were obtained from the slopes of linear regression.



(A)

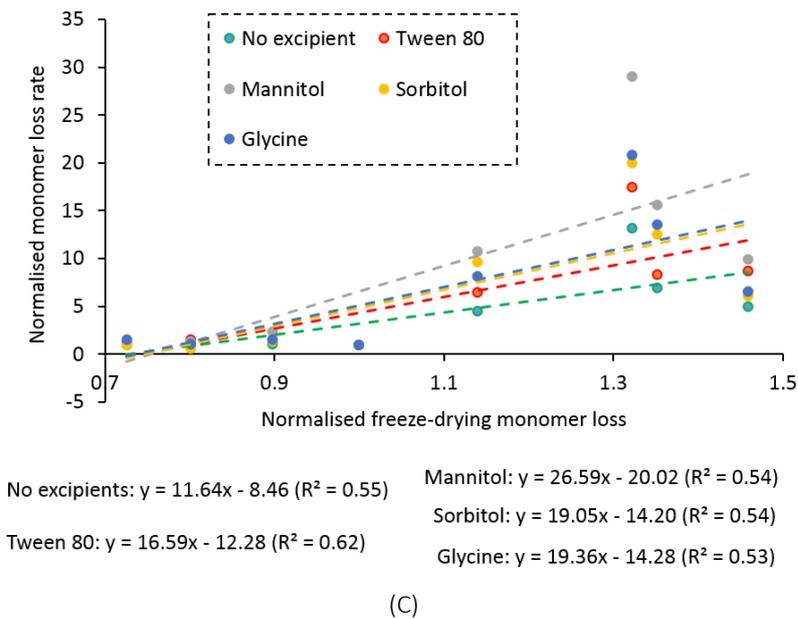
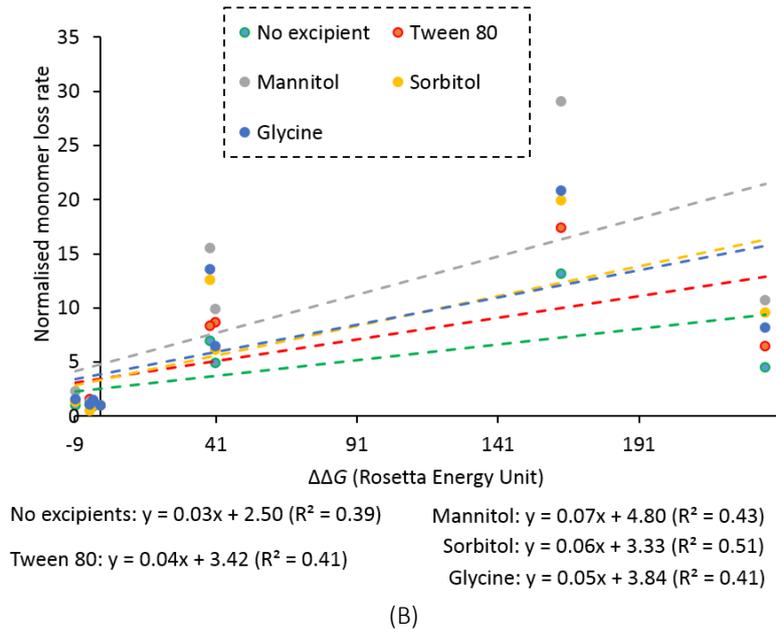


Figure 6.6 Correlations between  $\Delta T_m$ ,  $\Delta\Delta G$ , normalised freeze-drying monomer loss and normalised monomer loss rate.

Figure A: correlations between  $\Delta T_m$  and normalised monomer loss rate; Figure B: correlations between  $\Delta\Delta G$  and normalised monomer loss rate; Figure C: correlations between normalised freeze-drying monomer loss and normalised monomer loss rate.

Only the 8 mutants used for liquid aggregation study are shown. The normalised monomer loss rates,  $\Delta T_m$  from C226S,  $\Delta\Delta G$ , normalised freeze-drying monomer loss were derived from Figure 6.5, Figure 5.6, Figure 5.8 and Figure 5.3, respectively. The  $T_m$  data for LC-S176W was excluded as a sigmoid function could not be fitted. The  $T_m$  values were measured at 10 mM sodium phosphate at pH 7; the liquid aggregation was conducted at 20 mM citrate at pH 4 with NaCl to bring the total ionic strength to 200 mM; the freeze-drying was conducted at 20 mM sodium phosphate at pH 7 with NaCl to bring the total ionic strength to 200 mM.

It is interesting to see whether the impact of mutations upon aggregation could be predicted by their  $T_m$  and Rosetta-based  $\Delta\Delta G$ , and also if there was a connection between the aggregation in the aqueous phase and in freeze-drying. Figure 6.6 showed that there existed a strong correlation ( $R^2 = 0.83-0.92$ ) between  $T_m$  and the liquid aggregation rate, which indicated that the  $T_m$  could confidently reflect the global unfolding and subsequent aggregation that occurred during liquid aggregation. It has been shown previously (Chakroun et al. 2016) that the liquid-state aggregation rate of A33 Fab C226S only correlated well to  $T_m$  under conditions in which the protein is globally unstable, i.e. at temperatures close to the  $T_m$  of the molecules. This was observed for incubations at 65°C, but not at 45°C or below. The results for the mutants are consistent with this observation as the incubation was also at 65°C, and because most of the range in  $T_m$  explored here was due to mutations that were destabilising, and hence more likely to promote aggregation via a global unfolding pathway.

Compared to the correlation between freeze-drying aggregation and  $T_m$  ( $R^2=0.40$ , Figure 5.7), the large fitting decline in freeze-drying implied that aggregation due to freezing and drying involved at least one further mechanism for monomer loss, in addition to simple aggregation driven by global unfolding.

The correlations between liquid aggregation rate and the  $\Delta\Delta G$  predicted by Rosetta were much lower ( $R^2 = 0.39-0.51$ ), though a slight improvement compared to correlation with freeze-drying aggregation ( $R^2=0.33$ , Figure 5.10). This implied that the enthalpy free energy calculated *in silico* could capture one third to half of the denaturing stress occurred in aqueous phase. The poor correlation is most likely to be due to the limitation of the Rosetta calculations, and good correlation to an experimentally determined  $\Delta\Delta G$  cannot, and should not, yet be ruled out.

Figure 6.6 also shows that the monomer loss observed in the aqueous phase could partially be correlated to that in freeze-drying ( $R^2 = 0.53-0.62$ ). This suggests that the mutants undergo aggregation driven by global unfolding in the liquid conditions used

(pH 4, 65°C), and that this mechanism is partly attributable to monomer loss during freeze-drying.

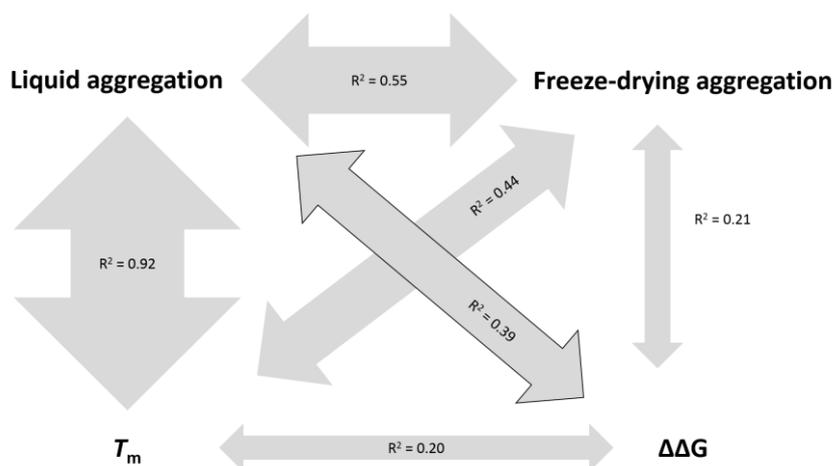


Figure 6.7 A summary of correlations between liquid aggregation rate, freeze-drying aggregation rate,  $T_m$  and  $\Delta\Delta G$ .

The widths of the arrows were proportional to their coefficient of determination ( $R^2$ ). The data were derived from the 8 mutants used for liquid aggregation kinetics. “Liquid aggregation” denoted the normalised aggregation rate for the data without excipient (Figure 6.6); “Freeze-drying aggregation” denoted the normalised monomer loss (Figure 5.3); “ $T_m$ ” denoted the  $\Delta T_m$  against C226S (Figure 5.6); “ $\Delta\Delta G$ ” was derived from Figure 5.8.

Figure 6.7 showed the correlations between liquid aggregation rate, freeze-drying aggregation rate,  $T_m$  and  $\Delta\Delta G$ , in which only the 8 mutants used for liquid aggregation were included. It could be seen that liquid aggregation was much more likely to be accurately predicted compared to aggregation in freeze-drying. The  $\Delta\Delta G$ , however, performed poorly for any of the other three factors, and the correlations were weaker compared to Figure 5.11 where 16-18 mutants were included. Rosetta proposed mutations that could improve the global enthalpy for the whole protein with 69% confidence (Kellogg 2011), which influenced  $\Delta G$  through  $\Delta H$ . However, the poor performance mainly came from not considering entropic contributions to stability, and did not separate out the overall global stability ( $\Delta G$ ) from local stability and fluctuations of surface features. It is well known that decreasing surface flexibility at aggregation-prone regions leads to improved global stability through entropic factors, but that this is not always necessarily the case (Kamerzell & Middaugh 2008; McClelland & Bowler 2016). For this work, a default scoring function was used in Rosetta for the  $\Delta\Delta G$  calculation. It may be worthwhile to manipulate the weights of the scoring function to customise it to this

A33 Fab. It might also be the reason that the mutants analysed here were very limited. So the performance might be improved if more mutants, which cover a wider range of  $\Delta\Delta G$ , were included in the fitting. Therefore,  $\Delta\Delta G$  could only be used qualitatively for the scope of this work. Other *in silico* tools could be considered if surface fluctuations and instabilities were properly addressed.

### **6.3.3 Effect of excipients upon liquid aggregation kinetics**

In this work, only one excipient was tested in each batch. During our experiment, it was observed that the temperature of the incubator dropped 5-10°C when samples were first put into the chamber. This was because the sample holders were kept on ice and the 65°C incubator needed to heat them together with the sample tubes. This partially accounted for the initial 30-60 min plateau where no significant monomer loss was detected. As a result, this plateau period was excluded when fitting the aggregation rate. In order to minimise the batch-to-batch variation, it would be ideal to include C226S without excipients as a common control in every batch, and to normalise incubation conditions. Unfortunately, this was not done and so the batch-to-batch variations were not considered when comparing the stabilising effects of different excipients. However, all mutants could be compared against C226S within the same formulation, as these were measured within a single batch.

All of the linear regression curves between liquid aggregation rate and  $T_m$  for the excipients are shown in Figure 6.6 (A). The C226S results were set as 1 and results for the other mutants were normalised relative to C226S. It can be seen from the figure that all the excipients widened the gaps between the mutants compared to the condition without excipient. This implied that if all the mutants were added with a same excipient, their relative aggregation rates with regard to C226S would be magnified. The sensitivity order from high to low followed the trend of mannitol > glycine > sorbitol > Tween 80. However, it may also be due to the concentration difference used that caused

the sensitivity difference as shown in Table 6.1, though mannitol, sorbitol and glycine shared close molarity.

To examine the absolute stabilising effect from excipients, a generalised criterion should be applied to quantify the aggregation among all the batches so as to draw an unbiased comparison. The aggregation rate between 1-1.5 hour was acceptable as this period could capture the monomer loss of most of the destabilising mutants. However, the LC-G66P mutant (no excipient) had no monomer at 1.5 hour, which implied that it might already lose all of the monomers before the 1.5-hour sampling point as it only retained 30% at 1-hour sampling point. So the aggregation rate was not generalisable for all the mutants to quantify the absolute stabilising effect from excipients. Instead, the absolute monomer retention was used in this work. Figure 6.8 lists the monomer retention at 2.5 hour for C226S and stabilising mutants, and at 1.5 hour for destabilising mutants. These two time points were chosen because they can maximally differentiate the monomer loss for all the mutants. The data in the figure were also normalised against the results with no excipients so as to present the contrast more clearly as influenced by the excipients.

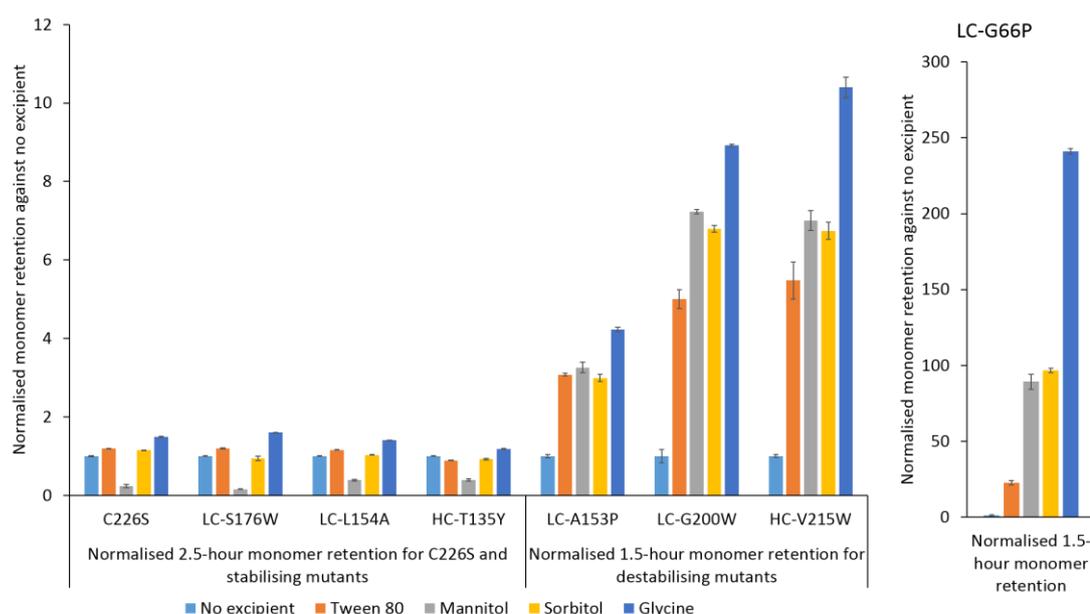


Figure 6.8 The impact of excipients on the liquid monomer retention of mutants. The monomer retention at 2.5 hour was used for C226S and stabilising mutants, and 1.5 hour for destabilising mutants. The data was normalised against results of no excipients for each mutant.

Figure 6.8 showed the impact of excipients on the liquid monomer retention for C226S and stabilising mutants at 2.5-hour sampling point, and destabilising mutants at 1.5-hour sampling point. The excipients predominantly increased the monomer retention of destabilising mutants and by 3-10 fold for LC-A153P, LC-G200W, HC-V215W, and 22-241 fold for LC-G66P. However, excipients exerted a considerably more limited impact upon the stabilising mutants and C226S. For these mutants, Tween 80 and sorbitol had little no difference upon monomer retention compared to the condition without excipient. By contrast, glycine gave an improvement of 50%, whereas mannitol surprisingly reduced the monomer retention to 15-40%.

In general, destabilising mutants were more susceptible to be improved through interacting with excipients. This was reasonable as the mutated residues in the unstable mutants were more flexible, disrupted the structure, and promote unfolding. Through associating with excipients, these flexibilities were decreased and the unfolding was slowed down, which reduced their chance to aggregate with other proteins. For the stabilising mutants, contrarily, the mutants are already globally very stable, and have potentially fewer flexible sites, and so the excipients have less opportunity to improve them under the conditions studied.

The distinctive monomer retention kinetics from mannitol are very interesting. C226S and the stabilising mutants lost more than 70-90% monomer in the first 2.5 hours. Then they increased slightly and remained at 30-40% for C226S, LC-L154A, HC-T135Y, and 12-15% for LC-S176W for at least 1.5 hour before completely aggregating. This relatively stable monomer preservation during the aggregation implied that mannitol might suppress Fab self-association as it unfolded. They may bind to the exposed hydrophobic core of Fab and prevent them from complete unfolding and aggregation. Sorbitol, which is the isomer of mannitol, did not show a similar preserving effect as the monomer concentration for stabilising mutants gradually decreased without a plateau state in their progress. In general, sorbitol preserved the stabilising mutant monomers best, while mannitol preserved the destabilising ones better. The only difference

between mannitol and sorbitol is the orientation of the hydroxyl group on carbon 2. Therefore, the structure of mannitol may be more likely to bind the hydrophobic regions when the protein is unfolded, and so protect the residues mutated to non-polar proline or aromatic tryptophan in the destabilising mutants. Sorbitol retained the monomer concentration of the stabilising mutants, potentially through binding to the polar surface of the folded protein.

Glycine performed better than any other excipients to retain the monomers for both stabilising and destabilising mutants. This was promising as it could serve as a universal stabiliser for Fab in the aqueous phase. It was found previously that glycine suppressed the pH change in sodium phosphate buffer during freezing and stabilised proteins through the preferential exclusion mechanism (Pikal-Cleland et al. 2002). In the preferential exclusion mechanism (Arakawa & Timasheff 1982; Kendrick et al. 1997) the stabilisers are preferentially excluded from the protein surface, which increases the free energy for proteins to denature and thus retains their native states. As shown in Figure 6.4, glycine extended considerably the time for which the monomers were retained, compared to other excipients. For stabilising mutants in glycine, monomer populations did not drop to less than 60% until after 4.5 hours. For two destabilising mutants in glycine, LC-A153P and LC-G200W, the monomer population was also not fully depleted for the same period. The molarity of glycine (266 mM) was higher than mannitol and sorbitol (220 mM) though its mass-concentration was only half (20 mg/ml versus 40 mg/ml, Table 6.1). Therefore, the stabilising effect may also be partly convoluted by the molarity difference.

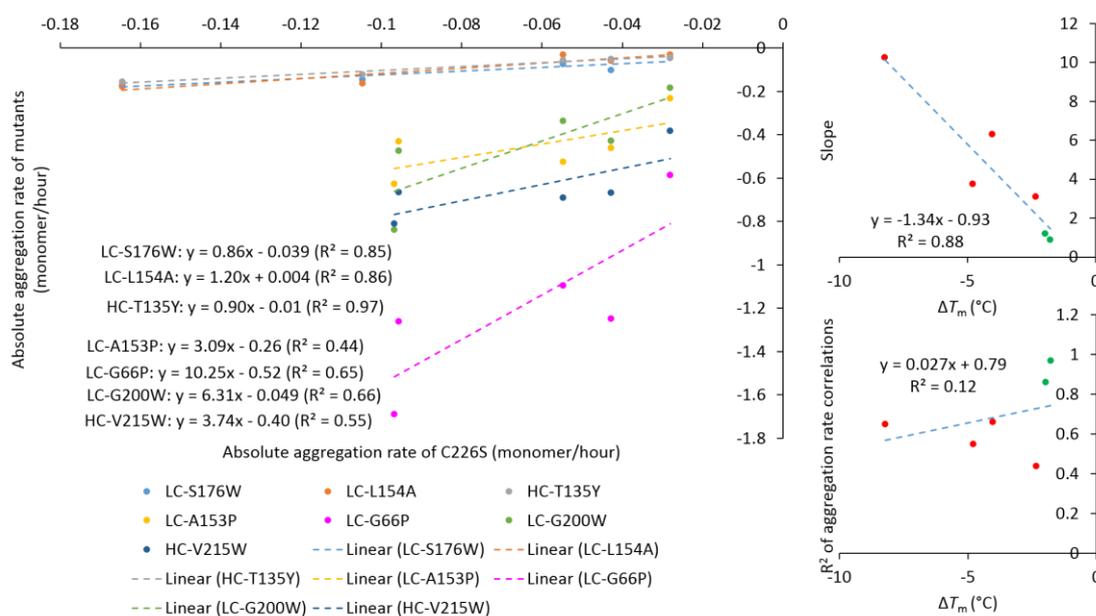


Figure 6.9 The absolute aggregation rate correlations between C226S and designed mutants, and their correlation accuracy as a function of  $\Delta T_m$ .

The aggregation rates were derived from the slopes of the right-hand graphs in Figure 6.4. The  $\Delta T_m$  values were derived from Figure 5.6 (LC-S176W was excluded as a sigmoidal  $T_m$  curve was not obtained). The slope and  $R^2$  values from the left-hand plots were used in the upper-right and lower-right graphs, respectively, in which green and red dots were used to indicate the stabilising and destabilising mutants, respectively.

As the excipients exhibited different effects upon the aggregation of the mutants, it was interesting to examine whether the ranking of excipients in their ability to stabilise C226S could be generalised to other mutants. If that were the case, then once a series of excipients had been tested on one variant, then new variants with similar structures might only need to be screened with the most beneficial excipients. Figure 6.9 correlates the aggregation rates for each mutant with those of C226S for the range of excipient formulations tested. It could be seen that the stabilising mutants had slopes close to 1 (i.e. 0.85-1.2), and  $R^2$  values of more than 0.85, whereas the destabilising mutants had slopes of 3-10, and much lower  $R^2$  values (0.43-0.65). This indicated that the ranking of excipients for C226S was similar to those of the stabilising mutants, but much less predictive of the rank order for the destabilising mutants. As shown on the lower-right of Figure 6.9, the  $R^2$  values of absolute aggregation rate correlations decreased further from 1 as the  $\Delta T_m$  increased in magnitude. This implies that the rank order of the excipients becomes less reliable as the difference in stability ( $\Delta T_m$ )

increases. Interestingly this was the case even though  $\Delta T_m$  was found to correlate well with the absolute aggregation rates of the mutants ( $R^2$  of 0.87), in the aqueous phase.

Admittedly, the present study was not completely rigorous. As mentioned before, there was not a benchmark condition, like C226S without excipient, used for every batch. Therefore, the batch variations were not effectively normalised. The aggregation for destabilising mutants occurred very rapidly so a 30 min difference would result in significantly different results, while different batches may take a different time to heat the tube rack to thermal equilibrium. The excipient effects for LC-G66P may be far exaggerated, as the monomer was nearly completely gone in the absence of excipients, at the 1.5-hour sampling time. Therefore, if the excipient influence needed to be precisely studied, Fab samples added with various excipients should be run in the same batch.

In conclusion, it was shown that 1) the excipients in general would stabilise the destabilising mutants, but had a more limited influence on the stable mutants (Figure 6.8); 2) the mutants' difference with regard to the stabilising effect of excipients was amplified as reflected by their different  $T_m$  (Figure 6.6A); 3) the rank-order of excipient effects for individual mutants, relative to that of wild type, became less similar as the mutant  $\Delta T_m$  magnitude increased (Figure 6.9).

### **6.3.4 The $T_m$ with excipients and its correlation with liquid aggregation kinetics**

In order to investigate if the stabilising roles imposed by excipients could be characterised and predicted by a more efficient measurement with less sample, the  $T_m$  was measured for C226S added with the tested excipients as well as other commonly used excipients as shown in Figure 6.10. It can be seen that most of the excipients increased the  $T_m$  0.5-1°C compared to the condition with no excipient added. Glycine increased 2°C, which outweighed any other excipients. Arginine, however, decreased

the  $T_m$  by nearly 8°C, which may due to its negative effect on protein thermostability, as it contains a guanidine moiety that also acts as a mild detergent (Barata et al. 2016).

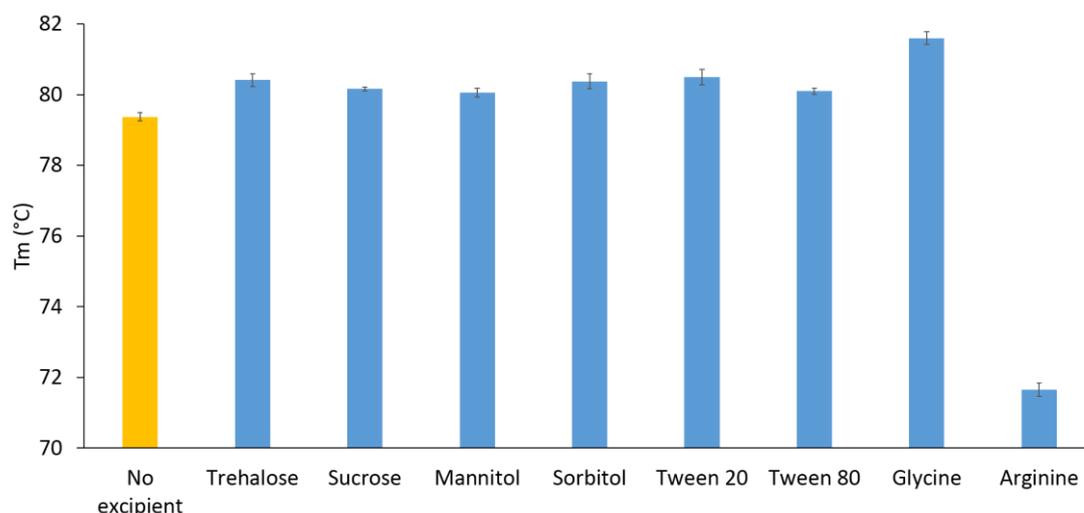


Figure 6.10 The  $T_m$  of 1 mg/ml C226S added with various excipients at 10 mM sodium phosphate, pH 7. The excipient concentration was shown in Table 6.2.

Figure 6.11 correlated the monomer retention for stabilising mutants (2.5 hour) and destabilising ones (1.5 hour) with the  $T_m$  values measured with the different excipients added. The correlations for stabilising mutants were fairly poor ( $R^2 = 0.17-0.30$ ), with the mannitol data being notable outliers. Due to mannitol's unnatural behaviour for the monomer retention during liquid aggregation study, those data need to be conducted again to confirm its preserving mechanism. If the mannitol data was excluded, the correlations were largely increased ( $R^2 = 0.48-0.95$ ). Therefore, the excipients' stabilising effect could be mostly reflected by the  $T_m$  values even for globally stable mutants. The correlations for the destabilising mutants were very strong ( $R^2 = 0.73-0.92$ ). Because the monomers of destabilising mutants were more likely to be preserved by the excipients, the extended improvement on monomer retention could be well captured. The results implied that using a relatively unstable mutant could provide more insight on the stabilising role from excipients. It was noted that all the correlations decreased for mutants compared to C226S. This might because the  $T_m$  values were measured only for C226S. The correlations would be expected to improve if  $T_m$  was

measured for each mutant. Barata (Barata et al. 2016) also conducted a molecular docking study to examine the excipients' role in interacting with aggregation-prone regions on this A33 Fab. It was found that increased  $T_m$  values resulted in decreased protein binding affinity with a strong correlation ( $R^2 = 0.743$ ).

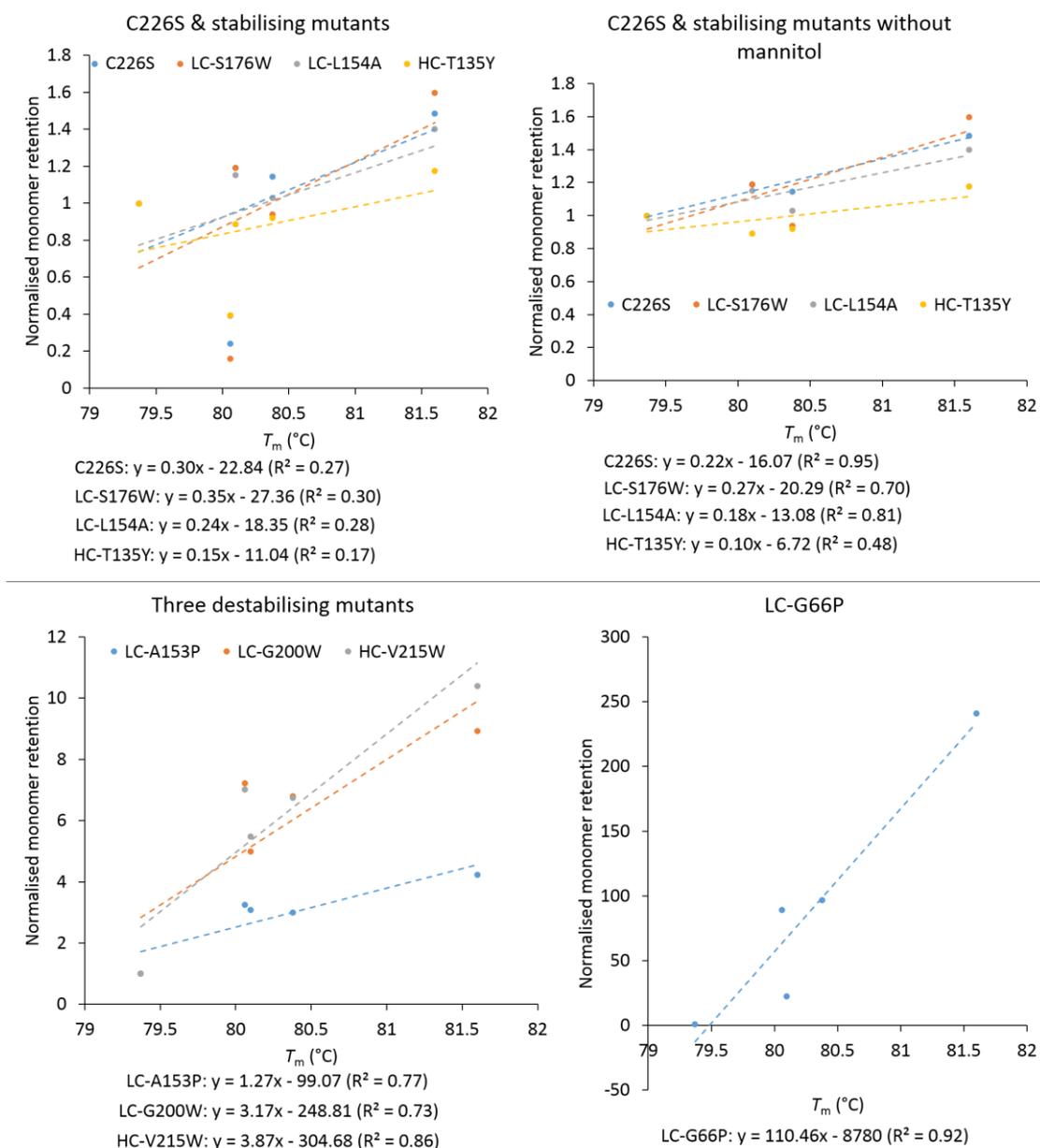


Figure 6.11 The correlations between monomer retention and  $T_m$  as impacted by the excipients. The normalised monomer retention was derived from Figure 6.8; the  $T_m$  was derived from Figure 6.10.

## 6.4 Conclusion

This chapter analysed the liquid aggregation for C226S, three stabilising and four destabilising mutants at pH 4 and elevated temperature at 65°C. Compared to freeze-

drying, a more distinguish difference was observed for aggregation in the aqueous phase as all the destabilising mutants aggregated more than 5 times faster than C226S and the stabilising mutants did. The different mutants' aggregation rates were fully captured by their differences in  $T_m$  ( $R^2 = 0.92$ ). There was also a good correlation between aggregation in aqueous phase and in freeze-drying with  $R^2$  of 0.55.  $\Delta\Delta G$  still behaved poorly in predicting the  $T_m$  ( $R^2 = 0.20$ ), and with a little better accuracy for liquid aggregation ( $R^2 = 0.39$ ) than freeze-drying aggregation ( $R^2 = 0.21$ ), indicating enthalpy stability was more important in aqueous phase than that in freeze-drying.

Excipients' effect was also assessed for Tween 80, mannitol, sorbitol and glycine. It was found that excipients barely exerted influence on the stable mutants but provided sufficient protection for the unstable ones, especially the most unstable LC-G66P mutant. Glycine outperformed than any other excipients probably through preferential exclusion mechanism. However, the excipients could not stabilise unstable mutants to the same level as stable ones at the tested concentration. The excipients stabilising effect was also reflected by their  $T_m$  values, and stronger correlations were found for unstable mutants ( $R^2 = 0.73$ – $0.91$ ) than stable ones ( $R^2 = 0.17$ – $0.30$ ) due to the outlier of mannitol. The rank-order of excipient effects for individual mutants, relative to that of wild type, became less similar as the mutant  $\Delta T_m$  magnitude increased.

## 7 Conclusion

This project determined the key factors that would influence the Fab stability in freeze-drying and aqueous phase. A rapid freeze-drying screening platform was developed with comparable drying rate, in which samples were filled in the middle 60 wells of a 96-well plate, while the outer wells, and surrounding trimmed plates were filled with water. Afterwards, the pseudo wild type C226S and four mutants with altered surface charge were freeze-dried over a range of solution conditions covering pH 4-9, ionic strength (IS) 50-200 mM, and using NaCl or Na<sub>2</sub>SO<sub>4</sub>. In general, losing one positive charge increased the degree of monomer loss during freeze-drying, while adding one slightly improved the stability. Freeze-drying at acidic pH resulted on average in 4-6% more monomer loss than at alkaline pH. Higher ionic strengths mostly caused more aggregation, and yet formulation with Na<sub>2</sub>SO<sub>4</sub> retained 2% more monomer than NaCl on average. An acceptable cake morphology was obtained at the higher ionic strengths, although Na<sub>2</sub>SO<sub>4</sub> again performed better than NaCl. The work suggested that colloidal stability due to electrostatic repulsions had a key role in stabilising proteins against monomer loss during the freeze-drying process.

Several *in-silico* methods were explored to identify potential mutational sites in order to improve the Fab stability. Rosetta software was used to build a homology model and screen for all the possible mutant candidates. RMSF and B-factor were used to restrict the mutants mainly located at the flexible sites. In the meantime, unstable mutants were also prepared as controls to validate the prediction accuracy.

The designed stabilising and destabilising mutants were assessed in freeze-drying and aqueous state, and the stabilising mutants showed significant less monomer loss than the destabilising ones in both of the processes. The  $T_m$  correlated much stronger for aggregation in liquid state than freeze-drying. This implied that the unfolding, which indicated conformation stability, dominated the aggregation in aqueous state at pH 4, 200 mM NaCl and 65°C; but was not prominent in freeze-drying where freezing and drying stress also caused the Fab degradation. The  $\Delta\Delta G$  did not show great impact on Fab

stability in the two processes. This implied that the global stability as depicted by enthalpy might not be sufficiently comprehensive to characterise the aggregation due to local flexibility on the Fab residues. Nevertheless, Rosetta was still a good *in-silico* protein modelling software to qualitatively propose the mutational effect on protein stability.

# 8 Future work

## 8.1 Short-term future work

Due to the limited time available in the PhD study, there were several perspectives that need to be further improved and explored so as to improve the Fab stability in freeze-drying and liquid formulations.

### 8.1.1 Improve the liquid kinetics operations

For liquid aggregation kinetic studies, Figure 6.4 showed there was approximately one hour lag phase in the beginning of incubation. This was because the 65°C incubator needed to heat the cold sample holders together with the sample tubes. As a result, only the data points after one hour were used to derive the aggregation rate. This was subjective as the lag phase might complete earlier or later at different batches due to the varying temperature of the cold sample holders. It would be better if all the samples could be loaded into the incubator to enable direct thermal contact to the air within the chamber. For example, they could be loaded onto several racks and put them into the chamber at a time. In this case, the aggregation could be studied from the very beginning of the incubation without arbitrary lag phase subtraction.

In Figure 6.4, it also showed that the destabilising mutants mostly aggregated in two hours. It turned out that the 30 min interval for sampling is not frequent enough as only 2-3 time points could be used to fit the aggregation kinetics. It would be better if a shorter sampling interval (e.g. 15 min) could be applied so as to capture the monomer loss in greater detail.

## 8.1.2 Compare aggregation driven by global unfolding and native states

Aggregation could be driven by the unfolding of protein and subsequent association of exposed hydrophobic residues. It could also occur when the protein native states were retained while association through local flexible residues. In Chapter 6, the low pH, high salt concentration and high incubation temperature for liquid aggregation study suggested that its aggregation was subject to be driven by global unfolding. However, in Figure 6.6 (A), the samples for  $T_m$  measurement were prepared at pH 7 phosphate buffer without additional salt; while the samples for liquid kinetics were prepared at pH 4 citrate buffer with NaCl to bring the total ionic strength to 200 mM. Therefore,  $T_m$  with conditions the same as liquid kinetics should be conducted so that the  $T_m$  could reasonably reflect the thermal stability of protein at an unfolding treatment.

A parallel study is aggregation driven by local flexibility in native states. This requires the incubation temperature is far from the  $T_m$  and with neutral and low salt concentration. As the Fab wild type is already very stable, it is difficult to increase its  $T_m$  to a large extent. Therefore, liquid incubation at pH 7, 45°C could be used to investigate the aggregation with local instability.

## 8.1.3 Examine the excipients' stabilising effect

Chapter 6 showed that the rank-order of excipient effects for individual mutants, relative to that of wild type, became less similar as the mutant  $\Delta T_m$  magnitude increased. This might be that different destabilising mutants may preferentially interact with typical excipients; while the stable ones were already very stable so their stability could not be further improved by interaction with excipients.

Excipients stabilise the proteins in different ways (Ohtake et al. 2011; Kamerzell et al. 2011). One hypothesis is they interact the flexible residues of proteins, preventing it from partially unfolding or association with other protein molecules. To verify this assumption, the protein-excipient docking energy could be screened by docking software

(e.g. iGEMDOCK, AUTODOCK), and protein-exciptent complex could be simulated to examine their flexibility.

## **8.1.4 Analyse the degradation stresses in freeze-drying**

Compared to the monomer loss in aqueous phase (Chapter 6), the monomer loss in freeze-drying could not be well correlated with  $T_m$  (Chapter 5). The monomer loss of was reduced for some stable mutants though their  $T_m$  did not increase compared to the wild type. This implies that freeze-drying involves more complex stresses (e.g. freezing and drying) that could not be well captured by the thermal stability of proteins. To better understand the degradation mechanism in freeze-drying, the following could be explored.

### ***8.1.4.1 More destabilising mutants for freeze-drying***

The freeze-drying analysis in Figure 5.7, Figure 5.9 and Figure 5.10 showed that the linear correlation was largely influenced by the destabilising mutants, which accounted for less than one third of the total mutants. Therefore, more mutants covering a wider range of  $T_m$ ,  $\Delta\Delta G$  and aggregation propensity could be designed to have a more comprehensive understanding for the Fab stability. This could result in a more unbiased conclusion if the mutants could be evenly distributed to study the predictions of  $T_m$ ,  $\Delta\Delta G$  upon Fab aggregation.

### ***8.1.4.2 Step-by-step study for the monomer loss in freeze-drying***

In this research, the monomer loss was determined based on the reconstituted sample. The whole freeze-drying process could generally be divided into freezing, drying, storage and reconstitution. As the denaturation of proteins could occur at different steps during freeze-drying, it is worthwhile to identify which step causes the most of monomer loss. For example, a comparison of monomer loss between freeze-thawing and freeze-drying could be carried out to determine if freezing or drying accounts

for more monomer loss. Once this is determined, the optimisation of operation parameters could be reduced and more relevant to the typical stress.

The molecular dynamic simulation could be tried at sub-zero temperature to reveal the flexibility of protein in supercooled or frozen state. It would also be beneficial if the shift of hydrogen bond in protein-water to protein-excipient could be simulated to provide *in-silico* evidence for the excipients' stabilising effect.

## 8.2 Long-term future work

### 8.2.1 Improve the mutagenesis strategy

The mutagenesis strategy needs to be improved. Rosetta could differentiate the stabilising effect from destabilising ones. However, its correlations with  $T_m$ , aggregation in liquid state and freeze-drying were fairly poor. As  $T_m$  strongly correlated with aggregation in the liquid state ( $R^2 = 0.92$ ), a protein modelling software that could precisely calculate the  $T_m$  would be very useful. One solution is to examine the software that achieved good performance in the contest Critical Assessment of Structure Prediction (CASP). In general, the prediction tools were developed based on a set of various proteins, which made the structural stability averaged from the extensive training dataset. As a result, they may not be adequately accurate for a particular protein. Therefore, with the  $T_m$  and aggregation data available for the Fab, a combination of several prediction tools, ideally orthogonally covering different structural aspects, could be developed so as to customise it to the protein of interest. In addition, double or triple mutants that encompass advantageous substitutions could be considered to improve the stability and reduce the aggregation propensity.

### 8.2.2 Standardised formulation studies

The formulation studies need to be carried out in more standardised conditions so as to enable cross-comparison with other people's work. The protein structures have already been standardised in PDB format, where identical terms are used to represent

the chain ID, residual numbering and atomic coordinates of different proteins. The formulation studies, however, are still performed without strict criteria. For example, concentration, molarity, ionic strength are used relatively interchangeably among different research groups. In addition, most of the formulation literature did not provide their PDB files, which made readers not straightforward to examine the relations between different formulation performance and their structural difference.

### **8.2.3 Study other antibody structures**

Once the formulation conditions are standardised to certain extent, it is promising to extend the Fab work to other antibody species like F(ab')<sub>2</sub>, scFv, sdAb and full antibodies, and investigate their unfolding and aggregation in the same formulations. This can provide insights about the antibody stability at different levels of molecular weight while maintaining the heterogeneity of protein species in a relatively low level.

### **8.2.4 Activity**

Due to the limited information for the binding affinity of Fab, the Fab activity was not analysed in this work. But it is one of the most important factors that determines the efficacy of the medical drug. For the future work, it would be valuable to develop an assay (e.g. Western blot, ELISA) to ensure the potency of the Fab. In that case, both thermal stability (e.g.  $T_m$ , aggregation) and activity could be used to evaluate the formulation performance.

## 9 Reference

- Abbas, S.A. et al., 2012. Opposite effects of polyols on antibody aggregation: Thermal versus mechanical stresses. *Pharmaceutical Research*, 29(3), pp.683–694.
- Abdelwahed, W. et al., 2006. Freeze-drying of nanoparticles: formulation, process and storage considerations. *Advanced drug delivery reviews*, 58(15), pp.1688–713.
- Abraham, M.J. et al., 2015. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2, pp.19–25.
- Ahmad, S., 2011. *Analysis of the precipitation and aggregation of engineered proteins*. University College London.
- Amini-Bayat, Z. et al., 2012. Relationship between stability and flexibility in the most flexible region of Photinus pyralis luciferase. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1824(2), pp.350–358.
- Angov, E., 2011. Codon usage: Nature's roadmap to expression and folding of proteins. *Biotechnology Journal*, 6(6), pp.650–659.
- Anisimov, M.P., 2003. Nucleation: theory and experiment. *Russian Chemical Reviews*, 72(7), pp.591–628.
- Arakawa, T. et al., 2007. Biotechnology applications of amino acids in protein purification and formulations. *Amino Acids*, 33(4), pp.587–605.
- Arakawa, T. & Timasheff, S.N., 1982. Preferential interactions of proteins with salts in concentrated solutions. *Biochemistry*, 21(25), pp.6545–6552.
- Arnold, K. et al., 2006. The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*, 22(2), pp.195–201.
- Awotwe-Otoo, D. et al., 2012. Quality by design: Impact of formulation variables and their interactions on quality attributes of a lyophilized monoclonal antibody. *International Journal of Pharmaceutics*, 438(1–2), pp.167–175.
- Baldwin, R.L., 1996. How Hofmeister ion interactions affect protein stability. *Biophysical journal*, 71, pp.2056–2063.

- Baneyx, F. & Mujacic, M., 2004. Recombinant protein folding and misfolding in *Escherichia coli*. *Nature Biotechnology*, 22(11), pp.1399–1408.
- Barata, T. et al., 2016. Identification of Protein–Excipient Interaction Hotspots Using Computational Approaches. *International Journal of Molecular Sciences*, 17(6), p.853.
- Barresi, A.A. et al., 2009. Monitoring of the primary drying of a lyophilization process in vials. *Chemical Engineering and Processing: Process Intensification*, 48(1), pp.408–423.
- Bava, K. a., 2004. ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Research*, 32(90001), p.120D–121.
- Beech, K.E. et al., 2015. Insights into the influence of the cooling profile on the reconstitution times of amorphous lyophilized protein formulations. *European Journal of Pharmaceutics and Biopharmaceutics*, 96, pp.247–254.
- Benedix, A., Becker, C. & Groot, B. de, 2009. Predicting free energy changes using structural ensembles. *Nature ...*, 6(1), pp.6–7.
- Bhatnagar, B.S., Pikal, M.J. & Bogner, R.H., 2008. Study of the Individual Contributions of Ice Formation and Freeze-Concentration on Isothermal Stability of Lactate Dehydrogenase during Freezing. , 97(2), pp.798–814.
- Bianco, S. et al., 2013. Bulk, surface properties and water uptake mechanisms of salt/acid amorphous composite systems. *International Journal of Pharmaceutics*, 456(1), pp.143–152.
- Biasini, M. et al., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research*, 42(Web Server issue), pp.W252-8.
- Bishop, B. et al., 2001. Reengineering granulocyte colony-stimulating factor for enhanced stability. *The Journal of biological chemistry*, 276(36), pp.33465–33470.
- Bloom, J.D. et al., 2006. Protein stability promotes evolvability. *Proceedings of the National Academy of Sciences of the United States of America*, 103(15), pp.5869–5874.
- Bloom, J.D. et al., 2005. Thermodynamic prediction of protein neutrality. *Proceedings of the National Academy of Sciences of the United States of America*, 102(3), pp.606–611.

- Bosshard, H.R., Marti, D.N. & Jelesarov, I., 2004. Protein stabilization by salt bridges: Concepts, experimental approaches and clarification of some misunderstandings. *Journal of Molecular Recognition*, 17(1), pp.1–16.
- Breen, E.D. et al., 2001. Effect of moisture on the stability of a lyophilized humanized monoclonal antibody formulation. *Pharmaceutical research*, 18(9), pp.1345–53.
- Brooks, B. & Brooks, C., 2009. CHARMM: the biomolecular simulation program. *Journal of ...*
- Brunauer, S., Emmett, P. & Teller, E., 1938. Adsorption of gases in multimolecular layers. *Journal of the American Chemical Society*, 60(2), pp.309–319.
- Burstein, E.A., Vedenkina, N.S. & Ivkova, M.N., 1973. Fluorescence and the location of tryptophan residues in protein molecules. *Photochem. Photobiol.*, 18, pp.263–279.
- Bye, J.W. & Falconer, R.J., 2014. Three stages of lysozyme thermal stabilization by high and medium charge density anions. *The journal of physical chemistry. B*, 118, pp.4282–6.
- Calloni, G. et al., 2005. Investigating the effects of mutations on protein aggregation in the cell. *The Journal of biological chemistry*, 280(11), pp.10607–13.
- Cao, W. et al., 2013. Rational design of lyophilized high concentration protein formulations-mitigating the challenge of slow reconstitution with multidisciplinary strategies. *European Journal of Pharmaceutics and Biopharmaceutics*, 85(2), pp.287–293.
- Capriotti, E., Fariselli, P. & Casadio, R., 2004. A neural-network-based method for predicting protein stability changes upon single point mutations. *Bioinformatics*, 20(SUPPL. 1), pp.63–68.
- Carr, D., 2002. The handbook of analysis and purification of peptides and proteins by reversed-phase HPLC D. Carr, ed. *Hesperia, CA, USA: Grace Vydac*, p.36.
- Case, D.A. et al., 2005. The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26(16), pp.1668–1688.

Cavallo, L., Kleinjung, J. & Fraternali, F., 2003. POPS: A fast algorithm for solvent accessible surface areas at atomic and residue level. *Nucleic Acids Research*, 31(13), pp.3364–3366.

Cavatur, R. & Suryanarayanan, R., 1998. Characterization of frozen aqueous solutions by low temperature X-ray powder diffractometry. *Pharmaceutical research*.

Chakroun, N. et al., 2016. Mapping the Aggregation Kinetics of a Therapeutic Antibody Fragment. *Molecular Pharmaceutics*, 13(2), pp.307–319.

Chang, . et al., 2005. Mechanism of protein stabilization by sugars during freeze-drying and storage: Native structure preservation, specific interaction, and/or immobilization in a glassy matrix? *Journal of Pharmaceutical Sciences*, 94(7), pp.1427–1444.

Chang, B. & Fischer, N., 1995. Development of an efficient single-step freeze-drying cycle for protein formulations. *Pharmaceutical research*, 12(6), pp.831–837.

Chang, B.S. & Hershenson, S., 2002. Practical approaches to protein formulation development. *Rationale Design of stable protein formulations-theory and practice*, pp.1–25.

Chang, B.S. & Randall, C.S., 1992. Use of subambient thermal analysis to optimize protein lyophilization. *Cryobiology*, 29(5), pp.632–656.

Chang, L. et al., 2005. Effect of sorbitol and residual moisture on the stability of lyophilized antibodies: Implications for the mechanism of protein stabilization in the solid state. *Journal of Pharmaceutical Sciences*, 94(7), pp.1445–1455.

Cheng, J., Randall, A. & Baldi, P., 2006. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins*, 62(4), pp.1125–1132.

Cheng, W. et al., 2012. Comparison of high-throughput biophysical methods to identify stabilizing excipients for a model IgG2 monoclonal antibody: Conformational stability and kinetic aggregation measurements. *Journal of Pharmaceutical Sciences*, 101(5), pp.1701–1720.

Chi, E.Y., Krishnan, S., Randolph, T.W., et al., 2003. Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. *Pharmaceutical research*, 20(9), pp.1325–36.

Chi, E.Y., Krishnan, S., Kendrick, B.S., et al., 2003. Roles of conformational stability and colloidal stability in the aggregation of recombinant human granulocyte colony-stimulating factor. *Protein science : a publication of the Protein Society*, 12(5), pp.903–13.

Chiti, F., Taddei, N., et al., 2002. Kinetic partitioning of protein folding and aggregation. *Nature structural biology*, 9(2), pp.137–43.

Chiti, F., Calamai, M., et al., 2002. Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. *Proceedings of the National Academy of Sciences of the United States of America*, 99(Suppl 4), pp.16419–26.

Chiti, F. & Dobson, C.M., 2006. Protein misfolding, functional amyloid, and human disease. *Annual review of biochemistry*, 75, pp.333–66.

Chivian, D. & Baker, D., 2006. Homology modeling using parametric alignment ensemble generation with consensus and energy-based model selection. *Nucleic acids research*, 34(17), p.e112.

Choi, Y. & Deane, C.M., 2010. FREAD revisited: Accurate loop structure prediction using a database search algorithm. *Proteins: Structure, Function and Bioinformatics*, 78(6), pp.1431–1440.

Chouvenc, P. et al., 2004. Optimization of the Freeze-Drying Cycle: A New Model for Pressure Rise Analysis. *Drying Technology*, 22(March 2015), pp.1577–1601.

Cicerone, M.T. & Soles, C.L., 2004. Fast dynamics and stabilization of proteins: binary glasses of trehalose and glycerol. *Biophysical journal*, 86(6), pp.3836–3845.

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.

Coleman, N.J. & Craig, D.Q.M., 1996. Modulated temperature differential scanning calorimetry: A novel approach to pharmaceutical thermal analysis. *International Journal of Pharmaceutics*, 135(1–2), pp.13–29.

Constantino, H.R. & Pikal, M., 2004. *Lyophilization of Biopharmaceuticals*,

- Corbellini, S., Parvis, M. & Vallan, A., 2010. In-process temperature mapping system for industrial freeze dryers. *IEEE Transactions on Instrumentation and Measurement*, 59(5), pp.1134–1140.
- Craig, D.Q. et al., 1999. The relevance of the amorphous state to pharmaceutical dosage forms: glassy drugs and freeze dried systems. *International journal of pharmaceutics*, 179(2), pp.179–207.
- Craig & Reading, 2007. *Thermal Analysis of Pharmaceuticals*, CRC Press.
- Crescent, 2004. *Moisture Measurement by Karl Fischer Titrimetry* 2nd Edit., GFS Chemicals, Inc.
- D'Amico, S. et al., 2003. Activity-stability relationships in extremophilic enzymes. *Journal of Biological Chemistry*, 278(10), pp.7891–7896.
- Das, R. & Baker, D., 2008. Macromolecular modeling with rosetta. *Annual review of biochemistry*, 77, pp.363–82.
- Daugherty, A.L. & Mersny, R.J., 2006. Formulation and delivery issues for monoclonal antibody therapeutics. *Advanced Drug Delivery Reviews*, 58(5–6), pp.686–706.
- DePristo, M. a, Weinreich, D.M. & Hartl, D.L., 2005. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nature reviews. Genetics*, 6(9), pp.678–687.
- Dill, K.A. et al., 2008. The protein folding problem. *Annual review of biophysics*, 37(1), pp.289–316.
- Dill, K. a & MacCallum, J.L., 2012. The protein-folding problem, 50 years on. *Science (New York, N.Y.)*, 338(6110), pp.1042–1046.
- Dillon, T.M. et al., 2006. Optimization of a reversed-phase high-performance liquid chromatography/mass spectrometry method for characterizing recombinant antibody heterogeneity and stability. *Journal of chromatography. A*, 1120(1–2), pp.112–20.
- Dobson, C.M., 2004. Principles of protein folding, misfolding and aggregation. *Seminars in Cell and Developmental Biology*, 15(1), pp.3–16.
- Doillon, C.J. et al., 1986. Collagen-based wound dressings: control of the pore structure and morphology. *Journal of biomedical materials research*, 20(8), pp.1219–28.

- Dolinsky, T.J. et al., 2004. PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research*, 32(WEB SERVER ISS.), pp.665–667.
- DuBay, K.F. et al., 2004. Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *Journal of molecular biology*, 341(5), pp.1317–26.
- Duddu, S.P. & Dal Monte, P.R., 1997. Effect of glass transition temperature on the stability of lyophilized formulations containing a chimeric therapeutic monoclonal antibody. *Pharmaceutical research*, 14(5), pp.591–5.
- Dupuis, N.F. et al., 2011. The amyloid formation mechanism in human IAPP: dimers have beta-strand monomer-monomer interfaces. *J Am Chem Soc*, 133(19), pp.7240–7243.
- Edwards, D. & Hrkach, J., 2000. STABLE SPRAY-DRIED PROTEIN FORMULATIONS. *WO Patent 2,000,010,541*, 14(21).
- Eijsink, V.G.H. et al., 2004. Rational engineering of enzyme stability. *Journal of biotechnology*, 113(1–3), pp.105–20.
- Ericsson, U.B. et al., 2006. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry*, 357(2), pp.289–298.
- Falconer, R.J. et al., 2011. Stabilization of a monoclonal antibody during purification and formulation by addition of basic amino acid excipients. *Journal of Chemical Technology and Biotechnology*, 86(7), pp.942–948.
- Fernandez-Escamilla, A.-M. et al., 2004. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nature biotechnology*, 22(10), pp.1302–6.
- Fields, P.A., 2001. Review: Protein function at thermal extremes: Balancing stability and flexibility. *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology*, 129(2–3), pp.417–431.
- Fonte, P. et al., 2012. Effect of cryoprotectants on the porosity and stability of insulin-loaded PLGA nanoparticles after freeze-drying. *Biomatter*, 2(4), pp.329–39.
- França, T.C.C., 2015. Homology modeling: an important tool for the drug discovery. *Journal of Biomolecular Structure and Dynamics*, 33(8), pp.1780–1793.

- Franks, F. & Auffret, T., 2007. *Freeze-drying of Pharmaceuticals and Biopharmaceuticals*, RSC Publishing.
- Fraternali, F. & Cavallo, L., 2002. Parameter optimized surfaces (POPS): analysis of key interactions and conformational changes in the ribosome. *Nucleic acids research*, 30(13), pp.2950–2960.
- Fu, H. et al., 2010. Increasing protein stability: importance of  $\Delta C(p)$  and the denatured state. *Protein science : a publication of the Protein Society*, 19(5), pp.1044–52.
- Gabrielsen, M. et al., 2010. Self-interaction chromatography as a tool for optimizing conditions for membrane protein crystallization. *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 1), pp.44–50.
- Garidel, P., Pevestorf, B. & Bahrenburg, S., 2015. Stability of buffer-free freeze-dried formulations: A feasibility study of a monoclonal antibody at high protein concentrations. *European Journal of Pharmaceutics and Biopharmaceutics*, 97, pp.125–139.
- Geigert, J., 2004. *The challenge of CMC regulatory compliance for biopharmaceuticals*, Gardners Books.
- Gill, P., Moghadam, T.T. & Ranjbar, B., 2010. Differential scanning calorimetry techniques: applications in biology and nanoscience. *Journal of biomolecular techniques : JBT*, 21(4), pp.167–93.
- Ginalski, K., 2006. Comparative modeling for protein structure prediction. *Current Opinion in Structural Biology*, 16(2), pp.172–177.
- Gó Mez, G., Pikal, M.J. & Rodríguez-Hornedo, N., 2001. Effect of Initial Buffer Composition on pH Changes During Far-From-Equilibrium Freezing of Sodium Phosphate Buffer Solutions. , 18(1), pp.90–97.
- Grant, Y., Matejtschuk, P., et al., 2012. Freeze drying formulation using microscale and design of experiment approaches: a case study using granulocyte colony-stimulating factor. *Biotechnology letters*, 34(4), pp.641–8.

Grant, Y., Dalby, P.A. & Matejtschuk, P., 2012. Use of design of experiment and microscale down strategies in formulation and cycle development for lyophilization. *American Pharmaceutical Review*, 15(2), pp.75–80.

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), pp.957–64.

Gunasekaran, K. & Nussinov, R., 2007. How Different are Structurally Flexible and Rigid Binding Sites? Sequence and Structural Features Discriminating Proteins that Do and Do not Undergo Conformational Change upon Ligand Binding. *Journal of Molecular Biology*, 365(1), pp.257–273.

Guo, B. et al., 1999. Correlation of second virial coefficients and solubilities useful in protein crystal growth. *Journal of Crystal Growth*, 196, pp.424–433.

Haas, C., Drenth, J. & Wilson, W.W., 1999. Relation between the Solubility of Proteins in Aqueous Solutions and the Second Virial Coefficient of the Solution. *The Journal of Physical Chemistry B*, 103(14), pp.2808–2811.

Hall, T., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, pp.95–98.

Hang, B.Y.S.C. & Endrick, B.R.S.K., 1996. Surface-Induced Denaturation of Proteins during Freezing and Its Inhibition by Surfactants. , 85(12), pp.1325–1330.

Haque, M.K. & Roos, Y.H., 2005. Crystallization and X-ray diffraction of spray-dried and freeze-dried amorphous lactose. *Carbohydrate Research*, 340, pp.293–301.

Harris, R.J., Shire, R.J. & Winter, C., 2004. Commercial manufacturing scale formulation and analytical characterization of therapeutic recombinant antibodies. *Drug Development Research*, 61(3), pp.137–154.

Hayashi, T. & Mukamel, S., 2007. Vibrational-exciton couplings for the amide I, II, III, and A modes of peptides. *The journal of physical chemistry. B*, 111(37), pp.11032–46.

Heljo, P., 2013. *Comparison of disaccharides and polyalcohols as stabilizers in freeze-dried protein formulations.*

- Henzler-Wildman, K. & Kern, D., 2007. Dynamic personalities of proteins. *Nature*, 450(7172), pp.964–972.
- Huang, Y.J. et al., 2014. Assessment of template-based protein structure predictions in CASP10. *Proteins: Structure, Function and Bioinformatics*, 82(SUPPL.2), pp.43–56.
- Izutsu, K. & Kojima, S., 2002. Excipient crystallinity and its protein-structure-stabilizing effect during freeze-drying. *The Journal of Pharmacy and Pharmacology*, 54(8), pp.1033–1039.
- Izutsu, K. & Kojima, S., 2000. Freeze-concentration separates proteins and polymer excipients into different amorphous phases. *Pharmaceutical research*, 17(10), pp.1316–22.
- Izutsu, K.I. et al., 2005. Effect of counterions on the physical properties of L-arginine in frozen solutions and freeze-dried solids. *International Journal of Pharmaceutics*, 301(1–2), pp.161–169.
- Jackel, C. et al., 2010. Consensus protein design without phylogenetic bias. *Journal of Molecular Biology*, 399(4), pp.541–546.
- Kabir, M.M. & Shimizu, K., 2003. Fermentation characteristics and protein expression patterns in a recombinant *Escherichia coli* mutant lacking phosphoglucose isomerase for poly(3-hydroxybutyrate) production. *Applied Microbiology and Biotechnology*, 62(2–3), pp.244–255.
- Kaiyal, W., Khan, U. & Mawlud, S., 2016. Influence of mannitol concentration on the physicochemical, mechanical and pharmaceutical properties of lyophilised mannitol. *International Journal of Pharmaceutics*, 510(1), pp.73–85.
- Kamerzell, T.J. et al., 2011. Protein-excipient interactions: mechanisms and biophysical characterization applied to protein formulation development. *Advanced drug delivery reviews*, 63(13), pp.1118–59.
- Kamerzell, T.J. & Middaugh, C.R., 2008. The complex inter-relationships between protein flexibility and stability. *Journal of Pharmaceutical Sciences*, 97(9), pp.3494–3517.

van der Kamp, M.W. & Daggett, V., 2010. Pathogenic Mutations in the Hydrophobic Core of the Human Prion Protein Can Promote Structural Instability and Misfolding. *Journal of Molecular Biology*, 404(4), pp.732–748.

Kararli, T.T., Hurlbut, J.B. & Needham, T.E., 1990. Glass-rubber transitions of cellulosic polymers by dynamic mechanical analysis. *Journal of pharmaceutical sciences*, 79(9), pp.845–8.

Kasper, J.C., Winter, G. & Friess, W., 2013. Recent advances and further challenges in lyophilization. *European journal of pharmaceutics and biopharmaceutics : official journal of Arbeitsgemeinschaft für Pharmazeutische Verfahrenstechnik e.V.*, 85(2), pp.162–9.

Kaufmann, K.W. et al., 2010. Practically useful: what the Rosetta protein modeling suite can do for you. *Biochemistry*, 49(14), pp.2987–98.

Kawakami, K. & Ida, Y., 2003. Direct Observation of the Enthalpy Relaxation and the Recovery Processes of Maltose-Based. , 20(9), pp.1430–1436.

Kelley, L.A. et al., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10(6), pp.845–858.

Kellogg, E., 2011. Role of conformational sampling in computing mutation - induced changes in protein structure and stability. *Proteins: Structure, ...*, 79(3), pp.830–838.

Kelly, S.M., Jess, T.J. & Price, N.C., 2005. How to study proteins by circular dichroism. *Biochimica et biophysica acta*, 1751(2), pp.119–39.

Kelly, S.M. & Price, N.C., 2000. The use of circular dichroism in the investigation of protein structure and function. *Current protein & peptide science*, 1(4), pp.349–84.

Kendrick, B.S. et al., 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), pp.11917–11922.

Kenneth, G.L., 2005. The physics of snow crystals. *Reports on Progress in Physics*, 68(4), p.855.

- Kerwin, B.A., 2008. Polysorbates 20 and 80 used in the formulation of protein biotherapeutics: Structure and degradation pathways. *Journal of Pharmaceutical Sciences*, 97(8), pp.2924–2935.
- Kett, V., McMahon, D. & Ward, K., 2004. Freeze-drying of protein pharmaceuticals: The application of thermal analysis. *CryoLetters*, 25(6), pp.389–404.
- Kett, V.V., 2001. Modulated temperature differential scanning calorimetry and its application to freeze-drying. *European Journal of Parenteral & Pharmaceutical Sciences*, 6, pp.95–99.
- Khan, S. & Vihinen, M., 2010. Performance of protein stability predictors. *Human Mutation*, 31(6), pp.675–684.
- Kheddo, P. et al., 2014. The effect of arginine glutamate on the stability of monoclonal antibodies in solution. *International Journal of Pharmaceutics*, 473(1–2), pp.126–133.
- King, D., Antoniow, P. & Owens, R., 1995. Preparation and preclinical evaluation of humanised A33 immunoconjugates for radioimmunotherapy. *British journal of ...*, 72, pp.1364–1372.
- King, J., Haase-Pettingell, C. & Gossard, D., 2002. Protein folding and misfolding. *American Scientist*, 90(5), pp.445–453.
- Kodama, T. et al., 2014. Optimization of secondary drying condition for desired residual water content in a lyophilized product using a novel simulation program for pharmaceutical lyophilization. *International Journal of Pharmaceutics*, 469(1), pp.59–66.
- Kourkoumelis, N., 2013. PowDLL, a reusable .NET component for interconverting powder diffraction data: Recent developments. *ICDD Annual Spring Meetings (ed. Lisa O'Neill)*, *Powder Diffraction*, 28, pp.137–48.
- Krasucka, D.M. et al., 2012. Karl Fisher determination of residual moisture in veterinary vaccines -- practical implementation in market monitoring. *Acta poloniae pharmaceutica*, 69(6), pp.1364–7.
- Kueltzo, L. & Wang, W., 2008. Effects of solution conditions, processing parameters, and container materials on aggregation of a monoclonal antibody during freeze–thawing. *Journal of pharmaceutical sciences*, 97(5), pp.1801–1812.

Kuhlman, B., Dantas, G. & Ireton, G., 2003. Design of a novel globular protein fold with atomic-level accuracy. *Science*, 302(November), pp.1364–1368.

Kumar, M.D.S. et al., 2006. ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions. *Nucleic acids research*, 34(Database issue), pp.D204-6.

Lau, Y.M., Taneja, A.K. & Hodges, R.S., 1984. Effects of High-Performance Liquid Chromatographic Solvents and Hydrophobic Matrices on the Secondary and quaternary Structure of a Model Protein. *Journal of Chromatography*, 317, pp.129–140.

Lauer, T. & Agrawal, N., 2012. Developability index: a rapid in silico tool for the screening of antibody aggregation propensity. *Journal of ...*, 101(1), pp.102–115.

Leaver-fay, A. et al., 2014. R 3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules.

Lee, S. et al., 2010. Conformational diversity in prion protein variants influences intermolecular beta-sheet formation. *The EMBO journal*, 29(1), pp.251–62.

Lehermayr, C. & Mahler, H., 2011. Assessment of net charge and protein–protein interactions of different monoclonal antibodies. *Journal of pharmaceutical sciences*, 100(7), pp.2551–2562.

Lehmann, M. et al., 2000. The consensus concept for thermostability engineering of proteins. *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, 1543(2), pp.408–415.

Lehmann, M. & Wyss, M., 2001. Engineering proteins for thermostability: The use of sequence alignments versus rational design and directed evolution. *Current Opinion in Biotechnology*, 12(4), pp.371–375.

Li, F. et al., 2010. Cell culture processes for monoclonal antibody production. *mAbs*, 2(5), pp.466–479.

Libbrecht, K., 2001. Morphogenesis on ice: The physics of snow crystals. *Engineering and Science*, 1, pp.10–19.

Linding, R. et al., 2003. Protein disorder prediction: Implications for structural proteomics. *Structure*, 11(11), pp.1453–1459.

Liu, J., 2006. Physical characterization of pharmaceutical formulations in frozen and freeze-dried solid states: techniques and applications in freeze-drying development.

*Pharmaceutical development and technology*, 11(1), pp.3–28.

López-Llano, J. et al., 2006.  $\alpha$ -helix stabilization by alanine relative to glycine: Roles of polar and apolar solvent exposures and of backbone entropy. *Proteins: Structure, Function and Genetics*, 64(3), pp.769–778.

Lowe, D. et al., 2011. *Aggregation, stability, and formulation of human antibody therapeutics* 1st ed., Elsevier Inc.

Manikwar, P. & Majumdar, R., 2013. excipient effects on conformational and storage stability of an IgG1 monoclonal antibody with local dynamics as measured by hydrogen/deuterium - exchange mass. *Journal of ...*, 102(7), pp.2136–2151.

Manning, M.C. et al., 2010. Stability of protein pharmaceuticals: an update. *Pharmaceutical research*, 27(4), pp.544–75.

Martínez, L.M. et al., 2016. Controlled water content for evaluation of denaturation temperature of freeze-dried enzymes. *Thermochimica Acta*, 638, pp.52–57.

Mary, N., 1967. Determination of moisture in crude drugs by gas - liquid chromatography. *Journal of pharmaceutical sciences*, 1(12), pp.1670–1672.

McClelland, L.J. & Bowler, B.E., 2016. Lower Protein Stability Does Not Necessarily Increase Local Dynamics. *Biochemistry*, p.acs.biochem.5b01060.

Mehl, A.F., Crawford, M.A. & Zhang, L., 2009. Determination of myoglobin stability by circular dichroism spectroscopy: Classic and modern data analysis. *Journal of Chemical Education*, 86(5), pp.600–602.

Meister, E. & Gieseler, H., 2008. Freeze-Dry Microscopy of Protein/Sugar Mixtures: Drying Behavior, Interpretation of Collapse Temperatures and a Comparison to Corresponding Glass Transition Data. *Journal of pharmaceutical sciences*, 98(9), pp.3072–3087.

Mensink, M.A. et al., 2015. In-line near infrared spectroscopy during freeze-drying as a tool to measure efficiency of hydrogen bond formation between protein and sugar, predictive of protein storage stability. *International Journal of Pharmaceutics*, 496(2), pp.792–800.

- Miller, M.P. & Kumar, S., 2001. Understanding human disease mutations through the use of interspecific genetic variation. *Hum.Mol.Genet.*, 10(21), pp.2319–2328.
- Millqvist-Fureby, a, Malmsten, M. & Bergenståhl, B., 1999. Surface characterisation of freeze-dried protein/carbohydrate mixtures. *International journal of pharmaceuticals*, 191(2), pp.103–14.
- Milton, N. et al., 2007. Vial breakage during freeze-drying: Crystallization of sodium chloride in sodium chloride-sucrose frozen aqueous solutions. *Journal of Pharmaceutical Sciences*, 96(7), pp.1848–1853.
- Monsellier, E. & Chiti, F., 2007. Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO reports*, 8(8), pp.737–42.
- Morimoto, A. et al., 2002. Aggregation and neurotoxicity of mutant amyloid  $\beta$  (A $\beta$ ) peptides with proline replacement: Importance of turn formation at positions 22 and 23. *Biochemical and Biophysical Research Communications*, 295(2), pp.306–311.
- Moult, J. et al., 2014. Critical assessment of methods of protein structure prediction (CASP) - round x. *Proteins: Structure, Function and Bioinformatics*, 82(SUPPL.2), pp.1–6.
- Munch, C. & Bertolotti, A., 2010. Exposure of hydrophobic surfaces initiates aggregation of diverse ALS-causing superoxide dismutase-1 mutants. *Journal of Molecular Biology*, 399(3), pp.512–525.
- Murase, N. & Franks, F., 1989. Salt precipitation during the freeze-concentration of phosphate buffer solutions. *Biophysical Chemistry*, 34(3), pp.293–300.
- Neal, B.L. et al., 1999. Why is the osmotic second virial coefficient related to protein crystallization? *Journal of Crystal Growth*, 196(2–4), pp.377–387.
- Neergaard, M.S. et al., 2014. Stability of monoclonal antibodies at high-concentration: Head-to-head comparison of the IgG1 and IgG4 subclass. *Journal of Pharmaceutical Sciences*, 103(1), pp.115–127.
- Newman, A.N.N. et al., 2008. Characterization of Amorphous API : Polymer Mixtures Using X-Ray Powder Diffraction. , 97(11), pp.4840–4856.

Ng, P.C. & Osawa, Y., 1997. Preparation and characterization of the F (ab)<sub>2</sub> fragments of an aromatase activity-suppressing monoclonal antibody. *Steroids*, 62(12), pp.776–81.

Nick Pace, C., Martin Scholtz, J. & Grimsley, G.R., 2014. Forces stabilizing proteins. *FEBS Letters*, 588(14), pp.2177–2184.

Niesen, F.H., Berglund, H. & Vedadi, M., 2007. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature protocols*, 2(9), pp.2212–2221.

Niroula, A. & Vihinen, M., 2016. Variation Interpretation Predictors: Principles, Types, Performance and Choice. *Human mutation*.

Ohmura, T. et al., 2001. Stabilization of hen egg white lysozyme by a cavity-filling mutation. *Protein science : a publication of the Protein Society*, 10(2), pp.313–320.

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), pp.1053–73.

Oxtoby, D.W., 1992. Homogeneous nucleation: theory and experiment. *Journal of Physics: Condensed Matter*, 4(38), pp.7627–7650.

Pace, C.N. et al., 2011. Contribution of hydrophobic interactions to protein stability. *Journal of Molecular Biology*, 408(3), pp.514–528.

Pace, C.N., Alston, R.W. & Shaw, K.L., 2000. Charge-charge interactions influence the denatured state ensemble and contribute to protein stability. *Protein science : a publication of the Protein Society*, 9(7), pp.1395–8.

Pace, C.N. & Scholtz, J.M., 1998. A helix propensity scale based on experimental studies of peptides and proteins. *Biophysical journal*, 75(1), pp.422–427.

Park, J. et al., 2013. Effect of pH and Excipients on Structure, Dynamics, and Long-Term Stability of a Model IgG1 Monoclonal Antibody upon Freeze-Drying. *Pharmaceutical Research*, 30(4), pp.968–984.

Parrini, C. et al., 2005. Glycine residues appear to be evolutionarily conserved for their ability to inhibit aggregation. *Structure*, 13(8), pp.1143–1151.

- Patel, S.M., Doen, T. & Pikal, M.J., 2010. Determination of End Point of Primary Drying in Freeze-Drying Process Control. *AAPS PharmSciTech*, 11(1), pp.73–84.
- Pavlopoulou, A. & Michalopoulos, I., 2011. State-of-the-art bioinformatics protein structure prediction tools (Review). *International Journal of Molecular Medicine*, 28(3), pp.295–310.
- Pawar, A.P. et al., 2005. Prediction of “aggregation-prone” and “aggregation-susceptible” regions in proteins associated with neurodegenerative diseases. *Journal of molecular biology*, 350(2), pp.379–92.
- Peters, B.H., Molnar, F. & Ketolainen, J., 2014. Structural attributes of model protein formulations prepared by rapid freeze-drying cycles in a microscale heating stage. *European Journal of Pharmaceutics and Biopharmaceutics*, 87(2), pp.347–356.
- Pikal-Cleland, K.A. et al., 2002. Effect of glycine on pH changes and protein stability during freeze-thawing in phosphate buffer systems. *Journal of Pharmaceutical Sciences*, 91(9), pp.1969–1979.
- Pikal-Cleland, K. a et al., 2000. Protein denaturation during freezing and thawing in phosphate buffer systems: monomeric and tetrameric beta-galactosidase. *Archives of biochemistry and biophysics*, 384(2), pp.398–406.
- Pikal, M. et al., 1990. The secondary drying stage of freeze drying: drying kinetics as a function of temperature and chamber pressure☆. *International Journal of Pharmaceutics*, 60(3), pp.203–207.
- Pikal, M. & Rigsbee, D., 1997. The stability of insulin in crystalline and amorphous solids: Observation of greater stability for the amorphous form. *Pharmaceutical research*.
- Pikal, M.J. et al., 1991. The effects of formulation variables on the stability of freeze-dried human growth hormone. *Pharmaceutical Research*, 8(4), pp.427–436.
- Pisano, R., Fissore, D. & Barresi, A. a., 2012. Quality by Design in the Secondary Drying Step of a Freeze-Drying Process. *Drying Technology*, 30(11–12), pp.1307–1316.
- Potapov, V., Cohen, M. & Schreiber, G., 2009. Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein engineering, design & selection : PEDS*, 22(9), pp.553–60.

- Prabakaran, P. et al., 2001. Thermodynamic database for protein-nucleic acid interactions (ProNIT). *Bioinformatics (Oxford, England)*, 17(11), pp.1027–1034.
- Pronk, S. et al., 2013. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7), pp.845–854.
- Pyne, A., Chatterjee, K. & Suryanarayanan, R., 2003. Solute crystallization in mannitol-glycine systems--implications on protein stabilization in freeze-dried formulations. *Journal of pharmaceutical sciences*, 92(11), pp.2272–83.
- Raman, S. & Vernon, R., 2009. Structure prediction for CASP8 with all - atom refinement using Rosetta. *Proteins: Structure, ...*, 77(0 9), pp.89–99.
- Rambhatla, S. et al., 2005. Cake shrinkage during freeze drying: a combined experimental and theoretical study. *Pharmaceutical development and technology*, 10(1), pp.33–40.
- Raymond C Rowe, P.J.S. and M.E.Q., 2009. *Handbook of Pharmaceutical Excipients* R. C. Rowe, P. J. Sheskey, & M. E. Quinn, eds., Pharmaceutical Press.
- Reetz, M.T. & Carballeira, J.D., 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nature protocols*, 2(4), pp.891–903.
- Rezácová, P. et al., 2008. Crystal structure and putative function of small Toprim domain-containing protein from *Bacillus stearothermophilus*. *Proteins*, 70(2), pp.311–319.
- Roberts, D. et al., 2015. Specific ion and buffer effects on protein-protein interactions of a monoclonal antibody. *Molecular Pharmaceutics*, 12(1), pp.179–193.
- Roberts, D. et al., 2014. The role of electrostatics in protein-protein interactions of a monoclonal antibody. *Molecular Pharmaceutics*, 11, pp.2475–2489.
- Rosa, N. et al., 2015. Rapid and Adaptable Measurement of Protein Thermal Stability by Differential Scanning Fluorimetry: Updating a Common Biochemical Laboratory Experiment. *Acta Crystallographica Section F Structural Biology Communications*, 15(7), pp.387–392.
- Ru, M.T. et al., 2000. On the salt-induced activation of lyophilized enzymes in organic solvents: Effect of salt kosmotropicity on enzyme activity. *Journal of the American Chemical Society*, 122(8), pp.1565–1571.

Salomon-Ferrer, R., Case, D.A. & Walker, R.C., 2013. An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 3(2), pp.198–210.

Samra, H.S. & He, F., 2012. Advancements in high throughput biophysical technologies: Applications for characterization and screening during early formulation development of monoclonal antibodies. *Molecular Pharmaceutics*, 9(4), pp.696–707.

Santana, H. et al., 2014. Stabilization of a recombinant human epidermal growth factor parenteral formulation through freeze-drying. *Biologicals*, 42(6), pp.322–333.

Sarciaux, J.M. et al., 1999. Effects of buffer composition and processing conditions on aggregation of bovine IgG during freeze-drying. *Journal of pharmaceutical sciences*, 88(12), pp.1354–61.

Savage, K.N. & Gosline, J.M., 2008. The effect of proline on the network structure of major ampullate silks as inferred from their mechanical and optical properties. *The Journal of experimental biology*, 211(Pt 12), pp.1937–1947.

Schersch, K. et al., 2010. Systematic Investigation of the Effect of Lyophilizate Collapse on Pharmaceutically Relevant Proteins I: Stability after Freeze-Drying. *Journal of pharmaceutical sciences*, 99(5), pp.2256–2278.

Schymkowitz, J. et al., 2005. The FoldX web server: An online force field. *Nucleic Acids Research*, 33(SUPPL. 2), pp.382–388.

Scott, K. a et al., 2007. Conformational entropy of alanine versus glycine in protein denatured states. *Proc. Natl Acad. Sci. USA*, 104(8), pp.2661–2666.

Searles, J. a, Carpenter, J.F. & Randolph, T.W., 2001. Annealing to optimize the primary drying rate, reduce freezing-induced drying rate heterogeneity, and determine T(g)' in pharmaceutical lyophilization. *Journal of pharmaceutical sciences*, 90(7), pp.872–87.

Seeliger, D. & de Groot, B.L., 2010. Protein thermostability calculations using alchemical free energy simulations. *Biophysical journal*, 98(10), pp.2309–2316.

Sharma, S.S., Chong, S. & Harcum, S.W., 2005. Simulation of large-scale production of a soluble recombinant protein expressed in *Escherichia coli* using an

intein-mediated purification system. *Applied biochemistry and biotechnology*, 126(2), pp.93–118.

Sharp, P.M. et al., 1988. Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity. *Nucleic Acids Research*, 16(17), pp.8207–8211.

Sheinerman, F.B., Norel, R. & Honig, B., 2000. Electrostatic aspects of protein-protein interactions. *Current Opinion in Structural Biology*, 10(2), pp.153–159.

Shire, S.J., Shahrokh, Z. & Liu, J.U.N., 2004. Challenges in the Development of High Protein Concentration Formulations. , 93(6), pp.1390–1402.

Shoichet, B. & Baase, W., 1995. A relationship between protein stability and protein function. *Proceedings of the ...*, 92(January), pp.452–456.

Shukla, D., Schneider, C.P. & Trout, B.L., 2011. Molecular level insight into intra-solvent interaction effects on protein stability and aggregation. *Advanced drug delivery reviews*, 63(13), pp.1074–85.

Singh, S.K. & Nema, S., 2010. Freezing and Thawing of Protein Solutions. In *Formulation and Process Development Strategies for Manufacturing Biopharmaceuticals*. pp. 625–675.

Singh, S.N. et al., 2014. Dipole-Dipole Interaction in Antibody Solutions: Correlation with Viscosity Behavior at High Concentration. *Pharmaceutical Research*, pp.2549–2558.

Smith, D.K. et al., 2003. Improved amino acid flexibility parameters. *Protein science : a publication of the Protein Society*, 12(5), pp.1060–1072.

Stefani, M. & Dobson, C.M., 2003. Protein aggregation and aggregate toxicity: New insights into protein folding, misfolding diseases and biological evolution. *Journal of Molecular Medicine*, 81(11), pp.678–699.

Steipe, B. et al., 1994. Sequence statistics reliably predict stabilizing mutations in a protein domain. *Journal of molecular biology*, 240(3), pp.188–192.

- Steward, A., Adhya, S. & Clarke, J., 2002. Sequence conservation in Ig-like domains: the role of highly conserved proline residues in the fibronectin type III superfamily. *Journal of molecular biology*, 318(2), pp.935–940.
- Strickler, S.S. et al., 2006. Protein stability and surface electrostatics: A charged relationship. *Biochemistry*, 45(9), pp.2761–2766.
- Tahiri-Alaoui, A. et al., 2004. Methionine 129 variant of human prion protein oligomerizes more rapidly than the valine 129 variant. Implications for disease susceptibility to Creutzfeldt-Jakob disease. *Journal of Biological Chemistry*, 279(30), pp.31390–31397.
- Tan, K.P. et al., 2013. Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. *Nucleic acids research*, 41(Web Server issue), pp.314–321.
- Tang, X. & Pikal, M.J., 2004. Design of freeze-drying processes for pharmaceuticals: practical advice. *Pharmaceutical research*, 21(2), pp.191–200.
- Taylor, M.J., Tanna, S. & Sahota, T., 2010a. Formulation design and high-throughput excipient selection based on structural integrity and conformational stability of dilute and highly concentrated IgG1 monoclonal antibody solutions. *Journal of pharmaceutical sciences*, 99(10), pp.4215–4227.
- Taylor, M.J., Tanna, S. & Sahota, T., 2010b. Systematic Investigation of the Effect of Lyophilizate Collapse on Pharmaceutically Relevant Proteins, Part 2: Stability During Storage at Elevated Temperatures. *Journal of pharmaceutical sciences*, 99(10), pp.4215–4227.
- Taylor, M.J., Tanna, S. & Sahota, T., 2010c. Systematic investigation of the effect of lyophilizate collapse on pharmaceutically relevant proteins I: Stability after freeze-drying. *Journal of pharmaceutical sciences*, 99(10), pp.4215–4227.
- Teilum, K., Olsen, J.G. & Kragelund, B.B., 2011. Protein stability, flexibility and function. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1814(8), pp.969–976.

Telikepalli, S. et al., 2015. Characterization of the Physical Stability of a Lyophilized IgG1 mAb after Accelerated Shipping-Like Stress. *Journal of Pharmaceutical Sciences*, 104(2), pp.495–507.

Timasheff, S.N., 2002. Protein hydration, thermodynamic binding, and preferential hydration. *Biochemistry*, 41(46), pp.13473–13482.

Tokuriki, N. & Tawfik, D.S., 2009. Stability effects of mutations and protein evolvability. *Current Opinion in Structural Biology*, 19(5), pp.596–604.

Tonnis, W.F. et al., 2015. Size and molecular flexibility of sugars determine the storage stability of freeze-dried proteins. *Molecular Pharmaceutics*, 12(3), pp.684–694.

Towns, J.K., 1995. Moisture content in proteins: its effects and measurement. *Journal of chromatography. A*, 705(1), pp.115–27.

Trnka, H., Palou, A., et al., 2014. Near-infrared imaging for high-throughput screening of moisture induced changes in freeze-dried formulations. *Journal of Pharmaceutical Sciences*, 103(9), pp.2839–2846.

Trnka, H., Rantanen, J. & Grohganz, H., 2014. Well-plate freeze-drying: a high throughput platform for screening of physical properties of freeze-dried formulations. *Pharmaceutical development and technology*, 7450(1), pp.1–9.

Tscheliessnig, A.L. et al., 2013. Host cell protein analysis in therapeutic protein bioprocessing - methods and applications. *Biotechnology Journal*, 8(6), pp.655–670.

Uchiyama, S., 2014. Liquid formulation for antibody drugs. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1844(11), pp.2041–2052.

Uversky, V.N., 2003. Protein folding revisited. A polypeptide chain at the folding - Misfolding - Nonfolding cross-roads: Which way to go? *Cellular and Molecular Life Sciences*, 60(9), pp.1852–1871.

Venselaar, H. et al., 2010. Homology modelling and spectroscopy, a never-ending love story. *European Biophysics Journal*, 39(4), pp.551–563.

Vottariello, F. et al., 2011. RNase A oligomerization through 3D domain swapping is favoured by a residue located far from the swapping domains. *Biochimie*, 93(10), pp.1846–57.

- Wahl, V. et al., 2015. The influence of residual water on the secondary structure and crystallinity of freeze-dried fibrinogen. *International Journal of Pharmaceutics*, 484(1–2), pp.95–102.
- Wang, W. et al., 2007. Antibody structure, instability, and formulation. *Journal of Pharmaceutical Sciences*, 96(1), pp.1–26.
- Wang, W., 1999. Instability, stabilization, and formulation of liquid protein pharmaceuticals. *International journal of pharmaceutics*, 185(2), pp.129–88.
- Wang, W., 2000. Lyophilization and development of solid protein pharmaceuticals. *International journal of pharmaceutics*, 203(1–2), pp.1–60.
- Wang, W., 2005. Protein aggregation and its inhibition in biopharmaceutics. *International journal of pharmaceutics*, 289(1–2), pp.1–30.
- Wang, W., Nema, S. & Teagarden, D., 2010. Protein aggregation--pathways and influencing factors. *International journal of pharmaceutics*, 390(2), pp.89–99.
- Wang, Z. & Moulton, J., 2001. SNPs, protein structure, and disease. *Human mutation*, 17(4), pp.263–70.
- Webb, B. & Sali, A., 2014. *Comparative protein structure modeling using MODELLER*,
- Wedemeyer, W.J., Welker, E. & Scheraga, H. a, 2002. Current Topics Proline Cis - Trans Isomerization and Protein Folding †. *Biochemistry*, 41, pp.14637–14644.
- Welt, S. et al., 2003. Phase I study of anticolon cancer humanized antibody A33. *Clinical cancer ...*, 9, pp.1338–1346.
- Wiederstein, M. & Sippl, M.J., 2007. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, 35(SUPPL.2), pp.407–410.
- Wijma, H.J., Floor, R.J. & Janssen, D.B., 2013. Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability. *Current Opinion in Structural Biology*, 23(4), pp.588–594.
- Williams, A.D. et al., 2004. Mapping A $\beta$  amyloid fibril secondary structure using scanning proline mutagenesis. *Journal of Molecular Biology*, 335(3), pp.833–842.

- Wu, S.J. et al., 2010. Structure-based engineering of a monoclonal antibody for improved solubility. *Protein Engineering, Design and Selection*, 23(8), pp.643–651.
- Xu, Y. et al., 2005. Conformational transition of amyloid  $\beta$ -peptide. *Proc. Natl. Acad. Sci. USA*, 102(15), pp.5403–5407.
- Yearley, E.J. et al., 2014. Observation of small cluster formation in concentrated monoclonal antibody solutions and its implications to solution viscosity. *Biophysical Journal*, 106(8), pp.1763–1770.
- Yoshioka, S. et al., 2011. Effect of sugars on the molecular motion of freeze-dried protein formulations reflected by NMR relaxation times. *Pharmaceutical Research*, 28(12), pp.3237–3247.
- Yu, H. & Huang, H., 2014. Engineering proteins for thermostability through rigidifying flexible sites. *Biotechnology Advances*, 32(2), pp.308–315.
- Yu, L., 2001. Amorphous pharmaceutical solids: preparation, characterization and stabilization. *Advanced drug delivery reviews*, 48(1), pp.27–42.
- Yu, L. et al., 1999. Existence of a mannitol hydrate during freeze-drying and practical implications. *Journal of Pharmaceutical Sciences*, 88(2), pp.196–195.
- Yuan, Z., Bailey, T.L. & Teasdale, R.D., 2005. Prediction of protein B-factor profiles. *Proteins: Structure, Function and Genetics*, 58(4), pp.905–912.
- Zhang, L. & Skolnick, J., 1998. What should the Z-score of native protein structures be? *Protein science : a publication of the Protein Society*, 7(5), pp.1201–7.
- Zhou, D. et al., 2002. Physical stability of amorphous pharmaceuticals: Importance of configurational thermodynamic quantities and molecular mobility. *Journal of pharmaceutical sciences*, 91(8), pp.1863–72.
- Zhou, X. et al., 1998. Gas chromatography as a reference method for moisture determination by near-infrared spectroscopy. *Analytical Chemistry*, 70(2), pp.390–394.

# 10 Appendix

## 10.1 Pymol visualisation for the PDB of C226S



Figure 10.1 The C226S structure predicted by Rosetta.

The crystal structure of human germline antibody 5-51/O12 (PDB ID 4KMT) was used for the homology modelling. Light chain and heavy chain were coloured in green and cyan, respectively.

## 10.2 Rosetta script

Examples of a mutfile (i.e. mutation file) and an options file were listed here to mutate the 1<sup>st</sup> residue aspartic acid into alanine.

### 10.2.1 Mutfile

```
total 1
1
D 1 A
```

## 10.2.2 Option file

```
-ddg::mut_file  
/home/ucbechz/Scratch/20150203_ddg_monomer_8398mutants/input/mutfile/D1A.mutfile  
-in::file:s  
/home/ucbechz/Scratch/20150203_ddg_monomer_8398mutants/input/C226S.pdb  
-constraints::cst_file  
/home/ucbechz/Scratch/20150203_ddg_monomer_8398mutants/input/input.cst  
-ddg::weight_file soft_rep_design  
-ddg::minimization_scorefunction talaris2013  
-ddg::iterations 50  
-ddg::dump_pdb true  
-ignore_unrecognized_res  
-ddg::local_opt_only false  
-ddg::min_cst true  
-ddg::suppress_checkpointing true  
-in::file::fullatom  
-ddg::mean false  
-ddg::min true  
-ddg::sc_min_only false  
-ddg::ramp_repulsive true  
-unmute core.optimization.LineMinimizer  
-ddg::output_silent true  
-override_rsd_type_limit
```

## 10.3 The $\Delta\Delta G$ of 8398 mutants

Light chain: residue 1-214

Heavy chain: residue 215-442

Residue number &	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	
1 D	-0.266	-4.554		1.972	1.269	0.311	-0.136	-1.118	-0.678	-1.883	-0.006	1.938	-1.892	-6.347	2.638	3.41	1.282	1.65	3.125	-2.594	
2 I	-0.128	-1.901	3.929	9.957	-1.674	0.211	-0.815		5.895	4.929	5.196	4.558	11.043	1.943	9.736	4.341	1.855	3.773	9.953	-0.817	
3 Q	3.197	-3.261	-1.744	0.381	-2.442	-1.581	5.996	-2.909	2.805	4.042	-1.316	0.462	5.491		5.504	4.669	0.597	-0.949	4.853	0.611	
4 M	3.76	2.499	1.524	4.424	-0.641	3.987	-1.662	-4.926	8.841	-3.609		2.259	2.74	6.894	8.301	-0.822	3.157	3.749	2.691	5.933	
5 T	0.562	-5.183	6.715	1.361	-0.661	-0.585	-0.105	-4.239	3.151	-2.572	-1.796	-0.755	4.557	-1.292	0.738	1.952		-6.372	-1.448	-1.872	
6 Q	1.227	6.909	9.483	-0.224	8.212	7.115	8.913	0.342	12.007	6.516	-0.516	3.944	0.075		12.171	4.013	2.556	7.845	20.21	16.298	
7 S	2.348	1.44	3.33	-0.903	3.754	-1.275	-0.413	-3.305	0.066	0.709	-1.131	-2.707	8.245		0.223	4.824		-1.436	0.62	-3.419	-1.327
8 P	4.321	3.238	-1.378	-0.645	3.893	2.683	0.1	0.009	5.783	-1.425	0.817	3.072		3.068	0.877	2.422	-1.583	9.678	6.173	2.638	
9 S	3.606	5.416	3.146	-1.269	1.274	-3.365	2.117	-1.737	1.952	1.294	-4.314	-5.102	-5.058	-0.654	5.296		-0.87	1.104	5.156	2.646	
10 S	-1.802	-4.587	1.761	-5.964	-0.814	1.001	-3.115	0.6	1.628	0.947	2.644	-0.412	10.272	-4.537	6.222		1.447	0.714	4.45	2.465	
11 L	3.596	5.346	3.273	2.679	-4.597	8.102	0.621	-1.873	2.235		2.088	5.816	0.121	-3.552	0.921	3.384	1.741	-3.619	-1.203	3.251	
12 S	-2.509	-2.707	-1.813	3.253	-7.845	0.891	-4.823	-3.207	5.228	-4.598	-1.76	-2.61	1.602	-7.841	2.028		0.347	-2.297	-4.706	-4.571	
13 A		4.629	7.161	5.799	15.624	2.497	14.752	3.92	1.165	-0.502	2.509	4.933	6.335	5.795	4.926	3.775	4.886	2.512	10.156	11.886	
14 S	2.662	-1.517	3.938	1.04	4.847	0.581	-1.546	6.206	-3.382	0.738	1.697	-2.534	4.228	0.47	1.725		-0.438	4.726	3.628	2.163	
15 V	-1.594	1.156	5.381	1.62	-1.71	1.439	6.972	-0.844	2.548	-2.048	0.334	1.101	3.906	-1.567	-2.728	-4.861	1.644	-3.695	0.84		
16 G	3.274	4.233	3.792	2.549	0.131		2.306	10.953	9.778	5.852	9.239	5.833	85.343	2.523	7.302	0.062	5.445	4.681	2.69	2.741	
17 D	-1.237	0.555		5.121	2.622	1.29	-4.281	-0.05	0.754	-2.524	-0.788	0.171	6.398	5.111	4.936	-2.039	-0.267	4.766	0.871	2.551	
18 R	3.014	-0.877	7.802	0.238	3.813	3.559	-2.257	3.131	3.145	-2.689	-1.535	-0.334	-4.204	-3.303		-5.717	-2.678	-5.599	-2.911	-0.398	
19 V	2.763	2.483	7.271	9.827	8.791	1.942	8.886	2.46	8.851	1.583	8.23	4.678	5.222	2.443	11.603	3.064	5.75	11.74	13.983		
20 T	-1.182	0.291	0.762	-1.151	-2.141	5.406	-2.531	-3.681	-0.433	-1.255	-1.647		1.3568	0.392	-3.755	0.395		0.147	-2.603	2.011	
21 I	4.141	3.931	12.792	6.05	3.387	8.208	7.726		14.739	2.792	5.671	3.944	10.231	10.215	20.277	2.639	4.375	0.668	14.821	6.472	
22 T	-1.06	-0.371	1.532	-4.987	-6.819	1.872	3.478	-0.412	-2.47	-5.608	-3.364	-4.023	6.685	0.479	1.614	1.736		-4.394	3.286	0.521	
23 C	2.327		-1.553	-3.133	-2.368	3.909	4.628	-0.853	-0.45	3.433	-0.214	-0.904	-0.89	1.081	2.132	-3.821	-0.392	-2.544	1.831	0.669	
24 K	3.334	1.978		7.279	-3.909	-6.542	5.815	-0.14	-3.285		-0.085	4.47	-1.48	7.705	-2.948	-3.921	0.196	1.42	-0.618	-5.713	
25 A		-0.004	0.882	6.645	2.387	4.181	-2.673	2.702	8.292	2.233	-1.374	0.481	3.467	0.897	13.808	-3.843	2.696	3.583	4.322	1.9	
26 S	4.901	1.008	0.864	6.125	7.337	4.549	-1.037	0.266	-3.323	2.13	4.205	0.885	1.13	-1.988	2.473		-4.16	5.52	1.725	2.439	
27 Q	1.547	1.516	4.776	-3.721	1.582	1.166	-5.748	3.653	2.739	1.585	-2.609	0.054	12.964		1.822	1.642	-2.784	1.957	4.06	-0.702	
28 N	-3.506	0.66	-0.11	0.589	0.195	2.889	-0.164	-1.253	-0.173	5.344	0.13		-0.084	-4.356	-1.068	0.598	1.53	-4.464	4.022	5.158	
29 V	2.861	-1.627	6.631	3.182	9.474	3.143	11.51	2.117	14.312	-2.158	6.516	4.493	3.199	8.016	14.471	4.115	1.745	10.759	20.66		
30 R	2.168	-0.368	4.148	-4.223	3.557	-3.207	0.516	1.332	-0.061	-3.618	-1.532	0.201	11.047	2.479		-3.495	-2.396	0.268	-1.683	-0.656	
31 T	4.395	2.935	1.272	5.805	-2.107	3.174	-2.909	-1.278	-3.077	4.514	3.095	-1.499	10.738	-0.42	-1.395	0.477		0.186	-4.255	0.211	
32 V	-2.245	3.526	4.121	3.326	0.178	5.497	-3.13	-2.612	4.754	3.595	0.185	4.741	7.702	-0.152	3.676	5.259	2.47		-6.436	-0.758	
33 V	4.799	2.599	11.587	2.559	-1.556	7.21	2.491	-1.159	9.613	-0.89	-2.561	4.163	8.745	1.578	10.191	2.867	7.185		17.732	16.336	
34 A		4.023	7.927	1.774	-3.083	0.592	-2.433	-1.814	5.324	-0.475	-1.443	0.259	7.705	-0.834	0.807	3.666	-1.638	-4.754	0.557	0.218	
35 W	8.603	5.454	18.78	13.949	2.137	11.723	9.541	6.767	15.933	4.176	7.102	10.27	15.68	13.381	14.259	8.144	14.575	8.237		3.389	
36 Y	7.84	10.719	10.036	13.08	-0.676	8.874	7.955	8.776	15.227	7.686	8.569	6.774	11.273	2.088	3.762	8.923	0.726	0.161	12.811		
37 Q	-0.139	1.244	1.83	0.303	3.328	3.554	-2.415	-2.566	5.819	4.588	-3.438	0.728	13.719		6.73	7.052	5.02	-0.394	5.699	-1.309	
38 Q	9.852	3.873	8.97	7.426	-1.864	4.731	4.012	8.682	4.843	0.293	-1.431	6.15	8.249		-0.694	6.419	0.919	9.775	5.491	2.842	
39 K	3.537	5.224	-0.897	6.144	-2.276	1.274	8.15	-0.044		-3.519	-0.01	-0.638	-0.111	-0.251	2.106	2.378	1.4	-1.267	-3.775	-3.563	
40 P	-1.554	4.531	5.445	4.741	0.68	7.287	4.519	0.657	9.288	0.673	7.568	2.143		0.927	-1.201	5.74	6.125	0.216	2.052	4.922	
41 G	-2.665	6.348	5.186	-1.168	4.163		-1.677	4.071	-2.716	-2.689	2.263	0.845	58.336	-2.34	3.077	1.941	4.666	8.184	2.628	2.52	
42 K	-4.19	0.058	0.052	4.238	-1.391	0.636	-0.774	2.178		-0.238	-0.951	0.775	4.767	-3.518	0.232	-3.332	0.55	-3.399	2.461	4.121	
43 A		3.476	6.942	2.185	-3.91	2.768	-5.153	1.423	1.829	6.898	5.554	-1.044	-1.355	-3.483	1.518	-3.072	4.969	1.769	0.209	-3.069	
44 P	-0.006	8.29	17.404	11.676	5.433	9.992	15.29	13.424	17.272	13.456	13.551	7.071		11.412	20.26	10.701	6.17	4.841	7.853	7.25	
45 K	1.345	0.145	5.039	2.012	-4.462	3.414	1.886	-1.791		1.548	-1.699	0.607	7.803	-0.404	6.445	1.889	0.747	-4.429	1.088	-6.191	
46 T	-0.397	3.145	8.332	8.983	5.618	5.663	2.834	-0.937	8.003	11.877	7.374	3.023	-1.815	5.757	7.33	1.936		5.574	5.633	10.892	
47 L	1.005	4.429	9.735	8.309	6.184	3.351	10.141	0.937	4.116		3.477	0.086	17.078	6.364	12.151	3.963	3.865	-2.75	6.005	3.797	
48 I	2.241	-2.519	4.891	6.957	24.972	7.374	0.623		5.585	-2.519	7.655	6.163	23.212	-2.671	13.77	3.228	6.185	0.961	10.386	14.392	
49 Y	5.873	9.771	8.713	8.047	0.171	11.38	0.848	-0.645	5.111	3.263	4.859	3.203	10.553	11.162	10.826	6.207	5.819	-1.888	12.258		
50 L	2.326	3.199	4.398	3.508	-4.695	7.132	-1.743	5.877	5.881		1.246	4.181	41.124	0.643	6.739	0.796	3.005	-0.059	4.481	-1.158	
51 A		3.141	5.525	3.052	15.917	-5.174	7.048	5.407	6.362	7.917	8.959	1.476	83.233	15.996	-1.756	4.674	2.127	0.429	12.683	19.457	
52 S	-1.957	1.996	-2.117	4.07	-4.688	1.188	1.695	8.448	0.511	-1.821	0.688	-0.823	14.943	4.988	6.399		4.722	1.856	-5.523	-1.117	
53 N	-0.616	2.02	8.227	1.681	0.265	2.195	0.404	-3.097	-0.652	3.928	-0.866		10.784	-4.422	-4.366	4.346	2.19	-3.295	1.904	-1.651	
54 R	2.087	-0.157	4.617	1.571	1.554	9.155	5.827	-3.712	0.084	-2.317	0.798	-1.399	-3.72	2.036		0.401	1.5	0.071	4.118	4.514	
55 H	4.134	2.448	13.249	9.207	0.858	8.896	11.782	12.879	1.784	7.249	7.1	13.364	8.751	15.932	7.116	7.46	3.714	3.535	3.819		
56 T	-1.777	2.329	4.42	0.005	0.935	8.411	3.676	1.44	-1.077	4.641	2.42	1.464	-2.134	4.406	3.362	-5.492		-2.93	3.39	3.113	
57 G	5.263	7.205	0.4	6.95	9.172		8.603	14.063	6.004	13.853	1.889	2.558	90.729	4.084	6.944	6.679	8.866	15.294	9.16	9.254	
58 V	1.919	2.475	4.919	4.975	6.604	3.426	12.476	-2.816	14.359	2.676	4.44	3.182	7.003	3.725	8.151	2.477	3.848		11.133	1.084	
59 P	5.73	6.979	5.779	8.468	1.215	-0.486	-0.204	9.329	9.396	4.925	3.538	0.367		6.617	9.986	0.274	2.679	3.021	1.643	2.612	
60 S																					

Residue number &	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	
101G	13.054	12.37	9.759	8.181	25.107		13.507	14.353	15.194	7.7	10.382	11.284	96.961	6.607	13.789	8.846	6.659	14.412	27.294	23.366	
102T	4.28	3.509	4.854	11.552	8.577	4.805	-0.086	11.115	7.247	3.7	2.396	4.731	11.28	2.266	22.218	-0.307	0.675	17.905	10.157		
103K	-1.458	-1.891	-4.823	0.251	-7.504	3.764	-3.64	0.338		-0.146	-2.279	-1.546	-2.368	-4.335	-5.07	-3.333	-7.243	-3.397	-8.296		
104V	0.84	-0.613	6.839	4.968	12.477	4.014	2.677	0.662	10.859	-0.067	9.854	4.651	9.748	3.069	18.406	9.578	-1.299	11.17	12.65		
105E	-5.507	-0.457	-0.13		-2.346	1.884	-1.226	-2.474	0.07	-6.243	-2.957	-2.574	2.848	-1.828	-2.89	2.201	0.03	-0.954	-3.515	-6.106	
106I	1.414	-1.205	0.865	2.727	-2.06	1.591	0.573		3.01	3.706	0.217	5.44	1.028	2.583	0.081	4.276	4.605	4.307	5.664	-4.526	
107K	4.276	-1.144	2.745	0.481	1.037	1.051	-1.277	5.724		-3.614	4.745	-4.891	3.336	-4.216	-1.583	-3.655	-2.434	4.766	0.794	3.555	
108R	0.212	1.04	0.302	4.633	1.508	3.429	-2.368	-4.834	4.327	3.948	0.6	-0.542	7.042	5.542		-0.098	-1.937	1.222	0.304	6.857	
109T	-5.691	-0.116	0.252	0.291	2.031	1.643	-2.536	-1.796	1.464	-0.564	4.916	0.14	-3.155	0.83	1.731	1.837		2.863	5.795	-0.852	
110V	0.657	1.809	-0.25	3.188	0.344	4.029	0.59	-0.734	0.457	2.789	-0.223	-2	-2.079	4.807	3.875	0.254	-0.733		2.434	-0.526	
111A		-2.295	1.771	1.258	3.251	4.031	0.705	6.03	3.241	-0.187	0.263	7.044	7.256	3.729	-0.604	1.626	-6.39	2.467	2.148	4.455	
112A		0.19	5.323	2.865	8.686	-0.909	6.439	3.659	7.249	7.352	6.917	3.348	-5.952	-0.861	4.926	3.999	2.996	5.542	6.349	10.259	
113P	3.493	6.409	15.724	8.202	25.765	6.19	15.965	10.403	20.1	19.051	18.039	12.209		2.843	24.409	3.177	10.916	11.67	26.718	18.775	
114S	0.801	-0.028	1.572	-5.707	-1.495	0.916	0.958	1.251	-0.699	-3.448	1.446	-1.611	11.519	-6.633	0.227		-2.016	-4.248	-0.234	-3.967	
115V	3.588	8.075	4.541	6.94	15.019	7.315	7.805	5.813	2.772	-0.728	7.956	4.622	2.217	7.076	18.839	8.479	2.719		10.585	17.299	
116F	6.595	-0.566	5.064	10.408		6.576	6.391	5.592	5.555	6.616	3.532	1.092	17.86	6.575	5.558	4.578	4.331	8.655	-0.194	5.3	
117I	0.18	7.326	10.099	5.332	9.002	8.011	4.941		12.012	0.757	6.361	0.097	7.332	8.713	19.069	3.381	5.291	2.815	6.801	5.05	
118F	6.804	7.239	14.087	10.876		9.557	3.577	4.383	5.533	3.194	10.033	6.375	16.205	4.426	8.077	9.313	9.918	-1.427	7.703	2.595	
119P	4.173	4.01	8.81	9.958	0.961	4.835	3.057	5.628	5.141	1.11	4.217	1.998		3.758	10.771	2.643	6.863	6.127	11.67	3.645	
120P	4.486	6.178	7.279	6.238	11.164	5.352	4.465	14.67	18.355	25.918	12.706	11.03		15.032	13.038	0.049	8.408	6.054	14.506	15.505	
121S	-0.259	4.618	0.881	0.261	-0.551	0.222	-4.479	0.177	5.529	6.485	-1.856	1.605	-6.927	0.087	4.018		-1.059	-2.551	-4.592	1.657	
122D	2.725	3.646		0.964	-0.89	3.666	4.272	3.205	1.079	3.69	4.944	0.479	-0.217	0.707	2.779	2.048	-0.735	-1.232	-0.928	-2.466	
123B	-0.928	4.022	6.483		-0.195	4.363	2.832	2.261	2.73	1.299	-3.529	4.677	1.928	-0.035	0.327	6.125	2.912	3.347	2.798	-1.371	
124Q	2.22	7.869	4.268	-0.785	-0.027	3.574	3.829	6.341	9.075	2.554	6.468	4.91	6.406			11.608	3.154	1.401	7.881	7.924	1.484
125L	1.693	0.066	2.574	8.536	5.202	2.858	5.008	-2.478	5.102		2.313	-2.471	12.432	1.32	2.771	3.123	-2.044	-0.068	-0.453	-1.81	
126K	-2.894	4.773	-0.135	3.407	-4.438	2.727	3.422	1.768		-0.402	5.853	0.69	4.97	-2.179	3.95	-2.816	4.623	2.849	-1.486	0.637	
127S	3.956	4.372	1.959	-2.683	-4.645	5.938	-5.417	2.571	2.379	1.624	-3.666	-5.147	10.284	0.301	0.978		-0.731	2.786	1.504	-2.03	
128G	3.302	-0.337	7.802	4.012	2.004		-0.38	2.9	5.334	0.309	1.673	5.118	86.686	2.141	1.93	4.07	2.055	7	2.805	6.264	
129T	-2.211	-0.637	0.389	0.311	-4.092	-3.124	-5.933	3.447	2.512	0.064	1.786	2.147	6.685	-1.544	-1.411	-2.202		-0.633	-4.772	2.396	
130A		5.981	22.259	11.98	21.896	1.304	16.065	15.172	11.734	7.327	2.801	9.823	26.043	3.649	15.871	9.058	10.657	12.254	24.553	27.23	
131S	-2.684	3.276	5.997	3.955	-2.238	4.851	-2.062	3.601	7.284	-0.875	-0.52	-0.936	-2.371	0.344	3.13		-4.363	1.884	-3.601	0.378	
132V	4.136	3.642	9.484	2.599	1.527	2.786	2.217	-0.779	12.611	-1.451	1.946	7.608	9.003	5.975	18.651	0.793	1.365		6.103	3.256	
133V	-0.477	5.078	8.601	5.14	-0.199	6.513	3.121	2.508	5.734	1.85	0.742	2.88	9.136	-1.788	14.114	6.663	-1.77		-2.142	3.053	
134C	-5.23		-2.191	0.55	0.284	-2.568	-1.663	-3.985	4.229	0.001	2.056	-5.106	-0.143	-4.635	3.913	3.172	2.056	3.974	5.117	4.774	
135L	-1.872	2.371	4.476	9.21	6.442	5.508	7.216	-0.909	9.426		0.063	9.292	5.319	2.983	10.618	-3.022	4.756	-0.247	16.132	5.349	
136L	1.648	2.975	3.511	1.593	6.296	6.406	4.981	3.438	7.285		2.501	6.841	14.221	0.348	13.621	3.227	8.098	-5.129	11.713	7.308	
137N	5.363	-0.288	2.937	0.215	6.698	5.096	-0.221	-3.892	3.247	-9.36	4.662		4.544	-2.923	7.767	3.256	3.856	-3.688	15.713	9.013	
138N	5.064	5.617	1.151	3.159	1.249	3.59	1.1	3.562	4.165	-1.676	0.326		15.047	-3.016	6.36	2.308	7.854	6.793	-4.253	4.262	
139F	11.228	12.31	15.996	13.397		7.715	0.859	10.979	17.049	8.901	4.725	8.07	13.563	9.363	15.097	9.94	8.732	10.948	4.944	3.973	
140Y	5.745	1.256	5.727	5.134	-1.958	12.668	2.679	4.587	8.357	7.915	4.675	6.321	26.956	2.898	7.786	5.774	8.042	9.917	2.267		
141P	6.345	2.002	6.54	9.49	2.778	7.051	4.43	7.405	6.929	1.212	3.341	6.196		5.501	0.404	5.575	-1.128	-1.285	2.725	-1.763	
142R	-1.091	-1.7	0.361	3.021	-2.039	-0.232	-6.123	-4.316	3.526	3.6	-3.227	-2.364	-6.007	-1.316		-4.924	-0.377	-2.566	-1.387	-1.173	
143E	-3.053	3.458	-0.511		-1.33	1.899	-2.416	-0.24	0.206	0.924	-1.243	0.017	2.429	0.565	3.807	-0.067	-1.03	1.256	-3.483	-0.517	
144A		2.773	5.649	9.497	5.713	4.281	16.401	1.663	10.568	9.77	10.308	5.376	4.333	7.298	6.684	5.429	3.366	-1.967	8.537	15.456	
145K	0.682	-3.666	-1.759	2.244	3.068	-1.826	1.129	-6.798		3.854	2.83	1.323	4.983	-0.205	1.082	0.27	2.688	-3.921	-0.116	2.185	
146V	1.372	1.131	14.142	3.027	7.973	2.971	7.201	0.846	4.465	2.519	3.649	6.371	2.921	10.565	8.883	2.6	2.17		15.012	16.691	
147Q	-0.16	-3.97	2.636	1.85	1.071	-3.97	5.206	0.636	4.042	-2.789	1.207	-2.546	0.36			-4.763	-2.17	1.917	-3.817	-1.787	-0.713
148W	12.885	10.357	15.639	14.13	0.671	12.461	11.294	7.93	9.642	3.311	10.186	9.004	12.742	10.69	13.767	15.833	9.319	11.633		4.994	
149K	-1.092	-2.289	8.244	-0.681	-6.44	-0.348	-4.041	-6.15		-0.623	0.141	4.728	2.659	-3.372	-0.075	-2.746	-0.529	-0.069	-2.22	-1.416	
150V	2.2	4.721	4.028	2.798	12.182	5.489	5.372	1.867	5.444	5.356	5.429	1.225	8.797	3.087	4.584		-4.54	0.833	7.146	9.072	
151D	1.858	2.517		4.301	-0.11	0.995	-0.063	1.726	5.776	2.719	1.923	-2.214	33.815	5.171	3.694	-1.111	1.335	0.864	1.147	1.033	
152N	-0.105	1.781	-0.341	-0.772	0.681	2.649	4.546	7.701	1.987	1.603	-0.986		79.539	-1.497	5.835	3.889	7.98	0.919	3.842	0.138	
153A		-2.811	-1.943	-1.155	2.815	3.395	-3.104	-2.454	-4.029	-0.441	-4.122	0.413	235.72	-2.44	3.775	-3.077	-1.669	2.748	-1.178	-3.701	
154L	-3.753	-6.041	1.324	-0.494	1.351	3.206	2.902	4.998	1.061		-3.59	-0.494	-2.706	5.139	-2.572	0.904	-0.325	0.575	0.626	-1.1	
155Q	5.388	3.622	2.498	2.789	2.22	1.47	0.082	2.618	7.34	0.993	2.239	0.035	1.11		3.031	2.028	-2.19	5.726	1.171	-1.239	
156S	0.65	4.875	-0.002	-0.858	0.818	4.727	-4.082	4.368	-2.411	4.98	-4.097	2.844	1.989	-1.024	0.223	0.103	5.57	3.476	-0.433		
157G	2.565	1.859	5.79	7.6	7.322		-3.783	9.835	3.999	5.048	3.966	-1.004	56.228	1.62	0.986	5.542	6.845	4.046	5.722	2.817	
158N	1.244	-0.569	0.251	-2.583	6.309	-0.597	-2.122	-0.646	0.127	2.315	4.158		-3.188	1.288	-0.306	6.566	3.826	0.201	1.327	0.352	
159S	0.708	3.036	5.593	7.029	-7.209	0.305	0.411	-5.331	4.632	0.64	-2.972	0.826	8.694								

Residue number &	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y		
201	L	5.055	3.608	9.303	2.094	-3.427	6.815	4.589	4.015	8.044		5.227	5.125	6.635	-0.129	10.021	7.111	0.52	0.025	0.883	2.338	
202	S	-3.315	-3.139	0.209	1.792	0.738	1.463	-4.489	-2.313	-4.111	-1.641	-2.812	-4.766	3.79	-2.094	-5.077		1.943	1.622	5.539	-4.524	
203	S	3.439	-0.308	1.024	3.831	-2.565	-4.826	-7.504	4.999	3.47	-2.613	-1.011	-0.928	11.828	-0.254	4.459		-2.373	5.878	1.356	-1.693	
204	P	-0.751	3.102	2.332	6.237	-0.963	0.283	7.375	4.446	7.05	3.01	1.909	4.155		2.384	-0.826	0.464	3.289	6.054	0.51	4.866	
205	V	5.825	-1.818	0.531	5.942	-1.417	3.062	7.657	-5.174	0.427	-1.215	1.175	1.101	9.804	4.417	7.884	3.524	3.752		-0.303	6.477	
206	T	-3.344	1.353	3.275	1.38	2.494	4.521	-2.089	0.495	2.652	-3.827	0.221	-3.377	0.423	-2.832	4.872	0.711		2.713	3.378	6.737	
207	K	0.877	-2.314	3.077	-2.296	0.137	5.552	-3.084	0.356		1.92	2.664	-0.447	7.087	0.001	5.174	-2.619	-3.697	5.136	3.639	-1.117	
208	S	-1.385	-4.34	2.588	1.964	-5.97	1.288	0.546	-2.597	0.302	-2.209	-5.973	5.081	-0.82	-0.645	1.251		-3.52	3.812	-2.196	-4.921	
209	F	3.478	5.349	12.656	5.226		12.456	7.134	7.645	9.304	5.223	6.424	13.573	20.905	13.244	9.052	9.701	12.259	8.141	4.814	1.059	
210	N	0.985	-0.713	1.676	3.154	-1.777	-0.145	0.477	-1.567	1.029	4.182	-3.659	7.691	-3.946	0.806	2.864	3.184	-2.439	1.293	-0.986		
211	R	-1.344	3.43	1.095	3.514	3.993	2.337	-0.071	0.832	0.616	-1.04	-3.966	2.927	8.996	1.262	0.378	-1.561	3.057	2.509	1.052		
212	G	7.349	1.516	4.531	8.209	1.447		2.91	4.758	5.282	2.401	6.557	-0.44	55.195	2.551	5.195	3.685	6.583	9.161	1.063	3.97	
213	E	-2.741	-5.161	4.657		-0.442	-3.117	2.949	-1.183	1.397	-1.538	6.211	-0.236	14.852	0.063	-0.291	-1.158	0.666	4.19	-4.899	-5.733	
214	C	3.875		0.32	4.468	4.461	1.211	-1.711	0.024	4.017	2.336	2.734	-4.276	4.505	-0.33	1.319	-3.815	-0.924	0.033	2.389	-0.431	
215	E	-2.563	-0.647	1.808		-3.05	2.354	-0.45	-2.617	-0.634	5.58	-0.421	4.996	-0.909	-2.726	1.085	-3.328	-1.724	2.211	3.74	2.225	
216	V	6.177	-3.277	5.496	-1.614	3.236	3.447	0.354	1.057	3.795	6.319	5.542	5.435	13.248	-0.386	2.36	2.962	-0.524		-0.521	-0.112	
217	Q	4.059	1.342	6.709	-2.011	4.652	1.736	0.83	2.577	3.623	2.43						0.818	2.991	-1.721	2.21	-3.383	-0.304
218	L	4.679	4.732	11.706	5.516	-0.879	3.456	5.785	6.282	13.54		0.708	4.437	8.97	-3.52	11.176	2.466	5.431	7.827	9.355	10.02	
219	V	1.495	4.318	1.718	0.317	7.727	3.705	-4.546	-0.611	7.4	0.828	-2.835	-1.667	7.437	-0.672	4.845	-3.475	4.305		0.022	-0.163	
220	E	0.67	-2.126	3.141		9.204	1.297	3.491	-3.495	6.77	-0.864	1.379	-1.64	-0.691	-2.228	11.205	0.842	-2.2	-0.951	11.169	5.034	
221	S	1.006	3.462	-0.178	0.674	3.271	-0.107	7.266	3.28	3.783	3.576	3.049	5.579	6.087	1.317	6.195	4.066	3.334	3.064	0.174		
222	G	-0.427	-0.347	4.699	-0.644	3.874		2.408	-1.807	3.724	-4.985	2.87	-3.896	2.107	-2.019	-2.144	-0.978	-0.124	-5.188	0.127	3.728	
223	G	3.6	8.517	9.336	15.417	31.868		4.711	23.509	4.217	19.819	20.335	12.011	6.588	21.971	4.769	5.642	10.255	9.251	8.804	25.239	
224	G	6.82	-2.96	5.909	3.488	11.018		3.313	2.697	7.57	3.721	2.282	13.15	87.695	-1.782	7.192	8.112	-3.737	5.862	5.702	4.439	
225	L	6.62	-1.856	9.993	8.142	-2.476	4.799	7.558	2.434	4.744		7.286	0.792	3.914	2.478	4.988	5.512	3.106	1.867	0.142	0.674	
226	V	1.031	1.096	6.042	1.297	-3.542	1.713	-5.04	4.521	2.811	4.033	-4.786	2.234	4.721	4.39	-0.971	6.938	-5.003		-6.374	-7.84	
227	Q	2.795	1.553	1.684	3.918	0.08	8.916	-0.015	-2.658	-0.176	-0.211	4.695	-4.515	2.031		-1.333	1.21	0.174	-1.805	-5.42	-0.372	
228	P	2.291	4.206	11.092	8.529	4.605	2.018	3.386	2.333	6.367	3.697	-1.249	4.495		0.637	7.21	2.858	10.121	7.228	9.548	1.635	
229	G	5.606	3.822	2.631	0.091	5.693		1.536	5.658	4.634	6.621	4.846	6.074	66.995	6.832	6.237	10.894	2.425	7.988	-0.745	1.286	
230	G	0.292	-0.551	5.64	-2.301	-3.291		-4.339	3.547	-1.17	-0.827	-0.107	4.945	3.828	-3.482	-3.136	-4.539	-2.408	-2.227	-4.198	0.613	
231	S	-3.957	-0.04	1.794	4.179	1.15	4.637	1.454	-1.37	-2.569	3.221	2.28	2.395	-0.368	-1.709	3.748		-1.588	-1.432	-5.785	3.505	
232	L	6.329	0.064	8.297	0.514	-1.096	5.811	6.1	-0.484	1.121		-3.288	-1.239	5.366	6.635	11.26	-0.643	7.528	-2.814	-0.533	9.368	
233	R	1.777	-2.238	2.314	-2.18	0.227	-0.991	1.703	-1.215	0.681	-5.695	3.061	1.274	2.524	-0.051		-0.253	-3.486	-3.506	6.016	2.238	
234	L	8.501	3.239	10.036	5.701	-1.68	5.882	0.761	1.613	0.999		5.039	0.451	8.867	3.552	10.57	1.961	4.794	-1.123	70.11	4.667	
235	S	1.648	-6.342	1.215	-2.667	-0.621	1.858	-0.269	0.442	-0.407	-1.619	1.777	2.618	3.711	3.311	0.348		-4.015	1.605	-0.618	-2.819	
236	C	0.133		-5.942	-1.295	-2.047	-0.056	-6.038	-0.878	-5.568	-6.069	-6.418	-1.207	-5.452	0.532	-5.927	-5.457	-0.938	-1.607	-2.875	-4.896	
237	A		-5.882	0.981	-0.442	-1.866	0.719	0.216	0.578	1.767	-1.277	0.903	-4.547	1.363	-5.098	-0.382	-0.682	-5.082	-4.33	-0.647	-3.568	
238	A		-1.089	8.514	6.696	-2.371	4.334	1.232	1.005	11.313	6.497	7.207	2.88	10.068	5.405	10.127	2.148	0.337	-0.981	6.815	2.619	
239	S	1.393	0.492	3.967	-3.328	0.504	5.96	1.03	-0.526	0.082	4.799	4.907	2.346	7.35	-0.72	-1.71		0.73	0.29	-1.71	0.788	
240	G	9.903	6.744	5.643	6.283	7.84		2.371	7.974	12.31	14.663	5.539	3.628	97.66	4.617	4.403	4.314	6.634	11.05	9.483	4.353	
241	F	3.308	6.522	9.213	11.828		8.736	6.552	4.397	10.073	8.89	-0.577	10.865	8.677	9.772	12.854	8.752	8.906	6.488	3.151	2.071	
242	A		0.476	-5.888	6.026	-1.453	-2.268	0.492	0.492	0.943	1.384	-1.88	1.404	1.733	0.396	1.895	-0.211	1.496	-1.242	-1.532	4.145	
243	F	8.325	8.229	7.62	13.97		6.601	1.933	4.043	9.756	12.931	12.909	11.647	9.922	8.522	17.578	4.699	7.908	5.782	6.203	0.242	
244	S	3.462	4.181	5.73	1.597	1.888	-3.651	0.402	5.182	-4.535	-0.148	-0.018	1.492	3.796	0.884	3.899		3.186	3.174	-5	1.303	
245	T	0.029	-2.146	-1.609	0.693	-5.797	-1.352	1.871	0.231	5.7	2.078	-0.986	2.325	13.532	0.328	5.514	3.387		-2.951	4.427	-0.547	
246	V	0.163	4.448	7.12	8.642	1.9	5.024	2.441	4.125	1.68	6.78	3.893	-0.296	29.182	3.327	8.476	0.36	3.307	-0.317	2.717		
247	D	-5.932	-5.458		1.523	-2.015	-0.447	-2.544	0.468	-1.541	-1.684	-4.739	-0.638	-2.933	-5.88	-3.942	-2.138	2.821	-0.977	-4.073	-9.929	
248	M	2.904	0.022	3.724	6.268	3.383	5.151	7.63	5.46	4.774	0.88		-0.62	7.224	0.229	11.831	9.109	2.221	3.596	3.593	3.008	
249	S	-0.046	-2.832	5.399	10.407	7.353	3.326	3.291	13.361	14.627	7.907	4.169	5.124	23.997	10.519	12.889		3.032	14.437	14.812	5.043	
250	W	8.605	7.945	14.279	12.758	5.855	11.218	5.731	4.856	9.438	9.889	5.098	6.189	14.484	10.66	15.431	8.818	8.665	2.735		2.962	
251	V	0.1	0.915	6.828	6.273	2.293	11.471	2.605	3.138	1.761	3.683	-2.747	2.443	17.99	5.621	10.145	4.739	0.506		7.724	-0.619	
252	R	0.308	2.667	9.712	4.303	-1.433	-1.386	3.557	1.506	-3.718	-1.316	1.903	5.636	20.691	-2.06		-0.486	4.919	-1.712	3.728	4.267	
253	Q	5.419	4.677	5.277	5.972	3.329	6.021	-6.037	2.302	11.63	1.894	2.145	5.992	18.436		5.601	10.11	6.711	3.791	4.46	6.248	
254	A		3.261	8.3	-1.853	3.146	1.112	-1.349	2.939	0.57	-3.101	0.522	2.73	0.112	6.194	1.777	2.122	5.043	1.357	1.408	-5.277	
255	P	7.424	0.324	0.854	1.087	0.808	5.77	5.072	4.842	1.574	1.642	2.723	3.077		1.824	2.438	5.965	-1.239	-3.339	3.468	2.801	
256	G	3.815	1.414	-0.374	0.105	1.194		3.015	3.736	-7.75	2.696	2.276	-2.474	60.577	0.117	6.339	-2.547	1.249	0.944	4.057	1.496	
257	K	2.854	-3.936	-2.525	1.919	-0.468	0.85	-2.636	0.835		-1.428	2.386	-0.774	6.806	0.059	-2.577	-2.852	0.851	4.676	2.959	-5.277	
258	G	3.547	6.142	1.171	2.278	-4.448		-3.412	4.353	1.129	8.24	4.613	-2.452	1.156	1.528	-0.914	-3.14					

Residue number &	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	
301 R	-1.96	1.24	1.892	-4.489	1.206	1.018	-4.466	-0.136	-0.224	-6.781	-5.188	-4.395	3.162	-3.425		0.559	-2.693	-2.251	-0.181	-2.595	
302 A	-2.843	1.898	8.867	2.997	-4.036	9.879	5.979	-3.949	4.809	0.447	3.875	7.1	8.572	0.043	5.792	4.157	4.604		0.569	-0.283	-1.238
303 E	2.271	4.294	3.975		-2.476	-1.8	1.488	1.744	-0.4	4.641	-1.871	-4.563	-1.388	2.234	3.612	0.186	-2.843	0.441	2.172	4.783	
304 D	-4.179	3.164		2.451	-2.048	2.78	-3.77	-1.955	1.432	-3.709	4.022	-0.965	9.094	-2.411	0.639	4.34	0.602	-2.776	0.568	-2.285	
305 T	2.631	-4.138	4.724	1.685	-2.717	0.199	1.743	2.853	3.762	0.969	3.681	2.205	9.67	1.18	7.268	3.949		3.984	1.228	-3.41	
306 A		-0.43	7.966	9.525	13.252	1.84	13.722	11.714	14.561	8.351	5.337	9.168	22.394	8.976	9.961	2.812	7.071	9.071	14.197	14.613	
307 V	4.882	1.898	8.867	2.997	-4.036	9.879	5.979	-3.949	4.809	0.447	3.875	7.1	8.572	0.043	5.792	4.157	4.604		0.957	-1.812	
308 Y	10.352	10.18	11.532	15.304	-3.136	13.814	6.254	10.359	12.365	12.953	10.247	7.879	9.776	7.834	11.851	9.86	12.003	10.164	-0.707		
309 Y	5.82	11.536	10.957	10.398	-3.822	14.271	1.402	6.171	8.462	0.183	2.549	8.445	17.207	6.354	12.349	12.209	9.621	11.545	5.241		
310 C	-6.426		-0.643	-1.371	-1.839	-5.633	-4.531	-5.58	-6.255	-2.386	-5.426	-6.037	-1.445	-1.133	-5.44	-2.42	-3.202	-6.412	-2.384	-1.147	
311 A		4.532	5.316	5.6	-3.228	-1.022	1.496	5.722	4.558	0.76	-1.953	1.271	2.083	0.352	3.249	0.857	3.47	3.219	6.806	-0.584	
312 P	0.096	4.195	4.235	8.446	17.043	0.436	9.146	0.652	15.335	3.056	11.285	10.364		4.763	12.039	4.831	-1.164	-0.451	17.843	18.004	
313 T	-2.614	1.209	4.573	3.573	11.005	-0.019	5.073	-1.905	6.24	-1.275	2.967	-1.894	7.086	4.231	6.952	-4.779		-0.251	4.073	17.993	
314 T	0.61	3.413	4.247	2.147	2.563	3.53	-2.005	0.918	4.116	4.725	4.826	3.241	3.635	2.803	0.423	4.669		2.078	2.767	6.828	
315 V	2.42	3.044	2.322	3.316	-1.914	0.442	2.413	5.628	5.707	1.433	-0.406	5.733	-0.886	-2.369	-0.664	1.513	0.185		-0.864	-2.004	
316 V	2.663	2.163	-4.701	-2.241	0.496	2.335	1.672	6.388	-5.132	-2.758	-1.106	-3.139	11.529	-4.267	-1.243	-0.222	-1.523		-0.204	-6.931	
317 P	1.935	5.397	3.624	2.966	7.131	4.686	10.18	8.637	1.289	5.531	4.62	-0.058		3.19	8.947	2.848	7.238	6.254	7.407	12.002	
318 F	7.091	7.031	-3.894	5.999		0.124	-0.134	4.777	4.585	-2.25	8.365	5.886	24.275	4.666	2.247	-1.478	7.21	-0.853	7.093	-1.059	
319 A	-4.637	-2.693	2.883	5.999	-6.237	5.481	-5.622	-1.955	0.58	-1.882	-4.484	4.425	16.521	-2.71	5.898	-2.841	2.119	-5.309	3.286	1.986	
320 Y	6.572	6.485	7.473	2.902	-0.502	10.224	-0.975	15.812	5.12	6.607	0.159	10.16	3.138	4.597	13.068	6.25	3.446	12.668	7.766		
321 W	13.347	13.176	8.905	12.298	8.15	15.297	4.51	4.796	14.68	11.116	7.992	7.901	14.278	8.949	12.988	12.376	9.284	9.319		5.148	
322 G	1.863	5.03	12.093	19.756	17.904		21.208	19.518	34.101	22.026	15.728	7.85	8.435	15.371	27.315	5.937	7.039	8.947	25.592	18.766	
323 Q	0.121	4.444	4.479	1.414	-2.11	7.283	0.955	1.797	1.353	2.718	2.432	7.032	-4.104		2.351	1.495	4.49	3.773	-4.635	-2.429	
324 G	9.756	7.105	14.956	13.506	14.91		18.57	15	11.57	2.349	10.44	13.954	86.35	9.484	12.608	3.016	7.938	16.3	30.193	21.462	
325 T	3.709	-1.429	3.801	6.093	13.172	5.19	14.561	-0.724	10.928	3.851	10.716	3.885	5.775	3.493	11.624	3.153		0.445	14.124	12.385	
326 L	4.161	1.45	4.057	1.325	-0.955	3.803	2.576	-0.813	3.106		-2.502	0.967	-2.445	-1.609	6.658	4.711	2.824	0.207	10.184	-3.427	
327 V	3.674	3.137	5.577	9.037	4.508	1.735	4.174	0.03	5.774	1.501	6.038	0.103	7.255	7.067	4.382	1.258	2.928		10.427	5.243	
328 T	-0.058	3.126	7.777	-1.925	2.908	3.767	3.111	-1.394	3.12	-3.42	-0.597	-3.43	-2.544	1.059	-1.537	-2.214		-2.855	-0.73	-1.699	
329 V	-1.83	5.822	1.654	11.546	3.476	5.707	8.868	6.655	14.39	8.777	4.026	4.447	9.646	9.614	15.089	6.47	5.59		4.822	4.929	
330 S	0.662	4.964	-0.03	2.677	4.151	4.334	4.341	-0.097	3.985	0.564	0.286	-1.349	16.058	3.392	8.262		8.793	-0.286	7.37	-1.157	
331 S	4.15	-4.532	5.684	0.385	0.421	1.34	5.098	0.178	1.12	3.36	-4.713	-0.845	-0.307	-5.073	-3.366		2.105	3.825	1.217	-4.102	
332 A		2.439	2.184	2.235	-4.416	1.615	3.727	-0.414	3.647	6.849	-4.162	1.663	4.59	-4.748	2.674	-0.701	-3.149	-3.748	5.085	1.749	
333 S	-1.905	0.623	0.225	2.275	-3.516	2.78	1.144	1.528	-1.971	-4.023	-0.074	2.004	-6.242	0.683	3.169		0.875	4.599	0.663	-1.092	
334 T	-2.521	-3.567	2.319	1.701	-1.035	1.102	0.932	-5.486	0.582	1.997	-1.08	3.891	0.186	3.97	3.121	-0.479		-1.427	1.703	-1.644	
335 K	-2.663	2.159	1.725	0.427	-4.239	8.479	0.16	-3.216		-0.402	-1.789	0.712	7.117	2.374	-0.167	2.217	5.451	-5.969	1.874	-4.355	
336 G	0.183	-3.716	3.05	-1.363	4.044		-5.007	1.193	0.05	0.753	-0.678	2.619	-2.983	0.996	-2.992	-0.435	0.297	5.104	1.301	2.388	
337 P	2.58	4.909	7.961	5.925	26.279	3.329	11.478	3.191	15.922	11.034	12.476	5.917		0.876	25.267	6.734	11.008	3.46	20.316	14.067	
338 S	2.092	-0.739	2.503	-3.27	-1.546	5.084	-5.578	3.835	5.197	-1.683	-3.412	-0.407	3.766	-5.995	-1.035		-2.591	1.122	2.214	-6.68	
339 V	2.816	7.492	7.52	13.752	21.513	4.249	11.217	1.33	20.842	4.979	9.098	8.579	5.353	13.572	15.043	4.605	-2.307		38.267	18.352	
340 F	5.084	8.157	10.945	1.669		2.367	-1.665	6.222	5.707	3.311	5.197	5.12	7.118	9.523	12.49	6.065	3.979	3.767	5.081	0.549	
341 P	6.366	7.842	7.968	10.992	9.694	1.638	4.517	5.252	7.338	0.587	3.583	2.022		1.727	4.812	5.754	5.631	6.575	3.989	6.534	
342 L	4.397	8.339	4.684	6.831	3.836	7.868	8.163	1.573	8.196		1.843	2.915	22.971	4.74	11.477	11.608	7.33	2.835	9.066	7.469	
343 A		5.183	1.059	-0.489	-0.06	-0.183	-3.511	3.299	0.868	-2.142	-2.747	1.57	2.037	-0.876	-1.73	-1.301	2.858	4.897	1.571	1.065	
344 P	9.748	6.451	12.866	7.361	13.013	8.121	1.881	1.931	10.198	9.13	0.923	3.211		7.509	10.82	7.474	4.67	5.94	0.909	-2.924	
345 S	3.483	7.298	2.499	-2.371	-2.552	0.418	-2.643	2.077	-3.533	1.815	-6.213	0.687	-6.888	3.358	1.221		1.948	-4.713	-2.327	-1.288	
346 S	0.408	7.892	4.569	1.632	-5.84	5.846	0.528	0.547	0.907	4.962	-1.716	1.393	7.702	4.124	-2.177		6.964	-2.276	1.173	0.345	
347 K	-2.443	-0.693	1.864	0.277	-3.455	3.909	-2.429	-1.971		0.1	1.105	4.555	3.708	-2.983	3.532	5.335	-0.06	4.544	-3.053	-0.493	
348 S	-2.429	-3.984	1.496	-1.544	0.178	4.888	-1.044	-3.43	4.382	2.679	-5.256	-0.655	-3.98	-3.864	-0.774		1.195	-1.841	1.178	-5.998	
349 T	-0.135	1.106	0.122	0.075	-2.524	6.212	2.809	-6.086	0.935	2.076	5.017	2.555	2.366	0.354	4.76	-2.07		-0.706	-0.529	-2.581	
350 S	-0.461	-1.79	3.483	1.607	2.923	-4.32	4.248	0.878	-0.286	0.914	1.478	-0.09	-0.747	-1.06	0.486		0.603	-0.529	-2.42	-3.073	
351 G	3.403	4.46	-0.075	2.362	5.962		2.439	5.52	2.02	4.478	-1.256	-1.879	65.807	3.305	4.568	3.865	4.43	6.28	4.863	4.164	
352 G	-2.159	-0.114	4.228	3.271	0.886		2.255	13.723	1.74	2.919	-0.009	3.465	6.802	0.991	0.79	8.136	-1.651	6.626	6.765	3.181	
353 T	-2.55	-1.537	7.594	-0.872	-1.169	-2.573	-2.759	2.631	0.001	-3.165	0.943	0.613	-1.482	-3.789	3.538	2.702		0.375	-2.628	-0.654	
354 A		-0.346	-0.972	-1.749	-1.367	-2.472	-2.838	3.976	0.935	-0.251	-0.728	1.58	10.867	-1.932	0.944	-0.357	-1.76	1.954	-0.514	-2.217	
355 A		3.146	2.576	8.244	5.792	4.634	5.637	9.376	9.358	6.002	5.424	-0.643	3.427	5.854	18.196	3.56	-0.015	0.434	13.222	10.229	
356 L	2.702	3.944	7.844	2.322	1.824	11.134	6.645	2.627	3.928		2.943	2.176	8.442	4.852	11.346	2.199	0.505	0.677	11.223	6.564	
357 G	3.587	10.458	22.68	21.316	11.794		19.176	21.146	28.698	18.318	13.724	16.762	22.313	17.372	23.798	12.433	15.537	21.871	23.938	18.775	
358 C	-5.42		1.917	-5.832	-0.544	0.705	0.151	-1.802	1.471	3.576	-3.029	-0.519	1.433	0.158	-0.136	-1.126	-3.247	0.887	3.235	-0.523	
359 L	9.35	5.985	4.067	8.24	-4.055	5.433	1.376	2.708	10.748												

Residue number &	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
401 T	-1.542	-0.762	5.505	0.865	-1.251	-2.055	1.806	-0.874	-6.369	2.898	-0.325	0.797	5.82	0.417	-2.159	-2.841		0.082	-6.34	0.96
402 V	-0.104	6.59	6.449	0.136	5.557	4.974	0.637	-1.725	5.036	0.452	3.517	8.733	2.196	5.887	-1.578	6.718	-0.578		7.884	4.983
403 P	0.387	0.623	2.845	6.157	2.199	2.909	-5.861	2.352	2.465	-2.01	3.283	-0.819		1.055	3.648	-1.357	-1.116	3.336	6.215	-1.779
404 S	-0.506	2.623	11.042	10.522	1.354	0.805	7.991	-2.656	3.235	0.577	2.114	6.901	10.015	2.137	8.627		5.682	7.659	7.589	5.107
405 S	-0.267	1.154	-0.862	1.225	0.377	1.761	-1.858	-0.05	-0.301	-0.284	-2.772	0.367	4.173	-4.852	0.698		-1.401	2.974	1.851	2.699
406 S	6.106	-2.109	5.23	1.411	-3.292	1.997	2.286	3.908	-1.374	3.339	4.063	-3.577	7.433	-3.429	1.559		4.301	-0.431	1.227	2.282
407 L	8.847	6.939	10.46	9.583	2.07	4.385	3.537	4.319	4.497		0.635	-0.095	8.797	2.417	4.909	1.592	4.026	2.106	5.305	10.114
408 G	-0.809	5.609	0.945	0.398	-3.397		-5.381	4.104	-4.179	-2.614	-3.04	-3.835	4.637	-0.289	-1.567	3.061	0.677	-3.663	1.646	-3.44
409 T	-3.235	-2.478	1.061	0.96	-1.6	2.912	1.268	0.248	-0.913	-5.064	3.749	4.434	4.627	-1.387	-0.279	-0.687		-1.663	-3.055	-1.041
410 Q	-3.643	1.001	2.518	7.743	0.414	3.178	3.655	-0.57	-0.032	-2.807	1.098	-2.365	13.214		5.639	1.07	-0.749	4.473	0.623	0.045
411 T	0.505	-1.895	0.861	2.849	-2.456	4.051	3.364	2.347	3.578	-4.003	4.563	2.059	-0.137	1.368	5.257	2.865		3.584	5.136	-1.087
412 Y	7.405	11.336	15.775	8.269	-0.946	5.521	3.363	1.373	8.233	0.491	8.924	10.445	11.947	10.782	11.602	5.824	10.189	6.715	2.755	
413 I	1.978	6.719	3.614	2.456	-0.263	2.329	1.163		5.078	3.557	4.772	4.386	7.798	1.464	0.841	6.234	2.241	0.279	3.182	6.118
414 C	0.372		0.845	-1.382	-4.565	1.662	-0.497	6.331	-3.386	-2.66	2.971	1.214	2.953	0.946	-4.587	-3.672	-3.143	0.414	-1.506	-1.017
415 N	-0.226	1.919	-3.1	-6.435	-2.755	-0.761	-1.971	-3.446	-3.671	0.642	-4.847		2.815	-5.925	7.304	-1.37	1.559	-2.546	-1.803	-7.173
416 V	2.709	3.345	5.661	13.272	16.157	9.367	8.84	1.392	14.141	8.793	8.43	8.227	9.614	3.375	24.482	4.896	3.869		11.213	14.279
417 N	6.536	0.823	-0.873	-3.627	1.01	0.746	-5.281	2.709	3.537	0.608	2.084		5.354	-2.197	2.179	1.876	-2.108	-4.793	-2.941	-4.772
418 H	12.037	9.4	13.09	8.018	3.043	10.385		9.753	8.601	1.024	6.397	11.635	15.75	9.745	13.291	11.248	13.564	9.33	11.258	10.45
419 K	-0.749	4.379	1.961	2.762	-4.333	1.584	-3.536	3.675		1.445	-1.352	-0.109	5.16	-2.864	5.502	-1.386	2.8	-0.589	-2.252	-0.051
420 P	-1.672	0.825	5.829	4.097	-0.146	3.207	5.014	1.176	2.365	8.681	2.297	1.983		-0.562	2.598	9.692	2.109	3.791	2.653	1.868
421 S	2.052	2.139	4.419	7.932	12.421	5.834	8.611	4.107	4.12	2.688	0.666	6.24	13.446	4.749	5.769		0.182	0.367	4.654	14.047
422 N	3.968	0.274	-1.538	1.862	-0.438	-1.745	-0.696	6.312	4.812	-2.326	4.73		52.194	3.564	0.713	3.932	-2.168	7.447	1.634	-3.047
423 T	2.116	2.014	7.725	5.033	4.278	4.136	0.725	0.886	-3.819	1.455	0.907	3.201	17.44	1.776	2.169	2.217		3.144	-1.28	0.694
424 K	-0.914	-2.712	-5.057	1.813	-2.83	-2.069	-2.2	-1.255		-2.443	-5.24	-0.172	0.059	-1.661	-1.397	2.015	-0.201	-2.895	-3.527	-0.361
425 V	0.331	2.2	-0.511	-0.602	-2.687	6.654	6.268	1.401	3.696	-2.142	1.996	5.114	4.654	0.921	-1.758	-1.844	0.133		4.253	3.18
426 D	-1.042	0.879		0.247	-3.885	-3.169	-0.184	1.043	-5.573	-3.332	-6.253	-7.402	0.876	-5.439	-2.326	-4.452	-6.664	-0.971	-3.695	-3.301
427 K	7.47	4.543	9.181	4.039	7.146	7.094	2.349	2.109		0.341	-0.51	-0.708	9.97	3.389	-3.586	-1.423	1.157	-3.589	0.238	2.084
428 K	0.827	0.94	7.348	1.894	-2.073	2.791	-1.054	4.606		4.333	2.132	1.742	-5.795	-0.719	1.447	-1.295	1.818	-1.31	-4.337	-2.76
429 V	4.894	9.44	12.986	14.507	20.067	3.89	12.27	4.473	19.453	10.238	9.542	4.472	3.423	4.9	19.017	7.419	0.793		39.057	24.498
430 E	0.515	2.423	-0.494		-5.275	-2.385	-0.32	-3.533	1.203	-3.471	2.774	3.933	2.521	-0.324	0.16	0.494	0.969	-4.263	-3.623	-6.213
431 P	3.711	2.116	1.753	7.11	6.144	7.747	4.894	6.337	-1.114	2.478	-2.701	5.56		6.442	6.377	5.582	0.81	5.649	-1.282	1.864
432 K	-6.012	-0.517	0.834	2.638	-2.114	-2.292	-1.247	-2.167		-0.975	-5.598	0.5	5.533	-1.204	-0.807	2.289	-0.688	-3.125	-1.995	-1.096
433 S	-3.607	-1.351	4.562	-2.604	2.349	4.451	0.871	-2.655	5.355	-4.214	0.161	0.097	4.278	0.224	-4.636		1.684	2.058	-1.191	-4.956
434 C	1.025		-0.421	0.506	-0.049	2.352	-1.051	-1.278	1.548	2.438	2.128	-2.286	-3.409	-0.417	0.615	3.837	1.582	5.584	1.91	1.025
435 D	0.975	2.269		1.889	-4.332	-2.314	-3.273	4.303	1.15	2.847	-1.953	-0.729	-6.658	-3.014	2.669	0.322	-2.818	-3.39	1.419	-0.642
436 K	1.317	0.68	5.31	2.438	2.228	-3.576	-0.5	-0.257		-4.638	0.162	3.375	6.277	1.757	-0.928	-1.54	-3.831	0.667	1.046	-1.214
437 T	-0.897	3.659	0.702	2.182	0.848	2.77	6.494	4.589	-1.073	6.186	2.598	3.879	3.242	4.901	-1.561	-3.336		0.985	6.294	6.223
438 H	-1.848	4.172	3.511	-1.676	-2.13	7.626		-0.287	-0.976	5.46	-1.056	1.99	0.769	-1.087	2.75	-2.998	-0.798	3.979	0.505	-1.119
439 T	1.531	-3.449	-1.738	-0.732	1.915	1.347	-2.76	4.773	-1.336	-0.365	0.346	3.078	2.657	-2.654	-4.31	-0.121		-3.815	-0.98	4.696
440 S	4.636	-4.364	-2.279	3.519	0.495	-2.046	2.986	3.555	-1.784	-0.833	-3.376	5.218	8.741	0.358	2.697		-0.663	6.009	-0.521	-1.045
441 A		3.746	-1.631	-4.42	0.703	3.923	-1.46	1.497	-0.786	-1.378	-0.709	2.074	14.309	2.58	5.542	1.589	-4.446	10.207	-1.902	4.989
442 A		3.296	4.79	1.074	-0.629	2.381	-3.535	1.792	0.287	1.81	-1.9	-3.324	1.512	0.186	2.373	5.212	0.193	3.159	4.68	1.336

## 10.4 The $\Delta\Delta G$ of all the stabilising mutants and their relation with CDR regions, RMSF and Depth

Residue	CDR	RMSF	Depth (Å)	Aliphatic				Non-polar				Aromatic				Polar				Chargeable			
				A	V	L	I	G	P	C	M	H	F	Y	W	N	Q	S	T	K	R	D	E
1D		0.163	343	-0.3	-1.9	-1.1		-1.9	-4.6	0.0	-0.1	-2.6			-6.3					-0.7			
2I		0.122	521	-0.1					-1.9		-0.8	-1.7	-0.8										
3Q		0.119	382		-0.9	-2.9	-1.6		-3.3	-1.3		-2.4											-1.7
4M		0.098	635		-3.6	-4.9					-1.7	-0.6											-0.8
5T		0.098	421	-6.4	-2.6	-4.2	-0.6		-5.2	-1.8	-0.1	-0.7	-1.9	-1.4	-0.8	-1.3							
6Q		0.085	622							-0.5													-0.2
7S		0.092	396							-1.1	-0.4		-1.3	-3.4	-2.7								-1.4
8P		0.099	395			-1.4																	-1.6
9S		0.117	375			-1.7	-3.4	-5.1		-4.3													-0.9
10S		0.110	390	-1.8					-4.6		-3.1	-0.8			-0.4	-4.5							-6.0
11L		0.099	424		-3.6		-1.9							-4.6	-1.2	-3.6							-0.9
12S		0.105	373	-2.5	-2.3	-4.6	-3.2		-2.7	-1.8	-4.8	-7.8	-4.6	-4.7	-2.6	-7.8							-1.8
13A		0.102	518																				
14S		0.113	429						-1.5		-1.5				-2.5								-0.4
15V		0.114	360	-1.6		-2.0	-0.8					-1.7		-3.7		-1.6	-4.9						2.7
16G		0.115	373																				
17D		0.103	386	-1.2		-2.5	-0.1				-0.8	-4.3											-0.2
18R		0.089	355		-5.6	-2.7			-4.2	-0.9	-1.5	-2.3		-0.4	-2.9	-0.3	-3.3	-5.7	-2.7				
19V		0.084	534																				
20T		0.074	373	-1.2		-1.3	-3.7			-0.3	-1.6	-2.5	-2.1		-2.6								-0.4
21I		0.071	622																				-3.8
22T		0.077	459	-1.1	-4.4	-5.6	-0.4		-0.4	-3.4		-6.8			-4.0								2.5
23C		0.077	741		-2.5		-0.9		-0.9		-0.2				-0.9								-1.6
24K		0.096	417			-0.1	-3.3					-0.1	-6.5	-5.7	-0.6	-1.5	-2.9						-0.9
25A		0.104	475							0.0	-1.4	-2.7											-3.8
26S		0.128	367									-1.0											-4.2
27Q		0.133	399								-2.6	-5.7		-0.7									-2.8
28N		0.129	377	-3.5	-4.5		-1.3		-0.1			-0.2											-4.4
29V	L1	0.112	650			-2.2					-1.6												-0.2
30R		0.121	356			-3.6		-3.2		-0.4	-1.5			-0.7	-1.7								-3.5
31T		0.107	424				-1.3					-2.9	-2.1		-4.3	-1.5	-0.4						-0.1
32V		0.099	524	-2.2			-2.6					-3.1		-0.8	-6.4		-0.2						-3.1
33V		0.078	911			-0.9	-1.2				-2.6		-1.6										-0.4
34A		0.064	1165		-4.8	-0.5	-1.8				-1.4	-2.4	-3.1										-0.8
35W		0.052	1085																				-1.6
36Y		0.053	1094																				
37Q		0.071	700	-0.1	-0.4		-2.6					-3.4	-2.4		-1.3								-0.7
38Q		0.081	501									-1.4		-1.9									-0.7
39K		0.122	416		-1.3	-3.5	0.0		-0.1		0.0		-2.3	-3.6	-3.8	-0.6	-0.3						-0.9
40P		0.175	361	-1.6																			-1.2
41G		0.193	331	-2.7		-2.7						-1.7											-2.7
42K		0.151	384	-4.2	-3.1	-0.2					-1.0	-0.8	-1.4										-3.5
43A		0.113	440									-5.2	-3.9	-3.1		-1.0	-3.5	-3.1					-3.1
44P		0.077	634	0.0																			
45K		0.091	494		-4.4		-1.8				-1.7		-4.5	-6.2									-0.4
46T		0.073	846	-0.4			-0.9		-1.8														
47L		0.071	864		-2.8																		
48I		0.072	816			-2.5					-2.5												-2.7
49Y		0.087	586		-1.9		-0.6																
50L		0.095	442		-0.1							-1.7	-4.7	-1.2									
51A		0.083	534																				-1.8
52S		0.087	405	-2.0		-1.8							-4.7	-1.1	-4.5	-0.8							-2.1
53N		0.096	411	-0.6	-3.3		-3.1				-0.9												-0.7
54R		0.093	375			-2.3	-3.7																-4.4
55H		0.114	646																				-1.4
56T		0.140	400	-1.8	-2.9																		-5.5
57G		0.138	348																				-1.1
58V		0.112	508			-2.8																	
59P		0.120	382									-0.5			-0.2								
60S		0.116	347		-0.8	-2.7	0.0	-1.3				-1.6	-0.7										2.7
61R		0.098	407	-0.3	-1.7	-3.4	-2.0				-0.8	-0.8											-0.7
62F		0.080	635																				
63S		0.078	410		-0.4	-1.8	-5.0					-1.0	-4.6		-1.1								
64G		0.073	486																				
65S		0.084	419	-1.4																			-2.3
66G		0.095	381																				-3.7
67S		0.110	381		-1.7																		
68G		0.120	426																				
69T		0.115	401				-2.0																
70D		0.103	399		-3.8	-5.5	-7.1	-1.8															
71F		0.083	663																				
72T		0.071	492	-2.0	0.0	-3.6	-2.2	-0.3				-3.8	-0.5	-3.0	-4.8	-8.7	-0.6	-6.5	-0.3				-0.4
73L		0.062	771																				5.5
74T		0.067	442	-4.3	-2.4		-1.7	-0.7				-1.3		-1.3	-2.7								-3.2
75I		0.074	695																				
76S		0.087	392	-0.3	-1.1	-0.6					-0.6	-5.3	-0.8	-2.8	-3.7	-2.6	-0.1						-0.4
77S		0.097	438	-0.1																			-3.4
78L		0.099	678																				
79Q		0.113	432	-1.9		-3.9																	
80P		0.121	363																				-3.8
81E		0.113	371		-5.2	-1.0	-2.2	-4.2		-1.8													-2.9
82D		0.097	650																				-2.1
83F		0.092	532																				-0.3
84A		0.098	630																				
85T		0.077	497	-3.8	-1.7	-0.5																	-1.3
86Y		0.061	886																				
87F		0.055	740																				
88C		0.055	968	-2.6		-4.2	-1.9	-0.5															-1.2
89L		0.059	1016	-0.1																			-4.2
90Q		0.075	841																				
91H		0.085	643																				
92W		0.106	412																				
93S		0.113	432		-2.1																		





Residue	CDR	RMSF	Depth (Å)	Aliphatic				Non-polar			Aromatic				Polar				Chargeable																
				A	V	L	I	G	P	C	M	H	F	Y	W	N	Q	S	T	K	R	D	E												
301R		0.138	3.93	-2.0	-2.3	-6.8	-0.1					-5.2	-4.5	-2.6	-0.2	-4.4	-3.4	-2.7	-0.2			-4.5													
302A		0.147	3.70		-5.3	-5.3				-2.8	-4.2			-1.2	-0.3		-3.4			-0.3		0.0													
303E		0.170	3.52					-1.8	-1.4		-1.9		-2.5			-4.6			-2.8		-3.6														
304D		0.138	6.26	-4.2	-2.8	-3.7	-2.0				-4.0	-3.8	-2.0	-2.3		-1.0	-2.4																		
305T		0.121	4.39							-4.1				-2.7	-3.4																				
306A		0.095	6.08							-0.4																									
307V		0.086	5.02				-3.9						-4.0	-1.8																					
308Y		0.070	8.87										-3.1		-0.7																				
309Y		0.070	7.45										-3.8																						
310C		0.073	10.2	-6.4	-6.4	-2.4	-5.6	-5.6	-1.4		-5.4	-4.5	-1.8	-1.1	-2.4	-6.0	-1.1	-2.4	-3.2	-6.3	-5.4	-0.6	-1.4												
311A		0.078	11.38					-1.0			-2.0		-3.2	-0.6																					
312P		0.087	9.88		-0.5																														
313T		0.099	6.38	-2.6	-0.3	-1.3	-1.9	0.0								-1.9		-4.8																	
314T		0.115	4.80									-2.0																							
315V		0.135	4.24					-0.9			-0.4		-1.9	-2.0	-0.9		-2.4					-0.7													
316V		0.138	4.58			-2.8					-1.1		-6.9	-0.2	-3.1	-4.3	-0.2	-1.5		-5.1	-1.2	-4.7	-2.2												
317P		0.133	4.42																																
318F		0.133	5.38		-0.9	-2.3						-0.1		-1.1				-1.5					-3.9												
319A		0.123	9.01		-5.3	-1.9	-2.0				-4.6	-4.5	-5.6	-6.2			-2.7	-2.8					-2.7												
320Y		0.114	6.41									-1.0	-0.5																						
321W		0.093	7.15																																
322G		0.095	5.54																																
323Q		0.105	3.87							-4.1			-2.1	-2.4	-4.6																				
324G		0.099	4.36																																
325T		0.107	5.04																																
326L		0.108	3.72								-1.4																								
327V		0.101	6.24							-2.4		-2.5	-1.0	-3.4				-1.6																	
328T		0.105	5.47	-0.1	-2.9	-3.4	-1.4								-1.7	-0.7	-3.4			-2.2															
329V		0.106	6.23	-1.8																															
330S		0.123	4.71		-0.3		-0.1								-1.2		-1.3						0.0												
331S		0.141	3.69							-0.3	-4.5	-4.7			-4.1		-5.1						-3.4												
332A		0.135	4.01		-3.7		-0.4					-4.2		-4.4				-4.7	-0.7	-3.1															
333S		0.134	3.74	-1.9		-4.0					-6.2	-0.1		-3.5	-1.1								-2.0												
334T		0.124	3.57	-2.5	-1.4		-5.5				-3.6	-1.1		-1.0	-1.6				-0.5				-0.6												
335K		0.116	3.99	-2.7	-6.0	-0.4	-3.2					-1.8		-4.2	-4.4			-0.7					-0.2												
336G		0.111	3.78							-3.0	-3.7	-0.7	-5.0						-0.4				-3.0	-1.4											
337P		0.098	4.87																																
338S		0.090	4.07			-1.7					-0.7	-3.4	-5.6	-1.5	-6.7		-0.4	-6.0		-2.6			-1.0	-3.3											
339V		0.082	4.98																																
340F		0.070	5.93										-1.7																						
341P		0.071	4.97																																
342L		0.065	6.33																																
343A		0.089	5.24			-2.1		-0.2				-2.7	-3.5	-0.1			-0.9	-1.3					-1.7	-0.5											
344P		0.134	5.97												-2.9																				
345S		0.166	4.40		-4.7						-6.9		-6.2	-2.6	-2.6	-1.3	-2.3						-3.5	-2.4											
346S		0.194	3.91		-2.3							-1.7												-2.2											
347K		0.223	3.66	-2.4			-2.0					-0.7		-2.4	-3.5	-0.5	-3.1		-3.0		-0.1														
348S		0.275	3.75	-2.4	-1.8		-3.4				-4.0	-4.0	-5.3	-1.0		-6.0		-0.7	-3.9					-0.8	-1.5										
349T		0.249	3.87	-0.1	-0.7		-6.1							-2.5	-2.6	-8.0			-2.1					-0.9	-0.3										
350S		0.255	3.37	-0.5	-0.5					-4.3	-0.7	-1.8			-3.1	-2.4		-0.1	-1.1																
351G		0.228	3.35									-1.3						-1.9							-0.1										
352G		0.178	4.15	-2.2								-0.1	0.0																						
353T		0.134	3.78	-2.6		-3.2		-2.6	-1.5	-1.5			-2.8	-1.2	-0.7	-2.6			-3.8						-0.9										
354A		0.100	5.19			-0.3				-2.5	-0.3	-0.7	-2.8	-1.4	-2.2	-0.5			-1.9	-0.4	-1.8				-1.0	-1.7									
355A		0.083	6.08															-0.6			0.0														
356L		0.079	8.69																																
357G		0.071	9.80																																
358C		0.071	10.08	-5.4			-1.8					-3.0		-0.5	-0.5			-0.5		-1.1	-3.2				-0.1	-5.8									
359L		0.065	9.74									-2.2		-4.1	-5.5																				
360V		0.077	9.12																																
361K		0.082	5.94	-2.7								-1.0			-0.8	-2.5			-0.5							-0.5									
362D		0.090	4.70									-1.6			-1.7			-2.4	-2.6							-0.3									
363Y		0.098	7.48																																
364F		0.111	6.20																																
365P		0.122	4.98									-1.4		-1.3																					
366E		0.122	4.22	-1.4							-4.4		-0.4	-1.5	-4.5											-1.0									
367P		0.128	4.02			-1.0	-2.7	-1.4					-1.2	-0.4	-4.9				-4.0	-3.3	-6.0					-3.4									
368V		0.116	5.17	-1.9																															
369T		0.127	4.02		-6.0	-1.8	-1.2							-2.2	-1.2		-1.8										-5.3	-0.2							
370V		0.114	4.90																									-0.1	-3.1						
371S		0.123	4.36			-0.2	-0.2					-0.5	-2.2		-3.8	-3.6		-0.2										-0.6							
372W		0.114	7.08																																
373N		0.132	3.93	-0.7		-3.1							-2.5																						
374S		0.152	3.52									-2.4		-1.3	-0.3	-6.9	-5.1			-4.1									-2.8	-1.6	-3.2				
375G		0.151	3.62																																
376A		0.155	3.44			-2.1																													
377L		0.135	4.42		0.0		-1.9																												
378T		0.137	3.70		-0.9	-4.4	-3.4	-1.6						-0.1		-0.7																			
379S		0.141	3.51		-0.5		-0.1								-3.8		-2.8		-0.7	-1.9															
380G		0.115	4.04			-1.4								-4.3		-1.3																			
381V		0.101	4.17																																
382H		0.095	4.83																																
383T		0.095	3.73	-0.2	-1.2	-1.4	-4.7																												
384F		0.087	6.25																																



	PWFADWGGQTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVT VSWNSGALT--SGVHTFPAVLQSSGLYSLSSVTVPSSSLGTQTYICNVNHKPSNT KVDKKVEPKSC-----
<b>1C5D</b>	EVKLLESGPGLVQPSQTLSTCTVSGFPLTTNG-VSWVRQPPGKGLEWIAAISS--- GGSPYYNSALKSRLSINRDTSKSQVFLKMNSLQTEDTAIFYCTREDGWN----- -FDYWGPMTMTVSSAQTAPSVYPLAPGCGDTSSTVTLGCLVKGYFPEPVT TWNSGALS--SDVHTFPAVLQS-GLYTLTSSVT--SSTWPSQTVTCNVAHPASSTKV DKKLER-----
<b>1DFB</b>	EVQLVESGGGLVQPGRSLRLSCAASGFTFNDYA-MHWVRQAPGKGLEWVSGIS W--DSSSIGYADSVKGRFTISRDNANKSLYLQMNSLRAEDMALYYCVKGRDYYDS G--GYFTVAFDIWGGTMVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDY FPEPVTVSWNSGALT--SGVHTFPAVLQSSGLYSLSSVTVPSSSLGTQTYICNVN HKPSNTKVDKKVEPKSC-----
<b>1DN0</b>	EVQLQQWGAGLLKPSETLSLTCVYGGSFSDYY-WSWIRQPPGKGLEWIGEINH- --SGSTNYNPSLKSRTISVDTSKNQFSLKLSSVTAADTAVYYCARPPHDTSG----- -HYWNYWGGQTLVTVSSGSASAPTLFPLVSC-----SSVAVGCLAQDFLPDSITFS WKYKNSDISSTRGFPSVLRG-GKYAATSQVLLPSKDVTEHVVCKVQHPNGNK EKNVPLPV-----
<b>1DQD</b>	EVQLQESGPSLVKPSQTLSTCSVTGDSITSGY-WNWIRKFPGNKLEYMGYISY--- SGSTYYNPSLKSRLSITRDTSRNQYYLQKSVTPEDTATYYCASPPGYGSGP---- -YAMDYWGQGTSTVTVSSAKTTPPSVYPLAPGS-AQTNSMVTGCLVKGYFPEPV TVTWNSSGSL--SGVHTFPAVLQS-DLYTLSSSVTVPSSTWPSETVTCNVAHPASS TKVDKKISPG-----
<b>1FGN</b>	EIQLQQSGAELVRPGALVKLSCKASGFNIKDY-MHWVKQRPEQGLEWIGLIDP-- ENGNTIYDPKFQKASITADTSSNTAYLQLSSLTSEDVAVYYCARDNS-----YY FDYWGGQTTLVSSAKTTPPSVYPLAPGSAAQTNSMVTGCLVKGYFPEPVT WNSGSL--SGVHTFPAVLQS-DLYTLSSSVTVPSSTWPSETVTCNVAHPASSTKV DKKI-----
<b>1IT9</b>	QVQLVQSGAEVKKPGASVKVSCASGYFTSYW-MQWVKQAPGQGLEWMGEI DP--SDSYTNYNQKFKGKATLTVDTSTSTAYMELSSLRSEDVAVYYCARNR----- YSNNWYFDVWGEGLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFP EPVTVSWNSGALT--SGVHTFPAVLQSSGLYSLSSVTVPSSSLGTQTYICNVNHK PSNTKVDKKV-----
<b>1L7I</b>	EVQLVESGGGLVQPGGSLRLSCAASGFTFTDY-MDWVRQAPGKGLEWVADV ---NSGGSIYNQRFKGRFTLSVDRSKNTLYLQM---RAEDTAVYYCARNLGPFS----- ----DYWGGQTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVT VSWNSGALT--SGVHTFPAVLQSSGLYSLSSVTVPSSSLGTQTYICNVNHKPSNTK VDKKVEPKSC-----
<b>1OPG</b>	EVQLVQSGGGLVNPGRSLKLSCAASGFTFSSYG-MSWVRQTPEKRLEWVA AISG --GGTYIHYPDSVKGRFTISRDNANKNNLYLQMSSLRSEDVAVYYCTRHPFYRYDG G---NYYAMDHWGQGTSTVTVSSAKTTPPSVYPLAPGSAAQTNSMVTGCLVKGY FPEPVTVWNSGSL--SGVHTFPAVLQS-DLYTLSSSVTVPSSPRPSETVTCNVA HPASSTKVDKKIVPRDC-----
<b>1T3F</b>	-VQLVQSGAELKKPGSSVKVSCASGYIFTSSW-INWVKQAPGQGLEWIGRIDP-- SDGEVHYNQDFKDKATLTVDKSTNTAYMELSSLRSEDVAVYYCARG-----FLP-- WFADWGGQTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVT VSWNSGALT--SGVHTFPAVLQSSGLYSLSSVTVPSSSLGTQTYICNVNHKPSNTK VDKKVEPKSC-----

<b>2Z4Q</b>	QVQLQQSGSEMARPASVKLPCKASGDTFTSYW-MHWVKQRHGHGPEWIGNIY P--GSGGTNYAEKFKNKVTLTVDRSSRTVYMHLSRLTSEDSAVYYCTRSGGP----- ----YFFDYWGQGTSLTVSSAKTTAPSVYPLAPVCGDTTGSSVTLGCLVKGYFPEP VTLTWNSSGSL--SGVHTFPAVLQS-DLYTLSSSVTVTSSTWPSQSITCNVAHPAS STKVDKKEPR-----
<b>2ZKH</b>	EVKLEESGGGLVQPGGSMKLSAASGFTFSDAW-MDWVRQSPEKGLEWVAEIR SKVNNHAIHYAESVKGRFTVSRDDSKSSVYLMNSLRAEDTGIYYCSGWSFL---- -----YWGQGTSLTVSAAKTPPSVYPLAP-----SMVTLGCLVKGYFPEPVTVTV NSGSL--SGVHTFPAVLQS-DLYTLSSSVTVPSSTWPSETVTCNVAHPASSTKVD KKIVPR-----
<b>3D69</b>	GVQLVESGGGVVQPGRSLRLSCLASGFTFSTYA-MHWVRQAPGKGLEWVAIIS-- -DGSKKYYADSVKGRFTISRDNKNTLYLQM---RAEDTAVYYCARASIAA----- -DYWGRGTMVTVSSASTKGPSVFPLAPS-----GTAALGCLVKDYFPEPVTVSWNS GALT--SGVHTFPAVLQSSGLYSLSSVTVPSSSLGTQTYICNVNHKPSNTKVDKK VEPKSCD-----
<b>3G6A</b>	QVQLVESGGGLVQPGGSLRLSCLASGFTFNYSYW-INWVRQAPGKGLEWVSGIA Y--DSSNTLYADSVKGRFTISRDNKNTLYLQMNSLRAEDTAVYYCARGLGAFHW D--MQP----DYWGQGTSLTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYF PEPVTVSWNSGALT--SGVHTFPAVLQSSGLYSLSSVTVPSSSLGTQTYICNVNH KPSNTKVDKKVEP-----
<b>3HC0</b>	QVQLVQSGAEVKKPGSSVKVSCASGYTFTTY-LHWVRQAPGQGLEWVWGIY P--GNVHAQYNEKFKGRVTITADKSTSTAYMELSSLRSEDVAVYYCARS----- WEGFPYWGQGTSLTVSSASTKGPSVFPLAPSS----GGTAALGCLVKDYFPEPVT VSWNSGALT--SGVHTFPAVLQSSGLYSLSSVTVPSSSLGTQTYICNVNHKPSNT KVDKKVEPK-----
<b>3HI5</b>	-VQLLESGGGLVQPGGSLRLSCLASGFTFSRYV-MWVVRQAPGKGLEWVSYIW P--SGGNTYYADSVKGRFTISRDNKNTLYLQMNSLRAEDTAVYYCASSYDFWSN A--F-----DIWGQGTSLTVSSASTKGPSVFPLAPCSRSTSESTAALGCLVKDYFPE PVTVSWNSGALT--SGVHTFPAVLQSSGLYSLSSVTVPSSSLGTQTYICNVNDHK PSNTKVDKRVES-----
<b>3VG0</b>	EVKLLESGLVAPSELSITCTISGFSLTDDG-VSWIRQPPGKGLEWLVGIWG--- GGSTYFNSLFSKRLSITRDNKSKQVLEMDSLQTTDAMYYCAKHDGHET----- -MDYWGQGTSLTVSSSKTTPPSVYPLAPGSAQTNSMVTGCLVKGYFPEPVTV TWNSSGSL--SGVHTFPAVLQS-DLYTLSSSVTVPSSTWPSETVTCNVAHPASSTK VDKKIVPRDC-----
<b>4GSD</b>	-VQLQESGGLVQKPSGTVSLTCAVSGGSISSSYWWSVVRQPPGKGLEWIGIYH ---SGNTNYPNLSKSRVTISVDKSKNLFSLKLSSVTAADTAVYYCARVALFDILTGG W----FDPWGQGTSLTVSSAGTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEP VTVSWNSGALT--SGVHTFPAVLQSSGLYSLSSVTVPSSSLGTQTYICNVNHKPS NTKVDKRVES-----
<b>4HBC</b>	-QSVEESGRLVTPGTPLTACTVSGFSLNTYS-MFWVRQAPGKGLQWIGIISN--- FGVIYYATWAKGRFTIS--KTSTTVDLKITSPTTEDTATYFCVRKYGSEWG-----G DLWGPGLTVTVSSGQPKAPSVFPLAPCCGDT--PTVTLGCLVKGYLPEPVTVTVN SGTLT--NGVRTFPSVRQSSGLYSLSSVSVTSP-----VTCNVAHPATNTKVDKTV APSTC-----
<b>4HH9</b>	QVQLVESGGGVVQPGRSLRLSCLASGFTFSNHG-MHWVRQAPGKRLEWVAVIS ---DGRHEHYADLVKGRFTISRDNKNTLYLQM---RAEDRALYFCAREGLSRD----- ----DYWGQGTSLTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVT

	VSWNSGALT--SGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNT KVDDKVEPKSCD-----
<b>4HIE</b>	LINLVESGGGVVQPGRSLRLSCAASGFTFSRYG-MHWVRQAPGKGLEWVAVVS- --DGRTTYADSVKGRFTISRDNKNLTYLQM---RAEDTAVFYCAKE---GGDN--K-- ----DYWGQGTTLVTVSSASTKGPSVFPLA-----SESTAALGCLVKDYFPEPVTVSWN SGALT--SGVHTFPAVLQSSGLYSLSSVVTVP-----QTYTCNVDHKPSNTKVDDKTV- -----
<b>4LKX</b>	QVQLQESGPGLVKPSSETLSLTCTVSGYSITSDYAW-WIRQPPGKGLEWIGSISY--- SGITGYNPSLKSRTISRDTSKNQFSLKL---TAADTAVYYCARMG-YDGLAY----- -WGQGTTLVTVSSASTKGPSVFPLAPS-----GGTAALGCLVKDYFPEPVTVSWNSG ALT--SGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDDKA EPKS-----
<b>4OCY</b>	DVQLQESGPGLVKPSQSLSLTCTVTGFSITSPYAWNIRQFPNTLEWMGYISY- --RGSTTYHPSLKSRISTRDTSKNQFFLQLNSVTTEDTATYFCSS-YGNYG----- -AYSGQGTTLVTVSAAKTPPSVYPLAPG-----SMVTLGCLVKGYFPEPVTVTWNS GSL--SGVHTFPAVLQS-DLYTLSSSVTVPSSTWPSETVTCNVAHPASSTKVDDKI VPR-----
<b>4OSU</b>	QVQLVQSGAEVRKPGASVKVSCASGYSLKDHY-MVWVRQAPGQGLEWMGWI NP--QSGGTGYGQKFQGRVTMTRDTSTNTAYMILSSLRSDDTAVYFCARDGAKT VNSLLYHNRLDAWGQGTMTVTVSSASTKGPSVFPLAPS-----GTAALGCLVKDY FPEPVTVSWNSGALT--SGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVN HKPSNTKVDDKVEPK-----
<b>7FAB</b>	AVQLEQSGPGLVRPSQTLTCTVSGTSFDDYY-WTWVRQPPGRGLEWIGYVYF ---TGTTLLDPSLRGRVTMLVNTSKNQFSLRLSSVTAADTAVYYCARNLIAGG----- --IDVWGQGSLLVTVSSASTKGPSVFPLAP-----TAALGCLVKDYFPEPVTVSWNS GALT--SGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDDK VEP-----

Table 10.2 The sequence alignment for light chain

<b>PDB ID</b>	<b>Sequence of light chain</b>
<b>C226 S</b>	DIQMTQSPSSLSASVGDRVTITCKAS-QNVR-----TVVAWYQQKPGKAPKTLIYLAS NRHTGVPSRFSGSGSGTDFTLTISSLQPEDFATYFCLQHWSYP--LTFGGQGTKVEI KRTV-AAPSVFIFPPSDEQLKSGTASVVCLLNNFYPRKAKVQWKVDNALQSGNSQ ESVTEQDSKSTYLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC-
<b>1A6T</b>	QSVLSQSPAILSASPGEKIVIMTCSPS-SSVS-----YMQWYQQKPGSSPKPWIYSTS NLASGVPRFSGGGSGTSFSLTISGVEAEDAATYYCQQYSSHP--LTFGGGKLEL KRAD-AAPTVSIFPPSSEQLTSGGASVVCFLNNFYPKDINVKWKIDGSRQNGVLN SWTDQDSKSTYSMSSTLTLTKDEYERHNSYTCEATHKTSTSPIVKSFNR----
<b>1B2 W</b>	DIQMTQSPSTLSASVGDRVTITCKAS-ENVD-----TYVSWYQQKPGKAPKLLIYGAS NRYTGVPSRFSGSGSGTDFTLTISSLQPDFFATYYCGQSYNYP--FTFGQGTKVEV KRTV-AAPSVFIFPPSDEQLKSGTASVVCLLNNFYPRKAKVQWKVDNALQSGNSQ ESVTEQDSKSTYLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC-
<b>1C5D</b>	DIQMTQSPSSLSASLGDKVTITCQAS-QDINK-----YIAWYQQKPGKAPRQLIRYTSIL VLGTPSRFSGSGSGRDFSFISINVAIEDIASYYCLQYGN-L--YTFGAGTKLEIKRA D-AAPTVSIFPPSTEQLATGGASVVCMLNNFYPRDISVKWKIDGTERRDGVLDVSVT DQDSKSTYSMSSTLTLTKADYESHNLTYCEVHKTSSSPVVKSFNRNEC-

<b>1DF B</b>	DIQMTQSPSTLSASVGDRVTITCRAS-QSIS-----RWLAWYQQKPGKVPKLLIYKASS LESGVPSRFSGSGSGTEFTLTISLQPDDEFATYYCQQ-YNSY--S-FGPGTKVDIKR TV-AAPSVFIFPPSDEQLKSGTASVVCLLNNFYPRKAVQWKVDNALQSGNSQES VTEQDSKDSTYLSSTLTLSKADYKHKVYACEVTHQGLSSPVTKSFNRGEC-
<b>1DN0</b>	EIVLTQSPATLSLSPGERATLSCGAS-QSVSS-----NYLAWYQQKPGQAPRLLIYDAS SRATGIPDRFSGSGSGTDFTLTISRLEPEDFAVYYCQQYGSPP--LTFGGGKTKVEIK RTV-AAPSVFIFPPSDEQLKSGTASVVCLLNNFYPRKAVQWKVDNALQSGNSQES SVTEQDSKDSTYLSSTLTLSKADYKHKVYACEVTHQGLSSPVTKSFNRGEC-
<b>1DQ D</b>	DIVLSQSPAIMSASPGEKVTITCSAS-SSVS-----YMHWFQQKPGTSPKLCIYTTSNL ASGVPARFSGSGSGTSYSLTISRMEAEADAATYYCQQRSTYP--PTFGSGTKLEIKR AD-AAPTVISIFPPSSEQLTSGGASVVCFLNNFYPRDINVKWKIDGSERQNGVLNS WTDQDSKDSTYSMSSTLTTLTKDEYERHNSYTCEATHKTSTSPIVKSFNRECA
<b>1FG N</b>	DIKMTQSPSSMYASLGERVTITCKAS-QDIRK-----YLNWYQQKPKWSPKTLIYYATS LADGVPSRFSGSGSGQDYSLTISLESDDTATYYCLQHGESPP--YTFGGGKTKLEIN RAD-AAPTVISIFPPSSEQLTSGGASVVCFLNNFYPRDINVKWKIDGSERQNGVLNS WTDQDSKDSTYSMSSTLTTLTKDEYERHNSYTCEATHKTSTSPIVKSFNRECA-
<b>1IT9</b>	EIVLTQSPGTLSPGERATLSCAS-QSVDYDGD-SYMNWYQQKPGQAPRLLIYA ASNLESGIPDRFSGSGSGTDFTLTISRLEPEDFAVYYCQQSNEDP--RTFGQGTKL EIKRTV-AAPSVFIFPPSDEQLKSGTASVVCLLNNFYPRKAVQWKVDNALQSGNS QESVTEQDSKDSTYLSSTLTLSKADYKHKVYACEVTHQGLSSPVTKSFN-----
<b>1L7I</b>	DIQMTQSPSSLSASVGDRVTITCKAS-QDVS-----IGVAWYQQKPGKAPKLLIYSASY RYTGVPSPRFSGSGSGTDFTLTISLQPEDFATYYCQQYYIYP--YTFGGGKTKVEIKR TV-AAPSVFIFPPSDEQLKSGTASVVCLLNNFYPRKAVQWKVDNALQSGNSQES VTEQDSKDSTYLSSTLTLSKADYKHKVYACEVTHQGLSSPVTKSFNRGEC-
<b>1OP G</b>	DELLTQSPATLSVTPGDSVLSLSCRAS-QSISNN-----LHWYQQKSHESPRLLIKYAS QSISGIPSRFSGSGSGTDFTLSINSVETEDFGMYFCQQSNSWP--LTFGGGSKLEIK RAD-AAPTVISIFPPSSEQLTSGGASVVCFLNNFYPRDINVKWKIDGSERQNGVLNS WTDQDSKDSTYSMSSTLTTLTKDEYERHNSYTCEATHKTSTSPIVKSFNRECA-
<b>1T3F</b>	DIQMTQSPSTLSASVGDRVTITCKAS-ENVN-----TYVSWYQQKPGKAPKLLIYGAS NRYTGVPSPRFSGSGSGTDFTLTISLQPDDEFATYYCGQSYNYP--FTFGQGTKVEV KRTV-AAPSVFIFPPSDEQLKSGTASVVCLLNNFYPRKAVQWKVDNALQSGNSQ ESVTEQDSKDSTYLSSTLTLSKADYKHKVYACEVTHQGLSSPVTKSFNRGEC-
<b>2Z4Q</b>	DILMTQTPLSLPVLGDQASISCRSS-QNIVHNNGITYLEWYLQRPQGSPKLLIYKV SDRFSGVPDRFSGSGSGTDFTLKISRVEAEDLGIYYCFQGSHIP--PTFGGGTKLEI KRAD-AAPTVISIFPPSSEQLTSGGASVVCFLNNFYPRDINVKWKIDGSERQNGVLN SWTDQDSKDSTYSMSSTLTTLTKDEYERHNSYTCEATHKTSTSPIVKSFNRECA-
<b>2ZK H</b>	QVVLQSPGIMSASPGEKVTITCSAS-SSVS-----YMYWFQQKPGTSPKLIYISTS NLASGVPARFRGSGSGTSYSLTISRMEAEADAATYYCQQRSGYP--RTFGGGKTKLEI KRAD-AAPTVISIFPPSSEQLTSGGASVVCFLNNFYPRDINVKWKIDGSERQNGVLN SWTDQDSKDSTYSMSSTLTTLTKDEYERHNSYTCEATHKTSTSPIVKSFN-----
<b>3D69</b>	--VLTQ-PPSVSAAPGQKVTISCSG--STIGNNY-----VSWYQQHPGKAPKLMYDVS RPSGVPDRFSGSKSGNSASLDISGLQSEDEADYYCAAWDDSLF--LFGTGKTLTVL GQPKAAPSVTLFPPSSEELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVET TTPSKQS-NNKYAASSYLSLTPEQWKSHKSYSCQVTHEG--STVEKTVAPT---
<b>3G6 A</b>	SYELTQ-PPSVSVAPGQTARISCSG--DNIGGTF-----VSWYQQKPGQAPVLIYDDN DRPSGIPERFSGSNGNTATLTISGTQAEDEADYYCGTWDMMVTN-NVFGGGKTKLT VLGQPKAAPSVTLFPPSSEELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGV ETTPSKQS-NNKYAASSYLSLTPEQWKSHRSYSCQVTHEG--STVEKTVAPT---

<b>3HC0</b>	DIQMTQSPSSLSASVGDRTITCKAS-QNVG-----INVAWYQQKPGKAPKSLISSAS YRYSVPSRFSGSGSGTDFTLTISLQPEDFATYFCQQYDTYP--FTFGQGTKVEI KRTV-AAPSVFIFPPSDEQLKSGTASVVCLLNNFYPPREAKVQWKVDNALQSGNSQ ESVTEQDSKDSTYLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRG---
<b>3HI5</b>	DIQMTQSPSSLSASVGDRTITCRAS-QSIG-----SYLNWYQQKTGKAPKALIYAASS LQSGVPSRFSGSGSGTDFTLTISLQLEDFATYYCQQSYSTP--S-FGQGTKVEIKR TV-AAPSVFIFPPSDEQLKSGTASVVCLLNNFYPPREAKVQWKVDNALQSGNSQES VTEQDSKDSTYLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRG---
<b>3VG0</b>	EIVMTQSPKFMSTSIGDRVNITCKAT-QNVRT-----AVTWYQQKPGQSPQALIFLAS NRHTGVPARFTGSGSGTDFTLTINNPKSEDLADYFCLQHWYYP--LTFGSGTKLEI KRAD-AAPTVISIFPPSSEQLTSGGASVVCFLNNFYPPKDINVKWKIDGSRQNGVLN SWTDQDSKDSTYSMSSTLTLTDEYERHNSYTCEATHKTSTSPIVKSFNRE--
<b>4GS D</b>	---LTQ-PPSVSVSPGQTVNITCSG--DTLGDKY----VCWYQQKPGQSPVLVIYQDTK RPSGIPERFSGSNSGDTATLTVSGTQAMDEADYYCQAWDSSS--FVFGTGTKVTV LRQPKANPTVTLFPPSSEELQANKATLVCLISDFYPGAVTVAWKADGSPVKAGVE TTKPSKQS-NNKYAASSYLSLTPEQWKSHRSYSCQVTHEG--STVEKTVAPTE--
<b>4HB C</b>	DVVMQTTPASVSEPVGGTVTIKCQAS-QSISS-----YLAWYQQKPGQRPRLLIYETS TLASGVPSRFRKSGSGTDFTLTISDLECAATYYCQSTYENPTYVVSFGGGTEVGV VKGDP-VAPTVLIFPPSADLVATGVTIVCVANKYFP-DVTVTWEVDGTTQTTGIEN SKTPQNSADCTYNLSSTLTLTSTEYNHKEYTCKVTQG--TTSVVQSFNRGDC-
<b>4HH9</b>	EVVLTQSPATLSLSPGERATISCRAS-QSVG-----GYLTWYQQKPGQAPRLLIYDAS NRATGIPARFSGSGSGTDFTLTISGLEPEDFAIYYCQQRGNWP--ITFGQTRLEIK RTV-AAPSVFIFPPSDEQLK-GTASVVCLLNNFYPPREAKVQWKVDNALQSGNSQE SVTEQDSKDSTYLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC-
<b>4HIE</b>	-GQLTQSPATLSLSPGERATLSCRAS-QSVT-----NYLAWYQQKPGQAPRLLIYGAS NRATGIPARFSGSGSGTDFTLTISLLEPEDFAVYYCQQRDNWP--ATFGQGTKVEI KRTV-AAPSVFIFPPSDEQLKSGTASVVCLLNNFYPPREAKVQWKVDNALQSGNSQ ESVTEQDSKDSTYLSSTLTLSKADYEKHKVYACEVTHQGLRSPVTKSFNR----
<b>4LKX</b>	DIVMTQTPLSLSVTPGQPASISCRSS-QNGN-----TYLEWYLQKPGQSPQLLIYKVS NRFSGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCFQGSHVP--PTFGGGTKVEI KRTV-AAPSVFIFPPSDEQLKSGTASVVCLLNNFYPPREAKVQWKVDNALQSGNSQ ESVTEQDSKDSTYLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC-
<b>4OC Y</b>	DVLLTQIPLSLPVSLGDQASISCRSS-QSIVHSNGNTYLEWYLQKPGQSPKLLIYKV STRFSGVPDRFSGSGSGTDFTLKISRVEAEDLGVYYCFQGSHVP--LTFGAGTQLE LKRAD-AAPTVISIFPPSSEQLTSGGASVVCFLNNFYPPKDINVKWKIDGSRQNGVL NSWTDQDSKDSTYSMSSTLTLTDEYERHNSYTCEATHKTSTSPIVKSFNRE----
<b>4OS U</b>	QSVLTQ-PPSVSAAPGQMTISCSGSSSNIGKNY----VSWYQQLPGAAPKLLIFDN NKRPSGTPDRFSGSKSGTSATLVITGLQTGDEADYYCGTPDRSLS-VIFGGGTKVT VLGQPKAAPSVTLFPPSSEELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGV ETTTPSKQS-NNKYAASSYLSLTPEQWKSHRSYSCQVTHEG--STVEKTVAPT---
<b>7FA B</b>	ASVLTQ-PPSVSGAPGQRVTISCTGSSSNIGAGHN---VKWYQQLPGTAPKLLIFHN N-----ARFSVSKSGTSATLAI TGLQAEDEADYYCQSYDRSLR--VFGGGTKLTVLR QPKAAPSVTLFPPSSEELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETT TPSKQS-NNKYAASSYLSLTPEQWKSHKSYSCQVTHEG--STVEKTVAP----

## 10.6 The correlation between $\Delta T_m$ , $\Delta\Delta G$

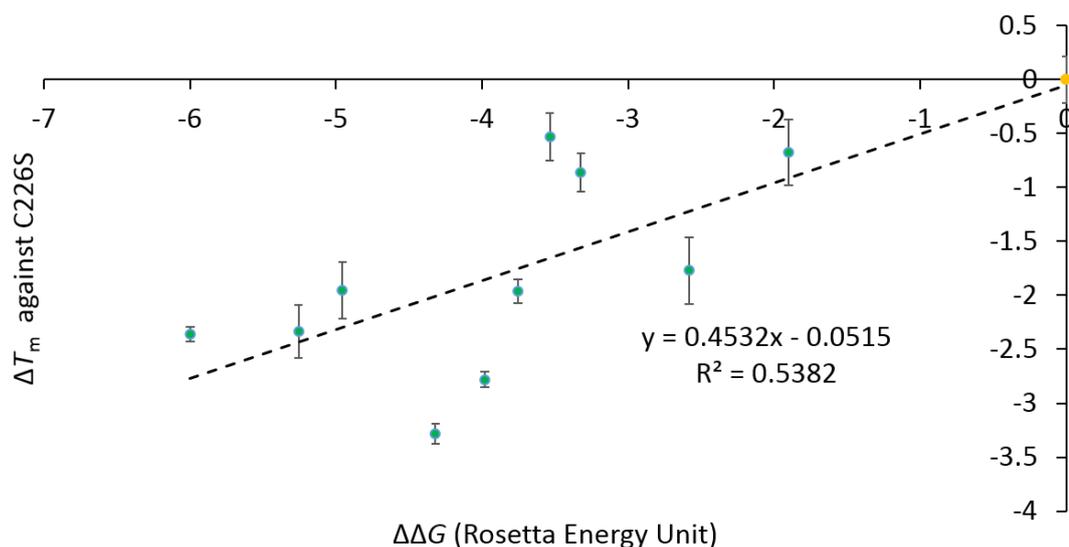


Figure 10.3 The correlation between  $\Delta T_m$  and  $\Delta\Delta G$  without destabilising mutants  
This is a supplementary information for Figure 5.9.  $\Delta T_m$  was calculated against C226S (Figure 5.6) and  $\Delta\Delta G$  is calculated by Rosetta (Figure 5.8). Error bar is SEM (standard error of the mean). Pseudo wild type C226S and stabilising mutants were coloured in yellow and green, respectively.

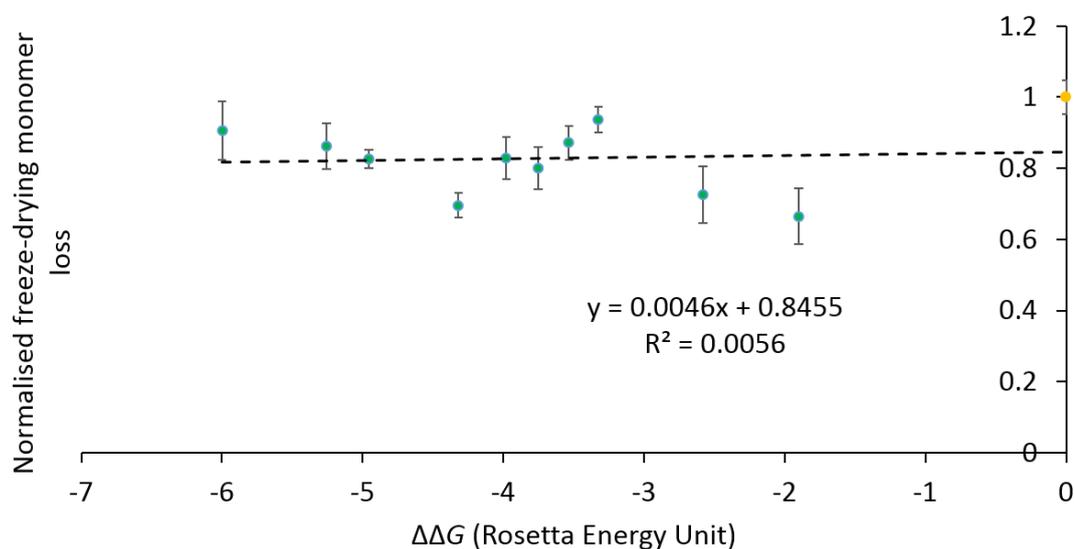


Figure 10.4 The correlation between normalised monomer loss in freeze-drying and  $\Delta\Delta G$  without destabilising mutants  
This is a supplementary information for Figure 5.10. Normalised monomer loss in freeze-drying was from Figure 5.3 and  $\Delta\Delta G$  calculated by Rosetta was from Figure 5.8. Error bar is SEM (standard error of the mean). Pseudo wild-type C226S and stabilising mutants were coloured in yellow and green, respectively.

## 10.7 GROMACS code

### 10.7.1 Code for “job.sh” file

```
#!/bin/bash -l
#$ -S /bin/bash
#$ -l h_rt=0:30:0
#$ -l mem=4G
#$ -l tmpfs=15G
#$ -N MD
#$ -pe openmpi 32
#$ -cwd

module unload compilers
module unload mpi
module unload mkl
module load compilers/intel/13.0/028_cxx11
module load mpi/openmpi/1.6.5/intel.13.0-028_cxx11
module load atlas/3.10.1/intel.13.0-028_cxx11
module load fftw/3.3.4/double/intel.13.0-028_cxx11
module load gromacs/5.0/openmpi/intel.13.0-028_cxx11

gerun convert-tpr -s md_0_1.tpr -o md_0_1.tpr
gerun mdrun_mpi -deffnm md_0_1 -cpi md_0_1.cpt -maxh 0.5 -append
```