Structural analysis of the cattle antibody repertoire

John Douglas Clarke

Linacre College
University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Michaelmas Term 2022

Abstract

Cattle present an extreme immunological model, wherein immunoglobulin sequence diversification is achieved not through recombinational diversity, as in humans and mice, but by great heterogeneity in the length of the complementarity determining region #3 of the heavy chain (CDR3H), and by somatic hypermutation independent of antigenic contact. Bovine CDR3H lengths span 5 - 72 residues, with those containing >48 labelled as 'ultralong' and assumed to interact with epitopes typically invisible to paratopes with shorter CDR loops. Cattle also demonstrate preferential combination of V(D)J rearrangements without antigenic stimulus, inferring a structural relationship between bovine immunoglobulins that is not yet understood. This project aims to advance understanding of these relationships from a structural and functional context.

The structures of several chain-exchanged ultralong antibodies were solved by X-ray crystallography, or predicted by AlphaFold, revealing an intimate contact between the CDR3L and the CDR3H. Varying the CDR3L sequence imparts an axial twist about the ultralong stalk domain, with minor accommodating atomic shifts at the distal, antigen-interacting knob domain, together with a subtle change in binding affinity by biolayer interferometry. Electron cryomicroscopy was used to assess whether chain-exchange resulted in any changes to the recognised epitope, but no significant differences were observed at the attained resolutions. Initial efforts with a non-ultralong antibody-antigen system are also described.

These observations imply bovine light chains principally contribute a structural role stabilising ultralong CDR3H, but their contribution in non-ultralong antibodies remains enigmatic. The limited impact of light chain selection on antibody function provides little insight to bovine B cell development. Bioengineers of the ultralong CDR3H may wish to consider the light chain sequence should their targeted epitope lie within deep groves of known geometry, however at present, light chain selection is only a minor consideration in the development of bovine ultralong CDR3H antibodies as diagnostic, therapeutic or research tools.

Structural analysis of the cattle antibody repertoire



John Douglas Clarke

Linacre College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas Term 2022

This thesis is dedicated
in loving memory of

Richard W. Causton,
whom championed academic pursuit.
1958 - 2018

and

Douglas C. Clarke,
whom inspired my fascination with innovation.
1938 - 2023

Acknowledgements

Foreword

This thesis serves as a reminder of intellectual curiosity to myself. I completed this thesis amidst a time of emerging conflict, a pandemic reverberating around the globe, the deterioration of our climate's stability and the subsequent onset of a sixth mass extinction event, a deepening resource crisis, the rise of antimicrobial resistance, a cost of living crisis, and a myriad of societal issues. In the face of these life-threatening worldwide problems, it is easy to assume an air of hopelessness. But there is hope. One must remember that for all of these concerns, be they scientific or humanitarian, there exists a significant community of people identifying, investigating and resolving these challenges, for their futures, and the rest of ours.

Not only must research be allowed to prevail, but the dissemination and realisation of research output must be encouraged and actively promoted. Humankind is truly in a race for survival, and now more than ever before, our success depends wholly on identifying novel medical treatments, efficient and sustainable solutions, and reconsidering age-old social structures. Our ancestors were arguably reliant on luck to survive, but our children will depend on science. It is on this gallant quest my love of science has emerged, and has given purpose and direction to my life. If I ever look back over this thesis in times of personal difficulty, I hope this serves as a reminder of why. Why a career in science is not only a passion, but a necessity.

Institutional

No research is conducted in isolation, and the same is true for this interdisciplinary project. I have hugely benefited from the support of collaborators around the world, whom have endured tedious laboratory work and administrative procedures to provide the samples, knowledge and expertise to complete this thesis. In particular, I acknowledge the support from James, William, Daniel and Kristel for their efforts in antibody sequence discovery; Kay and Kevan, for producing most of the Virus-like particles; Maud, Chelsea and Josh for their practical advice and assistance with experiments; Gemma and Angela, for their support with biophysics; and Jingshan, Jiandong, and Claudine for invaluable discourse and advice with structural biology.

Innumerable people have contributed to this project in some way, but in particular I'd like to thank all colleagues at The Rosalind Franklin Institute, The Pirbright Institute, Diamond Light Source, The Research Complex at Harwell, The Oxford Particle Imaging Centre, and the Wellcome Trust Centre for Human Genetics, for their patience, courtesy, and diligence especially in keeping me safe through the experimental work, and the COVID-19 pandemic. I'd also like to thank

the Doctoral Training Centre for their continuous support, and the opportunity to conduct this study.

This work was carried out with the support of Diamond Light Source (proposals mx19946, nt23570, nt29493, and bi28713). The computational aspects of this research were supported by the Wellcome Trust Core Award [Grant Number 203141/Z/16/Z] and the NIHR Oxford BRC. The views expressed are those of the author and not necessarily those of the NHS, the NIHR or the Department of Health. This work has been supported by funding from the United Kingdom Research and Innovation Biotechnology and Biosciences Research Council (UKRI-BBSRC) [Grant number: BB/M011224/1].

Personal

I have been truly lucky to complete this research project with the best mentors, collaborators and friends I could imagine. I am immensely grateful for my three supervisors, Ray Owens, Sir Dave Stuart, and John Hammond, for their unrivalled support, patience and insight. I joined the University of Oxford with little understanding of immunology, and even less experience in research. Through hours of discussions, Ray, Dave and John provided me the knowledge to contextualise my work, and led me into a world of curiosity. I am enormously fortunate to have been further supported by mentors Marie Di Placido, Liz Fry, Helen Duyvestyn, Halina Mikolajek, Alice Douangamath and Miriam Weckener whom guided me through basic principles and provided me with the basic experimental techniques. Their infectious enthusiasm has always inspired and motivated me, even when experimental results pose more questions than answers.

On a personal note, I really want to thank all of my friends whom have supported me throughout this project, notably Charlotte, Toby, Kat, Calum, Brad, Abi, Becky, and Audrey. I have always known I can count on them anytime for their knowledge, help and comfort, as well as being unfailingly dependable at raising pints or cocktails. I must also massively thank my family. Thank you for always welcoming a chat when I have been suffering, thank you for waiting and having faith in me when I've needed to focus, thank you for maintaining your interest in my somewhat alien work, thank you for not applying pressure or expectation, and most of all, thank you for showering me with your love through countless dinners, days out, daily titbits to make me smile, and seemingly limitless wine and gin.

Finally, I want to thank my now fiancée, Rebecca Bowles. You have been by my side through the highs and lows of this project and beyond. You have picked up the pieces of my struggles, and celebrated my successes. You always strive to put a smile on my face, and have always given all your heart in supporting me to complete this thesis. Thank you for bearing with me on the late nights, and the overnights, and the early mornings. Thank you for your understanding of the toll of negative results. Thank you for taking the time to understand and accurately paraphrase my research to others whom ask. Your unwavering belief in me alone has given me the confidence to finish.

Impact of COVID-19

During the COVID-19 pandemic and in support of progressing COVID-19 projects, laboratory access was restricted for non-COVID-19 research. Consequently, I had no laboratory access between 20th March 2020 and 3rd August 2020, and, to allow for social distancing in the laboratories, shift-based half-day access from the 3rd of August 2020 until 29th March 2021 when my group moved into the new Rosalind Franklin Institute building. The suspension of non-COVID-19 laboratory work also prevented the trial of a lentiviral expression system for the production of bRSVF at the STRUBI labs, and introduced a delay in my collaborators at The Pirbright Institute being able to generate novel antigen-specific antibody sequences for my project. Furthermore, I also missed several weeks of laboratory access due to self-isolating after i) being contacted by NHS Test and Trace on 23rd November 2020 and ii) testing positive for COVID-19 by NHS PCR on 18th March 2021 and again on 14th March 2022.

Except for a handful of health interruptions (symptomatic infections and for mental health purposes), I was able to mitigate most of the impact of losing laboratory access by completing desk-based research from home; performing data analysis, reading literature, and preparing research posters. However, significant laboratory time was still lost, sufficient to reduce the scope of my project within the original time-frame, especially with respect to the delayed sequence discovery from Pirbright. To this end, a three month funding extension was sought and approved from round five of Oxford's COVID-19 Scholarship Extension Fund, and subsequently, a thesis submission deadline extension was sought and approved to match this.

It was briefly discussed whether my project could be restructured to include some COVID-19 related research, but this was concluded not to be worth pursuing, due to the long time-frames associated with planning and running novel animal experiments, the resources being utilised for vaccine safety studies, and the success of SARS-CoV-2-binding nanobody projects.

Declaration of Work

I hereby declare, the work described herein is my own, except as described below.

Chapter 3: The cell sorting, cloning and sequencing experiments were conducted by Kristel Ramirez and James Nyagwange of The Pirbright Institute. Initial sequence analysis was also performed by Kristel, and clustering and in depth sequence analysis was conducted by Daniel Dorey-Robinson, also of Pirbright.

Chapter 4: Inactivated FMDV, used in the vaccination experiments, was produced by Erwin van den Born of MSD Animal Health, and purified by Claudine Porta of Pirbright. Eva Perez of Pirbright conducted the animal vaccination experiments. Kristel Ramirez and Angela Steyn of Pirbright performed the ELISpot assays to identify antibody-secreting cells. The fluorescent serospecificity assays were conducted by Cristina Faralla, Katherine Lam, Róisín McComb and Sherie Duncan of AbCellera Biologics Ltd. Daniel Dorey-Robinson completed the sequence analysis. Whole antibodies were expressed and purified by Elena Lokhman, Sylvia Crossley and Kevan Hanson of The Immunological Toolbox. The VLPs used during cryo-EM experiments, but not those used during Fab BLI experiments, were produced, quantified and integrity assessed by Kristel Ramirez, Benja Nzau, Kay Humphries and Kevan Hanson of Pirbright. Orthogonal IgG-antigen interaction assays were also conducted by Angela Steyn of Pirbright.

Chapter 5: Cell culture maintenance, expansion and aliquot freezing of stable cell lineages were kindly conducted by Maud Dumoux of The Rosalind Franklin Institute between the dates of 23$^{\text{rd}}$ March 2020 and the 14$^{\text{th}}$ May 2020, as part of the implemented institutional response to the COVID-19 pandemic. Graphene-oxide coated cryo-EM grids were kindly prepared by Miriam Weckener, of The Rosalind Franklin Institute, as part of the technical maturation of the Chameleon sample vitrification robot.

John Douglas Clarke
Linacre College,
University of Oxford

Abstract

Cattle present an extreme immunological model, wherein immunoglobulin sequence diversification is achieved not through recombinational diversity, as in humans and mice, but by great heterogeneity in the length of the complementarity determining region #3 of the heavy chain (CDR3H), and by somatic hypermutation independent of antigenic contact. Bovine CDR3H lengths span 5 - 72 residues, with those containing >48 labelled as 'ultralong' and assumed to interact with epitopes typically invisible to paratopes with shorter CDR loops. Cattle also demonstrate preferential combination of V(D)J rearrangements without antigenic stimulus, inferring a structural relationship between bovine immunoglobulins that is not yet understood. This project aims to advance understanding of these relationships from a structural and functional context.

The structures of several chain-exchanged ultralong antibodies were solved by X-ray crystallography, or predicted by AlphaFold, revealing an intimate contact between the CDR3L and the CDR3H. Varying the CDR3L sequence imparts an axial twist about the ultralong stalk domain, with minor accommodating atomic shifts at the distal, antigen-interacting knob domain, together with a subtle change in binding affinity by biolayer interferometry. Electron cryomicroscopy was used to assess whether chain-exchange resulted in any changes to the recognised epitope, but no significant differences were observed at the attained resolutions. Initial efforts with a non-ultralong antibody-antigen system are also described.

These observations imply bovine light chains principally contribute a structural role stabilising ultralong CDR3H, but their contribution in non-ultralong antibodies remains enigmatic. The limited impact of light chain selection on antibody function provides little insight to bovine B cell development. Bioengineers of the ultralong CDR3H may wish to consider the light chain sequence should their targeted epitope lie within deep groves of known geometry, however at present, light chain selection is only a minor consideration in the development of bovine ultralong CDR3H antibodies as diagnostic, therapeutic or research tools.

Contents

List of Abbreviations **xxi**

1 Introduction **1**
- 1.1 Canonical diversification of immunoglobulin repertoires 2
- 1.2 Genetics of the bovine heavy chain locus 7
  - 1.2.1 Organisation of the bovine heavy immunoglobulin locus . . . 7
  - 1.2.2 Limited VDJ recombinations 9
  - 1.2.3 Genetic sequence invariance 11
  - 1.2.4 Somatic diversification 11
  - 1.2.5 Exceptionally long CDR3H loops 12
- 1.3 Genetics of the bovine light chain locus 20
  - 1.3.1 Organisation of the bovine light immunoglobulin locus . . . 20
  - 1.3.2 Limited genetic diversity 22
  - 1.3.3 Somatic diversification of bovine light chains 22
- 1.4 Pairing peculiarities 22
- 1.5 A summary of bovine structural immunology 25
- 1.6 A curious immunological paradigm 26
- 1.7 Project aims and objectives 28

2 Methodology **29**
- 2.1 Core reagents 32
  - 2.1.1 Antibiotics 32
  - 2.1.2 Antibodies 32
  - 2.1.3 Cell lines 32
  - 2.1.4 Vectors 33
  - 2.1.5 Viruses 36
- 2.2 Laboratory Workflows 37
  - 2.2.1 Fab molecular biology workflow 37
  - 2.2.2 Plasmid propagation workflow 38
  - 2.2.3 Fab expression screening 38
  - 2.2.4 Fab expression 38
  - 2.2.5 bRSV F protein production 38

2.2.6 VLP production 39
2.3 Microbiology and cell culture techniques 39
2.3.1 Vector linearisation 39
2.3.2 In-Fusion ligation independent cloning 40
2.3.3 Preparation of LB media 41
2.3.4 Preparation of LB 2 % agar 41
2.3.5 Transformation of chemically competent *E. coli* cells 41
2.3.6 Glycerol stocks of transformed *E. coli* 41
2.3.7 Creation of a stable Expi293F™ cell lineage 42
2.3.8 Freezing down of stable cell lineages 43
2.4 Molecular biology techniques 43
2.4.1 Purification of plasmid DNA 43
2.4.2 Polymerase Chain Reaction screening 44
2.4.3 Agarose gel electrophoresis 45
2.4.4 Gel purification of DNA 45
2.4.5 Plasmid sequence verification 46
2.5 Protein expression 46
2.5.1 Transfection of Expi293F™ cells for Fab expression 46
2.5.2 Stable Expi293F™ cell line expression of bRSVF 47
2.5.3 Dual infection of HEK293T cells for VLP production 47
2.6 Protein purification and concentration 48
2.6.1 Immobilised affinity gel filtration of Fabs 48
2.6.2 Immobilised affinity gel Filtration of bRSVF 49
2.6.3 Chloroform extraction of VLPs 49
2.6.4 Sucrose density gradient (SDG) purification of VLPs 50
2.6.5 Concentrating proteins by centrifugation 50
2.6.6 Protein G re-purification 51
2.6.7 Protein desalting 52
2.7 Protein characterisation 52
2.7.1 Inference of protein molecular weight by SEC 52
2.7.2 SDS polyacrylamide gel electrophoresis (PAGE) 53
2.7.3 SDS-PAGE staining 54
2.7.4 Western blotting 54
2.7.5 Protein concentration quantification 55
2.7.6 Qubit™ quantification of VLPs 56
2.7.7 VLP Integrity assessment by VHH double sandwich ELISA . 56
2.7.8 Size exclusion chromatography with multi-angle laser light scattering 59
2.7.9 Prediction of protein glycosylation sites 60

2.7.10 Analytical Ultracentrifugation 61
2.7.11 Bio-layer Interferometry 62
2.8 Structural biology methods 64
2.8.1 ABodyBuilder structural prediction 64
2.8.2 AlphaFold structural prediction 64
2.8.3 Crystallographic X-Ray Diffraction 65
2.8.4 Single-particle cryogenic Electron Microscopy 74

3 Chain exchange of ultralong antigen-unknown antibodies 89
3.1 Immunoglobulin chain-exchange 92
3.2 Challenges of *in silico* approaches 92
3.2.1 ABodyBuilder *(formerly SAbPred)* 93
3.2.2 AlphaFold 95
3.2.3 Summary of *in silico* challenges for the prediction of unusual antibodies 96
3.3 Antigen-unknown antibody sequence discovery 96
3.4 Transient transfection of antigen-unknown ultralong Fabs 98
3.4.1 Fab molecular biology preparations 98
3.4.2 Ultralong Fab expression screening 99
3.4.3 Purification of Fab D08 100
3.5 Crystal structure of naturally paired ultralong Fab D08 101
3.5.1 Crystallography 101
3.5.2 X-Ray Diffraction data collection for Fab D08 102
3.5.3 Crystallographic structure of Fab D08 108
3.5.4 PISA crystal packing analysis 109
3.5.5 Performance of VMXi for structural determination 111
3.6 Chain-exchanged ultralong fragment antibody structures 116
3.6.1 Purification of chain-exchanged ultralong antibodies 116
3.6.2 Crystallography 118
3.6.3 Crystallographic structure of ultralong chain-exchanged Fab D08* 123
3.6.4 Pairing-mediated twist to the ultralong stalk 123
3.6.5 Pairing-mediated ultralong paratope warping 124
3.6.6 Crystallographic structure of ultralong chain-exchanged Fab 4K3E* 128
3.6.7 PISA crystal packing analysis 134
3.7 Assessment of AlphaFold 2 prediction accuracy 136
3.8 Discussion 140
3.8.1 Ultralong antibody D08 structure 140

3.8.2 Proposed immunological significance of specific ultralong heavy-light chain pairs . . . 140
3.8.3 Suggestions for further analysis . . . 142

4 Structural and functional characterisation of ultralong antibodies specific for Foot-and-mouth disease virus 143
4.1 Introduction to Foot-and-mouth disease . . . 146
4.1.1 Importance and epidemiology of FMD . . . 146
4.1.2 Foot-and-mouth disease virus . . . 148
4.1.3 Control of foot-and-mouth disease . . . 149
4.1.4 Challenges of inactivated FMDV vaccines . . . 151
4.1.5 Development of VLP vaccine . . . 152
4.2 Sequence Discovery . . . 153
4.2.1 Vaccination experiment . . . 153
4.2.2 Serospecificity assay . . . 155
4.2.3 Sequencing . . . 157
4.2.4 Bioinformatic sequence clustering . . . 161
4.3 Transient transfection of anti-FMDV Fabs . . . 164
4.3.1 FMDV-specific Fab expression screening . . . 164
4.3.2 Purification . . . 166
4.4 Fab apo-structures . . . 168
4.4.1 Crystallography . . . 169
4.4.2 Crystallography of Fabs #117 and #126 . . . 169
4.4.3 AlphaFold Structural Prediction . . . 171
4.4.4 Crystallography of ultralong Fab #30 . . . 181
4.4.5 Crystallographic structure of Fab #30 . . . 182
4.5 Production of FMD VLPs by vaccinia virus co-infection . . . 187
4.6 Fab holo-structures . . . 188
4.6.1 Electron cryomicroscopy specimen preparations . . . 188
4.6.2 Electron cryomicroscopy data collections . . . 189
4.6.3 Interaction of SAT2 12S pentamers by Fab #117 . . . 191
4.6.4 Interaction of SAT2 12S pentamers by Fab #126 . . . 197
4.6.5 Introduction of thermostable SAT2 VLPs . . . 202
4.6.6 Interaction of O1M VLPs by Fab #126 . . . 206
4.6.7 Interaction of O1M VLPs by Fab #117 . . . 212
4.6.8 Interaction of O1M VLPs by Fab #30 . . . 218
4.7 Quantifying ultralong antibody interactions with antigen . . . 223
4.7.1 BLI Experiment optimisation . . . 223
4.7.2 BLI characterisation of Fab #30 with O1M VLPs . . . 223

4.7.3 BLI characterisation of Fab #126 with O1M VLPs 224
4.7.4 BLI characterisation of Fab #117 with O1M VLPs 225
4.8 Discussion and future work 229
4.8.1 Systemic twist predicted by AlphaFold 229
4.8.2 Antigenic interactions mediated by ultralong knob domain . 231
4.8.3 Fab BLI behaviour 231
4.8.4 Quantifying the serospecificity of chain-exchanged Fabs #117 and #126 232
4.8.5 Future work 233

5 Preliminary work towards structural characterisation of chain-exchanged antibodies with the F protein of bovine respiratory syncytial virus 235
5.1 Introduction to bRSV 237
5.1.1 Importance and epidemiology of bRSV 237
5.1.2 Management of bRSV 238
5.1.3 The bovine respiratory syncytial virus 238
5.1.4 Vaccination against bRSV 239
5.1.5 Observations of pairing-mediated differential binding 240
5.1.6 X-Ray diffraction reveals alternative CDR3H conformations 241
5.2 Transient transfection of anti-bRSVF Fabs 242
5.2.1 Biosynthesis of anti-bRSVF Fabs by transient transfection . 242
5.2.2 Affinity chromatography purification of anti-bRSVF Fabs . . 243
5.3 Expression of bRSVF for structural study 245
5.3.1 Transient transfection trial 245
5.3.2 Establishment of a stable cell lineage for bRSVF production 246
5.3.3 Purification of bRSV F protein 248
5.4 bRSVF Biophysical Characterisation 249
5.4.1 Characterisation by SEC-MALS 249
5.4.2 Verification of Fab binding by AUC 253
5.5 Structural biology 260
5.5.1 bRSVF/B4 complexes demonstrate preferential orientation . 260
5.5.2 Resolution of bRSVF/B4 preferential orientation 263
5.5.3 Improving particle density 266
5.5.4 Repurification of F protein complexed with Fab B4 270
5.5.5 High resolution bRSVF/B4 data collection 280
5.6 Conclusions and Discussion 285
5.6.1 Recalcitrant protein purification 285
5.6.2 bRSVF exists in a dynamic concentration-dependent equilibrium 285

5.6.3 Variable degrees of Fab binding 286
5.6.4 Opportunity for structural determination of bRSVF in complex with other Fabs 287

6 General Discussion and Future Work 288
6.1 Thesis purpose 289
6.2 Conclusions from chain-exchanged ultralong antibodies chapter . . . 290
6.3 Conclusions from FMDV-specific chapter 293
6.4 Conclusions from the preliminary bRSVF cryo-EM structural chapter 295
6.5 Summary and final remarks 296

Appendices

A Sample Details 299

B Chapter 2: Methodology 308
B.1 Protein gel filtration standards 308

C Chapter 3: Chain exchange of ultralong antigen-unknown antibodies 310
C.1 Antigen-unknown antibody sequence discovery 310
C.2 Transient transfection of antigen-unknown ultralong Fabs 312
C.2.1 Antigen-unknown Fab expression screening 312
C.2.2 Heavy-light chain recombination affects recombinant expression efficiency 313

D Chapter 4: Functional and structural characterisation of ultralong antibodies specific for Foot-and-mouth disease virus 314
D.1 VHH ELISA confirms non-binding of Fabs #117 and #126 to intact SAT2 314
D.2 BLI experimental optimisation 316
D.2.1 On Fab purification buffer compatibility 316
D.2.2 On the improvement of Fab BLI responses 319
D.2.3 On the elimination of non-specific binding 321
D.3 Whole antibody BLI assessment of #117 an d#126 interaction with O1M 322

E Chapter 5: Preliminary work towards structural characterisation of chain-exchanged antibodies with the bRSV F protein 323
E.1 Inference of bRSVF/B4 complex MW from SEC elution volume . . 324

Bibliography 325

List of Figures

1.1 The anatomy of an antibody . . . 3
1.2 Somatic rearrangement of gene segments to form a viable antibody 5
1.3 Bovine IgH locus . . . 9
1.4 The ultralong CDR3H antibody paratope . . . 16
1.5 Bovine IgL locus . . . 21
1.6 Bovine IgK locus . . . 21
1.7 Multiple sequence alignment of ultralong-paired CDR3L loops . . . 23
1.8 Conservation of CDR3L conformation in ultralong antibodies . . . 24

2.1 pOPIN vector maps . . . 35
2.2 pOPINTTGNeo vector map . . . 36
2.3 VHH ELISA experimental schematic . . . 57
2.4 Chameleon grid vitrification robot . . . 78

3.1 SAbPred predicted CDR3H conformations for bovine antibodies B4 and B13 . . . 94
3.2 Performance of ABodyBuilder 2 predicted CDR3H conformations for ultralong bovine antibody D08 . . . 95
3.3 chain-exchanged ultralong Fab expression screening by western blot 99
3.4 Gel filtration of Fabs D08 . . . 101
3.5 Crystal morphology for Fab D08 . . . 103
3.6 Screening UV absorbance of Fab D08 crystal hits . . . 104
3.7 Experimental strategy for screening with VMXi . . . 104
3.8 X-Ray diffraction data collection for Fab D08 at I04 . . . 106
3.9 Crystal structure of the ultralong Fab D08 . . . 107
3.10 D08 atomic model geometric validations . . . 108
3.11 Aromatic stacking supports the D08 CDR3H stalk domain . . . 109
3.12 PISA interfacing residues of the D08 ultralong loop . . . 111
3.13 DISTL per-mage analysis for VMXi-collected datasets . . . 112
3.14 X-Ray diffraction data collection for Fab D08 at VMXi . . . 113
3.15 Comparison of antibody D08 structures at cryogenic and ambient temperatures . . . 115

3.16 Gel filtration of Fabs D08* and 4K3E* 117
3.17 Crystal morphology for Fab D08* and 4K3E* 120
3.18 X-Ray diffraction data collection for Fab D08* at I04 121
3.19 Crystal structure of the ultralong Fab D08* 122
3.20 D08* atomic model ramachandran plot 123
3.21 Disulphide network organising CDR3H knobs of Fab D08 and D08* 124
3.22 Aromatic stacking supports the D08* CDR3H stalk domain 125
3.23 Packing of CDR3L in Fabs D08 and D08* 126
3.24 Comparison of Fab D08 and Fab D08* knob regions 127
3.25 X-Ray diffraction data collection for Fab 4K3E* at I04 129
3.26 Crystal structure of the ultralong Fab 4K3E* 130
3.27 Molecular interactions affecting the 4K3E* CDR3H stalk domain . 131
3.28 Elbow Angles for Fabs D08, D08*, 4K3E*, and 4K3E 132
3.29 4K3E* atomic model ramachandran plot 133
3.30 PISA interfacing residues of the D08* ultralong loop 134
3.31 PDB Sequence coverage depth for scFv D08 137
3.32 Validation metrics for AF2 modelled scFv D08 structures 138
3.33 Predicted scFv D08 models versus experimental crystal structure . . 139
3.34 Disulphide networks supporting ultralong CDR3H 141

4.1 Serial vaccination to produce cross-protective antibodies 154
4.2 Somatic Hypermutation of AbCellera-discovered antibody sequences 158
4.3 CDR3H lengths of AbCellera-discovered antibody sequences 159
4.4 Sequence clustering of antibody sequences 160
4.5 Multiple Sequence Alignment of antibody #117 and #126 variable domains 162
4.6 Multiple sequence alignment of ultralong-paired CDR3L loops . . . 163
4.7 Expression efficiency screening for clustered ultralong antibodies . . 165
4.8 Purification of Fabs #117 and #126 167
4.9 Purification of Fab #30 168
4.10 Crystallisation screening for Fab #117 and #126 170
4.11 PDB Sequence coverage depth for scFv #117 172
4.12 Predicted models and alignment error for scFv #117 models 174
4.13 Predicted Local Distance Difference Test for scFv #117 175
4.14 PDB Sequence coverage depth for scFv #126 176
4.15 Predicted models and alignment error for scFv #126 models 177
4.16 Predicted Local Distance Difference Test for scFv #126 178
4.17 Structural superposition of AF2 predicted knob mini-domains . . . 179
4.18 Structural overlay of scFv #117 and #126 AlphaFold models 180
4.19 Crystal morphology for Fab #30 183

4.20 X-Ray diffraction data collection for Fab #30 at I03 184
4.21 Crystal structure of the ultralong Fab #30 185
4.22 Reducing SDS-PAGE analysis of O1M VLPs 187
4.23 Electron cryomicroscopy of Fab #117 in complex with FMD serotype SAT2 dissociated 12S pentamers 192
4.24 2D class averages for SAT2 in complex with Fab #117 193
4.25 Fitting of SAT2 pentamer and AF2 scFv #117 to observed cryo-EM density map 194
4.26 Comparison of observed and predicted #117 knob mini-domain structures 195
4.27 The SAT2 12S epitope for antibody #117 196
4.28 Electron cryomicroscopy of Fab #126 in complex with FMD serotype SAT2 dissociated 12S pentamers 198
4.29 2D class averages for SAT2 in complex with Fab #126 199
4.30 Comparison of density maps for SAT2 12S in complex with Fab #117 or #126 201
4.31 Electron cryomicroscopy of FMD serotype SAT2/Eth VLPs 203
4.32 Electron cryomicroscopy of FMD serotype SAT2/Eth VLPs mixed with Fab #117 204
4.33 Electron cryomicroscopy of FMD serotype SAT2/Eth VLPs mixed with Fab #126 205
4.34 Electron cryomicroscopy of Fab #126 in complex with FMD serotype O1M VLPs 207
4.35 2D class averages for O1M in complex with Fab #126 208
4.36 Observed O1M/Fab #126 density map coloured by determined local resolution 209
4.37 Inference of the O1M Fab #126 epitope site 211
4.38 Electron cryomicroscopy of Fab #117 in complex with FMD serotype O1M VLPs 213
4.39 2D class averages for O1M in complex with Fab #117 214
4.40 Observed O1M/Fab #117 density map coloured by determined local resolution 215
4.41 Difference map showing density not explained by O1M crystal structure 217
4.42 Electron cryomicroscopy of Fab #30 in complex with FMD serotype O1M VLPs 219
4.43 2D class averages for O1M in complex with Fab #30 220
4.44 Fab #30 interacts with O1M VLPs at the 3-fold symmetry axis . . 221
4.45 Fitting of O1M capsid and AF2 Fab #30 to observed cryo-EM density map 222

4.46 BLI characterisation of the Fab #30-O1M interaction 224
4.47 BLI characterisation of the Fab #126-O1M and #117-O1M interactions 226
4.48 BLI characterisation of the Fab #63-O1M interaction 227
4.49 Repeated BLI characterisation of the Fab #126-O1M interaction . . 228
4.50 Comparison of initial and revisited AF2 knob mini-domain predictions 230

5.1 Crystal structures for Fabs B4, B4*, B13 and B13* 242
5.2 Purification of Fab B4 244
5.3 Purification of Fab B13 244
5.4 Purification of Fab B13* 245
5.5 PCR Amplification of bRSVF insert from the pVRC8400 vector . . 246
5.6 Anti-His immunoblot of bRSVF stable cell lines 247
5.7 Anti-His immunoblot of bRSVF line #29 expression longevity . . . 247
5.8 Purification of bRSVF 248
5.9 SEC-MALS characterisation of purified bRSVF 250
5.10 Predicted N-glycosylation potential for bRSVF monomers calculated by NetNGlyc 252
5.11 Representative analytical ultracentrifuge sedimentation velocity scans 254
5.12 Analytical ultracentrifuge sedimentation coefficient characterisations of lone proteins 255
5.13 Analytical ultracentrifuge sedimentation coefficient characterisations of complexed bRSVF 257
5.14 Electron cryomicroscopy of Fab B4 in complex with bRSVF, prepared by Vitrobot 261
5.15 Electron cryomicroscopy of bRSVF/Fab B4 complex, prepared by Chameleon 264
5.16 *Ab-initio* 3D reconstruction of bRSVF/B4 complex, with overlaid observed sample pose distribution 265
5.17 Electron cryomicroscopy of bRSVF/Fab B4 complex; graphene-oxide coated grids prepared by Chameleon 268
5.18 Electron cryomicroscopy of bRSVF/Fab B4 complex; graphene-oxide coated grids with reduced Fab excess prepared by Chameleon 269
5.19 Size exclusion chromatography of bRSVF/B4 complex 270
5.20 SEC-MALS characterisation of purified bRSVF/B4 complex 273
5.21 Electron cryomicroscopy of bRSVF/B4 complex fraction 1, prepared by Chameleon 277
5.22 Electron cryomicroscopy of bRSVF/B4 complex fraction 2, prepared by Chameleon 278
5.23 High resolution electron cryomicroscopy of bRSVF/B4 complex fraction 2, prepared by Chameleon 281

5.24 B4 and bRSVF crystal structures fitted to cryo-EM density map . . 282
5.25 Interacting residues of B4 with bRSVF 284

B.1 Measurement of size exclusion chromatography standard elution volumes 309

C.1 Pairing promiscuity from A22-immunised and control animals . . . 311
C.2 Recombinant Fab expression efficiency screened by western blot . . 313

D.1 Homologous ELISA to assess topospecificity of IgG #117 within intact and disrupted serotype SAT2 VLPs 315
D.2 Homologous ELISA to assess topospecificity of IgG #117 within intact and disrupted serotype SAT2 VLPs 316
D.3 BLI optimisation: Fab purity 318
D.4 BLI optimisation: VLP loading concentration 319
D.5 BLI optimisation: Fab concentration 320
D.6 BLI characterisation of the IgG #126-O1M and #117-O1M interactions 322

E.1 Relationship between average distribution constant (K_{av}) and molecular weight 324

List of Tables

2.1 Antibodies used in this project 32
2.2 Recombinant vaccinia viruses used for VLP production 37
2.3 Restriction enzymatic activity for vector double digestion 40
2.4 Vector primers used in this project 44
2.5 Polymerase Chain Reaction Protocol 45
2.6 VHH ELISA Capture antibodies 57
2.7 VHH ELISA Detection antibodies 58
2.8 Biolayer Interferometry Experimental Schematic 63
2.9 List of accessed beamlines 67
2.10 X-Ray data collection parameters 70
2.11 Range of accepted R_{free} values for crystallographic XRD processing 73
2.12 Per-gridtype glow discharge and sample application parameters . . 76
2.13 List of accessed cryomicroscopes 79

3.1 Primers used in bovine antibody sequence discovery 98
3.2 PISA analysis for Fab D08 110
3.3 Individual and merged VMXi crystallographic data table for ultralong antibody structures 114
3.4 Synthetic chain-exchanged Fab purification metrics 118
3.5 Crystallographic data table for antigen-unknown ultralong antibody structures 133
3.6 PISA analysis for Fab D08* 135
3.6 PISA analysis for Fab D08* 136
3.7 Single chain variable fragment (scFv) sequences for antibodies D08 . 136

4.1 AbCellera sequencing summary 156
4.2 Predicted cross-reactive serospecificities from microfluidic approach 156
4.3 Clustering of identified FMDV-specific CDR3H sequences 161
4.4 Antibodies of interest and their serospecificities 164
4.5 Recombinant Fab yields obtained by transient transfection of Expi293F™ cells 166
4.6 Single chain variable fragment sequences for antibodies #117 and #126 172

4.7 Structural superposition matrix of AF2 scFv #117 knob mini-domains 175
4.8 Structural superposition matrix of AF2 scFv #126 knob mini-domains 178
4.9 Crystallographic data table for Fab #30 structure 186
4.10 Chapter 4 cryo-EM specimen preparation and data collection 190
4.11 Chapter 4 cryo-EM exposure curation and particle picking 190
4.12 BLI characterisation of the Fab #30-O1M interaction results table . 225
4.13 BLI characterisation of the Fab #126-O1M interaction results table 227

5.1 Synthetic chain-exchanged Fab purification yields 249
5.2 SEC-MALS characterisation of bRSVF 251
5.3 Sedimentation coefficient results and estimated molecular weights for lone proteins 256
5.4 Sedimentation coefficient results and estimated molecular weights for bRSVF complexes 258
5.5 Predicted species concentrations for bRSVF/Fab complexes 259
5.6 Inferred species molecular weight by SEC elution volume 271
5.7 Summary of bRSVF/B4 complex spin-concentration 272
5.8 SEC-MALS characterisation of bRSVF/B4 complex fraction #1 . . 275
5.9 High resolution bRSVF/B4 cryo-EM exposure curation 280
5.10 High resolution bRSVF/B4 cryo-EM particle picking 280

6.1 Summary of presented structural antibody work 290

1 Antibody sequences used in this project 299
A.2 bRSVF ProtParam Results 305
A.3 ProtParam Results 306
A.4 Antibody gene designations 307

B.1 Distribution constants for gel filtration protein standards 308

D.1 BLI optimisation: VLP loading concentration table 319

List of Abbreviations

2D . . . . . 2-dimensional (etc.)

AF2 . . . . . AlphaFold v2.0

APHA . . . . . Animal and Plant Health Agency

ASCs . . . . . antibody-secreting cells

AUC . . . . . analytical ultracentrifuge

BAC . . . . . bacterial artificial chromosome

BBSRC . . . . Biotechnology and Biological Sciences Research Council

βME . . . . . β-mercaptoethanol

BLI . . . . . bio-layer interferometry

BMRC . . . . Biomedical Research Computing

bp . . . . . base pairs

BRC . . . . . Biomedical Research Centre

BRD . . . . . bovine respiratory disease

bRSV . . . . . bovine Respiratory Syncytial Virus

bRSVF . . . . stabilised prefusion conformation bovine Respiratory Syncytial Virus fusion protein

BSA . . . . . bovine serum albumin

BSE . . . . . bovine spongiform encephalopathy

BSL . . . . . biosafety level

BTA21q24 . . *Bos taurus* chromosome 21, locus q24

BVDV . . . . . bovine viral diarrhoea virus

CASP . . . . . Critical Assessment of protein Structural Prediction

CCC . . . . . cross correlation coefficient

CDR . . . . . complementarity determining region

CDR3H . . . . complementarity determining region 3 of the heavy chain (etc.)

CH . . . . . constant domain of the immunoglobulin heavy chain

CL . . . . . constant domain of the immunoglobulin light chain

CMV . . . . . cytomegalovirus

cryo-EM . . . cryogenic electron microscopy

cryoSPARC . cryo-EM single-particle *ab-initio* reconstruction and classification

CTF . . . . . Contrast Transfer Function

DISTL . . . . . Diffraction Image Screening Tool and Library

DLS . . . . . Diamond Light Source

DMEM . . . . Dulbecco's Modified Eagle Medium

DNA . . . . . deoxyribose nucleic acid

dpi . . . . . days post immunisation

dRI . . . . . differential refractive index

eBIC . . . . . The electron Bio-Imaging Centre

EDTA . . . . . Ethylenediaminetetraacetic acid

ELISA . . . . . enzyme-linked immunosorbent assay

ELISPOT . . . enzyme-linked immunosorbent spot assay

ESS . . . . . effective sample size

etc . . . . . et cetera

eV . . . . . electron Volts

Fab . . . . . antibody fragment

FBS . . . . . foetal bovine serum

FCS . . . . . foetal calf serum

f/f_0 . . . . . frictional ratio parameter

FMD . . . . . foot-and-mouth disease

FMDV . . . . foot-and-mouth disease virus

FPS . . . . . frames-per-second

FR . . . . . framework region

FSC . . . . . Fourier Shell Correlation

GFCK . . . . . growth factor cysteine knot

GL . . . . . germline

GMO . . . . . Genetically Modified Organism

GMP . . . . . Good Manufacturing Practice

GO . . . . . graphene oxide

GPU . . . . . graphics processing unit

GPCR . . . . . G-protein coupled receptor

GSFSC . . . . gold-standard Fourier shell correlation

HBS-EP . . . HEPES buffer system supplemented with EDTA and Polysorbate-20

HEK . . . . . human embryonic kidney

HEPES . . . . 4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid

HIV . . . . . human immunodeficiency virus

HRP . . . . . horseradish peroxidase

hRSV . . . . . human Respiratory Syncytial Virus

hRSVF . . . . stabilised prefusion conformation human Respiratory Syncytial Virus Fusion Protein

Hz . . . . . hertz

ICK . . . . . inhibitory cysteine knot

IDT . . . . . Integrated DNA Technologies

IgA . . . . . immunoglobulin isotype A

IgD . . . . . immunoglobulin isotype D

IgE . . . . . immunoglobulin isotype E

IgG . . . . . immunoglobulin isotype G

IgH . . . . . immunoglobulin heavy chain

IgL . . . . . immunoglobulin light chain

Igλ . . . . . immunoglobulin isotype λ

Igκ . . . . . immunoglobulin isotype κ

IgM . . . . . immunoglobulin isotype M

IMAC . . . . . immobilised metal ion affinity chromatography

IMGT . . . . . international ImMunoGeneTics information system

IPTG . . . . . isopropylthio-β-galactoside

IRES . . . . . internal ribosome entry segment

ITB . . . . . The Immunological Toolbox

K . . . . . kelvin

K_a . . . . . association rate constant

K_{dis} . . . . . dissociation rate constant

K_D . . . . . equilibrium dissociation constant

kDa . . . . . kilodalton

LB . . . . . luria broth

LDS . . . . . lithium dodecyl-sulphate

mRNA . . . . messenger RNA

MOI . . . . . multiplicity of infection

MP . . . . . mass photometry

MSA . . . . . multiple sequence alignment

MSD . . . . . Merck, Sharp & Dohme Corporation

MWCO . . . . molecular weight cut off

MX . . . . . macromolecular crystallography

NC . . . . . nitrocellulose

NHS . . . . . National Health Service

NIHR . . . . . National Institute of Health and Care Research

NMR . . . . . nuclear magnetic resonance

NSB . . . . . non-specific binding

NVC . . . . . non-vaccinated control

OIE . . . . . Office International des Epizooties, former name of WOAH

OPIC . . . . . Oxford Particle Imaging Centre

OPPF . . . . . Oxford Protein Production Facility, former name of PPUK

PAE . . . . . Predicted Alignment Error

PBMCs . . . . peripheral blood mononuclear cells

PBS . . . . . phosphate buffered saline

PCR . . . . . polymerase chain reaction

PCT . . . . . Pre-Crystallisation Test

PDB . . . . . Protein Data Bank

PEI . . . . . Polyethylenimine Hydrochloride

PEG . . . . . polyethylene glycol

PFU . . . . . plaque forming units

pLDDT . . . . predicted Local Distance Difference Test

PPUK . . . . . Protein Production UK

QCM-D . . . . quartz crystal microbalance with dissipation

RCF . . . . . relative centrifugal force

RFI . . . . . Rosalind Franklin Institute

$\mathbf{R_{max}}$ . . . . . maximal response

RMSD . . . . . root-mean-square deviation

RNA . . . . . ribonucleic acid

RPM . . . . . revolutions per minute

RT . . . . . room temperature

(s) . . . . . sedimentation coefficient

SAPO . . . . . Specified Animal Pathogens Orders

scFv . . . . . single-chain antibody variable fragment

SDG . . . . . sucrose density gradient

SDS . . . . . sodium dodecyl-sulphate

SDS-PAGE . . SDS polyacrylamide gel elctrophoresis

SEC-MALS . size exclusion chromatography with multi-angle light scattering

SGD . . . . . stochastic gradient descent

SPR . . . . . surface plasmon resonance

STRUBI . . . Division of Structural Biology

SV . . . . . sedimentation velocity

SV40 . . . . . simian vacuolating virus 40

TdT . . . . . terminal deoxynucleotidyl transferase

TE . . . . . Tris-EDTA

TBE . . . . . Tris-borate-EDTA

TMB . . . . . 3,3',5,5'-Tetramethylbenzidine

UCSF . . . . . University of California, San Francisco

UK . . . . . United Kingdom

UKRI-BBSRC UK Research and Innovation

UV . . . . . ultraviolet

vCJD . . . . . variant Creutzfeldt–Jakob disease

VH . . . . . variable domain of the immunoglobulin heavy chain

VHH . . . . . antigen-binding fragment of heavy chain only antibody

VL . . . . . variable domain of the immunoglobulin light chain

VLP . . . . . virus-like particle

VMXi . . . . . versatile macromolecular crystallography *in situ* beamline

VNT . . . . . virus neutralisation titre assay

VP1 . . . . . viral protein 1 (etc.)

VV . . . . . vaccinia virus

(v/v) . . . . . volume-by-volume

(w/v) . . . . . weight-by-volume

WOAH . . . . World Organisation for Animal Health

WBVR . . . . Wageningen Bioveterinary Research

WPRE . . . . woodchuck hepatitis virus post-transcriptional regulatory element

XRD . . . . . X-Ray crystallographic diffraction

Life, uh. . . , finds a way.

— Dr Ian Malcolm, Jurassic Park

1 Introduction

Contents

1.1 Canonical diversification of immunoglobulin repertoires 2
1.2 Genetics of the bovine heavy chain locus 7
1.2.1 Organisation of the bovine heavy immunoglobulin locus 7
1.2.2 Limited VDJ recombinations 9
1.2.3 Genetic sequence invariance 11
1.2.4 Somatic diversification 11
1.2.5 Exceptionally long CDR3H loops 12
1.3 Genetics of the bovine light chain locus 20
1.3.1 Organisation of the bovine light immunoglobulin locus . 20
1.3.2 Limited genetic diversity 22
1.3.3 Somatic diversification of bovine light chains 22
1.4 Pairing peculiarities 22
1.5 A summary of bovine structural immunology 25
1.6 A curious immunological paradigm 26
1.7 Project aims and objectives 28

Figures

1.1 The anatomy of an antibody 3
1.2 Somatic rearrangement of gene segments to form a viable antibody 5
1.3 Bovine IgH locus 9
1.4 The ultralong CDR3H antibody paratope 16
1.5 Bovine IgL locus 21
1.6 Bovine IgK locus 21
1.7 Multiple sequence alignment of ultralong-paired CDR3L loops . 23
1.8 Conservation of CDR3L conformation in ultralong antibodies . 24

1.1 Canonical diversification of immunoglobulin repertoires

In mammals, protection against invading foreign bodies is achieved, in part, through the humoral immune system. As its eponymous name suggests, the humoral immune system utilises an extensive array of macromolecular chemical weapons, secreted into the body humours (fluids) to specifically target invaders, as opposed to the innate immune system, which comprises a non-specific array of cellular physicochemical barriers to infection.

The weapons of the humoral immune system include proteins of the complement system, a series of antimicrobial peptides, and predominantly, 'antibodies'. First described in 1890 [1], antibodies act as immune effector molecules, capable of recognising a specific foreign body, an 'antigen', with great precision, and initiating one of several defined adaptive immune responses. Such highly dynamic function is consequential of the antibody structure.

A single antibody typically constitutes a symmetric heterotetramer of two 'heavy' and two 'light' chains [2], joined by inter-chain disulphide bonds [3]. The chains comprise a repeated series of immunoglobulin (Ig) protein domains; light chains have only two Ig domains, whilst heavy chains have a variable, but no less than four, number of Ig domains [4–6]. Immunoglobulin domains that interact with antigen are named 'variable' domains ('VH' or 'VL' for short), whilst the remainder are labelled as 'constant' domains ('CH' or 'CL') [4–6] (Fig. 1.1). An antibody can be functionally divided into three fragments; two identical fragment variable ('Fv') regions, and a fragment constant ('Fc') region [4–6] (Fig. 1.1).

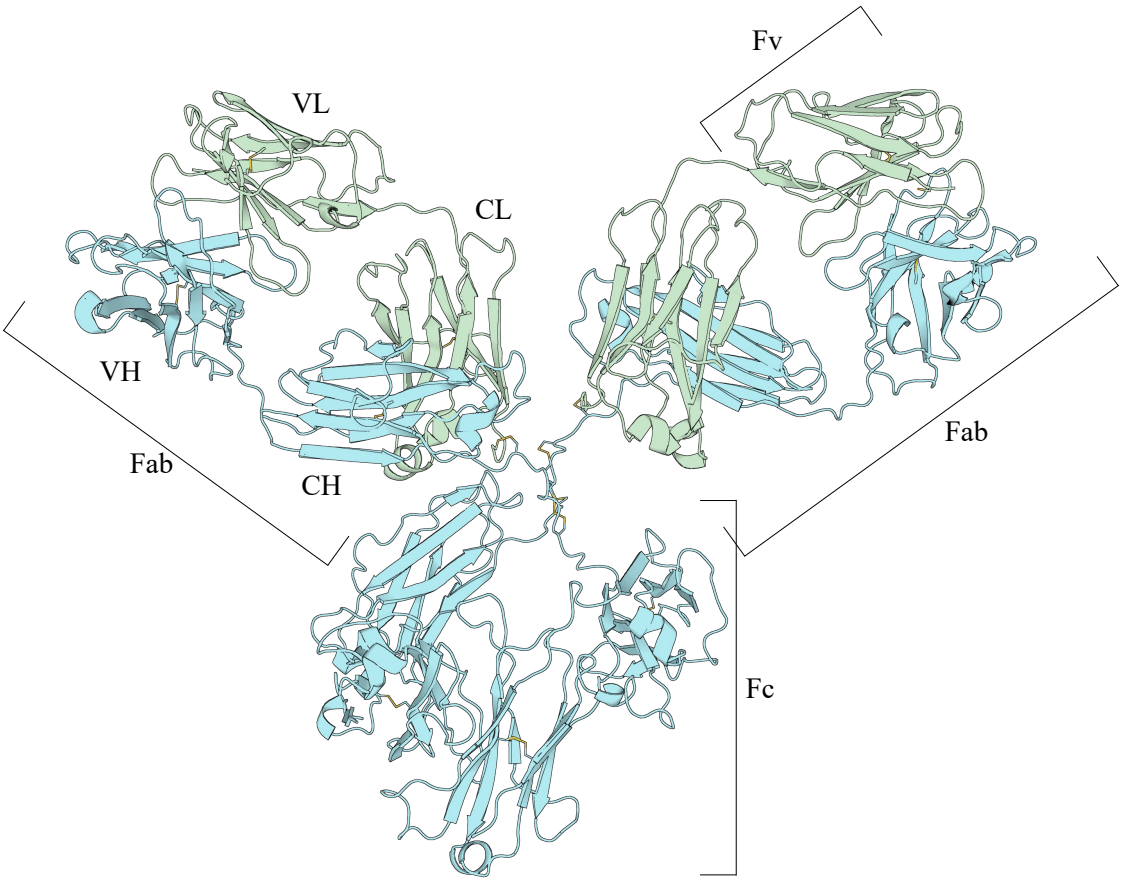


Figure 1.1: The anatomy of an antibody. Human immunoglobulin isotype G (PDB ID: 1IGY) [7] shown in cartoon representation. Heavy chains are coloured blue, light chains are coloured green. Disulphide bonds are indicated in orange stick notation. Fragments, and Fab domains are labelled for clarity. Figure prepared using PyMol.

The Fc region initiates a variety of immune responses; mediating antibody dependent cellular cytotoxicity via interaction with a complementary Fc receptor (FcR), or recruitment of complement protein C1 [4–6]. Antibodies may also block the interaction between antigen and host cell surface molecules. Antibodies are classified into distinct 'isoforms' by their effector function independent of antigen specificity [4–6], with humans having five isoforms.

Isoform M (IgM) predominantly constitute five (or infrequently six) copies of the heterotetrameric unit connected at the Fc regions by an additional J chain. IgM can also be expressed on B cell surfaces as monomers. IgM is especially prevalent in early infection; able to agglutinate multiple antigens together via their ten antigen recognition sites, and efficiently recruit and activate the complement system.

Isoform G (IgG) is the most abundant isoform in serum, constituting 70 - 85 % of total circulating immunoglobulin and has the longest circulating antibody half-life. Comprising one monomer of the heterotetrameric unit, IgG is subdivided into subclasses; IgG1 less efficiently activates the complement system than IgM, whereas IgG2 is almost unable to activate complement, and is therefore more beneficial in chronic infection to limit inflammation. IgG is also the only isoform capable of crossing the hemochorial placenta of humans, rodents and lagomorphs, and hence offers maternal immunity during foetal gestation [8].

Isoform A (IgA) is secreted into the mucosa in abundance, to defend the body's natural orifices from invaders. To this end maternal milk also contains a large quantity of IgA to coat the intestinal lining of offspring with IgA. IgA can exist either as a monomer, or a dimer joined by the J chain.

Isoform E (IgE) is mostly present on the cell surface of basophils and mast-cells, and has an additional two constant domains in place of a hinge region. IgE may have principal function in defending against larger invaders, such as parasitic worms, but is most commonly associated with immediate allergen hypersensitivity and resulting anaphylaxis.

Isoform D (IgD) has the same principal structure as IgG, though with a hinge region particularly sensitive to proteolytic digestion, and is mostly present on B cell surfaces. IgD is thought to act as an immune cell activator, but a precise function has not yet been determined.

Other isoforms do exist in nature (IgT [9], IgX [10, 11]), but in all cases except the heavy chain only (VHH) antibodies of camelids [12], and the new antigen receptors (NAR) of sharks [13, 14], contain an antigen recognition site comprising small portions of both a heavy chain and a light chain.

The Fv regions mediate this interaction with antigen via six hypervariable loops (three per variable domain), comprising the complementarity determining region (CDR). The remainder of the variable domain is called the framework region (FR), and acts as a scaffold in supporting the critical CDR structure.

There exists a potentially infinite array of antigen molecular structures, and a functional humoral immune response is therefore dependent on the ability of an organism to generate substantial antibody diversity, particularly at the CDR [15]. There are three core immunological mechanisms for achieving immunoglobulin sequence diversification; (1) access to, and recombination of several gene segments (recombination diversity), (2) flexibility at the junctions of gene segment recombination (junctional flexibility), and (3) via cellular introduction of mutation (somatic hypermutation). The contribution of both heavy and light chain CDRs to antigenic recognition means that with the exception of VHH's, heavy-light chain pairing offers a further mechanism of diversification.

Primary immunoglobulin diversification in humans and rodents is principally generated through somatic rearrangements [16] of the germline variable (V), diversification (D), and junctional (J) gene segments to form a viable VH domain (Fig. 1.2) [17, 18]. This process is mirrored in the formation of the VL domain, but without D gene segments [19–21]. Recombinational diversity is further diversified by imprecise junctional joining wherein nucleotides are removed or added *de novo*.

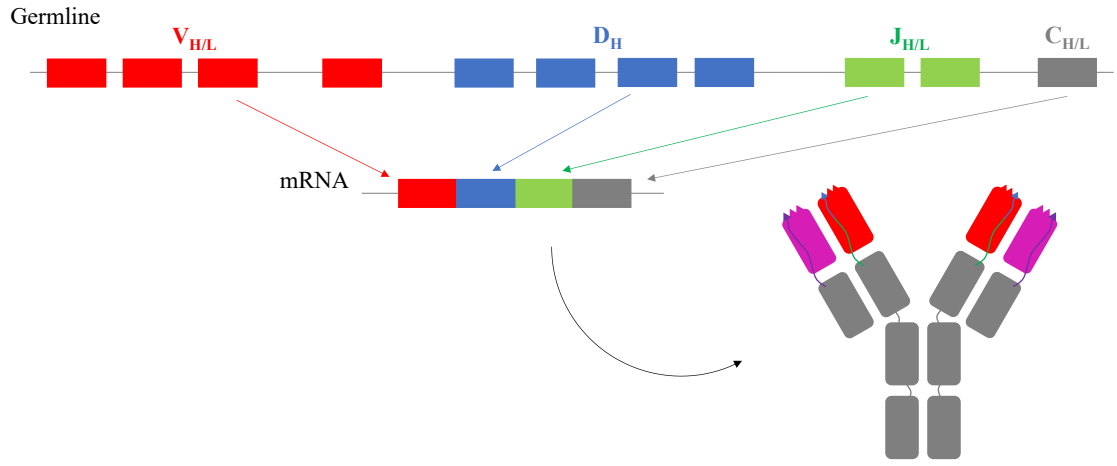


Figure 1.2: Somatic rearrangement of V, (D), J and C gene segments form a viable antibody. Antibody heavy chain variable domains are produced from rearrangement of V, D, and J gene segments, whilst light chain variable domains use only V and J gene segments. Cartoon indicates regions encoded by each gene segment for each chain, with immunoglobulin domains represented by rounded rectangles, and complementarity determining regions indicated by triangles.

Fully formed antibodies are expressed at the surface of mature naïve B cells as B cell receptors (BCR) [22]. The final mechanism of diversification is induced upon antigen stimulation of the BCR, with or without helper T cell co-stimulation, which activates the B cell [23]. Activated B cells express activation-induced cytidine deaminase (AID) [24] which selectively deaminates cytosine to uracil [25] within the V(D)J genes encoding the CDR of the antibody which facilitates antigen recognition. These mutations are repaired with low fidelity [26], with resulting mRNA described as 'somatically hypermutated' [27]. The mutated BCR is expressed, and the cycle repeats, until B cell clones expressing BCRs with high affinity against the antigen are achieved, outcompeting B cell clones with lower binding affinity BCRs. Iteratively performed, this process is called affinity maturation [27]. Imprecise recombination of germline gene segments also contributes to immunoglobulin diversification [6].

Establishing a diverse primary antibody repertoire is therefore highly dependent on an initial diverse germline repertoire of V, (D), and J gene segments. It is interesting then, to note the highly restrained genetic repertoire of immunoglobulin variable region gene segments found in cattle.

Despite recent innovations, a key limiting factor in understanding the bovine adaptive immune system is the lack of species-specific reagents and knowledge to enable high throughput characterisation of bovine antibody repertoires [28]. Maintaining populations of bovine B cells is not well evidenced, and cell sorting pipelines are not yet fully optimised for identification, and isolation of bovine B cells, despite recent advances for the isolation of bovine T lymphocytes [29] and B lymphocytes [30]. To further complicate matters, conventional sequencing pipelines typically applied to the characterisation of immune repertoires struggle with the extended bovine CDR3H transcripts [31], whilst genome assembly is complicated by the highly invariant and polymorphic nature of bovine immunoglobulin genes, particularly of the light chain [32]. Overall these limitations have slowed the advancement of bovine antibody understanding.

1.2 Genetics of the bovine heavy chain locus

For clarity, genes within this thesis are described using the IMGT® numbering system [33], adapted for the organisation of the bovine system [34–36] which applies the following formats:

V genes: Ig{H/L/K}V{gene family number}-{3' - 5' locus}
D genes: IgHD{gene family number}-{5' - 3' DH cluster number}
J genes: Ig{H/L/K}J{gene cluster number}-{5' - 3' locus}

For example, IgHV3-15 describes a heavy chain V gene segment; genetic family 3 at the 15^{th} locus, whilst IgHD1-2 describes a DH gene segment of genetic family 1, located within DH cluster 2. If appropriate, allelic information is suffixed to this nomenclature following an asterisk (i.e. IgLV1-47*01).

Constant domains are conventionally described by the isoform they encode (IgHM1, IgHM2, IgHD, IgHG3, IgHG1, IgHG2, IgHE, IGHA), and suffixed with 'ψ' for pseudogenes (IgHDD1ψ, IgHDD2ψ, IgHDD3ψ).

Deviations from the IMGT® standard, such as during citation of literature published prior to the adoption of the IMGT® standard, will be described as such.

1.2.1 Organisation of the bovine heavy immunoglobulin locus

The genes encoding bovine immunoglobulin heavy chains (IgH) exist within a single locus at q24 of the *Bos taurus* autosome 21 (BTA21q24) (Fig. 1.3) [34]. Forty six bovine V genes, which encode immunoglobulin framework regions (FR) 1 to 3, and CDRs 1 and 2, are subdivided to three gene families by sequence homology, of which only twelve genes of the IgHV1 family are functional [37, 38], the remainder of IgHV genes, including both the IgHV2 and IgHV3 gene families, being pseudogenes due to in-frame stop codons or sequence deletions [34, 39]. The most frequently utilised V genes, in both adult and foetal samples are IgHV1S39 and IgHV3/33 [40, 41], though this numbering does not correlate to the IMGT numbering scheme [33]. It is important to remember that although not directly expressed, pseudogenes may contribute to sequence diversification by gene conversion [40, 42, 43].

The modest bovine functional V gene segment library is followed by a -(D_n - J_n - membrane region-encoding domain 1)- gene segment cassette, and a triple tandem repeat consisting of a D pseudogene segment, a V pseudogene segment, and a functional D locus (Fig. 1.3) [34]. The four D gene segment loci (in distinct clusters spanning 68 kb) are subdivided into nine subgroups containing twenty-three IgHD gene segments in total, with sixteen of these being functional. Highly divergent recombination signal sequences explain the non-functionality of the remainder of the D gene segments. The IgH locus is concluded with a second -(J_n - membrane region-encoding domain 2)- segment, and finally, a 3' cassette containing the IgH constant (IgHC) domain genes for isoforms D, G, E, and A respectively (Fig. 1.3). Cattle express functional immunoglobulins of all common mammalian isoforms; IgM [44], IgG [45–47], IgE [48], IgA [49] and IgD [50].

The proximity of the 3' J_n locus to the IgHC domain genes results in greater expression efficiency than the 5' J_n locus, introducing a translational preference to J gene segments. Four J gene segments have been described as functional; two within the infrequently transcribed 5' J locus (IgHJ1-4, IgHJ1-6), and two within the more efficiently transcribed 3' J locus (IgHJ2-4, IgHJ2-6) (Fig. 1.3). Rearrangements of the D-J gene segment junction encodes a portion of the CDR3H, and FR4 of the immunoglobulin variable domain [51, 52], however the high observational invariance of FR4 implies a single JH gene segment, IgHJ2-4, is preferentially rearranged with multiple DH segments [37].

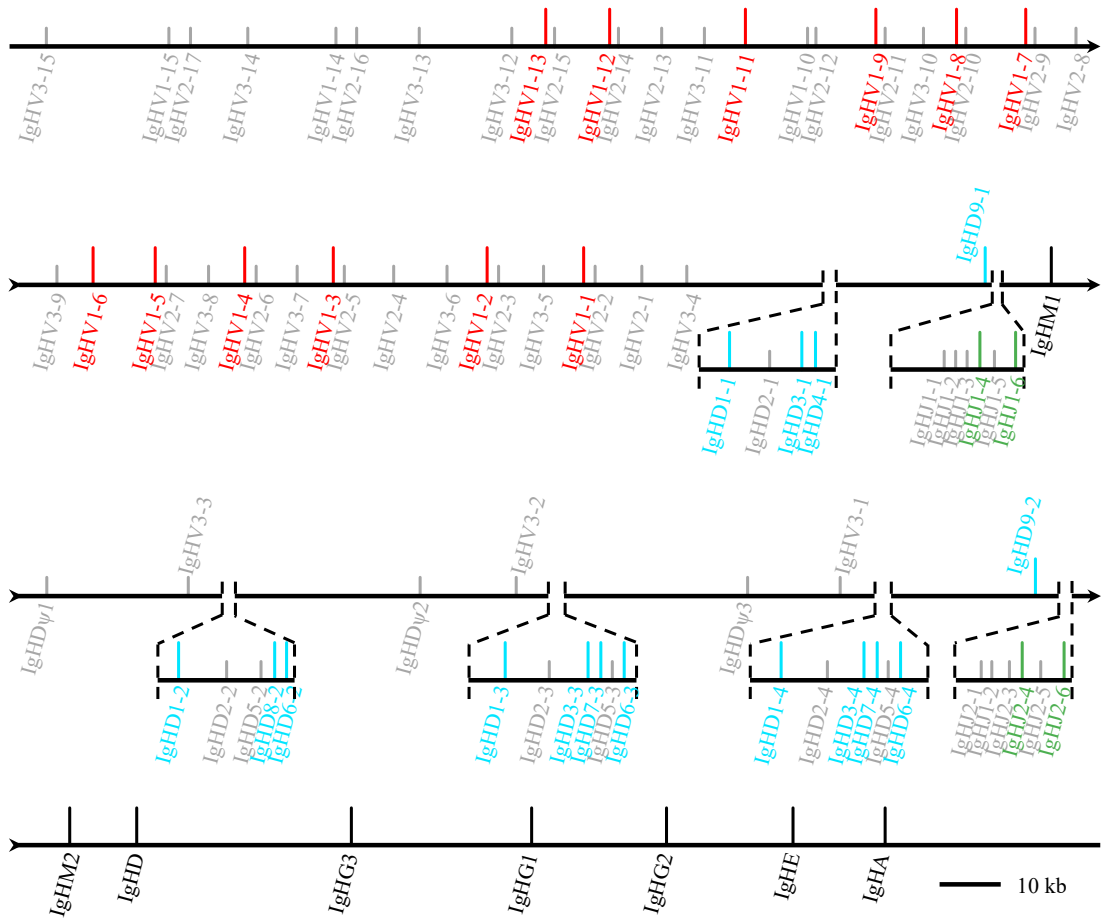


Figure 1.3: Organisation of the cattle antibody IgH chain loci in the UMD3.1 genome assembly. The IgH locus is located on *Bos taurus* autosome 21, locus q24. Vertical bars indicate the position of individual gene segments. Long bars indicate putatively functional genes, whilst short bars indicate pseudogenes and are coloured grey. V, D, J, and C gene segments are coloured red, blue, green, and black respectively for clarity. Scale bar indicates 10 kb. Figure adapted from Ma *et al.* [34] and Smider *et al.* [53].

1.2.2 Limited VDJ recombinations

In total, the twelve functional IgHV, sixteen functional IgHD and four functional IgHJ gene segments offer a theoretical combinatorial diversity of 768 possible recombinants for cattle. However it remains unknown how many of these 768 would resolve as productive heavy chain transcripts. Many may not form complete transcripts with absence of stop codons, or produce heavy chains that cannot pair with a light chain given bovine light chains are also genetically constrained (Sections 1.3.1 and 1.3.2). Additionally, it is intuitive that many transcripts may

encode self-targeting paratopes, and hence would be eliminated through clonal deletion. The scope for diversification through VDJ rearrangement is hampered further given that of the functional IgHV genes, IgHV1-21 and IgHV1-33 are identical at the nucleotide level, whilst IgHV1-25 and IgHV1-37 are identical at the amino acid level [34, 54]. Many bovine IgHV genes appear to be underutilised; depending on D gene classification, 162 rearrangements were observed across four cattle breeds (Aubrac, German Simmental, German Black Pied, Holstein-Friesian), with variation between breeds. By contrast, an estimated 7,590 potential recombinants exist in humans [40, 55] through recombination of 55 functional IgHV genes, 23 functional IgHD genes and 6 functional IgHJ genes [56, 57]. For all cattle breeds examined, IgHV36/29(not IMGT numbering) - IgHD8-2 - IgHJ2-4 was the most frequently observed rearrangement, but such a preference varies by breed and by CDR3H length [40, 58].

The broad distribution of CDR3H lengths will be discussed later (Section 1.2.5), but the cattle antibody repertoire demonstrates a trimodal distribution of CDR3H lengths [31, 59–61]. Distinct preferential rearrangements are observed for short, medium and long CDR3H in multiple bovine species [54, 58]. In Holstein-Friesian cattle (*Bos taurus*), IgHV1-10 dominates (72.7 % of sequences) the antibody landscape for short and medium length CDR3H's, whilst long CDR3H loops ubiquitously use IgHV1-7 rearranged with IgHD8-2 [54]. The IgHV1-7 - IgHD8-2 rearrangement also dominates long CDR3H antibodies in the Ankole and Boran breeds of Zebu cattle (*Bos indicus*) [58]. Short and medium length antibodies of *B. indicus* also appear to demonstrate some preferential gene usage and rearrangement pattern, however to much more limited extent, possibly owing to the smaller sample size and lack of complete *B. indicus* IgH assembly making analysis difficult [58]. In both *B. taurus* and *B. indicus*, IgHJ2-4 dominates all observed VDJ rearrangements [54, 58].

1.2.3 Genetic sequence invariance

Expansion of bovine antibody repertoires is not only constrained by the choice of VDJ rearrangement (Section 1.2.2); having multiple gene segments does not necessarily increase the diversity of the immunoglobulin repertoire. As mentioned previously (Section 1.2.2), bovine immunoglobulin germline sequences are highly invariant, with variation between genes almost exclusively occurring at or near the CDR [62, 63]. As previously introduced (Section 1.2.1), functional VH genes belonging to the same genetic family offer little diversity to the immune repertoire, with 83.6 – 93.2 % shared nucleotide sequence homology, after discounting the identical IgHV1-21 and IgHV1-33, and IgHV1-25 and IgHV37 [34, 54]. Indeed, CDR1H and CDR2H sequences, encoded by the IgHV gene, share >85 % sequence homology to the consensus germline [64]. Genes of the non-functional, pseudogene-encoding IgHV2 and IgHV3 families [37, 38] would offer more sequence diversity, perhaps via gene conversion as suggested with the IgL locus [51, 65], however gene conversion at the bovine IgH locus has not yet been described (Section 1.2.4).

The three IgHD loci within the triple tandem repeat (Section 1.2.1 and Fig. 1.3) are also homologous, with the second IgHD locus (DH2) gene cluster sharing 92.0 % nucleotide identity to DH3 and DH4, whilst DH3 and DH4 share 99.7 % identity at the germline nucleotide level [53]. Further still, most bovine DH gene segments are principally differentiated by varying numbers of repeated nucleotide motifs, further limiting the scope for diversification [53]. For example, nearly two thirds of IgHD8-2 nucleotides are the same repeating codons, resulting in a germline CDR3H sequence with a 58.8 % pairwise identity [54].

1.2.4 Somatic diversification

Human and mice, with many functional gene segments present in the germline, principally use V(D)J rearrangement to establish a diverse immunoglobulin repertoire [5, 17–20, 66]. A common trait of species with reduced genomic opportunity for immunoglobulin diversification is the ability to increase sequence diversity through somatic means without the requirement for B cell activation to

trigger mechanisms such as AID. Chickens for example, employ gene conversion and somatic hypermutation to diversify a single VDJ rearrangement [42, 43, 67], whilst rabbits achieve diversification by gene conversion in combination with somatic non-template point mutation [68–70] of a limited number of preferential VDJ rearrangements [71]. Finally, somatic hypermutation in absence of prior antigen exposure [72, 73] diversifies the limited repertoire of sheep [74].

Intuitively, given the presence of several psuedogenes within the germline IgH locus (Sections 1.2.1 and 1.2.2), gene conversion was initially proposed for cattle to achieve diversification [51, 65]. However, given the polymorphic nature of bovine heavy chain gene segments, and the incidence of mutations in both coding and non-coding regions, somatic hypermutation is generally favoured as a mechanism for heavy chain diversification [31, 37, 39, 59, 75, 76]. Immune sequence diversification is evident in animals from 1 day old. In day old calves, there is a tentatively higher mutational frequency in the antigen-interacting CDR than the supporting FR, that becomes clearer as the calves age [76]. Heavy chain sequence diversification by somatic hypermutation is also observed in VDJ recombinations of bovine foetal samples (hence collected during B cell ontogeny), implying diversification proceeds in cattle in absence of antigen exposure [39, 77], consistent with observations made in sheep [72, 73], or that terminal deoxynucleotidyl transferase (TdT), the enzyme responsible for non-template terminal nucleotide addition during V(D)J recombination, is not expressed until a later stage of calf development.

1.2.5 Exceptionally long CDR3H loops

With the limited opportunity for immunoglobulin sequence diversification by germline sequence (Section 1.2.3) or rearrangement (Sections 1.2.1 and 1.2.2), cattle have developed an alternative strategy to accomplish antibody diversity; by developing a broad range of CDR3H lengths. Bovine CDR3H average 22 - 26 residues in length [40, 60, 78, 79]; longer than the reported average of 15.5 amino acids in humans or 11.5 in mice [39, 80].

The high homology of bovine CDR1H and CDR2H to the germline sequence results in these loops as generally observed as invariant in length [64]. On the other hand, and as introduced earlier (Section 1.2.2), cattle present a broad trimodal distribution of CDR3H lengths, unique amongst mammalian immune systems [31, 59–61]. The first mode describes short loops with a CDR3H of <10 residues. The second mode, and highest populated, includes CDR3H loops of 11 - 35 residues. Finally, a small population (5 - 12 %, depending on breed) of bovine heavy chains harbour CDRH3 of >48 residues [31]. Within the bovine antibody system, antibodies containing a CDR3H of >48 amino acids are described as 'ultralong' [31, 34].

Such ultralong CDR3H loops arise, in part, from VDJ rearrangements exclusively involving IgHV1-7 and IgHD8-2. IgHV1-7 carries a 3' 8 bp extension compared to other VH gene segments, introducing three additional codons and a frameshift [54]. IgHD8-2, of the DH2 locus, has an extended length of 148 bp (encoding 49 residues) [34, 81]. The extraordinary IgHD8-2 gene segment is the fusion product of primordial IgHD6 and IgHD7 genes in the DH2 locus, as part of a larger chromosomal rearrangement in which a 5' IgHD3 paralog in the same locus was replaced with IgHD5-2 [53]. For reference, the DH3 and DH4 clusters are organised such that the IgHD segment belonging to family 5 follows IgHD segments belonging to families 3 and 7 respectively (Fig. 1.3). Such rearrangement is further corroborated by the repeated structure of the IgH locus in cattle [34] (Section 1.2.1 and Fig. 1.3), and the high nucleotide identity (89.4 %) between the 5' 85 nucleotides of IgHD8-2 with the 5' region of IgHD6-2, whilst the trailing 3' 96 bp of IgHD8-2 shares homology to IgHD7-3 (85.6 % identity) and IgHD7-4 (77.4 % identity) [53].

However, even the 154 bp IgHD gene segment does not fully explain observations of ultralong CDR3H exceeding 60 amino acids [31, 59, 78, 82–85]. Insertion of 'conserved short nucleotide sequences' (13 - 18 non-palindromic, A-rich, nucleotides) at the VH-DH junction, independent of TdT activity [66], further extends the ultralong CDR3H loop [76, 77]. N-region addition and P-nucleotide insertions at both the VH-DH and DH-JH recombination junctions, and junctional flexibility [77] also contribute to CDR3H transcript extension. Curiously, the insertion of

conserved short nucleotide sequences is not observed in foetal transcripts, and may therefore be a component of a response to antigen or, linked to the previously indicated delayed expression of TdT.

The evolution of ultralong CDR3H have been considered as a unique strategy to compensate for the limited germline diversity of the heavy locus in bovine immunology [31, 39, 64].

1.2.5.1 Hypermutability of ultralong CDR3H

Ultralong CDR3H are diversified in the same manner as non-ultralong antibodies (Section 1.2.4); principally via AID-mediated somatic hypermutation [75].

In other mammals, AID is induced upon B cell activation [23, 24]; however bovine foetal B cells show evidence of AID activity [86], indicating that AID-mediated somatic hypermutation may proceed in absence of antigen exposure as an additional compensatory mechanism in cattle. The syndesmochorial type of placenta of cows prevents the passage of maternal antibodies to the developing foetus [8, 59], and hence foetal serum represents the preimmune repertoire. This observation is corroborated by the observation of diversified ultralong CDR3H sequences from foetal specimens [77], and day old calves [76], though maternal antibodies are transferred to the calf via colostrum [8].

Additionally, cattle have a novel mechanism for diversification of long CDR3H loops encoded by IgHD8-2. Approximately half of transcripts produced by rearrangements with IgHD8-2 show in-frame CDR3H deletion events (up to 18 codons) compared to the germline sequence [54], with proclivity to overlap one of the many AID hotspot motifs ('RGYW'/'WRCY') encoded by IgHD8-2 [54, 87]. This mechanism is therefore possibly linked to AID activity, and alters the positions and numbers of cysteines within the ultralong CDR3H, yet preserves the structurally conserved 'CPDG' and 'YxYxY' motifs (Section 1.2.5.3) [54].

In ultralong antibodies, the ultralong CDR3H shows a higher mutational frequency than the CDR1H or CDR2H [54]. Indeed, the CDR1H and CDR2H of ultralong antibodies show a reduced mutational frequency compared to the CDR1H

and CDR2H of non-ultralong bovine antibodies [54]. This mutational propensity of ultralong CDR3H is possibly indicative of the localisation of antigen recognition to the CDR3H rather than other CDR loops in ultralong antibodies (Section 1.2.5.4), implying a structural conservation of CDR1H and CDR2H.

1.2.5.2 Mutational preference for cysteines in ultralong CDR3H

Another quirk of the IgHD8-2 gene segment, is that >80 % of codons are a single nucleotide change away from introducing a cysteine residue [54]. Disulphides are critical for supporting the ultralong structure (Section 1.2.5.3), as cysteines arise in pairs, generally increasing in number with CDR3H length [54]. A mutational preference towards cysteine, therefore brings an efficient low-energy strategy to greatly alter the ultralong CDR3H structure, and resulting antibody performance.

1.2.5.3 Ultralong structure

Generally the CDR3H loop, formed between β-strands F and G of the immunoglobulin fold is supported by a salt bridge [88]. The 8 bp extension and frameshift introduced by IgHV1-7 replaces this salt bridge with a 'CTTVHQ' motif [54], which elongates the F β-strand into the solvent above the typical antibody surface. Consequently, ultralong CDR3H protrude far above the canonical surface of the antibody; adopting a general β-ribbon stalk-knob-stalk structure that is conserved in all determined ultralong antibody structures by X-Ray crystallographic diffraction (XRD) [78, 82–84] or cryogenic electron microscopy (cryo-EM) [84, 85] (Fig. 1.4). Sequence conservation of the both the ascending and descending stalk 'YxYxY' motifs (where x is any residue), the knob mini-domain initiation motif 'CPDG', and the presence of internal cysteines suggests the stalk-knob-stalk structure (Fig. 1.4) is at least the dominant, and potentially the only structure adopted by bovine ultralong immunoglobulin transcripts [54, 58].

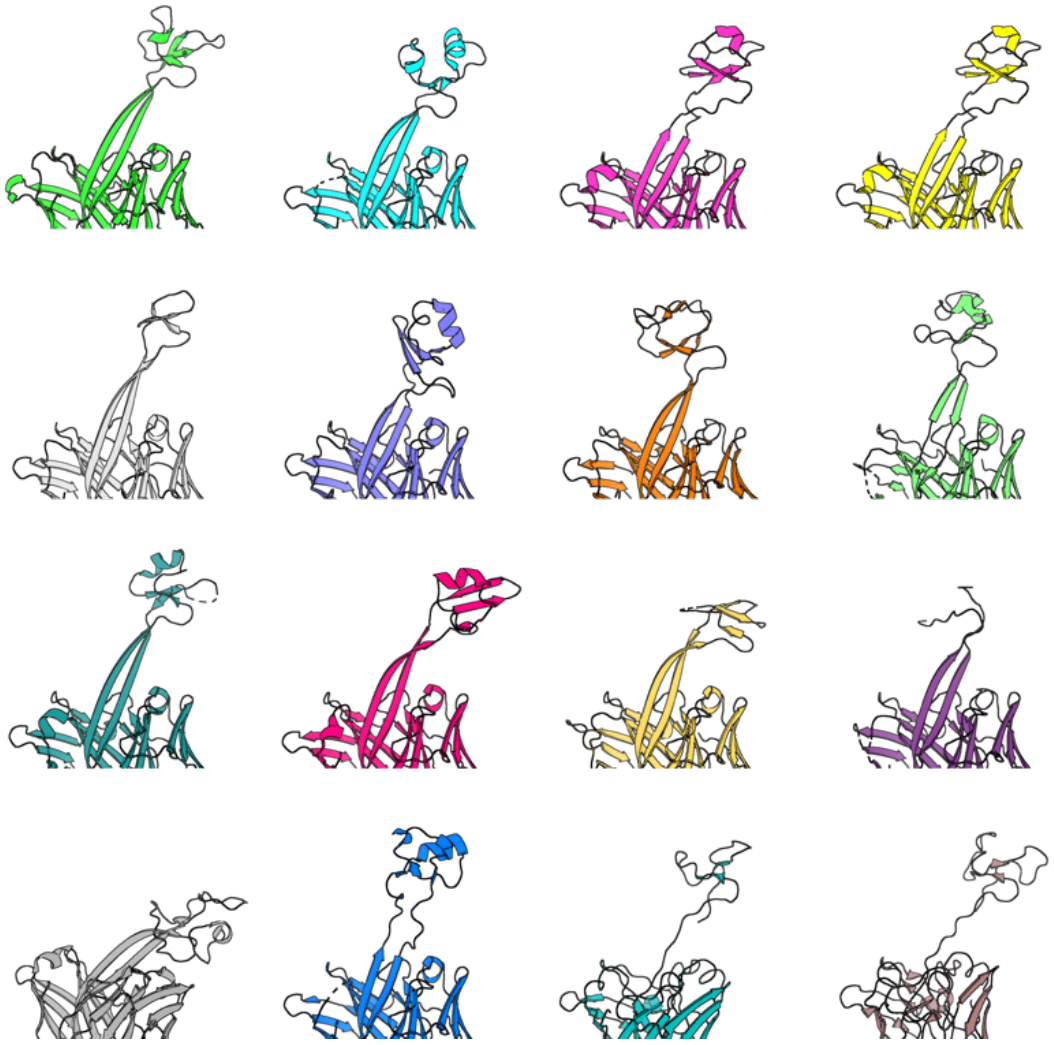


Figure 1.4: The ultralong CDR3H antibody paratope. Cartoon representation of the CDR and surrounding Fv region of ultralong antibodies BLV1H12 (PDB ID: 4K3D) [78], BLV5B8 (PDB ID: 4K3E) [78], F08 (PDB ID: 5E99) [82], B11 (PDB ID: 5IHU) [82], E03 (PDB ID: 5IJV) [82], A01 (PDB ID: 5ILT) [82], BOV-1 through BOV-7 (PDB IDs: 6E8V, 6E9G, 6E9H, 6E9I, 6E9K, 6E9Q, 6E9U) [83], NC-Cow1 (PDB IDs: 6OO0) [84], F145 (PDB ID: 7D3L) [85] and R50 (PDB ID: 7D3M) [85] in grid view, and aligned over the VH domain. Figure prepared using PyMOL.

The knob mini-domain is initiated by a characteristic 'CPDG' turn motif, encoded by the IgHD8-2 [54, 78] gene. The knob is a conserved three-pass anti-parallel β-sheet constrained by internal disulphide linkages between an even number of cysteines [78]. The loops between the knob mini-domain β-sheet strands are incredibly varied in both length and sequence, resulting in the knob mini-domain capable of supporting a diverse surface architecture and chemistry [54]. Thus

the paratope presented by the knob mini-domain is a direct consequence of the knob disulphide arrangement, and therefore may potentially reach an extensive conformational space given the mutational bias towards cysteines (Section 1.2.5.2).

An additional somatic diversification mechanism has been theorised, wherein B cells could reduce the disulphides of an existing ultralong knob mini-domain, reconfigure the cysteines and reestablish the constraining disulphides in an alternative arrangement to generate alternative paratopes from the same genetic sequence [78], but this has not been observed, nor does this seem likely given the additional enzymatic requirement and energetic penalty of knob mini-domain refolding.

The knob mini-domain is unique to bovines in the context of immunoglobulins, but other species demonstrate examples of small peptides and larger proteins whose structures are restrained by a network of internal disulphide bonds. The cysteine knot superfamily of proteins describes proteins containing an intertwined knot of disulphide bridges, usually three or greater in number. The cysteine knot superfamily is subdivided into inhibitor cysteine knots (ICK), growth factor cysteine knots (GFCK), and the eponymous cyclic cysteine knots, also named cyclotides [89]. Understandably, a significant body of work has been conducted on the cysteine knot superfamily, given members include human growth factors, and a variety of human-targeting toxins produced by venomous arthropods including scorpions and spiders.

Bovine ultralong CDR3H knob mini-domains are structurally similar to ICKs, owing to the compact structure with relatively tight inter-cysteine loops and typical β-sheet structure to the ICK core [90–105]. Cyclotides are similar to ICKs in size, however are enzymatically cyclised [106, 107]. By contrast, GFCKs have much larger inter-cysteine loops, such that the cysteine knot stabilises only a portion of the molecule [108–131].

1.2.5.4 Ultralong function

The observation of the ultralong structure revealed a new paradigm in antibody paratope conformational space [132], whilst their diversity indicates their

contribution to an adaptive immune response [78]. Long CDR3H loops have been observed in specific cases, such as antibodies targeting the V1V2 apex of HIV Env; the 'T'-shaped loop of antibody PG16 [133, 134] or the long β-ribbon of antibody VRC26.25 (which uses the 'YxYxY' motif) [135, 136], however these are the exception rather than the rule. Bovine ultralong CDR3H offer a novel paratope system in which extensive conformational diversity may be produced at a region distal to the antibody variable domains.

Separation of the paratope from the main antibody molecule allows surveying of concave epitopes typically inaccessible to conventional, flatter antibody paratopes [31] and therefore regarded as immunoquiescent, such as the recognition of pockets within the BC and GH loops of Foot and Mouth Disease (FMDV) virus viral protein 1 (VP1) by broadly neutralising ultralong antibody R50 [85]. The ultralong antibody paratope could even make available deep or buried epitopes, as exemplified by the recognition of the cluster of differentiation 4 (CD4) binding site of Human immunodeficiency virus 1 (HIV-1) through the glycan shield by ultralong antibody NC-Cow1 [84, 137]. The interaction of R50 and FMDV [85], remains the only complexed structure of a bovine ultralong antibody recognising a bovine-relevant antigen.

In typical, non-ultralong antibodies, interactions with antigen also principally involve the CDR3H, but generally additionally include binding contributions from the other CDR loops [63]. However, in ultralong antibodies, the knob mini-domain alone appears to confer the ability to bind antigen.

Somatic hypermutation targets sequence motifs such as AID hotspots. Hence, the relatively suppressed mutational frequency of CDR1H and CDR2H in ultralong antibodies (encoded by IgHV1-7; Sections 1.2.5 and 1.2.5.1) reflects an absence of such motifs, leading to sequence conservation; possibly in structural support of the ultralong CDR3H. Futhermore, CDR sequence conservation (as opposed to diversification) contributes little to increase paratope diversity, hence implying a more limited direct contribution to antigen recognition by CDR1H and CDR2H than the ultralong CDR3H [54].

Indeed, deletion of the knob domain abolished the activity of ultralong bovine antibody H12 to bovine viral diarrhoea virus (BVDV) [78], and removed activity of NC-Cow1 to HIV-1 Env [132], but did not otherwise impair the stability of the antibody molecule. Interestingly, binding is preserved in NC-Cow1 mutants lacking stalk regions or with highly flexible stalks, however these mutant antibodies suffered stability penalties. The Buchner group also confirmed that the knob mini-domain disulphide landscape is critical for restraining the paratope conformation, with disruption, or substitution to salt bridges resulting in a total loss of affinity to HIV-1 Env [132].

Going further, the knob mini-domain can bind to antigen autonomously, that is in absence of the antibody scaffold, evidenced by proteolytically excised knob mini-domains and human Complement component C5 with comparable binding affinity (but marginally increased kinetics) compared with that of the intact Fab molecule by SPR and FRET experiments [138, 139]. This may not be true in cases where the stalk region stabilises the knob mini-domain, as seen in antibody BOV-7, where the knob mini-domain makes a disulphide linkage with the stalk [83].

Further evidence the knob mini-domain can recognise antigen autonomously can be drawn from the cysteine knot proteins. Inhibitory cysteine knots (ICKs) in particular are structurally similar to ultralong knob mini-domains (Section 1.2.5.3), and are predominantly known ion channel inhibitors [90, 91, 94–99, 103, 104]. There is even a crystallographic structure to 3.0 Å of Psalmotoxin1 from the venom of the tarantula *Psalmopoeus cambridgei* binding to the deep proton-sensitive acidic pocket on the external surface of acid-sensing ion channel-1 from a chicken [101].

The novelty of the ultralong CDR3H fold makes them an attractive target for bio-engineering, with numerous reports of grafting other molecules in place of the knob mini-domain to generate antibodies with specified targeting [140–147], or the inverse situation of grafting an ultralong knob mini-domain onto another protein, to add targeting to a molecule that does not otherwise demonstrate it [148]. In all cases, the knob mini-domain is treated as an autonomous antigen-binding unit.

1.3 Genetics of the bovine light chain locus

The novelty of the ultralong CDR3H fold has attracted the majority of scientific study; indeed no primary literature focusing on the bovine light chain has been produced since 2012 [149], and even this is of limited relevance to this project as it concerns Igλ constant domains. Despite this, the bovine antibody light chain repertoire also demonstrates unique characteristics.

1.3.1 Organisation of the bovine light immunoglobulin locus

Assembly of the bovine light chain locus has proven challenging, due to gene polymorphism, and duplication. The most complete study of the bovine light chain locus was compiled using samples from the Hereford breed [32].

The bovine lambda (λ) locus exists within a single locus on chromosome 17 [32, 150–152]. Twenty-five polymorphic Vλ genes (encoding light chain FR 1 to 3, and CDRs 1, 2 and part of CDR3L) are organised into 3 loci separated by spacers of 126.8 kb and 138.3 kb respectively [32]. The twenty-five Vλ genes can be phylogenetically assigned into 1 of 3 families [61, 153], where family 1 includes seven Vλ pseudogenes, leaving the remainder 18 as functional.

In Holstein-Friesian cattle [152], 31 Vλ genes were assigned to *Bos taurus* autosome 17 (BTA17) (Fig. 1.5), with an additional 32 not assigned to any chromosome (likely either orphans or originating from uncharacterised germline genes), however the authors acknowledge uncertainty in their conclusions given the incomplete genome assembly. Polymorphism of the bovine λ locus may explain the variation between breeds [32].

Four Jλ genes are observed in the bovine λ locus (in both Hereford and Holstein-Friesian cattle), arranged into 4 cassettes each containing one Jλ and one Cλ gene segment [32, 151]. Two of the Cλ genes are psuedogenes, leaving 2 functional Jλ-Cλ gene pairs. The Jλ-Cλ region is 23.7 kb downstream of the Vλ region.

Information on the bovine Vκ locus is only available for Holstein-Friesians, where twenty-four Vκ genes are arranged upstream of a three-member Jκ cassette and a single Cκ gene, located on *Bos taurus* autosome 11 (BTA11) (Fig. 1.6) [152].

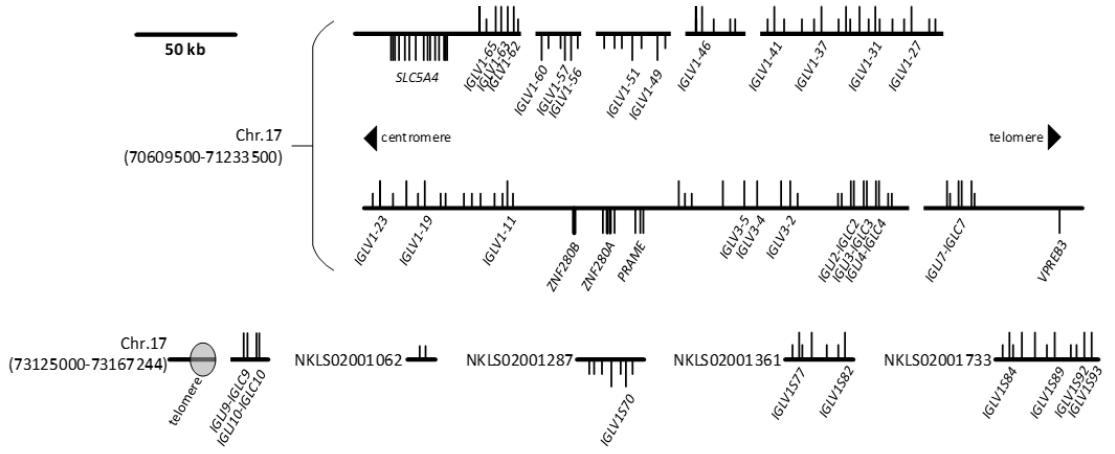


Figure 1.5: Organisation of the cattle antibody Igλ chain loci in the ARS-UCD1.2 genome assembly. The Igλ locus is fragmented into 12 contigs, four of which remain unplaced (accession numbers are indicated). Vertical bars indicate the position of individual gene segments, or Jλ-Cλ cassettes. Long bars indicate putatively functional genes, whilst short bars indicate pseudogenes. Relative gene orientation indicated by bar positioning relative to the horizontal backbone; sense genes are above the backbone, whilst anti-sense are below. Backbone breaks indicate sequence gaps. The non-Ig-like genes *SLC5A4, PRAME, ZNF280A* and *ZNF280B* are labelled to provide genomic context. Figure adapted from John Schwartz (unpublished, personal communication).

Eight of the 24 Vκ genes, and two of the three Jκ are predicted to be functional. Cross-breed transcriptional analysis revealed that the IgKJ1 gene is predominantly rearranged (88.75 - 97.5 %, dependent on breed) over the IgKJ3 gene [154]. The simplistic nature of the κ locus is reflected by the preferential usage (90 - 95 %) of λ-isotype light chains [153, 155, 156].

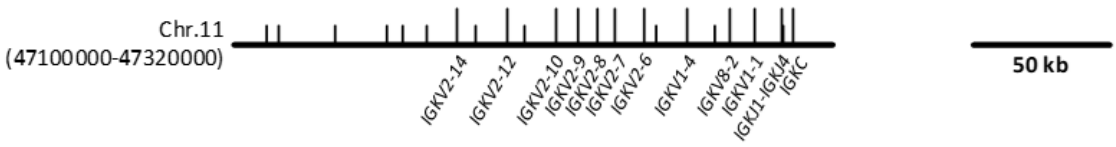


Figure 1.6: Organisation of the cattle antibody Igκ chain loci in the ARS-UCD1.2 genome assembly. The Igκ locus is intact on a single contig, belonging to *Bos taurus* chromosome 11 Vertical bars indicate the position of individual gene segments, or Jλ-Cλ cassettes. Long bars indicate putatively functional genes, whilst short bars indicate pseudogenes. All genes of the Igκ locus were observed in sense direction.below. Figure adapted from John Schwartz (unpublished, personal communication).

1.3.2 Limited genetic diversity

Despite other Vλ gene families in the germline, the IgLV1 gene family dominates (94 % prevalent) bovine λ chains [153]. The IgLV contribution to repertoire diversity is limited, as members of IgLV1 share no less than 91 % nucleotide identity to one another [153]. Furthermore, the Jλ-Cλ junction also contributes little to light chain diversity, as only a single Jλ-Cλ rearrangement (IgLJ4) has been observed as being expressed in cattle [51]. As a consequence, the circulating Igλ repertoire is highly homologous with almost all variation occurring at, or near the CDR [62, 63].

1.3.3 Somatic diversification of bovine light chains

As discussed for the IgH locus (Section 1.2.4), the presence of psuedogenes implies a role for gene conversion at the IgL locus [51, 65]. Indeed the observation of tracts of identical nucleotide mutational patterns [65] may be evidence for gene conversion, but may also be explained by gene polymorphism, contributions from unidentified genes, or introduced somatically as point mutations at mutational hotspots.

As observed with the bovine heavy chains, bovine λ chains are already diversified in foetal spleen samples (collected prior to formation of the ileal Peyer's patches, a site of antibody sequence diversification in bovids) [65, 72], indicating that somatic diversification, in absence of antigen or prior to TdT activity, is important for the establishment of the bovine preimmune repertoire. Maternal antibodies are not transferred across the syndesmochorial bovine placenta [8], hence bovine foetal samples are truly representative of the preimmune repertoire.

1.4 Pairing peculiarities

More recent analyses of human and murine immunoglobulin repertoires have contradicted an earlier view that all antibody heavy and light chain selection was random [157, 158] by demonstrating the existence of preferential pairings between specific VH and VL domains [159]. In total, eight pairings (one in human, and seven in mice) were favoured within 545 human and 1456 mouse sequences,

and all but one pairing were positively preferred, suggesting that optimum VH-VL pairing confers a selective advantage on an antibody, either in terms of stability or antigen recognition capability. Although preferential pairing has been shown to exist, it is relatively rare; only 1.2 % of human antibodies, and 2.7 % of murine antibodies showed significant pairing preferences.

Cattle however, may present a different situation, due to the requirement for structural accommodation of extended and ultralong CDR3H. Two members of the Vλ1 gene family are predominantly expressed over the others [31], however neither of these are observed as a pairing partner with ultralong CDR3H [31]. Instead, a single Vλ1 gene, IgL1-47 [31] (and close homolog IgL1-67 [61]), relatively infrequently expressed in the bovine antibody repertoire [153] exclusively pairs with ultralong heavy chains. Such a preference was observed in adult B cells, mitogen activated polyclonal B cells and foetal B cells, indicating the pairing preference is not the result of antigen-mediated light chain selection.

A tight, compact, and relatively featureless CDR3L may be a structural requirement for longer CDR3H loops. Indeed IgL1-47, and IgL1-67 encoded light chains provide conserved serines and alanines at positions 90, and 91 respectively within the CDR3L [61] (Fig. 1.7) filling a tight hydrophilic pocket between the CDR1L α-helix, and the descending CDR3H β-strand (Fig. 1.8).

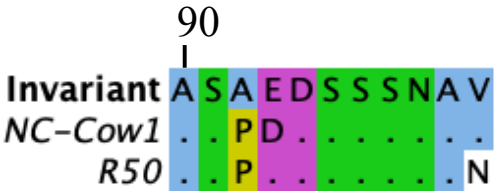


Figure 1.7: Multiple sequence alignment of CDR3L loops paired with ultralong heavy chains. 'Invariant' light chain utilised in ultralong antibodies BLV1H12 (PDB ID: 4K3D) [78], BLV5B8 (PDB ID: 4K3E) [78], A01 (PDB ID: 5ILT) [82], E03 (PDB ID: 5IJV) [82], F08 (PDB ID: 5E99) [82], B11 (PDB ID: 5IHU) [82], BOV-1 through BOV-7 (PDB IDs: 6E8V, 6E9G, 6E9H, 6E9I, 6E9K, 6E9Q, 6E9U) [83] and F145 (PDB ID: 7D3L) [85]. NC-Cow1 (PDB IDs: 6OO0, 6OPA, 6PW6) [84] and R50 (PDB IDs: 7D3M, 7D3R) [85] have light chains also encoded by IgLV1-47, but differ at two residues within the CDR3L. Conserved residues hidden for clarity.

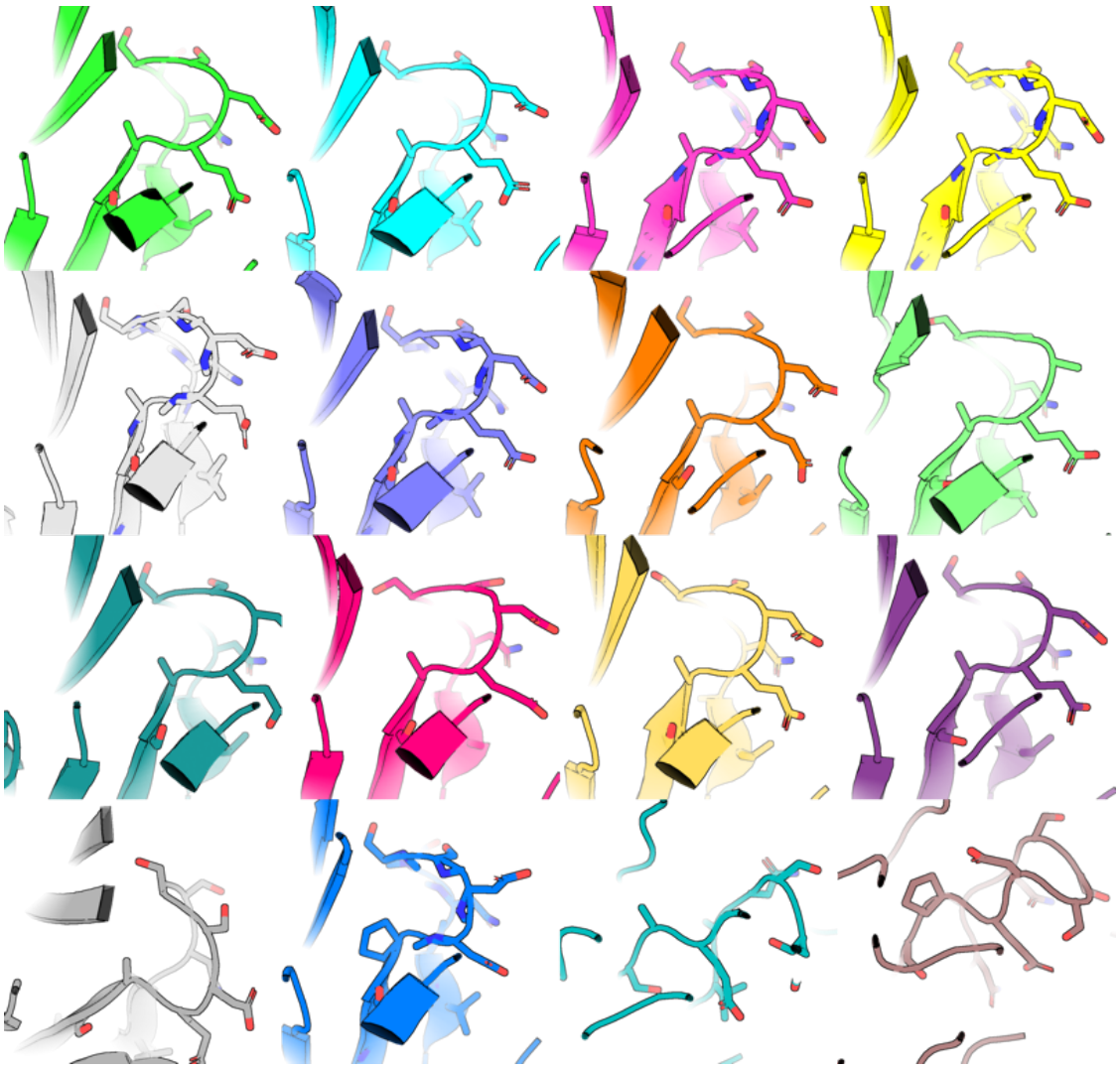


Figure 1.8: Conservation of CDR3L conformation in ultralong antibodies. Cartoon representation of the CDR3L and surrounding Fv region of ultralong antibodies BLV1H12 (PDB ID: 4K3D) [78], BLV5B8 (PDB ID: 4K3E) [78], F08 (PDB ID: 5E99) [82], B11 (PDB ID: 5IHU) [82], E03 (PDB ID: 5IJV) [82], A01 (PDB ID: 5ILT) [82], BOV-1 through BOV-7 (PDB IDs: 6E8V, 6E9G, 6E9H, 6E9I, 6E9K, 6E9Q, 6E9U) [83], NC-Cow1 (PDB IDs: 6OO0) [84], F145 (PDB ID: 7D3L) [85] and R50 (PDB ID: 7D3M) [85] in grid view, and aligned over the VH domain. Ultralong CDR3H can be seen in upper left of each image. Figure prepared using PyMOL.

In reported ultralong structures thus far, the serine hydroxyl is directed toward the centre of the VL domain, whilst the alanine points towards the CDR3H [78, 82, 83] (Fig. 1.8). Interestingly, ultralong antibodies specific for HIV-1 Env had a serine-proline motif within the CDR3L [84], but in all ultralong antibodies, the CDR3L is distinctly perturbed away from the ultralong CDR3H from position 92 onwards, filling the void region between the VH and VL domains. Structural conservation of this CDR3L conformation may explain the exclusive pairing of ultralong antibodies.

As yet, it is unclear whether non-ultralong antibodies demonstrate any preferential pairing behaviour in cattle. However, to test the contribution made by specific bovine light chains, a recent study [160] characterised the structural and functional implications of chain-exchange with two antibodies specific for the F fusion protein of bovine respiratory syncytial virus (bRSV); B4 and B13 [161]. This paper is covered in more depth in Sections 5.1.5 and 5.1.6, but briefly the activity of B4 is abolished when the B4 heavy chain is paired with the B13 light chain, explained by a stark reconfiguration of the CDR3H loop. Whilst not an ultralong antibody, this indicates that bovine light chain selection does appear to contribute to antibody function, even if indirectly by providing support to the CDR3H loop, though direct observation of the antibody-antigen interaction is lacking to exclude the possibility of light chain directly interacting with bRSV F protein (bRSVF). Chapter 5 describes initial work to characterise these interactions with bRSVF, whilst Chapter 3 and Chapter 4 will consider chain-exchange in the context of ultralong antibodies, first on antigen-unknown ultralong antibodies, and second on ultralong antibodies specific for Foot and mouth disease virus (FMDV).

1.5 A summary of bovine structural immunology

Relatively little is known about the bovine immune system from a structural perspective. To date, characterisation of the ultralong CDR3H has been prioritised, with 16 ultralong antibody structures (PDB IDs: 4K3D, 4K3E) [78], (PDB IDs: 5IHU, 5IJV, 5ILT) [82], (PDB IDs: 6E8V, 6E9G, 6E9H, 6E9I, 6E9K, 6E9Q, 6E9U) [83], (PDB IDs: 6OO0, 6OPA, 6PW6) [84], (PDB IDs: 7D3L, 7D3M, 7D3R) [85],

and two additional structures of isolated knob mini-domains (PDB IDs: 7AD6, 7AD7) [139] deposited in the Protein Data Bank (PDB)[162] thus far. Otherwise, the PDB contains structural information on the bRSVF-specific antibodies and their chain-exchanged progeny discussed previously (Section 1.4) (PDB IDs: 6QN7, 6QN8, 6QN9, 6QNA) [160], two bovine major histocompatibility complexes (PDB IDs: 2XFX, 3PWU, 3PWV) [163, 164] and four cell surface signalling molecules; cluster of differentiation (CD)1d (PDB IDs: 4F7C, 4F7E) [165], CD1b3 (PDB ID: 3L9R) [166], CD8aa (PDB ID: 5EBG) [167] and CD152 (also known as CTLA-4) (PDB: 4KKN) [168]. There is insufficient structural evidence to explain the observations of preferential usage of IgHV1-10 in short and medium CDR3H length bovine antibodies from a structural perspective (Section 1.2.2).

1.6 A curious immunological paradigm

The core principles of immunological understanding are mainly based on examinations of the human and murine systems [15–24, 27]. However, analyses of other species has revealed that these principles are not universal.

In humans and mice, diversity is principally achieved through V(D)J recombinatorial diversity, and further diversified by somatic hypermutation upon contact with antigen (Section 1.1). Junctional flexibility also significantly contributes to CDR variation, as it changes the CDR3 length, and can introduce frameshift mutations. Such recombinatorial diversity is unavailable to species with restricted germline repertoires of functional V, D, and J gene segments. For example, rabbits and chickens diversify their antibody repertoires through donation of genetic material from psuedogenes to functional gene segments through gene conversion [42, 43, 67–70].

Cattle demonstrate an extreme immunological position, wherein there exists a limited scope for V(D)J recombinatorial diversification through restricted libraries of functional IgH, Igλ and Igκ gene segments (Sections 1.2.1, 1.2.2 and 1.3.1), that is further hampered by sequence invariance between different genes (Sections 1.2.3 and 1.3.2), and observed preferential chain pairing activity (Section 1.4). To break

through these constraints on diverse immunoglobulin library generation, novel germline and somatic solutions have evolved in modern cattle.

Fusion of two IgHD genes, and the 8 bp extension of IgHV1-7, provided cattle with the means to generate ultralong CDR3H loops exceeding 48 residues (Section 1.2.5). Ultralong antibodies interact with antigen in a manner not previously demonstrated in the global antibody paratopic space (Section 1.2.5.4); they therefore offer perspective on an antigenic recognition system unique to bovids, and harbour novel therapeutic and diagnostic opportunities. By consequence of their structure (Section 1.2.5.3), ultralong antibodies may be able to access epitopes immunoquiescent to typical antibodies with shorter CDR loops. The diversity introduced via ultralong CDR3H loops is expanded further by the capacity for highly varied surface sequences (Section 1.2.5.3), a high proclivity for, and frequent observation of somatic mutation over the majority of the CDR3H sequence (Sections 1.2.5.1 and 1.2.5.2), and a novel AID-mediated codon deletion mechanism to alter the structure (Section 1.2.5.1).

Diversification of bovine CDR sequences proceeds in foetal tissue (Section 1.2.4), indicating a reliance on somatic hypermutation to establish a diverse preimmune repertoire. The presence of psuedogenes implies the activity of gene conversion (Sections 1.2.2 and 1.3.1), but conclusions must be interpreted with caution owing to incomplete assembly of the respective genome loci, and the polymorphic nature of gene segments. By contrast to the heavy chain CDR3H, bovine light chain sequences are highly invariant (Section 1.3.2). It has therefore been suggested, that bovine light chains contribute relatively little to antigenic recognition, potentially providing a structural role in supporting the extended CDR of bovine heavy chains [61, 153]. By extension, this argument suggests that most of the bovine immune response resides within the heavy chain contribution.

On the other hand, the light chain appears to modify the conformation of the CDR3H loop sufficiently to eliminate antigen binding in non-ultralong antibodies (Section 1.4), implying a non-trivial role for light chains in the bovine immune system. Hence, despite the VL sequence similarity (Section 1.3.2), there appears to

exist an element of selectivity in the pairing of bovine ultralong heavy and light chains, that is not yet fully understood. Structural information about the bovine immune system is sparse (Section 1.5), so presently only limited inferences can be made about the structural contributions of each chain.

Advancing the current understanding in this space would therefore provide an explanation for the restricted chain pairings of ultralong antibodies, and may provide the ability to modify the function of said antibodies in engineering efforts for use in therapy or diagnostics. Equally, improved understanding of the bovine adaptive immune response would lead to higher quality veterinary care of cattle; allowing for more efficient design of prophylaxis and reactive treatments against infection, and potentially to advise on herd maintenance routines, with impacts on animal welfare and food security.

1.7 Project aims and objectives

This project will examine bovine immunoglobulin heavy and light chain pairing at atomic resolution, and the implications of pairing on antigenic recognition. The overarching aim of this project is to better understand the structural basis of heavy and light chain heterodimeric pairing in bovine antibodies, in order to define the role of the highly invariant light chain, particularly when paired with ultralong heavy chains, by achieving the following objectives:

1. To observe the influence of chain-exchange on ultralong CDR3H conformation.
2. To measure functional implications of chain-exchange within ultralong antibodies.
3. To structurally resolve the interface formed by bovine antibodies (ultralong and otherwise) recognising a bovine-relevant antigen.
4. To contribute towards structural elucidation of spatial constraints that determine the unique architecture of cattle antibodies.
5. To investigate how such constraints impact the antigen-binding specificity of cattle antibodies.

If you want to find the secrets of the universe, think in terms of energy, frequency, and vibration.

— Nikola Tesla, inventor and engineer

2 Methodology

Contents

2.1 Core reagents 32
2.1.1 Antibiotics 32
2.1.2 Antibodies 32
2.1.3 Cell lines 32
2.1.4 Vectors 33
2.1.5 Viruses 36
2.2 Laboratory Workflows 37
2.2.1 Fab molecular biology workflow 37
2.2.2 Plasmid propagation workflow 38
2.2.3 Fab expression screening 38
2.2.4 Fab expression 38
2.2.5 bRSV F protein production 38
2.2.6 VLP production 39
2.3 Microbiology and cell culture techniques 39
2.3.1 Vector linearisation 39
2.3.2 In-Fusion ligation independent cloning 40
2.3.3 Preparation of LB media 41
2.3.4 Preparation of LB 2 % agar 41
2.3.5 Transformation of chemically competent *E. coli* cells . . 41
2.3.6 Glycerol stocks of transformed *E. coli* 41
2.3.7 Creation of a stable Expi293F™ cell lineage 42
2.3.8 Freezing down of stable cell lineages 43
2.4 Molecular biology techniques 43
2.4.1 Purification of plasmid DNA 43
2.4.2 Polymerase Chain Reaction screening 44
2.4.3 Agarose gel electrophoresis 45
2.4.4 Gel purification of DNA 45
2.4.5 Plasmid sequence verification 46

2.5 Protein expression 46
2.5.1 Transfection of Expi293F™ cells for Fab expression . . 46
2.5.2 Stable Expi293F™ cell line expression of bRSVF . . . 47
2.5.3 Dual infection of HEK293T cells for VLP production . . 47
2.6 Protein purification and concentration 48
2.6.1 Immobilised affinity gel filtration of Fabs 48
2.6.2 Immobilised affinity gel Filtration of bRSVF 49
2.6.3 Chloroform extraction of VLPs 49
2.6.4 Sucrose density gradient (SDG) purification of VLPs . . 50
2.6.5 Concentrating proteins by centrifugation 50
2.6.6 Protein G re-purification 51
2.6.7 Protein desalting 52
2.7 Protein characterisation 52
2.7.1 Inference of protein molecular weight by SEC 52
2.7.2 SDS polyacrylamide gel electrophoresis (PAGE) 53
2.7.3 SDS-PAGE staining 54
2.7.4 Western blotting 54
2.7.5 Protein concentration quantification 55
2.7.6 Qubit™ quantification of VLPs 56
2.7.7 VLP Integrity assessment by VHH double sandwich ELISA 56
2.7.8 Size exclusion chromatography with multi-angle laser light scattering 59
2.7.9 Prediction of protein glycosylation sites 60
2.7.10 Analytical Ultracentrifugation 61
2.7.11 Bio-layer Interferometry 62
2.8 Structural biology methods 64
2.8.1 ABodyBuilder structural prediction 64
2.8.2 AlphaFold structural prediction 64
2.8.3 Crystallographic X-Ray Diffraction 65
2.8.4 Single-particle cryogenic Electron Microscopy 74

Figures

2.1 pOPIN vector maps 35
2.2 pOPINTTGNeo vector map 36
2.3 VHH ELISA experimental schematic 57
2.4 Chameleon grid vitrification robot 78

Tables

2.1 Antibodies used in this project 32
2.2 Recombinant vaccinia viruses used for VLP production 37
2.3 Restriction enzymatic activity for vector double digestion . . . 40
2.4 Vector primers used in this project 44
2.5 Polymerase Chain Reaction Protocol 45
2.6 VHH ELISA Capture antibodies 57
2.7 VHH ELISA Detection antibodies 58
2.8 Biolayer Interferometry Experimental Schematic 63
2.9 List of accessed beamlines 67
2.10 X-Ray data collection parameters 70
2.11 Range of accepted R_{free} values for crystallographic XRD processing 73
2.12 Per-gridtype glow discharge and sample application parameters 76
2.13 List of accessed cryomicroscopes 79

2.1 Core reagents

2.1.1 Antibiotics

Carbenicillin (Melford) was used at an active concentration of 50 µg/mL to select for *Escherichia coli* successfully transformed with pOPIN vectors used in this study. Geneticin™ Selective Antibiotic (G418 Sulphate) (Gibco®, ThermoFisher Scientific) was used to select for Expi293F™ cells (Gibco®, ThermoFisher Scientific) successfully transformed with pOPINTTGNeo vectors for the creation of a stabilised cell lineage for the expression of bRSVF (Chapter 5).

2.1.2 Antibodies

Several antibodies were used in this project (Table 2.1).

Table 2.1: Antibodies used in this project.

Name	His Tag Antibody	Anti-Mouse IgG Peroxidase
Epitope	Polyhistidine	Mouse IgG
Conjugation	-	HRP
Species	Mouse	Rabbit
Clone	AD1.1.10	*poly*
Manufacturer	R&DSystems	Sigma-Aldrich
Usage	Western Blot	Western Blot
Dilution	1:2,500	1:100,000

2.1.3 Cell lines

2.1.3.1 Chemically competent *Escherichia coli*

Stellar chemically competent cells (TakaraBio) are an *Escherichia coli* HST08 strain which provide high transformation efficiency, and allow for blue-white screening when transformed with vectors containing the *lacZα* gene. Stellar cells were used for all preparations of nucleic acids.

2.1.3.2 Mammalian cell cultures

Expi293F™ cells (Gibco®, ThermoFisher Scientific) are derived from the 293F cell line; a permanent line derived from primary human embryonic kidney (HEK) [169].

Expi293F™ grow to high density in suspension cultures, are highly transfectable, and produce strong protein yields. Expi293F™ are maintained as suspension culture within Expi293™ Expression Medium (Gibco®, ThermoFisher Scientific) at 37 °C in 5 % CO_2. Expi293F™ cells were used for protein production by transient transfection, and in the creation of a stable cell lineage for the expression of stabilised prefusion conformation F protein from bovine respiratory syncytial virus.

HEK293T cells (kindly gifted by the Structural Virology group within the University of Oxford's Division of Structural Biology (STRUBI)) are a robust and low-maintenance culture also derived from immortalised HEK cells. HEK293T are maintained as adherent culture within Dulbecco's Modified Eagle Medium (DMEM) - High Glucose, 25 mM HEPES (Merck) supplemented with 10 % (v/v) filter sterilised foetal calf serum (FCS) (The Pirbright Institute (Pirbright) Stores), 2 mM L-Glutamine (Gibco™) and 50 U/mL Penicillin-Streptomycin (Gibco™) split across 8 175 cm^2 CellStar® Cell Culture Flasks, 550 mL, PS (Grenier Bio-One) at 37 °C in 5 % CO_2. DMEM is supplemented with 25 mM HEPES to provide a source of CO_2 when filter top lids are incompatible for biocontainment of dually infected cells during virus-like particle (VLP) production.

2.1.4 Vectors

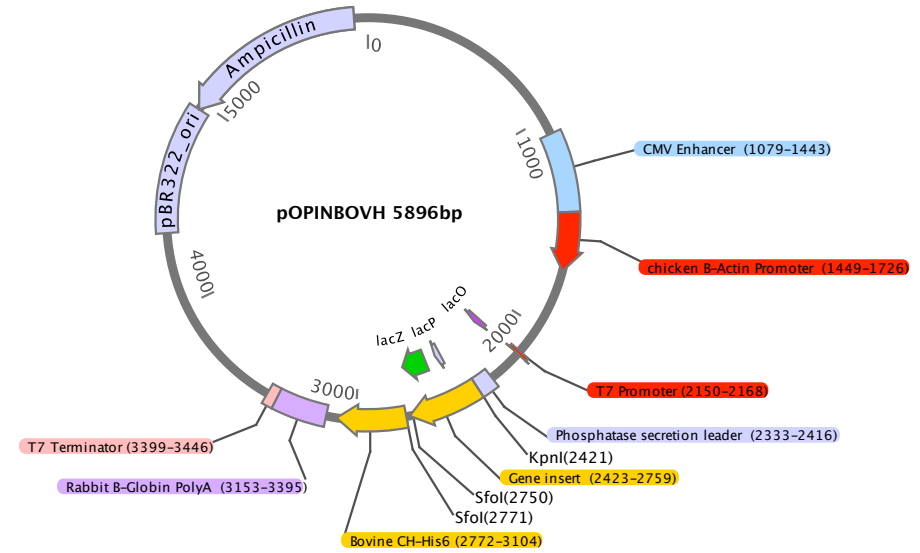
The pOPIN suite of vectors, developed by the Oxford Protein Production Facility (OPPF) [170, 171], permits protein expression in a range of cellular systems, and typically includes relevant intracellular trafficking signalling peptides and protein purification tags [171].

Bovine antibody heavy chain variable domains (VH), and bovine antibody light chain variable domains (VL) were cloned into vectors pOPINBOVH and pOPINBOVL respectively [160]. The cloning recipient site is located between an upstream phosphatase secretory leader sequence, and the desired downstream bovine immunoglobulin constant domain (CH/CL) (Figs. 2.1a and 2.1b). The lac operon is provided in both vectors to permit white/blue colony screening for identifying correctly digested and recombinant vectors, with *lacP* and *lacZα* located in the

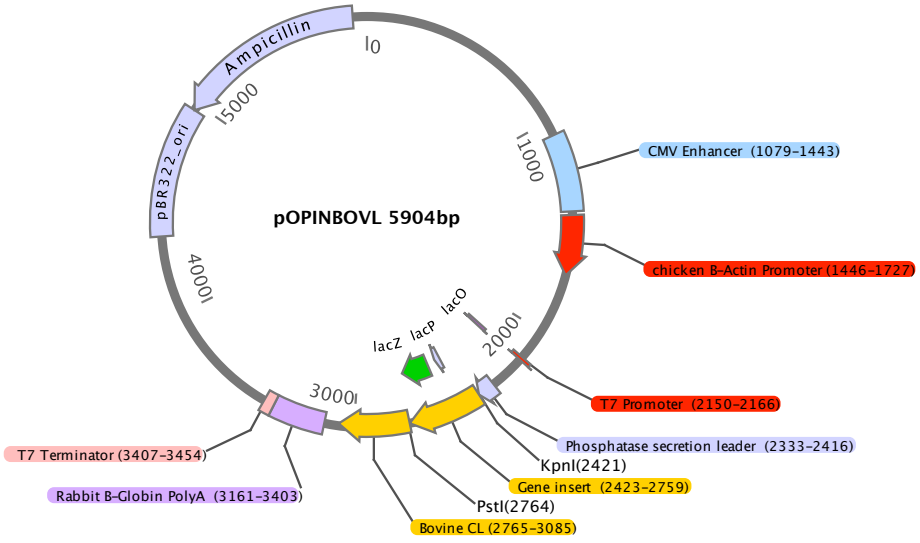
vector backbone between the restriction sites. Additionally, pOPINBOVH includes a C-terminal hexahistidine tag for purification and immunoblot identification of correctly assembled Fab (Fig. 2.1a). Expression in both pOPINBOVH and pOPINBOVL is primarily driven by the T7 promoter and terminator system, with the cytomegalovirus (CMV) early enhancer and chicken β-actin promoter system for long term expression efficiency. Finally, pOPINBOVH and pOPINBOVL include an ampicillin resistance gene for bacterial selection, and the origin of replication from pBR322 [172, 173].

For the creation of a stable lineage, a selection resistant gene is required, provided by pOPINTTGNeo vectors [174], which include a Geneticin/Neomycin resistance gene under expression control by the SV40 promoter system (Fig. 2.2). Similar to pOPINBOVH (Fig. 2.1a), pOPINTTGNeo encodes an N-terminal secretory(tripartite) leader sequence, and a C-terminal hexahistidine purification tag (Fig. 2.2a). Other similarities to pOPINBOVH and pOPINBOVL include the CMV promoter system, excision of *lacZ*α and *lacP* for white/blue colony screening, ampicillin resistance and pBR322 origin of replication [172, 173]. Finally, pOPINTTGNeo vectors include an origin of replication from Epstein-Barr virus, oriP, which establishes pOPINTTGNeo as a replicon once transfected into certain mammalian cell lines.

Stabilised prefusion conformation bovine respiratory syncytial virus F protein (bRVSF) [175](Chapter 5) was generously provided by Peter Kwong of the National Institutes of Health, within the pVRC8400 vector.

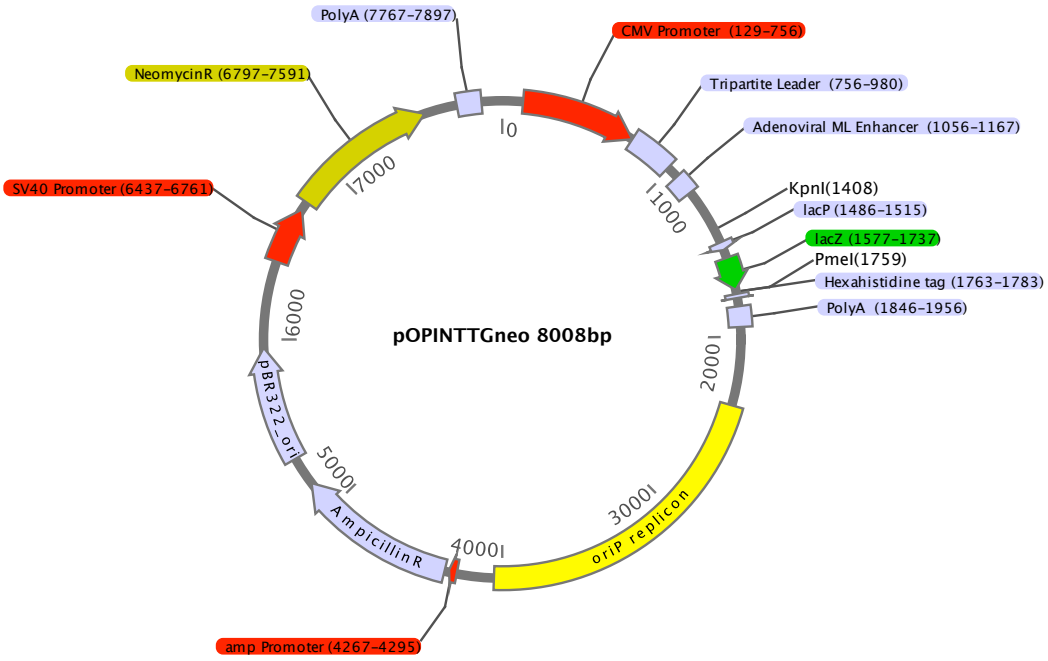


(a)



(b)

Figure 2.1: Schematic representation of the pOPIN vectors showing key components of the plasmid design and In-Fusion cloning system. Fig. 2.1a) pOPINBOVH, Fig. 2.1b) pOPINBOVL. Prepared using DNADynamo v1.569 (Blue Tractor Software Ltd)



(a)

Figure 2.2: Schematic representation of the pOPINTTGNeo vector showing key components of the plasmid design and In-Fusion cloning system. Prepared using DNADynamo v1.569 (Blue Tractor Software Ltd)

2.1.5 Viruses

FMD VLPs were produced by dual infection of HEK293T cells with recombinant vaccinia virus (VV), negating the need for high containment, and easily allowing production of FMD VLPs of different serotypes.

Recombinant VV were produced by, and kindly donated by the Structural Virology group within STRUBI, and the Viral Immunology group within The Pirbright Institute [176]. In brief, the P1-2A-3C FMDV cassette encoding FMDV viral proteins 1-4 (VP1-4) as a polyprotein, a reduced-activity FMDV 3C protease [177], and an FMDV internal ribosome entry segment (IRES), is cloned into a pBG200 vector, and transfected into CV-1 cells infected with VV strain WR. Recombinant VV were selected using HuTK-143 cells, and purified with 3 rounds of plaque purification [176]. Expression of the P1-2A-3C cassette is dependent on a T7 polymerase promoter. T7 polymerase is therefore provided by co-infection with recombinant VV vTF7.3. The sequence of the P1-2A cassette can easily be

altered to produce other FMDV serotypes, or introduce stability mutants such as the H2093C stability mutation of VP2, increasing inter-pentamer stability of the empty FMDV A22 capsid [176]. P1-2A-3C cassettes used in this project are summarised in Table 2.2.

Table 2.2: Recombinant vaccinia viruses carrying engineered P1-2A-3C cassettes used for the production of foot-and-mouth disease virus-like-particles used in this project. †PFU; plaque forming units.

Serotype	Topotype	Viral Titre (PFU†/mL)	Stability Mutant
A	22/Iraq/24/64	2.3×10^{11}	H2093C
Asia	1/Shamir	8.0×10^{10}	S2093C
O	1/Manisa/Turkey/69	3.0×10^{11}	S2093C
SAT	2/Zimbabwe/7/83	2.0×10^{10}	S2093Y

2.2 Laboratory Workflows

For clarity, this section will provide several core experimental workflows used within this project. For details on each step of the workflows, please see each respective section.

2.2.1 Fab molecular biology workflow

Candidate gene sequences were cloned into pOPINBOVH or pOPINBOVL vectors by In-Fusion cloning and subsequent transformation of Stellar cells (Section 2.1.3.1). Where appropriate, vectors were linearised (Section 2.3.1), gene sequences were codon optimised and inserted into the freshly linearised vectors (Section 2.3.2). Immediately following In-Fusion cloning, recombinant vectors were transformed into Stellar cells (Section 2.3.5), grown on LB 2 % agar (Section 2.3.4) and LB (Section 2.3.3). Plasmids were isolated by DNA Miniprep (Section 2.4.1). Insert length was verified using PCR Screening (Section 2.4.2) on an 0.8 % agarose gel (Section 2.4.3). If the inserts migrated as expected, samples were sequenced (Section 2.4.5).

2.2.2 Plasmid propagation workflow

Transient transfection requires a large quantity of highly pure DNA for effective expression. Accurately cloned constructs were again transformed into Stellar cells (Section 2.1.3.1 and Section 2.3.5), grown on LB 2 % agar (Section 2.3.4) and LB (Section 2.3.3). Glycerol stocks of transformed Stellar cells containing accurately cloned constructs were prepared (Section 2.3.6). Plasmids were isolated by DNA Midiprep (Section 2.4.1).

2.2.3 Fab expression screening

Expression of Fabs was screened using Expi293F™ cells (Section 2.1.3.2). Purified constructs for each chain were co-transfected into Expi293F™ cells (Section 2.5.1) in 1 mL cultures. Samples were separated by SDS-PAGE (Section 2.7.2), and expression level was assayed by Western blotting (Section 2.7.4).

2.2.4 Fab expression

Fabs were expressed using Expi293F™ cells (Section 2.1.3.2). Purified constructs for each chain were co-transfected into Expi293F™ cells (Section 2.5.1) in 100 mL, 150 mL, 200 mL, or 300 mL cultures. Fabs were purified by affinity chromatography (Section 2.6.1). Eluted fractions were separated by SDS-PAGE (Section 2.7.2), and fractions of interest identified by SDS instant staining (Section 2.7.3). Fractions containing Fab were pooled and spin-concentrated (Section 2.6.5). Fabs being used for BLI were additionally re-purified by Protein G (Section 2.6.6) and subsequent desalting (Section 2.6.7).

2.2.5 bRSV F protein production

To address bRSVF production issues, a stable cell lineage was established using Expi293F™ cells (Section 2.1.3.2). A stable cell lineage was created (Section 2.3.7) with aliquots flash frozen for later use (Section 2.3.8). Aliquots were thawed for expression (Section 2.5.2), with thawed cultures maintained following the same culture routine as the parental Expi293F™ cells (Section 2.1.3.2). Frozen cell

lineage stocks were replenished by flash freezing of spare culture (Section 2.3.8). bRSVF was purified by affinity chromatography (Section 2.6.2). Eluted fractions were separated by SDS-PAGE (Section 2.7.2), and fractions of interest identified by SDS instant staining (Section 2.7.3). Fractions containing bRSVF were pooled and spin-concentrated (Section 2.6.5).

2.2.6 VLP production

VLPs were produced by dual-infection of HEK293T cells (Section 2.1.3.2). HEK293T were dually-infected with recombinant vaccinia viruses (Section 2.1.5 and Section 2.5.3). VLPs were extracted using chloroform (Section 2.6.3) and fractionated by sucrose density gradient centrifugation (Section 2.6.4). Eluted fractions were separated by SDS-PAGE (Section 2.7.2), and fractions of interest identified by SDS instant staining (Section 2.7.3). VLPs were quantified using Qubit™ (Section 2.7.6), and integrity assessed by VHH ELISA (Section 2.7.7). Immediately prior to use, VLPs were spin-concentrated (Section 2.6.5).

2.3 Microbiology and cell culture techniques

2.3.1 Vector linearisation

In-Fusion cloning relies upon the annealing of complementary overhangs of a cloning insert and a linearised vector, ensuring seamless single step directional cloning, with complete freedom over the locus of insertion, insert and vector sequences.

Target vectors were linearised by restriction double digestion (using KpnI (New England BioLabs) and SfoI (New England BioLabs) for pOPINBOVH, and KpnI (New England BioLabs) and PstI (New England BioLabs) for pOPINBOVL). Optimised reaction conditions for double digests were calculated using the online NEBCloner® v1.12.0 Restriction Enzyme Double Digestion Tool (New England BioLabs) [178], with buffers selected for maximal activity of each enzyme. Enzyme activity for each double digest reaction is provided in Table 2.3. Linearisation of pOPINBOVH with KpnI and SfoI was performed in NEBuffer 2.1 (New England BioLabs), whilst linearisation of pOPINBOVL with KpnI and PstI was performed

in NEBuffer 1.1 (New England BioLabs). All double digestion reactions were performed for 1 hour at 37 °C, in volumes of 25 µL.

Table 2.3: New England BioLabs Restriction enzyme activity in varying buffers for vector double digestion.

		Activity (%)			
Vector	Enzyme	NEBuffer 1.1	NEBuffer 2.1	NEBuffer 3.1	CutSmart
VH	KpnI	100	75	10	50
VH	SfoI	50	100	100	100
VL	KpnI	100	75	10	50
VL	PstI	75	75	100	50

2.3.2 In-Fusion ligation independent cloning

Antibody sequences were discovered by The Pirbright Institute ('Pirbright') (Chapter 3)[179], and in (Chapter 4) (Ramirez Valdez, in preparation). These antibody variable domain sequences are provided in Table A.1.

The nucleotide sequences of all proteins were codon-optimised for expression in human cells using Integrated DNA Technologies (IDT)'s online Codon Optimization Tool, flanked by overhangs complementary to either pOPINBOVH or pOPINBOVL, and ordered as gBlocks™ from IDT. Codon optimised antibody variable domain sequences are described in Table A.1. The 500 ng freeze-dried double-stranded DNA fragments were re-suspended in 25 µL nuclease-free ddH_2O to achieve 20 ng/µL. 20 ng of re-suspended gBlocks™ DNA was mixed with 10 ng of the linearised vector, 1 µL of ClonExpress II Buffer (Vazyme), 0.5 µL of Exnase II (Vazyme), and water to a total volume of 5 µL. The In-Fusion reaction mixture was incubated at 42 °C for 30 minutes, followed by 4 °C indefinitely using a Veriti™ 96-Well Fast Thermal Cycler (Applied Biosystems). Subsequently, the In-Fusion products were diluted into 25 µL of Tris-EDTA (TE) buffer, which solubilises the DNA and sequesters divalent metal ions such as Mg^{2+}, terminating the In-Fusion reaction.

2.3.3 Preparation of LB media

To prepare Luria broth (LB) medium, either 25 g/L granulated Luria Broth (Miller's Composition) (Melford), or 10 g/L Tryptone (Melford), 5 g/L Yeast extract (Melford) and 10 g/L Sodium Chloride (NaCl) (Scientific Laboratory Supplies) are added to ddH_2O. Media were sterilised by autoclave following a 121 °C, 15 minute programme.

2.3.4 Preparation of LB 2 % agar

Agar (Fisher Chemical™, ThermoFisher Scientific)was added to LB medium to 20 g/L prior to autoclave sterilisation of 121 °C for 15 minutes. To prepare LB agar plates, the agar was microwaved until molten. Once the agar had cooled to hand hot (roughly 50 °C), appropriate antibiotics were added (Section 2.1.1), before pouring into the required plates.

2.3.5 Transformation of chemically competent *E. coli* cells

2.5 µL In-Fusion product, or miniprep-purified plasmid DNA, was added to a tube, followed by addition of 10 µL of Stellar chemically competent cells (TakaraBio). The mixture was incubated on ice for 30 minutes. The cells were then placed in a 42 °C water bath for 45 seconds, to facilitate genetic uptake by heat shock, and immediately returned to ice for a further 2 minutes. 200 µL of sterile LB was added to each tube, and incubated at 37 °C for 1 hour to permit recovery. Transformed cells were then plated onto sterile LB 2 % agar plates supplemented with appropriate antibiotic selection, and incubated overnight at 37 °C. For initial cloning experiments, LB agar was additionally supplemented 1:1000 (v/v) with X-Gal (Cayman Chemical Company) and IPTG (Sigma-Alrdich, Merck) for blue/white colony screening.

2.3.6 Glycerol stocks of transformed *E. coli*

Individual colonies were grown overnight in LB with appropriate antibiotic selection. 900 µL of culture were transferred to a cryovial. Sterilised glycerol was added to a final concentration of 25 % before storage at −80 °C.

2.3.7 Creation of a stable Expi293F™ cell lineage

When transient transfection failed to yield sufficient stabilised bRSV F protein for structural study (Section 5.3.1), a stable cell lineage was prepared to secure the production of F protein. Approximately 1×10^7 Expi293™ cells were introduced to Dulbecco's Modified Eagle Medium (DMEM), high glucose, GlutaMAX™ Supplement, pyruvate (Gibco®, ThermoFisher Scientific) supplemented with 10 % (v/v) Qualified Foetal Bovine Serum (Brazil)(Gibco®, ThermoFisher Scientific) to a total volume of 50 mL within a 75 cm^2 Cellstar® Tissue Culture treated flask with filter lid (Greiner Bio-One), and allowed to adapt to the new medium for 3 days at 37 °C with 5 % CO_2. Cells were trypsinised using Trypsin-EDTA solution (0.25 %) (Merck) and roughly diluted 1:5 in DMEM supplemented with FCS to 10 % (v/v). 2×10^5 cells/mL Expi293™ cells were seeded in a 75 cm^2 Cellstar® Tissue Culture treated flask with filter lid (Greiner Bio-One), and incubated overnight at 37 °C with 5 % CO_2. 25 µg of QIAprep® Plasmid Plus Midi-prep (Qiagen) purified DNA and 75 µL GeneJuice® Transfection Reagent (Novagen) were diluted into 50 mL Opti-MEM™ Reduced Serum Medium (Gibco®, ThermoFisher Scientific), before being applied to cells. Cells were returned to the 37 °C static incubator, and incubated for 24 hours with 5 % CO_2. Following trypsinisation, cells were serially diluted to 1:10, 1:100, 1:1000 and 1:10000 in DMEM supplemented with 10 % (v/v) FBS, and 100 µL of diluted cells were seeded within each well of a Cellstar® 96 well Tissue Culture treated plate (Greiner Bio-One) for each dilution titre, aiming to achieve a single cell per well. After 24 hours incubation at 37 °C with 5 % CO_2, Geneticin™ Selective Antibiotic (G418 Sulphate) (Gibco®, ThermoFisher Scientific) was applied to a concentration of 1.2 mg/mL to select for transformed cells. Cells were incubated under selection at 37 °C with 5 % CO_2 for 1 week. Thereafter, colony growth was monitored, with cell medium being replenished at weekly intervals. For each well, once confluence was achieved, cells were trypsinised and transferred to an incrementally larger vessel to expand the stabilised lineage volume; first as a 750 µL culture within a Cellstar® 24 well Suspension Culture treated plate (Greiner Bio-One), then as a 17 mL culture within a 25 cm^2 Cellstar®

Tissue Culture treated flask with filter lid (Greiner Bio-One), and finally as a 45 mL culture within a 75 cm^2 Cellstar® Tissue Culture treated flask with filter lid (Greiner Bio-One). 100 µL aliquots were collected prior to expansion to 17 mL cultures to test for protein expression by Western Blot. All cultures were maintained with DMEM supplemented with FBS to 10 % (v/v), and grown at 37 °C with 5 % CO_2 throughout.

2.3.8 Freezing down of stable cell lineages

Stable cell lineages were trypsinised using Trypsin-EDTA solution (0.25 %) (Merck), and mixed with 30 mL Dulbecco's Modified Eagle Medium (DMEM), high glucose, GlutaMAX™ Supplement, pyruvate (Gibco®, ThermoFisher Scientific) supplemented with 10 % (v/v) Qualified Foetal Bovine Serum (Brazil)(Gibco®, ThermoFisher Scientific). Cells were harvested by centrifugation of 1000 RCF for 5 minutes. The cell pellet was re-suspended in 10 mL Bambanker (Nippon Genetics Europe) and aliquoted to 1.8 mL within 2 mL cryophials. Controlled cell freezing was achieved by using a Nalgene® Mr. Frosty® Freezing Container (ThermoFisher Scientific) filled with 500 mL Isopropyl Alcohol (Fisher Chemical™, ThermoFisher Scientific) to achieve a uniform cooling rate of −1 °C/minute until reaching −80 °C. After 4 hours, to ensure freezing completion, cryovials were transferred to a liquid nitrogen vapour phase storage dewar for long term storage.

2.4 Molecular biology techniques

2.4.1 Purification of plasmid DNA

Individual colonies were grown overnight in 5 mL ('small scale') or 200 mL ('medium scale') sterile LB supplemented with appropriate antibiotic selection (Section 2.1.1). Cells were harvested by centrifugation at 6,000 RCF for 15 minutes. Isolation of plasmid DNA was carried out using either the QIAprep® Spin Miniprep Kit (Qiagen) (small scale) or QIAprep® Plasmid Plus Midi Kit (Qiagen)(medium scale) following the manufacturer's protocol. Purified DNA samples were quantified using either a NanoDrop™ 8000 Spectrophotometer (ThermoFisher Scientific) or a NanoDrop™ One Spectrophotometer (ThermoFisher Scientific). Plasmid DNA

must be highly pure for successful transfection of HEK293 cells. Nucleic acids have an absorbance maxima at 260 nm, whilst proteins maximally absorb 280 nm light. The ratio of 260 nm/280 nm absorbance is used as a nucleic acid quality metric, with a value of ≥1.8 considered pure. However, to negate a stress response in Expi293F™ cells, DNA purified with an A260/280 of ≥1.95 is required. Purified plasmid DNA was then stored at −20 °C.

2.4.2 Polymerase Chain Reaction screening

PCR screening was used to verify the length of cloned inserts, and for amplification of DNA encoding the stabilised prefusion conformation F protein from its original vector.

For verification of cloned insert sequence length, the standard OPPF primers pOPIN_FOR, and pTRIEX_Down, which each target respective sequences upstream and downstream of the cloning site in the pOPIN vector backbone were used. For amplification of DNA encoding the stabilised prefusion conformation F protein, appropriate primers were designed and used. Sequences of all primers are described in Table 2.4.

Table 2.4: Primers used in this project. Primer regions targeting the vector backbone are noted in lowercase, whilst regions complimentary to the target gene are capitalised.

Name	Orientation	Sequence	Usage
pOPIN_For	Forward	gaccgaaattaatacgactcactataggg	Screening
pTRIEX_Down	Reverse	tcgatctcagtggtatttgtg	Screening
OPPF21999_F	Forward	gcgtagctgaaaccggc CAGAACATTACCGAGGAATTCTAC	Subcloning
OPPF21999_R	Reverse	gtgatggtgatgttt TCCTCCCAGGAAGGTGCTCAG	Subcloning
T7 Forward	Forward	taatacgactcactataggg	Sequencing

All PCR reactions were carried out using Phusion Flash II High-Fidelity PCR Master Mix (ThermoFisher Scientific), which is provided as a 2x solution, and comprises the Phusion Flash II DNA Polymerase, all required nucleotides and buffering solution. 0.1 µL of each primer, and 1 µL of each DNA sample, were

combined with 12.5 µL of 2x Phusion Flash II High-Fidelity PCR Master Mix, and made up to 25 µL using water. PCR reactions were performed using a Veriti™ 96-Well Fast Thermal Cycler (Applied Biosystems), following a standard PCR protocol (Table 2.5). PCR products were analysed by 0.8 % agarose gel electrophoresis.

Table 2.5: Polymerase Chain Reaction Protocol.

Stage 1	Stage 2			Stage 3	
Denaturation	Denaturation	Annealing	Elongation	Elongation	Cooldown
98 °C	98 °C	60 °C	72 °C	72 °C	12 °C
30 s	30 s	30 s	30 s	30 s	5 min
x1		x30		x1	

2.4.3 Agarose gel electrophoresis

0.8 % (w/v) agarose gels were prepared by dissolving 0.8 g agarose (Bioline) in 100 mL 1x Tris-borate-EDTA (TBE) buffer (Sigma-Aldrich, Merck), and microwaving until homogeneously dissolved. Once the agarose had cooled to hand hot (roughly 50 °C), SYBR™ Safe DNA gel stain (Invitrogen™, ThermoFisher Scientific) was added to a final dilution of 1:10,000 (v/v) to visualise the DNA migration position within agarose gels. 5 µL of 6x DNA Loading Buffer Blue (Bioline) was added to each PCR amplified DNA samples prior to loading to the gel alongside HyperLadder™ 1kb DNA ladder (Bioline). 0.8 % agarose gels were submerged in 1x TBE buffer and run at 80 V for 1 hour. 0.8 % agarose gels were visualised by either a ChemiDoc or ChemiDoc XRS+ imaging system (Bio-Rad) using Image Lab Touch Software (Bio-Rad). Gel images were analysed using the Image Lab Software v6.1 (Bio-Rad).

2.4.4 Gel purification of DNA

For gel purification of PCR products, used in the subcloning of stabilised prefusion conformation bRSV F protein, relevant bands were excised by a clean, sharp scalpel blade atop a UV transilluminator for visualisation. DNA was extracted by a QIAquick Gel Extraction system (Qiagen), following the manufacturer's protocol.

2.4.5 Plasmid sequence verification

Once cloned inserts had been PCR screened for the correct length, aliquots were sent for Sanger sequencing at Source BioScience.

Samples were diluted to 100 ng/µL using nuclease-free water, and submitted for collection. Sequencing reactions were run with in-house T7Forward primers, and dGTP chemistry for heavy chain inserts to resolve any secondary structure that may otherwise impair read quality of longer sequences. Resulting sequence chromatograms were analysed using DNADynamo v1.569 (Blue Tractor Software Ltd) to assess cloned sequences for correct sequence, correct reading frame and sequence homogeneity.

2.5 Protein expression

2.5.1 Transfection of Expi293F™ cells for Fab expression

Expi293F™ cells were seeded at 1×10^6 cells/mL 1 day prior to transfection. Consequent to satisfying the criteria of achieving a culture density of $1.5-2 \times 10^6$ cells/mL, and a viability of ≥ 95 % at the point of transfection, 1 µg/mL of Expi293F™ culture of purified plasmid DNA with an A260/280 of ≥1.95 was diluted in a volume of Opti-MEM™ Reduced Serum Medium (Gibco®, ThermoFisher Scientific) constituting 10 % of the total culture volume, along with 135 nM Transfection Grade Linear Polyethylenimine Hydrochloride (MW 40,000) (PEI) (Polysciences) as a transfection reagent, and incubated at room temperature for 10 minutes, before being added to the culture. 16 - 18 hours after transfection, the culture was further supplemented with 50 mM Glucose (Sigma-Aldrich, Merck) which supplements the energy content, and 6.67 mM Sodium Propionate (Sigma-Aldrich, Merck), and 5 mM Valproic Acid (Sigma-Aldrich, Merck), which are Histone Deacetyase inhibitors, and therefore prolong protein expression. 3 days after transfection, cells were separated from the media by centrifugation of 6,000 RCF for 45 minutes at 4 °C.

2.5.2 Stable Expi293F™ cell line expression of bRSVF

For expressions, stably transfected Expi293F™ cells were seeded at 1×10^6 cells/mL. 40 - 42 hours after seeding, the culture was supplemented with 50 mM Glucose (Sigma-Aldrich, Merck), and 6.67 mM Sodium Propionate (Sigma-Aldrich, Merck), and 5 mM Valproic Acid (Sigma-Aldrich, Merck), and the incubation temperature was reduced to 30 °C. 6 days after transfection, cells were separated from the media by centrifugation of 6,000 RCF for 45 minutes at 16 °C.

2.5.3 Dual infection of HEK293T cells for VLP production

A large volume of adherent HEK293T cells is required to produce sufficient VLP for purification. Once HEK293T cultures achieved confluency during regular adherent culture maintenance, cells were trypsinised using Trypsin-EDTA solution (0.25%) (Merck), and roughly diluted 1:20 in cell growth medium (Dulbecco's Modified Eagle Medium (DMEM) - High Glucose, 25 mM HEPES (Merck) supplemented with 10 % filter sterilised foetal calf serum (FCS) (Pirbright Stores), 2 mM L-Glutamine (Gibco™) and 50 U/mL Penicillin-Streptomycin (Gibco™)). Approximately 100 mL of diluted cells were added to 2125 cm^2 Expanded Surface Roller Bottles (Gibco™), and incubated at 37 °C with very slow rolling. After 4 days, when the flasks have achieved confluency, cell growth medium is replaced with approximately 80 mL virus growth medium (DMEM - High Glucose, 25 mM HEPES (Merck) supplemented with 2 % (v/v) filter sterilised foetal calf serum (FCS) (Pirbright Stores), 2 mM L-Glutamine (Gibco™) and 50 U/mL Penicillin-Streptomycin (Gibco™)). 10 µL of vTF7.3 recombinant VV corresponding to a multiplicity of infection (MOI) of 5, and 10 µL of P1-2A-3C recombinant VV with an MOI of 10 were diluted in 5 mL virus growth media per roller bottle, before being added to the HEK293T cells. Infection was incubated for 24 hours at 37 °C with very slow rolling. Infected cells were detached from the flask walls by mechanical agitation, and harvested by centrifugation of 4,000 rpm for 15 minutes at 4 °C.

2.6 Protein purification and concentration

2.6.1 Immobilised affinity gel filtration of Fabs

Large volumes of cell culture were required to produce yields of Fab sufficient for structural study, presenting a challenge for protein purification. To exacerbate this issue, the crude media needs to be diluted 1:1 in Phosphate Buffered Saline (PBS)(Fisher Chemical™, ThermoFisher Scientific) as some media components displace the His-tagged Fab during immobilised metal ion affinity chromatography (IMAC).

To resolve this, an automated purification programme developed by OPPF [174] was used on an ÄKTA™ xpress system (GE Healthcare). Iteratively, 50 mL of 0.45 µm filtered media supernatant, diluted 1:1 in PBS, were loaded to a 5 mL His-Tag Protein Purification Column - AminTrap (Abcam) charged with 100 mM Nickel(II) Sulphate Hexahydrate (Sigma-Aldrich, Merck) and pre-equilibrated in column wash buffer (150 mM NaCl (Sigma-Aldrich, Merck), 15 mM Trizma-Hydrochloride (Sigma-Aldrich, Merck), 2 mM Imidazole (Sigma-Aldrich, Merck), 0.45 µm-filtered pH 7.50), followed by 10 mL of column wash buffer until the ÄKTA™ xpress system integrated air sensor detects all media had been loaded. Sample loading flow rates were adjusted to as low as possible, maximising the incubation period of His-Tagged protein with the His-Tag purification column within equipment booking availability timeframes. Typically, a flow rate of < 0.2 mL/minute was used for overnight purification. The 5 mL His-Tag Protein Purification Column is then further washed using column wash buffer before elution of the His-Tagged Fab with column elution buffer (150 mM NaCl (Sigma-Aldrich, Merck), 15 mM Trizma-Hydrochloride (Sigma-Aldrich, Merck), 500 mM Imidazole (Sigma-Aldrich, Merck), 0.45 µm-filtered pH 7.50).

The elution peak is retained in a sample loop and injected onto a HiLoad™ 16/60 Superdex 75 prepatory grade size exclusion chromatography (SEC) column (Cytiva) pre-equilibrated with size exclusion buffer (150 mM NaCl (Sigma-Aldrich, Merck), 150 mM Trizma-Hydrochloride (Sigma-Aldrich, Merck), 0.45 µm-filtered pH

7.50). As previously mentioned, proteins have an absorbance maxima at 280 nm. The absorbance of 280 nm light by the SEC column elution was recorded, and 2 mL fractions were collected over 1 SEC column volume.

2.6.2 Immobilised affinity gel Filtration of bRSVF

Stabilised prefusion conformation F protein aggregates at low temperature, and consequently, a modified protocol was used to purify bRSVF. 0.45 µm filtered media supernatant was diluted 1:1 with PBS (Fisher Chemical™, ThermoFisher Scientific), and 0.45 µm filtered Nickel Sulphate (Sigma-Aldrich, Merck) was added to 5 mM. The mixture was incubated for 2 hours at room temperature with stirring of 150 rpm before using the same automated purification programme on an ÄKTA™ xpress system (GE Healthcare) as used for Fab purification, with modified buffer composition and increased system temperature. The wash buffer used for purifying bRSVF was 1x PBS (Fisher Chemical™, ThermoFisher Scientific) with 40 mM Imidazole (Sigma-Aldrich, Merck) at pH 7.40, the elution buffer was 1x PBS (Fisher Chemical™, ThermoFisher Scientific)with 300 mM Imidazole (Sigma-Aldrich, Merck) at pH 7.40, and the size exclusion buffer was 1x PBS (Fisher Chemical™, ThermoFisher Scientific) at pH 7.40. bRSVF purification was performed at 16 °C. Additionally, due to the larger molecular weight of bRSVF over the Fabs (Tables A.2 and A.3), a HiLoad™ 16/60 Superdex 200 prepatory grade size exclusion column (Cytiva) was used for gel filtration.

2.6.3 Chloroform extraction of VLPs

The infected cell pellet was re-suspended in 28 mL ice-cold phosphate buffer (40 mM Sodium Phosphate, 150 mM Sodium Chloride, pH 7.6), and Igepal (Sigma-Aldrich, Merck) was added to 0.5 % (v/v) and mixed by inversion. The mixture was incubated at 4 °C for 20 minutes to lyse the cells. Cellular debris was pelleted by centrifugation at 4,000 rpm for 5 minutes at 4 °C. The soluble fraction was retained for a later step. Cellular debris was again re-suspended in 4 mL ice-cold phosphate buffer and 4 mL Chloroform (Sigma-Aldrich, Merck) was added to digest

remaining membranous vesicles, extracting the VLPs. After vortexing well, the mix was centrifuged at 4,000 rpm for 5 minutes at 4 °C. The resulting aqueous phase was mixed with the retained solubilised fraction from the cellular debris spin, and layered above a 30 % sucrose cushion in phosphate buffer within an Open-top Thin-walled Polypropylene Ultracentrifuge tube (Beckman Coulter). VLPs were harvested by ultracentrifugation of 28,000 rpm for 5 hours at 12 °C using a Sorvall WX Ultracentrifuge (ThermoFisher Scientific). Pelleted VLPs were covered with 0.5 mL phosphate buffer and incubated overnight at 4 °C, before being fully re-suspended using a micro-pestle.

2.6.4 Sucrose density gradient (SDG) purification of VLPs

Chloroform-extracted VLPs re-suspended in phosphate buffer were treated with Ribonuclease A from bovine pancreas (Sigma-Aldrich, Merck) to 0.2 mg/mL and Igepal (Sigma-Aldrich, Merck) to 0.1 %, and incubated at 4 °C for 20 minutes. The treated VLPs were then clarified by microcentrifugation at 14,000 rpm for 3 minutes. The resulting supernatant was layered above a 15-45 % Sucrose density gradient (SDG) prepared within phosphate buffer within an Open-top Thin-walled Polypropylene Ultracentrifuge tube (Beckman Coulter) using a Gradient Master (BioComp). VLPs were separated by isopycnic ultracentrifugation of 23,000 rpm for 17 hours at 12 °C using an SW32 rotor within a Sorvall WX Ultracentrifuge (ThermoFisher Scientific). 1 mL Fractions were manually collected by puncturing the bottom of the Open-top Thin-walled Polypropylene Ultracentrifuge tube (Beckman Coulter) using a Universal Fraction Recovery System (Hoeffer) with an Ultracentrifuge tube Needle Assembly (Beckman Coulter). Fractions of interest were analysed by SDS-PAGE, quantified by Qubit™, and assessed for capsid integrity by VHH double antibody sandwich ELISA.

2.6.5 Concentrating proteins by centrifugation

Proteins were concentrated using centrifuge ultrafiltration devices, where solvent and salts pass through the filter membrane during centrifugation, whilst species

larger than the defined cut off value are retained in the filter residue.

Prior to their usage, all Amicon® Ultra filter units were washed with 1x PBS (Fisher Chemical™, ThermoFisher Scientific) at 4,000 RCF for 5 minutes.

Ultralong and non-ultralong Fabs studied within this project are approximately 56 and 53 kDa in size respectively (Table A.3), and were concentrated using an Amicon® Ultra-15 centrifugal filter unit with an Ultracel-30 membrane MWCO 30 kDa (Merck) using iterative 5 or 10 minute spins of 4,000 RCF at 4 °C, until the Fabs reached a concentration appropriate for downstream processes. Fabs were then aliquoted and stored at 4 °C.

Stabilised prefusion conformation F protein forms a homotrimer, where each monomer is approximately 57 kDa, and the homotrimer is approximately 172 kDa (Table A.2). F protein was concentrated using an Amicon® Ultra-15 centrifugal filter unit with an Ultracel-100 membrane MWCO 100 kDa (Merck) using iterative 5 or 10 minute spins of 4,000 RCF at 16 °C, until the F protein reached a concentration appropriate for downstream processes. F protein was then aliquoted and flash frozen by liquid nitrogen submersal, before being stored at −80 °C.

VLPs are assembled with 60 replicates of a protomer of approximately 80 kDa, with the complete capsid being approximately 4845 kDa. VLPs were concentrated using an Amicon® Ultra-15 centrifugal filter unit with an Ultracel-100 membrane MWCO 100 kDa (Merck) using iterative 5 or 10 minute spins of 4,000 RCF at 4 °C, until the VLPs reached a concentration appropriate for downstream processes. VLPs were then aliquoted and stored at 4 °C.

2.6.6 Protein G re-purification

To help resolve non-specific binding issues with Fabs during BLI experiments, Fabs were re-purified using a NAb™ Protein G Spin Kit, 0.2 mL (ThermoFisher Scientific), following manufacturer's instructions except that samples were loaded to the NAb™ spin columns in two 500 µL batches and subsequent washing steps instead of one loading step. The resulting elution fractions were quantified by NanoDrop

using a NanoDrop™ Lite Spectrophotometer (Life Technologies), with the fractions containing the two highest protein concentrations being pooled for desalting.

2.6.7 Protein desalting

The NAb™ Protein G Spin Kit elution buffer contains amine and has a pH of 2.8 which may affect binding on the BLI. Consequently, Protein G re-purified Fabs were desalted using Zeba™ Spin Desalting Columns, 7K MWCO, 2 mL (ThermoFisher Scientific), following manufacturer's instructions. The resulting desalted protein was quantified by NanoDrop using a NanoDrop™ Lite Spectrophotometer (Life Technologies).

2.7 Protein characterisation

2.7.1 Inference of protein molecular weight by SEC

The molecular weight of a species can be inferred during SEC if the elution volume (indicated by 280 nm UV absorbance) of a known protein standards are measured (V_e). The SEC column limits (the void volume, V_0, and total column volume, V_C) also need to be determined. The void volume is determined using a large species, such as Dextran Blue, which does not penetrate the column resin, whilst, a small molecule such as acetone is used to determine the total column volume as it travels through all the resin pores. The average distribution constant, K_{av} (and in turn the species molecular weight) can then be calculated by the following equation:

$$K_{av} = \frac{V_e - V_0}{V_C - V_0}$$

K_{av} is then plotted against molecular weight (using a semi-logarithmic plot), such that K_{av} can be used to infer species molecular weight (Fig. E.1).

Gel Filtration Protein Standards (Bio-Rad) were re-suspended in 500 µL 0.45 µm-filtered PBS at pH 7.40, and 125 µL were injected to a Superdex 200 Increase 10/300 GL size exclusion chromatography column pre-equilibrated with pH 7.40, 0.45 µm-filtered PBS, loaded to an ÄKTA™ Pure chromatography system. To measure the

column limits, Acetone (Sigma-Aldrich, Merck) was diluted to 5 mg/mL and mixed with 2 mg/mL Dextran Blue (Sigma-Aldrich, Merck) in 500 µL 0.45 µm-filtered PBS at pH 7.40, and injected to the SEC column. Protein standard results reported in Appendix B.1 for the Superdex 200 Increase 10/300 GL SEC column.

2.7.2 SDS polyacrylamide gel electrophoresis (PAGE)

Sodium dodecyl-sulphate (SDS) uniformly denatures and binds to proteins, attaching negatively charged ions proportional to the protein length, and hence its mass. Disulphide bonds are not reduced upon addition of SDS, and so a reducing agent can optionally be included to differentiate species of a protein complex bound together by disulphide bond(s), such as Fab heavy and light chains. Samples can be separated by electrophoresis, allowing easy assessment of the protein molecular weight, when appropriate molecular weight standards are also run for comparison.

Samples were prepared by mixing with 2x SDS Loading Buffer (4.0 g SDS, 100 mM Tris-HCl pH 6.8, 20 % (v/v) Glycerol, 2.0 g/L Bromophenol blue) at a ratio of 1:1. When sample reduction was required, the SDS Loading Buffer was supplemented with 10 % (v/v) β-mercaptoethanol (βME)(ThermoFisher Scientific). To ensure proteins are completely denatured, samples were heated to 95 °C for 10 minutes, followed by 4 °C indefinitely using a Veriti™ 96-Well Fast Thermal Cycler (Applied Biosystems) prior to loading to either a NuPAGE™ 4 to 12 % Bis-Tris, 1.0 mm, Midi Protein Gel (Invitrogen™, ThermoFisher Scientific), or NuPAGE™ 4 to 12 % Bis-Tris, 1.0 mm, Mini Protein Gel (Invitrogen™, ThermoFisher Scientific). Gels were run in 1x NuPAGE™ MOPS SDS Running Buffer (Invitrogen™, ThermoFisher Scientific), at either constant 150 V for 90 minutes or 180 V for 40 minutes.

The SDS-PAGE performed to verify Fab composition prior to BLI optimisation experiments, after ProteinG re-purification, and to characterise purified VLPs were performed at Pirbright, following the same protocol with the following exceptions: samples were prepared by mixing with NuPAGE™ LDS Sample Buffer (4X)(Invitrogen™, ThermoFisher Scientific), and optionally NuPAGE™ Sample Reducing Agent (10X)(Invitrogen™, ThermoFisher Scientific). Gels were

run at constant 180 V for 40 minutes. SeeBlue™ Pre-stained Protein Standard (Invitrogen™, ThermoFisher Scientific) was used as a molecular weight standard.

2.7.3 SDS-PAGE staining

When a higher quantity of protein is expected, such as following protein purification, SDS-PAGE gels were stained using InstantBlue® Coomassie Protein Stain (Abcam) for 1 hour at room temperature with gentle agitation, and washed using ddH_2O for at least 15 minutes prior to imaging using either a ChemiDoc or ChemiDoc XRS+ imaging system (Bio-Rad) using Image Lab Touch Software (Bio-Rad). Gel images were analysed using the Image Lab Software v6.1 (Bio-Rad).

2.7.4 Western blotting

When a lower quantity of protein is expected, or the identity of protein bands needs to be confirmed, such as during protein expression screening, Western blotting was used to identify protein bands due to the increased sensitivity, and epitope targeting features.

Protein bands were transferred to a nitrocellulose (NC) membrane using either an iBlot™ 2 Dry Blotting System (Invitrogen™, ThermoFisher Scientific) with iBlot™ 2 Transfer Stacks, Nitrocellulose (Invitrogen™, ThermoFisher Scientific) following a 20 V, 7 minute programme, or using Power Blotter XL System (Invitrogen™, ThermoFisher Scientific). For the Power Blotter XL System, Gel Blot filter papers (Cytiva) were soaked in Transfer Buffer (40 % (v/v) methanol, 1x Tris/Glycine/SDS buffer) before layering in the transfer stack with a Amersham™ Protran™ 0.45 µm Nitrocellulose Blotting Membrane (GE Healthcare) cut to size. transfer was achieved using a 20 V, 7 minute programme. After a brief wash with water, NC membranes were blocked with 5 % (w/v) Semi-Skimmed Milk (Oxoid) dissolved in PBS (Fisher Chemical™, ThermoFisher Scientific) with 0.05 % (v/v) Tween20 (PBS-Tween20) (Sigma-Aldrich, Merck) for 1 hour with gentle agitation at room temperature. Alternatively, to accommodate half-day laboratory access during the COVID-19 pandemic, membranes were blocked overnight at 4 °C. Blocking

solution was drained and replaced with primary antibody solution (1:2,500 His Tag Antibody (MAB050)(R&DSystems; AD1.1.10 monoclonal) (Section 2.1.2) diluted in fresh blocking solution) and incubated for 1 hour with gentle agitation at room temperature. Membranes were washed 3 times with PBS-Tween20, each wash for 10 minutes, before secondary antibody solution (1:100,000 Anti-Mouse IgG (whole molecule) Peroxidase antibody produced in Rabbit (Sigma-Aldrich, Merck; polyclonal A9044-2ML) (Section 2.1.2) diluted in fresh blocking solution) was applied for 1 hour with gentle agitation at room temperature. Membranes were washed with 3 more washes of PBS-Tween20, before visualisation by chemiluminescence using the SuperSignal™ West Femto Maximum Sensitivity Substrate (ThermoFisher Scientific) and imaging using a ChemiDoc XRS+ imaging system (Bio-Rad), or a ChemiDoc MP imaging system (Bio-Rad) using Image Lab Touch Software (Bio-Rad). Gel images were analysed using the Image Lab Software v6.1 (Bio-Rad).

2.7.5 Protein concentration quantification

Except for VLPs, protein concentrations were assessed by measuring the absorbance maxima at 280 nm, using a variety of NanoDrop spectrophotometers: a NanoDrop™ 8000 Spectrophotometer (ThermoFisher Scientific) was used for quantifying the majority of Fab and bRSVF samples, and was later replaced with a NanoDrop™ One Spectrophotometer (ThermoFisher Scientific). A NanoDrop™ Lite Spectrophotometer (Life Technologies) was used for assessing Fab protein concentrations after NAb protein G re-purification and Zeba protein desalting, and a NanoDrop™ Spectrophotometer ND-1000 (Labtech) was used immediately prior to preparing sample grids for electron microscopy. When permitted by the spectrophotometers, protein extinction coefficients and molecular weights, predicted using the ExPASy ProtParam tool (Swiss Institute of Bioinformatics) [180], were entered for accurate concentration calculation, since protein concentration is determined as below:

$$Protein\ concentration = \frac{280\,\text{nm}\ Absorbance}{\left(\frac{Extinction\ coefficient}{1000}\right)}$$

Additionally, the protein molecular weight can be supplied to calculate the molar concentration of a sample. A list of key ProtParam results for each protein sequence studied is provided in Table A.3. In absence of the option to specify extinction coefficient, the '1 Abs = 1 mg/mL' mode was used.

2.7.6 Qubit™ quantification of VLPs

Due to their complexity, VLP concentration quantification was measured by Qubit™ protein assay, which provides more accurate quantitation by binding a fluorescein to the target protein and comparing the resultant fluorescence against standards of known concentration. However Qubit™ quantitation takes much longer, due to the 15 minute incubation period. VLPs were quantified at room temperature by a Qubit™ 2.0 Fluorometer (Invitrogen™, ThermoFisher Scientific), following manufacturer's protocols.

2.7.7 VLP Integrity assessment by VHH double sandwich ELISA

Virus-like particles (VLPs) promise a safer alternative to attenuated encapsulated viral vaccines, due to the lack of not needing high biocontainment facilities to handle live virus, and the impossibility of reversion (Section 4.1.5). Only the proteinaceous viral shell remains as complicated assembly. For *Picornaviridae*, the capsid is composed of 60 self-assembling copies of a protomer unit, itself a self-assembly of 4 viral proteins (VP1-4) [181, 182]. The 240-mer complete capsid therefore presents a challenge for recombinant expression systems, with engineered stability mutations developed to promote VLP integrity [177, 183].

To assess VLP integrity, a VHH double sandwich ELISA was performed with detection antibodies specific to internal and external epitopes of VLPs of each FMDV serotype. Responses from both detection antibodies are compared for the purified VLP using the untreated sample, and heat-treated VLPs to assess VLP integrity of the purified batch (Fig. 2.3).

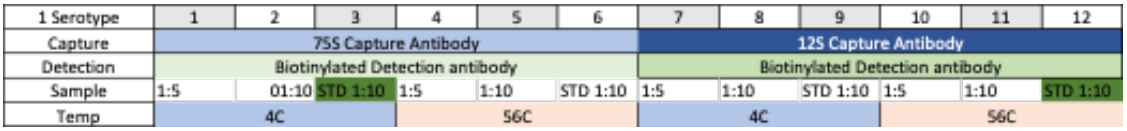


Figure 2.3: Schematic to show ELISA experimental set up.

Table 2.6: Double sandwich VHH ELISA capture antibodies or antigen. M3 and Ab117 were produced from The Immunological Toolbox (ITB) [184], the remainder were purchased from Wageningen Bioveterinary Research.

Serotype	Particle	Antibody/Antigen	Concentration	Dilution
All	75S/12S	Integrin	1 mg/mL	1:2000
A22	75S	M702	1 mg/mL	1:2000
A22	12S	M3	0.5 mg/mL	1:1000
Asia1	75S	M332	1 mg/mL	1:2000
Asia1	12S	M3	0.5 mg/mL	1:1000
O1M	75S	M170	1 mg/mL	1:2000
O1M	12S	M3	0.5 mg/mL	1:1000
SAT2	75S/12S	M377/M311	1 mg/mL	1:2000
SAT2	12S	Ab117	1 mg/mL	1:2000

The ELISA plate (Nunclon 96 Flat Bottom Transparent Polystyrene Plate (ThermoFisher Scientific)) was coated with 0.5 µg/mL VHH capture antibody diluted in 100 µL Coating Buffer (50 mM carbonate/bicarbonate buffer, pH 9.6)(Pirbright Stores) (Table 2.6), sealed and incubated either overnight at 4 °C, or for 1 hour at 37 °C. VHH Capture antibodies M3 and Ab117 were produced by The Immunological Toolbox (ITB) [184], the remainder were purchased from Wageningen Bioveterinary Research (WBVR). Ab117 was only used for the capture, and detection of SAT2 12S particles after cryo-EM confirmed binding to the internal leaflet of SAT2 dissociated pentamers (Fig. 4.23), before then, there was no satisfactory antibody specific for SAT2 12S pentamers. The plate was briefly washed 3 times using 200 µL VHH ELISA Wash Buffer (0.05 % (v/v) Tween20 (Sigma-Aldrich, Merck), PBS (Pirbright Stores) (PBS-Tween20)). Residual wash liquid was removed by strong agitation.

VLP samples were diluted 1:5 in 120 µL PBS (Pirbright Stores), whilst VLP standards were diluted 1:10 in 120 µL PBS (Pirbright Stores). 60 µL of each diluted sample/standard was incubated at 4 °C, or 56 °C for 15 minutes to thermally

disrupt the capsids, followed by 4 °C hold until use using a Veriti™ 96-Well Fast Thermal Cycler (Applied Biosystems).

50 µL of VHH Milk Buffer (1 % (w/v) Skimmed Milk (Marvel), VHH ELISA Wash Buffer) were added to each well of the ELISA plate. 25 µL of sample/standard were added of each well of row A of the ELISA plate. A serial 1:3 dilution down the plate was created by pipette mixing, and withdrawing 25 µL to the subsequent row until row G, where 25 µL was withdrawn and discarded, leaving row H as a blank. The ELISA plate was sealed and incubated for 1 hour at RT, with 150 rpm agitation using an ES-20 Orbital Shaker-Incubator (Grant-Bio). The plate was briefly washed 3 times using 200 µL VHH ELISA Wash Buffer and residual wash liquid was removed by strong agitation.

Table 2.7: Double sandwich VHH ELISA detection antibodies. Biotinylated M3 and Ab117 were produced from The Immunological Toolbox (ITB) [184], the remainder were purchased from Wageningen Bioveterinary Research.

Serotype	Particle	Antibody	Concentration	Dilution
A22	75S	M702-Biotin	1 mg/mL	1:5000
A22	12S	M3-Biotin	0.5 mg/mL	1:2500
Asia1	75S	M332-Biotin	1 mg/mL	1:5000
Asia1	12S	M3-Biotin	0.5 mg/mL	1:2500
O1M	75S	M170-Biotin	1 mg/mL	1:5000
O1M	12S	M3-Biotin	0.5 mg/mL	1:2500
SAT2	75S/12S	M377/M311-Biotin	1 mg/mL	1:5000
SAT2	12S	Ab117-Biotin	1 mg/mL	1:5000

Biotinylated detection antibodies (Table 2.7) were diluted to 0.2 µg/mL in VHH Milk Buffer. Biotinylated VHH detection antibodies M3 and ab117 were produced by The Immunological Toolbox (ITB) [184], the remainder were purchased from WBVR. 100 µL biotinylated detection antibody was added to each well of each condition (800 µL per column). The ELISA plate was sealed and incubated for 1 hour at RT, with 150 rpm agitation using an ES-20 Orbital Shaker-Incubator (Grant-Bio). The plate was briefly washed 3 times using 200 µL VHH ELISA Wash Buffer and residual wash liquid was removed by strong agitation.

Pierce™ High Sensitivity Streptavidin-HRP (Invitrogen™, ThermoFisher Scientific) was diluted 1:10,000 in VHH Milk Buffer, and 100 µL was added to each well. The ELISA plate was sealed and incubated for 1 hour at RT, with 150 rpm agitation using an ES-20 Orbital Shaker-Incubator (Grant-Bio). The plate was briefly washed 5 times using 200 µL VHH ELISA Wash Buffer and residual wash liquid was removed by strong agitation.

100 µL of TMB Substrate Solution (Invitrogen™, ThermoFisher Scientific) was applied to each well, and incubated until a full gradient of barely visible to intense blue colouring appeared, at which point 100 µL ELISA Stop Solution (Invitrogen™, ThermoFisher Scientific) was applied to each well. The plate was read immediately using an Infinite 200 Plate Reader (Tecan) using the endpoint absorbance 450/630 nm protocol on the Tecan i-control v.2.02.11.

Absorbance values at 650 nm are subtracted from those at 450 nm. Next the blank absorbance values from row H are subtracted in each column to account for background signal. Absorbance values for the standards across the dilution series are then plotted against known (calculated) standard concentrations to produce a linear regression model associating reference-subtracted 450 nm absorbance to detected sample concentration. These concentrations are normalised against the dilution factor and averaged per column. Mean sample detected by external epitope-targeting antibodies ($\overline{x}_{ext}$) divided by the total signal detected ($\overline{x}_{ext} + \overline{x}_{int}$), where $\overline{x}_{int}$ is the mean sample detected by internal epitope targeting antibodies, provides a figure for capsid integrity as a percentage:

$$Integrity(\%) = \left(\frac{\overline{x}_{ext}}{\overline{x}_{ext} + \overline{x}_{int}}\right) \times 100$$

2.7.8 Size exclusion chromatography with multi-angle laser light scattering

Size exclusion chromatography with multi-angle light scattering (SEC-MALS) is a powerful technique that first separates species by size, and then determines, amongst other things, the precise molecular weight, polydispersity and hydrodynamic radius

of a specimen by simultaneous recording of the scattering of incident light as it passes through a specimen to multiple spatially arranged detectors. Intensity provides information on molar mass, whilst the angular scattering informs on the size of the analyte.

Specimens were auto-injected to a Superdex 200 Increase 10/300 GL size exclusion chromatography column (Cytiva) pre-equilibrated with pH 7.40, 0.45 µm-filtered PBS (Fisher Chemical™, ThermoFisher Scientific), loaded to an ÄKTA™ Pure chromatography system (Cytiva) connected in series to a Dawn Heleos-II multi-angle light scattering detector (Wyatt Technologies) and an Optilab T-rEX refractometer (Wyatt Technologies). Data were analysed using the ASTRA® software v5.3 (Wyatt Technologies) using the Zimm method. A refractive index increment (a constant value indicating how refractive index changes with solute concentration), dn/dc, of 0.1850 mL/g was used for protein species, whilst a dn/dc of 0.1320 mL/g was used for conjugated species [185]. Human proteins have a mean dn/dc value of 0.1899 mL/g ± 0.003 mL/g [186]. Note that the true dn/dc value varies for each protein studied, consequently, some small systemic error was introduced to samples analysed by SEC-MALS, but this did not significantly effect the observed results.

2.7.9 Prediction of protein glycosylation sites

Sites of N-glycosylation can be predicted from the protein sequence with reasonable accuracy (76 % overall accuracy), using the artificial neural network NetNglyc [187]. Similarly, O-glycosylation sites can be predicted using the NetOGlyc artificial neural network [188]. NetNGlyc accurately predicts 86 % of glycosylated sequons, and 61 % of non-glycosylated sequons. Occupancy of O-glycosylation sites varies depending on the cells expressing the protein, and specificity of galactosaminyltransferase enzymes is yet to be elucidated, so NetOGlyc can only predict regions of a protein sequence likely to carry O-glycosylation, and hence identified residues are those in the identified region with a hydroxyl group. The glycosylation potential score is

the mean result from 9 independent neural networks. Protein sequences of interest were submitted to the NetNGlyc [189] or NetOGlyc [190] servers respectively.

2.7.10 Analytical Ultracentrifugation

Analytical Ultracentrifuge (AUC) investigates the behaviour of solutes within a strong gravitational field. Within sedimentation velocity mode, AUC generates a sedimentation coefficient, (s), for a sample by observing the change in absorbance of 280 nm UV light across the length of a cell within an ultracentrifuge rotor over the course of an experiment. Macromolecules will sediment at varying speeds dependent on particle size, shape, density, and conformation. Comparisons of (s) values can inform about these characteristics of a solute within solution. For example, to observe for a bimolecular interaction, the sedimentation coefficients (s) of the lone species are first determined, and compared against the (s) of the complexed sample. A shift in (s) indicates bimolecular interaction.

In order to accurately interpret AUC data, the density and viscosity of all sample buffers needs to be determined. To this end, 4 mL of buffer (PBS (Fisher Chemical™, ThermoFisher Scientific), pH 7.40, 0.45 µm-filtered) was manually injected to, and measured using a DMA 5000M Density Meter (Anton Paar) equipped with a 2000ME Viscometer module (Lovis). Samples were prepared to 2 µM, 4 µM, or 8 µM using PBS (Fisher Chemical™, ThermoFisher Scientific), and 196 µL sample, or 200 µL PBS (Fisher Chemical™, ThermoFisher Scientific) were injected to respective chambers of standard 2-sector epon centrepieces. Cells were loaded to an An-50Ti rotor within either a ProteomeLab XL-I or an Optima AUC (Beckman Coulter) analytical ultracentrifuge. Sedimentation velocity (SV) scans were recorded by measurement of 280 nm interference and absorbance measurements collected every 20 seconds at 20 °C at 50,000 rpm until the system memory reached capacity (999 scans for ProteomeLab XL-I, 1501 scans for Optima AUC). Following characterisation of single protein constructs, SV scans of 2 µM bRSVF in complex with increasing concentrations (2 µM, 4 µM, or 8 µM) of Fabs B4, B13, or B13* were recorded in the same manner. The partial specific volume of each protein

construct was calculated from the respective amino acid sequences using SEDFIT programme [191]. The partial specific volume of bRSVF was used during analyses of the complexes. AUC data were processed using SEDFIT [191], fitting to the c(S) model [192] to obtain Sedimentation coefficient distributions. Calculation of molecular weight from the sedimentation coefficient is dependent on the frictional ratio parameter (f/f_0). f/f_0 is fitted globally, thus when multiple species are observed, f/f_0 is an average value, and will not give true molecular weight values for each individual species. GUSSI v1.4.1 (The University of Texas South Western Medical Centre) was used to prepare figures.

2.7.11 Bio-layer Interferometry

Bio-layer interferometry (BLI) provides a label-free mechanism to interrogate bimolecular interactions. In BLI, dip-and-read biosensor probes are sequentially dipped into a series of solutions, whilst measuring the interference of light at the sensor's distal end. Sample interacting with the sensor tip introduces a phase shift to the reflected light, and hence bimolecular interactions can be measured in real time if one component of the interaction is first immobilised to the surface of the sensor tip.

Following BLI experimental optimisation for Fab experiments Section 4.7.1, the following kinetic analysis experiment protocol was determined. FMD VLPs were biotinylated using a Lightning-Link® Biotinylation Kit (Fast, Type B) (Abcam), following manufacturer's instructions. 5 µg/mL biotinylated FMD VLPs were then immobilised to Octet® Streptavidin Biosensor Probes (Sartorius) at 25 °C for 650 seconds, using an Octet® R8 BLI system (Sartorius). Probes were then blocked for 30 seconds with 10 µg/mL Biocytin (Sigma-Aldrich, Merck), before being introduced to Fab at varying concentrations (0 nM *(as a sample reference)*, 15 nM, 50 nM, 100 nM, 250 nM, 500 nM, and 1500 nM). Association and dissociation measurements were recorded over 1800 seconds. The optimised experimental schema is provided in Table 2.8.

Due to varying degrees of residual non-specific binding (NSB), each experiment was repeated with sensors dipped in buffer instead of FMD VLPs to provide a sensor

Table 2.8: Experimental schematic for Biolayer Interferometry characterisation of Fab antigenic interactions.

Stage	Step	Column	Time (s)
1	Baseline	1	60
2	Loading	2	650
3	Baseline	3	30
4	Blocking	4	30
5	Baseline	5	30
6	Association	6	1800
7	Dissociation	5	1800

reference. All stages were conducted in 200 µL volumes of HEPES buffer system supplemented with EDTA and Polysorbate-20 (HBS-EP) buffer (Teknova) within a 96 well black chimney microplate (Greiner Bio-One) and agitated at 1000 rpm throughout the experiment. The Octet® R8 BLI system was controlled by Octet® Discovery Studio v12.2.1.18 (Sartorius), and data analysis performed by Octet® Analysis Studio v12.2.1.23 (Sartorius). In order to fully account for residual NSB, the sensor reference data were first subtracted from the associated experimental data. The sample reference (0 nM Fab/IgG) was then subtracted from all probes to account for any sensor drift. The resulting data were aligned on the average of the baseline step immediately preceding association, and the end of dissociation stages. To aid data visualisation and interpretation, Savitzky-Golay filtering [193] was applied to the data. Samples were grouped by name, maximal response (R_{max}) values were unlinked (each analyte concentration will have a different R_{max}), and analysed according to the "Global Fit" method, which fits all traces against a predicted simple 1:1 Michaelis-Menten binding model and calculates kinetic constants using all specified concentrations. Alternatively, the "Local Fit" method calculates kinetic constants for each analyte concentration independently, assuming all interactions are completely reversible (given infinite time), whilst the "Steady State" method also calculates using all specified analyte concentrations, but requires "well behaved" 1:1 bimolecular interactions where each concentration achieves equilibrium. In this thesis, BLI was used to characterise the interaction of Fabs with icosahedral FMD VLPs (Section 4.6.1), where each VLP has 60 potential binding sites for Fabs. In

complex interactions as these, Sartorius still recommends applying the 1:1 kinetic model, and optimising the assay until the observed interaction *behaves* as a 1:1 reaction. Thus, systemic error will be introduced to the calculated kinetic constants.

2.8 Structural biology methods

2.8.1 ABodyBuilder structural prediction

ABodyBuilder 2 *(formerly SAbPred)*, hosted by the Oxford Protein Informatics Group [194], is a tool for the *in silico* prediction of antibody variable (Fv) regions with near equivocal accuracy to leading alternative publicly available pipelines examined during the most recent Antibody Modelling Assessment II competition [194, 195].

The ABodyBuilder 2 pipeline first employs ANARCI [196] to align and number input sequences to Fv regions from a variety of species from the SAbDab database [197]. Following orientation of the Fv domains by ABangle [198], CDR structures are predicted either by homology to known CDR structures using FREAD [199] or by satisfying known spatial restraints (Cα-Cα distances, main-chain N-O distances and Cα-Cβ χ dihedral angles) by MODELLER [200], followed by side chain conformation optimisation by SCRWL4 [201]. The ABodyBuilder 2 pipeline is fully automated, with options to use alternative antibody numbering schemes.

Antibody Fv regions amino acid sequences were submitted to the ABodyBuilder 2 [194] server for Fv paratope modelling, numbered according to IMGT [202].

2.8.2 AlphaFold structural prediction

In 2018, the biannual Critical Assessment of protein Structural Prediction (CASP) competition was won by DeepMind's AlphaFold (AF) [203], an artificial intelligence programme for protein structural prediction that leverages a deep-learning neural network to perform extensive multiple-sequence alignments and considers a wealth of known biophysical characteristics and known structures of proteins, harvested from the PDB [162] (approximately 100k protein structures). The code behind AlphaFold 2.0 (AF2), the second iteration of AF, has now been made public [204], and has been installed to Oxford's Biomedical Research Computing (BMRC) cluster. The amino

acid sequences of Fab VH and VL domains were connected via a polyglycine linker to make a single-chain variable fragment (scFv) and submitted to AF2 for structural prediction. AF2 provides two metrics to interpret AF models; the predicted Local Distance Difference Test (pLDDT), which provides a positional confidence score for each residue, based on neighbouring residue chemistry, and the Predicted Alignment Error (PAE), which provides the expected error at position x, if the predicted and true structures were aligned at residue y.

2.8.3 Crystallographic X-Ray Diffraction

2.8.3.1 Pre-crystallisation test

Following purification, initial Pre-Crystallisation Tests (PCT™)(Hampton Research) were conducted following manufacturer's instructions to assess appropriate protein concentration to begin initial crystallisation studies, since protein concentration is a significant crystallisation variable. The pattern of protein precipitations observed was interpreted against the expected results table, and protein concentrations adjusted as necessary before continuing with initial crystallisation screens.

2.8.3.2 Crystallisation screens

A number of commercial crystallisation screens were used for screening crystal hits. The Index™ screen and JCSG-*plus*™ screens were included for their breadth. The Morpheus® screen was selected due to the broad range of additives and systemic inclusion of cryo-protectants.

The Index™ screen (Hampton Research) is a sparse matrix, constituting a broad range of traditional salts (Sodium Chloride, Ammonium Sulphate), neutralised organic acids, low salt concentration mixed with high polymer concentration, high salt concentration mixed with low polymer concentration, and low ionic strength polymers covering a pH range of 3.0 to 9.0, aimed at providing initial indications for further crystallisation optimisation.

JCSG-*plus*™ (Molecular Dimension) is another sparse screen aimed at providing initial insight to sample crystallisation, composed of classic polymers (polyethylene

glycol; PEG) and salts (Sodium Chloride), organics and polyalcohols, neutralised organic acids and surveying a pH range of 4.0 to 10.5 [205].

The Morpheus® screen (Molecular Dimensions) allows surveying of novel chemical space not covered in traditional crystallisation screens (such as Index™ and JCSG-*plus*™), with a range of buffer systems, salt additives, polymers and ligands in a 3D matrix covering pH 6.5, pH 7.5 and pH 8.5. Additionally, all wells of the Morpheus® screen include cryo-protectant, simplifying the flash freezing process by eliminating the cryoprotective step [206].

2.8.3.3 Crystallisation trials

Crystallisation trials were set up using 200 nL or 300 nL sitting droplet vapour diffusion at 20 °C, following standard OPPF crystallisation protocols [207].

96 Well CrystalQuick™ X Plates (Greiner Bio-One) were used for initial crystallographic studies due to their ultra-thin walled well bottoms and material composition which is X-ray, and UV-compatible with low birefringence, allowing *in situ* X-ray analysis for rapid assessment of crystal diffraction quality. Subsequent crystallisation studies were set up within Swissci UV-transmissible Polymer 96-Well Triple Drop Plates (Molecular Dimensions), due to the wide concave sub-well profile making manual crystal retrieval easy, as well as providing a lens for optimal droplet placement and viewing.

40 µL of crystallisation screen solution was transferred to the solution reservoir sub-well using a Hydra96 liquid handling robot (Robbins Scientific). Protein and crystallisation screen solution droplets were prepared using a Mosquito liquid handling robot (SPT Labtech). 100 nL or 200 nL protein droplets were dispensed to 100 nL crystallisation screen solution.

Plates were sealed and incubated at 20 °C within a RockImager 1000 automated crystal imaging system (Formulatrix). Plates were imaged using 'Default' and 'UV-Mid' settings according to an inverse Fibonacci spiral schedule (+ 12 hours, + 1 day, + 2 days, + 3 days, + 4 days, + 5 days, + 8 days, + 13 days, and + 24 days since loading to the RockImager). Imaging using UV is useful to verify

whether visually observed crystals are composed of protein or salt, due to aromatic residue fluorescence under UV light.

Once proteinaceous crystals were observed, the plate was transferred to the VMXi RockImager for *in situ* data collections, or alternatively individual wells were opened using a scalpel blade, and, using a MDG41 digital microscope system (Leica) with DFC295 camera (Leica), crystals were retrieved using a variety of crystal mounts, roughly picked by size to suit each retrieved crystal, and flash frozen in liquid nitrogen. Since the Morpheus® screen contains cryoprotectant in all conditions, crystals retrieved from this screen did not need any additional cryoprotection. Crystals retrieved from either the Index, or JCSG-*plus* crystallisation screens were cryoprotected by briefly dipping the retrieved crystal into mother liquor supplemented with 10 mM glycerol before flash freezing. Crystal mounts included Dual Thickness MicroMounts™ (MiTeGen), Dual Thickness MicroLoops™ (MiTeGen) and Mounted CryoLoops™ (Hampton Research).

2.8.3.4 X-Ray diffraction data collection

All X-Ray Diffraction (XRD) data were collected at Diamond Light Source (DLS) within the macromolecular crystallography (MX) beamline village. Specific details of which beamlines were accessed for each sample are provided in Table 2.9.

Table 2.9: Diamond Light Source beamlines accessed throughout this project.

Chapter	Sample	Collected	Instrument	Proposal	Method
3	D08	02/12/2019	VMXi	mx19946-217	In Person
3	D08	09/12/2019	I04	nt23570-48	In Person
3	D08*	21/02/2020	I04	nt23570-70	In Person
3	4K3E*	21/02/2020	I04	nt23570-70	In Person
4	Fab #30	10/10/2021	I03	nt23570-97	Remote

The Versatile Macromolecular Crystallography *in situ* (VMXi) beamline is a room temperature beamline focused on fully automated high-flux, high-throughput *in situ* XRD data collection; without the need to retrieve crystals from the crystallisation plate, thereby preserving macromolecular crystal integrity, and allowing screening of many crystals for diffraction quality [208]. VMXi data

collections are limited to ± 30° oscillation, however the ability to collect XRD data even from micro-crystals still allows for structural determination from challenging samples. VMXi uses a beamsize of 10 µm x 10 µm at an energy of 16 keV (0.7748 Å), permitting a maximal sample resolution of 1.95 Å at the edge of the detector. VMXi has a custom goniometer to mount crystallisation plates in the beampath and uses an Eiger2 X 4M photon-counting X-Ray Detector (Dectris) with a maximum frame rate of 560 FPS (frames-per-second), and a sensor resolution of 75 μm^2.

I04 is a classically designed MX beamline, constructed during phase I of DLS's construction. I04 has a variable beamsize from 10 µm x 5 µm up to 110 µm x 100 µm, allowing beamsize selection for optimal macromolecular crystal illumination volume. The incident X-Ray wavelength is tuneable at I04, with a range of 0.69 - 2.07 Å (6.0 - 18.0 keV), though I04 is typically used at 0.9795 Å, delivering X-Rays at 12.658 keV. I04 uses an Eiger2 XE 16M photon-counting X-Ray Detector (Dectris) with a maximum frame rate of 560 FPS, and a sensor resolution of 75 μm^2.

I03 is another classically designed MX beamline constructed during phase I of DLS's construction. I03 has a beamsize of 80 µm x 20 µm. Similar to I04, I03 has a tuneable incident X-Ray wavelength, with a range of 0.60 - 2.48 Å (5.5 - 25.0 keV). though I03 is typically used at 0.9763 Å, delivering 12.7 keV. As a consequence of I03's optics, incident X-rays at the sample are highly parallel compared to those of I04. I03 also uses an Eiger2 XE 16M photon-counting X-Ray Detector (Dectris). The major difference between I03 and I04, is the capability for biological containment, allowing experiments involving hazard group 2 and 3 samples on I03.

Both I04 and I03 feature a DLS BART Robotic Sample Changer, capable of storing up to 592 crystal mounts in 37 pucks within a tank of 40 L liquid nitrogen. XRD data collections at I04 and I03 are conducted at 100 K under a CryoJet, and use a SmarGon™ Multi-Axis Goniometer (SmarAct) for sample positioning and oscillation.

Prior to the COVID-19 pandemic, XRD data were collected in person and experiments were iteratively optimised for exposure and radiation flux based on feedback from previous crystal diffraction using the Generic Data Acquisition

software (Diamond Light Source). Crystal centering was performed optically. After the COVID-19 pandemic, all data were collected remotely using beamline-standard experimental parameters and X-Ray centering for automated XRD data collection (Table 2.9). Experimental conditions are provided in Table 2.10. Data collections were curated, processing performed and compared using the SynchWeb interface [209] of the ISPyB database [210].

Table 2.10: X-Ray data collection parameters.

Chapter	Sample	Instrument	Flux (ph/s)	Oscillation (°)	Number of images	Exposure (s)	Wavelength (Å)	Transmission (%)	Beamsize (µm)	Pipeline
3	D08	I04	4.25×10^{11}	0.1	1800	0.008	0.9795	100	86 x 75	3dii
3	D08	VMXi	5.00×10^{13}	0.1	500	0.0018	0.9795	2	10 x 10	dials
3	D08*	I04	2.93×10^{11}	0.1	1800	0.008	0.9795	100	26 x 15	3dii
3	4K3E*	I04	2.91×10^{11}	0.1	3600	0.008	0.9795	100	26 x 15	dials
4	#30	I03	6.25×10^{11}	0.1	3600	0.004	0.9763	7.5	80 x 20	dials

2.8.3.5 X-Ray diffraction data processing pipeline

The molecular architecture of a sample is calculated by observations of XRD spots following the Arndt-Wonacott crystal rotation/oscillation method [211]. The number of diffraction spots are counted and resolution estimated on a per-image basis using the Diffraction Image Screening Tool and Library (DISTL) [212]. The diffraction intensity maxima ('spots') are indexed, scaled, and merged into a single reflections file. The X-Ray phase problem is resolved (either experimentally or by molecular replacement of a known crystal structure), producing an electron density map, into which atomic models are built and assessed.

2.8.3.6 Automated processing of XRD data

DLS has several parallel automated pipelines for processing of XRD data, including Fast DP [213], xia2 [214], and autoPROC [215]. Fast DP, xia2 and autoProc are run simultaneously and autonomously, with the results compared to select for the highest quality merged reflections data set, according to principal XRD metrics of: unit-cell completeness, spot uncertainty (R_{meas}) overall and over the outer-most (highest) resolution shell, expected precision of intensities 'I' (R_{pim}), signal-to-noise ratio (mean $I/\sigma(I)$), (expected) spacegroup, unit-cell parameters and data resolution.

Fast DP [213] offers a very rapid processing of XRD data by reading the header information from the first image, and then splitting the images into slices for parallel processing. Whilst Fast DP produces a result faster, useful for on-the-fly feedback, the limited data in each slice reduces the accuracy, and can result in disagreement between different slices. Consequently, other pipelines are preferred for downstream processing, but Fast DP is particularly helpful in providing an initial idea of crystal spacegroup within minutes of XRD data collection.

AutoProc [215] is a much more comprehensive pipeline. In brief, XRD images are analysed, spots are identified and indexed using XDS [216], and the beam centre is determined. AutoProc also checks for the presence of multiple crystallographic lattices. Diffraction quality and detector parameters are assessed, and initial unit-cell parameters are refined, as is the orientation and mosaicity of the crystal.

The most-likely spacegroup is selected by Pointless [217] analysis. All images are integrated, and the data are re-scaled with XScale from XDS [216] and merged with Aimless [218]. The autoPROC processing pipeline can be combined with STARANISO [219], which performs anisotropic cut-off of merged intensity data, Bayesian estimation of structure amplitudes, and applies an anisotropic correction to the XRD data. AutoPROC does not automatically generate the R_{free} control reflections dataset, so this will need to be generated separately (by programmes such as Aimless [218]) if autoPROC produces the best processing results.

xia2 [214] is another comprehensive processing tool implemented at DLS, designed for fully automated expert XRD processing. xia2 has several protocols, but DLS uses the DIALS and 3dii options. DIALS [220] offers a complete XRD data integration package covering spot finding, indexing, refinement and integration to produce background-subtracted, integrated intensities and associated error estimates. 3dii is similar, but forces xia2 to index identified peaks from all images using XDS and XScale [216]. xia2 also randomly selects R_{free} for 5 % of the reflections, generating the R_{free} control dataset.

2.8.3.7 Solving the X-Ray phase problem

Indexed, scaled, merged intensity data files were imported to CCP4 [221]. Where autoPROC [215] produced the highest quality indexed, scaled, merged data, Aimless [218] was separately performed to create the R_{free} dataset, otherwise the R_{free} dataset created by xia2 [214] was imported alongside the remainder merged intensity dataset. The Matthews correlation coefficient [222] was estimated for each crystal dataset, predicting the number of monomers in the asymmetric unit from the amino acid sequences. Molecular replacement was performed to solve for X-Ray phasing using Phaser [223]. Molecular replacement search models were identified by searching the PDB [162] by each antibody chain amino acid sequence, with the highest scoring model with the broadest sequence identity being selected as the search model. In the event PDB models gave identical E_{val} (the probability that matched sequences were matched by chance) and identity scores, the PDB validation metrics

were compared. All search models were manually inspected for clashes, disordered regions, and sensible geometries.

2.8.3.8 Crystallographic XRD model building, refinement and verification

Once a phasing solution had been identified, ARP/wARP [224–226] was used to build an initial protein model into the electron density, with mixed results. Alternatively, the molecular replacement solution was used, the placement of which was refined with 50 cycles of rigid body refinement using REFMAC5 [227] with a restrained resolution. Iterative refinements using REFMAC5 [227] and model re-building in COOT [228] were performed until satisfactory R_{work} and R_{free} statistics were achieved for each dataset's highest resolution shell without excluding any reflections, according to similar resolution structures in the PDB [162] (Table 2.11).

Table 2.11: Range of accepted R_{free} values of structures deposited in the PDB up to 2011 [162]**.** Adapted from Read *et al.* (2011), Figure 3C. [229]. As of July 2022, the mode R_{free} bin of all XRD structures in the PDB is 0.230 - 0.240, and the mode resolution bin is 1.8 - 2.0 Å.

		R_{free}	
Resolution (Å)	Best 25 %	Median	Worst 25 %
1.0	< 0.140	0.150	> 0.170
1.5	< 0.190	0.210	> 0.230
1.8	< 0.210	0.230	> 0.245
2.0	< 0.220	0.240	> 0.260
2.2	< 0.230	0.250	> 0.265
2.5	< 0.245	0.260	> 0.280
3.0	< 0.260	0.280	> 0.300
3.5	< 0.275	0.300	> 0.330
4.0	< 0.280	0.310	> 0.350

R_{work} describes the accuracy of the model in terms of explaining the observed data. Using R_{work} alone to measure the quality of a protein model is not recommended as the rebuilt model contributes to the recalculation of the associated electron density map, and therefore bias can be introduced. The R_{free} value, calculated using only randomly selected 5 % of reflections not used in model refinement, helps understand the level of bias introduced to the electron density map by comparing against the

R_{work} dataset. One would expect R_{free} to be a larger value than R_{work}, but the delta between R_{free} and R_{work} should not exceed 0.05. Both R_{work} and R_{free} are calculated as below, however remember that R_{free} uses a fraction of the data used by R_{work}:

$$R = \frac{\sum_{hkl} |F_{obs}(hkl) - F_{calc}(hkl)|}{\sum_{hkl} F_{obs}(hkl)}$$

XRD data constitutes the reflection spot intensities, which yield structural factor amplitudes with unique h, k, and l indices. $F_{obs}(hkl)$ and $F_{calc}(hkl)$ are the observed, and calculated structure factors, where $R = 0.0$ for perfect agreement, whilst a totally random model would produce $R \approx 0.59$ [230]. Consequently, both R_{work} and R_{free} are highly dependent on the observed data, which can be constrained by limiting the resolution; where small restrictions on data resolution can markedly improve R_{work} and R_{free} values [231]. Alternative metrics for model quality have been suggested that consider the number of unique reflections used to refine a structure per non-hydrogen atom in the asymmetric unit [231], but these were not used in this study. Instead, refinement was abandoned if the delta between R_{work} and R_{free} became greater than 0.05. Post-refinement model geometry was assessed using MOLPROBITY [232] within the CCP4 suite [221], and macromolecular interfaces (both biological and crystalline) were identified using PISA [233] available online on the ePDB (EMBL-EBI) website [234]. All figures were prepared using The PyMOL Molecular Graphics System v1.2r3pre (Schrödinger, LLC).

2.8.4 Single-particle cryogenic Electron Microscopy

Structural determination of aqueous samples can be achieved using cryogenic electron microscopy (cryo-EM) circumnavigating the challenges of protein crystallisation and and X-ray phasing, particularly if the sample of interest is not too small (roughly greater than 100 kDa to increase signal-to-noise ratio and to aid particle picking) and if a resolution higher than 2.0 Å is not absolutely required (the intrinsic flexibility and dynamism of samples in aqueous environments limits attainable resolution, though at the time of writing, 47 single-particle cryo-EM structures have been deposited in the PDB with a resolution of < 2.0 Å. The highest resolution

single-particle cryo-EM structure is the mouse apoferritin structure at 1.22 Å [235]. To achieve structural characterisation by cryo-EM, samples are immobilised on an ultra-thin film atop a supporting grid.

2.8.4.1 Available grid types

A number of commercial cryo-EM grid types are available within The University of Oxford's Division of Structural Biology (STRUBI), and the Oxford Particle Imaging Centre (OPIC). Quantifoil® Holey R2/1 Carbon Film on 300 Copper mesh Grid (Quantifoil) and Lacey Ultrathin Carbon film on Copper 400 mesh (Agar Scientific) gridtypes were screened for experiments involving VLPs, whilst C-Flat Ultra-thin Holey -1.2/1.3 Carbon film on Copper 200 mesh (Protochips) and Lacey gridtypes were screened for experiments involving bRSVF, due to anecdotal successes with SARS-CoV-2 Spike protein; a similar trimeric viral fusion protein. Additionally, self-wicking grids were used for automated grid vitrification, but this is discussed further in Section 2.8.4.4.

2.8.4.2 Grid glow discharge

The sample-binding surface of EM grids is an uncharged, hydrophobic surface, typically using a carbon film (though other substrates are available). Glow discharge treatment of carbon-coated grids with low-energy plasma generates radicals to react with and charge to the carbon surface, allowing it to take sample evenly with a uniform ice thickness [236]. Grids were deposited within the plasma chamber of a PDC-002-CE Plasma Cleaner Glow Discharger (Harrick Plasma). Vacuum was drawn by an RV3 Vacuum Pump (Edwards). A small amount of air was bled into the chamber before plasma was applied using the 'High' setting. Plasma was applied according to Table 2.12.

2.8.4.3 Semi-automated grid vitrification

Aqueous samples must be immobilised within ultra-thin vitreous ice to allow high-resolution cryo-EM [237]. Crystalline ice is antagonistic for cryo-EM imaging as crystalline ice can diffract the incident electrons, decreasing image quality.

Table 2.12: Per-gridtype glow discharge and sample application parameters.

Gridtype	Holesize (µm)	Spacing (µm)	Glow discharge (s)	Volume (µL)	Blot time (s)	Blot Force
Quantifoil®	2	1	30	3.0	6.0	+6.0
Lacey	NA	NA	30	3.5	6.0	+6.0
C-Flat	1.2	1.3	20	3.0	5.5	-1.0

Furthermore, crystalline ice dehydrates the specimen by withdrawing water molecules [238]. Ice thickness also needs to be considered, as electrons will be scattered by thick vitreous ice, reducing contrast, whilst very thick ice cannot be penetrated efficiently by electrons, reducing total signal and therefore impairing the image contrast. At the other end, the ice must still fully enclose the sample of interest, as sample distortion can occur at the ice-air interface [239], whilst samples can become denatured in very thin ice [240, 241]. Ideally, a layer of vitreous ice of uniform thickness equal to, or slightly above the specimen is produced. Ice thickness is controlled principally by blotting away excess liquid prior to immobilisation, however this wastes sample and risks sample integrity [242]. Sample vitrification is achieved by flash freezing within a liquid cryogen, with the speed of freezing preventing crystalline ice formation. Reproducible vitrification is achieved using a Vitrobot Mark IV automated vitrification system (ThermoFisher Scientific). Sample was applied to glow discharged cryo-EM grids (Table 2.12) within the Vitrobot chamber cooled to 4.5 °C with 100 % reported humidity. After a delay of 30 seconds, excess liquid was wicked away by blotting using Vitrobot Filter Papers (Whatman) according to Table 2.12, before being plunge frozen in liquid ethane. Once plunged, grids were stored under liquid phase nitrogen.

2.8.4.4 Automated grid vitrification

When a small sample volume is applied to a charged surface, it is easy for molecules to adhere in a preferential manner, especially when the sample is hydrophilic with more charged or polar regions exposed. *Ab-initio* reconstruction relies upon aligning particle projection images in 3D, in order to determine a 3D volume that best explains the particles as observed by their combined projections, however this means

the modelled volume is heavily dependent upon the 'pose' of the observed particles. When preferential orientation is observed, *ab-initio* reconstruction is constrained by signal dominating from a particular direction. Whilst not necessarily a bad thing, depending on the sample under study, preferential orientation can limit the resolution of single-particle cryo-EM specimens in a directional manner, and even lead to non-sensical *ab-initio* reconstruction, as seen with bRSVF (Section 5.5.1). Data of samples demonstrating preferential orientation may still be interpreted in some situations, but with caution due to potentially missing information. Therefore, reducing the delay between sample application and vitrification may immobilise sample molecules before they adopt a preferential orientation.

As introduced previously (Section 2.8.4.3), blotting is currently the gold standard technique in generating uniform vitreous ice, however blotting not only wastes the majority of applied sample volume, but introduces a time delay between application and vitrification. Recently, blotless techniques have been introduced to address these challenges, and are purported to benefit samples showing preferential orientation [243]. DLS and RFI are working in collaboration as user development partners for SPT Labtech in the development and technological maturation of one such technique, the Chameleon grid vitrification robot.

Chameleon combines a custom automated GloQube® (Quorum) glow discharger unit with a Spotiton piezo-electric transducer with a microcapillary channel dispenser [244] to deliver picoliter sample droplets to self-wicking nanowire cryo-EM grids (Fig. 2.4). Slow-motion video assessment of sample wicking is used to optimise glow discharge duration, applied sample volume, and wicking delay, and even provides a preliminary assessment of ice thickness [245–247]. Combined, these techniques reduce the time delay between sample application and vitrification to as low as 54 milliseconds.

300 Mesh Self-Wicking Copper grids (SPT Labtech) were used for this project. Hole spacing of chameleon grids are closer together (edge-to-edge spacing of 0.8 µm) than comparative non-self-wicking grids (edge-to-edge spacing of 1.0 µm for Quantifoil® and 1.3 µm for C-Flat grids), to increase the number of data

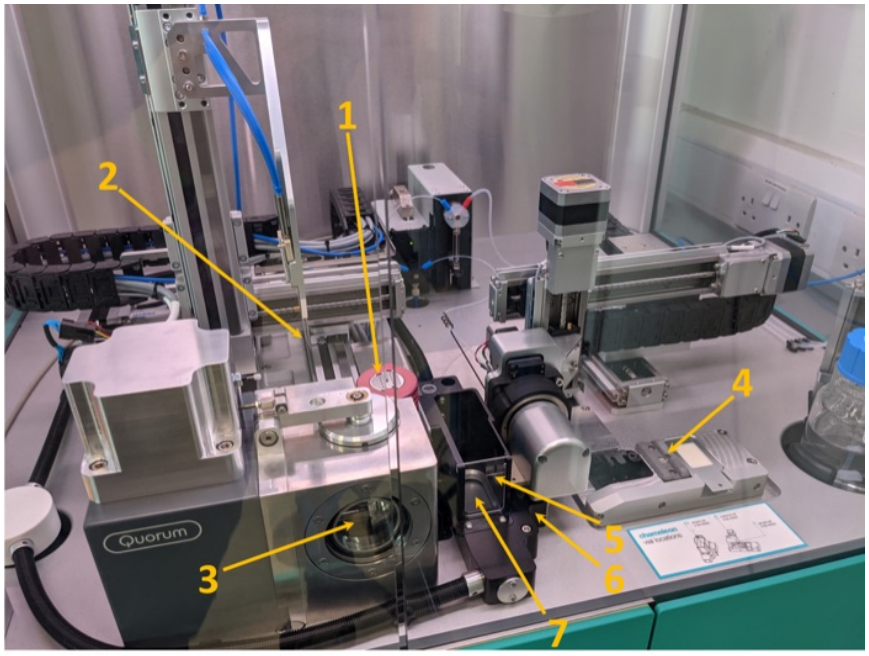


Figure 2.4: Chameleon gird vitrification robot (SPT Labtech) schematic. (1) Grid holder. (2) Robotic grid tweezers. (3) Automated GloQube® glow discharge unit (Quorum). (4) Temperature controlled specimen well. (5) Spotiton piezo-electric transducer picoliter sample handling robot [244]. (6) High-speed video module. (7) Sample application chamber; temperature and humidity controlled with trap door to liquid ethane reservoir for vitrification.

acquisition areas, compensating for the thin stripe of sample application versus the complete grid coverage in Vitrobot-prepared grids. During glow discharge, a current of 12 mA was used to generate plasma from air. Once plunged, grids were stored under liquid phase nitrogen.

2.8.4.5 Graphene Oxide coating of chameleon grids

Graphene oxide (GO) was applied to 300 Mesh Self-Wicking Copper grids (SPT Labtech) following the protocol described by Cheng and colleagues [248]. Briefly, 5 µL of Graphene oxide solution (2 mg/mL) (Sigma-Aldrich, Merck) were mixed with 5 µL 1.5 % (w/v) non-ionic detergent n-dodecyl-β-d-maltoside. The mixture was diluted 100 times with water, and 5 µL applied to the carbon film side of the grid. The GO was drawn through the grid by wicking the copper side of the grid with filter paper. Samples were applied to GO-coated grids on the same day as GO application to the cryo-EM grids.

2.8.4.6 Grid Clipping

All cryo-EM grids are mounted within copper AutoGrid/C-Clip Rings (ThermoFisher Scientific) with securing C-Clip (ThermoFisher Scientific) under vapour phase nitrogen to increase grid rigidity and aid manipulation, ensuring reliable mechanical manipulation by the electron cryomicroscope sample handling robotics. Clipped grids 'cartridges' were stored under liquid phase nitrogen until imaging. Prior to loading to an electron cryomicroscope cassette, cartridges were visually inspected for signs of grid damage, incorrect clipping, devitrification, or severe icing.

2.8.4.7 cryo-EM Data Collection

All cryogenic electron microscopy (cryo-EM) data were collected using either the Glacios or Titan Krios electron cryomicroscopes (ThermoFisher Scientific) within the Oxford Particle Imaging Centre (OPIC), or within the electron Bio-Imaging Centre (eBIC) at DLS (Table 2.13).

Table 2.13: Electron cryomicroscopes access throughout this project. For the eBIC data collection, the proposal number is provided between parentheses, rather than the STRUBI project number.

Chapter	Antigen	Fab	Collected	Instrument	Facility (Project)
4	SAT2 '12S	#117	16/04/2021	Titan Krios	STRUBI (P401)
4	SAT2 '12S	#126	27/06/2022	Titan Krios	STRUBI (P761)
4	SAT2 '75S	#117	02/09/2022	Titan Krios	STRUBI (P798)
4	SAT2 '75S	#126	03/09/2022	Titan Krios	STRUBI (P799)
4	O1M '75S	#117	29/06/2022	Titan Krios	STRUBI (P762)
4	O1M '75S	#126	03/09/2021	Titan Krios	STRUBI (P526)
4	O1M '75S	#30	01/03/2021	Titan Krios	STRUBI (P525)
5	bRSVF	B4	25/08/2021	Glacios	STRUBI (P524)
5	bRSVF	B4	28/09/2021	Glacios	eBIC (nt29493-27)
5	bRSVF	B4	05/08/2022	Glacios	eBIC (nt29493-76)
5	bRSVF	B4	10/08/2022	Glacios	eBIC (nt29493-78)
5	bRSVF	B4	22/09/2022	Glacios	eBIC (nt29493-86)
5	bRSVF	B4	21/10/2022	Titan Krios	eBIC (bi28713-31)

Glacios and Titan Krios are complementary high-resolution cryo-EM instruments owing to their shared multi-specimen cassette auto-loader, capable of storing 12

cartridges within an internal liquid phase nitrogen dewar within the cryomicroscope, therefore massively reducing time delay when switching between specimens (traditional side entry cryomicroscopes had no storage capacity and associated robotics, and therefore only a single cartridge could be loaded to the cryomicroscope at a time). Glacios and Titan Krios electron cryomicroscopes also share a similar objective optics design.

Titan Krios is usually operated at an accelerating voltage of 300 keV, whilst Glacios operates at 200 keV. Consequently, Glacios is often viewed as a screening microscope to select promising specimens for imaging with a Titan Krios. Cartridges can be inspected for ice thickness, vitrification preservation, integrity of Carbon film and rough particle count. Accelerating electrons to 300 keV has several benefits for high-resolution cryo-EM data collection over 200 keV; principally that the accelerated electron beam is less susceptible to inelastic scattering, so background noise can be reduced. Titan Krios additionally provides an energy filter to further eliminate electrons from elastic collisions, an additional condenser lens to increase the number of acquisition areas per cryo-EM grid hole due to smaller spot sizes, and has a more advanced stage with increased stability, for reduced global specimen motion during data collection over the Glacios. However, faster electron acceleration attracts more material disadvantages; the increased energy necessitates larger accelerating components and advanced safety equipment, combined with the additional optics and stage components, requires a significantly larger microscope housing for the Titan Krios. Additionally, the more advanced components attracts an inflated manufacturer cost, and furthermore, Titan Krios operational costs are significant, limiting accessibility of 300 keV cryomicroscopes.

Once a cassette is loaded to a cryomicroscope, a low-magnification 'atlas' of relative ice thickness is generated for each grid using EPU (ThermoFisher Scientific). Automated data collections were set up using EPU, selecting only promising grid holes (deselecting grid squares with damaged carbon film, and grid holes with large debris or disadvantageous ice) at least one grid square's distance inside the cartridge perimeter (due to physical stage limitations). Defocus range was selected

as −0.5 µm to −3.5 µm, with −0.5 µm being the minimum defocus that particles were still visible, and an objective aperture of 100 µm were used to increase the contrast of collected cryo-EM images.

Once microscope optics are set, several thousand cryo-EM movies are recorded from different loci about the sample grid, adjusting for the eucentric height; the specimen height in the optic axis, of each grid square. After correcting for any motion (Section 2.8.4.9) and estimating the specimen-specific contrast transfer function, (CTF) parameter (Section 2.8.4.10), particles are picked and classified (Section 2.8.4.11). Where sufficient particles were selected, an *ab-initio* reconstructed volume was calculated (Section 2.8.4.12) from the selected particle classes to provide an unbiased initial model for subsequent 3D classification, refinement (Section 2.8.4.13) and model building (Section 2.8.4.14).

2.8.4.8 Processing of single-particle cryogenic electron microscopic datasets

Data from STRUBI's Titan Krios cryomicroscope computer were transferred to Oxford's Biomedical Research Computing (BMRC) file system on-the-fly for cluster processing. cryo-EM single-particle *ab-initio* reconstruction and classification (cryoSPARC) v3.3.2 (Structura Biotechnology) [249] offers an intuitive, user-friendly graphics processing unit (GPU)-based approach to single-particle cryo-EM data processing, and is capable of real-time pre-processing of cryo-EM data (motion correction, CTF correction, particle picking and initial *ab-initio* reconstruction) using the 'Live' web interface, and intuitive downstream cryo-EM data processing using its 'static' web interface.

cryoSPARC Live pre-processing is used for producing a compiled particle stack, and principally involves motion correction, CTF estimation, exposure curation, particle picking and extraction. Once a particle stack of interest has been compiled that fits the anticipated sample density, cryo-EM processing is continued using the cryoSPARC Static interface.

2.8.4.9 Motion Correction

When exposed to the electron beam of an electron cryomicroscope, unavoidable, non-uniform beam-induced specimen translational movement is observed [250–252]. This motion diminishes attainable resolution by approximate Gaussian blurring [253]. High-frame rate direct electron detecting detectors (such as the Falcon IV (ThermoFisher Scientific) now installed on STRUBI's Titan Krios and capable of acquiring data at 320 fps) allow for the identification and correction of beam-induced motion. Since beam-induced motion is heterogeneous across the field of view, smaller image patches are cross-correlated to estimate local motion. Once the velocity field (the collective patch velocities per cryo-EM movie) has been assessed, frames are aligned and averaged in a dose-weighted manner to produce a single locally motion corrected 'micrograph' [254]. This method allows for correction of both local and global motion, and also allows rejection of individual frames showing extreme motion.

2.8.4.10 Contrast Transfer Function Estimation

Single particle cryo-EM specimens usually demonstrate only a weak signal to noise ratio, and consequently, direct illumination of a specimen with an electron beam is not sufficient to produce images clearly contrasting sample against background. To negate this, cryomicroscopes are operated in 'phase-contrast' mode, wherein the path lengths of electrons elastically scattered by the sample will vary dependent on the spatial frequency of the sample's electrostatic potential, introducing a phase-shift to the refracted electron beam wavelength. The interference of phase-shifted electrons (with one another) is modelled by the cryomicroscope's contrast transfer function (CTF), a parameter influenced by the incident electron beam wavelength, λ_{e}, the sample spatial frequency, f, the microscope's spherical aberration, C_{s}, a phase shift factor, ϕ, and dominantly, the sample defocus, Δz. In cryoSPARC, this is modelled as below:

$$CTF = -cos\left(\pi \Delta z \lambda_e f^2 - \frac{\pi}{2} C_s \lambda_e^3 f^4 + \phi\right)$$

Within cryoSPARC, multiple patch-based CTF estimations are automatically performed, to account for intra-micrograph differing eucentric height (and therefore defoci and CTF) caused by sample localisation to the air-water interface prior to freezing, and ice surface non-planarity (or mosaicity) and heterogeneity [241]. The CTF fit is assessed graphically by reviewing the CTF fit against the 2D power spectrum of the micrograph within the cryoSPARC Live interface. Exposures were then curated within the cryoSPARC Live interface, setting thresholds for total motion, maximum in-frame motion, defocus range, CTF fit, and relative ice thickness.

2.8.4.11 Particle Picking

Particles were then picked by default using the blob-picking module within cryoSPARC Live, which picks based on user-anticipated minimum and maximum particle diameters. The picks are filtered using Normalised Cross-Correlation and Power thresholds to filter out noise whilst leaving as many specimen particles as possible. Particles can also be picked manually if the blob-picker fails to correctly identify particles. On-the-fly 2D classification is run on the picked particles, the output of which was used as a template for template-based particle picking.

Topaz deep picking offers a machine learning approach to the particle picking challenge [255], and may enrich the number of good particle picks at the expense of time and convenience as Topaz is run on the cryoSPARC static interface, however Topaz requires a very high quality initial particle training set without missing any conformations or perspectives to train the neural network, and is consequently best used after several iteratively improved rounds of template-picked particles [255]. Topaz deep picking was used for the SAT2+Fab #126 dataset, due to the increased background and pre-processing challenges met (Section 4.6.6). The template-picking module was sufficient to pick enough *good* particles for the rest of the cryo-EM data collections. Finally, the picked particles are extracted using a box size for optimal processing.

2.8.4.12 *Ab-initio* 3D reconstruction

Extracted particles were imported to the cryoSPARC static interface. 2D classification is repeated once or more to produce a good quality set of particle classes whilst eliminating distracting junk, collapsed or aggregated particles. For all cryo-EM datasets, particles were classified into 50 particle classes. Good classes were decided based on particle count, indicated resolution and Effective Classes Assigned metrics provided. For the FMD VLP datasets, particle classes were only retained if they contained intact particles positioned at the centre of the class, if they appeared to show high resolution particle information, and in the case of VLPs mixed with Fabs, appeared to show Fab decoration.

Ab-initio 3D reconstruction from 2D particles represents a computational optimisation challenge, as programmes can easily get 'stuck' in incorrect optima, consequent to improper initialisation [249, 256]. cryoSPARC employs a stochastic gradient descent (SGD) approach to reference-free *ab-initio* 3D reconstruction wherein an arbitrary initial volume is used as a start model [249]. This model is assessed for how well it explains the observed particles, using a small randomised subset of particle images from the selected 2D classes. The optimisation equation for maximal likelihood of correct 3D reconstruction includes terms for class assignment, particle pose, and individual contribution of a particle to the overall likelihood of the reconstruction. SGD repeatedly approximates this equation using different random subsets of particles, updating the optimisation variables, and consequently the 3D map, with each iteration, dependent on the gradient. Consequently, this noisy SGD approach is insensitive to incorrect local optima, often identifying the correct 3D reconstruction solution [249]. Once an approximate map is achieved, a branch-and-bound algorithm is used for 3D image alignment with respect to the approximation, aiding downstream map refinement, and allowing a quick assessment of any preferred orientation, based on the viewing direction distribution. *ab-initio* reconstruction was run without a specified symmetry group to avoid introducing bias, and with the goal of producing 3 distinct classes, for use in a subsequent heterogeneous refinement step.

2.8.4.13 Map refinement

An important step in high resolution structural determination is the identification of distinct sample conformations within a particle dataset. The n outputted *ab-initio* 3D reconstructions are further refined by Heterogeneous Refinement. Heterogeneous refinement simultaneously reclassifies particles and refines the n *ab-initio* reconstructions, allowing identification of any distinct conformations within the dataset, and separating their associated particles. For Fab with intact FMD VLP samples, a symmetry of I2 was enforced, whilst for Fab with dissociated 12S SAT2 pentamers, C5 symmetry was enforced. C1 symmetry was imposed for the initial attempt at observing bRSVF in complex with Fabs, then C3 symmetry was imposed when processing the final high resolution structure. To limit GPU memory usage, particles are Fourier cropped (downsampled). All refined maps were inspected for refinement and reclassification artefacts (such as noise and streaks), sensible geometry and broadly sensible size for the anticipated sample using Chimera v1.15 (Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (UCSF)) [257].

In Fab-antigen interactions, it was assumed that Fab demonstrates some free rigid body motion relative to the rigid body of the larger antigen at a quaternary structural level (though more local flexibility to solvent-exposed loops is also anticipated). All the antigens in this study were anticipated to show symmetric native structures [176, 182, 183, 258, 259], with the FMD VLP capsid giving a broadly-smooth, near-spherical structure, and bRSVF giving a ternary 'lollipop' structure. Consequently Fab rigid body flexibility could get averaged out in subsequent refinement programmes. Dependent on the presence of preferred Fab pose(s) distributed about the 60 possible interaction sites on FMD VLPs, or three possible interaction sites on bRSVF, the cryoSPARC Local Refinement job can improve the particle alignment by performing and refining angular and translational searches within a user-defined masked volume, aiming to 'lock' on to the preferred Fab pose(s) that may otherwise get averaged out. In this instance, a rotational search of 5° and a translational shift search of 2.0 Å was applied to improve the particle

alignment, using a dynamic 'mask' volume of the Heterogeneous Refinement output volume. Symmetry was preserved from the preceding Heterogeneous Refinement job. Local Refinement refines the region enclosed within the mask, so in this case, the entire particle volume is slightly refined, albeit with modest improvement. Similar to XRD electron density map refinement (Section 2.8.3.8), to test against introduced bias, the dataset is randomly split, however in cryo-EM processing, a 50:50 split is created with each 'half map' produced by each half of the particles respectively instead of the 5 % subset of reflections for R_{free} calculation in XRD processing. Each half of the data is independently processed, with the correlation between the maps calculated. The final outputted map is the combination of these half-maps.

During the prior Heterogeneous Refinement job, particles were downsampled to reduce computational resource demand, thus increasing computational speed, however such downsampling severely limits the resolution of the data. To resolve this, the cryoSPARC Extract from Micrographs job is performed to re-extract particles using the full box size initially selected during the Particle Extraction. Using this method, the refined local particle alignments are preserved, resulting in a well aligned, full scale particle stack.

Following particle re-extraction, a Non-Uniform Refinement job is performed, which iteratively accounts for local regions of disorder or flexibility, which would otherwise reduce resolution. The loci of disordered regions are estimated, improving particle alignment further and improving downstream refinement outcomes by reducing the tendency for refinement algorithms to over-fit disordered regions [260]. Refinement programmes use regularisation by Fourier Shell Correlation (FSC) to penalise unnecessary model complexity, hence avoiding over-fitting, however FSC regularisation smooths a structure universally, and thus resolvable information is lost for some regions, whilst noise is introduced at others. cryoSPARC's Non-Uniform Refinement uses adaptive regularisation with a cross-validation approach to regularisation parameter optimisation for yielding improved map quality when one expects varied structure resolution [260].

Several other techniques have been tried within cryoSPARC to further refine map quality of all datasets beyond this initial processing pipeline, with varying success. The 3D Variability job assesses discrete and continuous heterogeneity, wherein particle images are used to create a continuous family of conformations, rather than trying to converge to a single conformation. However 3D Variability did not reveal the presence of such heterogeneity in the cryo-EM datasets produced within this project. Another tool is the Symmetry Expansion job, which creates a new particle stack by observing particle poses around the point group symmetry and duplicating into a single particle stack, with the aim of resolving symmetry-breaking features. 3D classification may then be run to separate out heterogeneity in the Symmetry Expanded particle stack. Symmetry expansion was trialled for all datasets with mixed results (Section 4.6.2).

Local refinements were repeated using iteratively improved masks produced with Chimera [257]. To prepare each mask, a model of Fab interacting with FMD VLP was produced by coarse fitting crystallographic XRD structures (where available) to the map, followed by the Fit in Map tool to optimise the positioning. As neither Fab117 nor Fab126 produce data collectable crystals, the structures of these were predicted using AF2 (Section 2.8.2 and Section 4.4.3). The cross-correlation of fitting the density of each of the AF predictions to the map were compared, with the highest scoring model selected for mask building. The map was split using the Color Zone method [257]. Both the map fraction corresponding to a region of interest to be locally refined, the 'refinement mask', and the residual map, 'residual mask', were uploaded to cryoSPARC by the Import 3D volume job as masks, holes filled and externally dilated by 10 Å, and soft padding added to 15 Å by the Volume Tools job. Mask dilation and soft padding allows inclusion of regions of density at the edges of the mask, to avoid symmetry clashes during model building. A Local Refinement is run to create the particle alignment. Following this, unwanted signal from the particle density corresponding to the residual mask was selected, convolved using the CTF, and subtracted from each particle image using the Particle Subtraction job. Local Refinement focused on the masked region of interest was

then performed to the subtracted particle stack. For each round of refinement, the symmetry group, rotational and shift search parameters were preserved from previous Local Refinement jobs, and the fulcrum of refinement was centred on the Fab-VLP interface within the refinement mask, using feedback from the Real Space Slice figures to iteratively assess fulcrum positioning on-the-fly.

To aid visualisation in post-processing, the Local Resolution Estimation job is used in combination with Local Filtering to present a noise-free map coloured by resolution, aiding interpretation.

2.8.4.14 cryo-EM Model Building, Refinement, and Validation

The cryoSPARC Local Resolution Estimation job can be used to appropriately filter the map with subsequent Local Filtering job in readiness for model building. The cryoSPARC Sharpening Tool was also used to prepare maps for model building by weighting the map B-factors according to the Guinier plot of the preceding refinement job.

Model building began in much the same way as mask creation (Section 2.8.4.13). Crystal structures (or AF models) of the VLP and Fab were approximately fitted 'by hand' to the map in Chimera [257], and positioning optimised by the Fit in Map tool. These models were imported to Coot [228] with the map for iterative rigid body refinement and model building. Models were validated using MOLPROBITY [232] within Phenix v1.19.2-4158-000 [261].

Figures were produced using Chimera v1.15 [257], ChimeraX v1.1.1 [257](Resource for Biocomputing, Visualization, and Informatics at UCSF), and The PyMOL Molecular Graphics System v1.2r3pre (Schrödinger, LLC).

Exegi monumentum aere perennius regalique situ pyramidum altius, quod non imber edax, non Aquilo impotens possit diruere aut innumerabilis annorum series et fuga temporum. non omnis moriar, multaque pars meivitabit Libitinam...

I have completed a monument more eternal than bronze, more lofty than the regal structure of the pyramids, one which neither corroding rain, nor the ungovernable north wind can ever destroy, nor the countless series of the years, nor the flight of time. I shall not wholly die, and a great part of me shall evade Libitinam...

— Horace's *Odes*

3

Chain exchange of ultralong antigen-unknown antibodies

Contents

3.1 Immunoglobulin chain-exchange 92
3.2 Challenges of *in silico* approaches 92
3.2.1 ABodyBuilder *(formerly SAbPred)* 93
3.2.2 AlphaFold 95
3.2.3 Summary of *in silico* challenges for the prediction of unusual antibodies 96
3.3 Antigen-unknown antibody sequence discovery 96
3.4 Transient transfection of antigen-unknown ultralong Fabs 98
3.4.1 Fab molecular biology preparations 98
3.4.2 Ultralong Fab expression screening 99
3.4.3 Purification of Fab D08 100
3.5 Crystal structure of naturally paired ultralong Fab D08 101
3.5.1 Crystallography 101
3.5.2 X-Ray Diffraction data collection for Fab D08 102
3.5.3 Crystallographic structure of Fab D08 108
3.5.4 PISA crystal packing analysis 109
3.5.5 Performance of VMXi for structural determination . . . 111
3.6 Chain-exchanged ultralong fragment antibody structures 116
3.6.1 Purification of chain-exchanged ultralong antibodies . . 116
3.6.2 Crystallography 118
3.6.3 Crystallographic structure of ultralong chain-exchanged Fab D08* 123
3.6.4 Pairing-mediated twist to the ultralong stalk 123
3.6.5 Pairing-mediated ultralong paratope warping 124

3.6.6 Crystallographic structure of ultralong chain-exchanged Fab 4K3E* 128
3.6.7 PISA crystal packing analysis 134
3.7 Assessment of AlphaFold 2 prediction accuracy 136
3.8 Discussion 140
3.8.1 Ultralong antibody D08 structure 140
3.8.2 Proposed immunological significance of specific ultralong heavy-light chain pairs 140
3.8.3 Suggestions for further analysis 142

Figures

3.1 SAbPred predicted CDR3H conformations for bovine antibodies B4 and B13 94
3.2 Performance of ABodyBuilder 2 predicted CDR3H conformations for ultralong bovine antibody D08 95
3.3 chain-exchanged ultralong Fab expression screening by western blot 99
3.4 Gel filtration of Fabs D08 101
3.5 Crystal morphology for Fab D08 103
3.6 Screening UV absorbance of Fab D08 crystal hits 104
3.7 Experimental strategy for screening with VMXi 104
3.8 X-Ray diffraction data collection for Fab D08 at I04 106
3.9 Crystal structure of the ultralong Fab D08 107
3.10 D08 atomic model geometric validations 108
3.11 Aromatic stacking supports the D08 CDR3H stalk domain . . . 109
3.12 PISA interfacing residues of the D08 ultralong loop 111
3.13 DISTL per-mage analysis for VMXi-collected datasets 112
3.14 X-Ray diffraction data collection for Fab D08 at VMXi 113
3.15 Comparison of antibody D08 structures at cryogenic and ambient temperatures 115
3.16 Gel filtration of Fabs D08* and 4K3E* 117
3.17 Crystal morphology for Fab D08* and 4K3E* 120
3.18 X-Ray diffraction data collection for Fab D08* at I04 121
3.19 Crystal structure of the ultralong Fab D08* 122
3.20 D08* atomic model ramachandran plot 123
3.21 Disulphide network organising CDR3H knobs of Fab D08 and D08* 124
3.22 Aromatic stacking supports the D08* CDR3H stalk domain . . 125
3.23 Packing of CDR3L in Fabs D08 and D08* 126
3.24 Comparison of Fab D08 and Fab D08* knob regions 127
3.25 X-Ray diffraction data collection for Fab 4K3E* at I04 129
3.26 Crystal structure of the ultralong Fab 4K3E* 130
3.27 Molecular interactions affecting the 4K3E* CDR3H stalk domain 131
3.28 Elbow Angles for Fabs D08, D08*, 4K3E*, and 4K3E 132
3.29 4K3E* atomic model ramachandran plot 133
3.30 PISA interfacing residues of the D08* ultralong loop 134

3.31 PDB Sequence coverage depth for scFv D08 137
3.32 Validation metrics for AF2 modelled scFv D08 structures . . . 138
3.33 Predicted scFv D08 models versus experimental crystal structure 139
3.34 Disulphide networks supporting ultralong CDR3H 141

Tables

3.1 Primers used in bovine antibody sequence discovery 98
3.2 PISA analysis for Fab D08 110
3.3 Individual and merged VMXi crystallographic data table for ultralong antibody structures 114
3.4 Synthetic chain-exchanged Fab purification metrics 118
3.5 Crystallographic data table for antigen-unknown ultralong antibody structures 133
3.6 PISA analysis for Fab D08* 135
3.6 PISA analysis for Fab D08* 136
3.7 Single chain variable fragment (scFv) sequences for antibodies D08 136

3.1 Immunoglobulin chain-exchange

Cattle demonstrate preferential immunoglobulin chain pairing amongst their ultralong antibodies, perhaps consequent on the need for structural accommodation of the large CDR3H (Section 1.4).

Such ultralong heavy chains ubiquitously pair with light chains encoded by IgLV1-47 [31] or IgLV1-67 [61], which present a highly conserved CDR3L, in both sequence (Fig. 1.7) and structure (Fig. 1.8). As previously discussed, the high sequence similarity of bovine light chains could be interpreted as providing a mere structural role (Section 1.4). On the flip side, chain-exchanged antibodies (synthetic pairings of a heavy chain from a parent antibody and the light chain of a second, distinct parent antibody) have recently shown that bovine light chain selection can impart a significant structural rearrangement to the CDR3H (Section 5.1.6) with subsequent loss of function (Section 5.1.5) [160].

This bovine chain-exchange study used non-ultralong antibodies, hence the likely impact of chain-exchange, or indeed alternative heavy-light chain pairing, on ultralong antibodies is unknown. The effect of alternative light chain CDR3L sequence on ultralong CDR3H conformation cannot be modelled from deposited ultralong antibody structures, as until 2020 [84], all reported ultralong antibody structures shared an invariant light chain CDR3L sequence of `ASAEDSSSNAV` (Fig. 1.7). Since 2020, the additional two ultralong antibody structures contain two substitutions to the CDR3L, but the loop conformation is conserved [84, 85]. The addition of Pro91L in both NC-Cow1 (PDB ID: 6OO0) [84] and R50 (PDB ID: 7D3M) [85] actually encourages tighter packing of the CDR3L through the introduction of a *cis*-peptide (Fig. 1.8). Clearly, more work is required to elucidate the restraints on this region.

3.2 Challenges of *in silico* approaches

It would be sensible to begin by attempting *in silico* homology modelling to answer the question of individual chain contribution to ultralong antibodies due to the

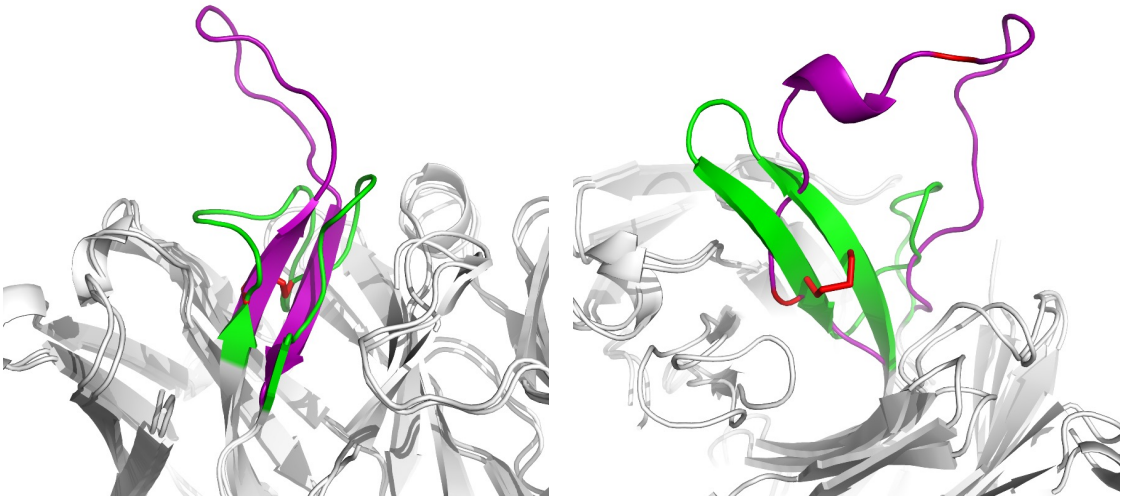
ease, speed, and availability of free software, however this is met with several challenges for antibodies with extended CDR3H loops [262]. ABodyBuilder 2 *(formerly SAbPred)* [194], is an in-house bioinformatic tool for the prediction of antibody paratopes, which, at the onset of this project had comparable accuracy to other publicly available prediction software. Briefly, ABodyBuilder 2 uses homology modelling against a database of known antibody structures, SAbDab [197], with *ab initio* modelling by MODELLER [200] for CDR loops with little homology to those within SAbDab to predict antibody structures. Subsequently, in 2021, DeepMind's AlphaFold v2.0 (AF2) was made available to the general public [204]. AlphaFold v2.0 uses neural networks trained on protein structures and biophysical information obtained from the extensive PDB repertoire [162], and outperforms all protein structure prediction programmes [203].

3.2.1 ABodyBuilder *(formerly SAbPred)*

In the first instance, the performance of SAbPred [194] was measured on the bovine antibodies B4 and B13 [161], and compared against crystal structures [160] (Fig. 3.1 and Section 5.1.6). B4 and B13 have CDR3H loops of 19 and 23 residues respectively, and therefore provide a modest challenge for antibody paratope prediction software, optimised for shorter length CDR3H loops.

SAbPred predicted a long, simplistic conformation for the CDR3H of both B4 and B13 (Fig. 3.1); with a long β-strand for B4, and a highly flexible, or largely disordered, CDR3H for B13. In both antibodies, MODELLER [200] failed to predict the internal disulphide bond that restrains the observed conformations of these loops.

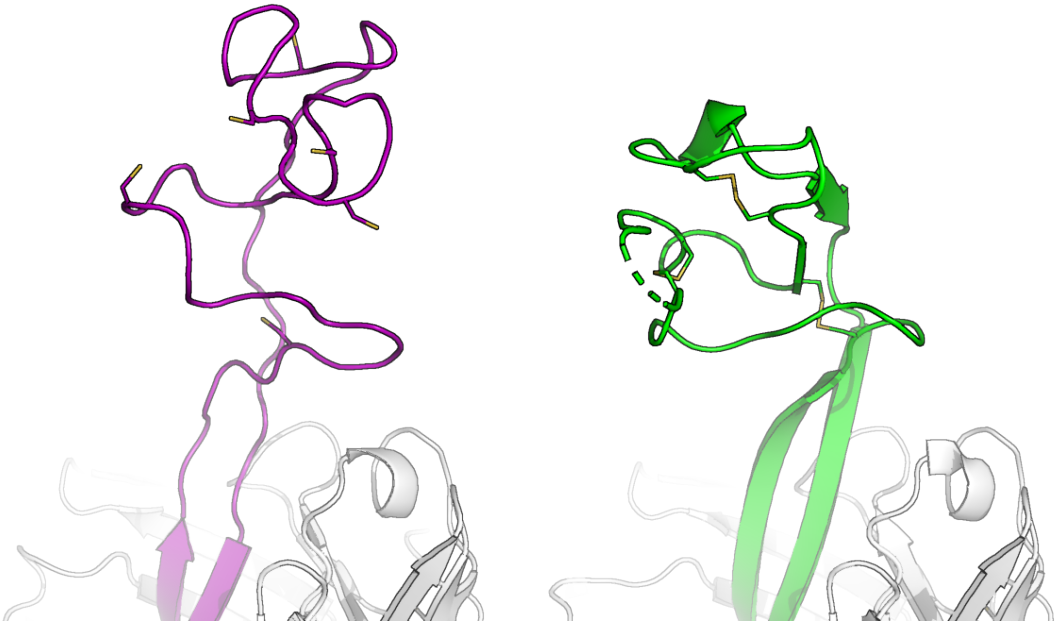
SAbPred was also trialled for the antigen-unknown ultralong antibody D08 of Section 3.3, however this initially failed due to the fact the CDR annotation tool ANARCI [196] only accepted CDR3H loop lengths of <32 residues, highlighting the unique nature of bovine ultralong antibodies, or perhaps reflecting the difficulty of ultralong loop modelling at the time. SAbPred [194] has since been updated (as ABodyBuilder 2) to accommodate ultralong submissions.



(a) B4 **(b)** B13

Figure 3.1: SAbPred predicted CDR3H conformations in cartoon representation for antibodies B4 (3.1a) and B13 (3.1b). Observed CDR3H conformation from crystal structures (PDB IDs: B4: 6QN9, B13: 6QN8) [160] are shown in green. SAbPred [194] predicted CDR3H conformations in purple. Cysteines are coloured red, with disulphide bonds shown in stick notation. Fv regions coloured white for clarity.

As with non-ultralong bovine antibodies (Fig. 3.1), ABodyBuilder 2 [194] struggles with cysteine handling, and fails to predict any disulphide linkages within the CDR3H (Fig. 3.2). Consequently, the predicted knob conformation is inaccurate, despite correct identification of the knob initiator motif 'CPDG', and the three-pass β-sheet at the centre of the knob region. However, prediction of such structures is likely artefactual due to the inclusion of ultralong knob mini-domains to the SAbDab database [197], rather than a knowledge-based prediction using the conformational restraints imposed by the internal disulphide linkages. Even if inclusion of the internal disulphides can be accommodated, prediction of which cysteine is bonded to which other cysteine is difficult, owing to the multiple theoretical possibilities and multitude of mini-domains with internal disulphide bonds deposited in the PDB [162]; comprising the bovine ultralong antibody structures [78, 82–85], and the cysteine knot proteins [90–131].



(a) (b)

Figure 3.2: Comparison of ABodyBuilder 2 predicted ultralong CDR3H conformation (3.2a) against observed crystal structure (3.2b). Cartoon representation for antigen-unknown ultralong antibody D08 (3.3). Observed CDR3H conformation from crystal structure is shown in green. ABodyBuilder 2 [194] predicted CDR3H conformation in purple. Cysteine sulphurs are coloured orange, with disulphide bonds shown in stick notation. Fv regions coloured white for clarity.

3.2.2 AlphaFold

As introduced earlier (Section 3.2), AlphaFold 2 has recently been made available to the public [204], after significantly outcompeting competitors in the 2016 and 2018 Critical Assessment of protein Structural Prediction (CASP) competitions [203]. AF2 predictions are covered in depth in Sections 3.7 and 4.4.3.

Critically, in the context of accurate paratope prediction for ultralong antibodies, AF2 performed similarly to ABodyBuilder 2, with accurate predictions of the knob initiator 'CPDG' motif and the presence of a three-pass β-sheet at the core of the knob mini-domain. AF2 was able to predict disulphide linkages within the CDR3H knob region (not shown in this thesis), but models did not converge on a single disulphide landscape, producing highly varied knob mini-domain structures.

However, AF2 did produce two models (out of five) with the correct predicted disulphide linkages, with an overall conformation similar to that observed in the D08 crystal structure (Fig. 3.33), however these were not the most confident predictions from AF2, showing caution is still required in the interpretation of AF2 predictions.

3.2.3 Summary of *in silico* challenges for the prediction of unusual antibodies

Overall, ultralong antibody structures are difficult to predict for several reasons. There are technical limitations such as the inability to accept and annotate ultralong loop lengths, challenges in the recognition and prediction of disulphide linkages internal to the CDR3H, and limited structural information of the knob mini-domain to infer inter-residue chemical restraints. Furthermore, the biological complexity and diversity of knob mini-domains further complicates the accurate paratopic prediction, despite recent advances in artificial intelligence protein structural prediction solutions such as AF2 or ABodyBuilder 2.

3.3 Antigen-unknown antibody sequence discovery

To investigate the structural impact of light chain selection on ultralong CDR3H architecture, a naturally occurring ultralong antibody was sought and sequenced.

Antibody-secreting cells (ASCs) were raised during an immunisation experiment to develop a novel low-throughput method to sequence paired bovine antibody chains (Ramirez-Valdez, personal communication). Briefly, bovine immunoglobulin variable domain forward primers, and constant region reverse primers specific to IgM, IgG, and Igλ were designed using the Bovine IgH 202 bacterial artificial chromosome (BAC) assembly [34] to cover all V gene segments (Table 3.1). Peripheral blood mononuclear cells (PBMCs) isolated from blood samples from an animal receiving adjuvant only. Viable, bovine CD3^{-} CD8a^{-} IgL^{low} ER^{high} single B cells were sorted by flow cytometry. CD3 is specific for T cells, whilst CD8a is specific for NK cells. IgL^{low} ER^{high} consequently gates specifically for antibody secreting B cells with

high levels of protein expression. Antibody heavy and light chains were reverse transcribed and amplified using the OneStep RT-PCR Kit (Qiagen) followed by a semi-nested PCR ('PCR-II' in Table 3.1), and amplicon cloned into pGEM-T Easy Vector (Promega), which were further Sanger-sequenced (Eurofins). An ultralong antibody, 'D08', with a CDR3H of 65 amino acids was observed, with the heavy chain encoded by IgHV1-7*02, IgHD8-2*01, IgHJ2-4*01, and the light chain encoded by IgLV1-47*01 (or IgLV1-52*01), and IgLJ4*01 (or IgLJ7*01, or IgLJ8*01). The ambiguity in identifying gene segment usage is resultant of the high resemblance between germline gene segments (Section 1.3.2). At 65 residues, the ultralong CDR3H is the same length as the CDR3H's of BOV-3 and BOV-7 ultralong antibodies previously reported (PDB ID's: 6E9H and 6E9U respectively) [83]. Also investigated, but not included in this thesis for brevity, were eleven non-ultralong antibodies demonstrating pairing promiscuity in either chain, described in more detail in Appendix C.1.

Table 3.1: Gene-specific primers used in amplification of bovine antibody sequences.

Primer	Primer Sequence	Orientation	RT-PCR	PCR-II
IgG_C3	GGCACCCGAGTTCCAGGTCA	RV	X	
IgG_C1	GGGTAGACTTTCGGGGCTGTGGTGG	RV		X
IgM_C2	CGAGCTCACGCAGGACACCA	RV	X	
IgM_C1	CTTTCGGGTGTGATTCACCT	RV		X
IgH_V1	ACTGTGGACCCTCCTCYTKGTGY	FW	X	X
IgH_V2	ACTGTGGACCCTSSTCTTTGTGC	FW	X	X
IgH_V3	ACTGTGGRCYCTCCTCTTTGTGC	FW	X	X
IgH_V4	VYYGTGGACCCTCCTCTTTGTGC	FW	X	X
IgH_V5	ACTGTGGACCCTCCTCTTTVTVC	FW	X	X
IgL_C2	CCGTTGAGCTCCTCCGTGGAG	RV	X	
IgL_C1	CGAGGGTGSGGACTTGGGCTGAC	RV		X
IgL_V1	CAGGCTSYACTGACTCAGCCR	FW	X	X
IgL_V2	CAGVCTGKSCTGACTCAGCCK	FW	X	X
IgL_V3	CAGGMTRTGCTGACKCAGCCG	FW	X	X
IgL_V4	CAGGMTVTRCTGACTCAGCCG	FW	X	X
IgL_V5	CAGGMTRTGCTGACKCAGCCG	FW	X	X
IgL_V6	CAGGCTGGYCTGACTCAGCCG	FW	X	X
IgL_V7	CAGGCTGTVCTRACBCAGCCG	FW	X	X
IgL_V8	CAGGCTGTGCTKRCTCARCCG	FW	X	X
IgL_V9	CAGGCTGTKYTGACTCAGCCR	FW	X	X
IgL_V10	CAGGGTGTGCTGACTCAGCCR	FW	X	X
IgL_V11	TCSTATGAACTGACMCAGYYG	FW	X	X
IgL_V12	TCTTCTCARCTGACTCAGCCG	FW	X	X

3.4 Transient transfection of antigen-unknown ultralong Fabs

3.4.1 Fab molecular biology preparations

The observed antibody D08 nucleotide sequence is provided in Table A.1. Retrieval of paired immunoglobulin sequence data has only recently become commonplace, and so until 2020, all ultralong antibody structures used heavy chains identified through single-chain sequencing [78, 82, 83] synthetically paired with a common light chain previously identified as pairing with ultralong heavy chains [31].

To investigate the impact of specific light chain sequences on the ultralong CDR3H structure, the D08 heavy chain was synthetically paired with the light

chain used in all pre-2020 reported ultralong Fab structures, creating a synthetic Fab named 'D08*'. A second synthetic pairing was also investigated, pairing the heavy chain from ultralong antibody 4K3E [78] (which donated the light chain for pre-2020 ultralong structures) with the D08 light chain, named '4K3E*'. Structural comparisons can be made between the structures of these hybrid chain-exchanged Fabs, and their parental, naturally-paired antibodies D08 and 4K3E. Candidate genes were codon optimised Table A.1, and inserted into the pOPINBOVH and pOPINBOVL expression vectors [160] (Section 2.2.1), for expression as recombinant Fabs. Correct constructs were propagated by Midiprep (Section 2.2.2). In total, 15 additional non-ultralong antibodies sourced from the sequencing optimisation study were explored, discussed in more detail in Appendix C.1.

3.4.2 Ultralong Fab expression screening

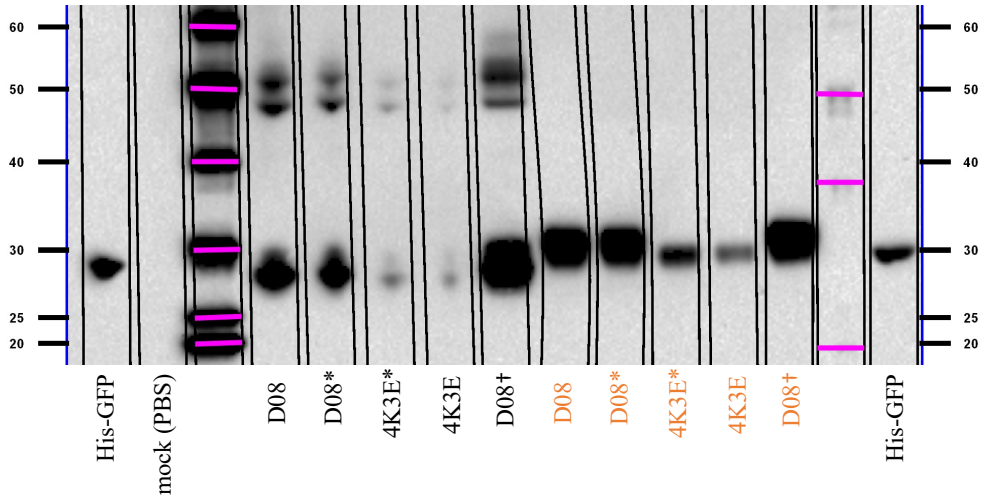


Figure 3.3: Anti-polyhistidine immunoblot to show relative recombinant expression efficiency of ultralong Fabs D08 and 4K3E, and chain-exchanged variants D08* and 4K3E* in Expi293F™ cells. All samples were heated to 95 °C for 10 minutes prior to loading. Orange text denotes samples prepared using reducing (10 % βME) SDS loading buffer. † - from 300 mL D08 expression culture. Protein standards are His-tagged Benchmark™ Precision Plus (Invitrogen™, ThermoFisher Scientific) on the left, and Precision Plus Protein Dual Color Standards (Bio-Rad) on the right. His-GFP included as a transfection control.

The expression of Fabs by transient transfection was screened in 1 mL Expi29F™ cultures, co-transfected with 0.5 µg/mL of each heavy/light chain encoding plasmid

(Section 2.2.3). Recombinant Fabs were detected in media samples by western blot, targeting the polyhistidine tail at the C terminus of the recombinant heavy chain (Fig. 3.3), consequently the migration of the bands under reducing conditions is associated with the the relative length of the CDR3H. In non-reducing conditions, a typical triple band pattern is observed for the Fabs, where the lower band corresponds to free heavy chain, dissociated at some point between expression and the heat treatment step, and the double band around 50 kDa is likely explained by two populations of Fabs; with, and without post-translational modification respectively. Strong expression efficiency was observed for the recombinant ultralong Fab D08 (Fig. 3.3 lanes 4 and 9). During expression screening (Section 2.2.3), both D08* and 4K3E* showed comparable expression efficiency to the heavy chain donor, with the D08 heavy chain showing more efficient expression than the 4K3E heavy chain (Fig. 3.3).

3.4.3 Purification of Fab D08

Recombinant Fabs were produced by co-transfection of VH and VL vectors into Expi293™ cells (Section 2.2.4) using 300 mL culture volumes. Recombinant Fab D08 was then recovered from the media using IMAC and further purified using gel filtration (Fig. 3.4). Eluted fractions were characterised by SDS-PAGE (Fig. 3.4), and those of interest were spin concentrated using Amicon® Ultracel-30 centrifugal filters (Merck). Fabs are generally highly soluble, and can easily be concentrated above 20 mg/mL. Initially, concentrated Fab D08 was flash frozen in liquid phase nitrogen and stored at −80 °C, however this may have hampered initial crystallisation efforts. Subsequent batches were stored at 4 °C, and proved to be stable for at least 9 months without demonstrating precipitation or significant degradation (not presented in this thesis). Purification of Fab D08 yielded 20.8 mg/L, spin-concentrated to 23.1 mg/mL (Table 3.4).

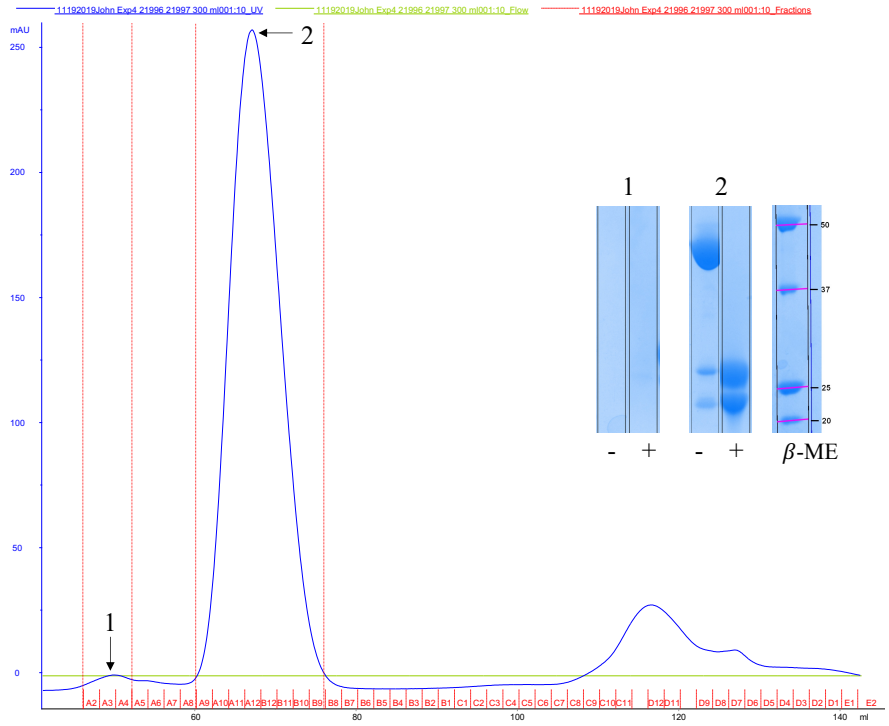


Figure 3.4: Size exclusion purification of Fab D08. 280 nm Absorbance trace of gel filtrate. Inset: Indicated fractions were analysed by SDS-PAGE. Samples prepared without ('-') and with ('+') 10 % β-ME. Band migration compared with Precision Plus Protein Dual Color Standards (Bio-Rad), with indicated molecular weights in kDa.

3.5 Crystal structure of naturally paired ultralong Fab D08

3.5.1 Crystallography

Prior to crystallisation, concentrated D08 was screened using the PCT (Section 2.8.3.1). PCT results (omitted from this thesis) indicated purified Fab had reached a concentration appropriate for trialling crystallisation experiments (23.1 mg/mL) (Table 3.4).

Sitting droplet vapour diffusion crystallisation experiments were set up at room temperature for Fab D08 (Section 2.8.3.3), using the Index™, JCSG-*plus*™, and Morpheus® commercial crystallisation screens within 96 well CrystalQuick™ X Plates (Greiner Bio-One). Fab D08 readily crystallised in all 3 tested screens. In both the Index™ and JCSG-*plus*™ screens, Fab D08 produced mainly small twinned crystals, with a couple of larger 3D crystal hits appearing (Figs. 3.5a

to 3.5d). The Morpheus® screen provided the best crystals for Fab D08, with plenty of 3D crystals observed in columns #04, #05, #09, and #12 (Figs. 3.5e and 3.5f). Nucleation and growth of Fab D08 crystals occurred rapidly; crystals first appeared within 24 hours, and most crystals had stopped growing by the + 5 day images. Crystals were initially screened by their morphology, and absorbance of 280 nm light to confirm proteinaceous composition (Fig. 3.6).

3.5.2 X-Ray Diffraction data collection for Fab D08

The plate was transferred to VMXi (Section 2.8.3.4) for *in situ* X-Ray Diffraction (XRD) screening (Fig. 3.7). Whilst VMXi has a micro-focused beamsize of 10 µm x 10 µm, the plate is oscillated following the Arndt-Wonacott method [211], meaning sub-well regions lateral to the beam target will be unintentionally oscillated into the beam path, introducing radiation damage to the sample, without the benefit of collecting XRD data. To resolve this issue during XRD screening with VMXi, crystals were selected at a vertical displacement away from candidate crystals for high-resolution cryogenic XRD data collection on I04 (Section 2.8.3.4) to avoid unintentionally irradiating candidate crystals during the horizontal *in situ* plate oscillation motion (Fig. 3.7).

Multi-crystal *in situ* XRD data were collected at 295 K using a Dectris Eiger2 X 4M detector on beamline VMXi at Diamond Light Source (DLS) [208]. The crystal quality of each well was assessed by morphology, size, UV absorbance and X-ray diffraction to select crystals for cryogenic data collection. Morpheus® condition C09 (0.1 M Tris (base):0.1 M Bicine buffer at pH 8.50 with precipitants 12 % (v/v) PEG 500 MME, 6 % (v/v) PEG 20000, and additives 0.09 M Sodium nitrate, 0.09 M Sodium phosphate dibasic, 0.09 M Ammonium sulphate)(Molecular Dimensions, calibre scientific) droplet 1 (1:1 ratio of sample:screen solution) produced the highest-resolution diffracting crystals (Figs. 3.7 and 3.8) for Fab D08, diffracting to 2.07 Å at beamline VMXi when processed by the xia2 dials pipeline (Section 2.8.3.6). Such crystals were cryocooled in liquid phase nitrogen and XRD data were collected

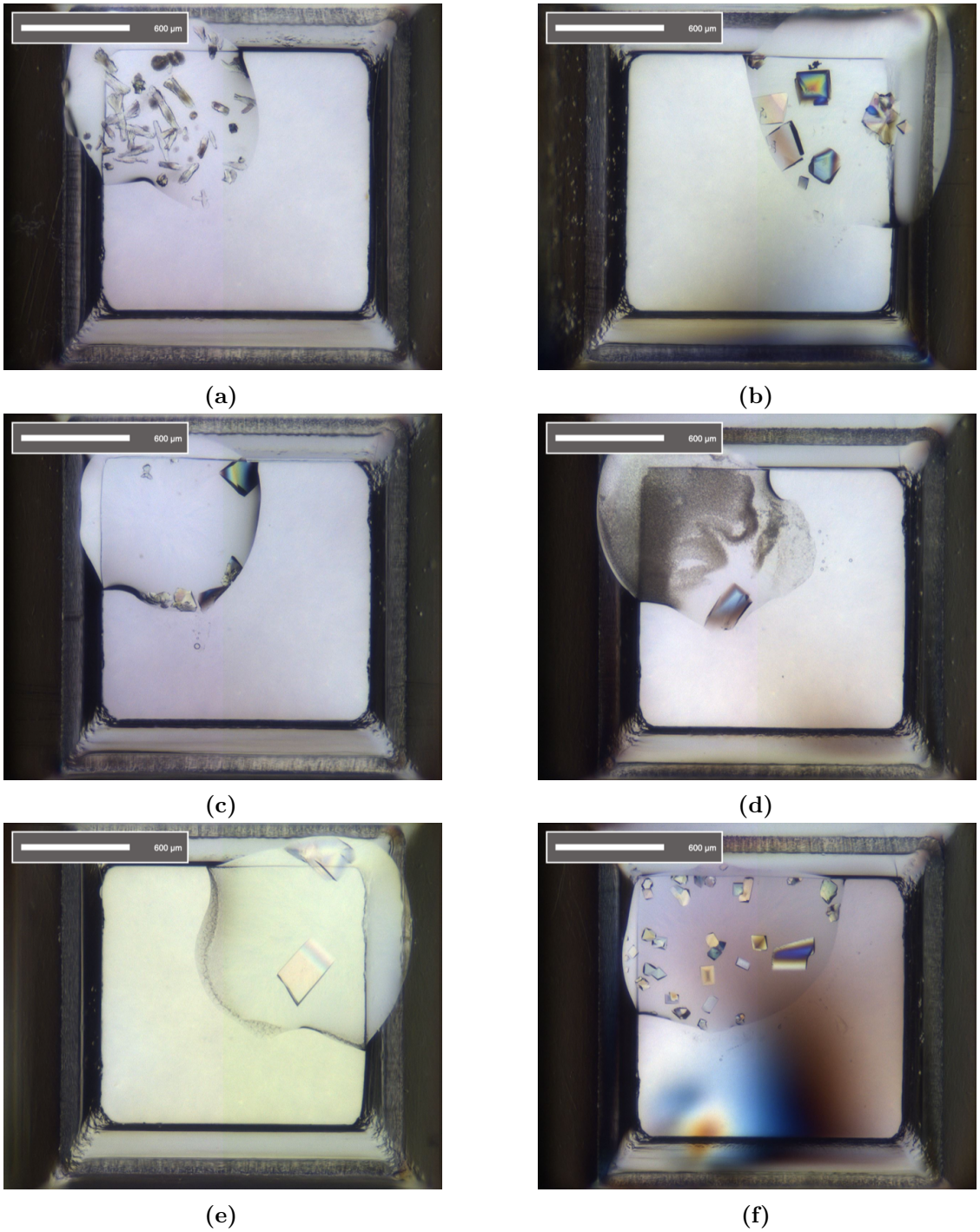
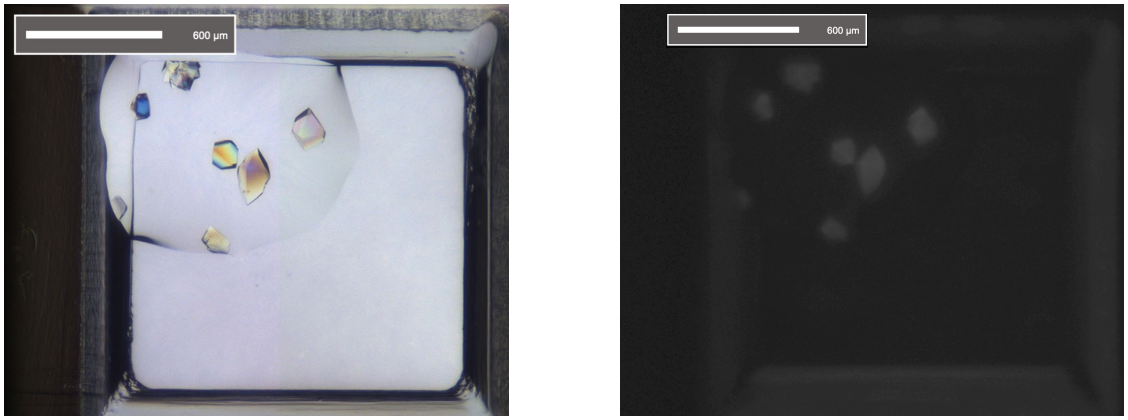
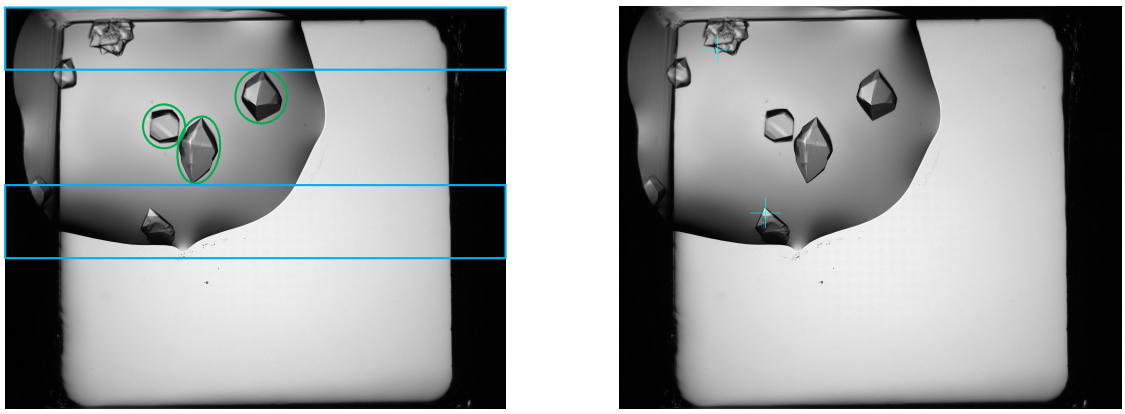


Figure 3.5: Exemplar crystal morphology for Fab D08 in various crystallisation screens. Index™ screen condition (3.5a) D09 droplet 1, and (3.5b) H01 droplet 2. JCSG-*plus*™ screen conditions (3.5c) H11 droplet 1, and (3.5d) D06 droplet 1. Morpheus® screen conditions (3.5e) B04 droplet 2 and (3.5f) G09 droplet 2. Images taken 8 days after preparing the crystallisation plate using a RockImager 1000 (Formulatrix) on imager default settings and inspected using RockMaker® web interface v3.17.7.1 (Formulatrix). Crystallisation plate was prepared and incubated at 20 °C throughout.



(a) (b)

Figure 3.6: (Fig. 3.6a) Visible and (3.6b) UV (280 nm) absorbance screening of crystal hits in Morpheus® crystallisation screen condition C09 droplet 1 for Fab D08. Images taken 3 days after preparing the crystallisation plate using a RockImager 1000 (Formulatrix) using imager default (3.6a), and UV High Resolution (3.6b) settings, and inspected using RockMaker® web interface v3.17.7.1 (Formulatrix). Crystallisation plate was prepared and incubated at 20 °C throughout.



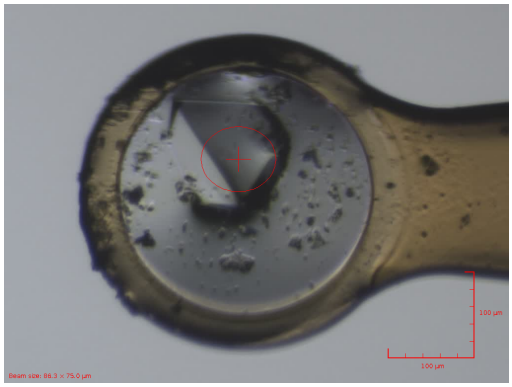
(a) (b)

Figure 3.7: Representative experimental strategy for X-Ray Diffraction (XRD) screening with beamline VMXi at Diamond Light Source. (3.7a) Candidate crystals for high-resolution cryogenic data collection were identified (circled green), and crystals with a vertical displacement away from these (blue boxes) were targeted for XRD screening using VMXi (3.7b). Images collected by the VMXi beam end sample camera of Fab D08 in Morpheus® condition C09 droplet 1. Control of beam targeting achieved through the SynchWeb interface [209].

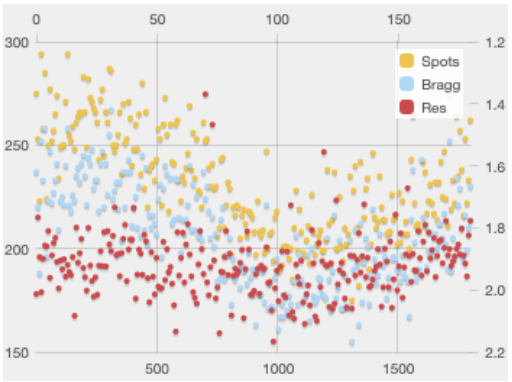
at 100 K using a Dectris Eiger2 XE 16M detector on beamline I04 at DLS. Further optimisation was not required.

Fab D08 diffracted to 1.59 Å on beamline I04 (Fig. 3.8 and Table 3.5). This single crystal dataset was characterised as belonging to the P 2_1 2_1 2_1 spacegroup,

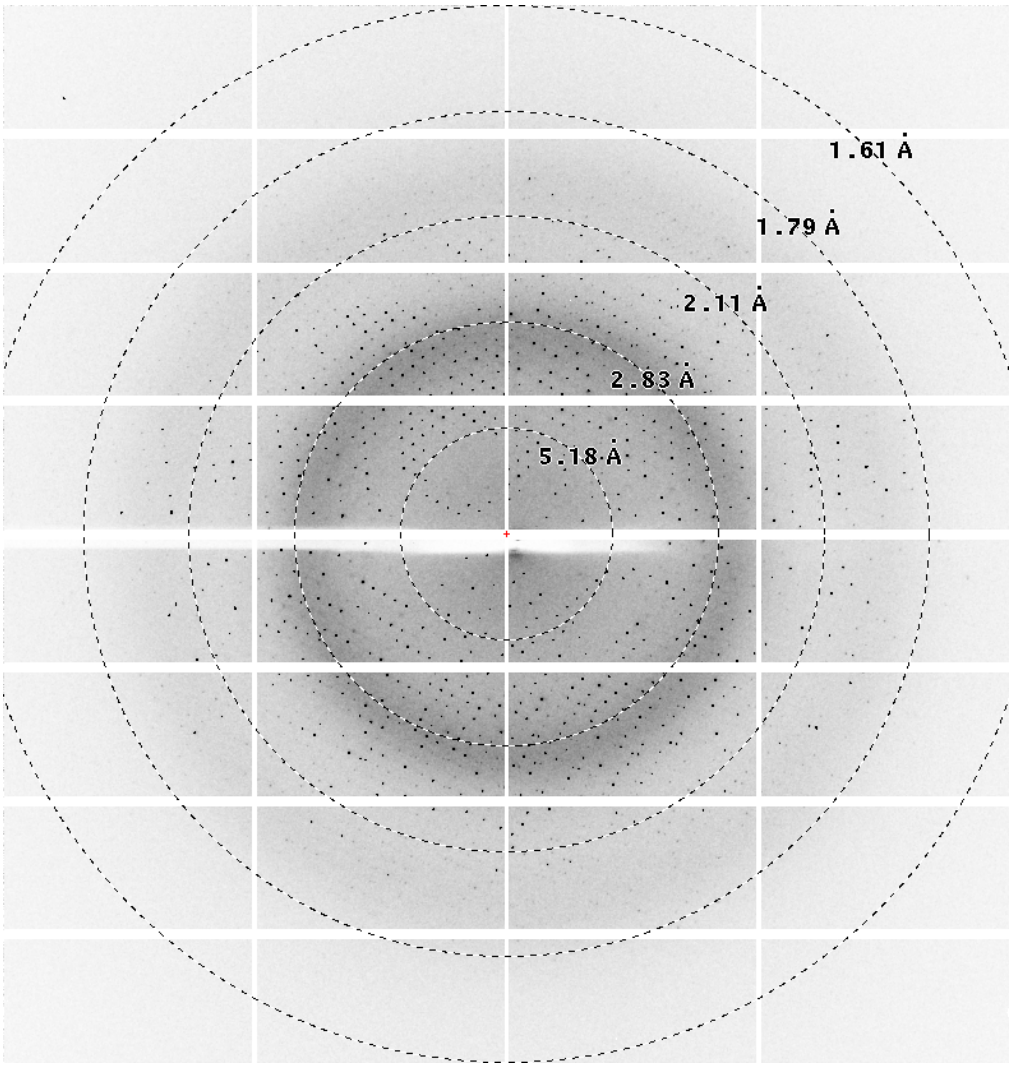
with unit cell dimensions of $a = 63.50\,\text{Å}$, $b = 71.11\,\text{Å}$, $c = 98.41\,\text{Å}$, and angles of $\alpha = \beta = \gamma = 90.00°$ by the xia2 3dii pipeline (Section 2.8.3.6). Data were collected at each 0.1° oscillation angle over 180° total oscillation, where each image was exposed for 0.008 s with an exposure flux of 4.25×10^{11} photons (Table 2.10). The asymmetric unit was predicted to contain one Fab D08 monomer, (subsequently validated by structural analysis). The structure was phased by molecular replacement using Phaser [223](Section 2.8.3.7) using 1 copy of each of the protein chain H monomer of bovine ultralong antibody BLV1H12 (PDB ID: 4K3D [78]) and the protein chain L monomer of bovine antibody B4HC-B13LC (PDB ID: 6QN7 [160]) as search models, and refined using REFMAC5 [227] to 1.59 Å (Figs. 3.9c and 3.9d and Table 3.5) within the CCP4 suite [221]. Model geometry was assessed by MOLPROBITY [232] within CCP4 [221], and by the PDB OneDep validation server [162] (Fig. 3.10). The final model constitutes 3546 protein atoms, and 597 water molecules (Table 3.5).



(a)



(b)



(c)

Figure 3.8: Single crystal X-ray diffraction data collection for ultralong Fab D08 at beamline I04 at Diamond Light Source. Fab D08 crystallised in Morpheus® condition C09 droplet 1. (3.8a) I04 beamline sample camera, showing position and size of incident X-Ray beam. (3.8b) DISTL [212] per-image analysis showing number of diffraction spots (total and Bragg)(left axis) and estimated resolution (right axis) for each image in the data collection. (3.8c) Example diffraction image composed of 15 summed frames.

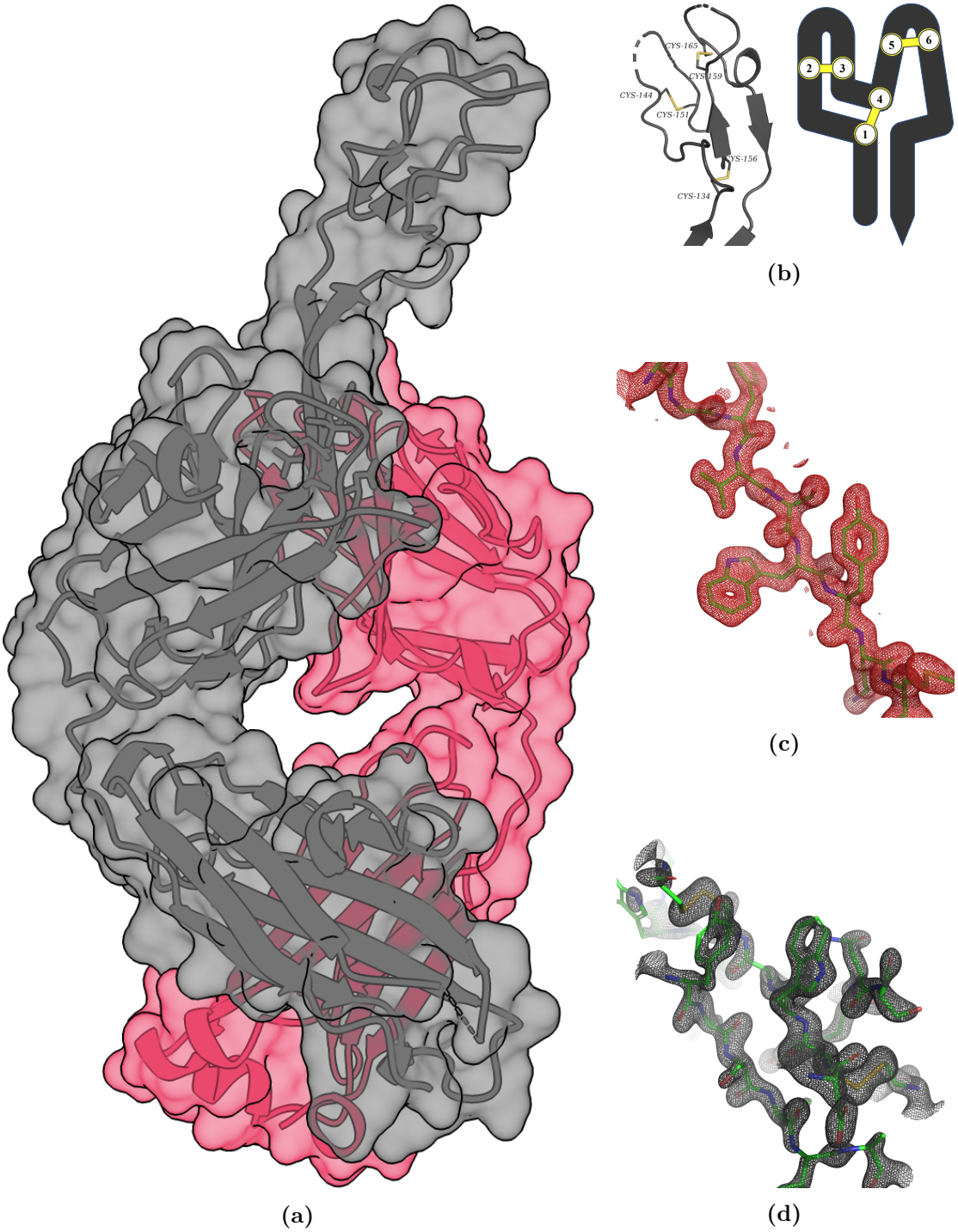
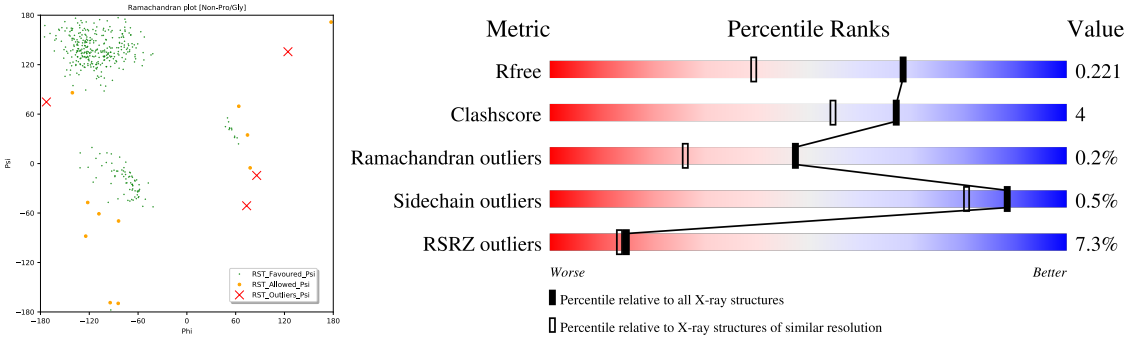


Figure 3.9: Ultralong Fab D08 crystal structure. (3.9a) Overall structure in cartoon representation with transparent surface overlay. The heavy chain is coloured in black, the light chain in red. (3.9b) Network of disulphide bonds within the ultralong knob domain, along with cartoon schematic. Representative electron density, contoured at 1.0 σ unit, for (3.9c) the tryptophan at the core of the VL domain and (3.9d) the 3-pass β-sheet of the ultralong knob domain. Surrounding structure hidden for clarity. Figures prepared using ChimeraX [257] (3.9a) and PyMOL (3.9b to 3.9d).



(a) (b)

Figure 3.10: D08 atomic model geometric validations. (3.10a) Ramachandran plot, and (3.10b) wwPDB Deposition overall quality metrics compared to all reported structures (filled boxes), and those of equivocal resolution (1.6 Å; open boxes).

3.5.3 Crystallographic structure of Fab D08

Modelling and refinement of the resulting Fab D08 XRD data revealed a stalk-knob structure to the CDR3H region that is typical for bovine ultralong antibodies (Fig. 3.9a) [78, 82, 83, 132]. The stalk conformation is modulated by a network of aromatic stacking interactions involving both intra- and inter-chain interactions involving both the heavy and light chains. Ascending from the base of the stalk toward the knob domain, Trp187H stacks with Pro73L, Tyr183H is stacked between His127H and Tyr78L, His131H stacks with Tyr61L (within CDR1L), and Tyr178H stacks with Tyr180H on the descending strand of the CDR3H (Fig. 3.11). Finally, the CDR3L is packed close to the ascending strand, but this is discussed in depth in Section 3.6.4. (Fig. 3.11) The knob domain is organised by a network of disulphide bonds (Fig. 3.9b) in a 1-4, 2-3, 5-6 configuration, establishing a 3-pass β-sheet mini-domain with connecting loops of varied length and structure. Continuous electron density was observed for both chains of Fab D08, except for the last eight residues of the heavy chain C terminus, a loop between 270 and 275 of the CH domain, and a few small breaks in the ultralong knob domain (Fig. 3.9d).

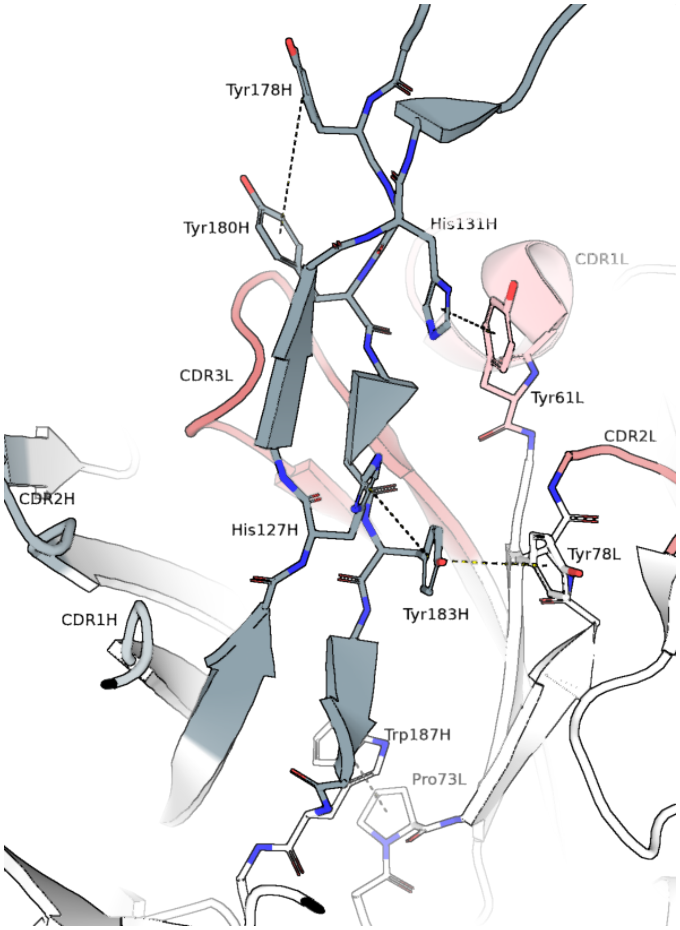


Figure 3.11: Molecular interaction network supporting the CDR3H stalk domain of ultralong Fab D08. CDR's coloured according to Fig. 3.9; D08-H CDR coloured black; D08-L CDR coloured red; increasing intensity for CDR1 to CDR3. For clarity, field of view focused on CDR3H stalk mini-domain and immediately neighbouring residues. π-stacking interactions between aromatics shown with dashed lines. Figure prepared using PyMol.

3.5.4 PISA crystal packing analysis

The ultralong CDR3H protrudes deep into the solvent [78, 82, 83] (Fig. 3.9), and would therefore intuitively make extensive contact with symmetry mates within a crystal structure. To assess this, Protein Interfaces, Surfaces and Assemblies (PISA) analysis [233] was performed on the refined atomic model for Fab D08 within CCP4 [221] (Table 3.2). The ultralong CDR3H of Fab D08 was involved in six of the 13 identified interfaces by PISA, involving 18 residues (Table 3.2 and Fig. 3.12).

Table 3.2: PISA analysis for Fab D08. NN; interface number. iNat and iNres; number of atoms and residues participating in the interaction respectively. 'x' indicates interface is crystallographically linked. Solvation free energy gain (ΔiG). Numbers of different types of bonds; hydrogen (NHB), salt bridges (NSB), disulphide bonds (NDS). Complexation significance score (CSS)..

	Structure 1					Structure 2					
NN	Range	iNat	iNres	Surface (Å^2)	x	Range	Symmetry op-n	Sym.ID	iNat	iNres	Surface (Å^2)
1	L	212	59	11128	-	H	x,y,z	1_555	220	59	14789
2	H	53	16	14789	-	L	x-1/2,-y+3/2,-z	3_465	50	19	11128
3	L	37	14	11128	-	H	-x+2,y-1/2,-z+1/2	4_745	39	12	14789
4	L	37	12	11128	-	H	-x+2,y-1/2,-z-1/2	4_744	32	9	14789
5	H	33	12	14789	x	H	x-1/2,-y+3/2,-z	3_465	31	11	14789
6	H	27	7	14789	x	H	-x+2,y-1/2,-z-1/2	4_744	23	9	14789
7	L	17	6	11128	x	L	-x+5/2,-y+1,z-1/2	2_764	23	9	11128
8	H	17	4	14789	x	H	x,y,z-1	1_554	18	7	14789
9	H	14	6	14789	x	H	-x+2,y-1/2,-z+1/2	4_745	13	8	14789
10	L	9	3	11128	x	L	-x+2,y-1/2,-z-1/2	4_744	11	2	11128
11	L	12	4	11128	-	H	x-1/2,-y+3/2,-z	3_465	11	5	14789
12	L	7	3	11128	-	H	x,y,z-1	1_554	6	2	14789
13	H	2	1	14789	-	L	-x+3/2,-y+1,z-1/2	2_664	1	1	11128

NN	interface area (Å^2)	ΔiG (kcal/mol)	ΔiG P-value	NHB	NSB	NDS	CSS
1	2007.2	-32.1	0.062	12	3	1	1.000
2	491.6	-2.9	0.564	6	0	0	0.000
3	343.2	-1.3	0.697	2	0	0	0.000
4	338.6	-0.7	0.736	6	0	0	0.000
5	322.4	-2.5	0.555	5	0	0	0.000
6	219.9	-0.9	0.689	2	0	0	0.000
7	187.8	-0.8	0.556	4	0	0	0.000
8	166.0	0.7	0.753	2	2	0	0.000
9	119.9	0.3	0.768	1	0	0	0.000
10	88.8	-0.8	0.449	0	0	0	0.000
11	84.7	-1.4	0.420	0	0	0	0.000
12	62.5	1.8	0.906	2	0	0	0.000
13	3.7	0.0	0.629	0	0	0	0.000

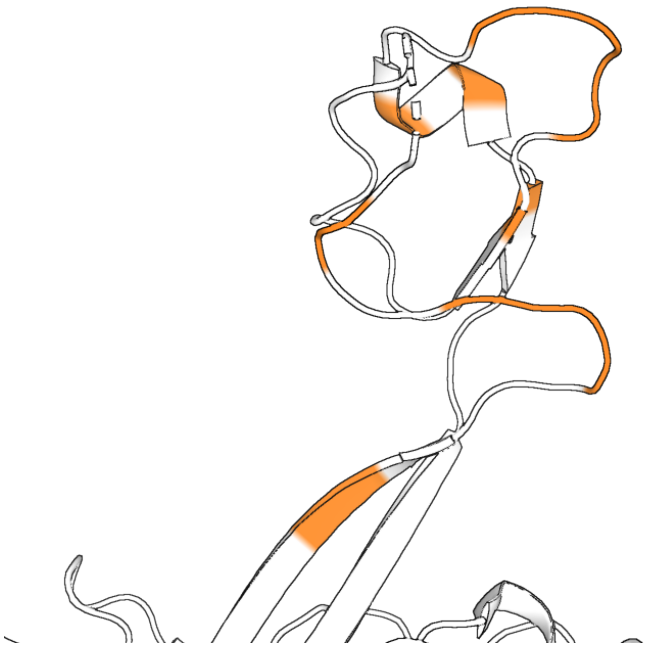
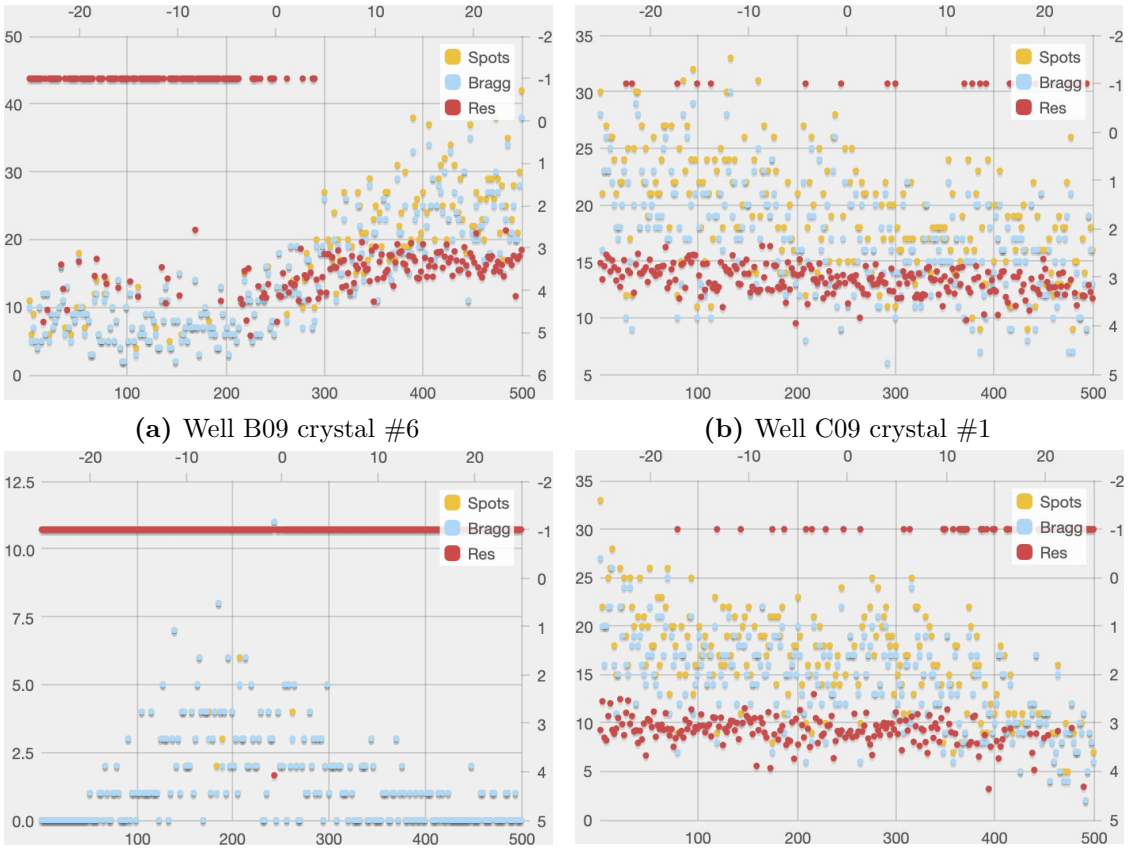


Figure 3.12: Ultralong CDR3H residues participating in a macromolecular interface. Residues in orange interface with a symmetry copy of Fab D08. Interfaces were identified using PISA [233].

3.5.5 Performance of VMXi for structural determination

The Versatile Macromolecular Crystallography *In situ* (VMXi) beamline was principally used in this project for screening crystals for diffraction quality, however, structural determination is possible through the merging of multiple datasets collected at VMXi to negate the limitations of collecting data over only ± 30° oscillation (Section 2.8.3.4). To assess the structural determination performance of VMXi, a multi-crystal, ambient temperature dataset was prepared, solved, and compared to the structure determined on I04 at cryogenic temperature.

Four single crystal VMXi data collections were merged using DIALS [220], with frame ranges selected by the presence of Bragg spots as analysed by DISTL [212] per-image analysis (Fig. 3.13). Frames 250:500 from crystal #6 of well B09 (Figs. 3.13a and 3.14a), frames 1:500 from crystal #1 (Figs. 3.13b and 3.14b), and frames 100:400 from crystal #2 of well C09 Figs. 3.13c and 3.14c) and frames 1:400 from crystal #4 of well G09 (Figs. 3.13d and 3.14d) were selected for merging.



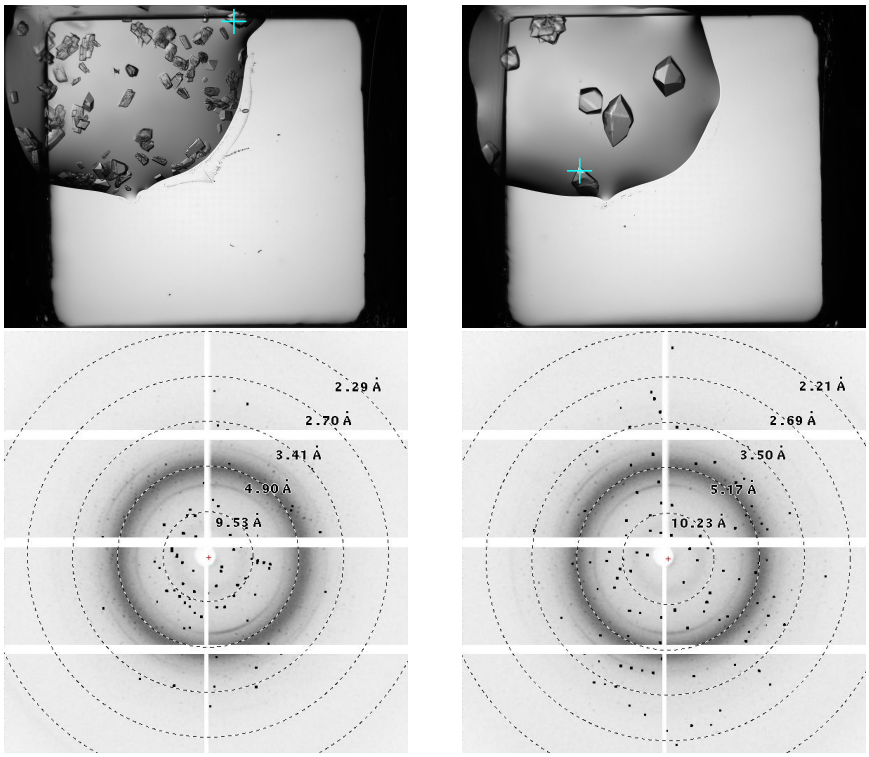
(a) Well B09 crystal #6

(b) Well C09 crystal #1

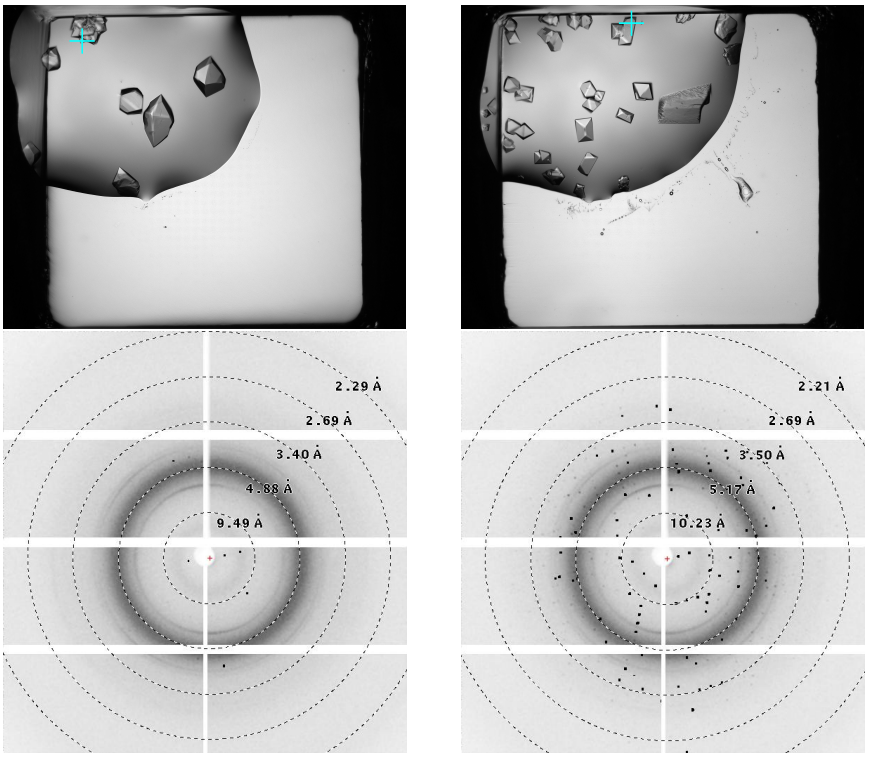
(c) Well C09 crystal #2

(d) Well G09 crystal #4

Figure 3.13: DISTL [212] **per-image analysis showing number of diffraction spots (total and Bragg)(left axis) and estimated resolution (right axis) for each image in each VMXi data collection.** (3.13a) well B09 crystal #6, (3.13b) well C09 crystal #1, (3.13c) well C09 crystal #2 and (3.13d) well G09 crystal #4.



(a) Well B09 crystal #6 **(b)** Well C09 crystal #1



(c) Well C09 crystal #2 **(d)** Well G09 crystal #4

Figure 3.14: Multiple crystal X-ray diffraction data collection for ultralong Fab D08 at beamline VMXi at Diamond Light Source. VMXi beamline sample camera, showing position of incident X-Ray beam, and example diffraction image composed of 15 summed frames for droplet 1 of wells B09 (3.14a), C09 (3.14b and 3.14c) and G09 (3.14d). Fab D08 crystallised using the Morpheus® screen at room temperature.

Table 3.3: Individual, and merged VMXi crystallographic diffraction datasets for antibody D08. Values in parentheses are for the highest resolution shell. Crystal #2 of well C09 failed pre-processing, so statistics could not be determined. Datasets were merged using DIALS [220]. †Parentheses indicate number of reflections with Free R flag.

	B09 #6	**C09 #1**	**G09 #4**	**Merged**
Data Collection				
Observations				
Total	32,604 (1446)	42,345 (797)	39,999 (959)	118,707 (3,896)
Unique	14,787 (758)	21,940 (619)	22,491 (757)	23,145 (959)
Space Group	P 2 2_1 2	P 2_1 2_1 2	P 1 2_1 1	P 2_1 2_1 2_1
Cell Dimensions				
a, b, c (Å)	64.16, 71.55, 100.19	64.23, 71.62, 100.29	71.55, 64.21, 100.22	64.23, 71.59, 100.23
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 89.97, 90.00	90.00, 90.00, 90.00
Statistics				
Resolution (Å)	47.70 - 2.43 (2.47 - 2.43)	58.27 - 2.07 (2.11 - 2.07)	58.23 - 2.17 (2.21 - 2.17)	58.26 - 2.20 (2.34 - 2.20)
R_{merge}	0.197 (1.227)	0.103 (0.831)	0.117 (1.405)	0.297 (1.851)
R_{pim}	0.117 (0.781)	0.064 (0.576)	0.072 (0.945)	0.130 (0.965)
$I/\sigma I$	3.5 (0.4)	6.0 (0.5)	5.1 (0.5)	12.5 (2.3)
Completeness (%)	82.4 (84.9)	76.0 (43.5)	46.5 (31.4)	96.1 (95.8)
Redundancy	2.2 (1.9)	1.9 (1.3)	1.8 (1.3)	5.1 (4.1)
$CC_{1/2}$	1.0 (0.3)	1.0 (0.5)	1.0 (0.3)	0.985 (0.225)
Refinement				
Resolution (Å)	-	-	-	58.32 - 2.20
No. reflections	-	-	-	23,012 (1127)†
R_{work}/R_{free}	-	-	-	0.177 / 0.228
No. atoms:				
Protein	-	-	-	3457
Water	-	-	-	52
B-factors:				
Protein	-	-	-	40.12
Water	-	-	-	47.59
R.m.s. deviations:				
Bond lengths (Å)	-	-	-	0.0091
Bond angles (°)	-	-	-	1.522

Following data merging (Table 3.3), the R_{free} flag was randomly raised for 5 % of reflections using AIMLESS [218]. Molecular replacement was performed in PHASER [223], using each chain monomer of the Fab D08 structure collected at cryogenic conditions as search models. Molecular replacement placed 1 monomer in the asymmetric unit. The phasing solution was used as an input model for ARP/wARP [224] automated model building, refined with 50 cycles of rigid body refinement with REFMAC5 [227]. The resulting structure was further refined using iterative REFMAC5 cycles [227] and manual model building using COOT [228] to 2.20 Å with

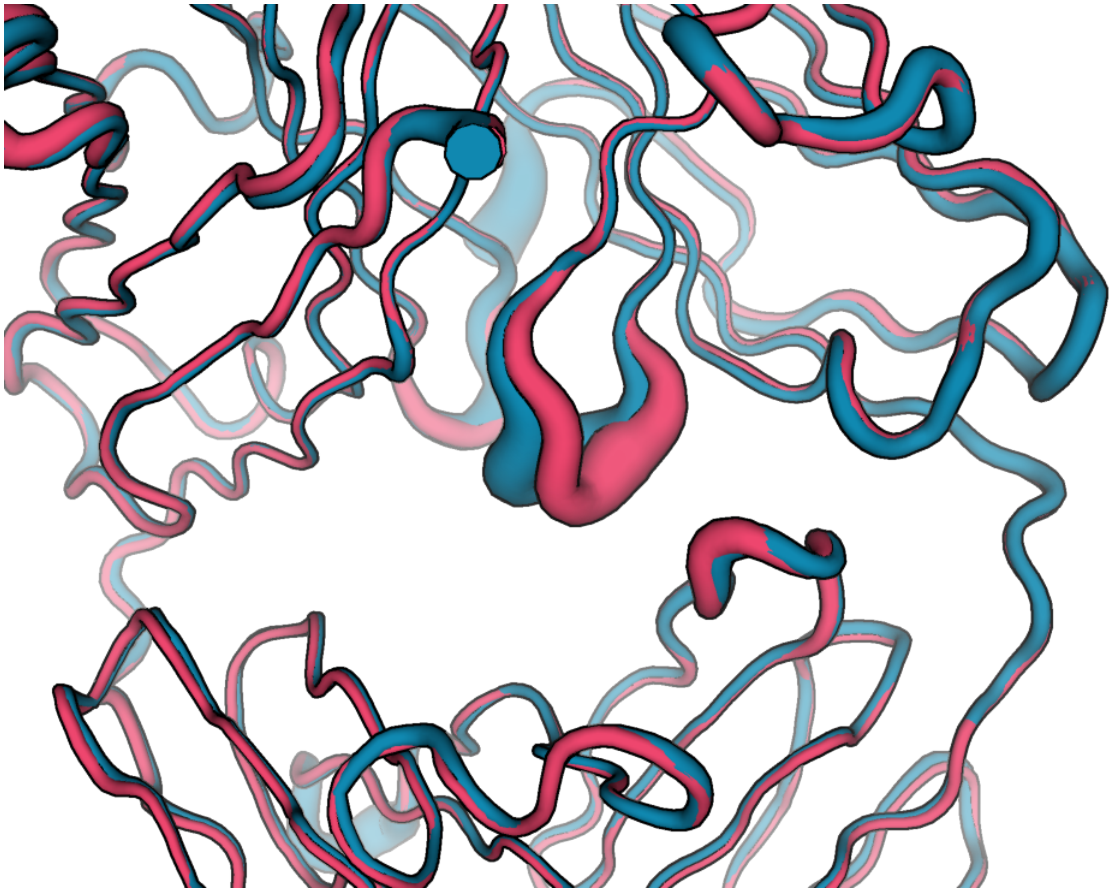


Figure 3.15: Comparison of antibody D08 structures at cryogenic and ambient temperatures. B factor putty representation of Fab D08 collected at cryogenic (blue) and ambient (fuchsia) temperatures on I04 and VMXi beamlines respectively. View centred on Cβ-Dβ loop of the VL domain, which showed the greatest RMSD between the structures.

a final R_{work} of 0.177, and R_{free} of 0.228 within the CCP4 suite [221] (Table 3.3).

The room temperature structure expectantly aligned very well to the cryogenic structure with a difference only observed at the Cβ-Dβ loop of the VL domain (RMSD of 3.5 Å versus 0.341 Å for the whole molecule), which protrudes into the solvent void between the four immunoglobulin domains of the Fab (Fig. 3.15). In the room-temperature dataset, the Cβ-Dβ loop is displaced 3.3 Å to the Fab molecule exterior surface, relative to that of the cryogenic structure. The B-factors of this loop in both structures is raised, as indicated by the increased putty diameter. Average Cα atom B-factors for this loop are 42.28 for the cryogenic, and 56.55 for the room temperature structures. For comparison, the equivocal Cβ-Dβ loop of the VH domain gave mean Cα atom B-factors of 33.18, and 39.78 for the cryogenic

and room temperature structures respectively.

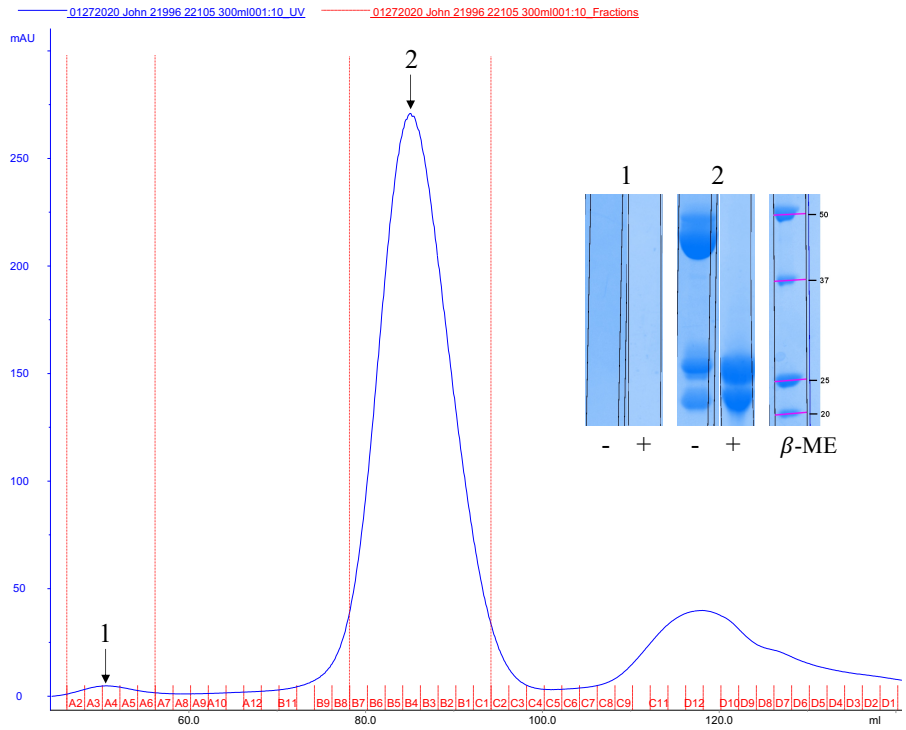
The conformation of the ultralong CDR3H loop was identical between the structures. Furthermore, B-factor amplitudes gave the same general pattern of increasing at the far distal loops of the ultralong CDR3H loop, or other inter-β-strand loops of the immunoglobulin domains, and decreasing when more proximal to the Fv region. Room temperature B-factors were elevated compared to those of corresponding atoms in the cryogenic dataset.

3.6 Chain-exchanged ultralong fragment antibody structures

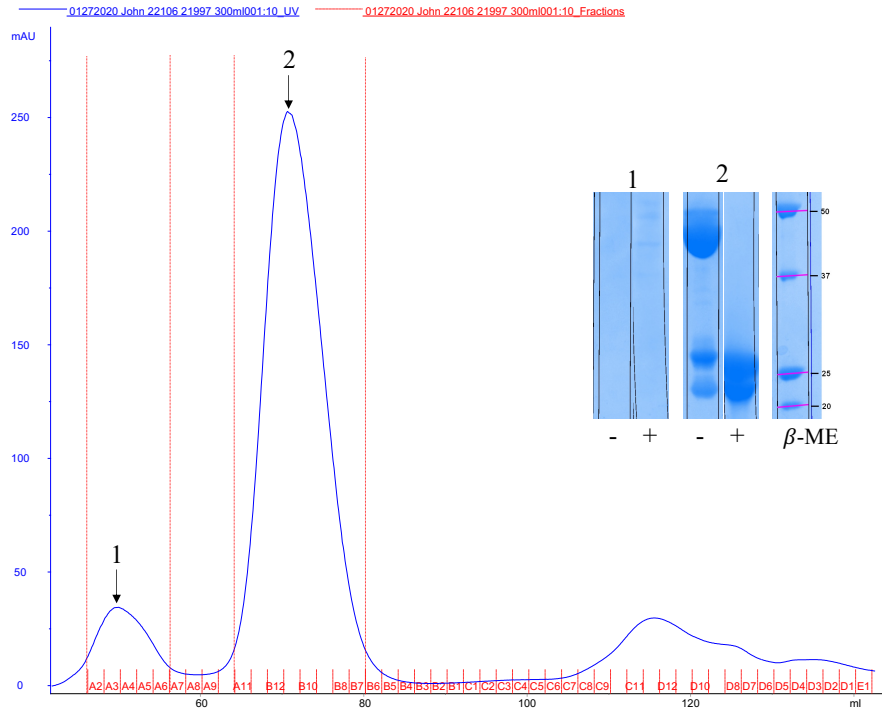
To examine the impact of light chain sequence on the ultralong CDR3H structure, each chain of Fab D08 was paired *in vitro* with its reciprocal from ultralong antibody BLV5B8 (PDB ID: 4K3E) [78], the light chain of which has also been paired *in vitro* with ultralong heavy chains in more recently published ultralong antibody structures [82, 83], and crystallographic structures pursued. As previously discussed (Section 3.4.1), chain-exchanged Fabs take the name of from the heavy chain donor, so these denoted Fab 'D08*' and '4K3E*'. Antibody variable domain sequences for both D08 and 4K3E are provided in Table A.1.

3.6.1 Expression and purification of chain-exchanged ultralong antibodies

Given that D08* and 4K3E* expressed well at small scale, 300 mL cultures were prepared for Fab purification to produce sufficient Fab for structural study (Section 2.2.4) (Fig. 3.16). Eluted fractions were characterised by SDS-PAGE (Section 2.7.2) (Fig. 3.16), and those of interest were spin-concentrated using Amicon® Ultracel-30 centrifugal filters (Merck) (Section 2.6.5) (Table 3.4). Purified, recombinant Fabs were stored at 4 °C.



(a) Fab D08*



(b) Fab 4K3E*

Figure 3.16: Size exclusion purification of Fabs D08* and 4K3E*. 280 nm Absorbance trace of gel filtrate. Inset: Indicated fractions were analysed by SDS-PAGE. Samples prepared without ('-') and with ('+') 10 % β-ME. Band migration compared with Precision Plus Protein Dual Color Standards (Bio-Rad), with indicated molecular weights in kDa.

Table 3.4: Comparative recombinant Fab yields obtained from transient transfection of Expi293F™ cells. Heavy and light chain ID refers to assigned sequence ID within the OPPF Target database. †Time taken for spin-concentration to achieve stated concentration.

Antibody name:	D08	D08*	4K3E*
Heavy chain:	21996	21996	22106
Light chain:	21997	22105	21997
Pooled fractions:	A09:B09	B07:C01	A11:B07
Pooled volume (mL):	16	16	16
Yield (mg/L):	20.8	17.6	9.2
Concentration (mg/mL):	23.1	26.0	24.4
Time taken† (minutes):	45	45	45

3.6.2 Crystallography

PCT (Section 2.8.3.1) results (omitted from this thesis) for concentrated Fabs D08* and 4K3E* indicated Fab concentrations (26.0 mg/mL and 24.4 mg/mL respectively) were appropriate for trialling crystallisation experiments (Table 3.4).

Sitting droplet vapour diffusion crystallisation experiments were set up at room temperature for synthetic chain-exchanged Fabs D08* and 4K3E* (Section 2.8.3.3), using the JCSG-*plus*™ and Morpheus® commercial crystallisation screens within Swissci UV-transmissible Polymer 96-Well Triple Drop Plates (Molecular Dimensions). The Index™ screen was omitted due to the low yield of Fabs D08* and 4K3E*, and the previous success of the JCSG-*plus*™ and Morpheus® screens in producing crystal hits sufficient for data collection with ultralong Fab D08.

JCSG-*plus*™ produced unsatisfactory crystals for both D08* and 4K3E*, with 1D-, and 2D-crystals, as well as a few 3D microcrystals (Figs. 3.17a to 3.17d). Morpheus® produced 1D crystals, but also produced many larger 3D crystals for both D08* and 4K3E* that were sufficient for data collection without further optimisation with plenty of 3D crystals observed in columns #01, #04, #05, #08, #09, and #12 (Figs. 3.17e and 3.17f). Nucleation and growth of Fab D08* and 4K3E* crystals occurred rapidly; crystals first appeared within 30 minutes, and most crystals had stopped growing by the +2, or + 3 day images for 4K3E* or D08* respectively. Crystals were screened by the absorbance of 280 nm light (not

presented in this thesis). Crystals diffracted to 1.45 Å and 1.46 Å on beamline I04 for D08* and 4K3E* respectively (Figs. 3.18 and 3.25 and Table 3.5).

The D08* single crystal dataset (Fig. 3.18 and Table 3.5) was characterised as belonging to the P 2_1 2_1 2_1 spacegroup, with unit cell dimensions of a = 65.03 Å, b = 70.86 Å, c = 96.01 Å and angles of $\alpha = \beta = \gamma = 90.00°$. Data were collected at each 0.1° oscillation angle over 180° total oscillation, where each image was exposed for 0.008 s with an exposure flux of 2.93×10^{11} photons (Table 2.10). Data were integrated and scaled automatically using the xia2 3dii pipeline (Section 2.8.3.6). The asymmetric unit was predicted to contain one Fab D08* monomer, (subsequently validated by structural analysis). The structure was phased by molecular replacement using Phaser [223](Section 2.8.3.7) using 1 copy of each of the D08 chain H monomer and the protein chain L monomer of bovine antibody B4HC-B13LC (PDB ID: 6QN7 [160]) as search models, and refined using REFMAC5 [227] to 1.45 Å (Figs. 3.19c and 3.19d and Table 3.5) within the CCP4 suite [221]. The final model constitutes 3696 protein atoms, and 493 water molecules (Table 3.5).

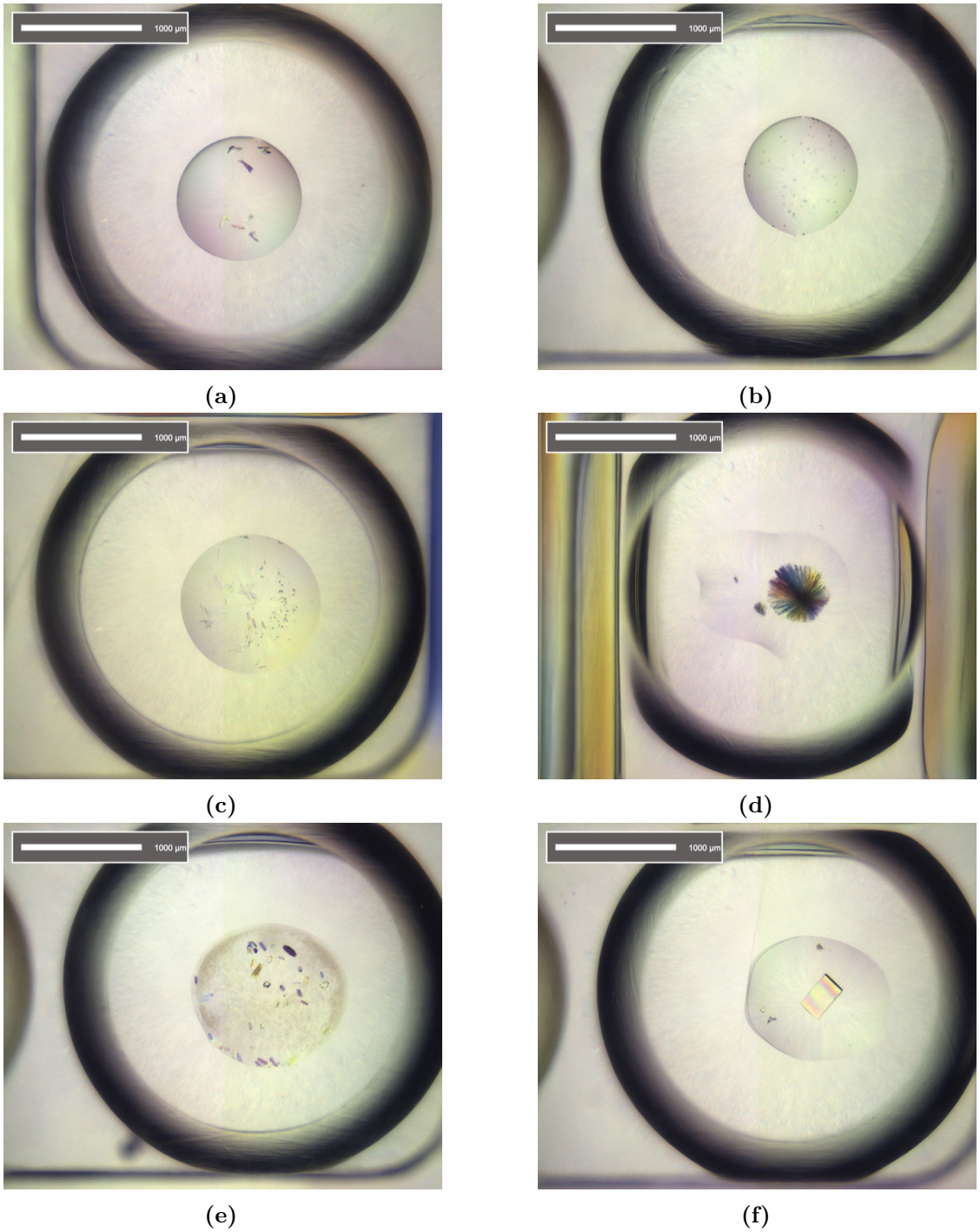
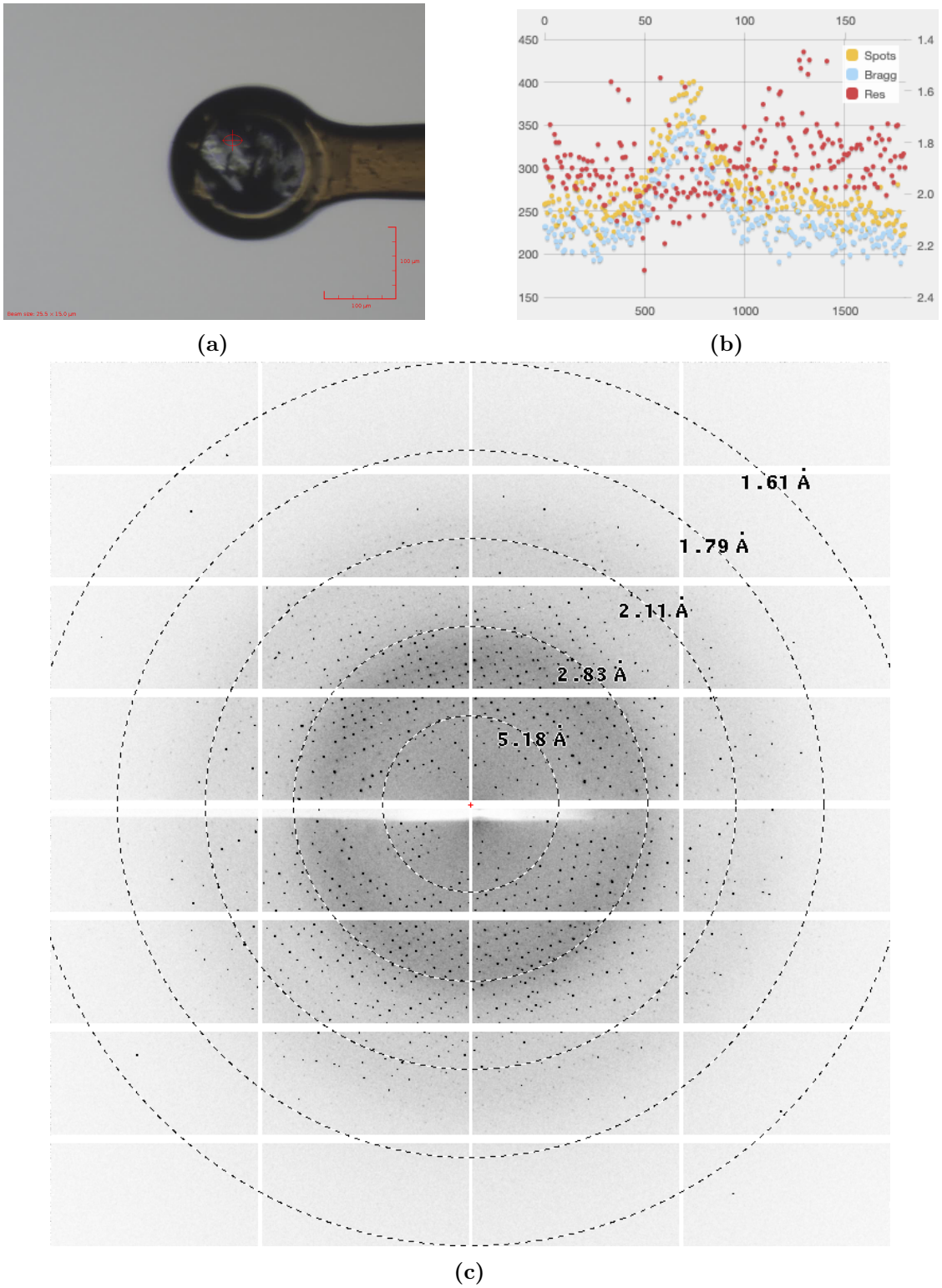


Figure 3.17: Exemplar crystal morphology for Fab D08* and 4K3E* in various crystallisation screens. Fab D08* in JCSG-*plus*™ screen condition (3.17a) D01 droplet 2, and (3.17b) H11 droplet 3. Fab 4K3E* in JCSG-*plus*™ screen conditions (3.17c) A01 droplet 3, and (3.17d) C03 droplet 1. (3.17e) Fab D08* in Morpheus® screen condition G12 droplet 3 and (3.17f) Fab 4K3E* in Morpheus® condition G09 droplet 3. Images taken 5 days after preparing the crystallisation plate using a RockImager 1000 (Formulatrix) on imager default settings and inspected using RockMaker® web interface v3.17.7.1 (Formulatrix). Crystallisation plate was prepared and incubated at 20 °C throughout.



(a) (b)

(c)

Figure 3.18: Single crystal X-ray diffraction data collection for ultralong Fab D08* at beamline I04 at Diamond Light Source. Fab D08* crystallised in Morpheus® condition B08 droplet 2. (3.18a) I04 beamline sample camera, showing position and size of incident X-Ray beam. (3.18b) DISTL [212] per-image analysis showing number of diffraction spots (total and Bragg)(left axis) and estimated resolution (right axis) for each image in the data collection. (3.18c) Example diffraction image composed of 15 summed frames.

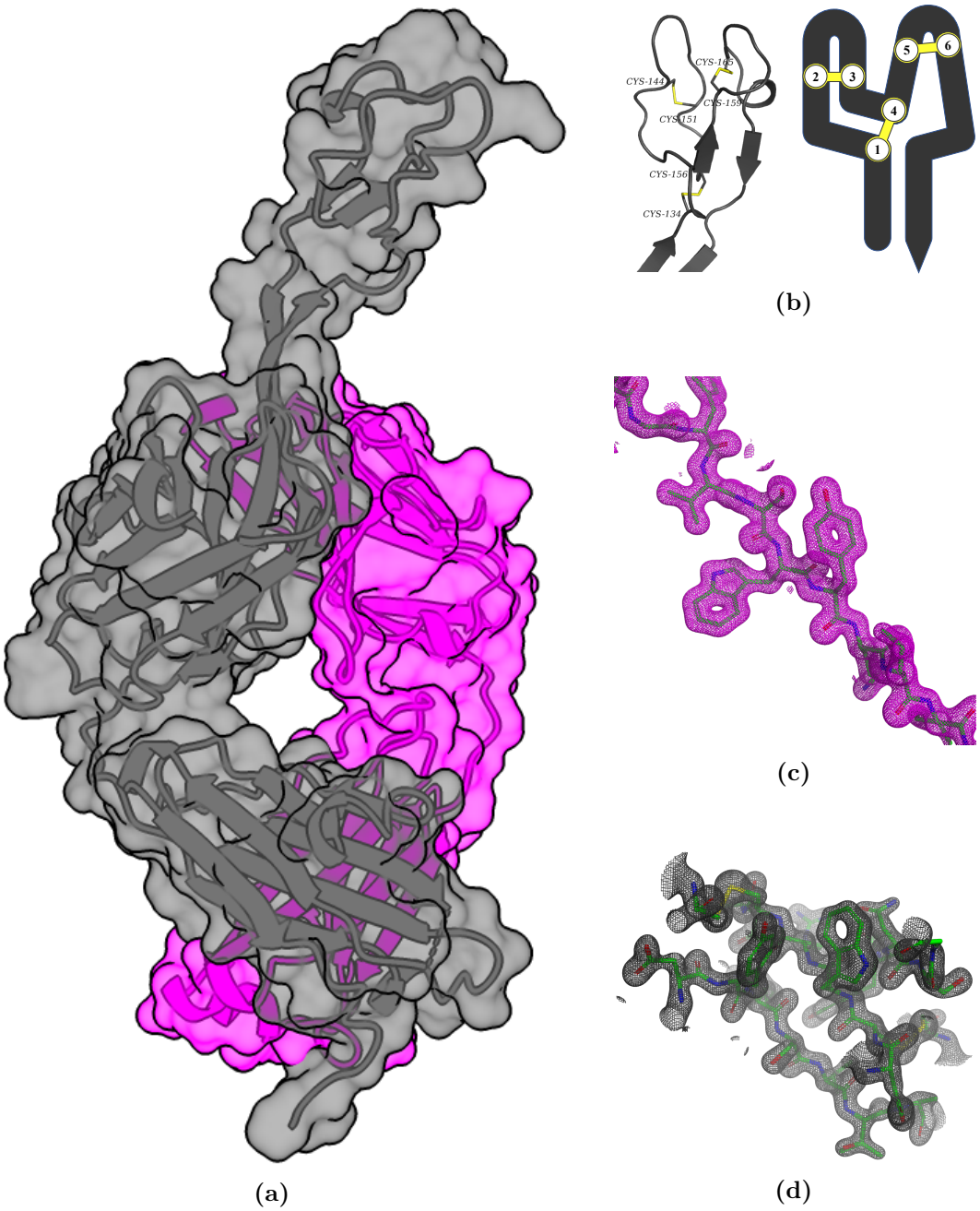


Figure 3.19: Ultralong Fab D08* crystal structure. (3.19a) Overall structure in cartoon representation with transparent surface overlay. The D08 heavy chain is coloured in black, the 4K3E light chain in magenta. (3.19b) Network of disulphide bonds within the ultralong knob domain, along with cartoon schematic. Representative electron density, contoured at 1.0 σ unit, for (3.19c) the tryptophan at the core of the VL domain and (3.19d) the 3-pass β-sheet of the ultralong knob domain. Surrounding structure hidden for clarity. Data were collected at beamline I04 at Diamond Light Source. Figures prepared using ChimeraX [257] (3.19a) and PyMOL (3.19b to 3.19d).

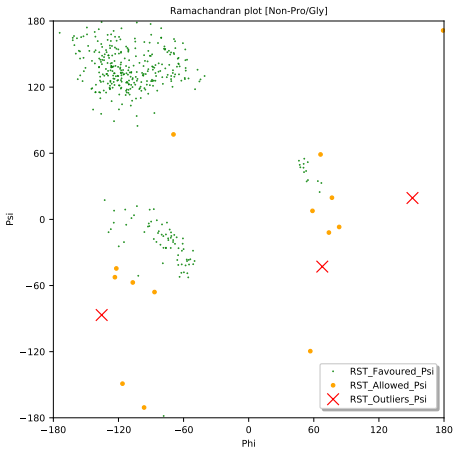


Figure 3.20: D08* atomic model ramachandran plot.

3.6.3 Crystallographic structure of ultralong chain-exchanged Fab D08*

As expected, Fab D08* presents a similar overall structure as Fab D08 (Figs. 3.19 and 3.24e), with the ultralong CDR3H adopting the canonical ultralong antibody arrangement of a 3-pass β-sheet knob domain, organised by the same network of disulphide bonds (Figs. 3.19b and 3.21), projected into the solvent by a β-ribbon stalk, supported by a series of aromatic stacking interactions involving both chains (Figs. 3.19a, 3.19b and 3.22). Continuous electron density was observed for both chains for Fab D08*, except for the last 4 Histidines of the heavy chain C terminus, a loop between 268 and 276 of the CH domain, and the side chain density of a few residues in the ultralong CDR3H loop (Figs. 3.19c and 3.19d).

3.6.4 Pairing-mediated twist to the ultralong stalk

Comparing the D08* structure to D08, a twist of 13.72°, and a translational shift of 2.47 Å is introduced to the knob region, resulting in a total displacement of 3.08 Å of the Fab D08* knob domain mass centre relative to that of Fab D08 (Fig. 3.24c), calculated using the PyMOL RotationAxis script [263]. The D08 light chain CDR3L adopts a tight turn conformation, allowing close packing against the side chain of Gln128H and the Glu129H backbone of the D08 CDR3H ascending strand. The 4K3E light chain CDRL3 adopts a β-turn structure, too large to pack against the

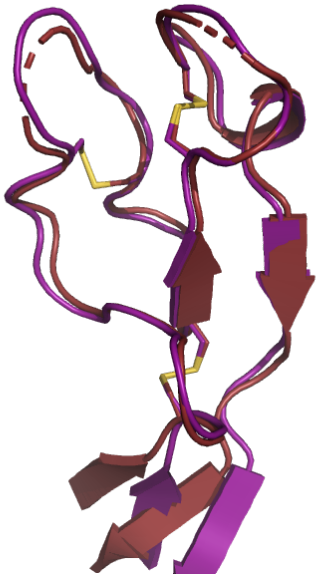


Figure 3.21: Structural overlay showing disulphide network organising the ultralong CDR3H knob domain of Fab D08 (dark red) is maintained in chain-exchanged Fab D08* (dark pink). Data were collected at beamline I04 at Diamond Light Source. Figure prepared using PyMOL.

CDR3H ascending strand, and thereby allowing Tyr180H to pack between the 4K3E CDRL3 and the D08 CDR3H stalk region (Fig. 3.22). The close packing of Tyr180H, instead of the CDR3L, introduces a twist to the stalk (Fig. 3.24), by disrupting the Tyr178H-Tyr180H stacking interaction.

3.6.5 Pairing-mediated ultralong paratope warping

The core of the D08* CDR3H knob domain is nearly identical to that of D08, with the cysteine positioning preserved between the two structures (Fig. 3.21). As a likely consequence of the aforementioned light-chain mediated twist, a shift is introduced to relative positions within the knob domain of D08*. This shift warps the presented paratope by 2.8 Å to 4.3 Å (Fig. 3.24), and may therefore result in modified antigen affinities, or even altered specificities of ultralong antibodies.

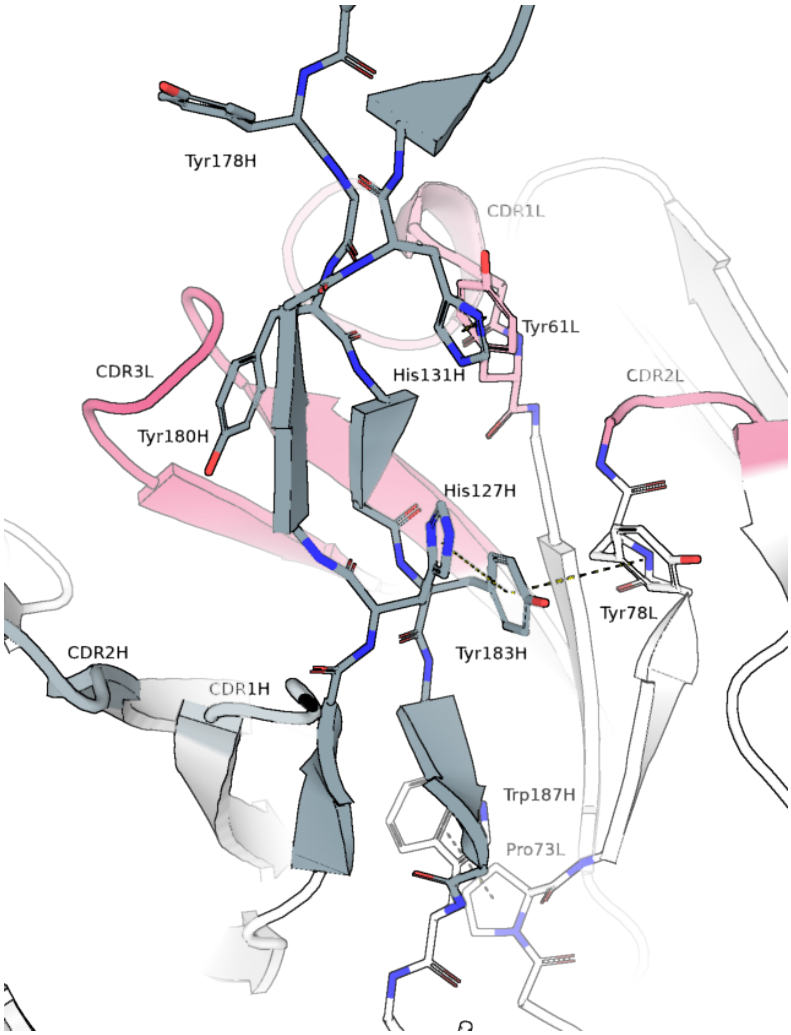
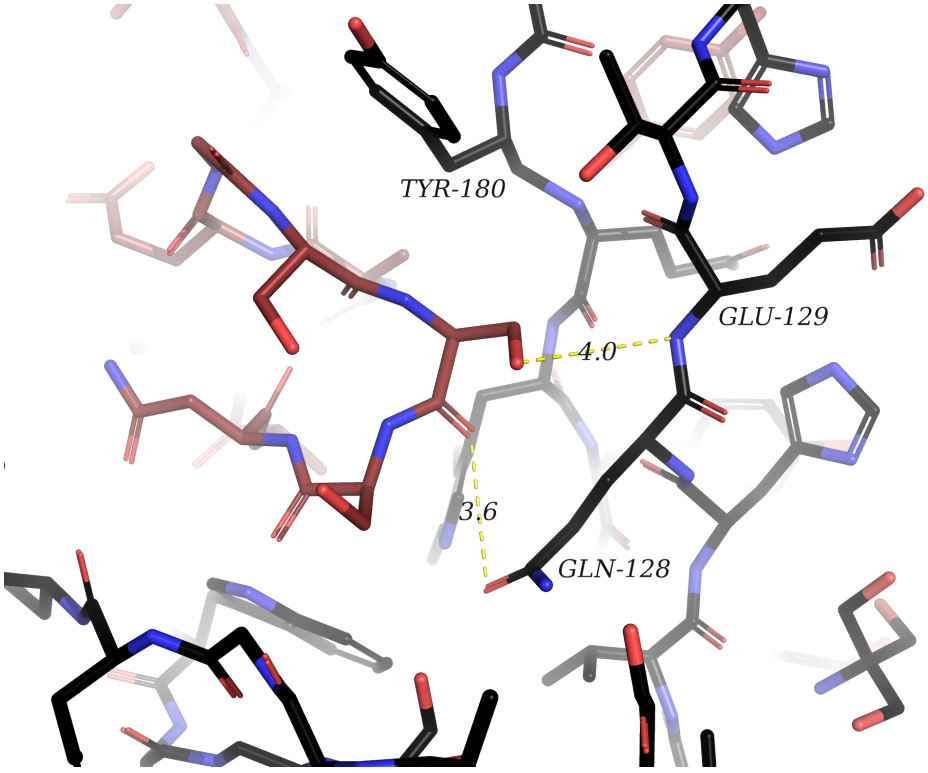
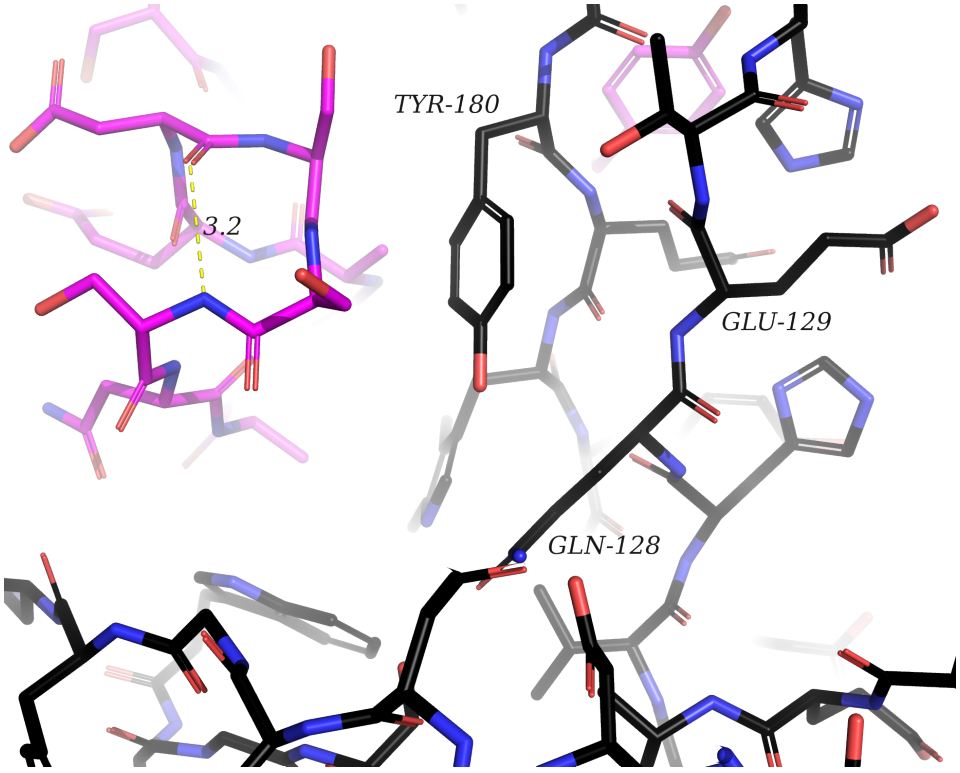


Figure 3.22: Molecular interaction network supporting the CDR3H stalk domain of ultralong Fab D08*. CDR's coloured according to Fig. 3.19; D08-H CDR coloured black; 4K3E-L CDR coloured pink; with increasing intensity for CDR1 to CDR3. For clarity, field of view focused on CDR3H stalk mini-domain and immediately neighbouring residues. π-stacking interactions between aromatics shown with dashed lines. Figure prepared using PyMol.

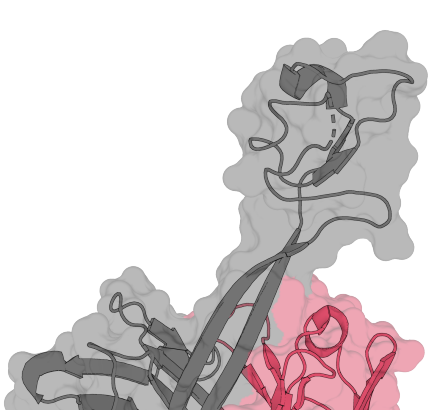


(a) Fab D08

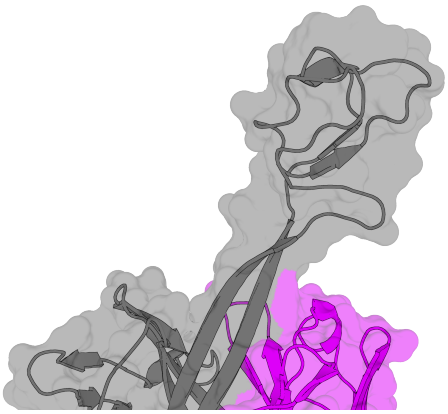


(b) Fab D08*

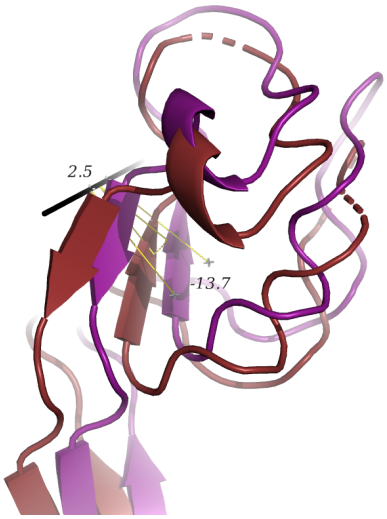
Figure 3.23: Structural comparison of the CDR3L packing of (3.23a) Fab D08 and (3.23b) D08*. Distances between the CDR3L and CDR3H are indicated, as is the β-turn hydrogen bond of the Fab D08* CDR3L. Chains coloured as in Figs. 3.9, 3.19 and 3.26.



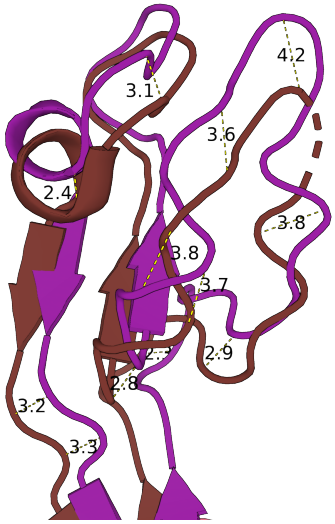
(a) Fab D08



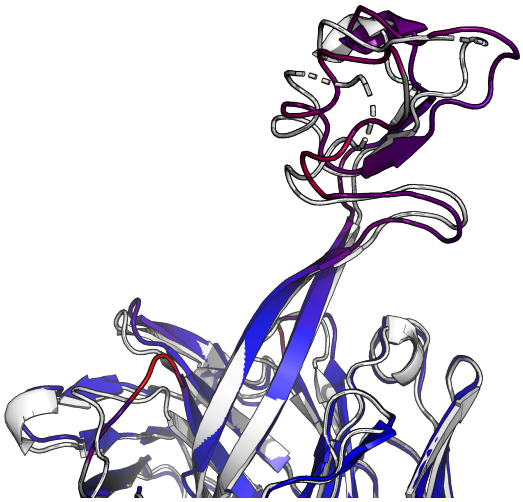
(b) Fab D08*



(c)



(d)



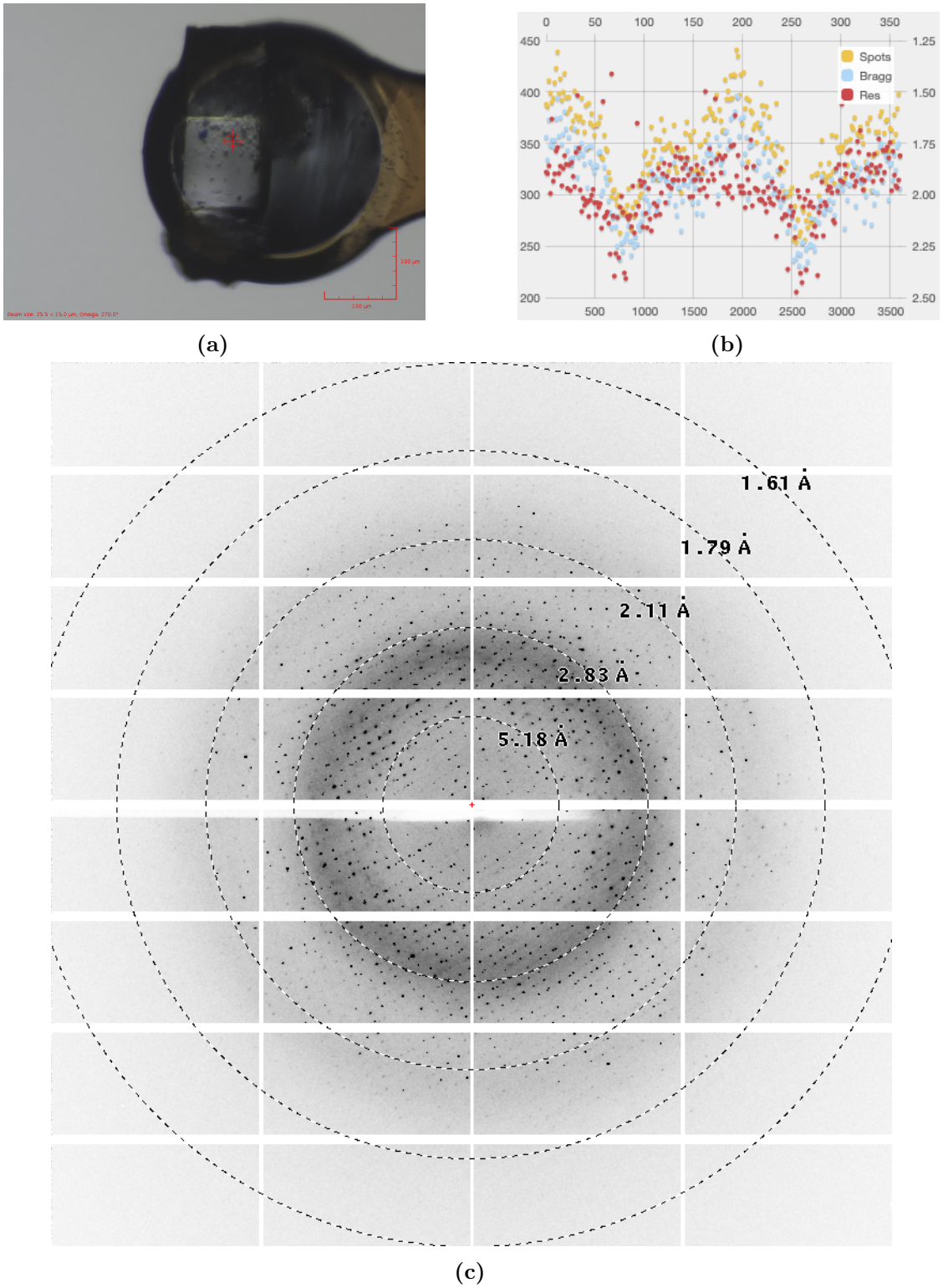
(e)

Figure 3.24: Comparison of Fab D08 and Fab D08* knob regions. *(Figure legend is continued on the next page...)*

Figure 3.24: Comparison of Fab D08 and Fab D08* knob regions. *(Figure legend is continued from the previous page.)* Side-by-side cartoon representation for (3.24a) D08 and (3.24b) D08* with transparent surface representation overlay. Chains are coloured as in Figs. 3.9 and 3.19. (3.24c) Rotational shift of 13.71° and translational shift of 2.47 Å were calculated using the PyMOL RotationAxis script [263]. Black bar indicates axis of rotation. (3.24d) Indicative atomic position shifts between Fab D08 and Fab D08* knob regions. The heavy chain is coloured dark red in the D08 Fab, and dark magenta in Fab D08* in 3.24c and 3.24d. (3.24e) Visualisation of RMSD differences between superimposed Fab D08 (white) and D08* (coloured blue-to-red for RMSD differences ranging from 0.05 Å to 6.93 Å respectively), prepared using the ColorByRMSD PyMol script [264].

3.6.6 Crystallographic structure of ultralong chain-exchanged Fab 4K3E*

The 4K3E* single crystal dataset (Fig. 3.25 and Table 3.5) was characterised as belonging to the C 1 2 1 spacegroup, with unit cell dimensions of $a = 161.17$ Å, $b = 49.05$ Å, $c = 60.40$ Å and angles of $\alpha = 90.00°$, $\beta = 90.83°$, $\gamma = 90.00°$. Data were collected at each 0.1° oscillation angle over 360° total oscillation, where each image was exposed for 0.008 s with an exposure flux of 2.91×10^{11} photons (Table 2.10). Data were integrated and scaled automatically using the xia2 dials pipeline (Section 2.8.3.6). The asymmetric unit was predicted to contain one Fab 4K3E* monomer, (subsequently validated by structural analysis). The structure was phased by molecular replacement using Phaser [223](Section 2.8.3.7) using 1 copy of each of the protein chain H monomer of bovine ultralong antibody BLV1H12 (PDB ID: 4K3D [78]) and the protein chain L monomer of bovine antibody B4HC-B13LC (PDB ID: 6QN7 [160]) as search models, and refined using REFMAC5 [227] within the CCP4 suite [221]. The final model constitutes 3201 protein atoms, and 303 water molecules (Table 3.5).



(a) (b)

(c)

Figure 3.25: Single crystal X-ray diffraction data collection for ultralong Fab 4K3E* at beamline I04 at Diamond Light Source. Fab 4K3E* crystallised in Morpheus® condition H05 droplet 3. (3.25a) I04 beamline sample camera, showing position and size of incident X-Ray beam. (3.25b) DISTL [212] per-image analysis showing number of diffraction spots (total and Bragg)(left axis) and estimated resolution (right axis) for each image in the data collection. (3.25c) Example diffraction image composed of 15 summed frames.

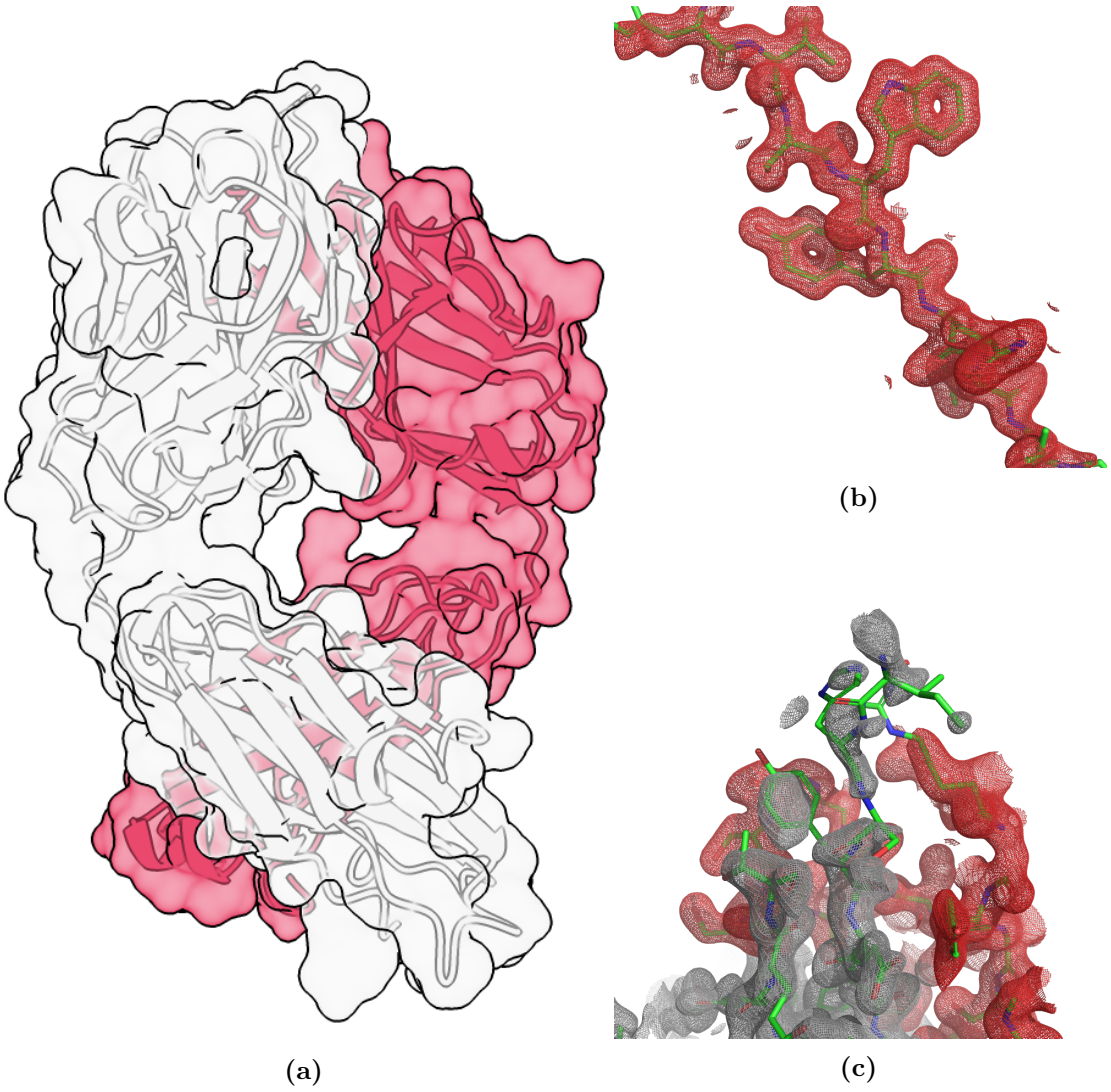
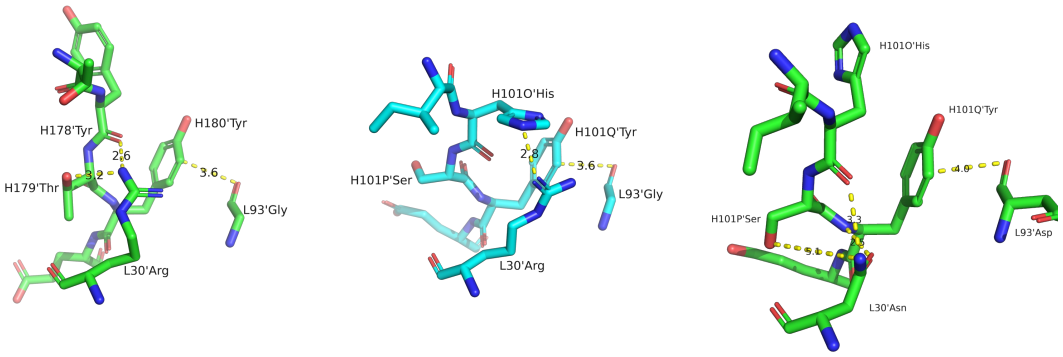


Figure 3.26: Ultralong Fab 4K3E* crystal structure. (3.26a) Overall structure in cartoon representation with transparent surface overlay. The 4K3E heavy chain is coloured in white, the D08 light chain in red. Representative electron density, contoured at 1.0 σ unit, for (3.26b) the tryptophan at the core of the VL domain and (3.26c) the stalk of the ultralong CDR3H, with surrounding (≤ 5 Å) residues from both the heavy chain (mesh coloured white) and the light chain (mesh coloured red). Data were collected at beamline I04 at Diamond Light Source. Figures prepared using ChimeraX [257] (3.26a) and PyMOL (3.26b and 3.26c).

Interestingly, for the Fab 4K3E* structure, the CDR3H is not visible in the electron density map (Fig. 3.26c) in any crystal XRD datasets, and hence may be too flexible to crystallise in a single conformation. The increased flexibility of the 4K3E* CDR3H may be explained by a backbone kink introduced at the base of the CDR3H stalk region between His101O and Ser101P, due to a hydrogen bond interaction between a Nitrogen of His101O and the lone pair of the guanidine group of Arg30 of the D08 light chain (Fig. 3.27).This large missing region (42 amino acids) contributes to the poorer refinement statistics. Elsewise, continuous electron density was observed for both chains for Fab 4K3E*, except the last eight residues of the heavy chain C terminus, and the loops between residues 208 and 212, and between residues 263 and 271 of the CH domain.



(a) D08 **(b)** 4K3E* **(c)** 4K3E

Figure 3.27: Molecular interactions at the base of the ultralong CDR3H loop. Fabs (3.27a) D08, (3.27b) Fab 4K3E*, and (3.27c) PDB ID: 4K3E [78]. Only interacting residues are shown for clarity, with the Fab D08 and 4K3E CDR3H loops extend beyond the top of the figure, whilst variable domains extend below. Note that the 4K3E heavy chain is numbered using the Chothia numbering scheme [265, 266].

The spacegroup of the 4K3E* crystal grown in Morpheus® condition H05 droplet 3 was characterised as C 1 2 1, unlike the P 2_1 2_1 2_1 spacegroup assigned to Fab D08 and D08* crystals grown in Morpheus® condition C09 droplet 1 and H05 droplet 3 respectively (Table 3.5), and to Fab 4K3E [78]. Additionally, the elbow angle between the variable and constant domains of 4K3E* was expanded to 157°, compared to the 124°, 133°, or 132° of Fabs D08, D08* and 4K3E respectively (calculated using the PyMOL Elbow Angle script [267]), though it remains unclear as to whether this is inherent or consequent of the alternative crystal packing (Fig. 3.28).

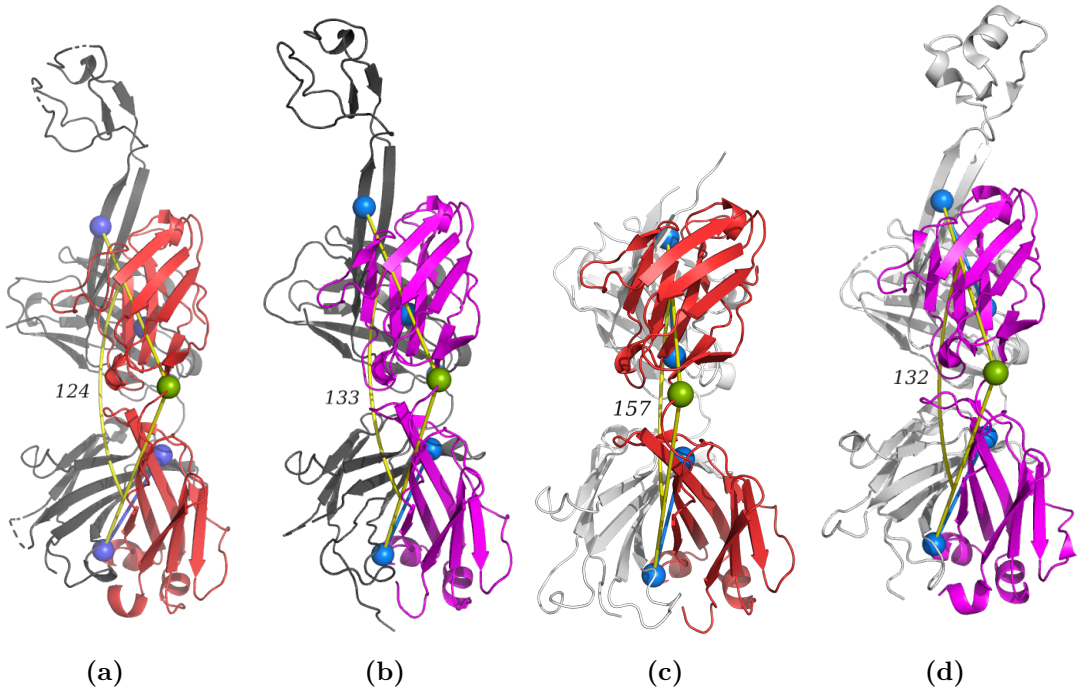


Figure 3.28: Elbow angles for (3.28a) Fab D08, (3.28b) Fab D08*, (3.28c) Fab 4K3E*, and (3.28d) Fab 4K3E, as calculated by the PyMOL Elbow Angle script [267]**.** Chains are coloured as in Figs. 3.9, 3.19 and 3.26. Blue dumbbells indicate calculated centre of mass for the variable and constant domains, green spheres show selected hinge residue. Calculated elbow angle shown between variable and constant domains.

Table 3.5: Crystallographic data of antigen-unknown ultralong CDR3H-containing Fabs. Values in parentheses are for the highest resolution shell. †Parentheses indicate number of reflections with Free R flag.

	D08	**D08***	**4K3E***
Data Collection			
Spacegroup	P 2_1 2_1 2_1	P 2_1 2_1 2_1	C 1 2 1
Cell Dimensions			
a, *b*, *c* (Å)	63.50, 71.11, 98.41	65.03, 70.86, 96.01	161.17, 49.05, 60.40
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.83, 90.00
Resolution (Å)	57.64 - 1.59 (1.62 - 1.59)	39.74 - 1.45 (1.48 - 1.45)	80.57 - 1.46 (1.49 - 1.46)
R_{merge}	0.062 (1.144)	0.076 (1.946)	0.076 (2.019)
R_{pim}	0.033 (0.710)	0.029 (0.806)	0.029 (0.847)
$I/\sigma I$	14.0 (1.3)	10.7 (0.5)	12.0 (0.4)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	100.0(100.0)
Redundancy	6.7 (6.8)	6.6 (5.7)	6.6 (5.7)
$CC_{1/2}$	1.0 (0.6)	1.0 (0.3)	1.0 (0.4)
Refinement			
Resolution (Å)	57.64 - 1.59	39.74 - 1.45	80.57 - 1.46
No. reflections	60540 (3079)†	79199 (5487)†	78110 (5807)†
R_{work}/R_{free}	0.178 / 0.221	0.185 / 0.228	0.224 / 0.259
No. atoms:			
Protein	3546	3696	3201
Water	597	493	303
B-factors:			
Protein	23.26	29	23
Water	42.32	41	43
R.m.s. deviations:			
Bond lengths (Å)	0.0081	0.012	0.012
Bond angles (°)	1.481	1.7	1.9

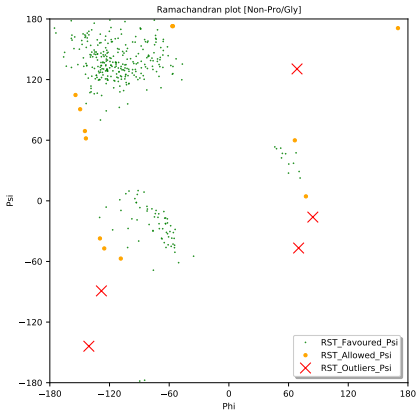


Figure 3.29: 4K3E* atomic model ramachandran plot.

3.6.7 PISA crystal packing analysis

Again, to assess the extent of crystal packing affecting the CDR3H structure, PISA analysis [233] was performed on the refined atomic model for Fab D08* within CCP4 [221] (Table 3.6).

The ultralong CDR3H was involved in six of the 16 identified interfaces by PISA, involving 13 residues (Table 3.6 and Fig. 3.30), compared to that of D08; in which 18 residues were identified as participating in a macromolecular interface.

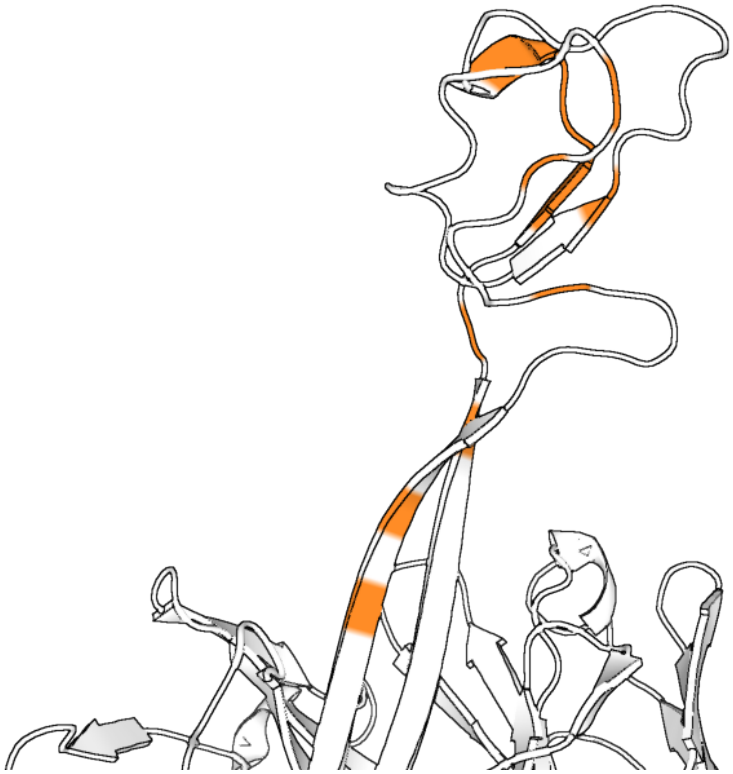


Figure 3.30: Ultralong CDR3H residues participating in a macromolecular interface. Residues in orange interface with a symmetry copy of Fab D08*. Interfaces were identified using PISA [233].

Table 3.6: PISA analysis for Fab D08*.

(a) NN; interface number. iNat and iNres; number of atoms and residues participating in the interaction respectively. 'x' indicates interface is crystallographically linked. 'f' indicates ligands fixed in place during assembly searching.

	Structure 1					Structure 2					
NN	Range	iNat	iNres	Surface (Å^2)	x	Range	Symmetry op-n	Sym.ID	iNat	iNres	Surface (Å^2)
1	L	216	62	10990	-	H	x,y,z	1_555	225	59	14961
2	H	72	23	14961	-	L	x-1/2,-y+3/2,-z	3_465	61	21	10990
3	H	52	14	14961	x	H	x,y,z-1	1_554	53	14	14961
4	H	41	14	14961	x	H	-x,y-1/2,-z+1/2	4_545	41	15	14961
5	L	38	13	10990	-	H	-x,y-1/2,-z+1/2	4_545	35	13	14961
6	L	37	13	10990	-	H	-x,y-1/2,-z-1/2	4_544	31	12	14961
7	H	23	11	14961	x	H	x-1/2,-y+3/2,-z	3_465	23	7	14961
8	[MRD]H:312	7	1	276	f	H	x,y,z	1_555	31	15	14961
9	H	21	6	14961	x	H	-x,y-1/2,-z-1/2	4_544	21	9	14961
10	L	19	7	10990	x	L	-x+1/2,-y+1,z-1/2	2_564	24	10	10990
11	L	9	3	10990	x	L	-x,y-1/2,-z-1/2	4_544	8	3	10990
12	L	12	5	10990	-	H	-x-1/2,-y+1,z-1/2	2_464	9	3	14961
13	L	10	3	10990	-	H	x-1/2,-y+3/2,-z	3_465	12	6	14961
14	[NA]H:311	1	1	125	f	H	x,y,z	1_555	12	8	14961
15	L	7	2	10990	-	H	x,y,z-1	1_554	3	1	14961
16	[NA]H:311	1	1	125	-	L	x-1/2,-y+3/2,-z	3_465	1	1	10990

Table 3.6: PISA analysis for Fab D08*.

(b) Solvation free energy gain (ΔiG). Numbers of different types of bonds; hydrogen (NHB), salt bridges (NSB), disulphide bonds (NDS). Complexation significance score (CSS).

NN	interface area (Å^2)	ΔiG (kcal/mol)	ΔiG P-value	NHB	NSB	NDS	CSS
1	2002.1	-32.0	0.054	11	4	1	0.860
2	556.9	-0.1	0.862	10	2	0	0.000
3	447.8	-6.1	0.285	7	0	0	0.000
4	371.8	-0.7	0.754	7	0	0	0.000
5	320.7	0.4	0.830	7	0	0	0.000
6	316.4	-0.5	0.744	6	0	0	0.000
7	197.9	0.3	0.810	3	0	0	0.000
8	197.7	2.6	0.537	0	0	0	0.000
9	172.4	-0.6	0.704	5	0	0	0.000
10	171.0	0.1	0.640	4	0	0	0.000
11	82.7	-0.5	0.462	1	0	0	0.000
12	80.8	0.6	0.747	2	4	0	0.000
13	76.1	-0.6	0.474	1	0	0	0.000
14	67.6	-9.4	0.000	0	0	0	0.195
15	46.9	1.1	0.779	1	0	0	0.000
16	3.6	-0.3	0.000	0	0	0	0.000

3.7 Assessment of AlphaFold 2 prediction accuracy

To assess the accuracy of AF2 structural prediction for ultralong antibodies, the variable domains of antibody D08 (Section 3.3) were joined *in silico* with a polyglycine linker to form an scFv D08 (Table 3.7), and submitted to AF2 on the BMRC cluster.

Table 3.7: Single chain variable fragment (scFv) sequences for antibodies D08. Polyglycine linker is emphasised.

scFv D08	1	QVQLRESGPS LVKPSQTLSL TCTASGSSLS DEAVGWVRQA PGKSLEWLGS 50
	51	IDTGGNTGYN PNLKTRLSIT KDNSKSQVSL SMSSVTPEDS ATYFCATVHQ 100
	101	ETHQTCPDGY NSGDDCGRGN WDCGTLDCWR CDWGGFCRAS TDYRSVTATY 150
	151	TYEWYIDTWG QGLLVTVSS**G GGGGGGGG**QG VLTQPSSVSG SLGQRVSITC 200
	201	SGSSSNVGRG YVSWYQMTPG SAPRTLIYGD TNRASGVPDR FSASRSGNTA 250
	251	TLTISSLQAE DEADYFCASA EGSSSNAVFG SGTTLTVL 288

AF2 predictions depend on the training database used to train to the neural network, which for the BMRC instance of AF2, includes a vast dataset reflective of the PDB [162]. It is no surprise then, that antibody sequences are well represented, though ultralong CDR3H sequences were less well represented (Fig. 3.31)

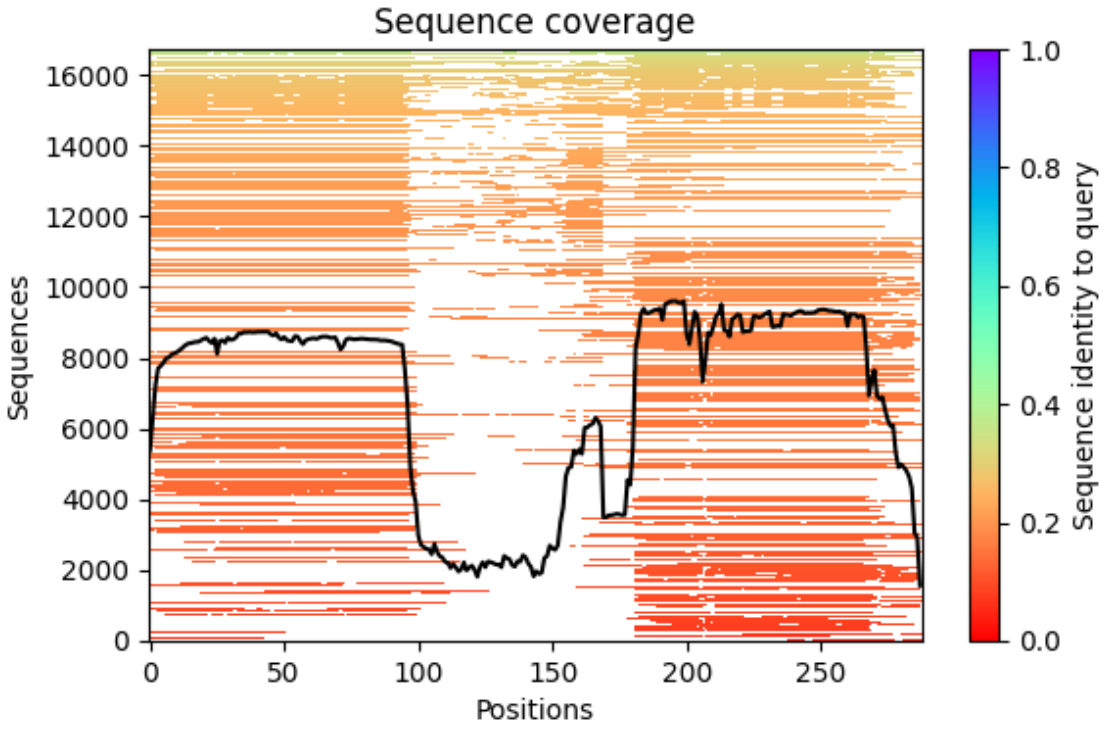


Figure 3.31: PDB Sequence coverage depth for scFv D08.

The predicted Local Distance Difference Test (pLDDT) provides a per-reside confidence level, based on neighbouring chemistry. AF2 predicted the variable domains of scFv D08 with strong accuracy, but showed a drop in accuracy over the ultralong CDR3H, so caution is advised in the interpretation of atomic positions in this locus (Fig. 3.32a). Though, as described in (Section 4.4.3.3), this may be resultant of the reduced local chemistry that contributes to pLDDT calculation.

The Predicted Alignment Error (PAE) is another metric offered for interpretation of AF2 predictions. PAE is useful for assessing the placement of protein domains relative to one another, based on similarly positioned protein folds in the PDB [162]. PAE of scFv D08 indicates strong prediction of the Ig domains, but the

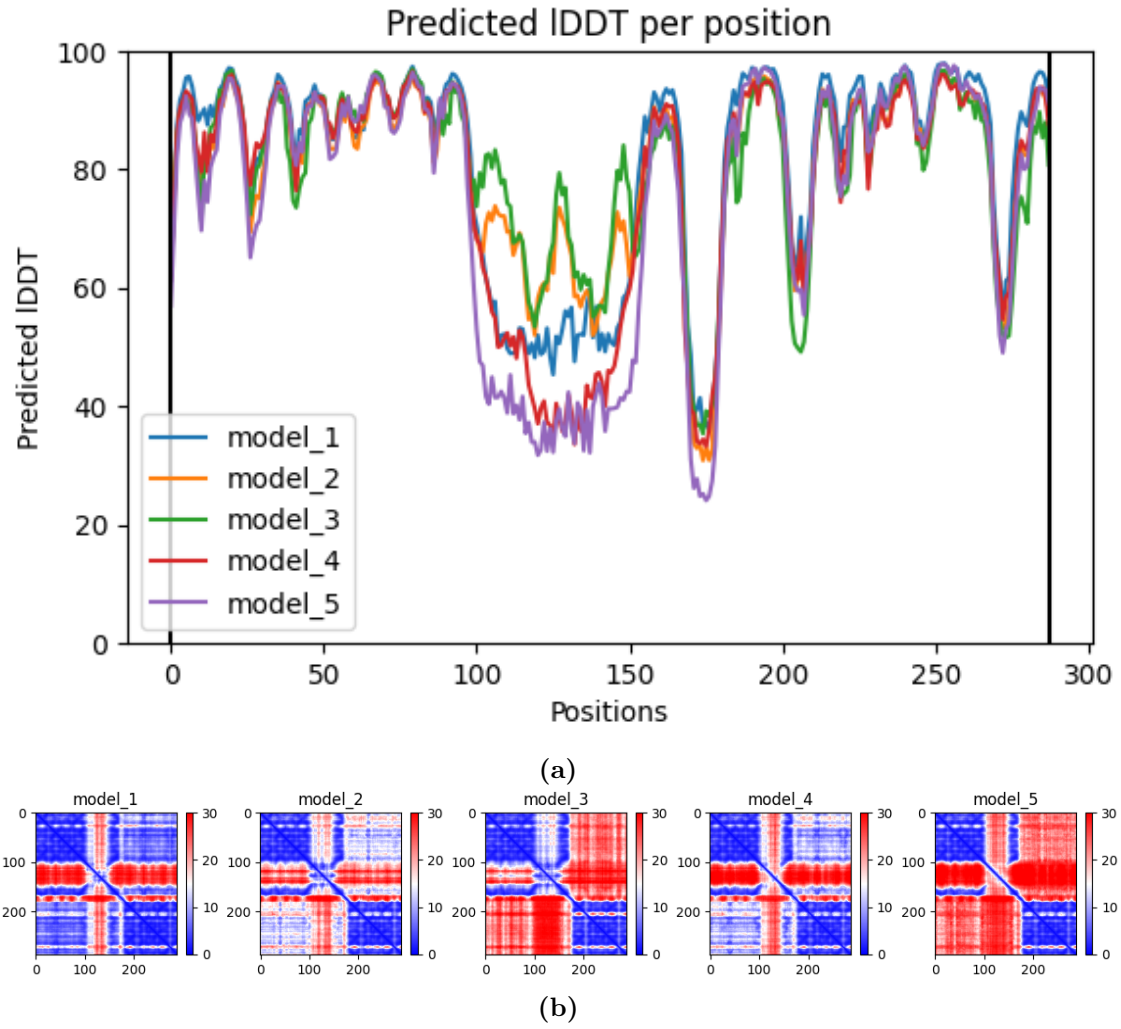


Figure 3.32: Validation metrics for AF2 modelled scFv D08 structures. (3.32a) Predicted local distance difference test (pLDDT) per-residue scores, and (3.32b) predicted alignment error (PAE) for each AlphaFold v2 [204] model of scFv D08.

relative position of the CDR3H loop has a greater predicted error relative to the position of the Ig domains (Fig. 3.32b).

The five AF2 models were structurally aligned to the VH domain of the Fab D08 crystal structure (Section 3.5.3), and coloured by pLDDT, where blue indicates strong confidence in the prediction, reducing down towards red indicating no confidence regions. The models were ordered in decreasing RMSD when aligned to the Fab D08 crystal structure (Fig. 3.33).

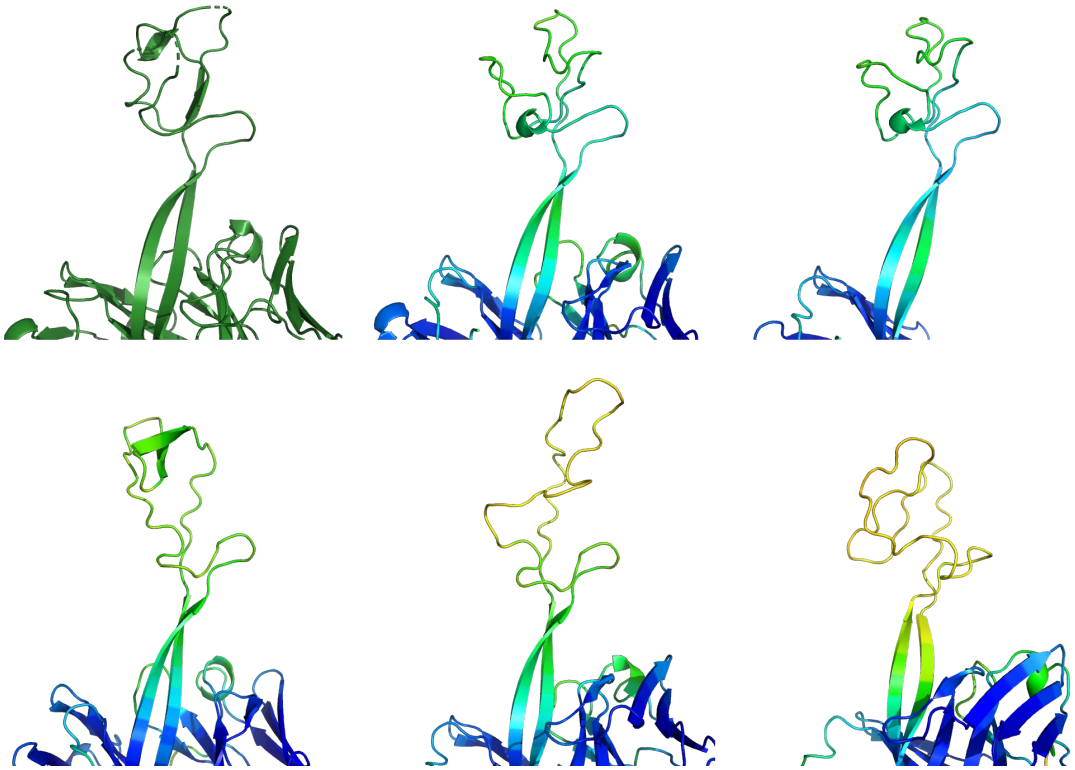


Figure 3.33: Predicted scFv D08 models versus experimentally determined crystal structure. Models are coloured by pLDDT, where blue indicates strong confidence (> 90%), green is good confidence (> 70%), amber is low confidence (>50%), and red is no confidence (< 50%). The models are ordered by decreasing RMSD when aligned to the VH domain of the Fab D08 crystal structure. Fab D08 crystal structure is box 1, coloured green.

In boxes #2 and #3 (Fig. 3.33), the threading and overall conformation of the ultralong CDR3H is remarkably similar to that of the observed crystal structure (box #1), however these were not the highest confidence models produced by AF2; indeed these where the 3rd and 2nd highest ranked models respectively. In all models, the CDR3H stalk had a very similar twist to the crystal structure, interestingly, even in absence of the VL domain, as seen in box #3. Overall, these findings show that AF2 can now predict ultralong CDR3H with reasonable accuracy, however experimental evidence is still required to identify the best model, which may not be the highest confidence prediction.

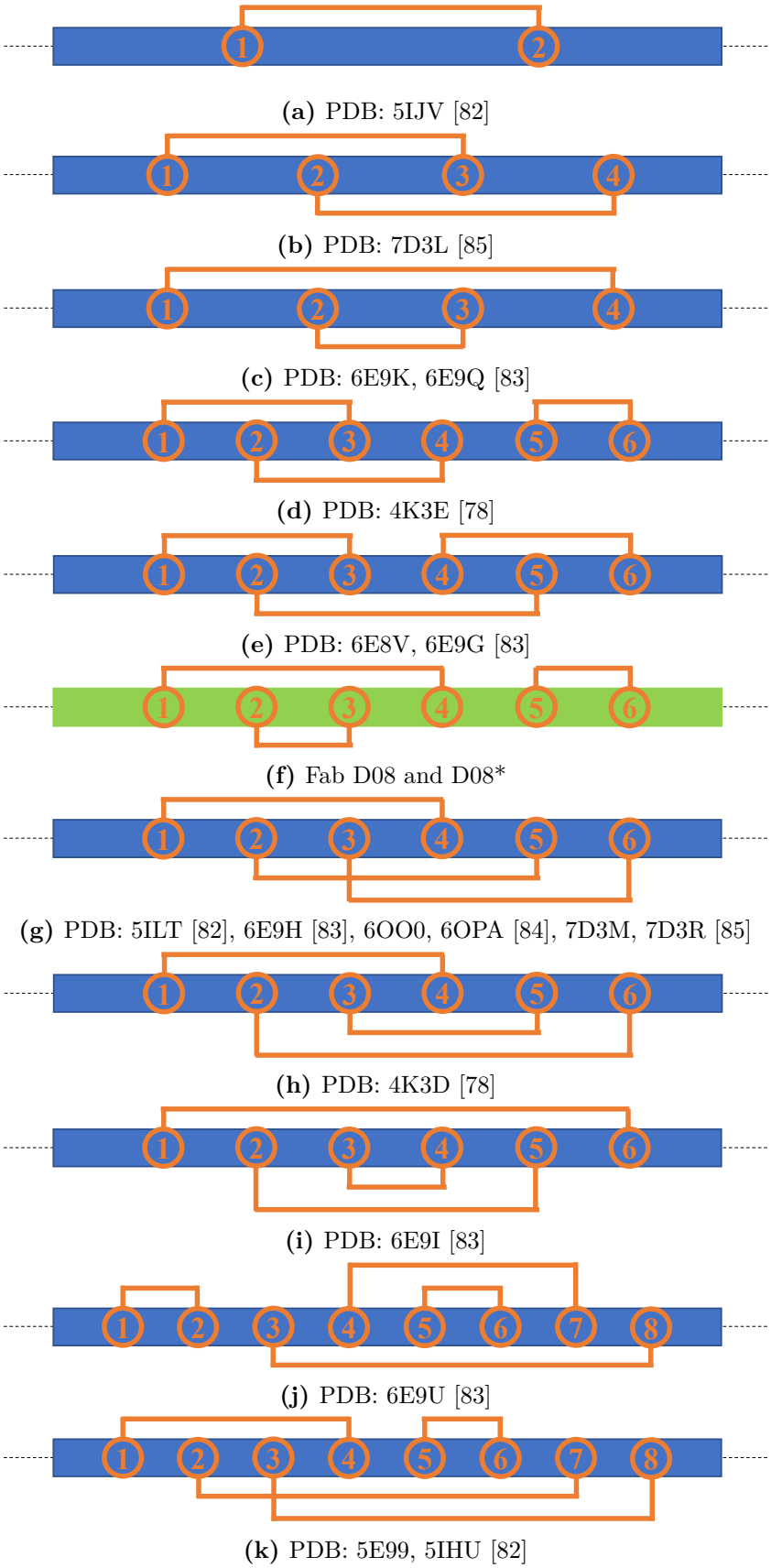
3.8 Discussion

3.8.1 Ultralong antibody D08 structure

The crystal structure of ultralong Fab D08 constitutes only the second observation of a naturally paired ultralong fragment antibody structure (the other being the NC-Cow1 structure of an anti-HIV antibody, PDB: 6OO0 and 6OPA [84]), given all other ultralong antibody structures (PDB IDs: 4K3D, 4K3E [78], 5E99, 5IHU, 5IJV, 5ILT [82], 6E8V, 6E9G, 6E9H, 6E9I, 6E9K, 6E9Q, 9E9U [83]) were synthetically paired with the same light chain originally identified in [31], or were engineered as scFv (PDB IDs: 7D3L, 7D3M, 7D3R, 7FEI, 7FEJ [85]). This antibody also demonstrates a novel arrangement (1-4, 2-3, 5-6) of disulphide bonds supporting the 3-pass β-sheet knob domain (Fig. 3.34), further confirming the structural diversity, and hence immunological potential of the bovine ultralong CDR3H.

3.8.2 Proposed immunological significance of specific ultralong heavy-light chain pairs

Within bovine ultralong antibodies, the interface chemistry between paired antibody heavy and light chains contributes to the structure of the ultralong CDR3H loop, highlighting the importance of paired-chain sequencing when surveying bovine antibody repertoires, despite the restricted light chain sequence diversity. Residues neighbouring the CDR3H loop from both chains contribute to the positioning of the loop, which subtly perturbs the presented ultralong paratope by warping the knob domain. Considering the ultralong CDR3H loop of bovine antibodies makes the major contact with antigen [78, 132, 138, 139], these paratope modifications may be sufficient to alter the performance of the antibody, without changing the sequence of the antigen-contacting CDR3H.



(a) PDB: 5IJV [82]

(b) PDB: 7D3L [85]

(c) PDB: 6E9K, 6E9Q [83]

(d) PDB: 4K3E [78]

(e) PDB: 6E8V, 6E9G [83]

(f) Fab D08 and D08*

(g) PDB: 5ILT [82], 6E9H [83], 6OO0, 6OPA [84], 7D3M, 7D3R [85]

(h) PDB: 4K3D [78]

(i) PDB: 6E9I [83]

(j) PDB: 6E9U [83]

(k) PDB: 5E99, 5IHU [82]

Figure 3.34: Disulphide networks supporting ultralong CDR3H. Fab D08 and D08* are highlighted (3.34f).

Equally, it should be remembered that where the light chain does not directly contact antigen in ultralong antibodies, this ultralong CDR3H warping effect is applied at distance, and thus light chains make a greater contribution to the ultralong antibody than currently attributed to them as a mere scaffold component [78] (though the flexible CDR3H loop of Fab 4K3E* confirms the structural importance of the light chain). Therefore, their selection, gene segment usage, and maturation, may lead to specific influence particularly in respect to antigen affinity. To explore this further, ultralong antibodies (and chain-exchanged versions thereof) must be examined in the context of antigen, as explored in Chapter 4.

3.8.3 Suggestions for further analysis

To expand on this study, mutagenesis of both the CDR3H stalk and surrounding amino acids of both the heavy and light chains could be used to identify residues providing essential support for the ultralong CDR3H loop, and perhaps more broadly, to explain why light chains which pair with ultralong heavy chains are highly conserved within the bovine antibody repertoire. In combination with biophysical assessment of CDR3H order or flexibility (such as by nuclear magnetic resonance (NMR) spectroscopy), synthetic chain-exchanged ultralong antibodies could be predicted, allowing functional optimisation of ultralong CDR3H in antibody engineering efforts.

When you're curious, you'll find lots of interesting things to do...

— Walt Disney, animator and entrepreneur

4

Structural and functional characterisation of ultralong antibodies specific for Foot-and-mouth disease virus

Contents

4.1 Introduction to Foot-and-mouth disease 146
4.1.1 Importance and epidemiology of FMD 146
4.1.2 Foot-and-mouth disease virus 148
4.1.3 Control of foot-and-mouth disease 149
4.1.4 Challenges of inactivated FMDV vaccines 151
4.1.5 Development of VLP vaccine 152
4.2 Sequence Discovery 153
4.2.1 Vaccination experiment 153
4.2.2 Serospecificity assay 155
4.2.3 Sequencing 157
4.2.4 Bioinformatic sequence clustering 161
4.3 Transient transfection of anti-FMDV Fabs 164
4.3.1 FMDV-specific Fab expression screening 164
4.3.2 Purification 166
4.4 Fab apo-structures 168
4.4.1 Crystallography 169
4.4.2 Crystallography of Fabs #117 and #126 169
4.4.3 AlphaFold Structural Prediction 171
4.4.4 Crystallography of ultralong Fab #30 181
4.4.5 Crystallographic structure of Fab #30 182
4.5 Production of FMD VLPs by vaccinia virus co-infection 187
4.6 Fab holo-structures 188
4.6.1 Electron cryomicroscopy specimen preparations 188
4.6.2 Electron cryomicroscopy data collections 189

4.6.3 Interaction of SAT2 12S pentamers by Fab #117 191
4.6.4 Interaction of SAT2 12S pentamers by Fab #126 197
4.6.5 Introduction of thermostable SAT2 VLPs 202
4.6.6 Interaction of O1M VLPs by Fab #126 206
4.6.7 Interaction of O1M VLPs by Fab #117 212
4.6.8 Interaction of O1M VLPs by Fab #30 218
4.7 Quantifying ultralong antibody interactions with antigen 223
4.7.1 BLI Experiment optimisation 223
4.7.2 BLI characterisation of Fab #30 with O1M VLPs . . . 223
4.7.3 BLI characterisation of Fab #126 with O1M VLPs . . . 224
4.7.4 BLI characterisation of Fab #117 with O1M VLPs . . . 225
4.8 Discussion and future work 229
4.8.1 Systemic twist predicted by AlphaFold 229
4.8.2 Antigenic interactions mediated by ultralong knob domain 231
4.8.3 Fab BLI behaviour 231
4.8.4 Quantifying the serospecificity of chain-exchanged Fabs #117 and #126 232
4.8.5 Future work 233

Figures

4.1 Serial vaccination to produce cross-protective antibodies 154
4.2 Somatic Hypermutation of AbCellera-discovered antibody sequences 158
4.3 CDR3H lengths of AbCellera-discovered antibody sequences . . 159
4.4 Sequence clustering of antibody sequences 160
4.5 Multiple Sequence Alignment of antibody #117 and #126 variable domains 162
4.6 Multiple sequence alignment of ultralong-paired CDR3L loops . 163
4.7 Expression efficiency screening for clustered ultralong antibodies 165
4.8 Purification of Fabs #117 and #126 167
4.9 Purification of Fab #30 168
4.10 Crystallisation screening for Fab #117 and #126 170
4.11 PDB Sequence coverage depth for scFv #117 172
4.12 Predicted models and alignment error for scFv #117 models . . 174
4.13 Predicted Local Distance Difference Test for scFv #117 175
4.14 PDB Sequence coverage depth for scFv #126 176
4.15 Predicted models and alignment error for scFv #126 models . . 177
4.16 Predicted Local Distance Difference Test for scFv #126 178
4.17 Structural superposition of AF2 predicted knob mini-domains . 179
4.18 Structural overlay of scFv #117 and #126 AlphaFold models . 180
4.19 Crystal morphology for Fab #30 183
4.20 X-Ray diffraction data collection for Fab #30 at I03 184
4.21 Crystal structure of the ultralong Fab #30 185
4.22 Reducing SDS-PAGE analysis of O1M VLPs 187
4.23 Electron cryomicroscopy of Fab #117 in complex with FMD serotype SAT2 dissociated 12S pentamers 192

4.24 2D class averages for SAT2 in complex with Fab #117 193
4.25 Fitting of SAT2 pentamer and AF2 scFv #117 to observed cryo-EM density map 194
4.26 Comparison of observed and predicted #117 knob mini-domain structures 195
4.27 The SAT2 12S epitope for antibody #117 196
4.28 Electron cryomicroscopy of Fab #126 in complex with FMD serotype SAT2 dissociated 12S pentamers 198
4.29 2D class averages for SAT2 in complex with Fab #126 199
4.30 Comparison of density maps for SAT2 12S in complex with Fab #117 or #126 201
4.31 Electron cryomicroscopy of FMD serotype SAT2/Eth VLPs . . 203
4.32 Electron cryomicroscopy of FMD serotype SAT2/Eth VLPs mixed with Fab #117 204
4.33 Electron cryomicroscopy of FMD serotype SAT2/Eth VLPs mixed with Fab #126 205
4.34 Electron cryomicroscopy of Fab #126 in complex with FMD serotype O1M VLPs 207
4.35 2D class averages for O1M in complex with Fab #126 208
4.36 Observed O1M/Fab #126 density map coloured by determined local resolution 209
4.37 Inference of the O1M Fab #126 epitope site 211
4.38 Electron cryomicroscopy of Fab #117 in complex with FMD serotype O1M VLPs 213
4.39 2D class averages for O1M in complex with Fab #117 214
4.40 Observed O1M/Fab #117 density map coloured by determined local resolution 215
4.41 Difference map showing density not explained by O1M crystal structure 217
4.42 Electron cryomicroscopy of Fab #30 in complex with FMD serotype O1M VLPs 219
4.43 2D class averages for O1M in complex with Fab #30 220
4.44 Fab #30 interacts with O1M VLPs at the 3-fold symmetry axis 221
4.45 Fitting of O1M capsid and AF2 Fab #30 to observed cryo-EM density map 222
4.46 BLI characterisation of the Fab #30-O1M interaction 224
4.47 BLI characterisation of the Fab #126-O1M and #117-O1M interactions 226
4.48 BLI characterisation of the Fab #63-O1M interaction 227
4.49 Repeated BLI characterisation of the Fab #126-O1M interaction 228
4.50 Comparison of initial and revisited AF2 knob mini-domain predictions 230

Tables

4.1 AbCellera sequencing summary 156
4.2 Predicted cross-reactive serospecificities from microfluidic approach 156

4.3 Clustering of identified FMDV-specific CDR3H sequences . . . 161
4.4 Antibodies of interest and their serospecificities 164
4.5 Recombinant Fab yields obtained by transient transfection of Expi293F™ cells 166
4.6 Single chain variable fragment sequences for antibodies #117 and #126 172
4.7 Structural superposition matrix of AF2 scFv #117 knob mini-domains 175
4.8 Structural superposition matrix of AF2 scFv #126 knob mini-domains 178
4.9 Crystallographic data table for Fab #30 structure 186
4.10 Chapter 4 cryo-EM specimen preparation and data collection . 190
4.11 Chapter 4 cryo-EM exposure curation and particle picking . . . 190
4.12 BLI characterisation of the Fab #30-O1M interaction results table 225
4.13 BLI characterisation of the Fab #126-O1M interaction results table 227

4.1 Introduction to Foot-and-mouth disease

The results reported in Chapter 3 show that chain exchange within ultralong antibodies can vary the axial twist of the β-strand stalk region of the ultralong CDR3H loop. The ultralong knob mini-domain at the end of the β strand stalks was only marginally affected by this twist, with preservation of the core disulphide bond positions, and accommodating side chain positional shifts. A similar effect was observed in the non-ultralong B13* chain-exchanged structure (Fig. 5.1), resulting in two-fold difference in affinity [160].

To investigate the functional implications of light chain sequence on antigen recognition within the ultralong antibody system, a chain-exchange experiment was devised using antibodies identified as part of a different project of the Immunogenetics group at Pirbright; investigating host responses to foot-and-mouth disease (FMD), particularly in cattle (Section 4.2).

4.1.1 Importance and epidemiology of FMD

Foot-and-mouth disease is a highly contagious viral infection that principally affects cloven-hoofed ruminants. Infection is characterised by lesions in and around the

mouths, teats, and between the hooves of infected animals, as well as fever. Lesions around the mouth and fever can reduce appetite, affecting milk production and weight gain, whilst lesions at the hooves can cause lameness. The disease causes myocarditis in young animals (with up to 100 % morbidity) which can be fatal (up to 20 % mortality) [268]. Mortality is less common in adults (1 - 5 %), typically resulting in severe and irrecoverable milk [269] and meat production losses, and decreased herd fertility. FMD is clinically indistinguishable from other vesicular diseases (such as vesicular stomatitis), necessitating urgent laboratory confirmation by observation of FMD virus, viral antigen or nucleic acid. Additionally, FMD infection can cause a prolonged asymptomatic infection that persists, resulting in some animals becoming carriers.

Foot-and-mouth disease is transmitted by direct contact with an infected animal, or contaminated surfaces such as fomites (including clothing, vehicles and equipment), long-range and short range airborne droplet transmission, transfer of contaminated bodily fluids (such as milk, sputum, and including artificial insemination) or consumption of contaminated waste as food (swill feeding).

The contagiousness of FMD, where an estimated 77 % of all livestock carry the disease, resulted in the unprecedented establishment of an official list of FMD-free countries by the World Organisation for Animal Health (WOAH), formerly the Office International des Epizooties (OIE). Wildlife species are more susceptible to FMD than domesticated species, equally high production breeds (such as Holsteins) are more susceptible than low production breeds.

Long ago, FMD was prevalent worldwide, but now North America, Western Europe and Oceania have been declared FMD-free regions. Conversely, FMD is endemic in Africa, Asia, and the Middle East, and has sporadic outbreaks in South America. Widespread endemic FMD, little-to-none cross-serotypic protection (Section 4.1.4), and persistent viral reservoirs in asymptomatic or subclinical wildlife carriers such as the African buffalo (*Syncerus caffer*) [270], prevent global eradication.

At present, vaccines exist, but there is no treatment for FMD. Outbreaks can be devastating (Section 4.1.3). In endemic regions, the estimated visible cost of FMD resulting from production losses and vaccination costs was estimated at US$6.5 - 21 billion in 2013 [271], whist outbreaks in FMD-free regions result in losses exceeding US$1.5 billion each year.

4.1.2 Foot-and-mouth disease virus

The causative agent of FMD is foot-and-mouth disease virus (FMDV), belonging to the *Aphthovirus* genus of the *Picornaviridae* family. The virus is geographically grouped into seven 'pools', where each pool constitutes different combinations of six FMDV serotypes; A [272], O [272], SAT1 [273], SAT2[273], SAT3[273], and Asia 1 [274]. A seventh FMDV serotype, C [275], has not presented any new cases since 2004 [276], and is now considered extinct [277]. Within each serotype, strains are further classified into topotypes denoting the location of initial isolation, and lineages denoting the specific strain [278].

The FMDV genome is an 8.5 kb non-segmented positive sense, single stranded, RNA chromosome, encoding a 2332 residue polyprotein. The polyprotein is ultimately cleaved to four structural proteins; viral protein 1 (VP1) through 4, and eight non-structural proteins including two proteases (L^{pro} and $3C^{pro}$, which both cleave the polyprotein) and an RNA polymerase ($3D^{pol}$) that lacks proof-reading activity [279]. Consequently, the FMDV genome has a high mutational frequency due to error-prone RNA [280, 281].

One copy of each FMDV structural protein self assembles into a protomer unit comprising the precursor VP0, VP1 and VP3. The protomer units then self assemble to form 12S pentameric units, which package the replicated genome within a 140S icosahedral capsid [181, 182]. In total, 60 copies of the protomeric unit form an intact FMDV capsid [182]. Dependent on the presence of viral RNA, the precursor VP0 is cleaved to VP2 and VP4, resulting in the viral capsid principally being composed of VP1:3, with VP4 present at the internal surface of the capsid. Viral proteins 1:3 are truncated relative to other *Picornaviridae*, with notable differences

at VP1 and VP2 loops external to the capsid [182]. As a result, the FMDV capsid is unusually featureless for a *Picornavirus* [182], lacking the canyon present in other *Picornaviridae* genera; for example human rhino virus 14 of the *Rhinovirus* genus [282], poliovirus of the *Enterovirus* genus [283] and Mengo virus of the *Cardiovirus* genus [284]. Truncation of VP1 loops, and associated loss of external capsid density, also exposes a deep hydrophobic tunnel at the 5-fold symmetry axis [182].

The virion engages cells via the GH loop (residues 133-158) of VP1, specifically an Arg-Gly-Asp tripeptide which interacts with integrin, principally αvβ6 of epithelial cells, but can also enter other cells by recognition of integrins αvβ1, αvβ3, and αvβ8 in other tissues [285]. Many neutralising antibodies target this loop, therefore achieving neutralising by blocking cell entry.

4.1.3 Control of foot-and-mouth disease

In absence of FMD-specific treatment options, the disease is controlled by (1) surveillance, (2) preventative policymaking, (3) prophylaxis and (4) euthanasia.

All confirmed cases of FMD must be reported to the WOAH. In the United Kingdom, FMD is a notifiable disease, and suspected cases must be reported to the Animal and Plant Health Agency (APHA). Suspected cases must also be reported to the relevant veterinary authority within North America. Efficient reporting allows for rapid responses to outbreaks; monitoring the occurrence and prevalence of FMDV infections, and characterising the circulating viral strain for vaccine selection. Intense surveillance also substantiates the FMD-free status of a given region.

It is commonplace for FMD-free regions to enforce strict importation regulations on both livestock and animal products to prevent FMD entry. Furthermore, farms are actively encouraged to practice good biosecurity measures to prevent infection. The WOAH has published policies for the inactivation of FMDV in animal foodstuffs, however these policies are difficult to enforce, and many countries prefer to adopt a blanket ban on the use of swill feeding for livestock. There are also strict policies on disinfection, restriction of herd movement, and quarantine which are enforced upon case identification. Suspected animals, those that may have been in contact with

the suspected animal, as well as any workers, may be subject to these policies, as are researchers investigating FMDV, even if sufficiently biosafety level (BSL) contained.

Vaccines are used differently to control FMDV dependent on several factors; scientific need, geopolitics, economics, and societal factors. Typically vaccination is used in endemic regions to mitigate the severity of illness, whilst in FMD-free regions, vaccination is selectively used to protect specific animals (such as those belonging to zoological collections) or to hamper the spread of an outbreak due to vaccine costs.

In cases of outbreaks in FMD-free regions, it is typical to humanely destroy any and all animals with either suspected, contacted, confirmed or recovered cases of FMDV, and to appropriately dispose of the resulting carcasses, typically by incineration. Potentially exposed livestock are monitored with heightened surveillance. Rodents and other wildlife may also be culled to prevent mechanical transmission. Such action can devastate entire populations of cattle, as apparent in the United Kingdom 2001 FMD epidemic, where 6.5 million animals were destroyed, and £8 - 9 billion of losses were incurred to prevent the spread, where positive cases were observed from 2026 premises across Great Britain [286].

This estimate does not account for the invisible emotional cost of euthanasia to farmers, the veterinary profession, and those directly working with animals, nor the emotional strain introduced to the general public through various means; the closure of livestock petting centres and rural footpaths, the suspension of events involving working animals such as horse racing, and the appearance (and smell) of mass carcass pyres [287]. Additionally, the public may have been concerned about zoonosis; the recent 1980's and 1990's outbreaks of bovine spongiform encephalopathy (BSE; colloquially termed 'mad cow disease') would have been forefront of people's minds, especially given consumption of BSE-contaminated beef had recently been confirmed as a causal agent for variant Creutzfeldt–Jakob disease (vCJD) [288], and that vCJD had started to cause human fatalities [289]. Finally, the estimated losses does not account for the reputational damage to the UK animal products industry and resulting international trade embargoes.

4.1.4 Challenges of inactivated FMDV vaccines

Live attenuated vaccines are not used for controlling FMDV due to the risk of reversion; instead inactivated vaccines are used [290].

Due to the high rate of FMDV genome mutation, hence low capsid epitope conservation, inactivated FMDV vaccines are maximally effective when strain-matched against the circulating variant of concern. Consequently, effective FMDV vaccination is reliant on accurate surveillance, and requires bespoke responsive culturing of specific strains, hence FMDV vaccination cannot be an off-the-shelf solution, increasing costs.

Preparations of vaccine from live virus will likely contain contaminating viral non-structural proteins, but the greater issue is the need to handle live FMDV prior to inactivation [290]. Handling live FMDV requires specialist facilities as the virus is listed as a category 4 pathogen under the Specified Animal Pathogens Orders (SAPO) 2008 legislation, thus set up and running costs are high.

Vaccine costs are further increased as the integrity of FMDV capsids is principally driven through networks of weak intermolecular bonds between capsid pentamers [182]. Hence inactivated FMD antigen, without thermostability mutation, requires cold chain distribution with associated costs and logistical challenges when reaching more rural communities. The addition of stability mutations is not possible due to the need for capsids to dissociate during the 'live' portion of manufacture. Furthermore, whilst single vaccination doses may be sufficient to quell an emerging outbreak, routine vaccination, especially in FMD-endemic regions, relies on a multidose vaccination regime [291], compounding the cost of vaccination.

Finally, there is some FMDV vaccine hesitancy amongst farmers and the general population. Whilst vaccination may help to prevent an outbreak, there is a delay (typically 14 days in cattle [292, 293]) between vaccination and onset of protection. At the onset of an outbreak, the responses may not be fast enough to protect cattle. Furthermore, if cattle are deemed susceptible enough to warrant vaccination, they may be euthanised eventually anyway if a relevant veterinary authority deems a return to FMD-free status is the target of outbreak control measures. Vaccination is

therefore considered by many among the farming community as a stay of execution, rather than a solution to an outbreak [291]. Additionally, public support for bovine vaccination programmes is questionable since some studies have shown emergency vaccination with inactivated FMDV and subsequent FMDV challenge can result in animals becoming persistent carriers, though such studies used high viral challenges [292]. On the other hand, vaccination has been effective at eliminating FMD from large regions of South America [294].

4.1.5 Development of VLP vaccine

As previously discussed (Section 4.1.4), there is a need to improve FMDV vaccination, with benefits to both FMD-free and endemic regions. Poliovirus capsids can be synthesised recombinantly by co-infection with vaccinia virus (VV) [295], yielding self-assembling empty (hence non-infectious) virus-like particles (VLPs) free from non-structural protein contaminants (except the required FMDV non-structural protease $3C^{pro}$). However difficulties arose when the system was modified for FMDV capsid expression; chiefly, that $3C^{pro}$ (or its catalytic products) is toxic to cells in culture [296–298]. Toxicity of $3C^{pro}$ can be remedied by switching from a constitutive promoter [297] to an inducible element, via the provision of the bacteriophage T7 polymerase by co-infection [299], thereby limiting $3C^{pro}$'s expression. Toxicity of $3C^{pro}$ is further limited by functional knock-down bio-engineering of $3C^{pro}$ [177]. Due to the necessary auto-cleaving function of $3C^{pro}$ (Section 4.1.2), $3C^{pro}$ cannot be wholly eliminated from expression cassettes.

The lack of any FMDV genomic information means VLPs cannot be infectious, and negates the need for high containment in the production of FMDV capsids [176, 177]. Furthermore, capsid-stabilising mutations may be investigated as there is no need to preserve capsid dissociation function. Another benefit is that the VV P1-2A-3C cassette can be adapted to produce field-matched strains of VLPs as necessary with relative ease. X-Ray crystallographic [176], and electron cryomicroscopy experiments [177] have revealed VLPs provide essentially an identical antigen to inactivated FMDV vaccine of the same strain, however vaccination experiments

have subsequently shown that VLPs are not as immunogenic as inactivated virus [296, 300–302], though this difference may be artefactual of antigen quantity [290]. Virus-like particles therefore offer an attractive antigen for FMDV vaccination.

4.2 Sequence Discovery

4.2.1 Vaccination experiment

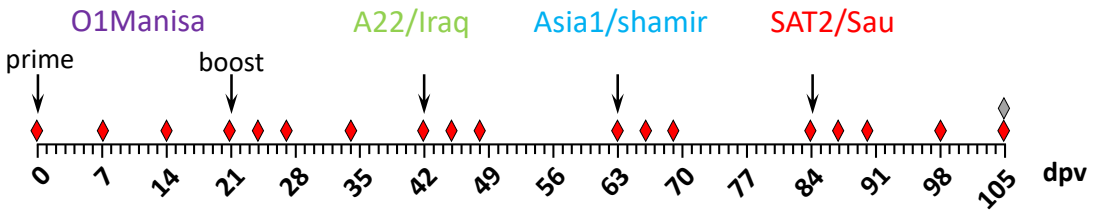
Protection against FMDV is largely conferred by humoral, rather than cellular immunity, but inter-serotype (and even occasionally intra-serotype) antigenic variation is sufficient to evade immune protection [303, 304]. However, sequential challenge of animals with multiple serotypes of FMDV results in short-lived increased resistance to FMDV infection in subsequent challenges [305], highlighting that cross-protection against FMDV serotypes is possible.

To improve cross-protection in the development of next generation vaccines, a serial vaccination experiment was devised at Pirbright, with recovery and characterisation of antibody-secreting PBMCs aiming to produce a cross-reactive paratope (Ramirez Valdez, in preparation).

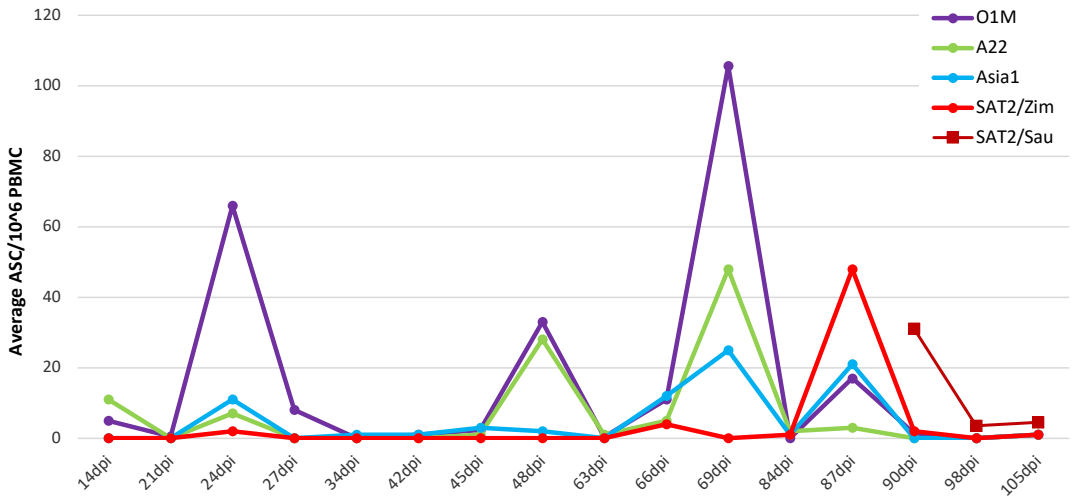
Cattle were serially vaccinated with classical, inactivated FMDV vaccines. Strain O/1/Manisa/Turkey/69 'O1M' classical antigen was used as the prime immunogen (Fig. 4.1a). Every subsequent 21 days, a 'boost' was administered of differing serotype (arrows); O1M on 21 days post immunisation (dpi), strain A/22/Iraq/24/64 'A22' on 42 dpi, strain Asia/1/Shamir 'Asia1' on 63 dpi, and finally strain SAT/2/Saudi Arabia/6/200 'SAT2' on 84 dpi (Fig. 4.1a). Insufficient inactivated SAT2 virus was produced, and so the SAT2 vaccination utilised baculovirus prepared SAT2 serotype VLPs (Section 2.2.6). However, this was performed prior to the observation of dissociated SAT2 VLPs on EM grids (Section 4.6.3). Routine VLP integrity testing by VHH double sandwich ELISA was being performed at the time, however a satisfactory antibody for the capture and detection of dissociated SAT2 VLPs was unavailable, so SAT2 VLP integrity was not assessed. It is likely then, that the administered SAT2 'boost' contained mostly dissociated 12S pentamers.

Antibody #117 (isolated during the sequential vaccination experiment; Section 4.2.4) was later adopted as the detection antibody for dissociated SAT2 12S particles.

Bloods (red diamonds) were taken on 0, 7, and 14 dpi following the O1M prime, and on the day of boost administration, as well as on 3, and 6 days post each boost. Additional bloods were collected on 34 and 98 dpi, and on 105 dpi, when the animals were culled. Left and right prescapular lymph nodes (grey diamonds) were also collected during post-mortem examination of the animals (Fig. 4.1a). PBMCs were isolated from blood samples and cryopreserved in liquid nitrogen.



(a)



(b)

Figure 4.1: Serial vaccination study to produce cross-protective antibodies against FMDV. (4.1a) Serial vaccine regime. Vaccination with indicated strain of inactivated FMDV antigen for O1M, A22, and Asia1. SAT2 immunisation used vaccinia virus-produced VLPs. blood samples were collected on days indicated with red diamonds, left and right prescapular lymph nodes were collected during post-mortem examination; grey diamonds. (4.1b) Representative ELISpot for animal # 3 of the vaccinated group.

ELISpot was used to quantify the frequency of FMDV-specific IgG antibody-secreting cells specific for each serotype at each sampling point (Fig. 4.1b), with a

few caveats. Unlike the classical antigen used in the immunisation experiment, the ELISPot assay used strain-matched stabilised VLPs produced by vaccinia virus due to antigen availability (Section 2.2.6). No vaccinia virus compatible FMD P1-2A-3C cassette is available to produce topotype SAT/2/Saudi Arabia/6/200 VLPs, so these were substituted with SAT/2/Zimbabwe/7/83 VLPs (Section 2.1.5). Initially, bloods taken 87 dpi from animal # 3 appeared the most promising for identifying cross-protective antibodies, as this time point provided evidence for reactivity with all tested serotypes following the final immunisation. However, retrospective knowledge of the SAT2/Zim VLP dissociation clouds this view as the VLP dissociation, likely prior to immunisation, could result in antibodies raised to unintended targets, such as the internal or inter-protomer surfaces of the viral proteins.

4.2.2 Serospecificity assay

In collaboration with AbCellera Biologics Limited (*2215 Yukon Street, Vancouver, British Columbia, V5Y 0A1, Canada*), serospecificity of antibody-secreting cells, (ASCs), were screened using a proprietary microfluidics approach. In brief, ASCs were single cell sorted from the cryopreserved PBMCs into a microwell plate and cultured to produce antibodies. Antigen-specific IgG ASCs were then identified using VLPs conjugated to beads with an additional fluorescence-based marker, providing a qualitative assessment of antibody serospecificity for each sorted ASC (Table 4.2). This microwell system allows visualisation of very low avidity interactions, and allows identification of very infrequent antibodies.

The strains of each serotype used in serospecificity testing matched the vaccination strains, except SAT2, where SAT/2/Zimbabwe/7/83 strain vaccinia virus VLPs were used for the serospecificity assay. All VLPs included stability mutants (Table 2.2). ASCs were then recovered, and antibody heavy and light chain sequences were sequenced.

Serospecificity screening discovered antibodies demonstrating all modalities of serotype cross-reactivity (except triple cross-reactive for O1M-A22-SAT2, and double cross-reactive for O1M-SAT2; and A22-SAT2). For downstream experimentation,

antibodies were prioritised according to several criteria; the multiplicity of predicted cross-reactivity (higher orders of cross reactivity were prioritised), CDR3H length, variety of cysteine content and normalised cluster abundance respective of the total antibody repertoire.

Following serospecificity screening, 349 cells were recovered for sequencing, 209 of which produced both a heavy chain and a light chain, with 143 unique paired chain sequences discovered (Table 4.1).

Table 4.1: AbCellera sequencing summary.

Animal ID	Sample Day	Cells recovered	Paired chains	Unique sequences
2	66 dpi	31	16	11
3	69 dpi	64	49	39
3	87 dpi	248	144	93
	Total:	343	209	143

Table 4.2: Predicted cross-reactive serospecificities from microfluidic approach.

Cross-reactivity	O1M	A22	Asia1	SAT2	Unique Sequences
Quadruple	Y	Y	Y	Y	17
Triple	Y		Y	Y	6
	Y	Y		Y	0
	Y	Y	Y		12
		Y	Y	Y	3
Double	Y	Y			3
	Y		Y		10
	Y			Y	0
		Y	Y		8
		Y		Y	0
			Y	Y	14
Single	Y				51
		Y			8
			Y		4
				Y	7
				Total:	143

4.2.3 Sequencing

No large differences in the degree of somatic hypermutation were observed between animals (Fig. 4.2a), nor different predicted serospecificities (Fig. 4.2b), corroborating the hypothesis of antigen-independent immunoglobulin sequence diversification (Section 1.2.4).

Nine bovine IgHV genes were represented in the sequencing analysis, with IgHV1-10 being the most well represented, and IgHV1-7 encoding 22 antibodies with an ultralong CDR3H (Fig. 4.4) (15.4 % of all heavy chain sequences). The enrichment of ultralong sequences was potentially a consequence of AbCellera's sequencing approach, optimised for longer heavy chain sequences. Additionally, antibodies with a typical length CDR3H (for cattle) constituted 115 of the 143 retrieved sequences, and six unique antibodies with CDR3H loops shorter than 10 residues were identified (Fig. 4.3). The distribution of CDR3H lengths was not significantly different between the sampling points (Fig. 4.3), and is broadly in line with whole repertoire sequencing results. That said, the concordance of these distributions may also be artefactual due to the small sampling numbers.

In order to visualise the diversity of antibody sequences achieved, a pairwise distance matrix was calculated for the antibody heavy chain nucleotide sequences, and visualised as a 2D distance map. Seven IgHV1-7 encoded ultralong antibodies are represented surrounding a highly preserved light chain (Fig. 4.4), reflecting literature observations of sequence invariance for light chains paired with ultralong CDR3H-containing heavy chains [61]. Given this map (Fig. 4.4) was calculated primarily on heavy chain sequence, the diverse ultralong CDR3H's are mostly presented as singletons.

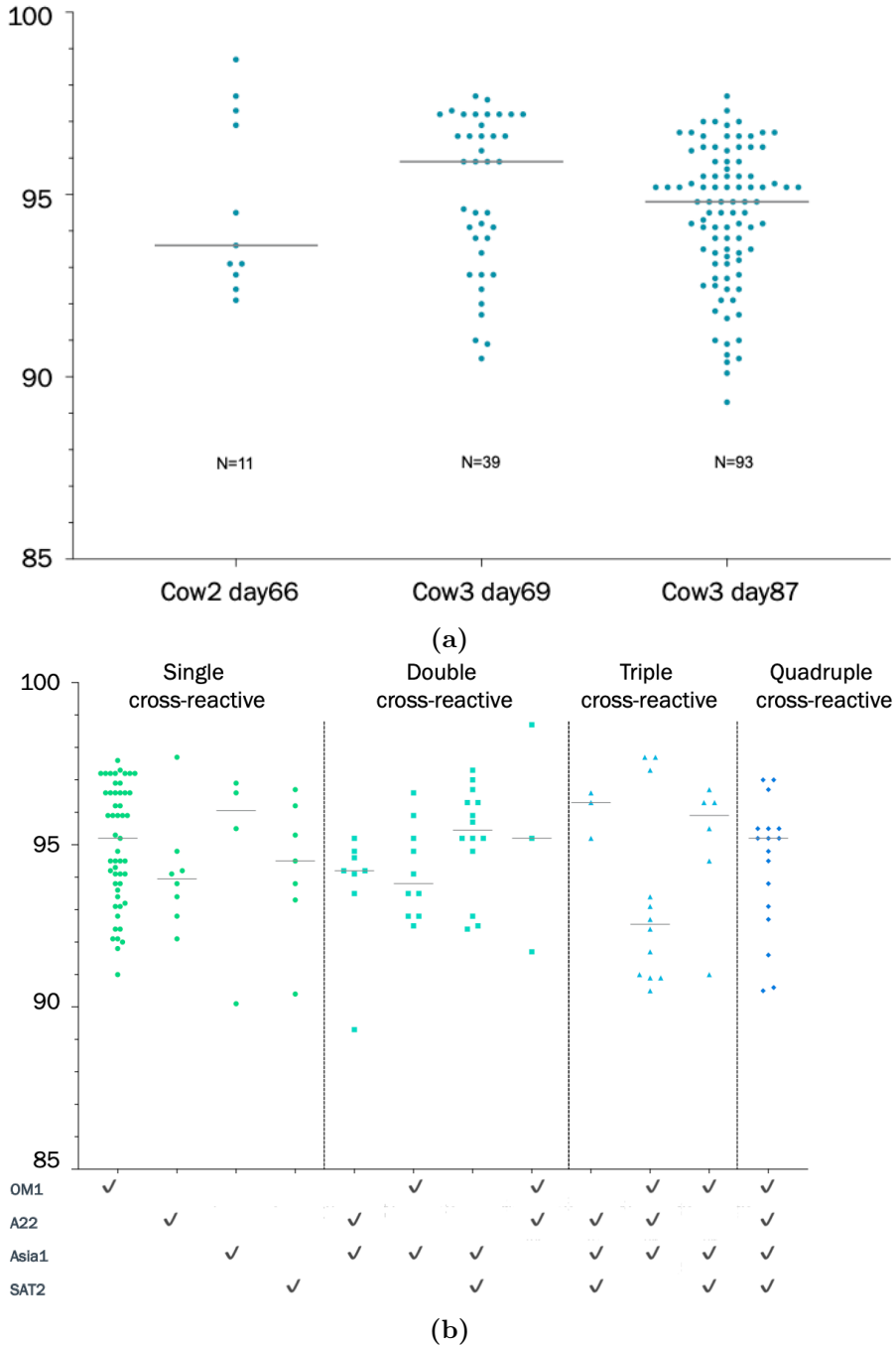


Figure 4.2: Percentage identity to germline for discovered antibodies by sample point (4.2a), and by serospecificity (4.2b). Figure produced by AbCellera.

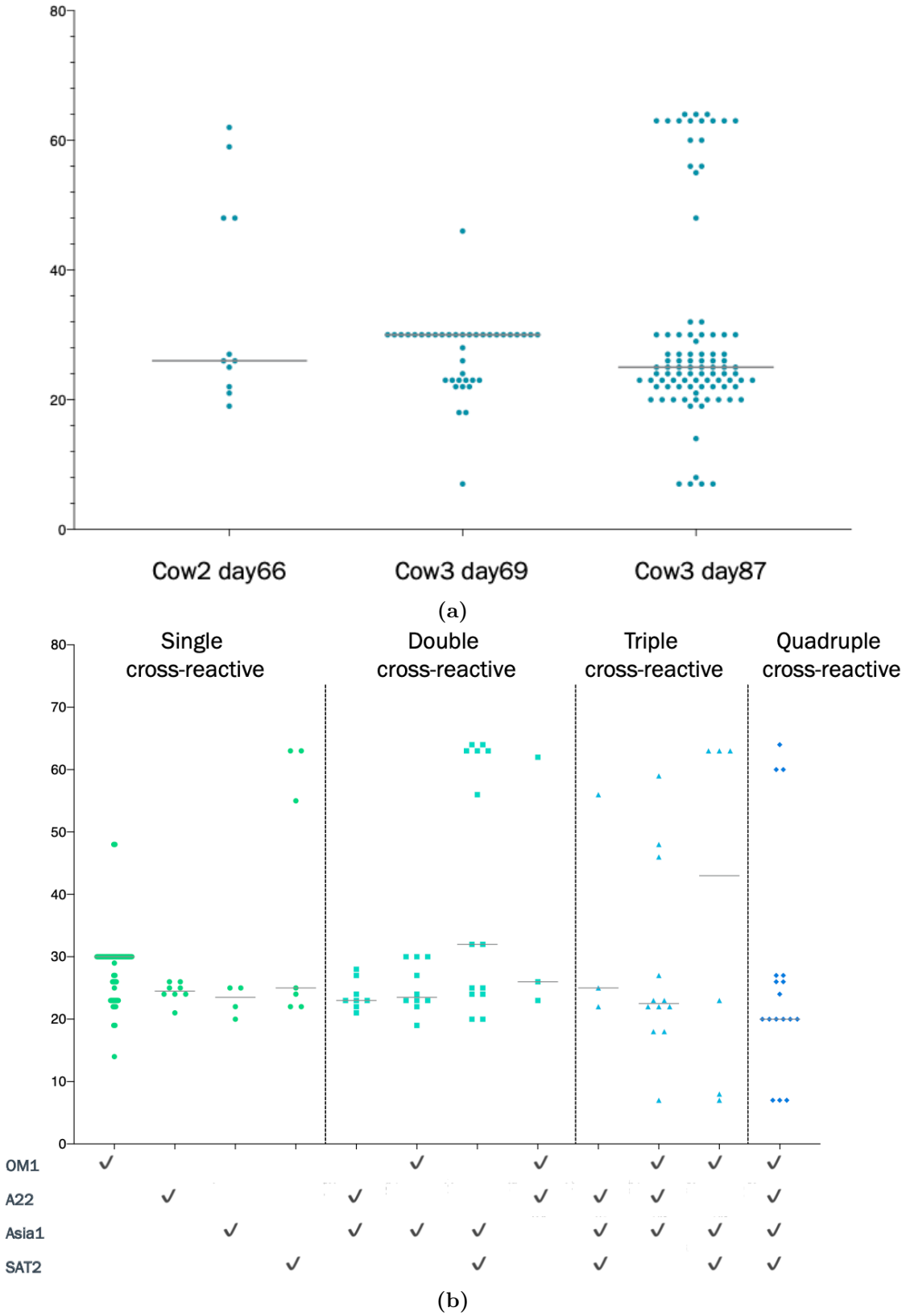


Figure 4.3: CDR3H length for discovered antibodies by sample point (4.3a), and by serospecificity (4.3b). Figure produced by AbCellera.

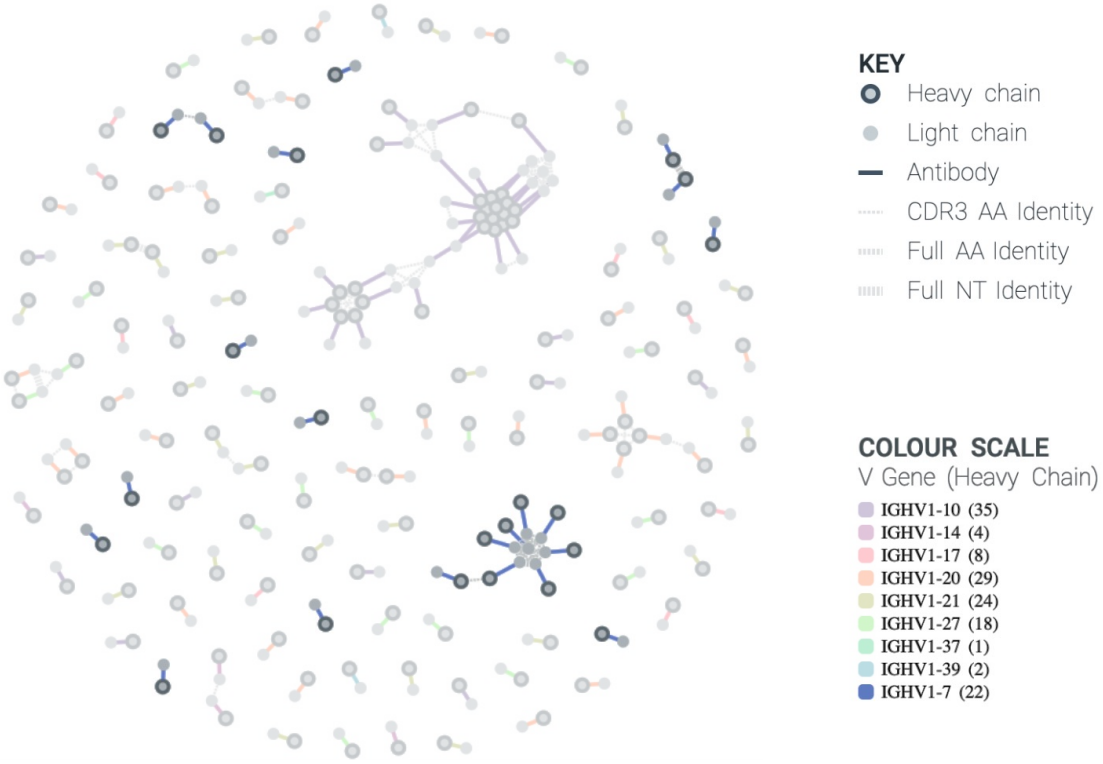


Figure 4.4: Sequence clustering of heavy and light chain sequences, coloured by VH encoding gene. Identity between sequences indicated with dashes. Antibody sequences clustered by amino acid similarity, followed by nucleotide similarity, clustering was performed principally on heavy chain sequences, and subsequently to the light chain sequences. Figure produced by AbCellera.

4.2.4 Bioinformatic sequence clustering

Viable sequences were annotated using IgMAT [306], and CDR3H sequences were clustered together using UCLUST v11.0.667 [307]. CDR3H sequences were clustered at $\geq$ 94 % amino acid identity, producing 10 clusters (Table 4.3). Clustering was performed for heavy chains only, due to light chain pairwise similarity (Section 1.3.2), with subsequent alignment of light chain sequences.

Table 4.3: Clustering of CDR3H sequences from the Pirbright-AbCellera collaboration at $\geq$ 94 % amino acid identity. All CDR3H clusters were identified from animal #3 at 69, and 87 dpi. Serospecificity assayed by AbCellera. CDR3H length in amino acids.

Cluster	Count	Predicted Serospecificity	CDR3H Length
1	26	O1M-Asia1; O1M	30
2	6	O1M-A22-Asia1-SAT2; Asia1	20
3	2	Asia1-SAT2	25
4	2	Asia1	25
5	2	O1M-A22-Asia1-SAT2; Asia1-SAT2	64
6	2	O1M-A22-Asia1-SAT2; O1M-A22-Asia1	27
7	2	Asia1-SAT2	20
8	2	O1M-Asia1-SAT2; SAT2	63
9	2	A22-Asia1	23
10	2	O1M-A22-Asia1-SAT2	7
Singletons	*95*	*Various*	*7 - 64*

Clusters 5, 6, 8, and 10 provide evidence of pairing promiscuity, where identical heavy chains are paired with different light chains (i.e. clusters 5 and 6). Both cluster 5 and 6 were predicted as demonstrating alternative serospecificities for each antibody within these clusters. Cluster 5 is especially interesting as these identical CDR3H's are ultralong (Fig. 4.5), and Fv region pairing promiscuity appears to effect a functional serospecificity shift by changes introduced at a region distant to the ultralong antigen-interacting CDR3H knob mini-domain (Section 1.2.5.4). Within cluster 5, antibody #117 was predicted as dual cross-reactive serospecificity for Asia1 and SAT2, whilst antibody #126 was predicted as quadruple cross-reactive against all tested serotypes (Tables 4.3 and 4.4). The serospecificity assay is a novel technique however, and the error rate is unknown. Given this, there is interest in verifying

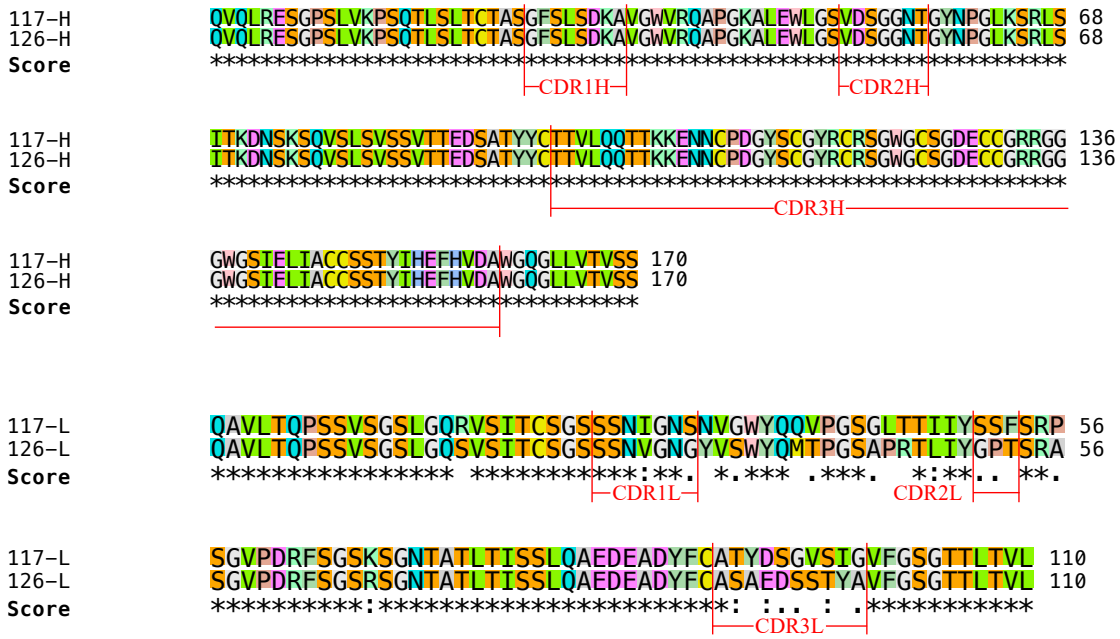


Figure 4.5: Multiple Sequence Alignment of antibody #117 and #126 variable domains. Sequences aligned using Clustalω in DNADynamo v1.569 (Blue Tractor Software).

whether indeed two identical heavy chains, paired to distinct light chains, can exhibit real differences in serospecificity.

As discussed previously (Section 1.2.5.4), the knob mini-domain principally interacts with antigen [78, 84, 85, 132], however, since the differences in the β-stalk geometry resulting from chain-exchange with ultralong antibodies (Section 3.6.4) are subtle, the functional consequence, if any, of light chain selection is uncertain. Consequently, comparing the structures of antibodies #117 and #126, particularly when complexed with antigen would inform on the light chain contribution to antigen recognition. The light chain of antibody #117 is encoded by IgLV1-67, and resolving the structure of antibody #117 would therefore be the first ultralong antibody structure with the IgLV1-67 encoded light chain. Early reported ultralong structures are described as pairing to an 'invariant' light chain [78, 82], whilst all current structures have light chains encoded by IgLV1-47 [78, 82–85]. In the CDR3H, antibody #126 has two amino acid substitutions compared to the invariant light chain, whereas the IgLV1-67 encoded CDR3L of antibody #117 has eight (Fig. 4.6).

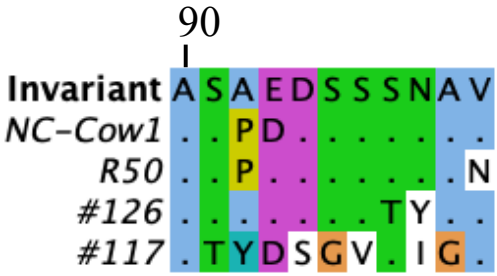


Figure 4.6: Multiple sequence alignment of CDR3L loops paired with ultralong heavy chains. 'Invariant' light chain utilised in ultralong antibodies BLV1H12 (PDB ID: 4K3D) [78], BLV5B8 (PDB ID: 4K3E) [78], A01 (PDB ID: 5ILT) [82], E03 (PDB ID: 5IJV) [82], F08 (PDB ID: 5E99) [82], B11 (PDB ID: 5IHU) [82], BOV-1 through BOV-7 (PDB IDs: 6E8V, 6E9G, 6E9H, 6E9I, 6E9K, 6E9Q, 6E9U) [83] and F145 (PDB ID: 7D3L) [85]. NC-Cow1 (PDB IDs: 6OO0, 6OPA, 6PW6) [84], R50 (PDB IDs: 7D3M, 7D3R) [85] and antibody #126 have light chains also encoded by IgLV1-47, but differ at two residues within the CDR3L. the antibody #117 is encoded by IgLV1-67. Conserved residues hidden for clarity.

The CDR3L loop, partially encoded by the IgLV gene (Section 1.3.1), likely makes intimate contact with the ultralong CDR3H stalk region due to its proximity in the reported ultralong antibody structures (Sections 1.4 and 3.6.4); an alternative CDR3L conformation from IgLV1-67 may therefore impact the ultralong CDR3H conformation of antibody #117 and #126 in a previously unobserved manner. Furthermore, Serine 91, within the CDR3L (Ser91L), has previously been described as an absolute requirement for ultralong-paired light chains, due to its conservation in ultralong-paired light chains, but substitution in light chains not paired with ultralong heavy chains [61]. The premise being, that Ser91L confers the ability for CDR3L to pack tightly between the VL domain N terminus and the CDR1L. Encoded by IgLV1-67, the #117 light chain is the first ultralong-paired light chain to deviate from Ser91L, instead encoding a threonine in this position. The additional volume of the methyl group may disrupt this tight packing, which could therefore influence the ultralong CDR3H architecture. For these reasons, there is an evident need to resolve the structure of ultralong antibody #117.

Therefore the two antibodies within cluster 5 were included in this thesis to investigate the functional implications of alternative antibody pairings (Tables 4.3 and 4.4). Antibody #30 (a CDR3H sequence singleton) was included as candidate

mono-reactive ultralong antibody against FMDV serotype O1M VLPs. Also investigated, but not included in this thesis for brevity, were; seven additional ultralong antibodies sourced from the serial vaccination study (antibodies #63 and #86 from cluster 8, and singletons #58, #75, #107, #127, #134), as well as FMDV-specific ultralong antibodies F145, F166, and B74 reported by others [308], and non-ultralong, O1M mono-reactive candidate antibody #14 from cluster 1.

Table 4.4: Serospecificity of clustered ultralong antibodies of interest in this thesis. Antibody #30 is a CDR3H sequence singleton, bearing < 94 % identity to any other CDR3H detected. Specificity predicted by AbCellera Biologics Ltd.

Cluster	ID	O1M	A22	Asia1	SAT2	Identities
5	#117			Y	Y	#126-L
5	#126	Y	Y	Y	Y	#117-L
-	#30	Y				

4.3 Transient transfection of anti-FMDV Fabs

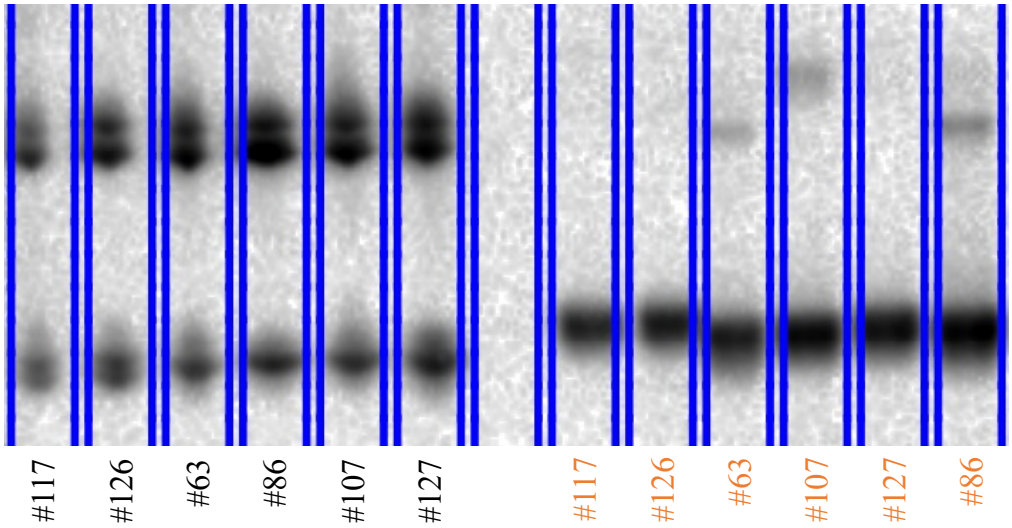
These examples of naturally-occurring antibodies showing pairing promiscuity gave some confidence for their efficient expression *in vitro*.

The observed antibody nucleotide sequences are provided in Table A.1. Candidate genes were codon optimised Table A.1, and inserted into the pOPINBOVH and pOPINBOVL expression vectors [160] (Section 2.2.1), for expression as recombinant Fabs. Successful cloning was verified by PCR screening (Section 2.4.2), and sequencing (Section 2.4.5). Correct constructs were prepared by Midiprep (Section 2.2.2).

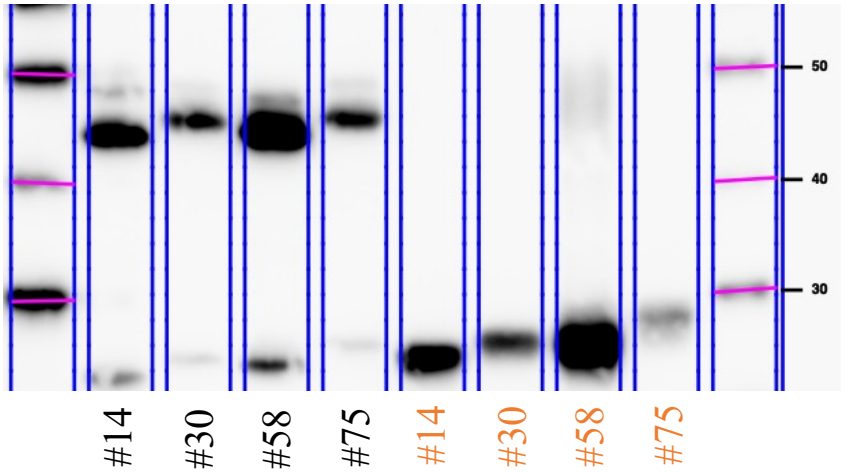
4.3.1 FMDV-specific Fab expression screening

The expression of Fabs by transient transfection was conducted in 1 mL Expi29F™ cultures, co-transfected with 0.5 µg/mL of each heavy/light chain encoding plasmid (Fig. 4.7). Recombinant Fabs were detected in crude media samples by western blot, targeting the polyhistidine tail at the C terminus of the recombinant heavy chain (Fig. 4.7).

The expression of ultralong recombinant Fabs #117 and #126 appeared fairly uniform, though subtle differences between expression efficiencies may of course exist. However, given the identical heavy chains, and high similarity of light chains, similar expression efficiency may be expected. Fab #30 expressed with less efficiency than Fabs #117, #126, and most other screened ultralong Fabs (Fig. 4.7).



(a)



(b)

Figure 4.7: Anti-polyhistidine immunoblot to show relative recombinant expression efficiency of antibodies identified from the Pirbright-AbCellera collaboration, expressed as Fabs. Orange text indicates samples run in reducing (10% βME) SDS loading buffer. All samples were heated to 95 °C for 10 minutes prior to loading. Data collected using (4.7a) a ChemiDoc XRS+, and (4.7b) a ChemiDoc MP system (Bio-Rad).

4.3.2 Purification

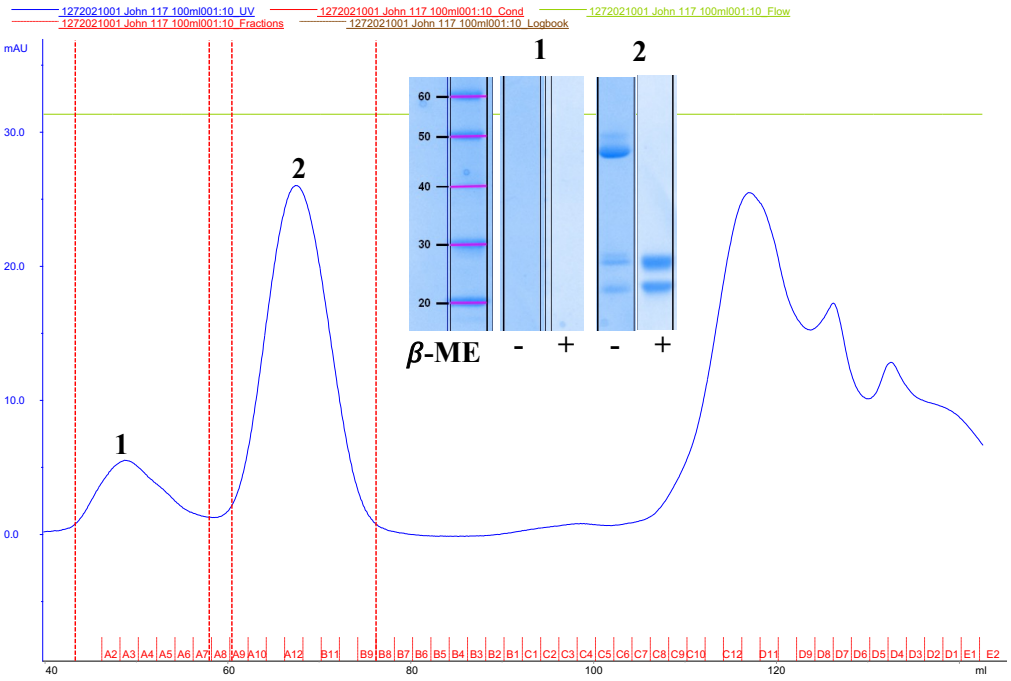
Initial expression screening indicated Fab #117 and Fab #126 are promising candidates for efficient expression without further optimisation (Fig. 4.7). Fab #30 expressed with less efficiency, and larger expression volumes were needed for purification of this Fab.

Recombinant Fabs were produced by co-transfection of VH and VL vectors into Expi293F™ cells (Section 2.2.4). Initially, 100 mL of Expi293™ cells, seeded at 1×10^6 cells/mL, were used to synthesise Fabs #117 and #126, to save resources following high expression yields of recombinant ultralong Fabs D08, D08*, and 4K3E* (Table 3.4). These preparations were supplemented with antibody from subsequent transient transfections using 215 mL, and 300 mL cultures respectively. Recombinant Fab #30 was expressed using a culture volume of 215 mL Expi293F™ cells.

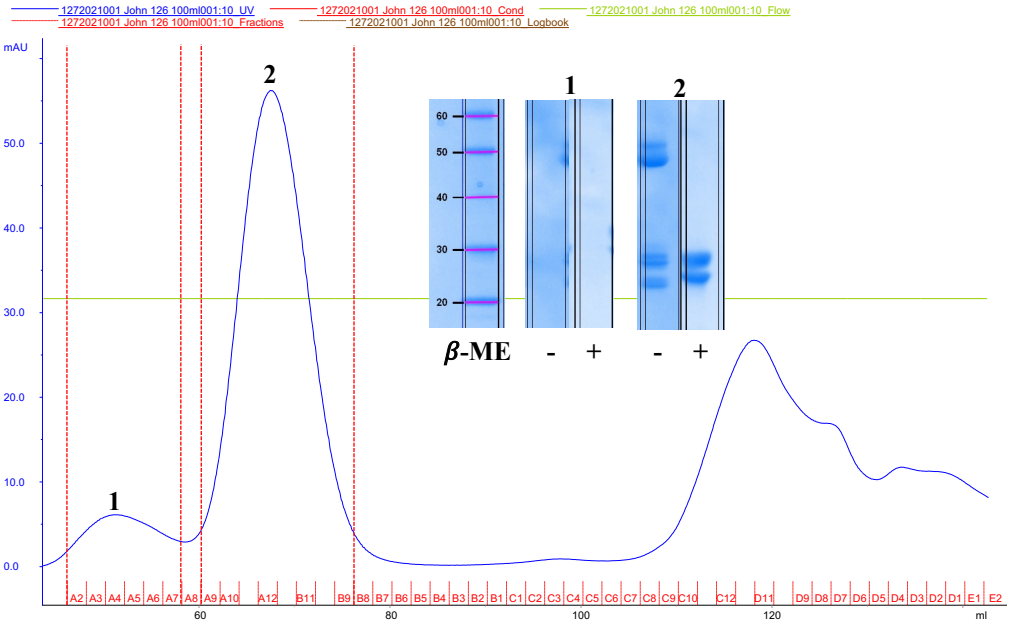
Recombinant Fabs were then retrieved from the media using IMAC and further purified using gel filtration (Section 2.6.1). Gel filtration elution peak UV absorbance traces are provided in (Fig. 4.8). Eluted fractions were characterised by SDS-PAGE, and those fractions of interest were spin concentrated using Amicon® Ultracel-30 centrifugal filters (Merck) (Table 4.5). As in Section 3.4.3, concentrated Fabs were stored at 4 °C, and proved to be stable for at least 9 months without demonstrating precipitation or significant degradation (not presented in this thesis).

Table 4.5: Recombinant Fab yields obtained by transient transfection of Expi293F™ cells. Heavy and light chain ID refers to assigned sequence ID within the OPPF Target database. †Time taken for spin-concentration to achieve stated concentration.

Fab ID:	#117	#126	#30
Heavy chain ID:	22404	22404	22447
Light chain ID:	22405	22406	22448
Pooled fractions:	A09:B09	A09:B09	A08:B08
Pooled volume (mL):	16	16	20
Yield (mg/L):	9.08	27.20	27.44
Concentration (mg/mL):	17.62	17.77	25.20
Time taken† (minutes):	40	40	40



(a)



(b)

Figure 4.8: Size exclusion chromatography of Fabs #117 and #126. Fabs #117 (4.8a) and #126 (4.8b). 280 nm Absorbance trace of gel filtrate. Inset: Indicated fractions were analysed by SDS-PAGE. Samples prepared without (left) and with (right) 10 % β-ME. Band migration compared with Benchmark™ Precision Plus (Invitrogen™, ThermoFisher Scientific), with indicated molecular weights in kDa.

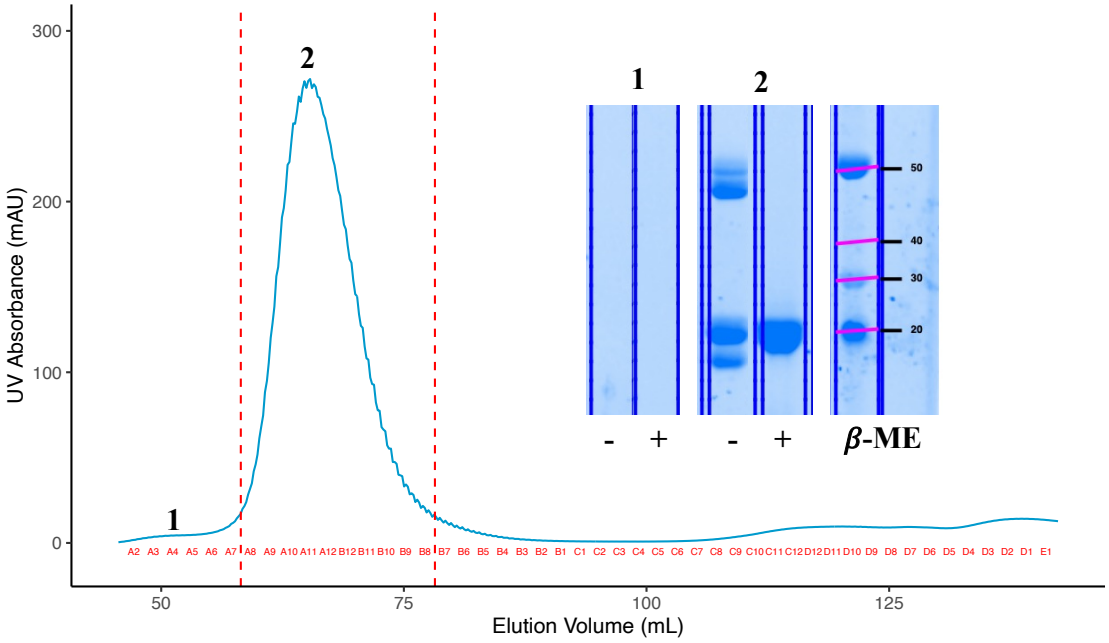


Figure 4.9: Size exclusion chromatography of Fab #30. 280 nm Absorbance trace of gel filtrate. Inset: Indicated fractions were analysed by SDS-PAGE. Samples prepared without (left) and with (right) 10 % β-ME. Band migration compared with Benchmark™ Precision Plus (Invitrogen™, ThermoFisher Scientific), with indicated molecular weights in kDa.

4.4 Fab apo-structures

Given the lack of published research, there is an inherent need to resolve the structure of the CDR3L conformation provided by IgLV1-67 encoded light chains paired with an ultralong heavy chain. Equally, experimental determination of these structures provides reliable models for interpreting cryo-EM datasets of these antibodies in complex with antigen. Given preliminary serospecificity data was recorded (Table 4.2), antibodies #117 and #126 offer an opportunity to learn about the functional impact of bovine light chain selection.

Crystallographic X-ray diffraction experiments were performed with the *apo*-Fabs (in absence of antigen), due to the small size of Fab molecules, the previous successes of crystallising ultralong bovine Fabs (Chapter 3), and the relative ease of Fab expression by transient transfection, providing sufficient material for crystallography (Section 4.3.2).

Cryogenic electron microscopy (cryo-EM) was used to investigate the structural nature of ultralong antibody interaction with antigen due to the size of the VLP/Fab complex, the reduced sample demand, and to avoid the challenges of crystallisation.

4.4.1 Crystallography

Prior to crystallisation, concentrated Fabs were screened using the PCT (Section 2.8.3.1). PCT results (omitted from this thesis) indicated purified Fabs had reached a concentration appropriate for trialling crystallisation experiments. Sitting droplet vapour diffusion crystallisation experiments (Section 2.8.3.3) were set up at room temperature for cluster 5 Fabs #117, #126, and O1M-mono-specific Fab #30 within Swissci UV-transmissible Polymer 96-Well Triple Drop Plates (Molecular Dimensions).

4.4.2 Crystallography of Fabs #117 and #126

Crystallisation of Fabs #117 and #126 were screened using JCSG-*plus*™, Morpheus®, and Index™ screens, however no crystal hits of sufficient quality for XRD data collection were produced. Crystal plates were prepared at both 20 °C and 4 °C using Fab #117 at concentrations of 17.62 mg/mL, and 21.68 mg/mL, and Fab #126 at concentrations of 17.77 mg/mL, and 30.54 mg/mL, and observed for crystal growth up to 40 days after plate preparation. Fab #117 produced a few wells with thin needle-like crystals, whilst crystal trials of Fab #126 mainly resulted in phase separation (Fig. 4.10). In either case, lack of significant UV absorbance at 280 nm indicated lack of 3D crystal formation or aggregation (not presented in this thesis). Screening with further commercial screens, or optimising conditions that produced 1D crystals could yield crystal hits of sufficient quality for data collection, as could crystal seeding. Alternative methods were explored to provide the structures of antibodies #117 and #126; *in silico* prediction using AlphaFold 2 [204] and cryo-EM (Section 4.6).

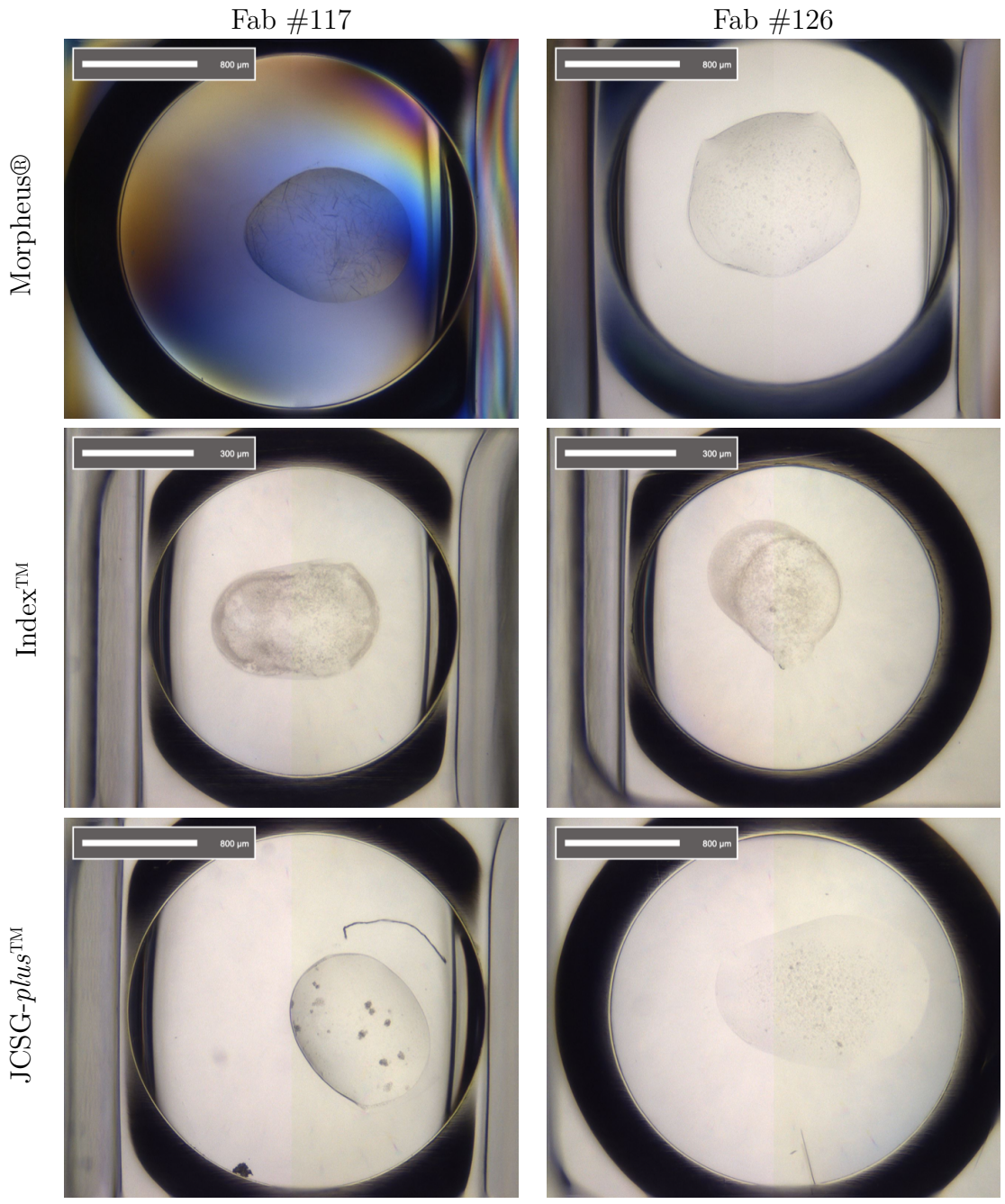


Figure 4.10: Crystallisation screening for Fabs #117 and #126 in various crystallisation screens. Fab #117: Morpheus® screen condition E01 droplet 1, Index™ screen condition F03 droplet 1, and JCSG-*plus*™ screen condition H03 droplet 1. Fab #126: Morpheus® screen condition A04 droplet 1, Index™ screen condition B09 droplet 2, and JCSG-*plus*™ screen condition A01 droplet 2. Images taken 40 days after crystallisation plate preparation using a RockImager 1000 (Formulatrix) on imager default settings and inspected using RockMaker® web interface v3.17.7.1 (Formulatrix). Crystallisation plates were prepared and incubated at 20 °C throughout.

4.4.3 AlphaFold Structural Prediction

Around the same time as the failed crystallisation screening of Fabs #117 and #126 (Fig. 4.10), the code behind CASP-18 winning structural prediction programme AlphaFold was made publicly available (Section 2.8.2)[204] as AlphaFold v2 (AF2). At the time, only single sequences could be submitted to AF2, but more recent updates now allow submission of distinct chains, which anecdotally predicts structures with slightly better accuracy when compared to experimental crystal structures, however at the time of writing, the BMRC instance of AF2 does not yet include these updates.

Structures predicted by AF2, and indeed by any predictive software, are of limited relevance in isolation. A wealth of biophysical and physicochemical information retrieved from the PDB [162] is considered by AF2, which is used for both structural prediction, and calculation of per-residue confidence metrics to aid interpretation of predicted AF2 models (Section 2.8.2).

The accuracy of AF2 was first assessed by submitting antigen-unknown ultralong Fab D08 (Section 3.3) as a single-chain variable fragment antibody (scFv), with heavy and light chains joined using a polyglycine linker, and compared against the crystal structure described previously (Sections 3.5.3 and 3.7). When compared to the experimentally determined crystal structure (Section 3.5.3), AF2 generally performed well, even producing a model with remarkable similarity to the observed D08 structure, though experimental determination was still essential to identify which of the AF2 models were closest to the true structure.

Heavy and light chain variable domains from antibody #117 and #126 were joined *in silico* using a polyglycine linker to form scFv #117 and scFv #126 (Table 4.6).

Table 4.6: Single chain variable fragment (scFv) sequences for antibodies #117 and #126. Polyglycine linker is emphasised.

scFv #117	1	QVQLRESGPS	LVKPSQTLSL	TCTASGFSLS	DKAVGWVRQA	PGKALEWLGS	50
	51	VDSGGNTGYN	PGLKSRLSIT	KDNSKSQVSL	SVSSVTTEDS	ATYYCTTVLQ	100
	101	QTTKKENNCP	DGYSCGYRCR	SGWGCSGDEC	CGRRGGGWGS	IELIACCSST	150
	151	YIHEFHVDAW	GQGLLVTVSS	**GGGGGGGGG**Q	AVLTQPSSVS	GSLGQRVSIT	200
	201	CSGSSSNIGN	SNVGWYQQVP	GSGLTTIIYS	SFSRPSGVPD	RFSGSKSGNT	250
	251	ATLTISSLQA	EDEADYFCAT	YDSGVSIGVF	GSGTTLTVL		289
scFv #126	1	QVQLRESGPS	LVKPSQTLSL	TCTASGFSLS	DKAVGWVRQA	PGKALEWLGS	50
	51	VDSGGNTGYN	PGLKSRLSIT	KDNSKSQVSL	SVSSVTTEDS	ATYYCTTVLQ	100
	101	QTTKKENNCP	DGYSCGYRCR	SGWGCSGDEC	CGRRGGGWGS	IELIACCSST	150
	151	YIHEFHVDAW	GQGLLVTVSS	**GGGGGGGGG**Q	AVLTQPSSVS	GSLGQSVSIT	200
	201	CSGSSSNVGN	GYVSWYQMTP	GSAPRTLIYG	PTSRASGVPD	RFSGSRSGNT	250
	251	ATLTISSLQA	EDEADYFCAS	AEDSSTYAVF	GSGTTLTVL		289

4.4.3.1 AlphaFold Predicted scFv #117 Structure

The AlphaFold neural network relies on published structural information available within the PDB [162]. It is no surprise then, that the antibody variable domains are well represented during the homology search step (4.11). As expected, the ultralong CDR3H loop is much less well covered.

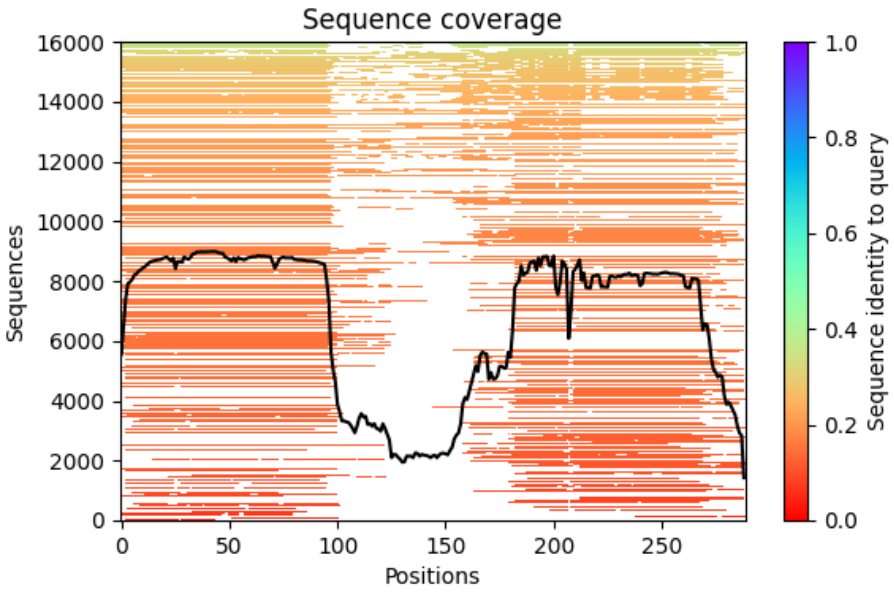
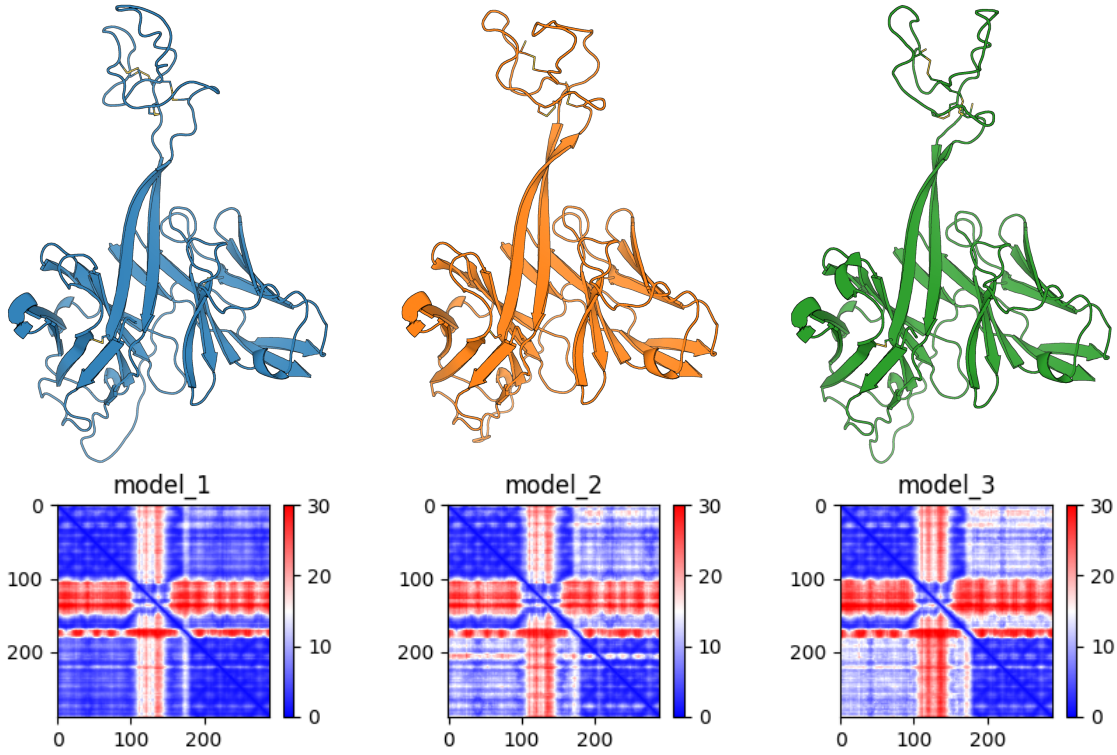


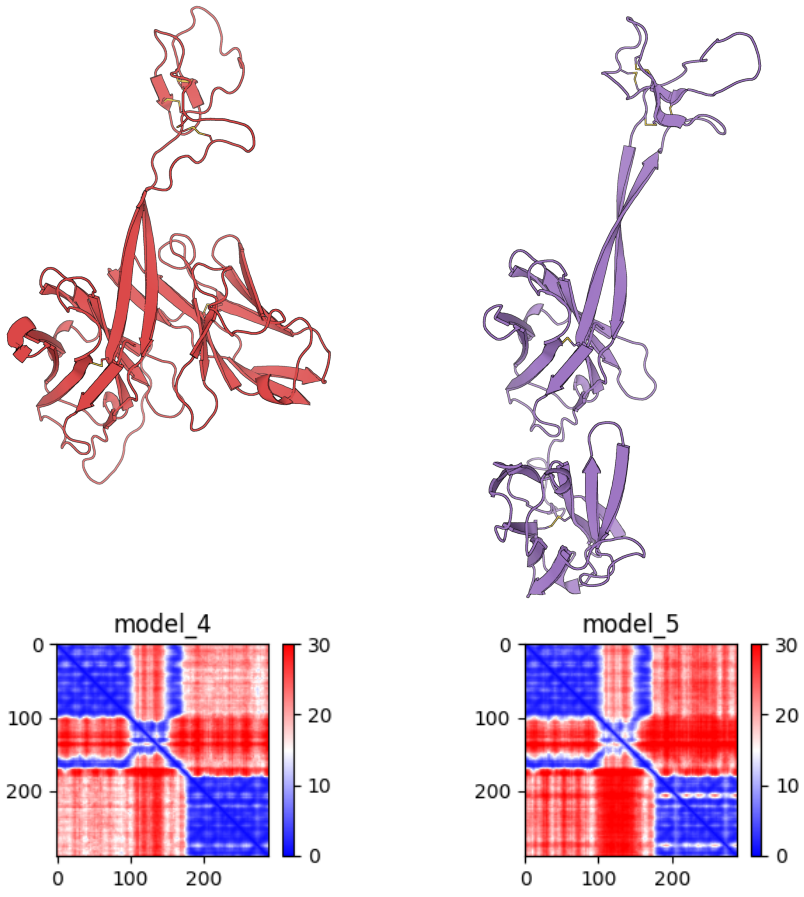
Figure 4.11: PDB Sequence coverage depth for scFv #117.

As introduced in Section 2.8.2, the predicted aligned error (PAE) reflects the confidence of a given modelled residue in the AF2 output, and is therefore informative for identification and assessment of the relative modelled positions of predicted

protein subunits. In all AF2 models of scFv #117, the PAE shows a greater predicted positional uncertainty for the ultralong CDR3H loop (residues 96 - 159), and the polyglycine linker (residues 171 - 179), than the two variable domains flanking these sequences (Fig. 4.12). In models #1 (Fig. 4.12a), #2 (Fig. 4.12b), and #3 (Fig. 4.12c), the positioning of the variable domains, relative to one another, is generally predicted with the highest confidence, likely consequent on the abundance of reported antibody structures in the PDB [162]. At the time of writing, 10,528 antibody, Fab or scFv structures were reported in the PDB. Model #4 (Fig. 4.12d) was predicted with a small displacement of the VL domain, explaining the score penalty to the PAE. The prediction of model #5 was non-sensical with a misplaced VL domain, which is reflected in the greater PAE over the modelled VL domain (Fig. 4.12e). In all models, the positions of the ultralong CDR3H residues, relative to one another, is broadly as confident as pairs of residues within the same variable domain, indicating confidence in the CDR3H structure, even if its relative position to the variable domains is predicted to show greater error. Conversely, there was a drop in the predicted Local Distance Difference Test (pLDDT) scores corresponding to the ultralong CDR3H loop, however, this may be due to there being fewer residues contributing to the local chemistry for pLDDT calculation, due to the the greater solvent exposure of residues of the ultralong CDR3H loop than compared to those comprising the immunoglobulin domains (Fig. 4.13).



(a) Model #1 **(b)** Model #2 **(c)** Model #3



(d) Model #4 **(e)** Model #5

Figure 4.12: Predicted models and alignment error for AlphaFold [204] **scFv #117 models.** Models are structurally aligned on scFv #117 VH domain.

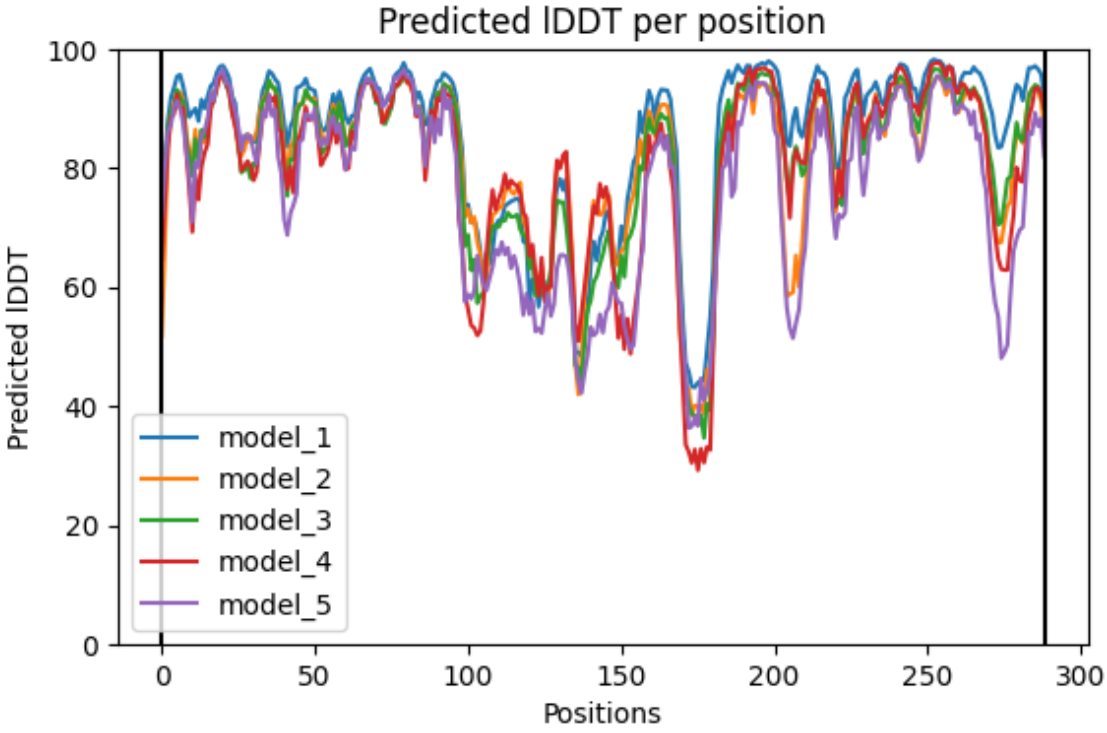


Figure 4.13: Predicted Local Distance Difference Test per-residue scores for each AlphaFold [204] **model of scFv #117.**

In all models except the C-terminal half of the model #5 knob, the knob mini-domains were predicted with similar geometry (Table 4.7).

Table 4.7: Structural superposition RMSD of AF2 scFv #117 knob mini-domains. Modelled CDR3H knob mini-domain (residues 104 - 150; 496 atoms) were aligned using PyMol.

Model	#1	#2	#3	#4	#5
#1	-				
#2	4.195	-			
#3	5.030	2.207	-		
#4	0.980	4.499	4.834	-	
#5	5.180	5.080	5.700	5.857	-

4.4.3.2 AlphaFold Predicted scFv #126 Structure

Overall, AF2 performed very similarly for scFv #126 as it did scFv #117. The database sequence coverage for scFv #126 is, unsurprisingly, very similar to that of scFv #117, with strong coverage for each of the variable domains, and weaker

coverage for the ultralong CDR3H (Fig. 4.14. The principle difference between these search coverages is a slightly improved representation for the IgLV1-47 encoded light chain of Ab #126, given the lack of representation of IgLV1-67 encoded light chain (of Ab #117) within the PDB [162].

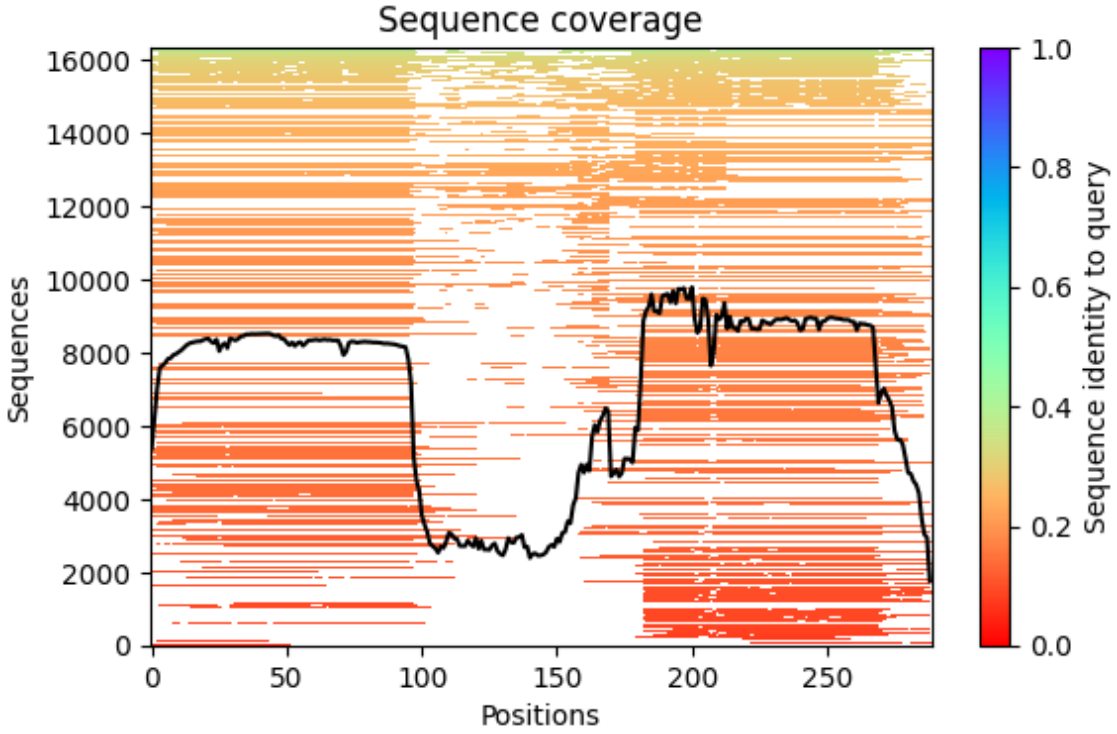
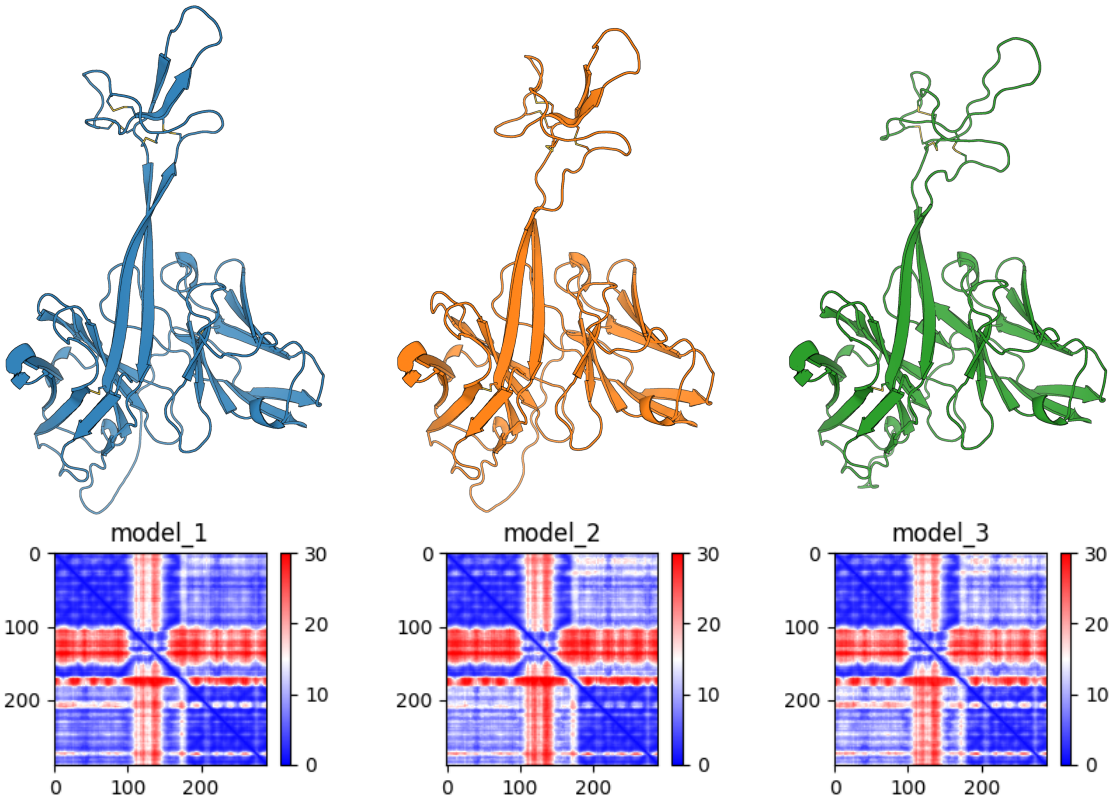
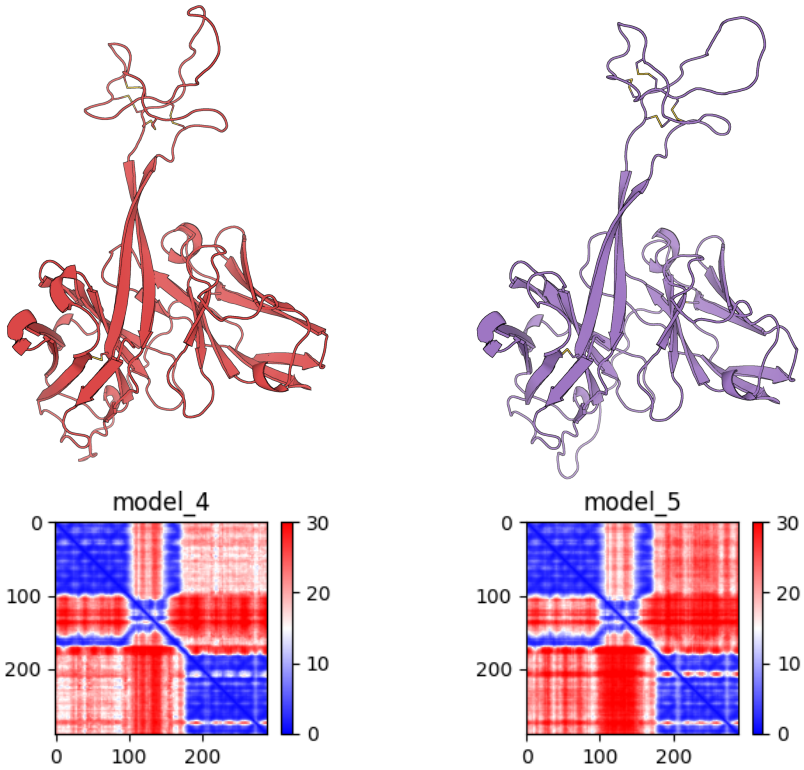


Figure 4.14: PDB Sequence coverage depth for scFv #126.

Similarly, PAE calculation show a similar situation for the scFv #126 models; relative positions of the variable domains to one another are fairly well predicted for models #1 (Fig. 4.15a), #2 (Fig. 4.15b), and #3 (Fig. 4.15c), whilst models #4 and #5 had a higher PAE, owing to a small displacement of the VL domain (Figs. 4.15d and 4.15e). The ultralong CDR3H residue positions were again predicted with good accuracy relative to one another, but with greater error relative to the variable domains (Fig. 4.15 and Table 4.8).



(a) Model #1 **(b)** Model #2 **(c)** Model #3



(d) Model #4 **(e)** Model #5

Figure 4.15: Predicted models and alignment error for AlphaFold [204] **scFv #126 models.** Models are structurally aligned on scFv #126 VH domain.

Table 4.8: Structural superposition RMSD of AF2 scFv #126 knob mini-domains. Modelled CDR3H knob mini-domain (residues 104 - 150; 496 atoms) were aligned using PyMol.

Model	#1	#2	#3	#4	#5
#1	-				
#2	0.699	-			
#3	0.942	0.953	-		
#4	2.658	1.085	1.839	-	
#5	1.260	1.291	1.210	1.318	-

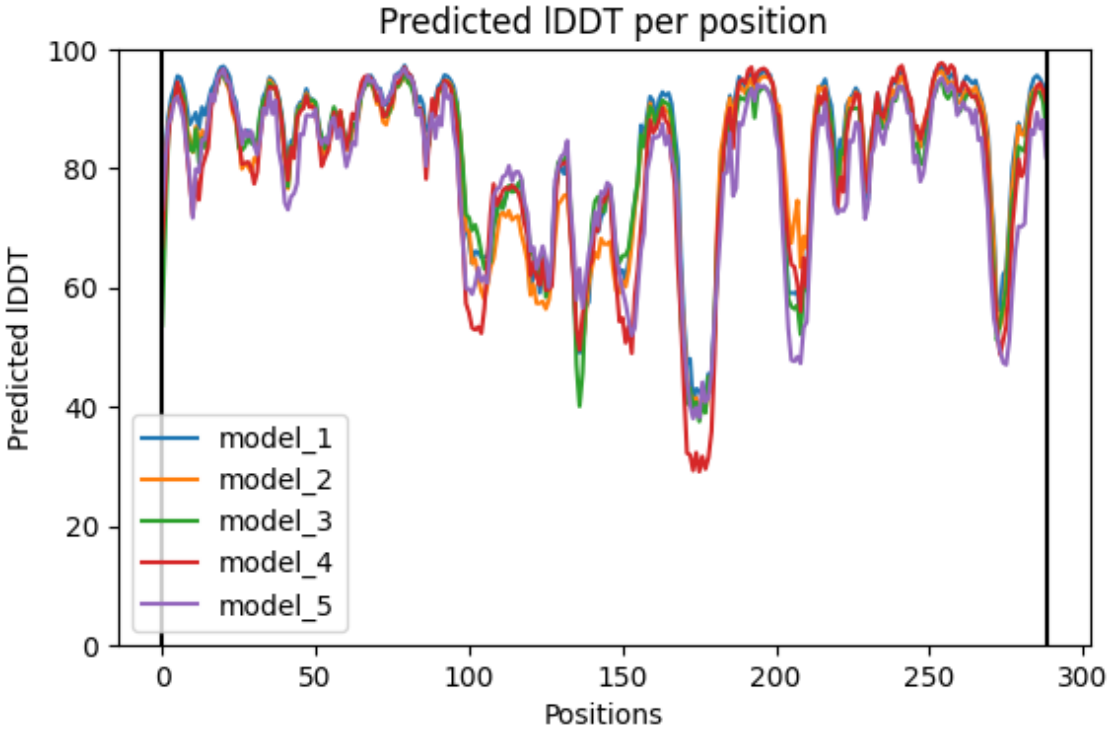


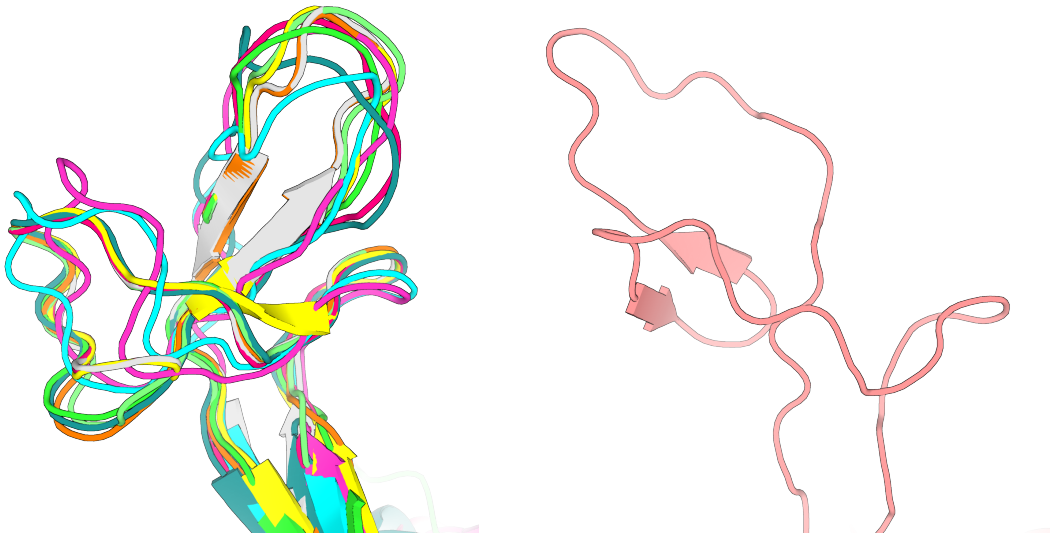
Figure 4.16: Predicted Local Distance Difference Test per-residue scores for each AlphaFold [204] **model of scFv #126.**

The pLDDT scores for scFv #126 models (Fig. 4.16) were also very similar to that of scFv #117 models (Fig. 4.13), indicating high (>90) local (chemical) confidence in the prediction of variable domain residues, and low-to-good (50-90) confidence in the local chemistry of the ultralong loop. The reduced pLDDT confidence of the ultralong loop predictions warrants caution in their interpretation [309]. Overall, the lowest scoring region on both scFv #117 and #126 was the polyglycine linker between the C' terminus of the VH domain and the N' terminus of the VL domain,

though the CDR1L and CDR3L loops of scFv #126 also scored relatively poorly.

4.4.3.3 Comparing AlphaFold #117 and #126 models

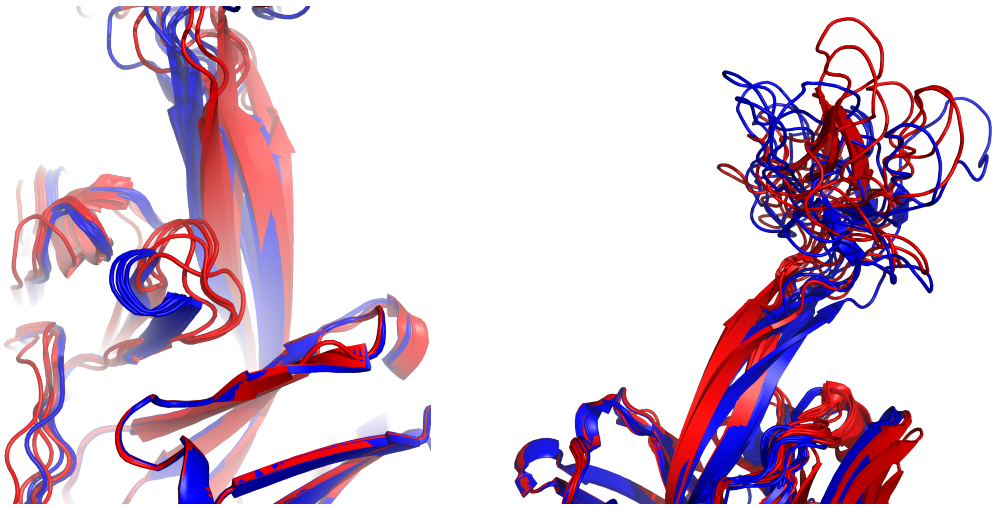
Despite AF2 performing well when predicting the ultralong CDR3H loop of Fab D08 (Section 3.7), the pLDDT scores for all AF2 predicted ultralong CDR3H models are too low for reliable interpretation of the CDR3H knob mini-domain structure (Figs. 4.13 and 4.16), especially in absence of experimental verification. Indeed, the most accurate prediction for antibody D08 (with an overall alignment RMSD of 1.465 Å), was not the model with the highest overall predicted confidence, indicated by the model ranking. Interestingly, whilst low confidence scores were ascribed to residues of the ultralong CDR3H, models converged on a common architecture (except scFv #117 model #5), with correct prediction of the knob-initiating CPDG motif, the three-pass β-strand at the core of the knob mini-domain and the same threading of loops between the central β-strands (Fig. 4.17).



(a) **(b)** scFv #117 model #5

Figure 4.17: Structural superposition of AF2 predicted ultralong antibody knob mini-domains of scFv #117 and #126. Modelled CDR3H knob mini-domain (residues 104 - 150; 496 atoms) were aligned using PyMol. (4.17b) Outlier scFv #117 model 5 was predicted with an alternative knob conformation, and the VL was entirely misplaced in this model (4.12e).

The prediction of five models for each scFv offers the opportunity to look at any predicted systemic differences. When all AF2 scFv are structurally aligned over their whole sequences, two such systemic differences are observed; first that the scFv #117 (in blue) CDR3L loop has a different conformation than that of the scFv #126 (in red) CDR3L loop (Fig. 4.18a), and second that the ultralong CDR3H stalk axial twist varies between scFv #117 (blue) and scFv #126 (red) models (Fig. 4.18b).



(a) **(b)**

Figure 4.18: Superposition of scFv #117 (in blue ribbon) and scFv #126 (in red ribbon) AF2 models. scFv #117 and scFv #126 models were structurally aligned by VH domain. Perspective focused on (4.18a) the CDR3L, and (4.18b) the CDR3H loops.

The pLDDT scores for all ultralong CDR3H stalk domains decrease with distance from the scFv variable domains (Figs. 4.13 and 4.16), however it is interesting to note the systemic twist of the identical ultralong CDR3H in scFv #117 relative to that of scFv #126, possibly consequent on the VL sequence.

As described previously, the CDR3L is partially encoded by the VL gene. The light chain of antibody #126 is encoded by the IgL1-47 gene, and the AF2 predicts a conformational conservation of CDR3L, reflecting other IgLV1-47 structures in the PDB [78, 82, 83] (Fig. 1.8). Interestingly, the #117 light chain is encoded by IgLV1-67. In absence of a CDR template in the PDB, AF2 modelled the #117 CDR3L as even more distant to the ultralong CDR3H stalk region than the CDR3L of #126. In antigen-unknown ultralong antibody D08, pairing with the 4K3E light

chain opened up a pocket between the CDR3L and CDR3H, allowing access to a CDR3H descending strand tyrosine, disrupting the YxYxY motif, and introducing an axial twist. If the IgLV1-67 is truly even further abstracted from the CDR3H, it is reasonable to suggest a descending strand tyrosine may also insert itself between the CDR3L and CDR3H, once again introducing a twist to the ultralong CDR3H.

A curious observation is that the confidence score for the CDR3L (indicated by the pLDDT) was actually greater for the IgLV1-67 encoded CDR3L of antibody #117 than that of the IgLV1-47 encoded CDR3L of antibody #126. The reason behind this disparity, given the prior reporting of IgLV1-47 encoded CDR3L structures [78, 82, 83] is unclear but may be resultant of the predicted closer packing, and hence slightly closer chemistry proximity contributing to pLDDT calculation.

4.4.4 Crystallography of ultralong Fab #30

Fab #30 produced several sub-optimal crystal hits in both the JCSG-*plus*™ and Morpheus® screens; producing mostly large twinned crystals, and a couple of large 3D crystals (Fig. 4.19). The JCSG-*plus*™ screen produced more crystal hits for Fab #30, but the proportion of untwinned crystals was greater when using the Morpheus® screen. Nucleation of Fab #30 crystals occurred rapidly; most crystals first appeared within 24 hours, and most had stopped growing by the + 13 day images. The Morpheus® screen had a few slower growing crystals, such as the large 3D observed in condition H05 droplet 2 (Fig. 4.19b), which only appeared in the + 21 day images, and had stopped growing by the + 29 day images. Crystals were initially screened by their morphology, and absorbance of 280 nm light to confirm proteinaceous composition (Fig. 4.19f). In the interest of time, crystallisation of Fab #30 was not optimised further. Such crystals were cryoprotected with 10 mM glycerol (Section 2.8.3.3) and cryocooled in liquid nitrogen and XRD data were collected at 100 K using a Dectris Eiger2 XE 16M photon-counting X-Ray detector on beamline I03 at DLS (Section 2.8.3.4).

4.4.5 Crystallographic structure of Fab #30

JCSG-*plus*™ condition C06 (0.1 M Potassium Phosphate Citrate buffer at pH 4.20 with precipitant 40 % (v/v) PEG 300) droplet 2 (2:1 ratio of sample:screen solution)(Figs. 4.19e and 4.19f) produced the highest-resolution diffracting crystals for Fab #30, diffracting to 2.65 Å at beamline I03 (Fig. 4.20) when processed by the xia2 dials pipeline (Section 2.8.3.6). This single crystal dataset was characterised as belonging to the P 6_1 2 2 spacegroup, with unit cell dimensions of $a = 69.51$ Å, $b = 69.51$ Å, $c = 442.68$ Å, and angles of $\alpha = 90.00°$, $\beta = 90.00°$, $\gamma = 120.00°$ by the xia2 dials pipeline (Section 2.8.3.6). Data were collected at each 0.1° oscillation angle over 360° total oscillation, where each image was exposed for 0.004 s with an exposure flux of 6.25×10^{11} photons (Table 2.10). The asymmetric unit was predicted to contain one Fab #30 monomer, (subsequently validated by structural analysis). The structure was phased by molecular replacement using Phaser [223](Section 2.8.3.7) using 1 copy of each of the protein chain H monomer of Fab D08 (Section 3.5.3) and the protein chain L monomer of bovine antibody B11 (PDB ID: 5IHU [82]) as search models, and refined using REFMAC5 [227] to 2.65 Å (Figs. 3.9c and 3.9d and Table 4.9) within the CCP4 suite [221].

Refinement is underway with R_{work} of 0.261 and R_{free} of 0.311 for all data. Substantial changes are still required to the knob mini-domain when compared to AF2 predicted models used in fitting to the cryo-EM density map of Fab #30 complexed with O1M VLPs (Section 4.6.8).

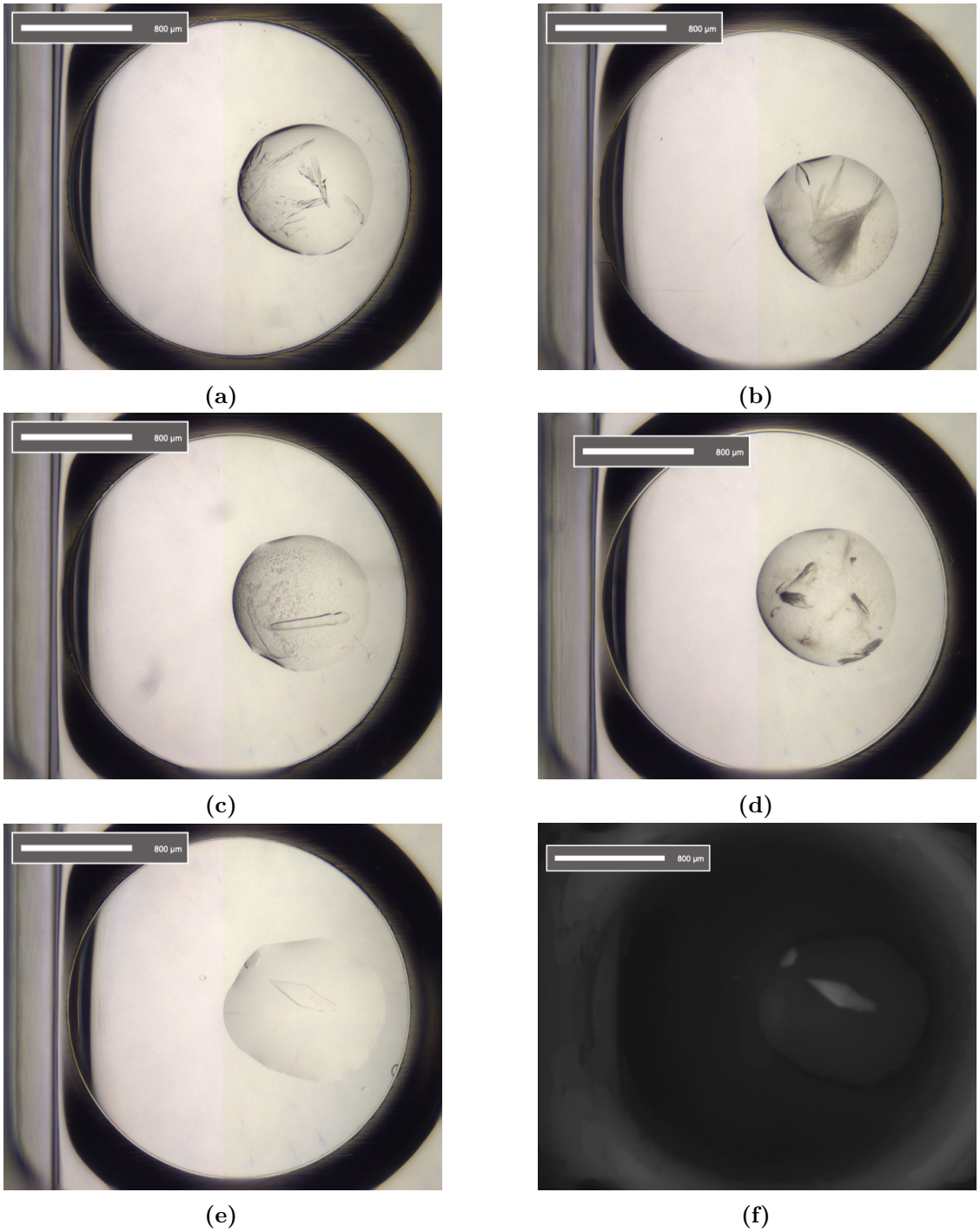
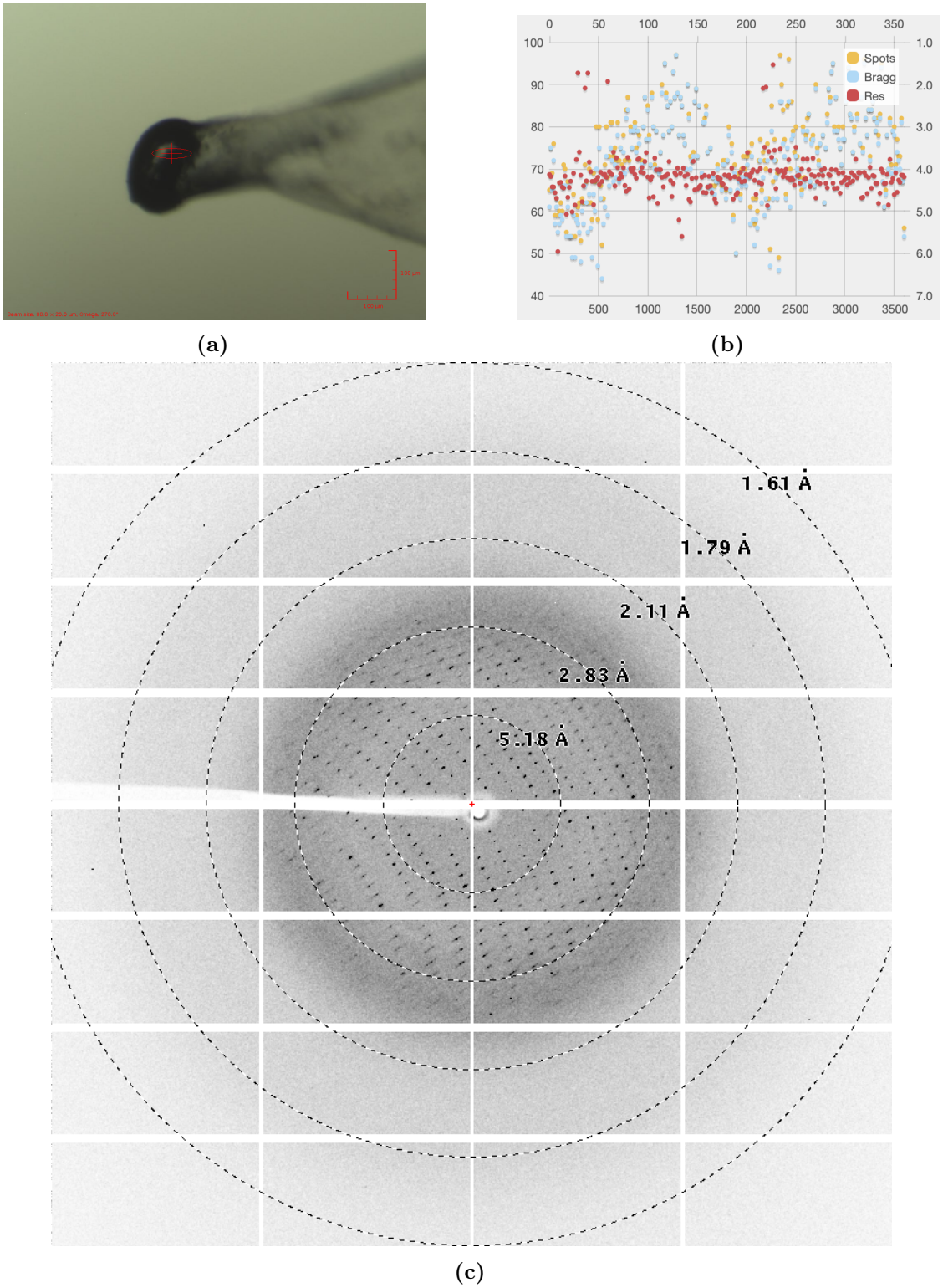


Figure 4.19: Exemplar crystal morphology for Fab #30 in various crystallisation screens. Morpheus® screen condition (4.19a) D09 droplet 2, (4.19b) H05 droplet 2, and (4.19c) F05 droplet 2. JCSG-*plus*™ screen conditions (4.19d) G06 droplet 2, and (4.19e) C06 droplet 2. UV absorbance image for (4.19f) JCSG-*plus*™ condition C06 droplet 2. Fig. 4.19b image taken 32 days after crystallisation plate preparation, the remainder of the images were taken 18 days after preparing the crystallisation plate using a RockImager 1000 (Formulatrix) on imager default or UV Mid settings and inspected using RockMaker® web interface v3.17.7.1 (Formulatrix). Crystallisation plate was prepared and incubated at 20 °C throughout.



(a) **(b)**

(c)

Figure 4.20: Single crystal X-ray diffraction data collection for ultralong Fab #30 at beamline I03 at Diamond Light Source. Fab #30 crystallised in JCSG-*plus*™ condition C06 droplet 2. (4.20a) I03 beamline sample camera, showing position and size of incident X-Ray beam. (4.20b) DISTL [212] per-image analysis showing number of diffraction spots (total and Bragg)(left axis) and estimated resolution (right axis) for each image in the data collection. (4.20c) Example diffraction image composed of 15 summed frames for ultralong Fab #30.

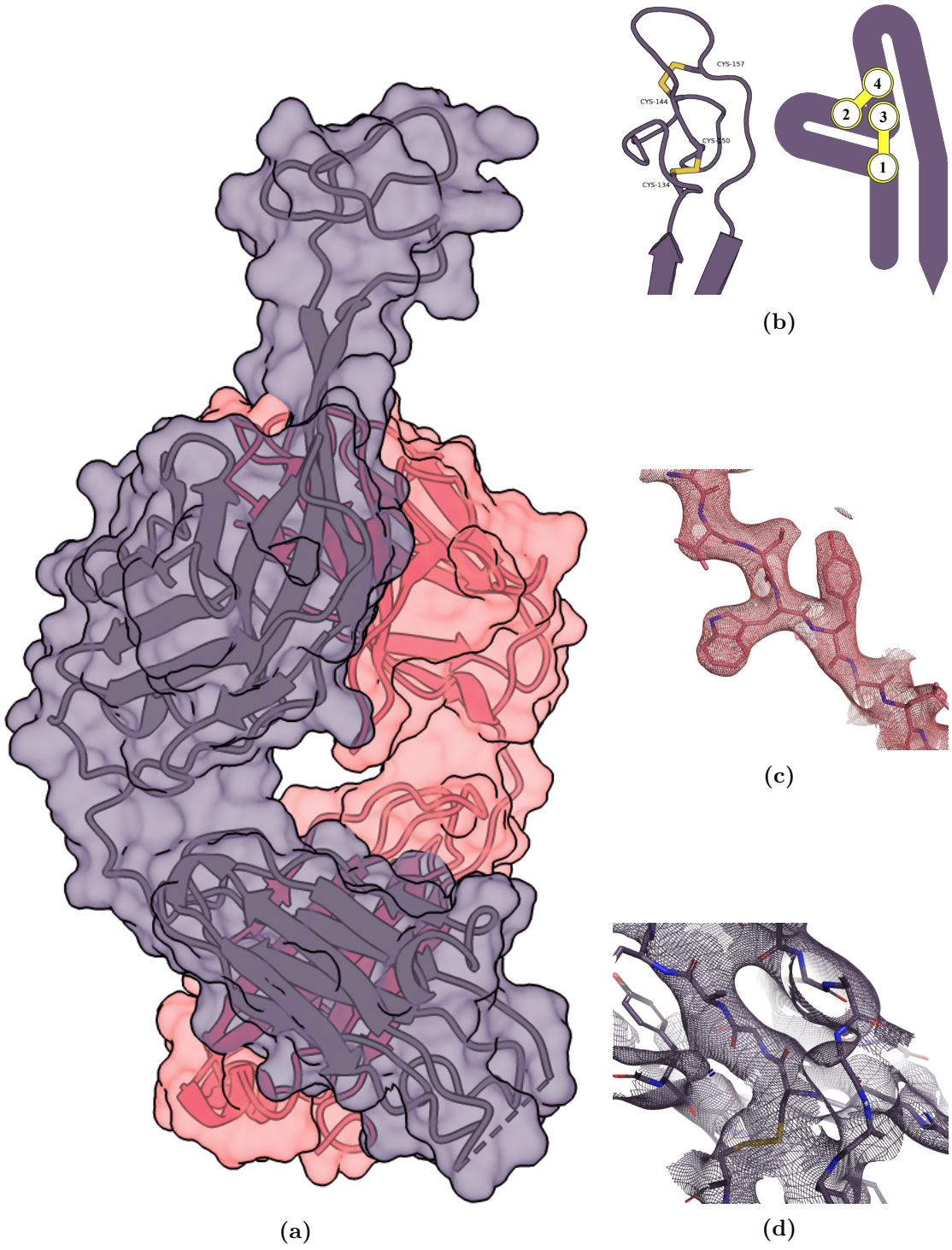


Figure 4.21: Ultralong Fab #30 crystal structure. (4.21a) Overall structure in cartoon representation with transparent surface overlay. The Fab #30 heavy chain is coloured purple, with the light chain coloured orange. (4.21b) Network of disulphide bonds within the ultralong knob domain, along with cartoon schematic. Representative electron density, contoured at 1.0 σ unit, for (4.21c) the tryptophan at the core of the VL domain and (4.21d) the 3-pass β-sheet of the ultralong knob domain. Surrounding structure hidden for clarity. Data were collected at beamline I03 at Diamond Light Source. Figures prepared using ChimeraX [257] (4.21a) and PyMOL (4.21b to 4.21d).

Table 4.9: Crystallographic data collection summary of Fab #30. Values in parentheses are for the highest resolution shell. Note further refinement of this model is required, and these statistics therefore do not represent a fully refined model.

Data collection	
Space Group	P 6_1 2 2
Cell Dimensions	
a, b, c (Å)	69.51, 69.51, 442.68
α, β, γ (°)	90.00, 90.00, 120.00
Resolution (Å)	60.20 - 2.65 (2.70 - 2.65)
R_{merge}	0.210 (5.072)
R_{pim}	0.034 (0.804)
$I/\sigma I$	11.1 (0.4)
Completeness (%)	100.0 (100.0)
Redundancy	37.5 (39.1)
$CC_{1/2}$	1.0 (0.2)
Refinement	
Resolution (Å)	43.569 - 2.710
No. Reflections	18275 (1600)
R_{work}/R_{free}	0.2610 / 0.3107
No. atoms:	
Protein	6704
Water	0
B-factors:	
Protein	112.96
Water	NA
R.m.s. deviations:	
Bond lengths (Å)	0.026
Bond angles (°)	1.519

4.5 Production of FMD VLPs by vaccinia virus co-infection

Cryogenic electron microscopy experiments of this chapter used VLPs produced at Pirbright on my behalf by the Immunogenetics group. The protocol to produce VLPs involves co-infection of HEK293T cells (Section 2.2.6), and therefore requires biosafety level 2 (BSL-2) containment. Additionally the protocol is an intermittent multi-week procedure, which would have proved logistically challenging whilst also performing experiments at the RFI.

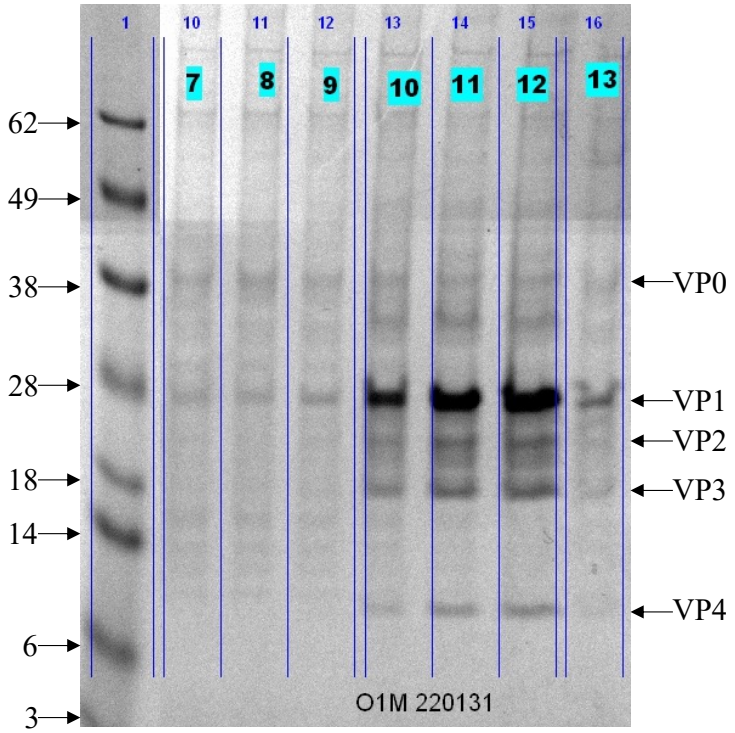


Figure 4.22: Reducing SDS-PAGE analysis of O1M VLPs. Fractions of the sucrose gradient purification were mixed with NuPAGE™ (Invitrogen™, ThermoFisher Scientific) sample reducing agent and loading buffer before being heated to 95 °C for 10 min before elctrophoresis. Constituent VLP components indicated on right edge. SeeBlue™ Pre-stained Protein Standard (Invitrogen™, ThermoFisher Scientific) was used as a molecular weight standard, with marker molecular weights provided in kDa.

Biophysical assessments of the interaction between Fabs #117, #126, and #30 and O1M VLPs using biolayer interferometry (Section 4.7) used VLPs prepared by the author, a representative SDS-PAGE analysis is provided in Fig. 4.22. This

particular batch was assessed by VHH ELISA (Section 2.7.7) as 99.96 % intact upon purification, and yielded 1088 µg/mL VLP, as measured by Qubit™ (Section 2.7.6).

4.6 Fab holo-structures

Cryogenic electron microscopy (cryo-EM) was used to investigate the structural aspect of antigen interaction of FMD VLPs by bovine ultralong antibodies #30, #117, and #126 as Fab fragments. Either O1M or A22 VLPs may be used to verify the predicted differential serospecificity of Fabs #117 and #126, since both Fabs are predicted to bind FMD serotypes Asia1 and SAT2 (Table 4.4). Both O1M and A22 VLPs include engineered stability mutants (Table 2.2), however at the time of experimentation, only antibody #30 had been identified as a high affinity, neutralising mono-specific antibody for O1M (ITB, personal communication), providing a positive control for comparisons of antibodies #117 or #126. Conversely, experiments with Asia1 or SAT2 may also be informative, as one would anticipate conservation of binding epitopes between antibodies with identical ultralong loops.

4.6.1 Electron cryomicroscopy specimen preparations

All specimens in this chapter were prepared by first mixing FMD VLPs with a $>$ 3-fold molar excess of each respective Fab under study for each VLP protomer. Given *Picornaviridae* viral capsids are an icosahedral arrangement of 60 protomer units [182] (Section 4.1.2), the Fab excess was calculated as $>$ 180-fold of the FMD VLP molar concentration. The mixtures were then applied to plasma-cleaned (Section 2.8.4.2) Quantifoil® Holey R2/1 Carbon film on 300 Copper mesh cryo-EM grids, and to plasma-cleaned (Section 2.8.4.2) Lacey ultrathin Carbon film on 400 Copper mesh cryo-EM grids (Section 2.8.4.1), and plunge frozen using a Vitrobot Mk IV (Section 2.8.4.3 and Table 4.10). Specimens were screened, and data were collected using a 300 keV Titan Krios electron cryomicroscope within STRUBI (Tables 2.13 and 4.10).

4.6.2 Electron cryomicroscopy data collections

All cryo-EM datasets in this chapter were acquired following a similar regime (Section 2.8.4.7). Data were collected using either a Gatan GIF Quantum K2 or Falcon IV Selectris X direct electron detectors at a nominal magnification of 130,000x (except for data collections involving SAT2/Eth, in which a nominal 165,000x magnification was used). The STRUBI Titan Krios detector was upgraded from the Gatan K2 to the Falcon IV in early 2022. In all acquisitions, a mild total exposure dose of 30 - 50 $e/Å^2$ was used since this provided sufficient signal for particle picking and downstream processing, and meant data could be collected rapidly with minimal radiation damage. A defocus range roughly spanning -3 µm to -1 µm in 0.3 µm increments was used to deconvolute the full range of spatial frequencies (Section 2.8.4.10). Acquisition parameters are shown in Table 4.10. Exposures were manually curated by application of filters (Table 4.11).

Table 4.10: Chapter 4 cryo-EM specimen preparations and data collection parameters. Det. - Detector. Mag. - Magnification. Exp. - Total exposure time. C2 - C2 aperture size. K2 - Gatan GIF Quantum K2 detector. F.IV - Falcon IV Selectris X detector. †A 0.2 µm increment was used in this case. All data collections used an objective aperture of 100 µm, and were performed at 300 keV using the STRUBI Titan Krios electron cryomicroscope.

Section	Antigen	Fab	Grid Type	Volume (µL)	Det.	Mag. (kX)	Pixel size (Å/px)	Exp. (s)	Dose rate (e/Å^2/s)	Defocus (µm)	C2 (µm)	Energy filter (eV)
4.6.3	SAT2 12S	#117	Quantifoil®	3.0	K2	130	1.05	4.4	9.09	-2.8 : -1.4†	70	20
4.6.4	SAT2 12S	#126	Lacey	3.5	F.IV	130	0.96	6.2	8.05	-3.5 : -0.8	50	10
4.6.5	SAT2 75S	#117	Quantifoil®	3.0	F.IV	165	0.73	3.4	11.64	-2.9 : -0.8	50	10
4.6.5	SAT2 75S	#126	Quantifoil®	3.0	F.IV	165	0.73	3.4	11.64	-2.9 : -0.8	50	10
4.6.6	O1M 75S	#126	Quantifoil®	3.0	K2	130	1.05	3.9	8.71	-2.7 : -1.2	70	20
4.6.7	O1M 75S	#117	Quantifoil®	3.0	F.IV	130	0.96	6.2	8.05	-3.5 : -1.1	50	10
4.6.8	O1M 75S	#30	Quantifoil®	3.0	K2	130	1.05	3.9	8.47	-2.7 : -1.2	70	20

Table 4.11: Chapter 4 cryo-EM exposure curation and particle picking. †Particle template picking algorithm picked excessive junk particles that could not be eliminated by algorithm parameterisation. Subsequent deep picking using Topaz [255] was required (Section 4.6.4.1).

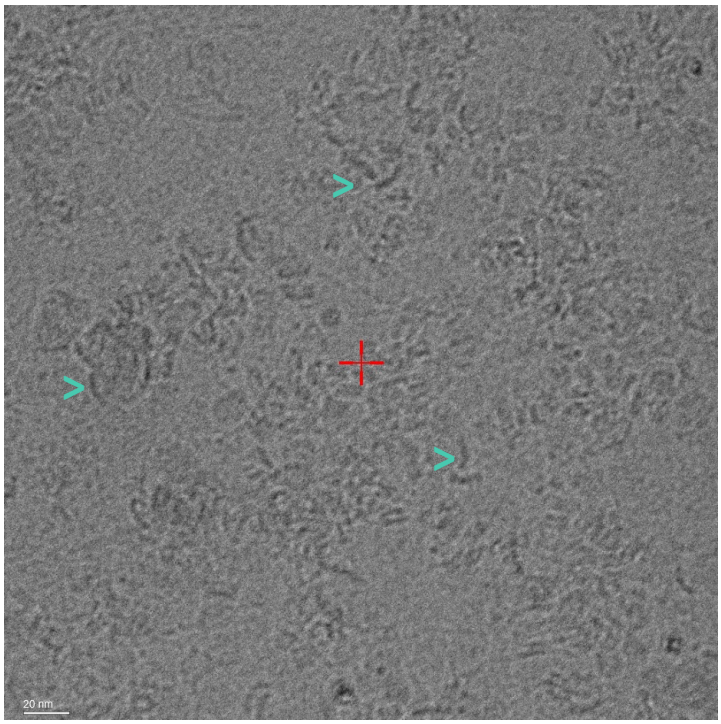
Section	Initial Exp.	Defocus average (Å)	CTF fit (Å)	Rel. ice thickness	Max. in-frame motion	Total motion (px)	Defocus range (Å)	Final Exp.	Template dia. (Å)	Box (px)	Bin (px)	Observed particles	Picked particles
4.6.3	13,553	-	2.105 - 10.055	0.981 - 1.066	-	-	27 - 4075	13,327	210	288	72	4,674,465	493,093
4.6.4	13,163	7737 - 42001	0.917 - 7.307	0.934 - 1.200	0.000 - 14.223	0.00 - 106.23	51 - 5023	13,163	210	288	72	8,288,923	91,372†
4.6.6	4,039	-	-	-	-	-	-	4,039	315	448	56	5,361,135	97,544
4.6.7	10,801	-	1.924 - 27.065	0.965 - 1.053	0.122 - 4.152	1.94 - 27.45	-	10,965	425	480	56	10,751,713	267,890
4.6.8	4,618	-	-	-	-	-	-	4,170	315	448	56	2,769,390	46,778

4.6.3 Interaction of SAT2 12S pentamers by Fab #117

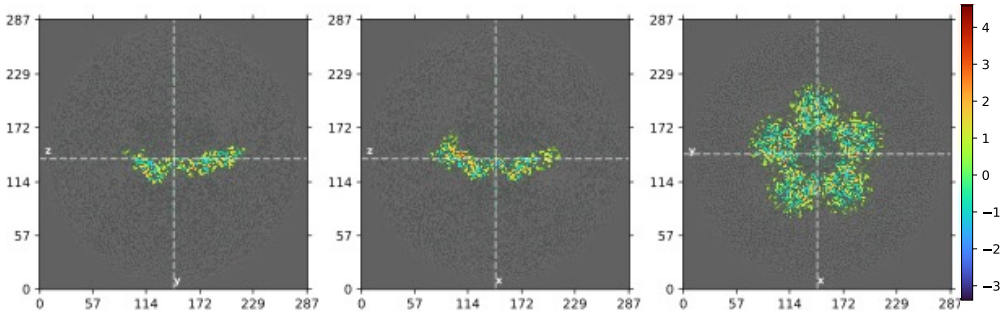
First, to confirm whether antibodies #117 and #126 (as Fabs) interact with a shared epitope, cryo-EM was used to characterise and compare the binding epitopes accessed by these antibodies. SAT2 VLPs were selected over Asia1 due to antigen availability, and as SAT2 biology is relatively under reported in the literature.

4.6.3.1 SAT2 12S-Fab #117 Data Collection

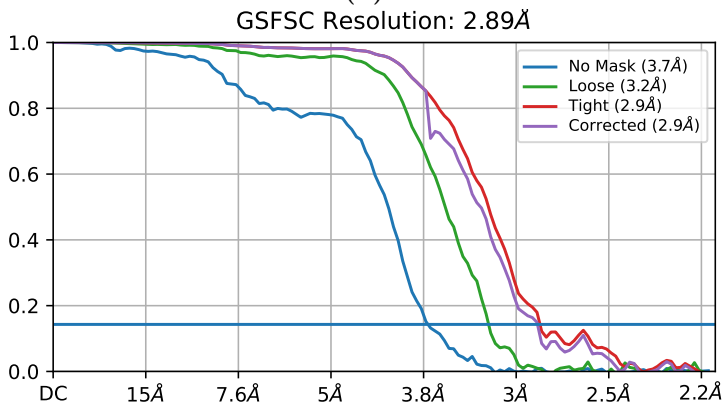
Specimen preparation constituting Fab #117 mixed with SAT2 VLPs (produced by vaccinia virus; Section 2.2.6) were prepared and data collected as previously indicated (Tables 4.10 and 4.11). A representative micrograph is provided in Fig. 4.23. Immediately clear in the micrographs is the lack of intact SAT2/Zim VLPs (Fig. 4.24); indeed no intact particles were observed in any specimen or control (without addition of Fab #117) grid, indicating that capsid dissociation is not an effect of Fab #117 addition (Fig. 4.23). Subsequent analysis confirmed the instability of SAT2/Zimbabwe/7/83 VLPs, with the majority of particles showing dissociation in a VHH ELISA within hours of purification (ITB, personal communication).



(a)



(b)



(c)

Figure 4.23: Electron cryomicroscopy of Fab #117 in complex with FMD serotype SAT2 dissociated 12S pentamers. (4.23a) Example micrograph, with particles marked by green arrowheads. Scale bar indicates 20 nm. (4.23b) Central slices from density maps after final round of refinement in C5 symmetry. (4.23c) Autotightened GSFSC plot indicating determined resolution. 0.143 FSC is marked by a horizontal line.

Even from the 2D projection images, diffuse density to the capsid internal surface was observed, which could correspond to Fab density (Fig. 4.24). A curious observation is the presence of additional diffuse density to the external surface of the capsid density. This region was not resolved in 3D reconstructions, however it is conceivable that it represents non-specific interactions between Fab #117 towards the rim of the pentamer. Finally, particle poses demonstrated a preferential orientation for side views, however, this meant the fivefold symmetry axis was perpendicular to the field of view, and hence imposition of C5 symmetry would effectively sample all particle orientations, permitting full structural determination without optimisation to alleviate preferential orientation.

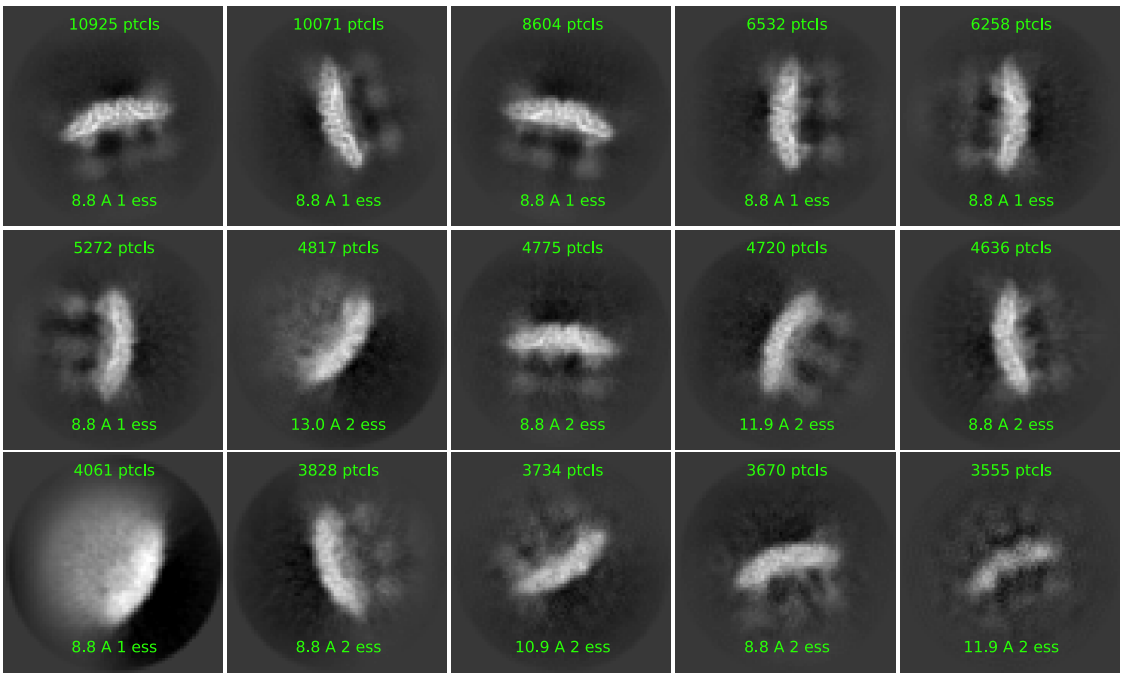


Figure 4.24: 2D class averages for SAT2 in complex with Fab #117. Highest populated 2D class averages with indicated population, estimated resolution and effective sample size (ESS).

4.6.3.2 Fitting structures to the observed cryo-EM density

Following exposure and particle curation (Table 4.11), 110,210 particles were used in the refinement of a 2.89 Å map bearing C5 symmetry and consistent with a pentamer subunit. The only available SAT2 structure (PDB ID: 5ACA [183]) for

strain SAT2/Zim/7/83 was fitted to the map (Fig. 4.25) and used for model building of the pentamer, as was each of the AF2 models for scFv #117 (Fig. 4.15). Each AF2 model was aligned to the observed density, and cross correlation coefficient (CCC) of the fit noted. The highest scoring model, ranked 4th in prediction confidence by AF2, was used for further refinement and model building.

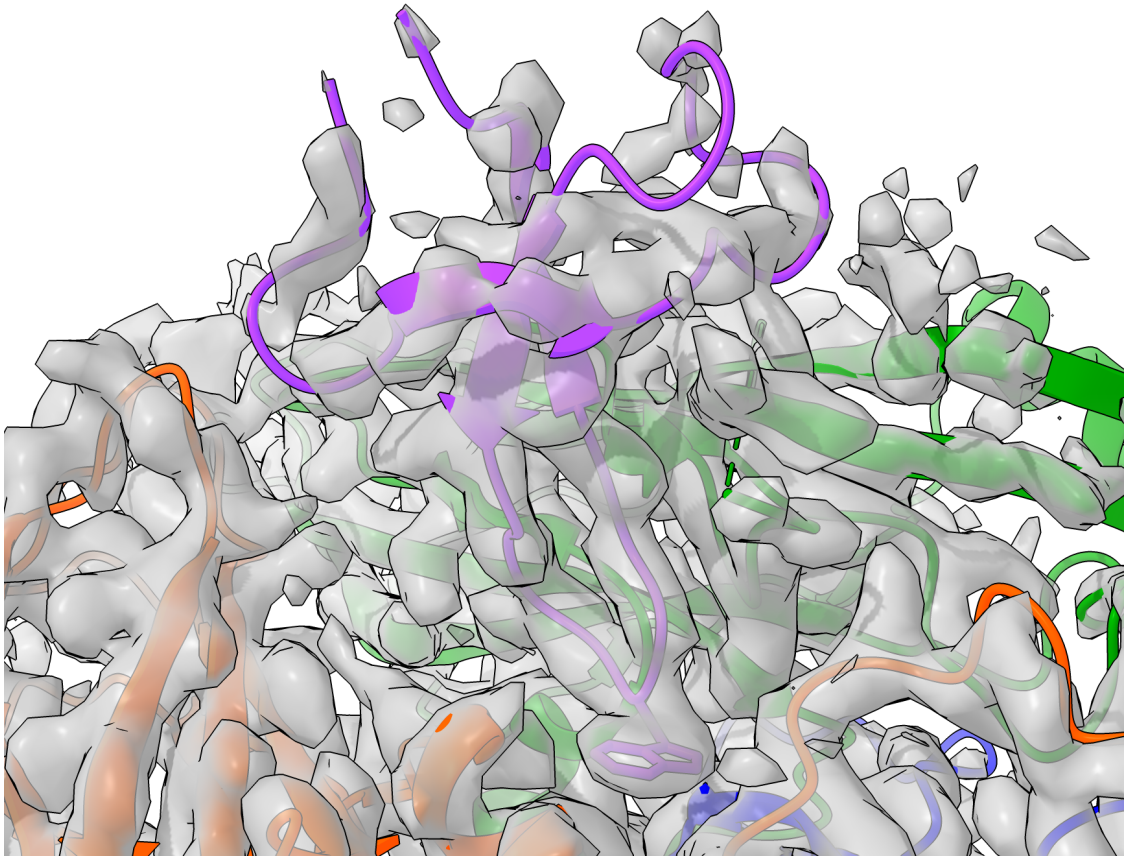
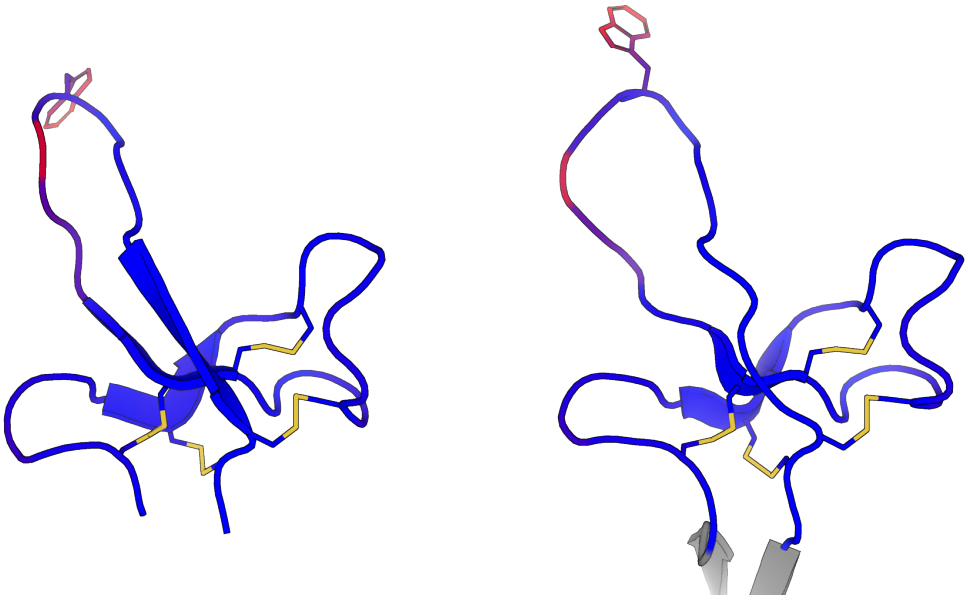


Figure 4.25: Fitting of SAT2 pentamer and AF2 scFv #117 to observed cryo-EM density map. SAT2 12S pentamer structure (PDB ID: 5ACA [183]) and AF2 scFv #117 model #4 (purple) fitted using Chimera [257]. Observed density is transparent white, VP1 is blue, VP2 is green and VP3 is orange. Perspective from the centre of the internal surface of the 12S pentamer, looking outward towards the 3-fold symmetry axis.

The tryptophan at the distal tip of the #117 CDR3H subloop (Trp138) can clearly be seen in the observed density (Fig. 4.25), allowing refinement of the remaining loop positions into the density. The 12S model closely aligned to the density, and only minor repositioning of loops was required. No density was observed for VP4, nor the Fv or CDR3H stalk regions of Fab #117.

4.6.3.3 Comparing observed against predicted #117 CDR3H structure

Observation of the #117 knob mini-domain by cryo-EM allows assessment of the AF2 predicted structure accuracy. The AF2 model #4 gave the best CCC when fitted to the observed density, with a small (6.2 Å) repositioning of the long subloop (residues 132:145), and rotation of the Trp138 side chain (Fig. 4.26).



(a) Observed **(b)** Predicted

Figure 4.26: Comparison of observed and predicted #117 knob mini-domain structures. Structures aligned and coloured by the PyMol ColorByRMSD script [264]; blue indicates low RMSD, whilst red indicated high RMSD. Knob mini-domain cysteines, and tyrosine138 shown in stick notation.

4.6.3.4 The SAT2 12S epitope accessed by Fab #117

Antibody #117 binds to the internal surface of SAT2 12S dissociated pentamers at a deep pocket of complimentary geometry between VP3 and VP2 (Fig. 4.27). PISA analysis [233] was used to identify interacting residues of Fab #117 with SAT2 12S pentamers.

As previously mentioned, a tryptophan (Trp138) can be seen protruding from the #117 CDR3H, making a hydrogen bond between the pyrrole nitrogen and the backbone carbonyl of Arg40 of VP3, and held in place via T-shaped π-stacking

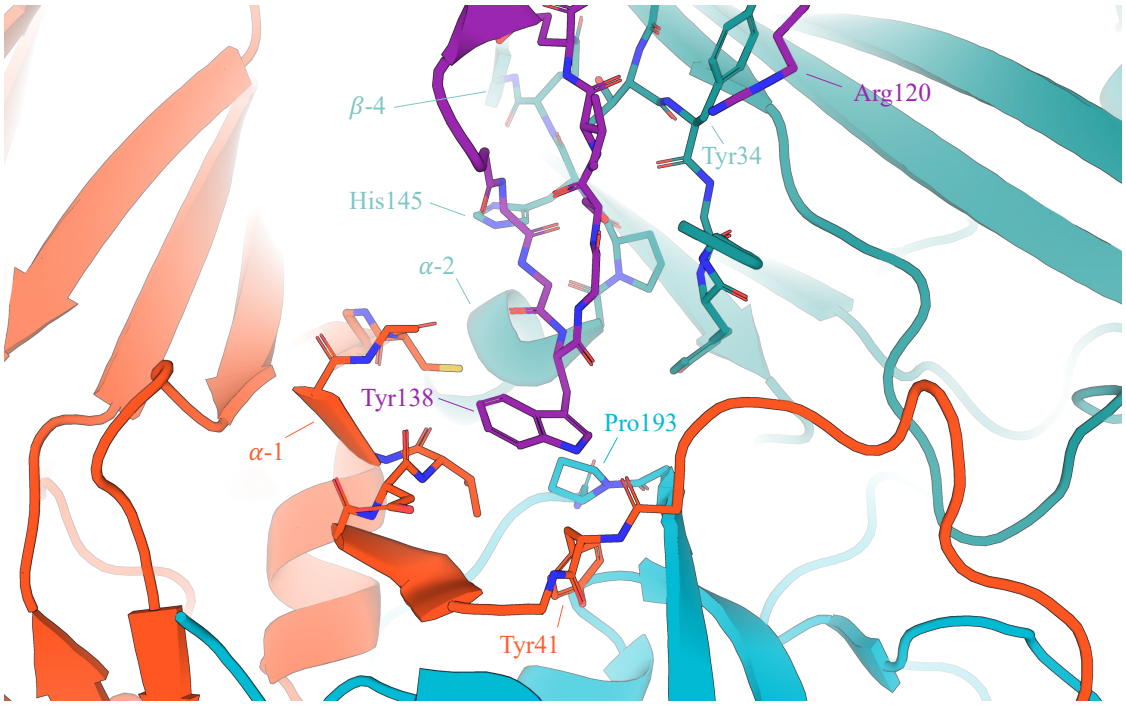


Figure 4.27: The SAT2 12S epitope for antibody #117. Stick notation for the long subloop of the #117 CDR3H, and SAT2 residues within 5 Å of the #117 CDR3H. Annotation colour indicates chain. Perspective and colouring as in Fig. 4.25. Figure produced using ChimeraX [257].

interaction with proline 193 of VP1. Trp138, and the preceding two glycines, fit tightly within a pocket flanked by α-1 of VP3, the N-terminus and a turn between α-2 and β-4 of VP2, and is bordered by the N-terminus of VP3, and His145 of VP2. Glycine 136 and 137 make hydrogen bonds with His145 of VP2, and Cys51 of VP3 respectively. Fab #117 makes further hydrogen bonds with VP2; Glu142 bonds to Gly31 and Ile32, and Ser140 bonds to Tyr36, and additionally a salt bridge between Arg120 and Asp38. Finally, Tyr34 of VP2 extends toward the CDR3H knob domain, finding a pocket between the side chains of Arg120 and Leu143, and the backbone carboynl of Ile141, and hydrogen bonding with the pyrrole nitrogen of Trp123.

The identification of an antibody that binds specifically to an epitope on the internal surface of the VLP capsid implies antibody #117 was produced in response to dissociated VLP antigen.

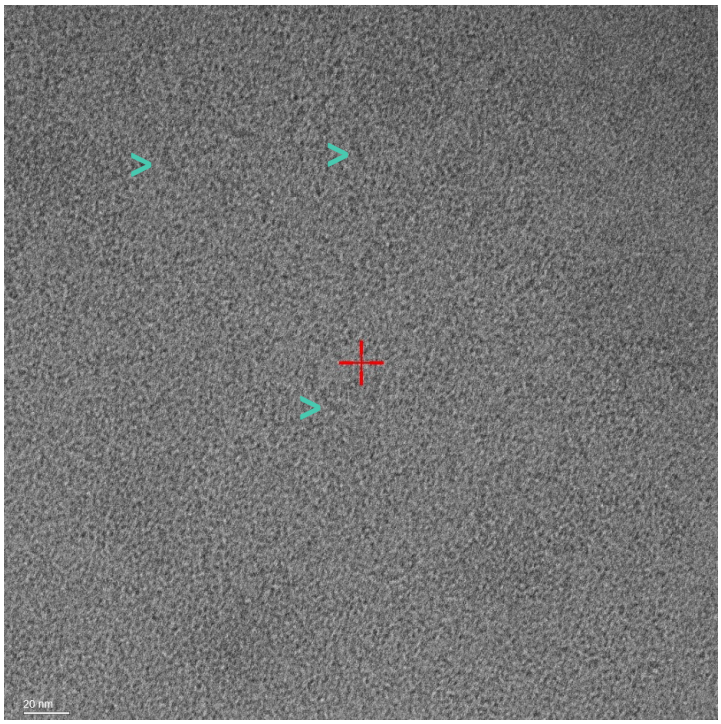
4.6.4 Interaction of SAT2 12S pentamers by Fab #126

It remains uncertain whether the light chain conformationally constrains ultralong CDR3H loops. Given the recognition of the SAT2 12S internal epitope by Fab #117 appears to be entirely mediated via the CDR3H loop (Section 4.6.3.4), a structure of SAT2 12S in complex with Fab #126 was determined to test this hypothesis.

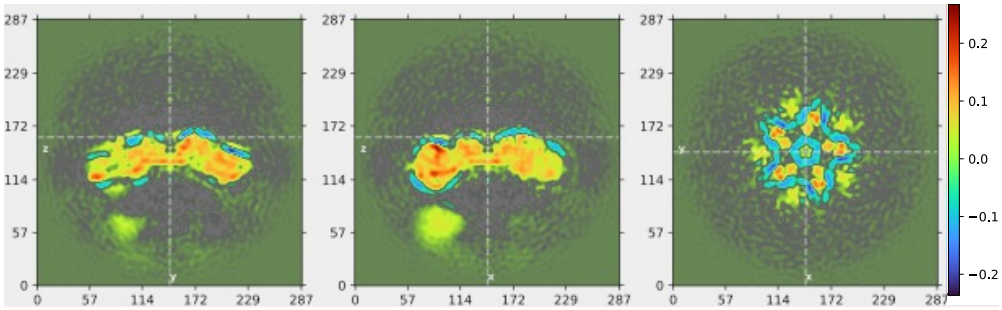
4.6.4.1 SAT2 12S-Fab #126 Data Collection

Specimen preparation constituting Fab #126 mixed with SAT2 VLPs (produced by vaccinia virus; Section 2.2.6) were prepared and data collected as previously indicated (Tables 4.10 and 4.11). A representative micrograph is provided in Fig. 4.28.

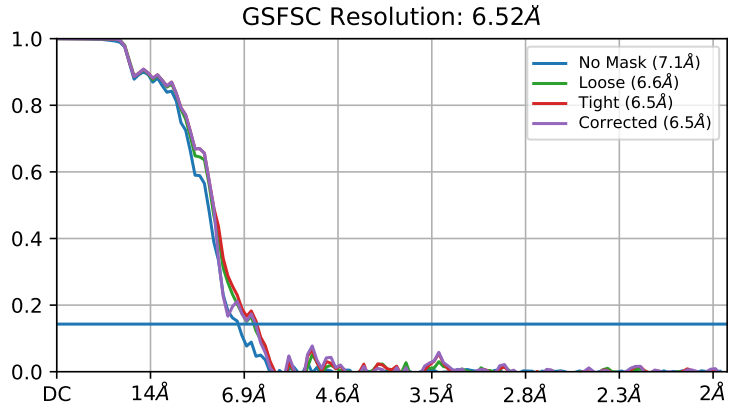
As observed previously (Section 4.6.3.1), SAT2/Zim/7/83 VLPs carrying the S2093Y mutation dissociated to 12S pentamers prior to cryo-EM grid plunging, with characteristic thin curved densities represented in the 2D class averages (Fig. 4.29).



(a)



(b)



(c)

Figure 4.28: Electron cryomicroscopy of Fab #126 in complex with FMD serotype SAT2 dissociated 12S pentamers. (4.28a) Example micrograph, with particles marked by green arrowheads. Scale bar indicates 20 nm. (4.28b) Central slices from density maps after final round of refinement. (4.28c) Autotightened GSFSC plot indicating determined resolution. 0.143 FSC is marked by a horizontal line.

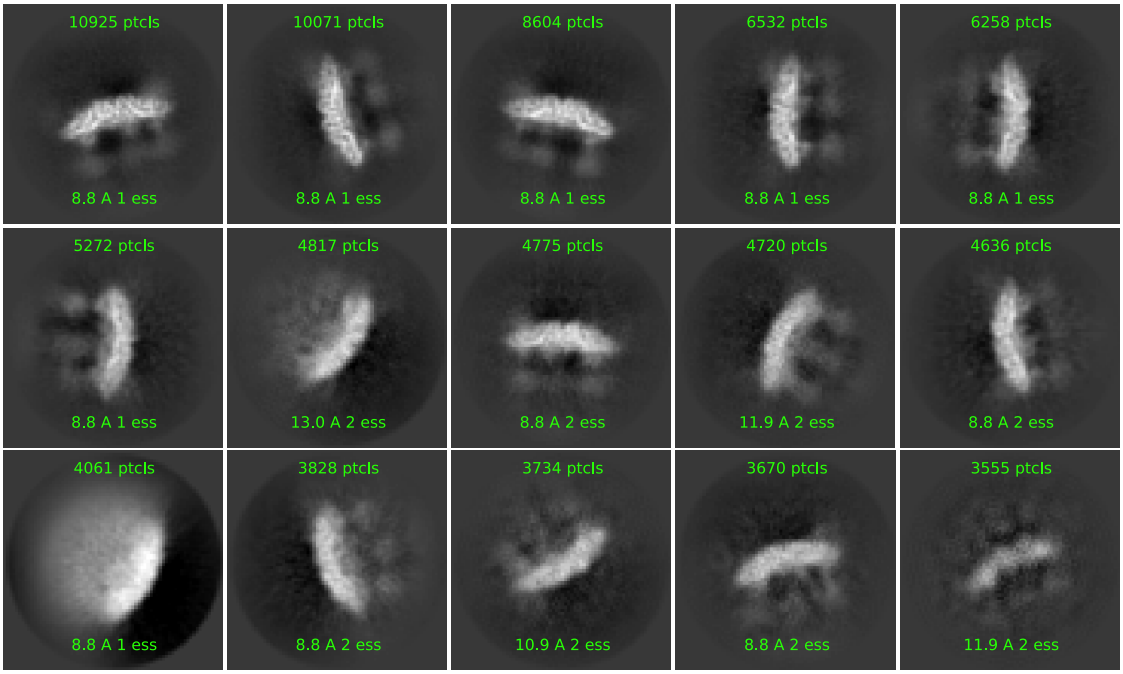


Figure 4.29: Initial 2D class averages for SAT2 in complex with Fab #126. Particles picked by template picking algorithm. Highest populated 2D class averages with indicated population, estimated resolution and effective sample size (ESS).

The 2D class averages for SAT2 12S pentamers in complex with Fab #126 appear indistinguishable to those of SAT2 12S in complex with Fab #117 (Figs. 4.24 and 4.29), with the same main features; a heavy preference for side views, decoration to the internal pentamer surface and additional diffuse density to the external pentamer rim.

Template picking of particles resulted in 'junk' particles (signal originating from patterned noise, non-sensical densities, edges of grid holes etc) being selected with high frequency (Fig. 4.29). As a result, the 91,732 total template-picked particles were used to train the Topaz deep picking algorithm [255], resulting in a total of 154,455 extracted particles. This secondary particle stack was reclassified into 2D classes, of which 116,542 particles were retained for downstream analysis. Despite the deep picking, atomic resolution density was not determined within this project's time frame for the SAT2 12S/Fab #126 dataset, with the processing efforts thus far achieving a 6.52 Å map (Fig. 4.28c). Further structural refinement will be performed in the future. The similarity of 2D classes to that of SAT2 12S

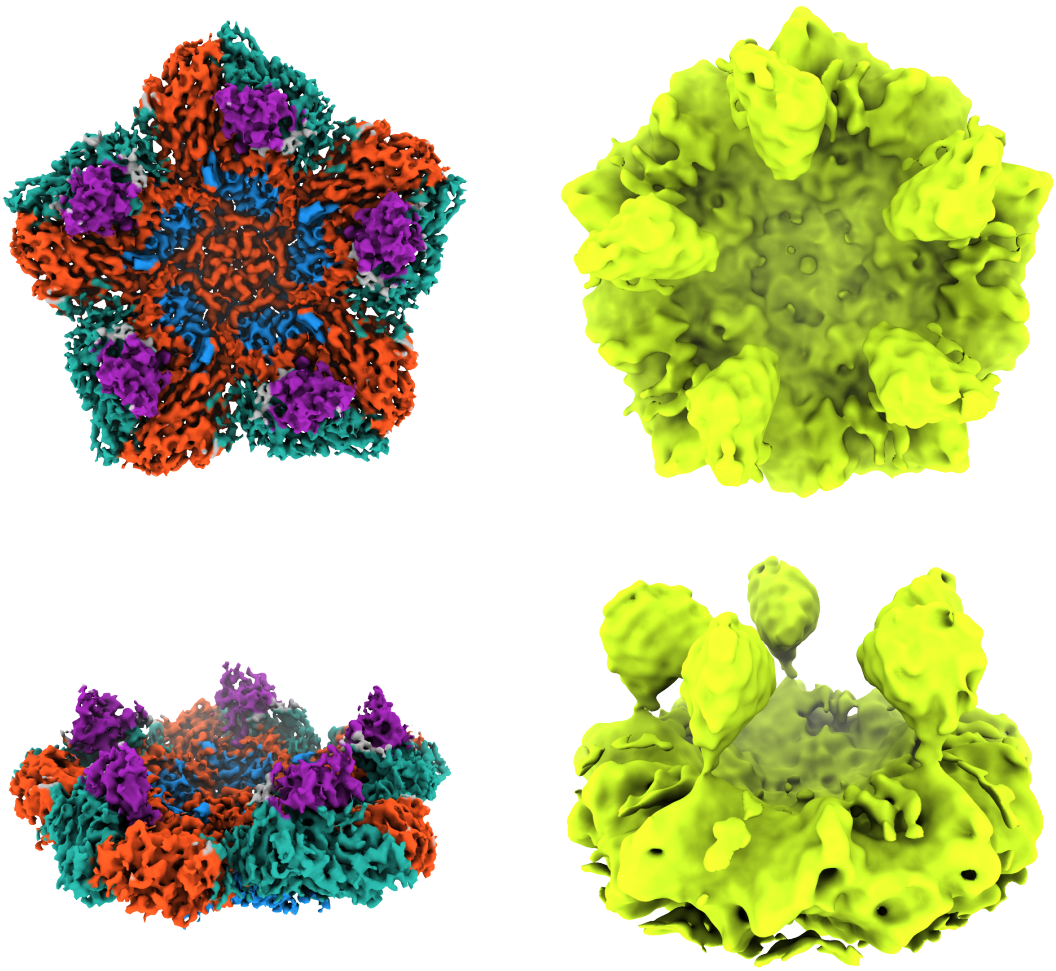
in complex with Fab #117, and particularly the presence of secondary structural information, warrants hope that near atomic resolution may also be achieved for Fab #126 binding SAT2 12S pentamers.

Since the observation of VLP dissociation (Section 4.6.3.1), production of these VLPs was halted, consequently this cryo-EM experiment used a residual batch of SAT2/Zim/7/83 VLPs that had been stored at 4 °C for an extended period. As a result, the pentamers may have further dissociated or otherwise become increasingly degraded, introducing a heterogeneity to the sample, and complicating both particle picking and reconstruction.

4.6.4.2 The SAT2 12S (approximate) epitope accessed by Fab #126

Additional density can be observed to the internal surface of the pentamer, but the resolution achieved thus far was deemed too low to attempt model building. Despite this, the observed density may be visually compared to that of SAT2 12S being recognised by Fab #117 to infer the approximate antigenic site with the caveat that map handedness cannot be confirmed and data must therefore be interpreted with extreme caution at this stage.

Upholding this caveat, comparison of the SAT2 12S/Fab #126 dataset to the SAT2 12S/Fab #117 dataset showed additional density to the internal pentamer surface that broadly corresponds with the region occupied by the Fab #117 knob mini-domains (purple) (Fig. 4.30), implying the same approximate antigenic site being used, as anticipated by the sequence identity of both the CDR3H knob mini-domains (Fig. 4.5) and the SAT2 proteins.



(a) SAT2 12S/Fab #117 **(b)** SAT2 12S/Fab #126

Figure 4.30: Comparison of cryo-EM density maps for SAT2 12S in complex with Fab #117, or #126. Maps thresholded to 0.25 and 0.0235 for SAT2 12S/Fab #117 and SAT2 12S/Fab #126 respectively. SAT2 12S/Fab #117 coloured as in Fig. 4.27; blue; VP1, green; VP2, orange; VP3, purple; Fab #117 knob mini-domain. Top and bottom images are related by a 70° rotation in the plane of the figure. Figures prepared using ChimeraX [257].

4.6.5 Introduction of thermostable SAT2 VLPs

Since the earlier VLP work, a new batch of SAT2 VLPs had been engineered by Claudine Porta of STRUBI. This strain of VLPs was prepared by baculovirus system, is topotype Ethiopia/65/2009, and contains a K93C stability mutation to VP2, rendering the VLP thermostable. To clarify whether Fab #117 or #126 is able to recognise intact 'SAT2/Eth' VLPs, additional cryo-EM experiments were performed.

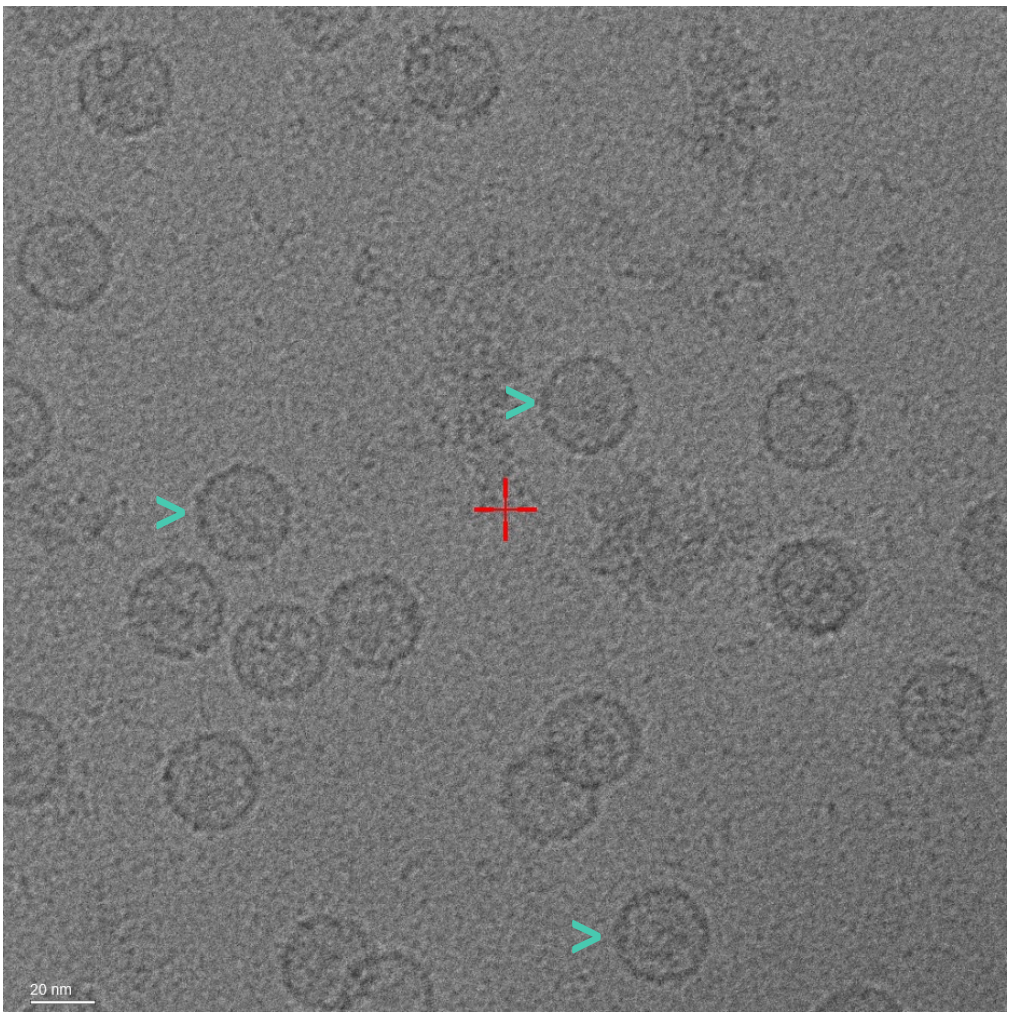
4.6.5.1 SAT2-Fab #117 and SAT2-Fab #126 Data Collections

Specimen preparations constituting either Fab #117 or #126 mixed with SAT2/Eth VLPs (produced by vaccinia virus; Section 2.2.6) were prepared and data collected as previously indicated (Table 4.10). Representative micrographs are provided in Figs. 4.31 to 4.33.

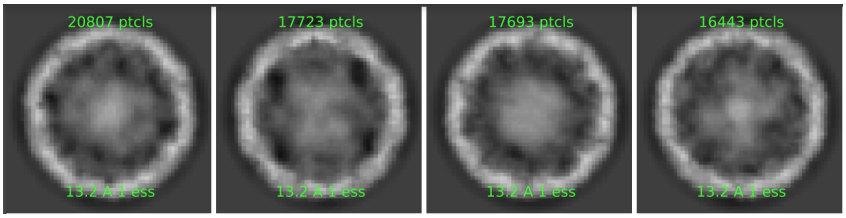
4.6.5.2 VHH ELISA confirms non-binding of Fabs #117 and #126 to intact SAT2

The 2D class averages for the intact SAT2/Eth datasets show only unbound SAT2 VLPs, with no clear differences between the control (without Fab) micrograph (Fig. 4.31b), and those for either Fab #117 (Fig. 4.32b) or #126 (Fig. 4.33b). These datasets were only processed up to the 2D classification step, which gave enough information to answer this hypothesis, on the presentation of circular capsid projections devoid of substantial external decoration.

FMD VLPs are relatively featureless at low resolution [182], hence the particle 2D alignment algorithm struggles to align the particles, and secondary information of the VLP capsid is averaged out. Independent experiments performed at Pirbright confirmed these data by homologous ELISA, where IgG #117 is only able to recognise dissociated SAT2 of the Zimbabwe and Egypt (classical vaccine antigen) topotypes, but neither intact or dissociated SAT2/Eth VLPs (Appendix D.1). Lack of binding to intact SAT2 corroborates the notion of SAT2/Zim VLP dissociation, as these VLPs were used in the vaccination experiment and if dissociated, possibly even prior to immunisation, the isolation of an antibody specific to the internal surface of the VLP is explained.

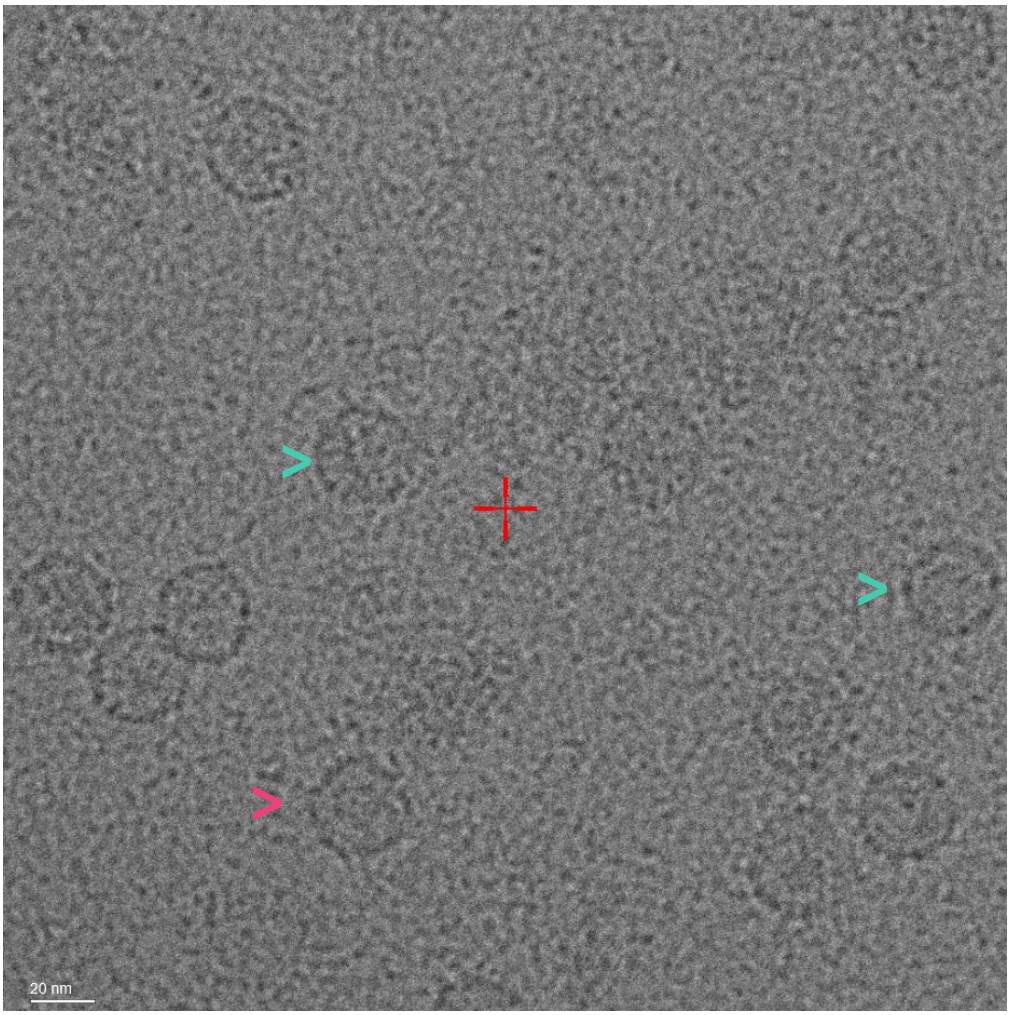


(a)

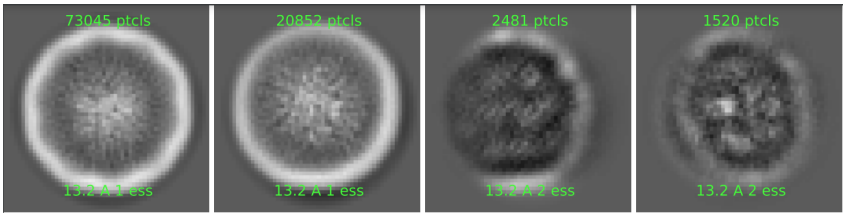


(b)

Figure 4.31: Electron cryomicroscopy of FMD serotype SAT2/Eth VLPs. (4.31a) Example micrograph, with particles marked by green arrowheads. Scale bar indicates 20 nm. (4.31b) Exemplar 2D class averages.

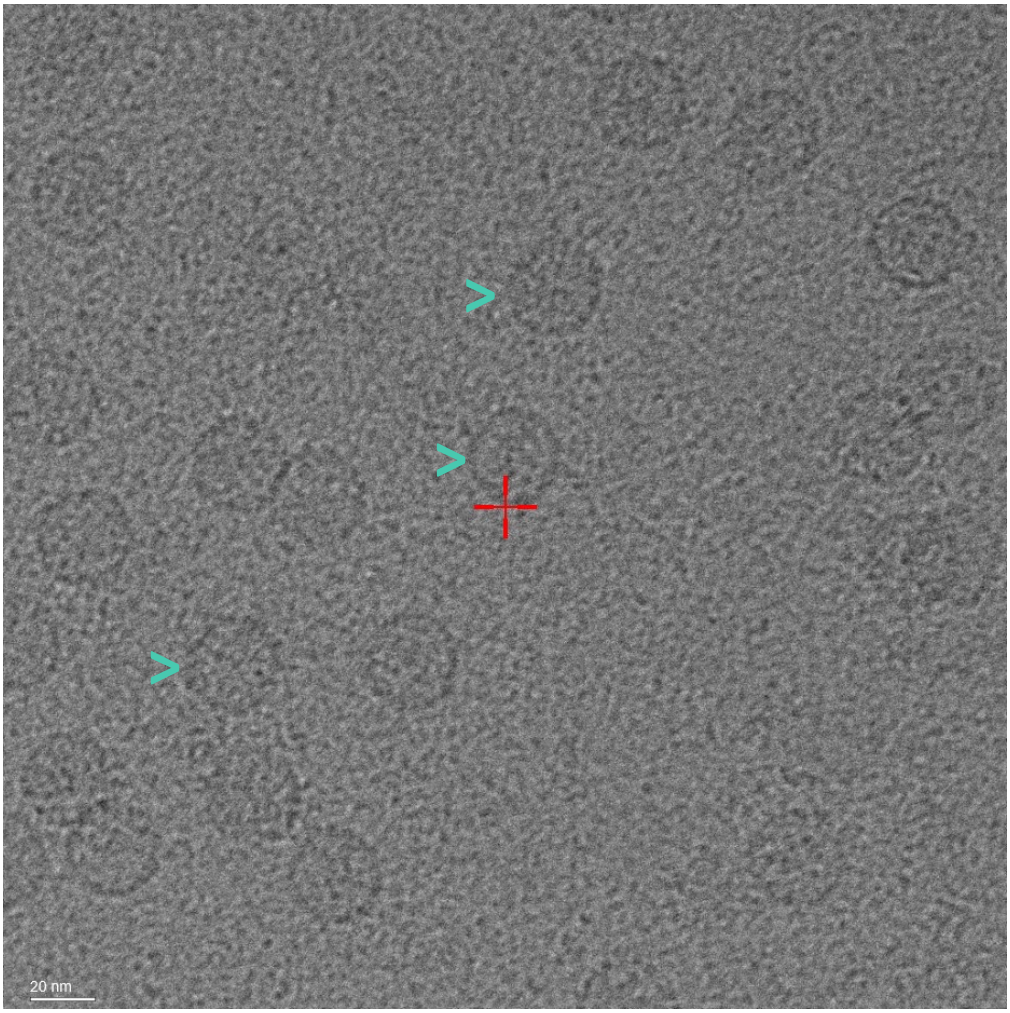


(a)

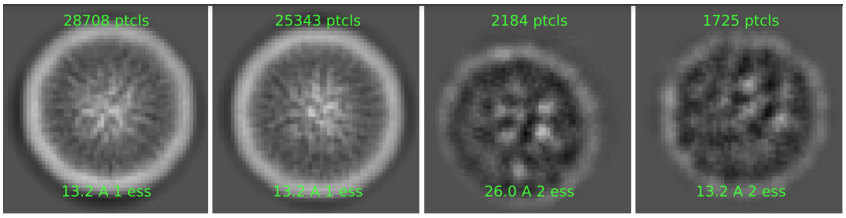


(b)

Figure 4.32: Electron cryomicroscopy of FMD serotype SAT2/Eth VLPs mixed with Fab #117. (4.32a) Example micrograph, with particles marked by green arrowheads. Pink arrowhead shows SAT2/Eth particle with potential Fab #117 decoration to exterior surface. Scale bar indicates 20 nm. (4.32b) Exemplar 2D class averages.



(a)



(b)

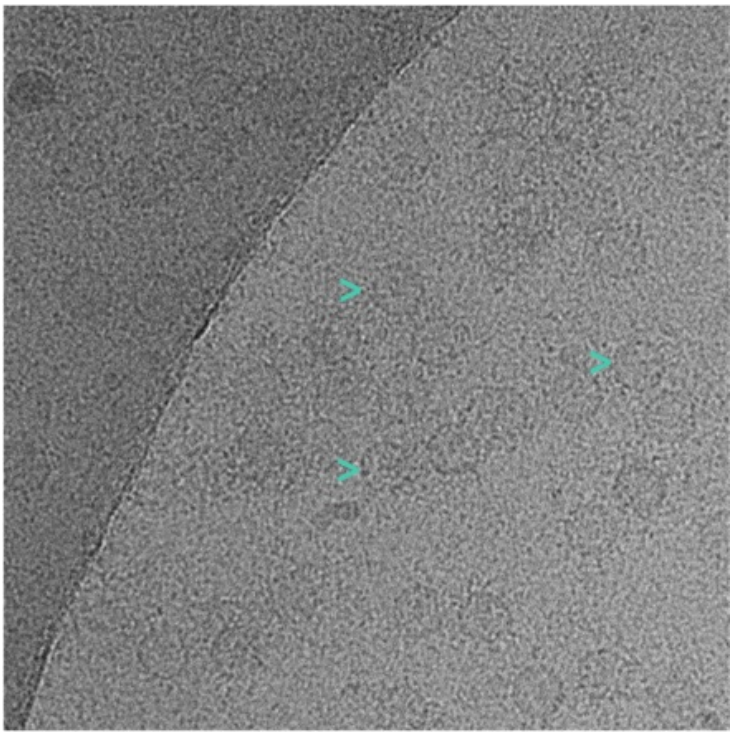
Figure 4.33: Electron cryomicroscopy of FMD serotype SAT2/Eth VLPs mixed with Fab #126. (4.33a) Example micrograph, with particles marked by green arrowheads. Scale bar indicates 20 nm. (4.33b) Exemplar 2D class averages.

4.6.6 Interaction of O1M VLPs by Fab #126

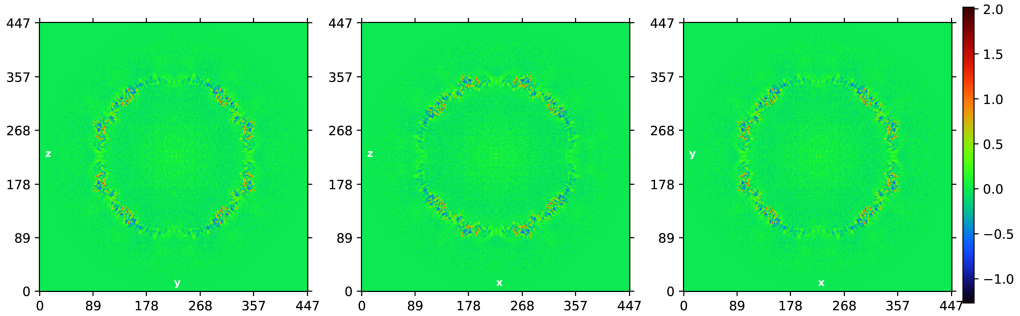
Contrary to the SAT2 experiments, Fab #126 was predicted to bind to O1M, whereas Fab #117 was predicted not to recognise O1M (Table 4.4), despite sharing an identical ultralong heavy chain with Fab #126, implying a long-range contribution to structure, and therefore, paratope by the light chain. Intuitively, with a conserved ultralong heavy chain, known to interact autonomously with antigen [78, 84, 85, 132, 137–139], one would expect equivalent antigenic recognition behaviour (as inferred with SAT2 pentamers; Fig. 4.30). Cross-serotype recognition of FMD VLPs is interesting as it implies either epitopic conservation between serotypes, or that a different epitope is being accessed by the same ultralong CDR3H by binding with a different paratope on opposing sides of the knob mini-domain. To resolve this question, further cryo-EM experiments were performed. Note that A22 VLPs could also have been used for this study, but as explained before, were not as reliably produced as O1M VLPs at the time of experimentation (Section 4.6).

4.6.6.1 O1M-Fab #126 Data Collection

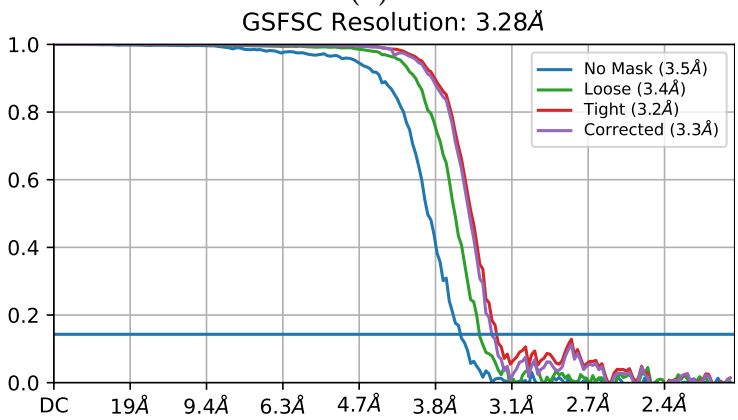
Specimen preparation constituting Fab #126 mixed with O1M VLPs (produced by vaccinia virus; Section 2.2.6) were prepared and data collected as previously indicated (Tables 4.10 and 4.11). A representative micrograph is provided in Fig. 4.34.



(a)



(b)



(c)

Figure 4.34: Electron cryomicroscopy of Fab #126 in complex with FMD serotype O1M VLPs. (4.34a) Example micrograph, with particles marked by green arrowheads. (4.34b) Central slices from density maps after final round of refinement. (4.34c) Autotightened gold-standard Fourier shell correlation (GSFSC) plot indicating determined resolution. 0.143 FSC is marked by a horizontal line.

2D class averages consisted of circular projections bearing secondary structure, consistent in size and shape with that anticipated for intact particles (Fig. 4.35). Additional diffuse density can be seen around the exterior of the particles, which may be commensurate with some Fab decoration.

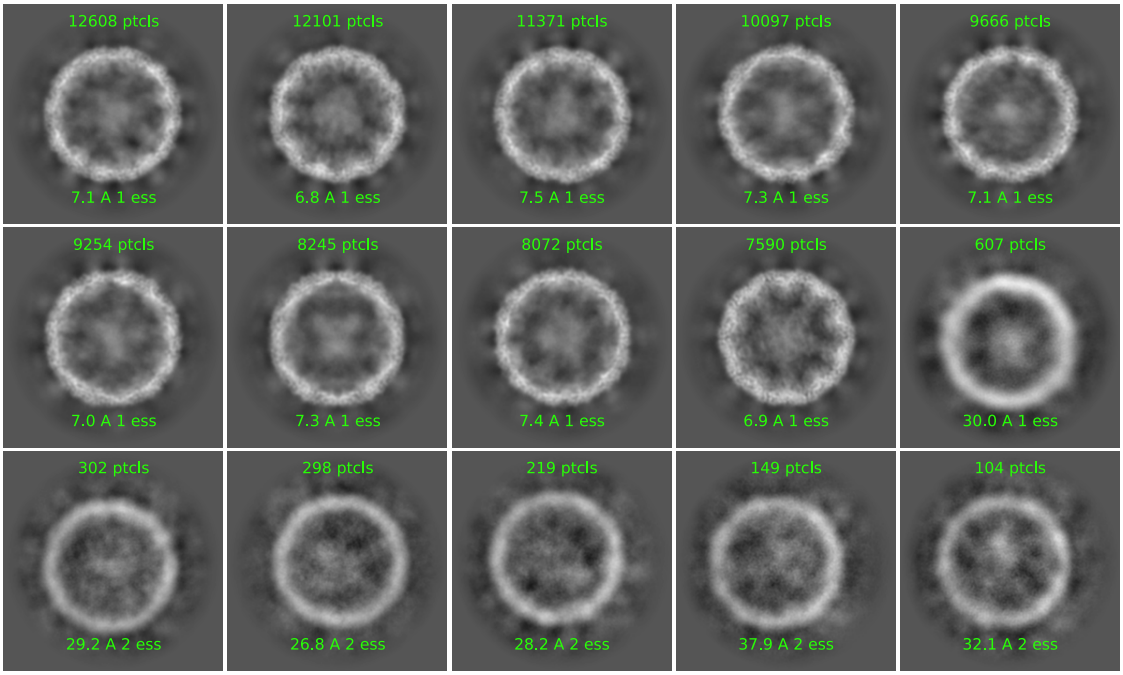


Figure 4.35: 2D class averages for O1M in complex with Fab #126. Highest populated 2D class averages with indicated population, estimated resolution and effective sample size (ESS).

Once picking parameters were optimised (Table 4.11), 97,544 particles were selected from 4,039 accepted exposures for downstream processing. Following several rounds of 3D refinement, a map with indicative resolution of 3.28 Å was achieved using 96,701 particles (Fig. 4.34c).

The observed resolution of the O1M/Fab #126 density map was more varied, with good resolution (2.5 Å - 3.5 Å; coloured red to yellow) for the capsid shell, but lower resolution for the ultralong knob mini-domain of Fab #126 (4.5 Å to exceeding 6.5 Å; coloured green to blue) (Fig. 4.36).

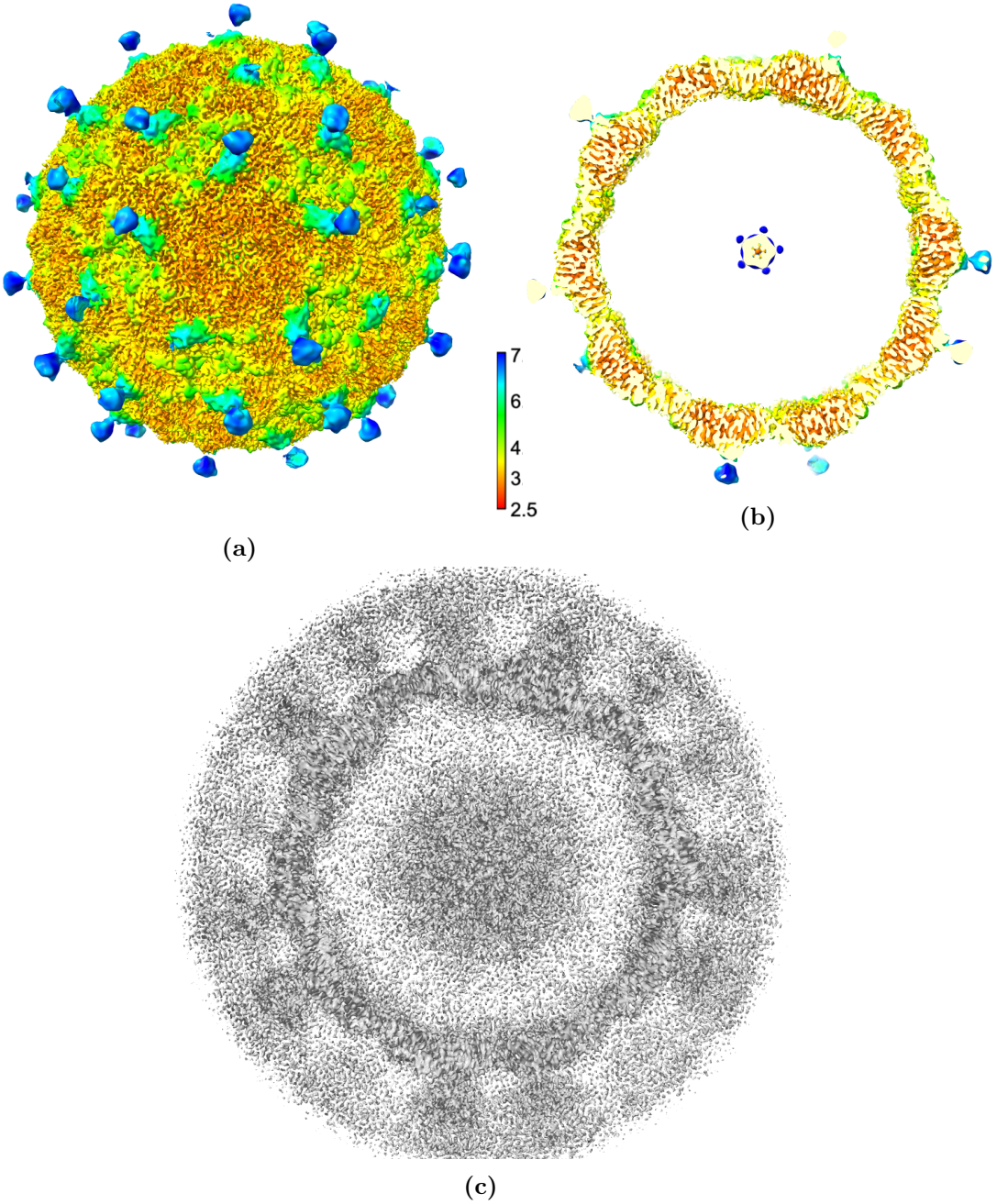


Figure 4.36: Observed O1M/Fab #126 cryo-EM density map coloured by determined local resolution. Locally filtered map (4.36a), and central slice thereof (Fig. 4.36b) reconstructed using 96,701 particles initially picked by template picking using cryoSPARC [249]. Map thresholded to 0.0491. Resolution scale bar in Å. (4.36c) Central slice through unfiltered map thresholded to 0.0491 for comparison. Figure produced using ChimeraX [257].

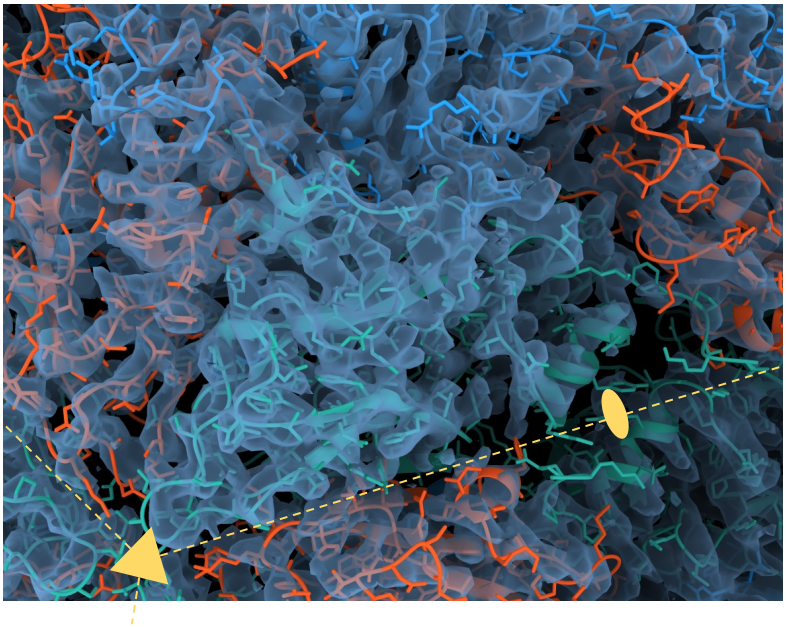
4.6.6.2 The proposed interaction between Fab #126 and O1M

By contrast to the O1M/Turkey/69 93C stabilising mutation at the two-fold axis (VP2 residue 93) used in this study (Table 2.2), a 3.5 Å resolution crystal structure of stabilised O1M/Turkey/69 VLPs (PDB ID: 5DDJ [183]) bearing a tyrosine stabilising mutation at the same site, VP2 residue 93, was fitted to the observed electron density. Except for an additional cysteine introduced at S1134 of VP1, 5DDJ is otherwise identical to the VLPs examined here by cryo-EM.

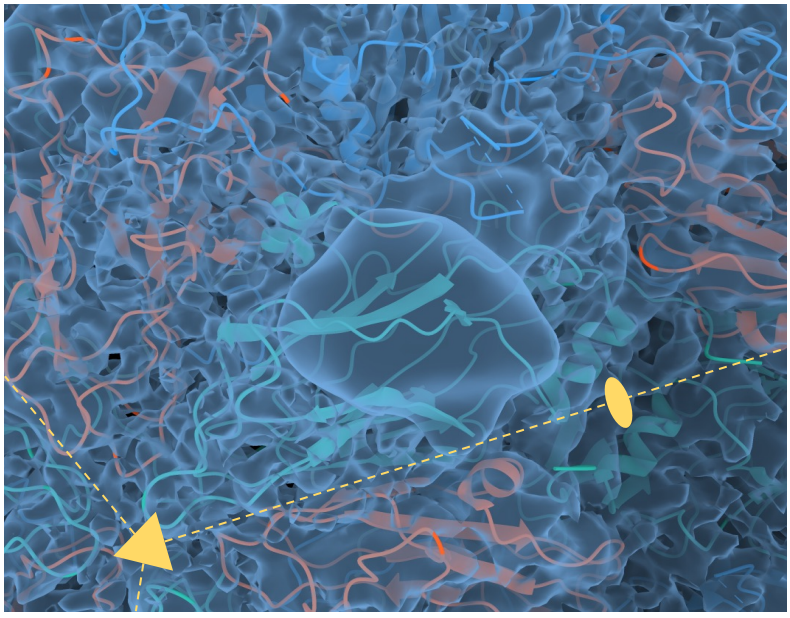
Whilst structural alignment of the O1M VLP crystal structure to the observed capsid density was possible, modelling of the ultralong CDR3H was prevented by insufficient resolution at this locus (Fig. 4.36). Indeed there is no evidence the observed density is actually an ultralong Fab at all.

Additional low level diffuse density can be seen between the 3-fold and 2-fold axes (Fig. 4.37). It is conceivable that this represents weak, possibly non-specific, interactions of Fab #126. Note that in the case of high affinity specific interactions with a well ordered capsid, even if the ultralong stalk region were flexible, the knob mini-domain should still be well resolved, as is the case with Fab #117 binding to dissociated SAT2 (Section 4.6.3.2), and Fab #30 binding to O1M (Section 4.6.8.2).

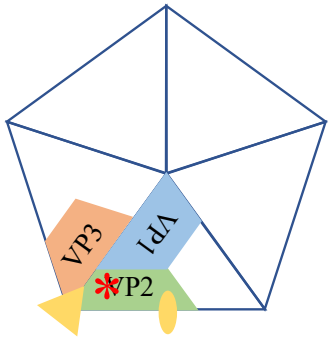
With this, the data presented offer no firm evidence of Fab #126 binding to O1M, as other explanations accounting for the additional observed diffuse density cannot be excluded at present. However, one may still speculate that Fab #126 may interact at this region, potentially interacting with the flexible GH loop of VP1 and accounting for the lack of local resolution.



(a) 0.195



(b) 0.0491



(c)

Figure 4.37: Inference of the Fab #126 epitope on O1M serotype VLPs. O1M VLP crystallographic structure (PDB ID: 5DDJ [183]) was fitted into the observed cryo-EM density map (transparent blue). Map threshold set to (4.37a) 0.195 to show fitting of 5DDJ, and (4.37b) 0.0491 to show additional external density (central blob). Triangle and oval indicates approximate locations of threefold and twofold symmetry axes respectively. (4.37c) Capsid schematic indicating proposed region of Fab #126 interaction with an asterisk. Cartoon chain colouring matches schematic; VP1 is blue, VP2 is green and VP3 is orange. Figure produced using ChimeraX [257].

4.6.7 Interaction of O1M VLPs by Fab #117

Data were also collected for Fab #117 mixed with O1M 93C VLPs, to test whether chain-exchange of the light chain alters the binding epitope, on the assumption that ultralong antibodies sharing an identical heavy chain would interact in a conserved manner (Section 4.6.6), contrary to the preliminary serospecificity findings (Table 4.4).

4.6.7.1 O1M-Fab #117 Data Collection

Specimen preparation constituting Fab #117 mixed with O1M VLPs (produced by vaccinia virus; Section 2.2.6) were prepared and data collected as previously indicated (Tables 4.10 and 4.11). A representative micrograph is provided in Fig. 4.38.

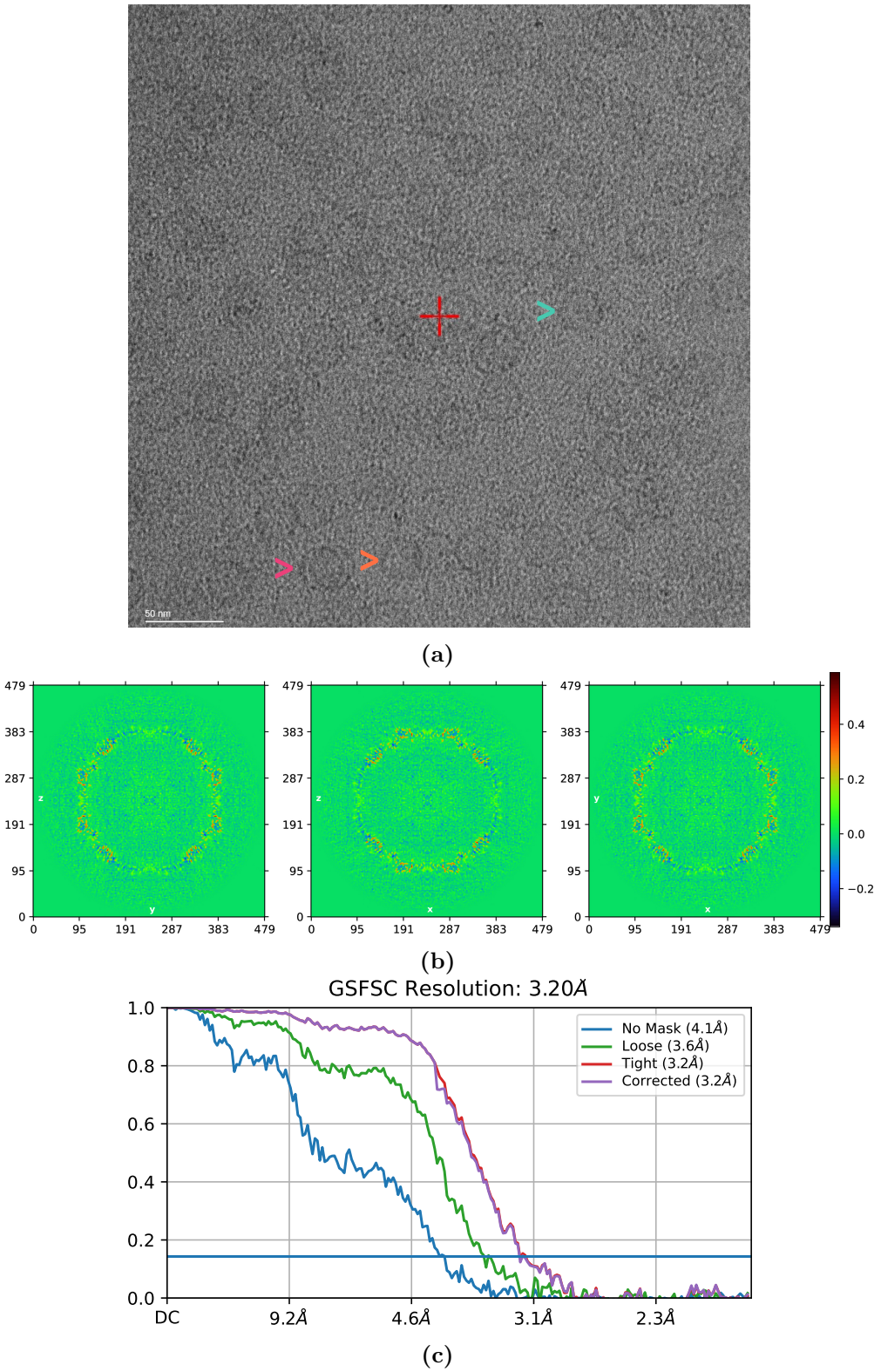


Figure 4.38: Electron cryomicroscopy of Fab #117 in complex with FMD serotype O1M VLPs. (4.38a) Example micrograph, with particles marked by arrowheads. Scale bar indicates 50 nm. O1M observed as distinct (green), or overlaid (orange and pink). (4.38b) Central slices from density maps after final round of refinement. (4.38c) Autotightened GSFSC plot indicating determined resolution. 0.143 FSC is marked by a horizontal line.

As with Fab #126 (Section 4.6.6.1), addition of Fab #117 to O1M VLPs produced 2D class averages showing additional periodic diffuse density to the external surface of O1M VLPs (Fig. 4.39).

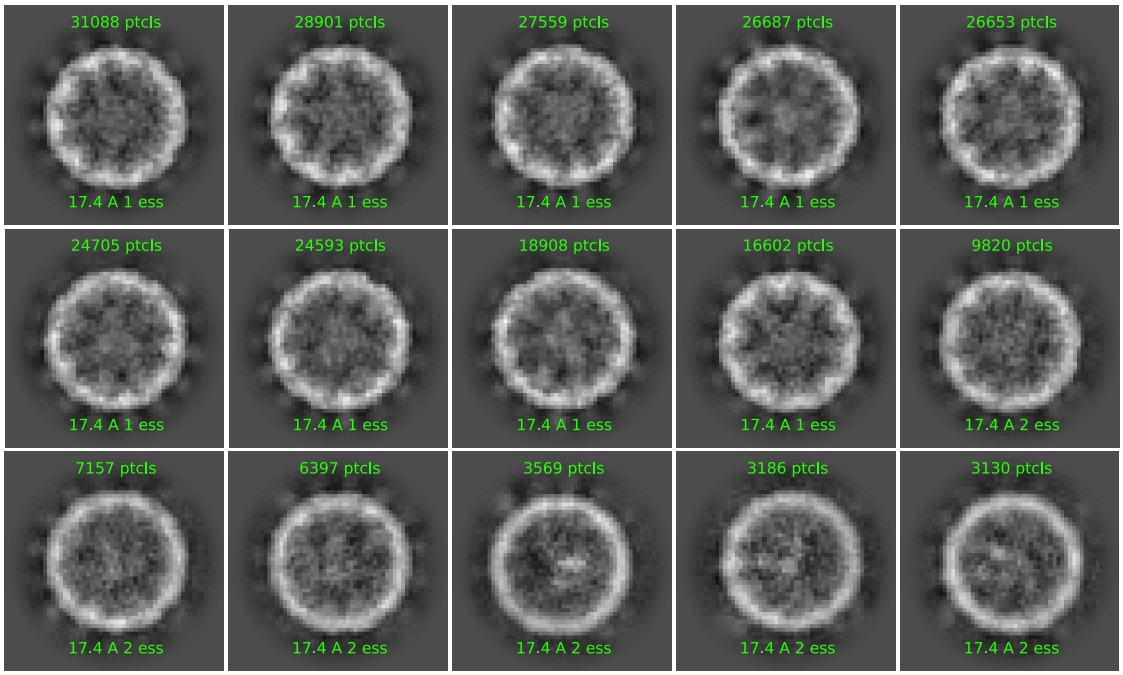


Figure 4.39: 2D class averages for O1M in complex with Fab #117. Highest populated 2D class averages with indicated population, estimated resolution and effective sample size (ESS).

In total, 267,890 particles were picked from 10,965 exposures (Table 4.11). Following particle curation (via iterative 2D classification and heterogeneous refinement) 1,378 particles were used in the final round of refinement. A 3.20 Å map was achieved following refinement (Fig. 4.38c), however the signal for the diffuse density appears to have become attenuated, with only very weak, low resolution evidence remaining of external decoration (Fig. 4.40).

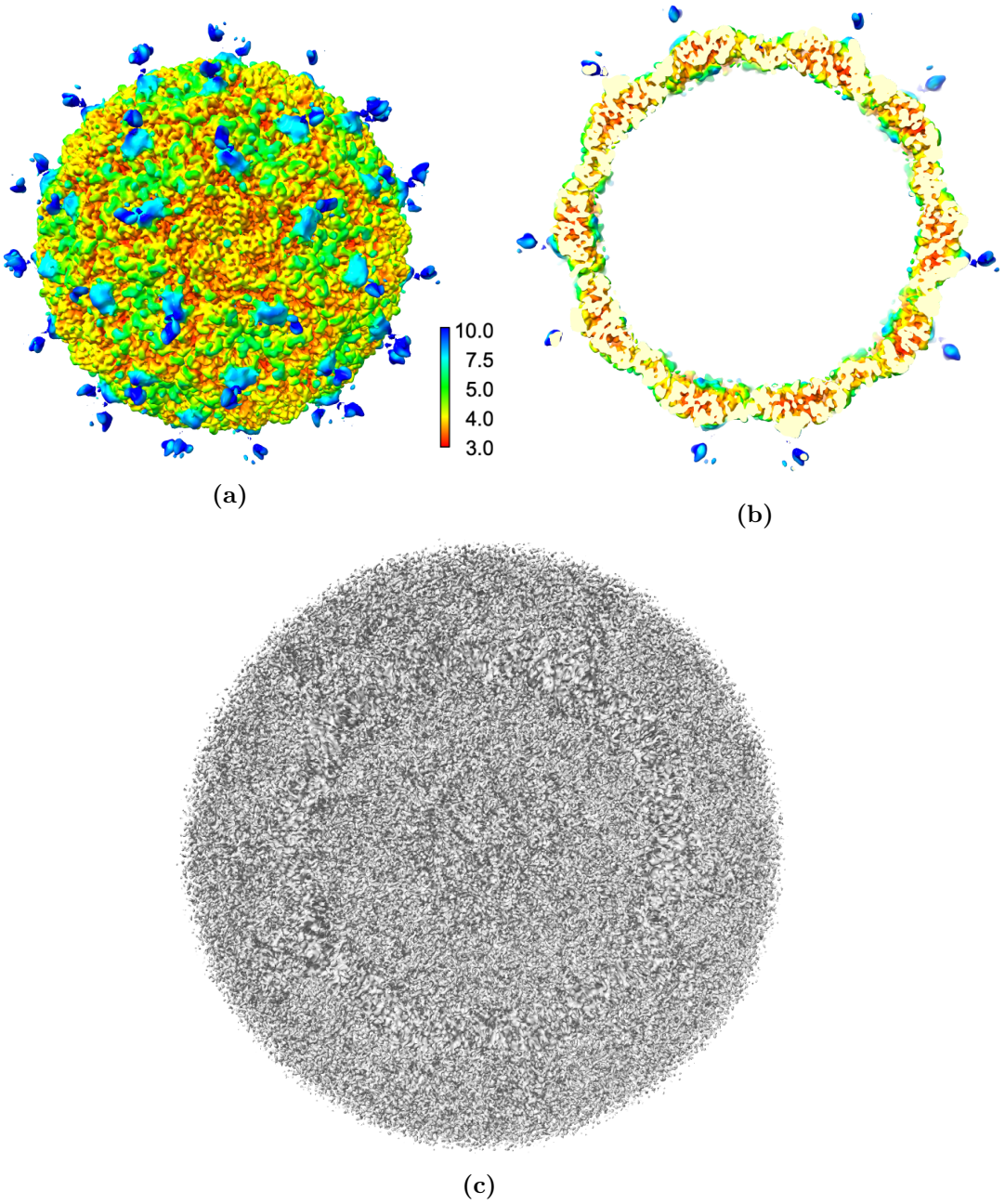


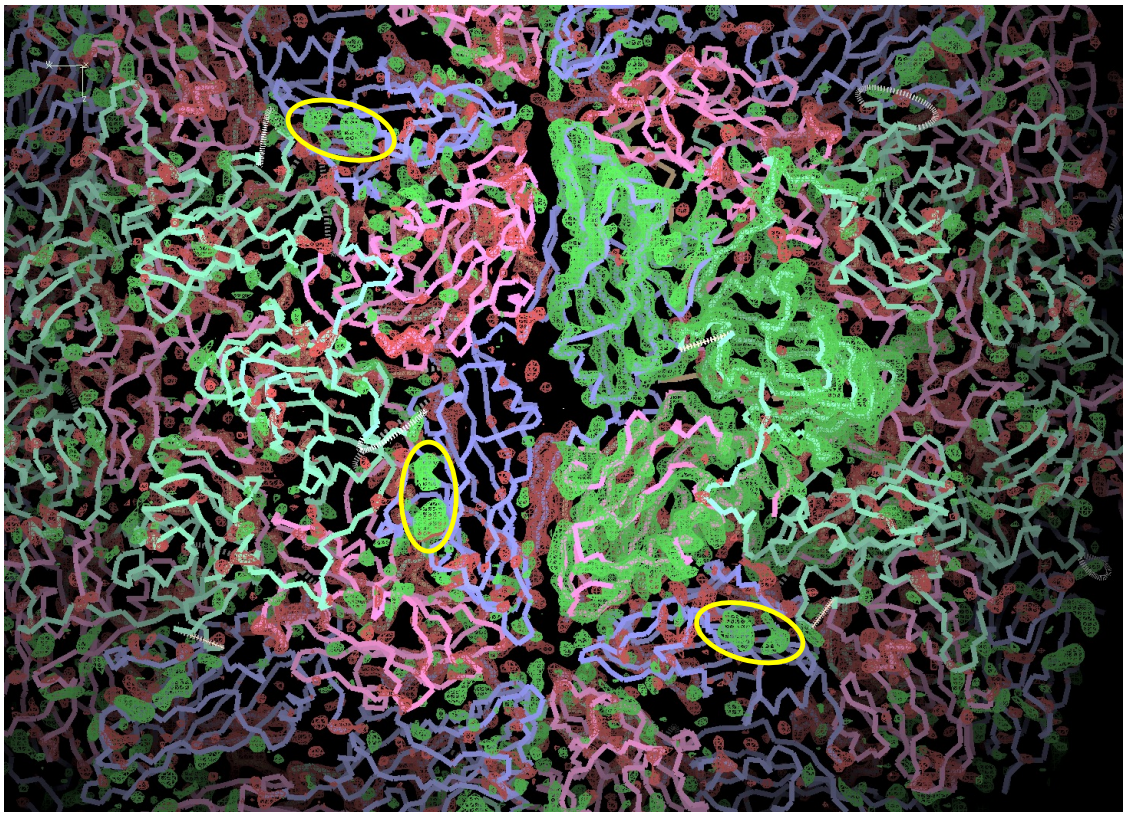
Figure 4.40: Observed O1M/Fab #117 cryo-EM density map coloured by determined local resolution. Locally filtered map (4.40a), and central slice thereof (Fig. 4.40b) reconstructed using 1,378 particles initially picked by template picking using cryoSPARC [249]. Map thresholded to 0.0251. Resolution scale bar in Å. (4.40c) Central slice through unfiltered map thresholded to 0.0251 for comparison. Figure produced using ChimeraX [257].

4.6.7.2 The O1M epitope accessed by Fab #117

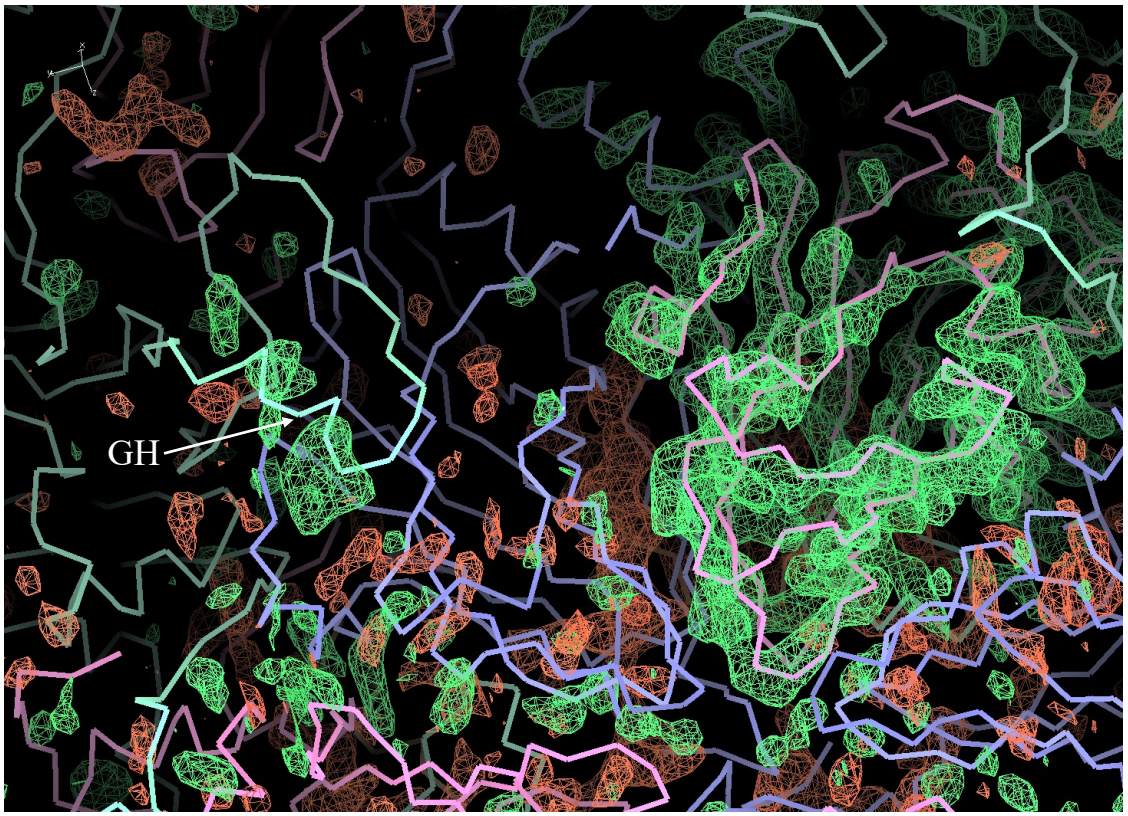
The density maps for O1M in complex with Fab #126 and #117 cannot simply be superimposed to assess whether the observed additional density, and hence approximate binding epitope, is similarly positioned about the capsid's exterior, as the superposition algorithm might align based on prominent features, i.e. the external density rather than aligning on the central featureless capsid density. To determine whether the external features were significant and above the global noise, maps were first aligned using a capsid-only map, generated from fitting the PDB capsid structure into the capsid density before calculation of difference maps.

Both maps for the #117 and #126 datasets were first aligned to the same O1M reference capsid structure (PDB ID: 5DDJ [183]), but it was observed that our cryo-EM capsids were slightly expanded relative to that of the reference capsid. While this could have been due to incorrect pixel size calibration, it has been shown previously that VP2 93C mutation can give rise to a slightly expanded reconstruction in cryo-EM [310].

The reference PDB capsid structure was adjusted to reflect the typical chain composition of VLP maps; the N termini of VP1 and VP2 were deleted (residues 1-26, and 1-54 respectively), as was VP4. This modified capsid model was rigid body fitted into the observed cryo-EM capsid density. One VLP protomer was omitted to act as a positive control for difference density calculation. The Chimera molmap command was used to generate a reference map from the expanded fitted model to a resolution of 3.5 Å. Difference maps were calculated using EMDA (Electron Microscopy Data Analysis) software [311].



(a)



(b)

Figure 4.41: Difference map showing density not explained by O1M crystal structure. *(Figure legend is continued on the next page...)*

Figure 4.41: Difference map showing density not explained by O1M crystal structure. *(Figure legend is continued from the previous page.)* Difference map calculated for O1M with Fab #117, as described in Section 4.6.7.2 and contoured to 7.50 RMSD. 1 protomer, right, was removed prior to difference map generation as a positive control. (4.41a) Top view showing positive and negative difference density in red and green mesh respectively. PDB 5DDJ [183] was fitted to the map in Chimera [257], and the difference map was inspected using COOT [228]. Additional, non-capsid density is circled in yellow. (4.41b) Close perspective of the negative non-capsid density, with fitting of PDB 1FOD [312]; a crystal structure with resolved GH loop in down conformation, represented in wire notation. GH-loop associated density is indicated.

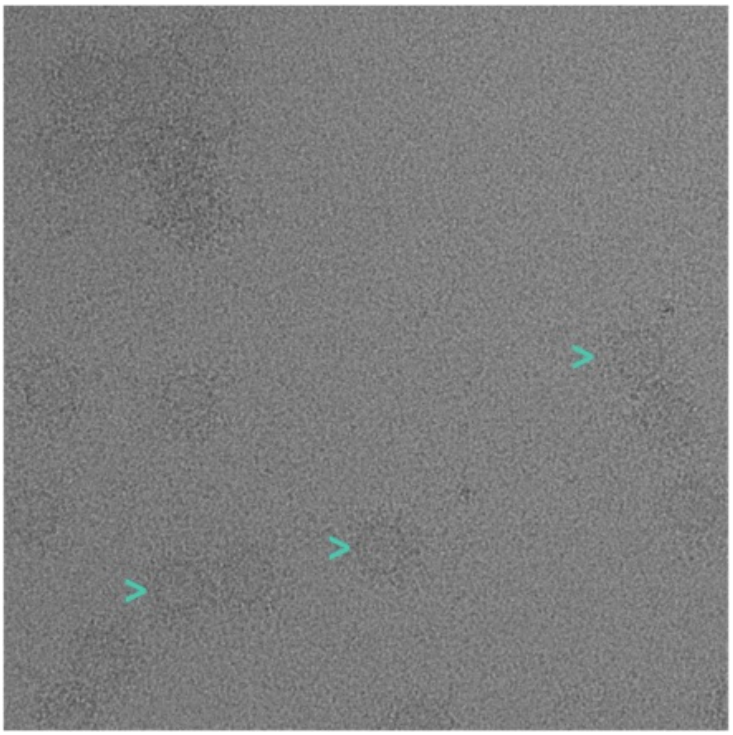
The calculated difference map is far from optimal, given the reference model was only rigid body fitted to the observed map, and hence will not account for variation in capsid expansion, or any sequence differences, introducing noise to the difference map. Regardless, difference density can still be seen between the twofold and threefold symmetry axes at a contour level only just above noise. This density is sufficiently diffuse that no confident conclusions may be drawn, however it is consistent with the down conformation VP1 GH loop from the aligned PDB 1FOD structure [312]. The additional density decorating O1M VLPs in 2D classifications may therefore be from a flattened VP1 GH loop, but no additional density that could be ascribed to the Fab could be seen. Further work is necessary to fully delineate this additional density.

4.6.8 Interaction of O1M VLPs by Fab #30

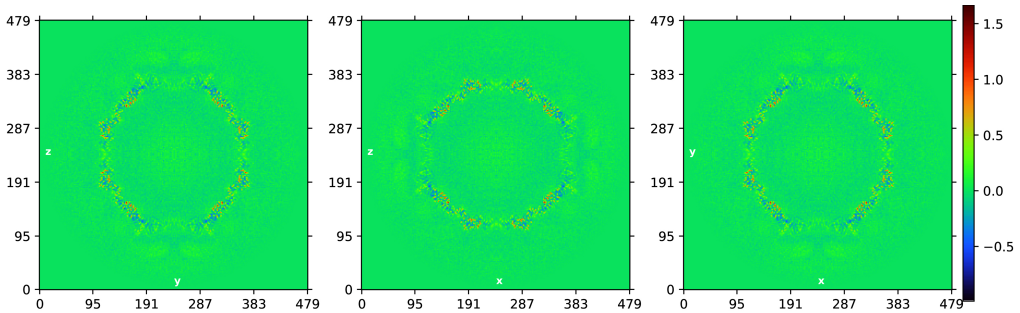
Fab #30 is an ultralong antibody mono-specific for FMD serotype O1M (Table 4.4). Subsequent analysis confirms a high affinity, neutralising function of IgG #30 (ITB, personal communication). It would therefore be interesting to compare against Fab #30 binding O1M as a positive control, and to resolve antigenic recognition between a bovine antibody and an antigen relevant to bovids.

4.6.8.1 O1M-Fab #30 Data Collection

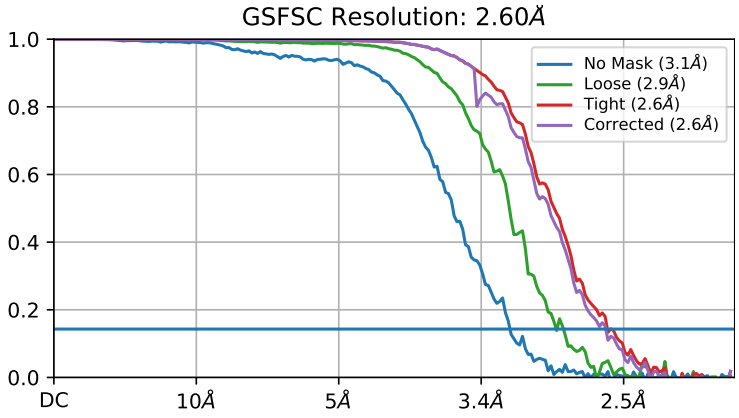
Specimen preparation constituting Fab #30 mixed with O1M VLPs (produced by vaccinia virus; Section 2.2.6) were prepared and data collected as previously indicated (Tables 4.10 and 4.11). A representative micrograph is provided in Fig. 4.42.



(a)



(b)



(c)

Figure 4.42: Electron cryomicroscopy of Fab #30 in complex with FMD serotype O1M VLPs. (4.42a) Example micrograph, with particles marked by green arrowheads. (4.42b) Central slices from density maps after final round of refinement. (4.42c) Autotightened GSFSC plot indicating determined resolution. 0.143 FSC is marked by a horizontal line.

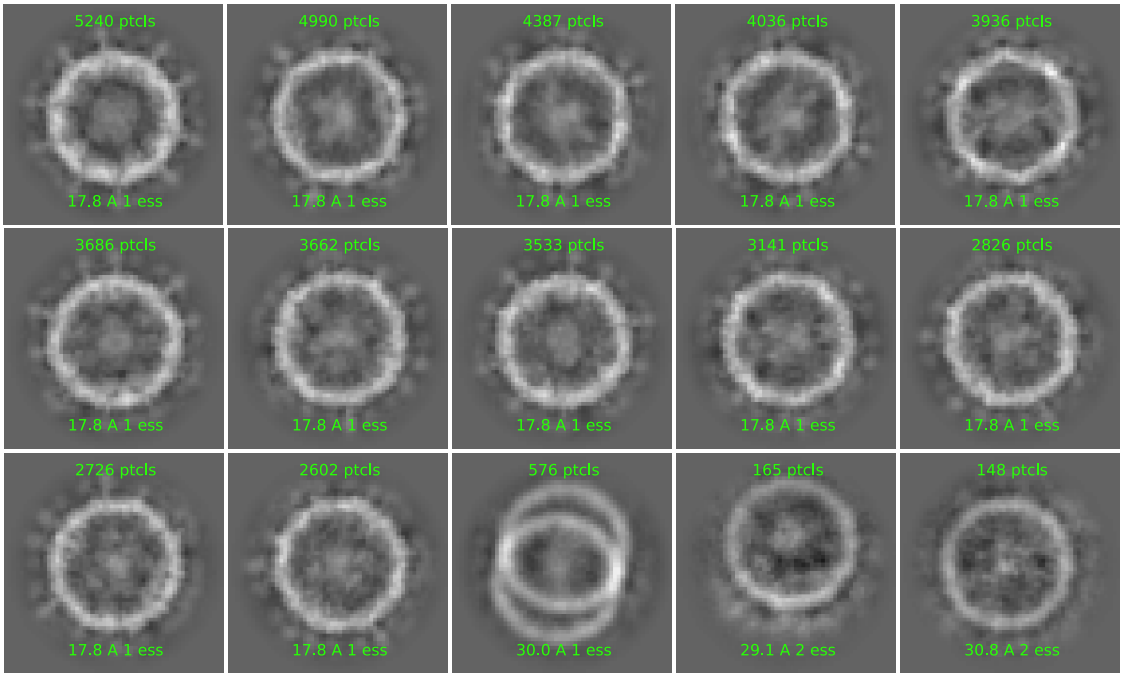


Figure 4.43: 2D class averages for O1M in complex with Fab #30. Highest populated 2D class averages with indicated population, estimated resolution and effective sample size (ESS).

2D class averages again presented decorated circular projections for O1M in complex with Fab #30 (Fig. 4.43).

4.6.8.2 The O1M epitope accessed by Fab #30

46,778 particles were initially selected from 4,170 exposures (Table 4.11), with 36,451 particles used in refinement of the final cryo-EM density map to a nominal 2.60 Å resolution (Fig. 4.44).

The epitope for Fab #30 on O1M is positionally similar to the additional density of the previous O1M unexplained densities (Sections 4.6.6.2 and 4.6.7.2) with additional density (white) being observed above VP2 (green), however for Fab #30, this additional external density is grouped much more closely to the threefold symmetry axis (Fig. 4.44).

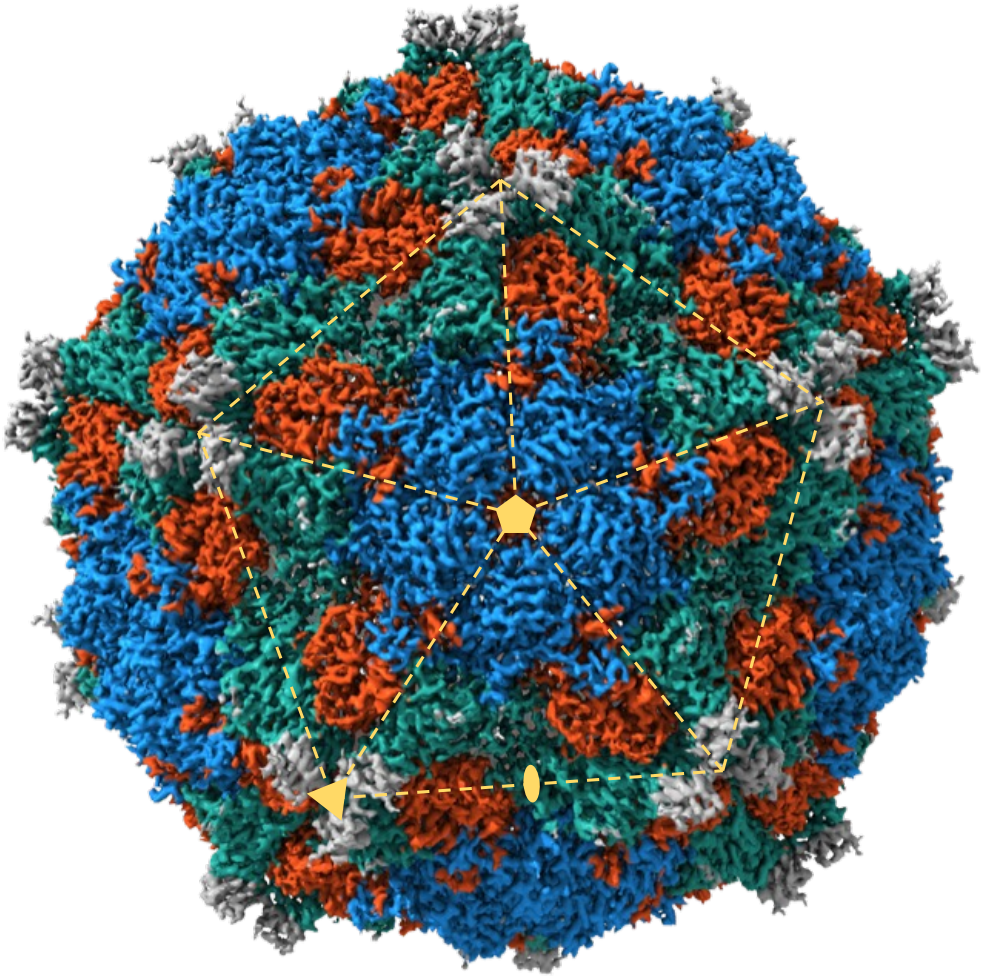


Figure 4.44: Fab #30 interacts with O1M VLPs at the 3-fold symmetry axis. Refined density map for O1M in complex with Fab #30. O1M crystal structure (PDB ID: 5DDJ [183]) was fitted, and used to colour the observed density by viral protein; blue for VP1, green for VP2 and orange for VP3. Unexplained density (white) indicates site of Fab #30 binding. Icosahedral symmetry axes are indicated in yellow shapes: twofold; oval, threefold; triangle, fivefold; pentagon. Figure prepared using ChimeraX [257].

The crystal structure of Fab #30 did not reach satisfactory refinement statistics to enable fitting into the cryo-EM density map within the time constraints of this project. As a result, Fab #30 was predicted by AF2, and the CDR3H knob mini-domains fitted into the map. As with Fab #117 (Section 4.6.3.2), the model with the highest CCC was retained (Fig. 4.45).

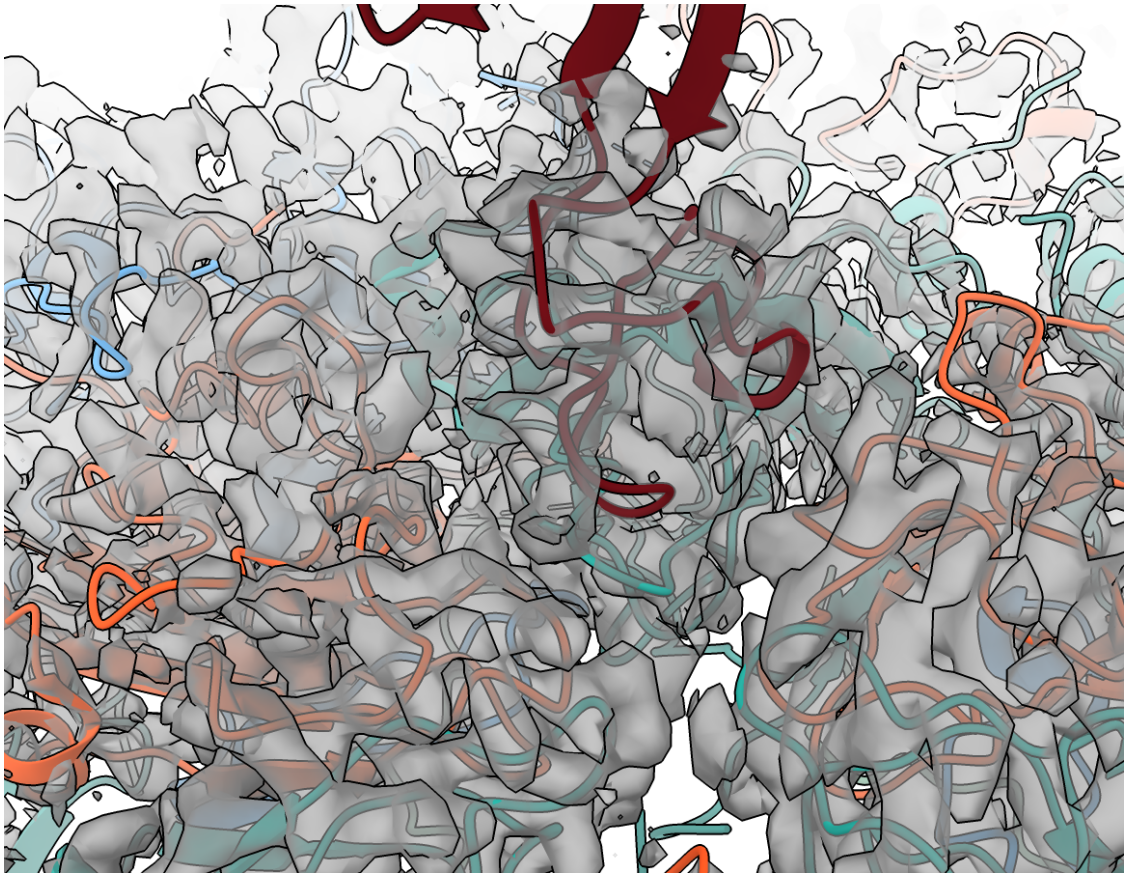


Figure 4.45: Fitting of O1M capsid and AF2 Fab #30 to observed cryo-EM density map. O1M capsid structure (PDB ID: 5DDJ [183]) and AF2 Fab #30 model (dark red) fitted using Chimera [257]. Observed density is transparent grey, VP1 is blue, VP2 is green, and VP3 is orange. Perspective from the centre of the threefold symmetry axis, looking down twofold axis.

The predicted Fab #30 model (dark red) globally fits the density at the threefold axis end of VP2 (green), and does not appear to make contact with either adjacent VP3, or Fab #30 knob mini-domains. The proximity of this epitope to its own symmetry-related copies on adjacent VP2s may introduce a steric clash to non-ultralong antibodies, but observation of diffuse density twisting away from the threefold axis may indicate some flexibility of the stalk domain to accommodate such situations. Further work is needed to assess the flexibility or rigidity of ultralong stalk domains. Model building of the Fab #30 interaction with O1M VP2 will be continued in future.

4.7 Quantifying ultralong antibody interactions with antigen

The disparity of predicted binding to O1M by antibody #126, yet lack of any specific interaction observed by cryo-EM warranted orthogonal verification of binding (or lack thereof). The Immunogenetics Group at Pirbright routinely uses biolayer interferometry (BLI) for quantitative assessment of antibody interaction with antigen, therefore a natural step would use BLI to quantify ultralong Fab-antigen interactions.

4.7.1 BLI Experiment optimisation

First, the BLI assay was optimised for studying interactions involving Fab, based on Pirbright's protocol for studying interactions with whole antibody (IgG) and O1M VLPs. Principally, three challenges had to be surmounted to improve the signal-to-noise ratio sufficient to enable measurement of antigen binding. (1) Fab purification buffer compatibility (Section D.2.1), (2) Fabs only produced a modest BLI response compared to IgG counterparts (Fig. D.5), which was compounded by (3) high non-specific binding (NSB), obscuring the Fab response (Section D.2.3). Optimisations were performed using O1M serotype VLPs and O1M mono-specific Fab #30 (Section D.2), following observations of strong association responses in previous BLI experiments performed by ITB [184] (personal communication).

4.7.2 BLI characterisation of Fab #30 with O1M VLPs

Encompassing the described optimisations (Section D.2), the interaction of Fab #30 with O1M VLPs was successfully characterised by BLI (Fig. 4.46). An R_{max} of 0.2415 ± 0.0003 nm was observed at 1500 nM Fab #30 in complex with 5 µg/mL O1M VLPs. The observed data were modelled using 1:1 Michaelis-Menten reaction kinetics (red traces; Fig. 4.46) with an overall R^2 of 0.9967 describing the fit of the model to the data, and giving an association rate constant (K_a) 1.875×10^5 ± 3.742×10^3 1/Ms, and a dissociation rate constant (K_{dis}) of 9.521×10^{-3} ±

4.842×10^{-5} 1/s, providing an overall equilibrium dissociation constant (K_D) of $5.078 \times 10^{-8} \pm 1.046 \times 10^{-9}$ M (Table 4.12).

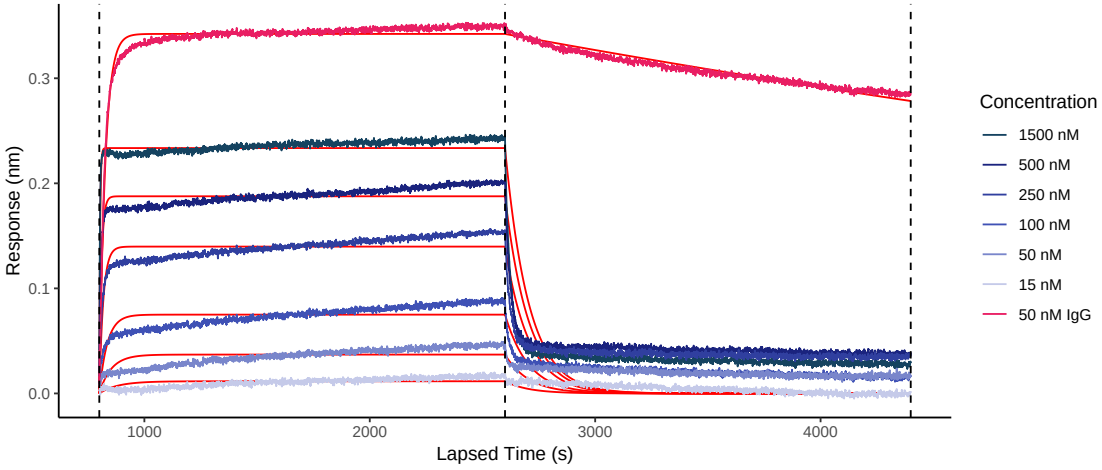


Figure 4.46: BLI characterisation of the Fab #30-O1M VLP interaction. Biotinylated O1M VLPs were immobilised to streptavidin biosensor probes (Sartorius) at 5 µg/mL for 650 s before being introduced to Fab #30, or IgG #30 at the concentrations shown. Both Fab #30 and IgG #30 prepared with Protein G re-purification and subsequent desalting step. 1:1 Michaelis-Menten kinetics modelled in red. After VLP loading, streptavidin biosensor probes were blocked for 30 s with 10 µg/mL biocytin to block non-specific binding. Association and dissociation steps were each performed for 1800 s. All assays run in HBS-EP buffer at 25 °C on an Octet R8 (Sartorius), as described in Section 2.7.11.

4.7.3 BLI characterisation of Fab #126 with O1M VLPs

The BLI assay, as optimised with Fab #30 (Fig. 4.46), was applied to characterise the predicted interaction between Fab #126 and O1M (Table 4.2).

An R_{max} of 0.0608 ± 0.0001 nm was observed with 1500 nM Fab #126 in complex with 5 µg/mL O1M VLPs. The observed data were modelled using 1:1 Michaelis-Menten reaction kinetics (red traces; Fig. 4.47a) with an overall R^2 of 0.8750 describing the fit of the model to the data, and giving a reported association rate constant (K_a) $1.424 \times 10^4 \pm 1.360 \times 10^2$ 1/Ms, and a dissociation rate constant (K_{dis}) of $8.429 \times 10^{-4} \pm 2.453 \times 10^{-6}$ 1/s, providing an overall reported equilibrium dissociation constant (K_D) of $5.919 \times 10^{-8} \pm 5.908 \times 10^{-10}$ M. The poor fitting of the observed data to a 1:1 model (R_{max} <0.95) is likely due to the observed dissociation curves reaching a value below the pre-association baseline. The poor fitting to a 1:1

Table 4.12: BLI characterisation of the Fab #30-O1M interaction results table. Observed data modelled using 1:1 Michaelis-Menten reaction kinetics by Octet® Analysis Studio v12.2.1.23 (Sartorius).

Conc. (nM)	R_{max} (nm)	R_{max} err. (nm)	R^2
1500	0.2415	0.0003	0.9269
500	0.2067	0.0005	0.9269
250	0.1682	0.0007	0.9269
100	0.1131	0.0009	0.9269
50	0.0745	0.0009	0.9269
15	0.0508	0.0012	0.9269

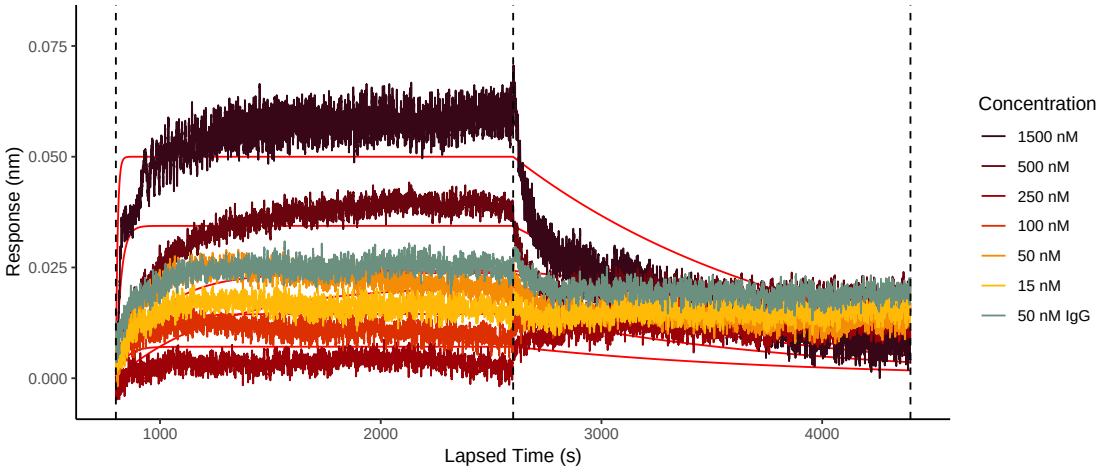
K_a (1/Ms)	K_a error (1/Ms)	K_{dis} (1/s)	K_{dis} error (1/s)	K_D (M)	K_D error (M)
1.875×10^5	3.742×10^3	9.521×10^{-3}	4.842×10^{-5}	5.078×10^{-8}	1.046×10^{-9}

model could be an artefact of the complex interaction with an FMD VLP with 60 potential binding sites. The closer fitting of observed data of Fab #30 to the 1:1 model may be explained by Fab #30 interacting at the 3-fold symmetry axis, and therefore steric clashes may prevent epitope saturation such that the ratio of bound Fab #30 to available O1M epitopes my be closer to 1:1 than with Fab #126.

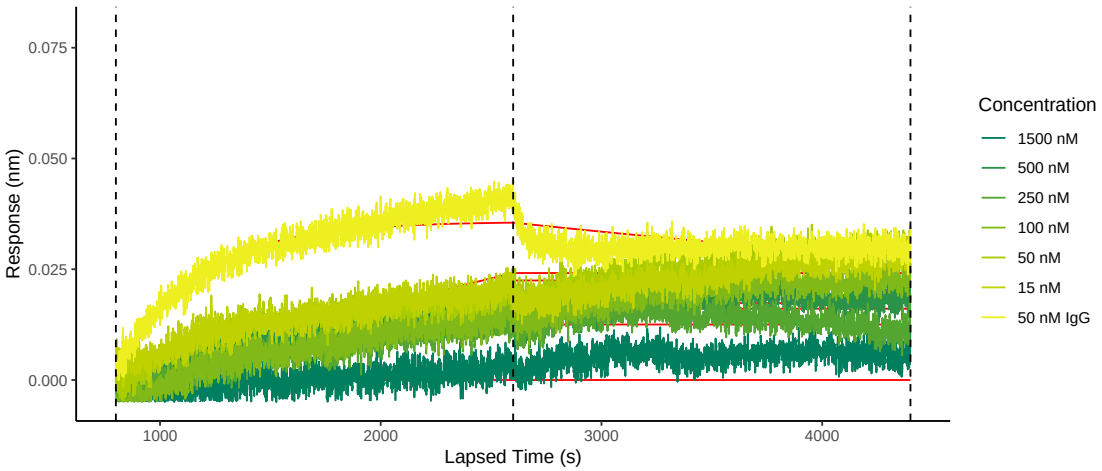
The low association response (<0.1 nm), and the poor fitting (<0.95), classifies Fab #126 as a non-binder of O1M, just as IgG #126 was classified as a non-binder (Section D.3) (ITB, personal communication). BLI is not best suited to characterising weak interactions, so the reported kinetic constants should be interpreted with extreme caution.

4.7.4 BLI characterisation of Fab #117 with O1M VLPs

As previously described, Fab #126 was predicted to recognise O1M in the AbCellera serospecificity assay, and gave a very weak binding response by BLI (Fig. 4.47a). On the other hand, Fab #117 was not predicted to bind O1M VLPs, despite sharing an identical ultralong heavy chain with Fab #126 (Table 4.2). Despite these different serospecificity predictions, cryo-EM could not provide any firm evidence of binding (Sections 4.6.6 and 4.6.7). BLI was used to investigate this observation further.



(a) Fab #126



(b) Fab #117

Figure 4.47: BLI characterisation of the O1M VLP interaction with Fab #126 (4.47a), and with Fab #117 (4.47b). Biotinylated O1M VLPs were immobilised to streptavidin biosensor probes (Sartorius) at 5 µg/mL for 650 s before being introduced to (4.47a) Fab #126 or (4.47b) Fab #117 at the concentrations varying 15 - 1500 nM. Both Fab #126 and Fab #117 was prepared with Protein G re-purification and subsequent desalting step. 1:1 Michaelis-Menten kinetics modelled in red. After VLP loading, streptavidin biosensor probes were blocked for 30 s with 10 µg/mL biocytin to block non-specific binding. Association and dissociation steps were each performed for 1800 s. All assays run in HBS-EP buffer at 25 °C on an Octet R8 (Sartorius), as described in Section 2.7.11.

Table 4.13: BLI characterisation of the Fab #126-O1M interaction results table. Note all sensograms produced an R_{max} of < 0.1 nm, and all reported kinetic constant values should be interpreted with extreme caution. Observed data modelled using 1:1 Michaelis-Menten reaction kinetics by Octet® Analysis Studio v12.2.1.23 (Sartorius).

Conc. (nM)	R_{max} (nm)	R_{max} err. (nm)	R^2
1500	0.0608	0.0001	0.8750
500	0.0508	0.0001	0.8750
250	0.0076	0.0001	0.8750
100	0.0108	0.0001	0.8750
50	0.0452	0.0003	0.8750
15	0.0096	0.0008	0.8750

K_a (1/Ms)	K_a error (1/Ms)	K_{dis} (1/s)	K_{dis} error (1/s)	K_D (M)	K_D error (M)
1.424×10^4	1.360×10^2	8.429×10^{-4}	2.453×10^{-6}	5.919×10^{-8}	5.908×10^{-10}

None of the Fab #117 concentrations produced an association response significantly greater than the baseline (R_{max} at 1500 nM was 0.0191 nm). This non-response is very similar to the negative control Fab #63 (Fig. 4.48), an ultralong antibody mono-reactive for SAT2 (ITB, personal communication), belonging to cluster 8 (Table 4.3).

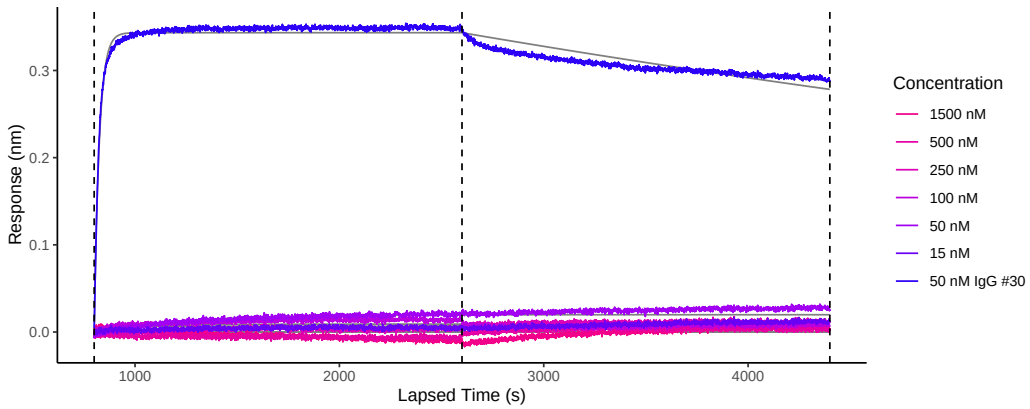


Figure 4.48: BLI characterisation of the Fab #63-O1M VLP interaction. Biotinylated O1M VLPs were immobilised to streptavidin biosensor probes (Sartorius) at 5 µg/mL for 650 s before being introduced to Fab #63, or IgG #30 (positive control) at the concentrations shown. Both Fab #63 and IgG #30 prepared with Protein G re-purification and subsequent desalting step. 1:1 Michaelis-Menten kinetics modelled in black. After VLP loading, streptavidin biosensor probes were blocked for 30 s with 10 µg/mL biocytin to block non-specific binding. Association and dissociation steps were each performed for 1800 s. All assays run in HBS-EP buffer at 25 °C on an Octet R8 (Sartorius), as described in Section 2.7.11.

Fab #126 appeared to demonstrate a concentration dependent series of sensogram amplitudes, albeit small, whereas Fab #117 produced no discernible amplitudes upon introduction of VLPs at any concentration (Fig. 4.47a), mirroring the response of irrelevant Fab #63 to O1M VLPs (Fig. 4.48). To confirm the validity of this observation, a technical replicate was performed. The Fab #126 replicate did not reproduce the concentration series of Fig. 4.47a, but sensograms did indicate association and dissociation events at all concentrations (Fig. 4.49), potentially explaining the differential predicted specificity's of antibody #117 and #126 as a function of fine affinity. Interestingly, this observation of a marginally stronger interaction by Fab #126 is corroborated by whole antibody BLI experiments which also show association and dissociation curves for IgG #126, but not for IgG #117 (ITB, personal communication) (Section D.3).

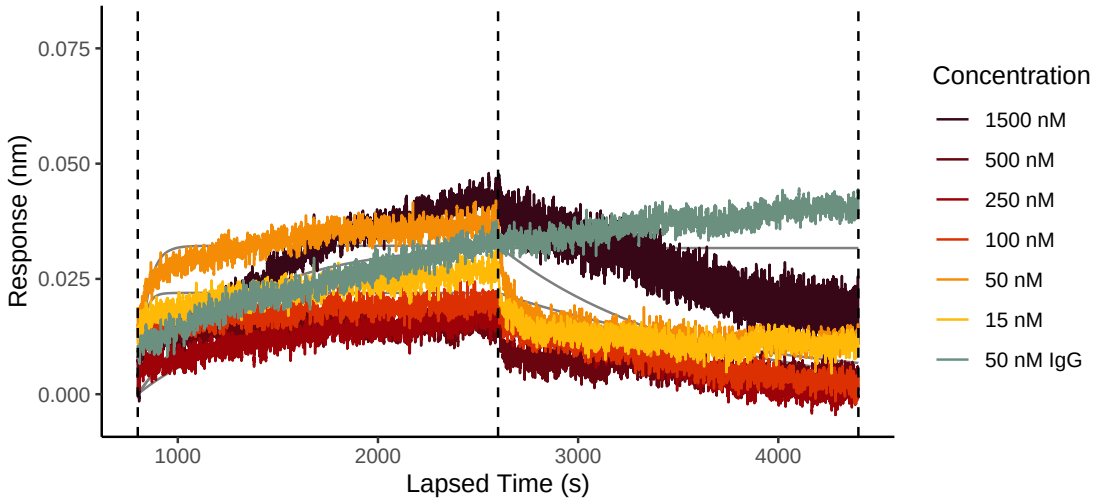


Figure 4.49: Repeated BLI characterisation of the Fab #126-O1M VLP interaction. Biotinylated O1M VLPs were immobilised to streptavidin biosensor probes (Sartorius) at 5 µg/mL for 650 s before being introduced to Fab #126, or IgG #126 at the concentrations shown. Both Fab #126 and IgG #126 prepared with Protein G re-purification and subsequent desalting step. 1:1 Michaelis-Menten kinetics modelled in red. After VLP loading, streptavidin biosensor probes were blocked for 30 s with 10 µg/mL biocytin to block non-specific binding. Association and dissociation steps were each performed for 1800 s. All assays run in HBS-EP buffer at 25 °C on an Octet R8 (Sartorius), as described in Section 2.7.11.

The reason for the slightly stronger attraction with Fab #126 is unclear, especially given the lack of a structure that includes the Fv region. Still, any difference between Fab #126 and #117 is too subtle to be accurately determined by BLI.

Overall, the response amplitudes generated by both Fab #117 and #126 are negligible, and, corroborating the cryo-EM results, provide no evidence for interaction between these Fabs and O1M serotype VLPs. The marginal signals infrequently observed between Fab #126 and O1M may be picking up interactions between a small population of dissociated O1M capsids, possibly binding at the crevice between VP2 and VP3 as observed with dissociated SAT2 pentamers. Sequence alignment of FMD VPs indeed indicate cross-serotype conservation of all residues participating in the interaction of Fab #117 binding SAT2, except Ile32 and Asp38 of VP2, which are substituted for chemically similar residues in O1M, though of course sequence conservation of interacting residues does not necessarily result in structural conservation of the binding epitope.

4.8 Discussion and future work

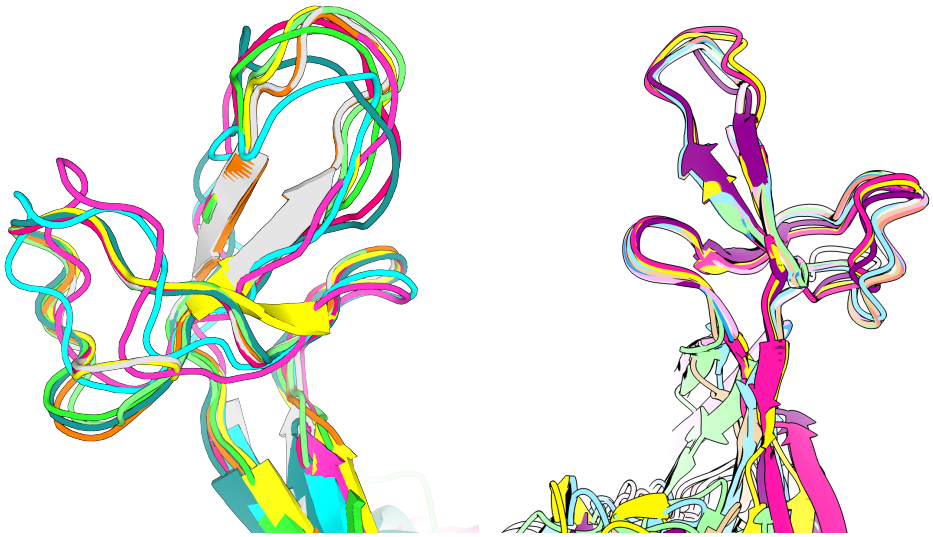
4.8.1 Systemic twist predicted by AlphaFold

In absence of crystal hits, AF2 was used to predict the structures for Fab #117 and #126. AF2 struggled to predict the ultralong CDR3H loops with great confidence, likely consequential of the scarcity of ultralong structures in the PDB [162], and the multiple cysteine options.

The PAE and pLDDT plots indicate that final models (except scFv #117 model #5) show some sensible chemistry and geometry within the CDR3H, but AF2 was less confident in the geometry of the CDR3H stalk region, resulting in knob regions modelled at a variety of positions relative to the Fv region (Figs. 4.12, 4.13, 4.15 and 4.16). Despite this, the ultralong CDR3H was systemically twisted between the scFv #117 and #126 models. In scFv #117 models, both the CDR3L and CDR3H are perturbed further away from the VH domain, and toward the VL domain, than in scFv #126 (Table 4.6).

Whilst the stalk region was modelled with varied conformations, modelling of the knob mini-domain converged on a single architecture for nine of the ten models, all of which were modelled with a three-pass β-strand at the core of the knob mini-domain (Fig. 4.17). Chapter 3 presented evidence that twisting the ultralong CDR3H stalk can subtly alter the presented paratope, and this may be sufficient to alter the fine affinity and possibly specificity of the antibody, which was investigated by cryo-EM and BLI experiments.

As a curiosity, AF2 predictions for Fab #117 and #126 were remodelled during the final stages of thesis construction, using the more accurate multimer mode. There was not enough time to fully compare the output models to those produced earlier in the project, but a side-by-side comparison shows a tighter convergence on the knob conformation across all models (Fig. 4.50), possibly a consequence of the multimer mode, or the increased availability of deposited ultralong antibody structures.



(a) **(b)**

Figure 4.50: Comparison of initial and revisited AF2 knob mini-domain predictions. (4.50a) Fig. 4.17a represented for comparison. AF2 run in single chain mode on 11th March 2022. (4.50b) AF2 run in multimer mode on 10th January 2023.

4.8.2 Antigenic interactions mediated by ultralong knob domain

To investigate whether the ultralong CDR3H perturbation was mediated distally by the light chain in the context of antigen, cryogenic electron microscopy experiments were prepared. Side chain resolution (approximately 3 Å) density maps were achieved for O1M mixed with Fab #126, #117, and #30, and for SAT2 12S pentamers in complex with Fab #117. No firm evidence of Fab binding O1M was recorded for Fabs #117 and #126. The density assigned to binding Fab was sufficiently diffuse to preclude reliable conclusions being drawn. Instead, evidence for the resolution of the VP1 GH loop was observed from the calculated difference maps.

Side chain resolution (nominal 2.89 Å) was also achieved for Fab #117 in complex with dissociated SAT2 pentamers. The Fab is locked in place via several interactions involving the Trp138 at the distal tip of the CDR3H loop, and binds a deep crevice on the capsid internal surface between VP2 and VP3, that extends down to VP1. An extensive hydrogen bonding network with VP2 also contributes to the binding of Fab #117 to SAT2 pentamers. The interacting residues from SAT2/Zim/7/83 are largely conserved in O1M/Tur/69, and may therefore be able to bind #117 if disrupted.

Fab #30 recognises the threefold end of O1M VP2 without appearing to make contact with other protomers or threefold symmetry copies of Fab #30, though completion of model building and PISA analysis will confirm the extent of interactions made by Fab #30 with O1M.

4.8.3 Fab BLI behaviour

When compared to whole antibody studies, Fabs provide only a modest response to antigen in BLI experiments, requiring greater concentrations to achieve comparable response amplitudes. This may be a facet of the BLI technique, where the interference between incident and reflected light is measured. The interference is produced by changes in the path length, corresponding to bimolecular interaction at the sensor tip. Fabs are approximately one third of the size of an IgG molecule, and so it follows that larger analytes would interfere with light to a greater extent, thereby

producing a larger BLI signal. Furthermore, Fabs do not benefit from the avidity advantage of whole antibodies, and consequently appear to show slower on-rates, and faster off-rates; indicating ultralong Fab binding appears to be more transient than ultralong whole antibody, as well demonstrated by IgG and Fab #30 (Fig. 4.46).

It has previously been implied that different antibody CH regions can alter the secondary structure of the antibody paratope, with consequent effects on fine specificity [313–318]. It is reasonable then, to suggest that complete loss of the Fc region (as with Fab fragments) would also likely influence antigenic interaction through destabilisation and could therefore account for the observed difference in BLI response magnitudes between Fab and IgG experiments with identical Fv regions.

4.8.4 Quantifying the serospecificity of chain-exchanged Fabs #117 and #126

The prediction (Table 4.4) of differing serospecificity for antibodies sharing an identical ultralong heavy chain is most intriguing, and warranted further study. This prediction should be interpreted with caution, given the assay has not been verified, and may be subject to false positives for one, or more, FMDV serotypes. Intuitively, given ultralong CDR3H principally interacts with antigen [78, 84, 85, 132, 138, 139, 319], antibodies that share an identical ultralong CDR3H should bind antigen in a conserved, or near conserved manner. On the other hand, the fine specificity of these interactions, as measured by BLI, indicates a subtle long range influence of the light chain to the function of antigen-binding ultralong CDR3H. Given long range effects can be sufficient to abolish interaction with antigen [318], there may be grounds to suggest specific light chains paired with ultralong antibodies may be able to demonstrate similar abrogation.

Despite cryo-EM observations, neither Fab #117 nor #126 achieved the arbitrary BLI association response threshold of 0.1 nm to be deemed as true binders of O1M VLPs. Indeed the lack of well-resolved Fab density in the O1M datasets corroborates this fact by presenting a transient interaction with incomplete occupancy.

4.8.5 Future work

Clearly a significant amount of further processing is required to fully define whether light chain sequence impacts the binding epitope accessed by ultralong antibodies. Preliminary experiments presented in this thesis aimed to characterise the structural and functional implications of chain exchange on antigen-specific antibodies. However the achieved resolutions, constrained by diffuse density, do not permit such conclusions to be drawn.

It would be interesting to assess the thermostability of different chain-exchanged ultralong antibodies, such as #117 and #126, given the importance of antibody stability on affinity and specificity [320]. The proximity of the CDR3L loop to the CDR3H, at the interface between the heavy and light chains, may impact the enthalpy of immunoglobulin domain dissociation, which could inform on the observed pairing preference associated with ultralong antibodies.

Furthermore, AF2 prediction of all identified ultralong antibodies in this project would be interesting to compare against the predicted and orthogonally assessed serospecificity to observe for any commonly predicted features, with the caveats already discussed. Similarly, the multimer mode could be used to predict binding epitopes, akin to molecular docking or dynamics simulations.

In parallel with the functional characterisation of Fabs #117 and #126, the Immunogenetics group at Pirbright has been performing orthogonal characterisation assays of IgG #117 and #126, amongst others, using BLI, ELISA, and virus neutralisation titre assay (VNT). A multitude of alternatives exist for characterising bimolecular interactions, however a more obvious experiment to do, would be to measure the kinetics of these Fabs interacting with dissociated SAT2 12S pentamers, particularly as Fab #117 demonstrates a specific interaction with SAT2 12S (Table 4.2).

Equally, synthetic chain exchange, or targeted mutagenesis to the light chain, could reveal individual contributions of CDR3L residues to the conformation, and potentially specificity, of the ultralong CDR3H of bovine antibodies. Indeed seven other ultralong, and one non-ultralong bovine antibody, identified from the serial

vaccination experiment, were worked on over the course of this project, and a wealth of structural data has been collected for these; seven additional XRD datasets and one additional cryo-EM dataset. Analysis of these data, and subsequent cryo-EM, XRD, and BLI experimentation may yield further revelations linking immunoglobulin chain pairing to the structure and function of ultralong antibodies.

What's the use in doing all this work, if we don't get some fun out of this?

— Rosalind Franklin, crystallographer

5

Preliminary work towards structural characterisation of chain-exchanged antibodies with the F protein of bovine respiratory syncytial virus

Contents

5.1 Introduction to bRSV 237
5.1.1 Importance and epidemiology of bRSV 237
5.1.2 Management of bRSV 238
5.1.3 The bovine respiratory syncytial virus 238
5.1.4 Vaccination against bRSV 239
5.1.5 Observations of pairing-mediated differential binding . . 240
5.1.6 X-Ray diffraction reveals alternative CDR3H conformations 241
5.2 Transient transfection of anti-bRSVF Fabs 242
5.2.1 Biosynthesis of anti-bRSVF Fabs by transient transfection 242
5.2.2 Affinity chromatography purification of anti-bRSVF Fabs 243
5.3 Expression of bRSVF for structural study 245
5.3.1 Transient transfection trial 245
5.3.2 Establishment of a stable cell lineage for bRSVF production 246
5.3.3 Purification of bRSV F protein 248
5.4 bRSVF Biophysical Characterisation 249
5.4.1 Characterisation by SEC-MALS 249
5.4.2 Verification of Fab binding by AUC 253
5.5 Structural biology 260
5.5.1 bRSVF/B4 complexes demonstrate preferential orientation 260
5.5.2 Resolution of bRSVF/B4 preferential orientation 263
5.5.3 Improving particle density 266

5.5.4 Repurification of F protein complexed with Fab B4 . . . 270
5.5.5 High resolution bRSVF/B4 data collection 280
5.6 Conclusions and Discussion 285
5.6.1 Recalcitrant protein purification 285
5.6.2 bRSVF exists in a dynamic concentration-dependent equilibrium 285
5.6.3 Variable degrees of Fab binding 286
5.6.4 Opportunity for structural determination of bRSVF in complex with other Fabs 287

Figures

5.1 Crystal structures for Fabs B4, B4*, B13 and B13* 242
5.2 Purification of Fab B4 244
5.3 Purification of Fab B13 244
5.4 Purification of Fab B13* 245
5.5 PCR Amplification of bRSVF insert from the pVRC8400 vector 246
5.6 Anti-His immunoblot of bRSVF stable cell lines 247
5.7 Anti-His immunoblot of bRSVF line #29 expression longevity . 247
5.8 Purification of bRSVF 248
5.9 SEC-MALS characterisation of purified bRSVF 250
5.10 Predicted N-glycosylation potential for bRSVF monomers calculated by NetNGlyc 252
5.11 Representative analytical ultracentrifuge sedimentation velocity scans 254
5.12 Analytical ultracentrifuge sedimentation coefficient characterisations of lone proteins 255
5.13 Analytical ultracentrifuge sedimentation coefficient characterisations of complexed bRSVF 257
5.14 Electron cryomicroscopy of Fab B4 in complex with bRSVF, prepared by Vitrobot 261
5.15 Electron cryomicroscopy of bRSVF/Fab B4 complex, prepared by Chameleon 264
5.16 *Ab-initio* 3D reconstruction of bRSVF/B4 complex, with overlaid observed sample pose distribution 265
5.17 Electron cryomicroscopy of bRSVF/Fab B4 complex; graphene-oxide coated grids prepared by Chameleon 268
5.18 Electron cryomicroscopy of bRSVF/Fab B4 complex; graphene-oxide coated grids with reduced Fab excess prepared by Chameleon 269
5.19 Size exclusion chromatography of bRSVF/B4 complex 270
5.20 SEC-MALS characterisation of purified bRSVF/B4 complex . . 273
5.21 Electron cryomicroscopy of bRSVF/B4 complex fraction 1, prepared by Chameleon 277
5.22 Electron cryomicroscopy of bRSVF/B4 complex fraction 2, prepared by Chameleon 278
5.23 High resolution electron cryomicroscopy of bRSVF/B4 complex fraction 2, prepared by Chameleon 281

5.24 B4 and bRSVF crystal structures fitted to cryo-EM density map 282
5.25 Interacting residues of B4 with bRSVF 284

Tables

5.1 Synthetic chain-exchanged Fab purification yields 249
5.2 SEC-MALS characterisation of bRSVF 251
5.3 Sedimentation coefficient results and estimated molecular weights for lone proteins 256
5.4 Sedimentation coefficient results and estimated molecular weights for bRSVF complexes 258
5.5 Predicted species concentrations for bRSVF/Fab complexes . . 259
5.6 Inferred species molecular weight by SEC elution volume 271
5.7 Summary of bRSVF/B4 complex spin-concentration 272
5.8 SEC-MALS characterisation of bRSVF/B4 complex fraction #1 275
5.9 High resolution bRSVF/B4 cryo-EM exposure curation 280
5.10 High resolution bRSVF/B4 cryo-EM particle picking 280

5.1 Introduction to bRSV

Chapter 1 introduced a study by Ren and colleagues [160], suggesting that chain-exchange was sufficient to drastically effect the antigenic recognition of the F fusion protein of bovine respiratory syncytial virus (bRSV) by non-ultralong bovine antibodies. This chapter will extend on this work, in an effort to resolve the idiotypic differences between parental and hybrid chain-exchanged non-ultralong antibodies.

5.1.1 Importance and epidemiology of bRSV

Bovine respiratory syncytial virus is a contagion that principally infects ruminants [321, 322], and is one of the most important pathogens associated with bovine respiratory disease (BRD) complex [323]. Symptoms of BRD include shortness of breath, increased breathing rate, spontaneous dry cough, and fever, as well as increased mucous production, decreased appetite, apparent depression, and difficulty drinking [324]. Upon infection, animals can become more susceptible to subsequent co-infection; indeed BRD is commonly recognised as a polymicrobic disease [325].

As with FMDV (Section 4.1.1), bRSV is transmitted upon direct contact with an infected animal or contaminated surface such as fomite, or aerosol transmission. The virus is present worldwide [326], likely consequential of global cattle movement, rather than via insect vectors [327].

Approximately 1.9 million calves are affected by bRSV each year in the UK, at an annual cost of £54 million due to loss of production, reduced carcass value, treatment and prevention costs [328]. Globally, annual losses attributable to bRSV approach US$1 billion [329]. Bovine RSV is genetically, clinically, and antigenically similar to human RSV [324, 330], and therefore offers a model for therapeutic development where RSV is only semi-permissive in typical model organisms [324]. At present, neither vaccine or anti-viral is routinely available for human RSV infection, though recent GSK clinical trials have shown promise. Palivizumab is available for at-risk patients as passive prophylaxis, whilst Ribavirin has had controversial successes, owing to underpowered studies, but also appeared slightly teratogenic, and treatment is expensive [331].

5.1.2 Management of bRSV

There is no specific treatment for bRSV. Cattle symptomatic for BRD are typically non-specifically treated, owing to the complexity of the condition. Mass treatment by herd-level prescription of prophylaxis and/or anti-microbials is common, with obvious implications for increasing anti-microbial resistance. Medicinal interventions have improved recently, but morbidity and mortality of BRD has not improved, warranting questions about the efficacy of such management strategies [332].

5.1.3 The bovine respiratory syncytial virus

Bovine RSV belongs to the *Pneumovirus* genus of the *Paramyxoviridae* family [333]. The bRSV genome is a 15 kb non-segmented, negative sense, single stranded RNA chromosome, encoding three surface glycoproteins; glycoprotein G, the F fusion protein and the small hydrophobic protein SH, as well as two viral nucleocapsid

proteins, N and P, a viral RNA-dependent polymerase, L, and four additional non-structural proteins [333].

5.1.4 Vaccination against bRSV

Viral vaccines for bRSV have shown mixed responses, where live, attenuated vaccines can exacerbate BRD if administered during live bRSV infection [334], whilst inactivated vaccines may also enhance BRD [335, 336]. Targeting RSV proteins therefore avoids these risks, whilst adding benefits of a targeted immune response, as well as being safer and cheaper to produce [175].

The prefusion conformation F fusion protein of bRSV is an attractive choice for vaccination [337], antibody therapy [338, 339] and small molecule antiviral agents [337].

Whilst bRSV displays antigenic variation between isolates [340–343], there is strong conservation of F protein sequences (up to, and exceeding 90 % identity) between bRSV isolates [344] even when belonging to different antigenic subgroups [342]. The conservation of F may be explained by its essential role in membrane fusion of the enveloped virion to host cell membranes. Furthermore, the most potent neutralising antibodies against RSV identified thus far target prefusion F protein [345, 346]. Finally, F is presented on the virus surface, allowing access to therapeutics.

Three translated F gene products ('F0') self-assemble into a homotrimer prior to host cell furin-like proteolytic cleavage within the Golgi apparatus [347, 348]. Each F0 monomer is processed to two covalently linked subunits; F1 and F2 [349–352], forming a single F protein monomer. At the virion surface, processed bRSV F homotrimers are metastable in the prefusion conformation. In the prefusion conformation, the hydrophobic N-terminus of the F1 subunits (involved in host cell membrane insertion) are buried within the central trimer cavity [353]. Following an unknown trigger event (thought to be related to osmolarity or virus-cell proximity), prefusion F protein homotrimer dramatically refolds into the postfusion conformation, where the hydrophobic F1 N-termini are inserted into a host cell

membrane, and the collapse of heptad repeats to a 6-helix bundle brings the membranes sufficiently close for fusion to occur [354]. The triggered, and hence non-spontaneous, structural reconfiguration of bRSV F explains the 'metastable' label for the prefusion state.

The Kwong group recently engineered bRSV F protein (herein bRSVF) to stabilise the prefusion conformation, eliminating the spontaneous shift to postfusion conformation [175]. Engineered bRSVF (PDB: 5TDL [175]) is structurally analogous to other stabilised prefusion conformation RSV F proteins (PDB: 4JHW [353]; 4MMQ, 4MMR, 4MMS, 4MMT, 4MMU, 4MMV [355]) elicits high-titer neutralising responses in calves without exacerbating disease, and are tolerant of osmolarity, pH and temperature [175]. The engineered DS2-v1 variant of bRSVF [175] is used in this chapter for structural study.

5.1.5 Observations of pairing-mediated differential binding

Ren and colleagues [160] tested the suggestion that in cattle immune responses, the limited genetic diversity of the Vλ repertoire makes a reduced contribution to antigen recognition compared with the heavy chain (Sections 1.3.1 and 1.3.2).

The previously reported antibodies B4 and B13 [161], specific for bRSVF, were used in this study. Immunological approaches have determined antibody B4 targets a single F monomer, recognising it at a pair of α helices composed of residues 255-275 of the F1 subunit [356, 357], together described as antigenic site II [358]. On the other hand, B13 binds to bRSVF between antigenic sites IV and V [358] on adjacent F monomers [359]. B4 and B13 were produced as recombinant Fabs, as were chain-exchanged variants, named 'B4*' and 'B13*'; comprising the heavy chains of B4 and B13, synthetically paired with the B13 and B4 light chains respectively. The interaction between these recombinant Fabs and bRSVF was examined by both ELISA and SPR.

Fab B4 bound bRSVF with an apparent K_D of 86 nM, whilst an accurate K_D of the chain-exchanged B4* Fab could not be determined as bRSVF saturation was

not achieved at the maximal tested concentration (5 µM), but the trend suggests a binding affinity in the low micromolar range.

Fab B13 bound bRSVF with an apparent K_D of 125 nM, whilst B13* demonstrated a roughly two-fold drop in affinity to bRSVF of 234 nM. To explain these observations, X-ray crystallography was pursued to investigate the CDR structures of these Fabs.

5.1.6 X-Ray diffraction reveals alternative CDR3H conformations

Fab B4 crystallised in spacegroup P 2_1 2_1 2, and diffracted to 1.92 Å, whilst Fabs B4* and B13 both crystallised in spacegroup P 2_1, and diffracted to 2.15 Å and 2.12 Å respectively. Fab B13* crystallised in P 1, and diffracted to 2.62 Å [160].

The CDR's of Fabs B13 and B13* were very similar, with a small deflection of 2.50 Å being introduced to the B13* CDR3H due to Tyr49 in the B4 light chain, but the B13* CDR3H is otherwise identical to that of B13 (Fig. 5.1). Additionally, a small difference was observed to the light chain interface with the B13 light chain residues 47-52 being buried deeper than the same residues of the B4 light chain in the B13* structure.

By contrast, the B4 CDR3H conformation is more severely impacted by chain-exchange to the B13 light chain. In the B4 Fab, the CDR3H takes a clockwise orientation (from the antigen perspective) due to the close packing of Tyr100C (Kabat numbering) with Trp47, also of the B4 heavy chain (Fig. 5.1). In the Fab B4* structure, Tyr100C makes a π-stacking interaction with Phe31 of the B13 CDR1L, with the remainder of the B4 CDR3H forced between the B13 light chain and B4 heavy chain in an anti-clockwise orientation (when viewed from the antigen perspective)(Fig. 5.1).

These observations of disrupted CDR3H support the biophysical characterisations of their interaction with bRSVF, but leave the question of the molecular mechanism unanswered. This chapter will present preliminary work towards structural determination of the bRSV F protein bound by these Fabs

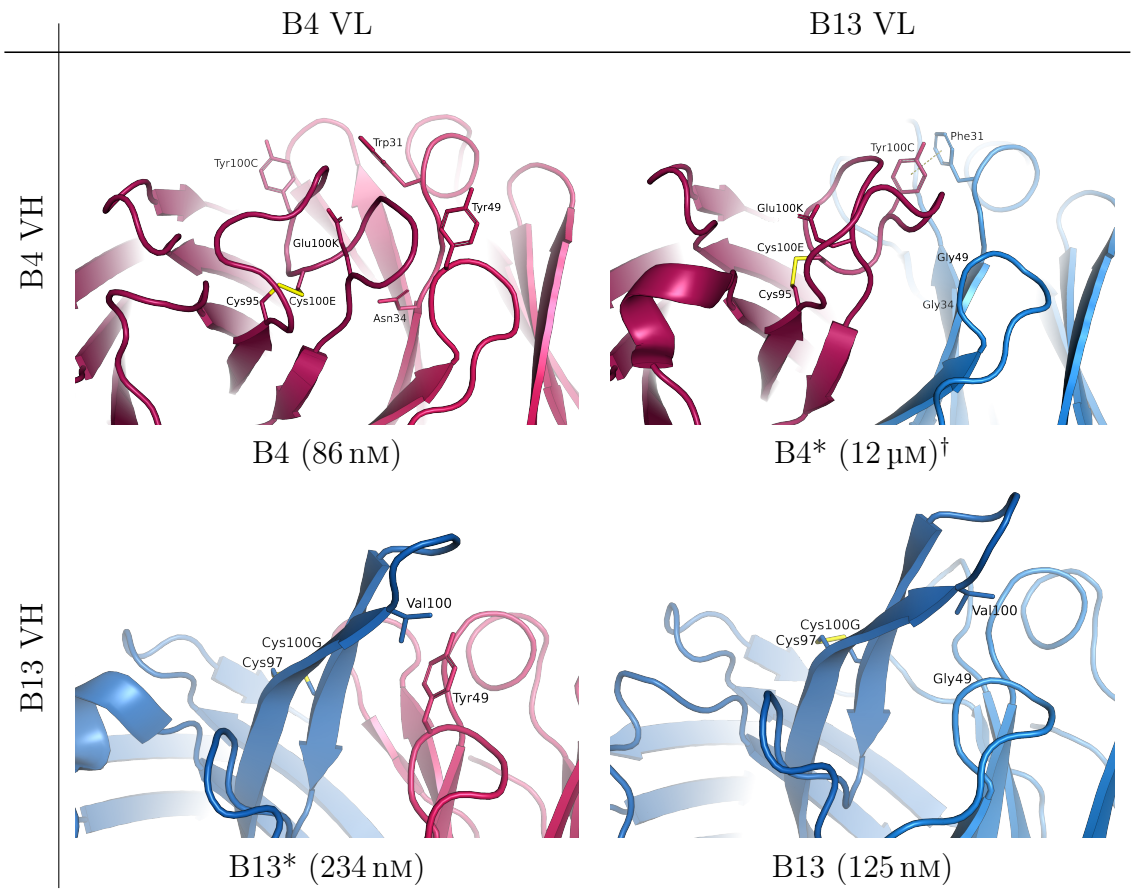


Figure 5.1: Crystal structures for Fabs B4, B4*, B13 and B13* (in clockwise order from the top left position) [160]**.** Residues contributing to CDR3H (centre) are highlighted in stick notation. Residues numbered following Kabat numbering scheme. Reported affinities for bRSVF from SPR. †Estimated affinity for B13*.

first, to provide the molecular mechanism behind recognition of the epitopes initially identified through escape mutagenesis, trypsin digestion and peptide synthesis [356, 359], and by competitive binding assay [357]; and further, to determine how these B4 and B13 epitopes are affected when accessed by the chain-exchanged hybrid Fabs B13* (and B4*).

5.2 Transient transfection of anti-bRSVF Fabs

5.2.1 Biosynthesis of anti-bRSVF Fabs by transient transfection

The heavy and light chain Fab constructs for antibodies B4 and B13, used to express Fabs B4, B13, and B13* [160] were kindly donated by Ray Owens. No further

optimisation was required for efficient expression of these Fabs. Recombinant Fabs were produced by co-transfection of VH and VL vectors into Expi293F™ cells (Section 2.2.4). Initially, 100 mL of Expi293™ cells were used to synthesise Fabs B4, B13 and B13*, to save resources following high expression yields of recombinant ultralong Fabs D08, D08*, and 4K3E* (Table 3.4). These were supplemented with subsequent transient transfections using 100 mL cultures. Fab B4* was not expressed due to the low reported affinity with bRSVF [160].

5.2.2 Affinity chromatography purification of anti-bRSVF Fabs

Recombinant Fabs were retrieved from the media using IMAC and further purified using gel filtration (Section 2.6.1). Gel filtration elution peak UV absorbance traces are provided in (Figs. 5.2 to 5.4). Eluted fractions were characterised by SDS-PAGE (inset), and those of interest were spin concentrated using Amicon® Ultracel-30 centrifugal filters (Merck) (Table 5.1). As in Section 3.4.3, concentrated Fabs were stored at 4 °C.

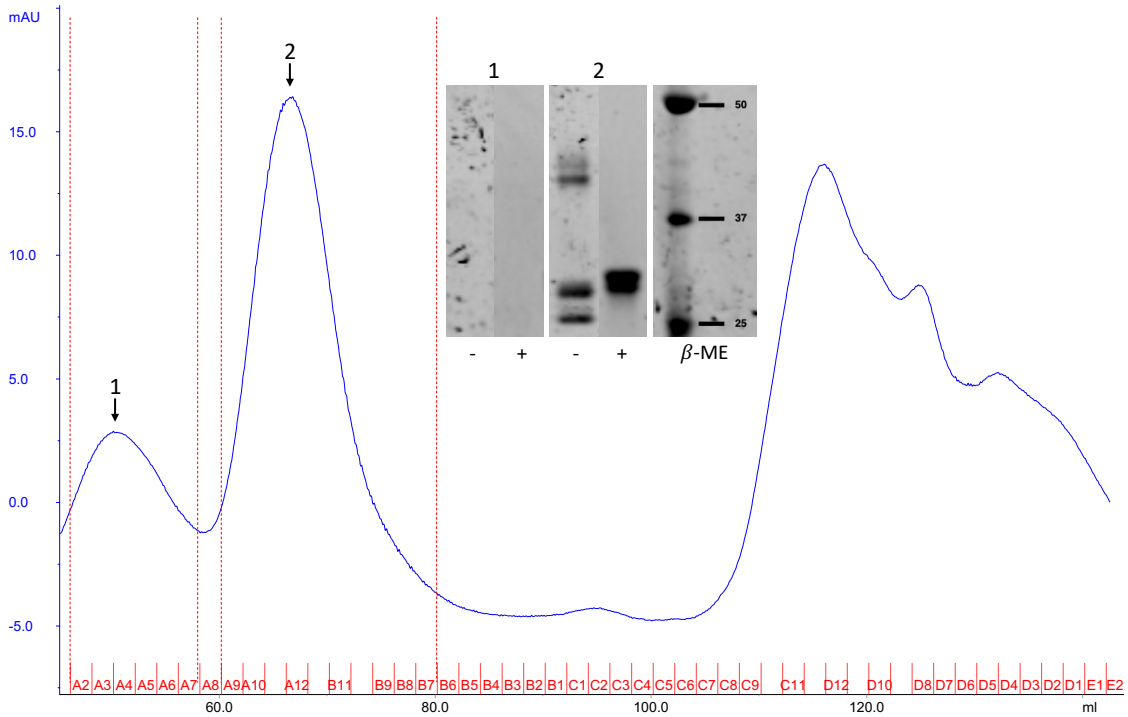


Figure 5.2: Size exclusion chromatography of Fab B4. 280 nm Absorbance trace of gel filtrate. Inset: Indicated fractions were analysed by SDS-PAGE. Samples prepared without (left) and with (right) 10 % β-ME. Band migration compared with Precision Plus Protein Dual Colour Standards (Bio-Rad), with indicated molecular weights in kDa.

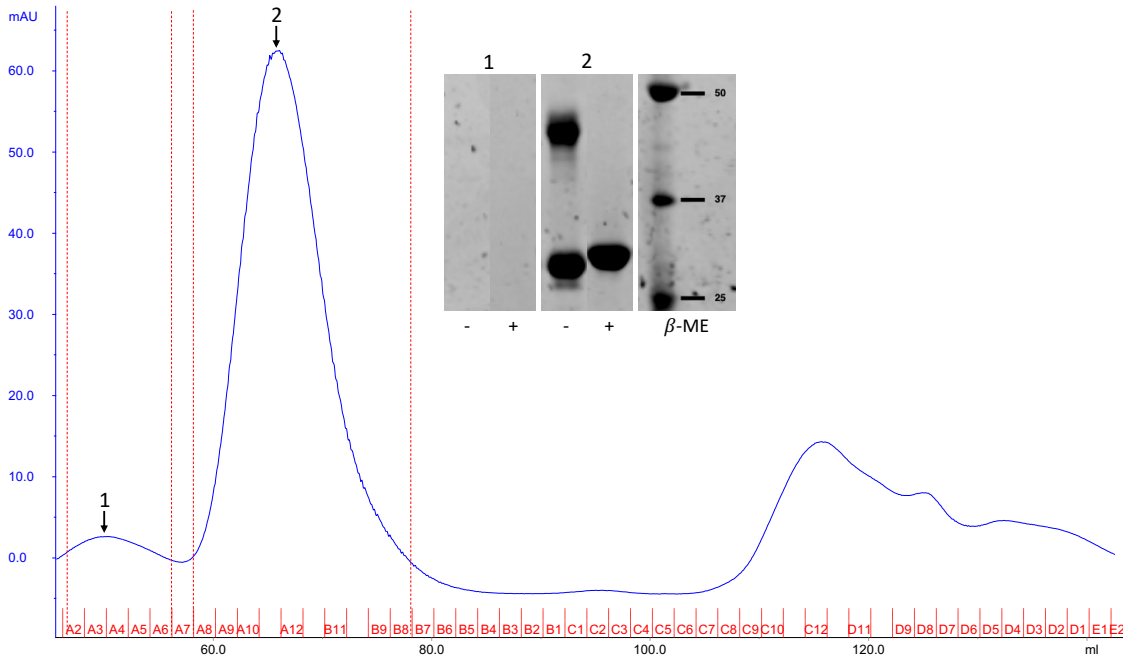


Figure 5.3: Size exclusion chromatography of Fab B13. 280 nm Absorbance trace of gel filtrate. Inset: Indicated fractions were analysed by SDS-PAGE. Samples prepared without (left) and with (right) 10 % β-ME. Band migration compared with Precision Plus Protein Dual Colour Standards (Bio-Rad), with indicated molecular weights in kDa.

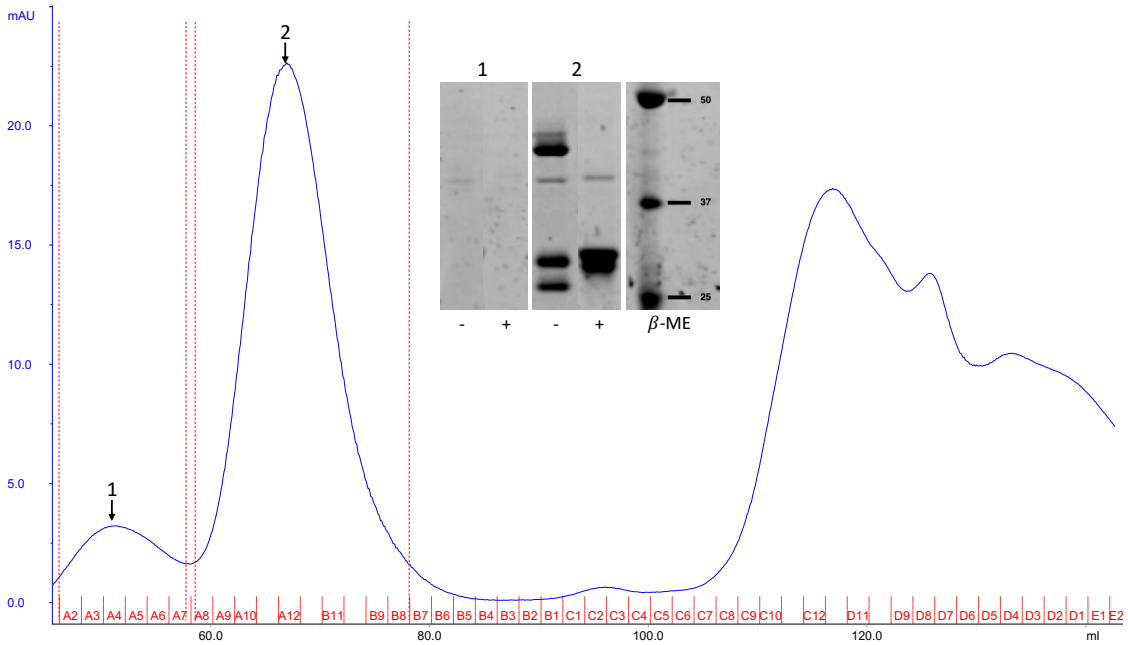


Figure 5.4: Size exclusion chromatography of Fab B13*. 280 nm Absorbance trace of gel filtrate. Inset: Indicated fractions were analysed by SDS-PAGE. Samples prepared without (left) and with (right) 10 % β-ME. Band migration compared with Precision Plus Protein Dual Colour Standards (Bio-Rad), with indicated molecular weights in kDa.

5.3 Expression of bRSVF for structural study

The gene encoding a stabilised prefusion conformation bRSVF variant DS2-v1 [175] was generously provided by Peter Kwong of the National Institutes of Health, within the pVRC8400 vector.

5.3.1 Transient transfection trial

Initially transient transfection (Section 2.2.4) was trialled for the expression of bRSVF using the pVRC8400 vector, however a 300 mL culture of Expi293™ cells only produced a very small UV 280 nm absorbance peak (Section 2.6.1), with the concentration too low in the respective eluted fractions to produce bands in a coomassie-stained SDS-PAGE gel (not presented in this thesis) (Sections 2.7.2 and 2.7.3). Instead, a stable cell line was developed.

5.3.2 Establishment of a stable cell lineage for bRSVF production

To create a stable cell line, the gene encoding the stabilised bRSVF was amplified by PCR (Section 2.4.2) using primers designed complimentary to the pOPINTTGNeo vector (Section 2.1.4); OPPF21999_F and OPPF21999_R (Table 2.4). These primers were designed to amplify an insert of 1486 base pairs. After PCR amplification from the pVRC8400 vector, the PCR products were assessed by 0.8 % agarose gel (Section 2.4.3) (Fig. 5.5), and extracted by a QIAquick Gel Extraction system (Section 2.4.4) before following the established molecular biology workflow of transformation, DNA harvest and sequence verification (Section 2.2.1).

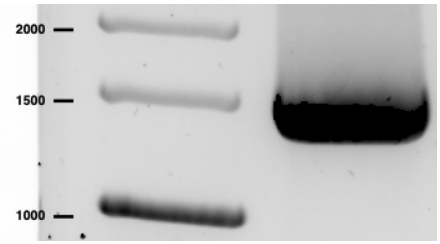


Figure 5.5: PCR amplification of stabilised prefusion conformation bRSV F protein from the pVRC8400 vector using OPPF21999_F and OPPF21999_R primers (Table 2.4). Standard PCR parameters (Section 2.4.2) were used, and analysed on a 0.8 % agarose gel (Section 2.4.3). HyperLadder™ 1 kb DNA Standards (Bioline) used for band migration interpretation, with sizes in base pairs.

Stable lines of Expi293™ cells were prepared as described previously (Section 2.2.5). Cell lines prepared by transformation with pOPINTTGNeo were successful, with 12 distinct lines expressing F protein prepared across all dilutions (Fig. 5.6). Cell lines #14 and #29 showed the strongest levels of expression, with lines #4, #15, and #24 also expressing well. Cell line #29 was used for long term expression of stabilised bRSV F protein, with strong, stable expression up to 12 passages (Fig. 5.7).

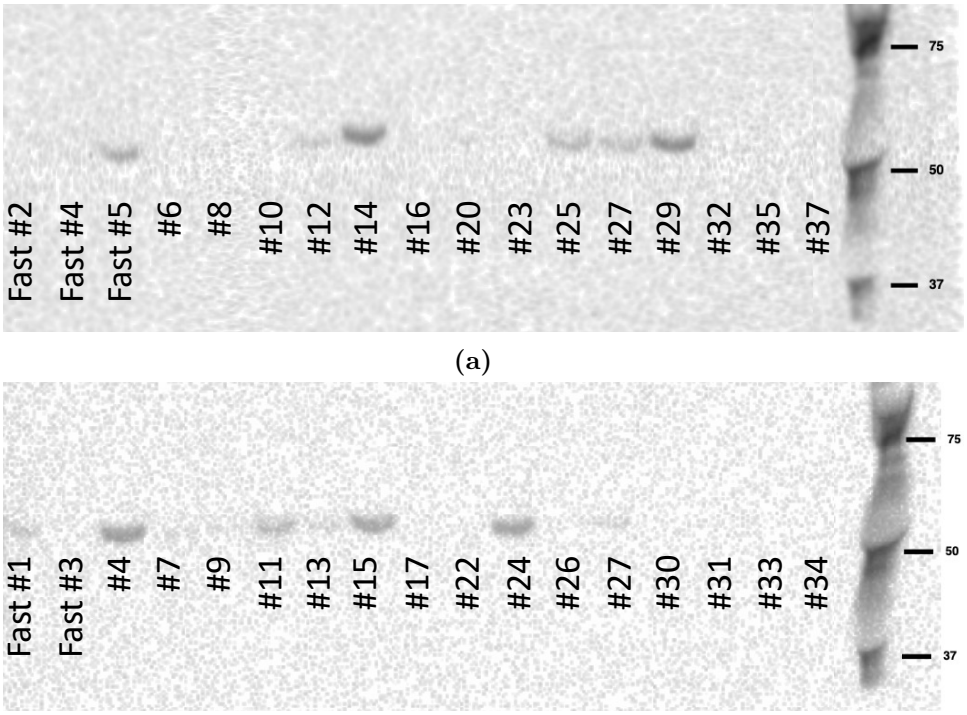


Figure 5.6: Reducing anti-polyhistidine immunoblot of bRSVF cell lines transformed with recombinant pOPINTTGNeo. Aliquots of each cell line collected once 24 Well-plate cultures achieved confluence. Samples prepared with 10 % β-ME, and heated to 95 °C prior to loading. Precision Plus Protein Dual Colour Standards (Bio-Rad) loaded to right-most lane for inference of migrated band molecular weights in kDa.

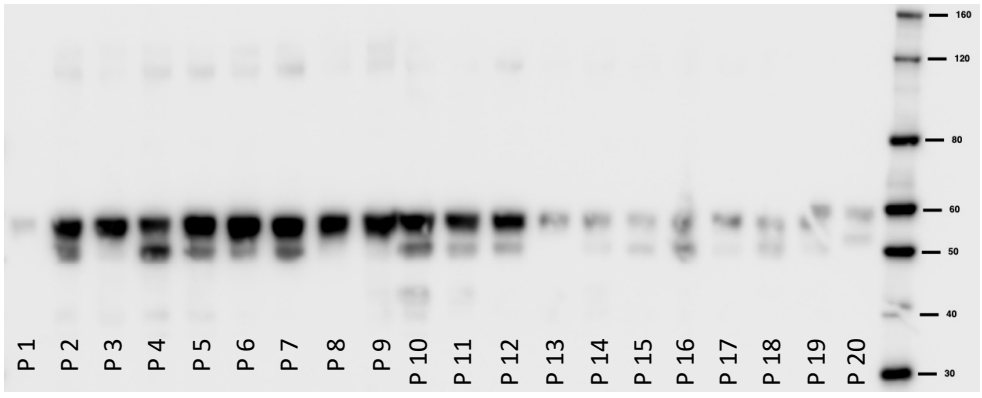


Figure 5.7: Reducing anti-polyhistidine immunoblot of bRSVF cell line #29 (transformed with recombinant pOPINTTGNeo) passages (P). Samples collected 7 days after each passage, and prepared with 10 % β-ME, and heated to 95 °C prior to loading. BenchMark His-Tagged Protein Standards (ThermoFisher Scientific) loaded to right-most lane for inference of migrated band molecular weights in kDa.

5.3.3 Purification of bRSV F protein

Cell line #29 was expanded to 3x 170 mL cultures to produce F protein at scale (Section 2.5.2). bRSVF was purified by combination IMAC and gel filtration at 16 °C following the addition of 5 mM Nickel Sulphate (Section 2.6.2 and Fig. 5.8), however only 2/3 of the 170 mL cultures produced a UV absorbance peak during purification. Eluted fractions were analysed by SDS-PAGE (Section 2.7.2 and Fig. 5.8; inset), and those of interest were spin concentrated (Section 2.6.5) and flash-frozen for storage at −80 °C.

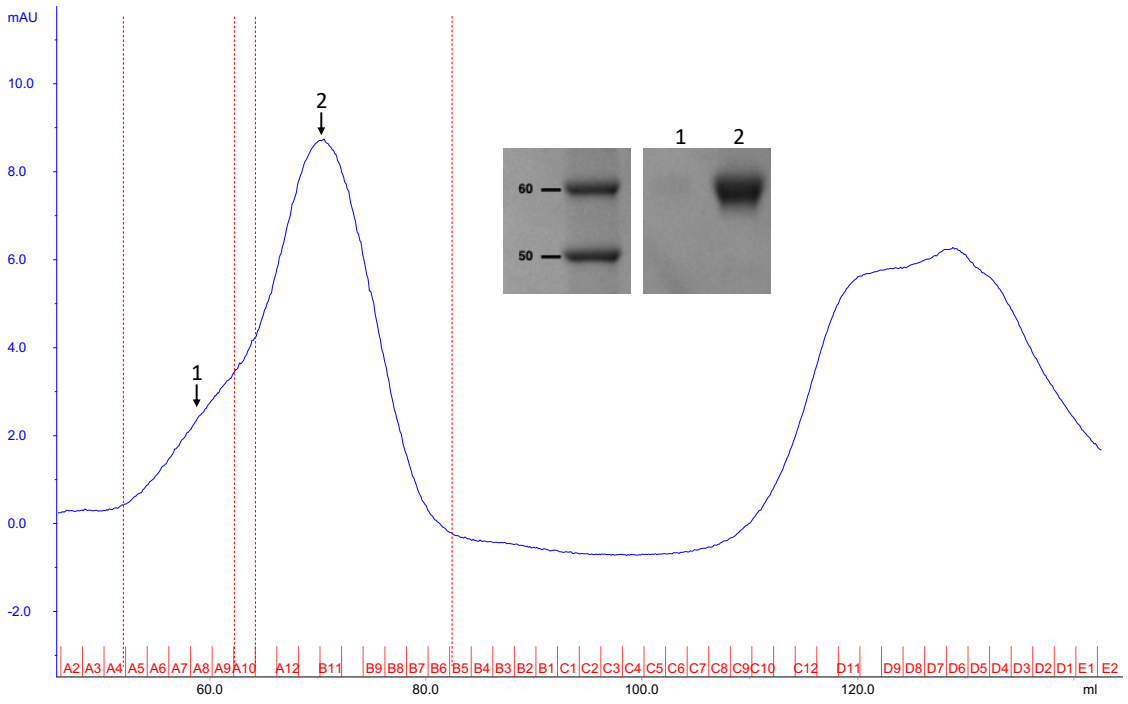


Figure 5.8: Size exclusion chromatography of bRSVF. 280 nm Absorbance trace of gel filtrate. Indicated fractions were analysed by reducing SDS-PAGE (inset). Samples prepared in the presence of 10 % β-ME. Band migration compared with BenchMark His-Tagged Protein Standards (ThermoFisher Scientific).

Table 5.1: Comparative protein yields obtained for Fabs (from transient transfection) and bRSVF (stable cell line) of Expi293F™ cells. Heavy and light chain ID refers to assigned sequence ID within the OPPF Target database. †Time taken for spin-concentration to achieve stated concentration. ‡bRSVF monomer OPPF target ID.

Name:	B4	B13	B13*	bRSVF
Heavy chain:	18771	18773	18773	21999‡
Light chain:	18766	18767	18766	n/a
Pooled Fractions:	A09:B09	A09:B09	A09:B09	2x A05:B06
Pooled volume (mL):	16	16	16	2x 20
Yield (mg/L):	18.624	64.684	18.182	4.12
Concentration (mg/mL):	9.312	32.342	9.091	2.578
Time taken† (minutes):	30	30	30	25

5.4 bRSVF Biophysical Characterisation

5.4.1 Characterisation by SEC-MALS

Further characterisation of the purified protein to confirm successful expression and assembly of the F protein homotrimer is necessary prior to further work. Size exclusion chromatography with multi-angle light scattering (SEC-MALS) is a powerful technique that determines, amongst other things, the precise molecular weight, polydispersity and hydrodynamic radius of a specimen by recording the interactions of the sample with light.

Purified bRSV F protein was diluted with PBS (Fisher Chemical™, ThermoFisher Scientific) at pH 7.40 to achieve a final concentration of 1.25 mg/mL, and 100 µL injected to an ÄKTA Pure chromatography system connected in series to a multi-angle light scattering detector and a refractometer (Section 2.7.8).

SEC-MALS results are typically described in order of ascending molar mass, hence in inverse elution volume order. The bRSVF SEC-MALS peak #1 was characterised with a molar mass moment of 1.928×10^5 g/mol ± 1.695×10^4 g/mol (Fig. 5.9 and Table 5.2), compared to the reported mass of >150 kDa [360], and a theoretical molar mass of 1.725×10^5 g/mol for the F protein homotrimer, as calculated using ProtParam (Swiss Institute of Bioinformatics) [180] (Table A.3). Peak #2 was characterised with a molar mass moment of 3.096×10^5 g/mol ±

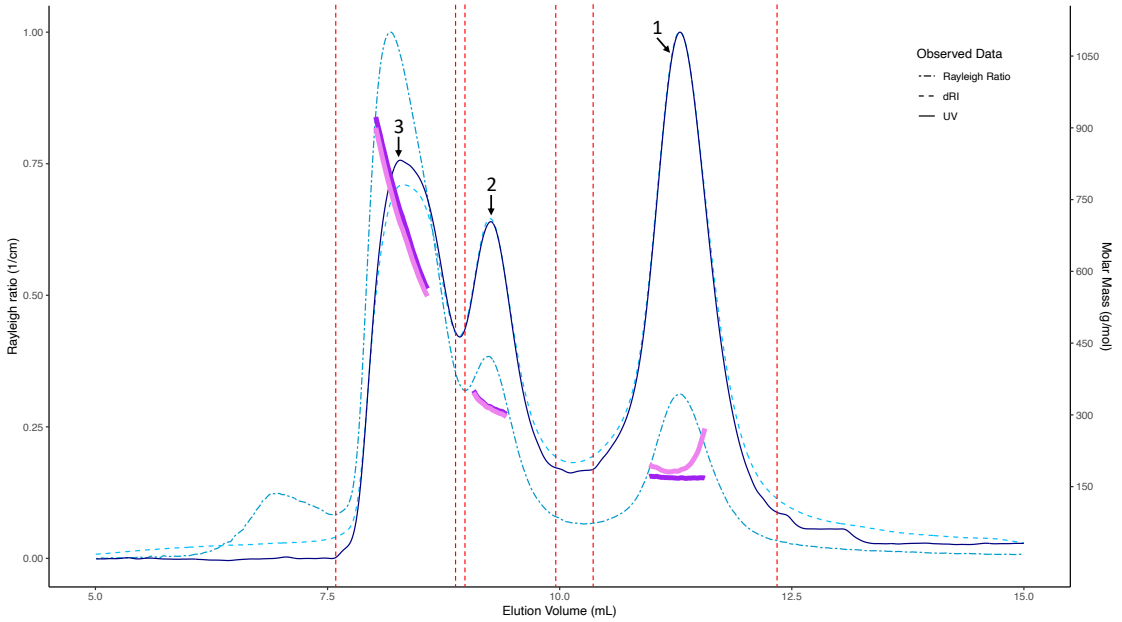


Figure 5.9: Size-exclusion chromatography with multi-angle light scattering characterisation of purified bRSVF homotrimers. Calculated molar masses for the protein (thick purple) and conjugated protein (thick pink) are overlaid measured data traces for Rayleigh ratio, differential refractive index (dRI), and UV absorbance. bRSVF was injected at 1.25 mg/mL, and measured using a Dawn Heleos-II multi-angle light scattering detector (Wyatt Technologies) and an Optilab T-rEX refractometer (Wyatt Technologies). Data were analysed using the ASTRA® software v5.3 (Wyatt Technologies).

3.000×10^3 g/mol, likely corresponding to a dimer of F protein homotrimers, (theoretical molar mass of 3.450×10^5 g/mol). The final proteinaceous peak is likely a tetramer of F protein homotrimers, with a calculated molar mass moment of 7.468×10^5 g/mol $\pm$ 6.347×10^3 g/mol (theoretical molar mass of 6.900×10^5 g/mol) (Fig. 5.9 and Table 5.2). The final peak in the dRI trace corresponds to higher order species in the sample, but at too low a concentration to meaningfully absorb UV.

For all species, only a very small degree of glycosylation was observed (1.7 %, and 1.9 % of the total species molar mass moments), with conjugate molar mass moments of 3.396×10^2 g/mol and 6.059×10^2 g/mol for peaks #1 and #2 respectively. The conjugate molar mass moment for peak #3 could not be calculated (Table 5.2). The steep gradient of calculated molar mass across this peak indicates a more heterogeneous species (Fig. 5.9), making determination of conjugated species impossible. The high errors on the conjugate molar mass moment simply reflects the much smaller contribution conjugates make to the total species mass.

Table 5.2: SEC-MALS characterisation of bRSVF.

	Peak #1	Peak #2	Peak #3
Masses			
Total (μg)	24.583	10.403	17.720
Protein (μg)	24.162	10.205	18.942
Conjugate (μg)	00.421	00.198	-
Molar Mass Moments			
Total (g/mol)	1.962×10^5 (± 8.791 %)	3.157×10^5 (± 0.891 %)	6.962×10^5 (± 0.774 %)
Protein (g/mol)	1.928×10^5 (± 8.792 %)	3.096×10^5 (± 0.969 %)	7.468×10^5 (± 0.850 %)
Conjugate (g/mol)	3.396×10^2 (± 508.101 %)	6.059×10^2 (± 46.440 %)	n/a
Fractional Moments			
Protein	0.983 (± 0.267 %)	0.981 (± 0.378 %)	n/a
Polydispersity			
Total (Mw/Mn)	1.010 (± 12.041 %)	1.002 (± 1.264 %)	1.020 (± 1.099 %)
Protein (Mw/Mn)	1.009 (± 12.046 %)	1.001 (± 1.373 %)	1.022 (± 1.205 %)
Conjugate (Mw/Mn)	1.261 (± 854.650 %)	1.159 (± 78.702 %)	n/a

The low extent of observed bRSVF glycosylation by SEC-MALS is consistent with predictions of glycosylation sites using the NetNGlyc artificial neural network for predicting N-glycosylation [187, 189], and the NetOGlyc network for predicting O-glycosylation [188, 190](Section 2.7.9). NetNGlyc identified two Asn-Xaa-Ser/Thr sequons within the bRSVF monomer (where 'Xaa' is any non-proline residue), but only one of these asparagines (at position 33) was predicted to be glycosylated with a glycosylation potential of 0.7209 (all 9 neural networks in agreement and above the threshold of 0.5000)(Fig. 5.10). Despite 7/9 of the neural networks predicting N-glycosylation, asparagine at position 469 did not achieve the threshold, with an averaged glycosylation potential of 0.4418; however, the crystal structure of stabilised bRSV prefusion F protein (PDB ID: 5TDL [175]) showed glycosylation only at the 469 position. No other glycosylation sites were observed in the crystallographic structure nor were predicted by NetNGlyc. NetOGlyc v4.0 [188, 190] was also

performed, identifying 89 possible O-glycosylation sites in the bRSVF monomer, however none of these achieved the averaged glycosylation potential threshold of 0.5000. The mean O-glycosylation potential value for all potential sites was 0.0902, with a standard deviation of 0.0743, ranging from 0.0039 for Thr280 to 0.3652 for Ser471.

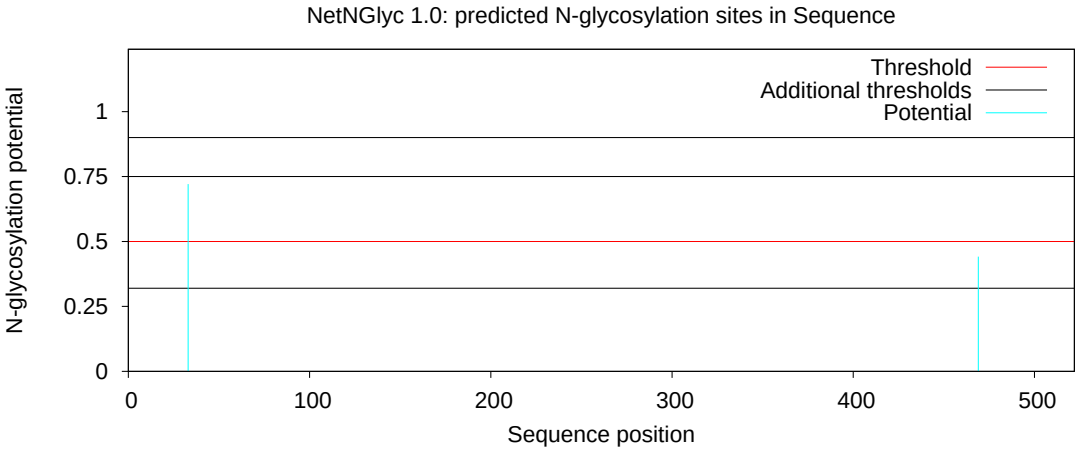


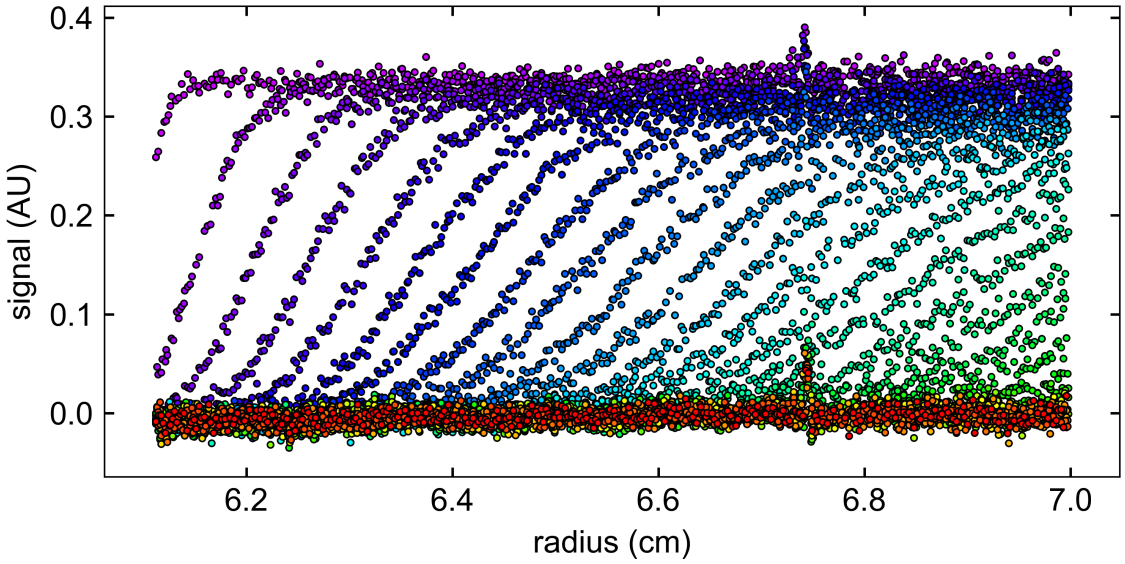
Figure 5.10: Predicted N-glycosylation potential for bRSVF monomers. Glycosylation potential calculated by NetNGlyc v1.0 [187, 189].

The observed pattern of molar mass moment heterogeneity within a SEC separated species, as demonstrated strongly by those with a larger hydrodynamic radius, could be explained by a concentration-dependent multimer self assembly or aggregation effect, where the extent of self assembly decreases as the sample is diluted further down the SEC column. This may indicate that the concentration achieved during protein spin-concentration is too high for bRSVF, however some concentration is necessary to push the equilibrium towards trimerisation.

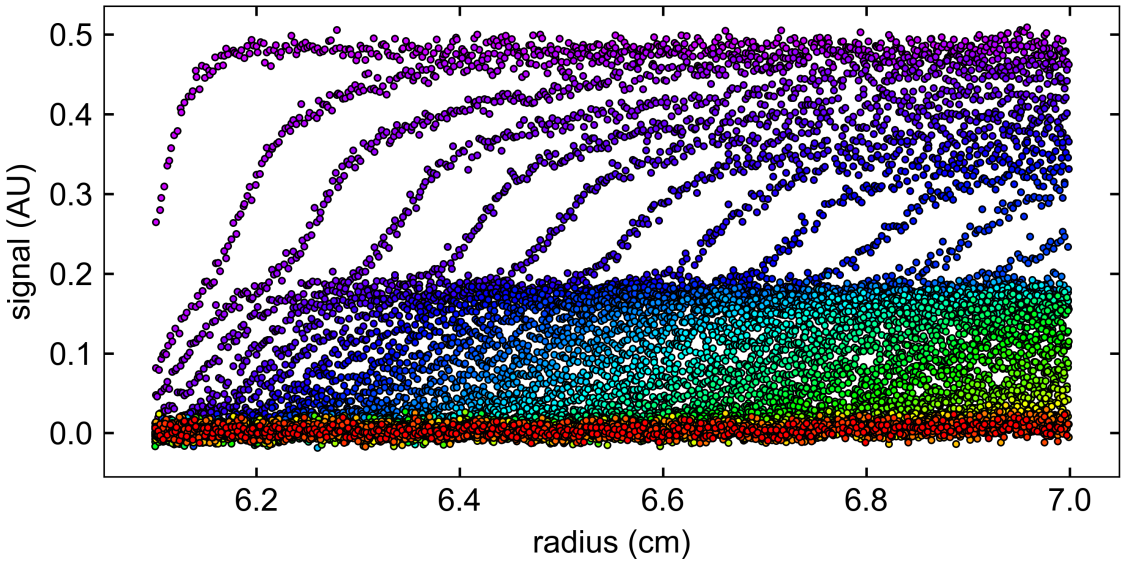
5.4.2 Verification of Fab binding by AUC

Upon satisfactory production and assembly of bRSVF homotrimers, analytical ultracentrifugation (AUC) was performed to confirm interaction between the stably-produced bRSVF and Fabs B4, B13 and B13*, and to add an orthogonal technique to the SPR experiments reported earlier (Section 5.1.5) [160]. The sedimentation coefficients, (s), of the lone species are first determined, and compared against the (s) of the complexed sample. A shift in (s) indicates bimolecular interaction.

Sedimentation velocity scans were collected every 20 seconds across the AUC centrepiece at 50,000 rpm (Section 2.7.10) Each time point is distinctly coloured, providing the distribution of protein within the centrepiece at each scan, and allowing calculation of the sedimentation velocity of the samples. In a complexed system, molecular species will sediment at different rates, producing a double sigmoid pattern (Fig. 5.11b). Finally, a minor observation is the slight inflection to all data traces at approximately 6.75 cm in Fig. 5.11a, likely due to an imperfection, such as a scratch, on the centrepiece lens.



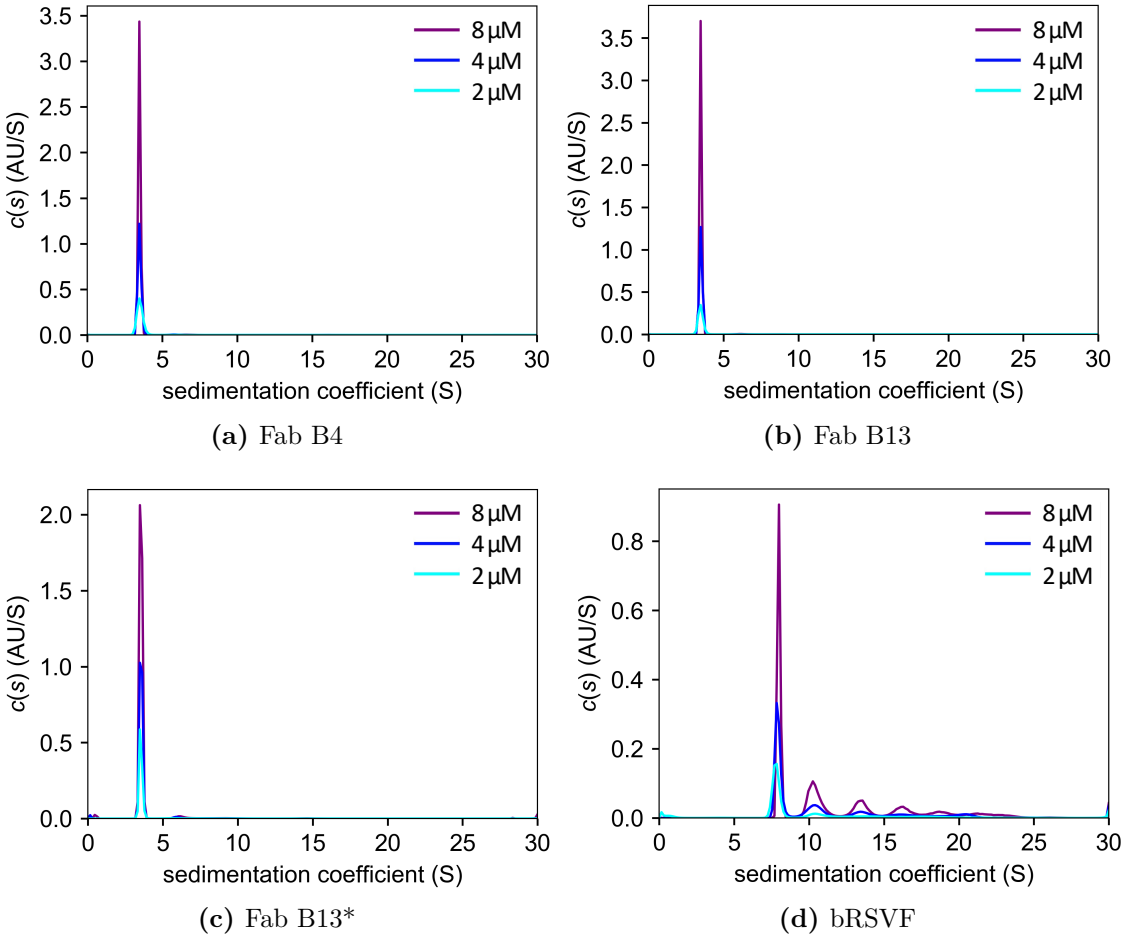
(a) 4 µM Fab B4



(b) 2 µM bRSVF with 4 µM Fab B4

Figure 5.11: Representative analytical ultracentrifuge 280 nm **absorbance scans for** 4 **µm Fab B4 (5.11a) and** 2 **µm bRSVF in complex with** 4 **µm Fab B4 (5.11a).** For clarity, only even numbered scans are shown, representing data collections at 40 s intervals. Data were collected at 50,000 rpm using a ProteomeLab XL-I (Beckman Coulter) analytical ultracentrifuge.

The (s) distributions are consistent with prior SEC characterisations of the individual species. Individual purified Fabs present a single (s) peak of around 3.5 S, corresponding to calculated molecular weights of 44.5 kDa to 49.6 kDa (Fig. 5.12 and Table 5.3), and consistent with published molecular weights for these Fabs



(a) Fab B4 **(b)** Fab B13

(c) Fab B13* **(d)** bRSVF

Figure 5.12: Sedimentation coefficient distribution characterisations for Fab B4 (5.12a), B13 (5.12b), and B13* (5.12c), and bRSVF (5.12d). Data were collected at 50,000 rpm using either a ProteomeLab XL-I (Fabs B4 and B13) or an Optima AUC (Fab B13* and bRSVF) (Beckman Coulter) analytical ultracentrifuge. AUC data were processed using SEDFIT [191], fitting to the c(S) model [192] to obtain sedimentation coefficient distributions. Figures prepared using GUSSI v1.4.1.

(B4: 46 kDa, B13: 48 kDa, B13*: 48 kDa) when characterised by SEC-MALS [160]. F protein appears as multiple discrete peaks, indicative of higher order multimer self assembly, and consistent with the observation of higher order species by SEC-MALS (Figs. 5.9 and 5.12d). The dominant (s) peak of bRSVF homotrimers appears at around 7.8 S, with a calculated molecular weight between 117.8 kDa and 145.0 kDa (Table 5.3). The frictional ratio, f/f_0, of bRSVF is lower than that of the Fab characterisations, but as explained in Section 2.7.10, this is consequent of the multiple (s) species, and hence bRSVF calculated molecular weights from these AUC data should be interpreted with caution. Nevertheless, the bRSVF

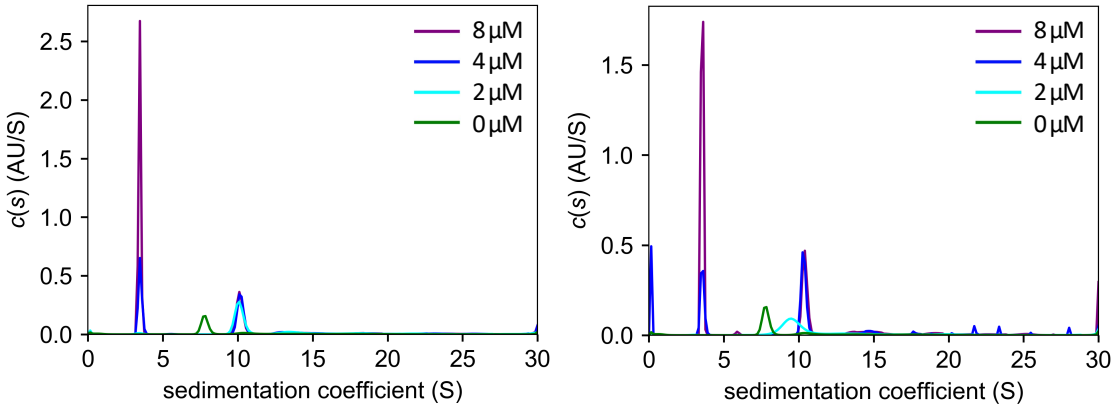
calculated molecular weight by AUC is broadly consistent with the reported mass of bRSVF of >150 kDa by SDS-PAGE [360], and the calculated molecular weight by SEC-MALS (Tables 5.2 and 5.3), confirming the F protein has been successfully purified, most likely as a homotrimer.

Table 5.3: Sedimentation coefficient results and estimated molecular weights for lone proteins. f/f_0 is the best-fit frictional ratio for the (s) distribution. Results for dominant species from the c(s) analysis.

	Molarity (μM)	Molar Mass (kDa)	(s)	f/f_0
Fab B4	8.0	46.8	3.48	1.30
	4.0	48.6	3.48	1.33
	2.0	49.3	3.51	1.33
Fab B13	8.0	47.9	3.48	1.31
	4.0	48.1	3.49	1.31
	2.0	49.6	3.48	1.34
Fab B13*	8.0	44.5	3.54	1.23
	4.0	45.6	3.54	1.25
	2.0	48.4	3.53	1.31
bRSVF	8.0	117.8	7.89	1.01
	4.0	129.4	7.83	1.09
	2.0	145.0	7.80	1.19

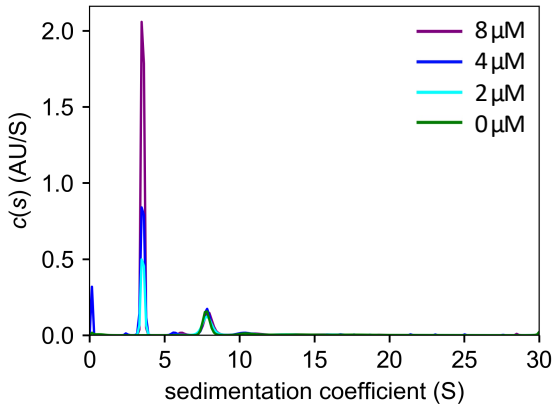
Interestingly, upon introduction with Fab, the multipeak distribution of bRSVF (s) values, indicative of multimer self assembly, appears to resolve, leaving just (s) peaks corresponding to complexed bRSVF and free Fab (Fig. 5.13). Addition of either Fab B4 or B13 shifts the bRSVF (s) from 7.8 S to around 10.1 S to 10.4 S, implying a strong bimolecular interaction.

At equimolar concentrations of Fab B4 and bRSVF monomers, a minute (s) peak at 3.5 S is observed, corresponding to a small population of free B4 (Fig. 5.13). Together with the (s) shift to 10.1 S, this implies that nearly all B4 is bound to bRSVF (Table 5.4). As the Fab B4 concentration is further increased, the bRSVF/B4 complex (s) does not shift any further to the right, nor does this population significantly increase, implying saturation at 2 µM B4. Instead, the (s) peak corresponding to free B4 just increases (Fig. 5.13a). These observations suggest a tight interaction between Fab B4 and bRSVF, corroborating the reported 86 nM affinity reported by SPR (Section 5.1.5) [160].



(a) bRSVF + Fab B4

(b) bRSVF + Fab B13



(c) bRSVF + Fab B13*

Figure 5.13: Sedimentation coefficient distribution characterisations for 2 µm bRSVF in complex with increasing concentrations of Fab B4 (5.13a), B13 (5.13b), and B13* (5.13c). Data were collected at 50,000 rpm using either a ProteomeLab XL-I (bRSVF + B4) or an Optima AUC (bRSVF + B13 and bRSVF + B13*)(Beckman Coulter) analytical ultracentrifuge. AUC data were processed using SEDFIT [191], fitting to the c(S) model [192] to obtain Sedimentation coefficient distributions. Figures prepared using GUSSI v1.4.1.

At equimolar concentrations of Fab B13 and bRSVF monomers, the bRSVF/B13 complex (s) peak is observed at an S-value intermediate between the free and bound S-values for bRSVF (9.58 S), implying a reasonably tight interaction between Fab B13 and bRSVF, but one that is likely to be either weaker than, or kinetically faster than the interaction between bRSVF and Fab B4. The SPR characterisation of this interaction reported an affinity of 125 nM (Section 5.1.5) [160]. At equimolar concentration, there is no (s) peak corresponding to free B13. As the concentration of Fab B13 is increased, bRSVF becomes saturated, with a (s) of 10.3 S to 10.4 S,

and a (s) peak appears at 3.5 S for free B13.

By contrast, addition of Fab B13*, even at the highest concentration, only produced a minor shift to the bRSVF (s) peak from 7.80 S to 7.93 S, implying a much weaker interaction than Fabs B4 or B13, if any at all. Again, this is consistent with the SPR characterisation of 234 nM binding affinity of B13* (Section 5.1.5) [160]. At all concentrations of Fab B13*, the (s) peak corresponding to free B13* is strongly populated.

Table 5.4: Sedimentation coefficient results and estimated molecular weights for 2 µm bRSVF complexed with increasing concentrations of Fabs B4, B13 and B13*. f/f_0 is the best-fit frictional ratio for the (s) distribution. Results for complexed species from the c(s) analysis.

	Molarity (μM)	Molar Mass (kDa)	(s)	f/f_0
Fab B4	8.0	242.1	10.14	1.29
	4.0	250.0	10.18	1.31
	2.0	254.4	10.12	1.34
Fab B13	8.0	242.0	10.40	1.26
	4.0	254.1	10.33	1.31
	2.0	183.7	9.58	1.14
Fab B13*	8.0	151.1	7.93	1.20
	4.0	157.9	7.84	1.25
	2.0	151.2	7.85	1.22
bRSVF	-	145.0	7.80	1.19

Sedimentation velocity experiments such as these could be used to calculate equilibrium dissociation constants (K_D) for direct comparison with SPR, but subsequent experiments with nanomolar Fab concentrations would be required to negate the large error of extrapolating the micromolar data.

The sedimentation coefficients of the interacting species are too similar to deconvolute the exact concentrations of each protein in the fully saturated bRSVF/Fab complexes. However, if one assumes that the concentration of bRSVF in each experiment is exactly equal (which of course it would not have been, but it should have been close), then one can subtract the absorbance attributable to bRSVF from the overall absorbance, and calculate the concentration of Fab present in the complex. For both B4 and B13, the results of this suggest the fully bound

complex (s) peak has roughly equal concentration of both bRSVF monomer and Fab, implying one bRSVF homotrimer binds three Fab molecules. The molecular weights obtained for the bRSVF/Fab complex are a little too low to fully support this, but, as described before (Sections 2.7.10 and 5.4.2), are subject to the limitations of global f/f_0 fitting for multiple observed species.

Table 5.5: Predicted species concentrations for bRSVF/Fab complexes.

	Molarity (μM)	[bRSVF] (μM)	[Fab] (μM)	f/f_0
Fab B4	8.0	1.45	1.12	1.29
	4.0	1.45	1.03	1.31
	2.0	1.45	1.25	1.34
Fab B13	8.0	1.45	1.41	1.26
	4.0	1.45	1.30	1.31
	2.0	1.45	0.80	1.14
Fab B13*	8.0	1.45	0.20	1.20
	4.0	1.45	0.11	1.25
	2.0	1.45	0.09	1.22
bRSVF	-	1.45	-	1.19

5.5 Structural biology

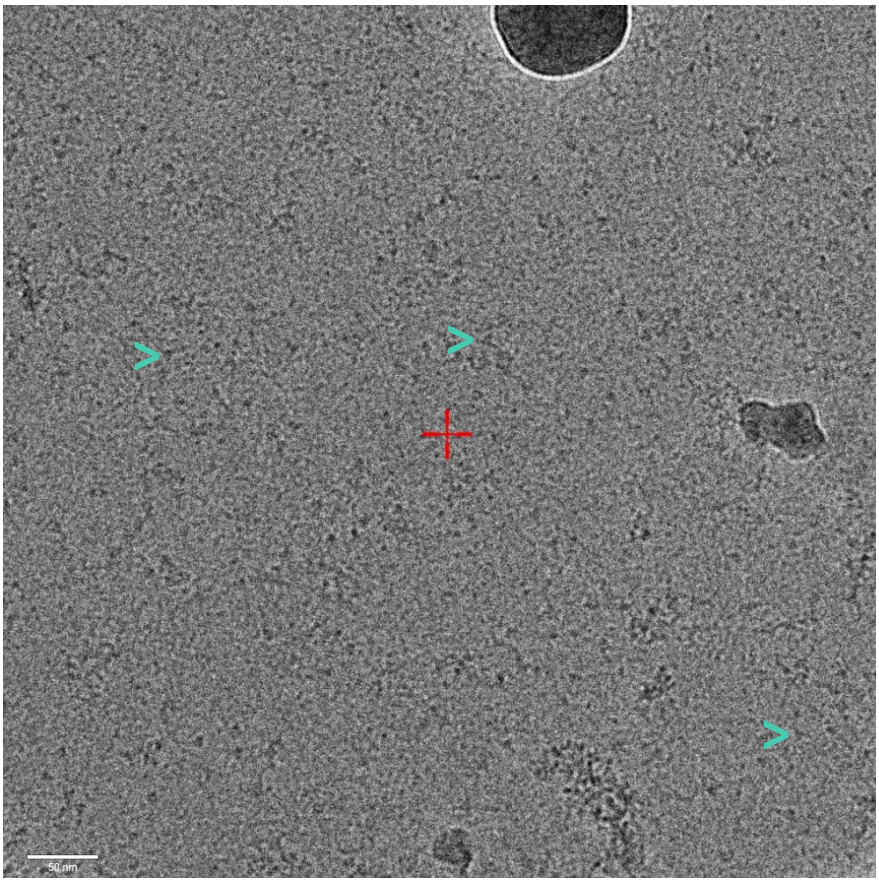
The epitopes accessed by antibodies B4 and B13 have been previously identified by escape mutagenesis, trypsin digestion and peptide synthesis [356, 359], and by competitive binding assay [357]. B13 binds residues 417 to 438, corresponding to an overlap of antigenic sites IV, V and VI, whilst B4 binds residues 262 to 272 within antigenic site II, but B4 recognition is additionally dependent on Asparagine at residue 216. Given this structural contribution to the bRSVF epitope, structural study would reveal the mechanism of action of antibody binding. It may also provide an explanation to the observed difference in binding affinity with B13* and B4* compared to B13 and B4 respectively, especially given the structural similarity of the B13 and B13* CDR3H loop conformation (Fig. 5.1 and Section 5.1.6) [160].

Cryogenic Electron Microscopy (cryo-EM) was used to investigate the structural aspects of bRSVF recognition by bovine antibody B4 as Fab fragments due to the size of the bRSVF/B4 complex, the reduced sample demand, and non-requirement of crystallisation. The bRSVF/B4 complex was selected for cryo-EM specimen optimisation, as B4 demonstrated the tightest binding with bRSVF by SPR (Section 5.1.5)[160], and AUC (Section 5.4.2) analyses.

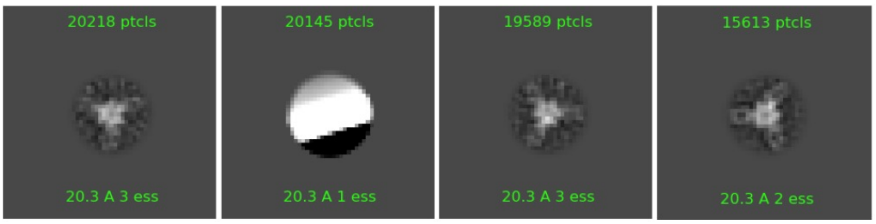
5.5.1 bRSVF/B4 complexes demonstrate preferential orientation

5.5.1.1 cryo-EM specimen preparation

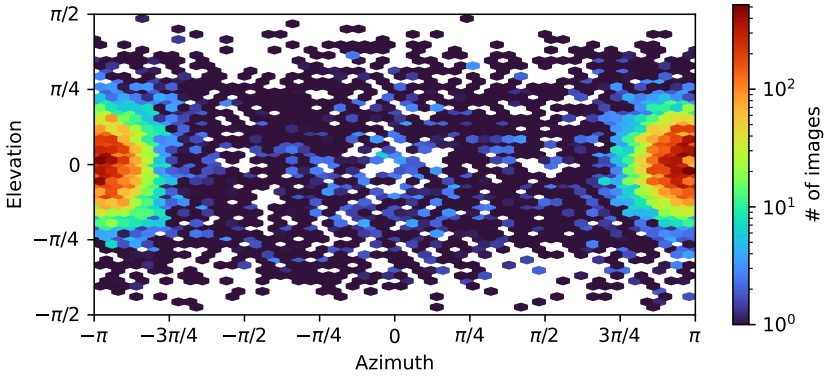
The F protein was mixed with Fab B4 at a 2-fold molar excess per bRSVF monomer, achieved through adding Fab B4 to 6-fold the bRSVF molar concentration (3.6 µM and 0.6 µM respectively). The mixture was then applied to plasma-cleaned (Section 2.8.4.2) C-Flat Ultra-thin Holey -1.2/1.3 Carbon film on 200 Copper mesh cryo-EM grids, and to plasma-cleaned (Section 2.8.4.2) Lacey ultrathin Carbon film on 400 Copper mesh cryo-EM grids (Section 2.8.4.1), and plunge frozen using a Vitrobot Mk IV (Section 2.8.4.3). Specimens were screened using a 200 keV Glacios electron cryomicroscope within STRUBI (Table 2.13).



(a)



(b)



(c)

Figure 5.14: Electron cryomicroscopy of Fab B4 in complex with bRSVF homotrimers. Specimen plunge frozen by Vitrobot. (5.14a) Example micrograph, with particles marked by green arrowheads. (5.14b) Representative 2D class averages with indicated population, estimated resolution and effective sample size (ESS). (5.14c) Viewing distribution heat map.

5.5.1.2 Results

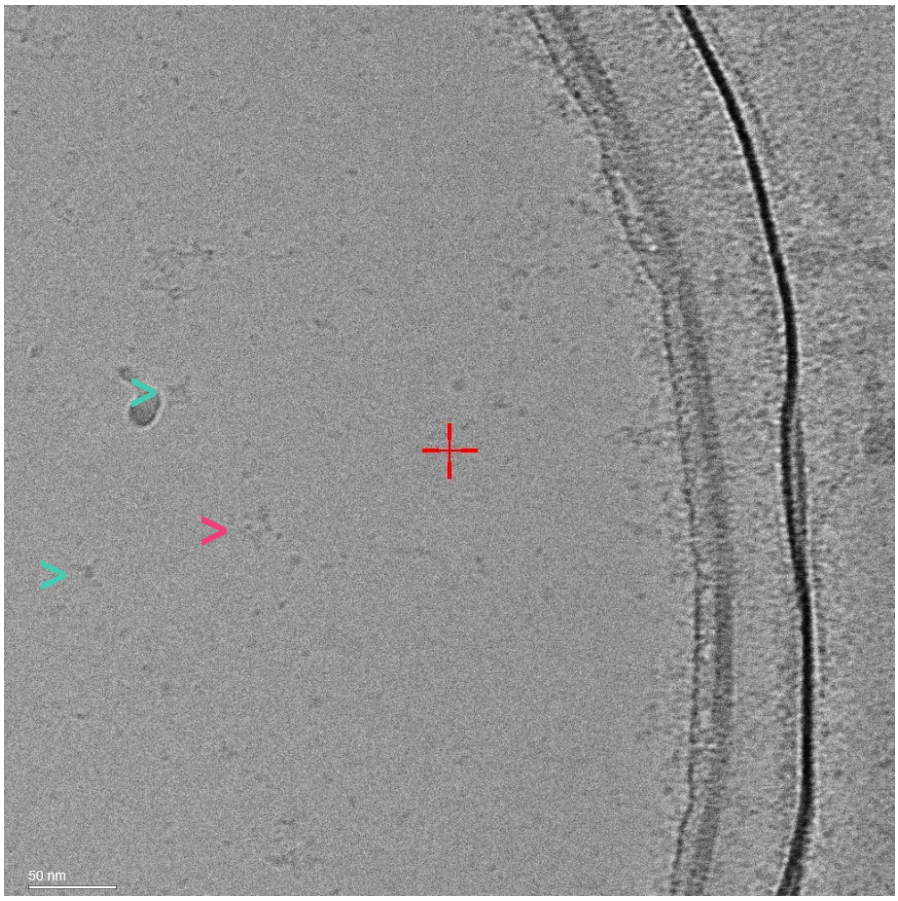
Top-down views dominate the 2D class averages (Fig. 5.14b) of bRSVF, with a few erroneous classes for pseudo-particles, picking the cryo-EM grid hole boundaries. In these top-down views, B4 can been seen decorating the central bRSV F protein homotrimer in an equally spaced (120°), ternary manner. One can even see distinction between the variable and constant Fab domains (Fig. 5.14b). Top-down preferred orientation of an F antigen complex was also observed when studying the interaction of stabilised prefusion conformation human RSV F protein (hRSVF) and humanised monoclonal antibody motavizumab [361], though no such preferential orientation was observed with hRSVF in complex with neutralising human antibody CR9501. One would expect side views would also show a triangular structure, but with the Fab decoration opposite one another, and perpendicular to the F protein, hence producing a near 90° morphology, based on the location of the reported B4 and B13 epitopes (Section 5.1.5) [356, 357, 359]. Though some side views are observed, the vast majority of views are top-down (Fig. 5.14c). This is reflected in the initial 3D *ab-initio* reconstruction, where the limited side view signal is virtually lost in the reconstruction, leaving just the incomplete 'top' portion (Fig. 5.16). Unlike with the dissociated SAT2 12S pentamers (Section 4.6.3.1), the threefold symmetry axis of bRSVF is parallel to the field of view, and so not all particle poses can be sampled simply through imposing C3 symmetry.

5.5.2 Resolution of bRSVF/B4 preferential orientation

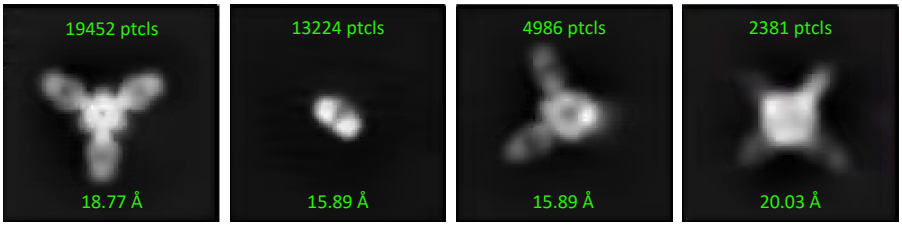
Preferential orientation arises from the interaction between the sample under study, and either the air-liquid boundary, or the liquid-grid boundary [243]. One strategy to alleviate preferred orientation, is to reduce the delay between sample application and grid vitrification, in attempt to immobilise the specimen before it has localised to either of these interfaces in a preferential manner. The Chameleon sample vitrification robot (SPT Labtech) can achieve this by combining a picoliter dispenser (to reduce applied volume) with self-wicking grids (to negate the need for blotting) (Section 2.8.4.4). Another strategy is to add detergent (0.05 % (v/v) Tween-20) to the sample preparation, to encourage particle tumbling but this failed to improve the observed preferential orientation (omitted from this thesis).

5.5.2.1 cryo-EM specimen preparation

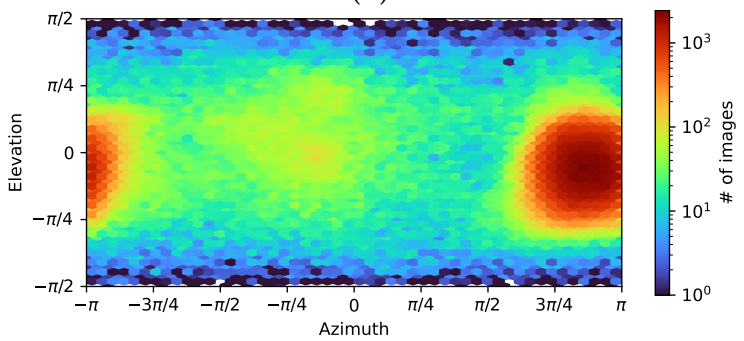
The F protein was again mixed with Fab B4 at a 2-fold molar excess per bRSVF monomer, however higher sample concentrations are required for Chameleon-prepared grids, owing to the concentrating effect of self-wicking nanowires (and hence diluting effect to sample applied over data acquisition areas). F protein was prepared to 10 µM, and mixed with Fab B4 at a final concentration of 60 µM. Following optimisation of glow discharge duration (which affects the speed of nanowire self-wicking), and the plunge delay (to allow time for self-wicking), 300 Mesh Self-Wicking Copper grids were glow discharged for 50 s, before 100 nL application of bRSVF/B4 complex as a single stripe, and subsequent plunge freezing into liquid ethane after a 150 ms plunging delay. Specimens were screened using a 200 keV Glacios electron cryomicroscope within eBIC at DLS (Table 2.13).



(a)

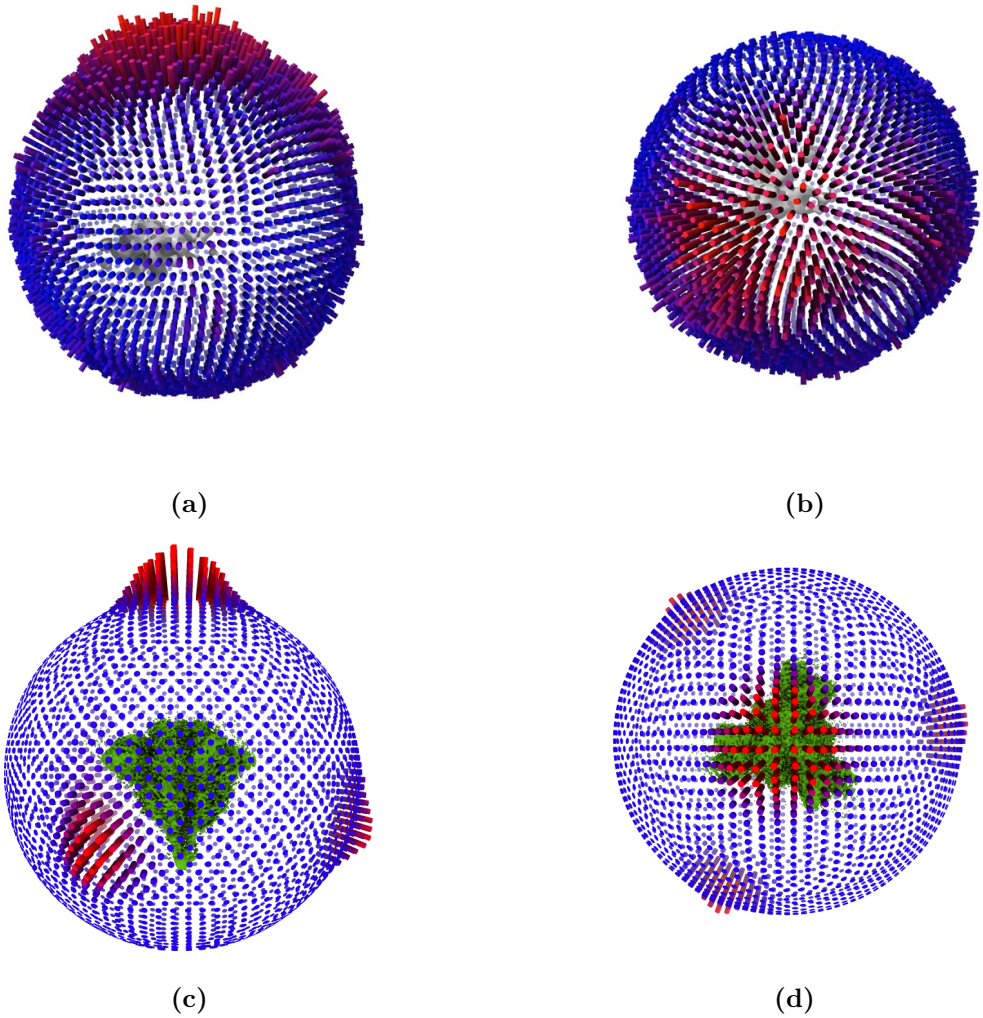


(b)



(c)

Figure 5.15: Electron cryomicroscopy of Fab B4 in complex with bRSVF homotrimers. Specimen plunge frozen by Chameleon. (5.15a) Example micrograph, with particles marked by green arrowheads. Pink arrowhead shows suspected bRSVF/B4 complexes in side view pose. (5.15b) Representative 2D class averages with indicated population and estimated resolution. (5.15c) Viewing distribution heat map. Representative 2D class averages are provided in Fig. 5.15b.



(a) **(b)**

(c) **(d)**

Figure 5.16: *Ab-initio* 3D reconstruction of bRSVF/B4 complex, with overlaid observed sample pose distribution. Specimen prepared by vitrobot (5.16a and 5.16b), and by chameleon (5.16c and 5.16d). Viewed from the side (5.16a and 5.16c), and from the top (5.16b and 5.16d).

5.5.2.2 Results

Top-down views of the bRSVF/B4 complex are again well represented in the cryo-EM grids prepared by cryo-EM (Fig. 5.15b), however a small population (2,381 particles; approximately 8 % of the total 29,380 picked particles) presented in the side view. The viewing distribution has improved, as seen with a more homogeneous viewing distribution heat map (Fig. 5.15c). Interestingly, free Fab B4 could also be classified, which may indicate the molar excess of Fab is too high, as this signal would contribute

more to the background than picked bRSVF/B4 particles. A distinct top-down view class average of bRSVF/B4 was also observed, which appears to present Fab B4 with a reduced occupancy, and also leaving one bRSVF epitope unbound.

In any case, more signal would be needed to produce a high resolution structure of the bRSVF/B4 complex.

5.5.3 Improving particle density

Increasing the sample concentration is the easiest way of increasing observed particle density on the cryo-EM grids, but this is materially costly, and there is no guarantee highly concentrated sample will behave similarly to lower concentrations (with the formation of aggregates etc). Furthermore, bRSVF was concentrated to 2.578 mg/mL (14.95 µM), so trialling higher concentrations would require pooling and further spin-concentration steps (including associated sample loss due to sticking to the filter membrane and due to filter imperfections), and risks a greater proportion of aggregated species (as seen at the edges of the grid squares, where sample accumulates due to the self-wicking nanowires, even if AUC experiments suggest multimer self assembly is reduced once Fab is added). An alternative is to trial with alternative cryo-EM grid surface chemistry. Graphene-oxide (GO) can be applied to cryo-EM grids to produce a uniformly flat, charged surface (once glow discharged) to adsorb sample against [243].

5.5.3.1 cryo-EM specimen preparation

The F protein was once again mixed with Fab B4 at a 2-fold molar excess per bRSVF monomer; bRSVF was prepared to 10 µM, and mixed with Fab B4 at a final concentration of 60 µM. Following optimisation of glow discharge duration, and the plunge delay, GO-coated 300 Mesh Self-Wicking Copper grids (Section 2.8.4.5) were glow discharged for 30 s, before 100 nL application of bRSVF/B4 complex as a single stripe, and subsequent plunge freezing into liquid ethane after a 140 ms plunging delay. Specimens were screened using a 200 keV Glacios electron cryomicroscope within eBIC at DLS (Table 2.13).

5.5.3.2 Results

Figure 5.17a was captured during cryo-EM screening, and makes immediately clear the sample preference for GO coating. A small tear to the GO layer was observed (orange arrowheads), allowing an easy side-by-side comparison of particle density on the same cryo-EM grid (and consequently the same amount of applied glow discharge, plunging delay, sample preparation conditions, and copper nanowire density, all affecting sample wicking). Where the GO is absent, a very low density of particles was observed, reminiscent of the non-coated grid screened previously (Fig. 5.15a). Conversely, when on the GO coating, particles appeared as nearly confluent to the point where manually picking bRSVF/B4 particles becomes a challenge. The additional GO layer will contribute more background noise to the observed micrograph than non-coated regions, but not to such an extent as observed in Fig. 5.17a. Automated particle picking struggled to discern bRSVF/B4 complexes from the high signal background (Fig. 5.17b), and erroneously picked a lot of 'junk' particles; random noise interpreted as particles during particle picking. High signal background can be caused by the high molar-excess of Fab in the sample mixture, or by heterogeneity or other contaminants within the sample.

To test whether the high signal background was due to the high Fab molar excess, GO-coated grids were prepared with a reduced Fab molar excess of 1.2-fold the molar concentration of bRSVF monomers (36 µM B4 mixed with 10 µM bRSVF). GO-coated 300 mesh self-wicking Copper grids were glow discharged for 30 s, before 100 nL application of bRSVF/B4 complex as a single stripe, and subsequent vitrification into liquid ethane after a 150 ms plunging delay (Fig. 5.18). Specimens were screened using a 200 keV Glacios electron cryomicroscope within eBIC at DLS (Table 2.13).

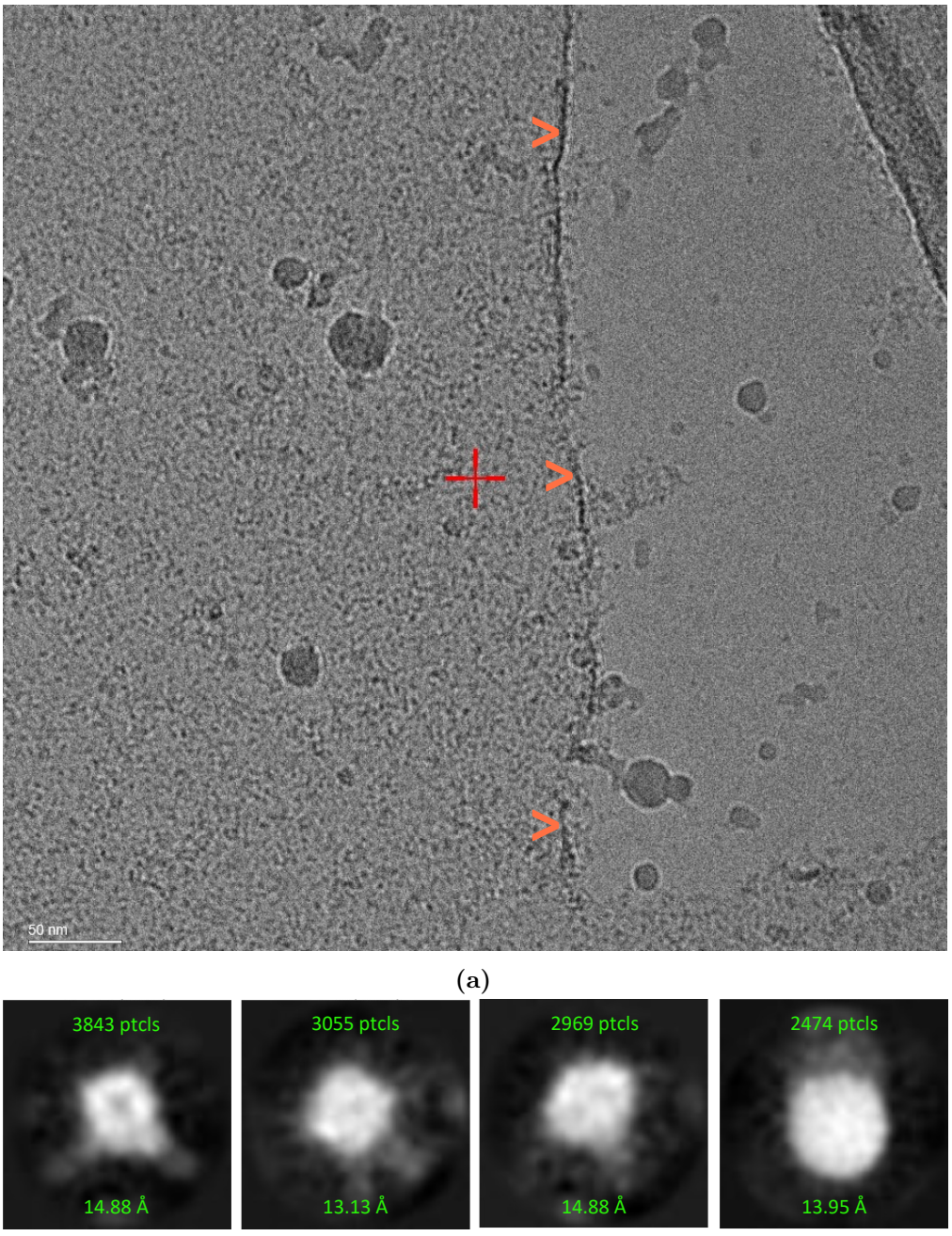
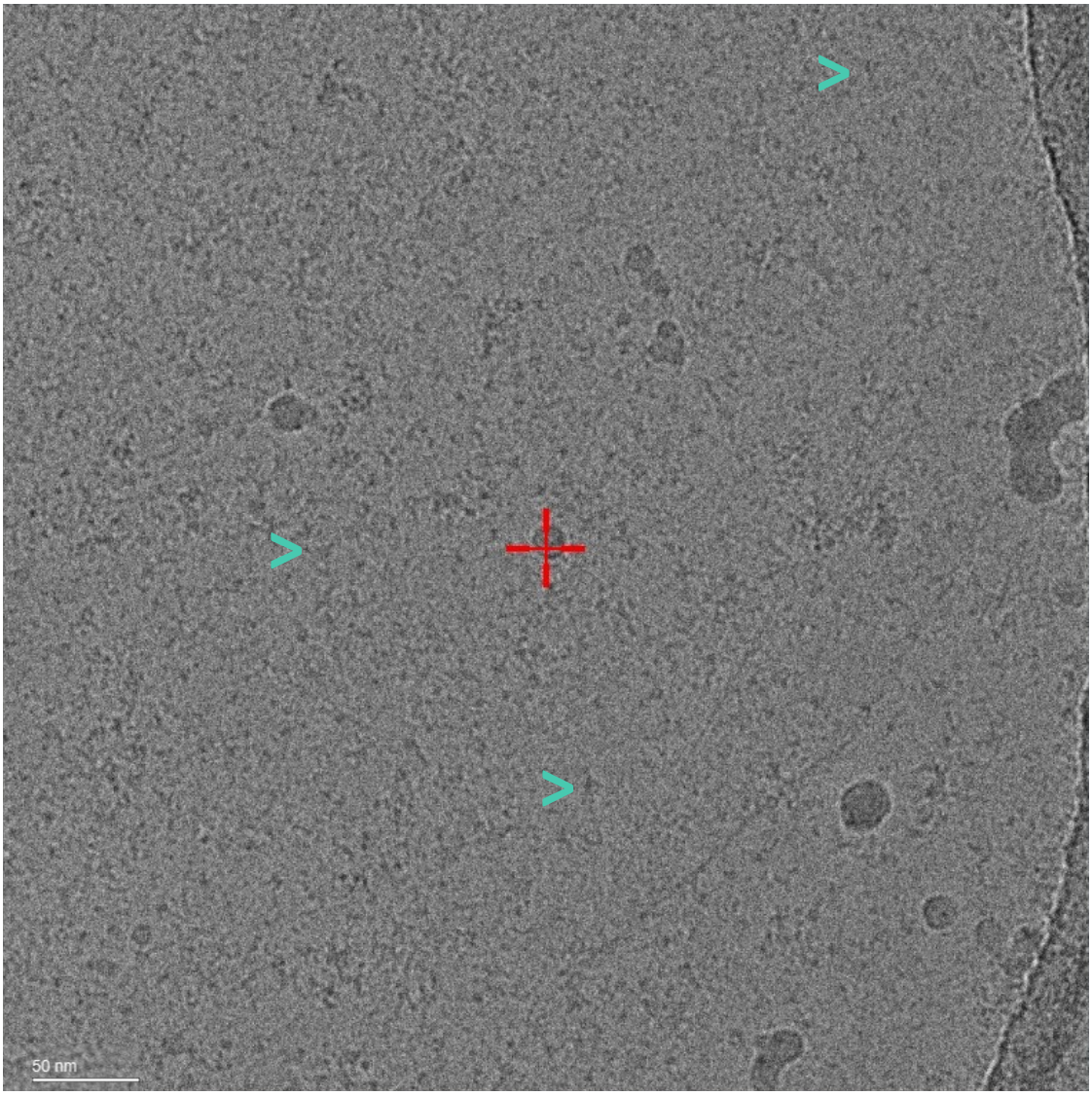
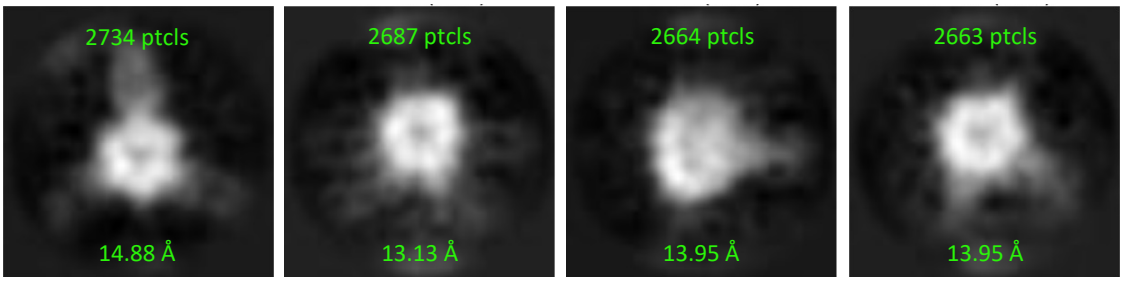


Figure 5.17: Electron cryomicroscopy of Fab B4 in complex with bRSVF homotrimers. Specimen plunge frozen by Chameleon. (5.17a) Example micrograph, with graphene oxide tear indicated by orange arrowheads. (5.17b) Representative 2D class averages with indicated population and estimated resolution.



(a)



(b)

Figure 5.18: Electron cryomicroscopy of Fab B4 in complex with bRSVF homotrimers, with reduced Fab excess. Specimen plunge frozen by Chameleon. (5.18a) Example micrograph, with graphene oxide tear indicated by orange arrowheads. (5.18b) Representative 2D class averages with indicated population and estimated resolution.

Even with a reduced Fab excess, background signal is still high on the GO-coated grids (Fig. 5.18a), and junk particles are still erroneously picked (Fig. 5.18b). In this instance, junk particles (2D class average #3) had a greater population than the suspected side-view particles of 2D class average #4. Although the background signal issue has been improved, the prevalence of junk particles would still make downstream processing difficult.

5.5.4 Repurification of F protein complexed with Fab B4

To improve sample homogeneity on GO-coated self-wicking cryo-EM grids, the bRSVF/B4 complex was purified by SEC to remove any contaminants, degradation products, or self assembled higher order multimer species. Fab B4 was mixed with bRSVF at a molar excess of 1.2-fold the bRSVF monomer molar concentration to a total volume of 130 µL, and manually injected to a Superdex 200 Increase 10/300 GL size exclusion chromatography column pre-equilibrated with pH 7.40, 0.45 µm-filtered PBS, loaded to an ÄKTA™ Pure chromatography system.

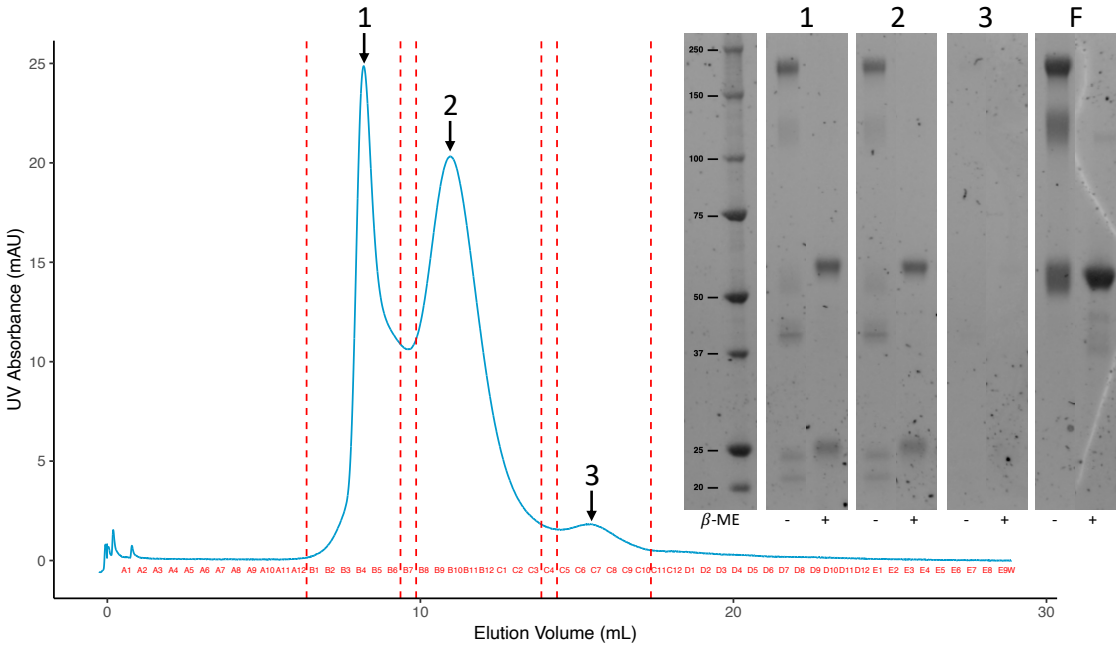


Figure 5.19: Size exclusion chromatography of bRSVF/B4 complex. 280 nm Absorbance trace of gel filtrate. Indicated fractions were analysed by SDS-PAGE (inset). F - lone bRSVF for comparison. Samples prepared without (left) and with (right) 10 % β-ME. Band migration compared with Bio-Rad Precision Plus Protein Standards.

Three peaks of UV (280 nm) absorbance were observed for the bRSVF/B4 complex (Fig. 5.19). SDS-PAGE (Section 2.7.2) analysis gave the same band migration profile for fraction #1 and #2. In non-reducing conditions (left lane), bands mirroring those observed for lone bRSVF ('F'), with bands appearing at approximately 200 kDa, 130 kDa, and (weakly) 60 kDa. The band around 200 kDa likely corresponds to the bRSVF homotrimer, but the weak, smeared band around 130 kDa is not explained. The ~60 kDa band becomes more prominent upon addition of 10 % β-M (right lanes), and corresponds to bRSVF monomers. Additionally, bands typical of Fab are observed at 40 kDa, 25 kDa and 20 kDa in the non-reduced lanes (left), which resolve into a single band at 25 kDa in reducing conditions. The SDS-PAGE of the SEC-purified bRSVF/B4 complex indicate that both F protein and Fab B4 are present in fractions #1 and #2 of the SEC UV absorbance peaks. Fraction #3 did not produce an observable band by SDS-PAGE.

Samples were mixed with SDS and heated to 95 °C prior to loading to the gel, likely disrupting any complexes or higher order species. The SEC elution volume (indicated by the UV absorbance peak) can be used to infer species molecular weight if the elution volumes of SEC protein standards are known (Table B.1 and Fig. B.1). The average distribution constant K_{av} is calculated using the elution volume V_e as described previously (Section 2.7.1), and used to infer the species molecular weight (Fig. E.1 and Table 5.6).

Table 5.6: Inferred species molecular weight by SEC elution volume.

Peak	Elution Volume (V_e) (mL)	Avg. Distribution Constant (K_{av})	Inferred Molecular Weight (Da)
1	8.19	0.0067	N/A
2	10.96	0.2379	366,155
3	15.45	0.6127	32,185

Peak #2 broadly correlated to the bRSV F homotrimer, with 3 Fabs bound (theoretical MW of 332.2 kDa), whilst peak #3 correlates roughly with the molar mass of lone Fab (theoretical MW of 53.2 kDa). Peak #1 eluted before the highest molecular weight marker, and consequently could not be reliably inferred. Peak

#1 eluted at 8.19 mL, nearly eluting in the column void volume of 8.11 mL. SEC separates species by hydrodynamic radius (which roughly correlates with molecular weight in globular proteins). Where the bRSVF/B4 complex was observed with a rod-like shape with Fabs equally decorated perpendicular around the F protein exterior, there will be some systemic error when using SEC elution values to estimate molecular weight, due to the incomplete hydrodynamic sphere. Additionally, if peak #1 is assumed to be self-assembled into a multimeric species, the multimer will have a much larger hydrodynamic radius, and may well be eluted near the void volume. It is likely then, that the first peak is a higher order multimer than the second bRSVF/B4 SEC peak, but this can only be confirmed by further analysis, such as by Förster resonance energy transfer (FRET) or cryo-EM.

5.5.4.1 Characterisation of purified bRSVF/B4 complex fractions by SEC-MALS

To resolve why two overlapping SEC UV absorbance peaks were observed, despite having the same composition by SDS-PAGE, SEC-MALS was performed. Each fraction was pooled and spin-concentrated separately using 100 kDa MWCO Amicon® centrifugal filter units (Section 2.6.5 and Table 5.7).

Table 5.7: Summary of bRSVF/B4 complex spin-concentration. [1]Time taken for spin-concentration to achieve stated concentration.

Fraction:	1	2
Pooled fractions:	B01:B06	B08:C01
Pooled volume (mL):	3.0	3.0
Concentration (mg/mL):	1.962	2.322
Time Taken[1] (minutes):	8	8

To conserve material, each fraction was diluted 1:4 in 0.45 µm-filtered PBS at pH 7.40, in total volumes of 100 µL and auto-injected to a Superdex 200 Increase 10/300 GL size exclusion chromatography column pre-equilibrated with pH 7.40, 0.45 µm-filtered PBS, loaded to an ÄKTA™ Pure chromatography system connected in series to a multi-angle light scattering detector and a refractometer (Section 2.7.8).

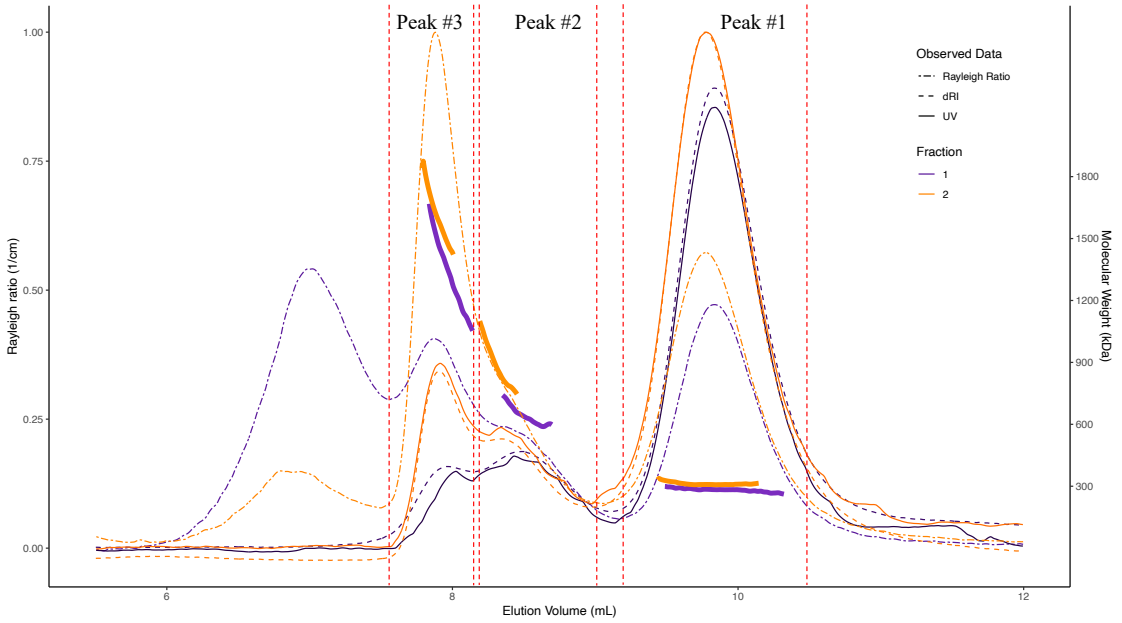


Figure 5.20: Size-exclusion chromatography with multi-angle light scattering characterisation of purified bRSV F homotrimers in complex with Fab B4. Calculated molar masses (bold) for the protein peaks are overlaid measured data traces for Rayleigh ratio (dot-dash), differential refractive index (dRI)(dashed), and UV absorbance (solid). Data measured using a Dawn Heleos-II multi-angle light scattering detector (Wyatt Technologies) and an Optilab T-rEX refractometer (Wyatt Technologies). Data were analysed using the ASTRA® software v5.3 (Wyatt Technologies) using the Zimm method.

The first SEC-MALS peak (last eluted, around 9.75 mL) for both purified fractions eluted at a very similar position, and were characterised with molar mass moments of 2.825×10^5 g/mol $\pm$ 1.392×10^3 g/mol for fraction #1 and 3.118×10^5 g/mol $\pm$ 8.512×10^2 g/mol for fraction #2 (Fig. 5.20 and Table 5.8). These values lie between the theoretical molecular weights for bRSVF decorated with 2 Fabs (279.0 kDa) and 3 Fabs (332.2 kDa) as calculated using ProtParam (Swiss Institute of Bioinformatics) [180], and may therefore represent a dynamic system at equilibrium between a smaller, and a larger species.

Upon the addition of Fab B4, the middle peak observed during SEC-MALS characterisation of lone bRSVF (Fig. 5.9), thought to be due to dimers of F protein homotrimers (Table 5.2), appears to largely have been resolved for both purified bRSVF/B4 complex fractions.

The 3rd peaks, eluting around 7.90 mL, gave the greatest difference between the

fractions. These peaks are much larger for fraction #2, but still present in fraction #1, indicating a greater proportion of self assembled multimer species in fraction #2 than in fraction #1, at odds with the conclusions drawn from inferring species molecular weight by comparison to SEC protein standards (Table 5.6). This disparity may be explained by observing the gradient (described by the polydispersity value) of the calculated molecular weights as a function of the elution volume. For all species, the molecular weight decreases as elution volume increases. Species will experience an increasing dilution effect the longer it takes to pass the SEC column. Therefore this may be explained as a dynamic, concentration-dependent self assembly, actively dissociated down the SEC column. Furthermore, during spin-concentration of the fractions, fraction #2 achieved a higher concentration (2.322 mg/mL) than fraction #1 (1.962 mg/mL) (Table 5.7). As a consequence, fraction #2 may then have re-assembled to a greater multimer extent than fraction #1. When characterised by SEC-MALS, the higher concentration fraction #2 demonstrated a greater degree of self assembly than the lower concentration fraction #1, despite eluting after fraction #1 during the initial purification of the bRSVF/B4 complex.

For both fractions, scattering was also observed at the column void volume around 7.0 mL, indicating the presence of trace amounts of highly aggregated species that did not penetrate the SEC resin pores, interestingly this species was proportionally larger for fraction #1 than fraction #2 (Table 5.8).

Note that peaks were selected for molar mass moment calculation for each fraction separately by reviewing the UV (280 nm) absorbance peaks, and will consequently have some lateral shift between them, dependent on the specific elution volumes of each species in each fraction. Protein conjugate analysis was performed, but not included in Figure 5.20 for clarity. For all species of both fractions, conjugate contributed insignificantly to molar masses (Table 5.8), as observed in lone bRSVF characterisation (Fig. 5.9 and Table 5.2).

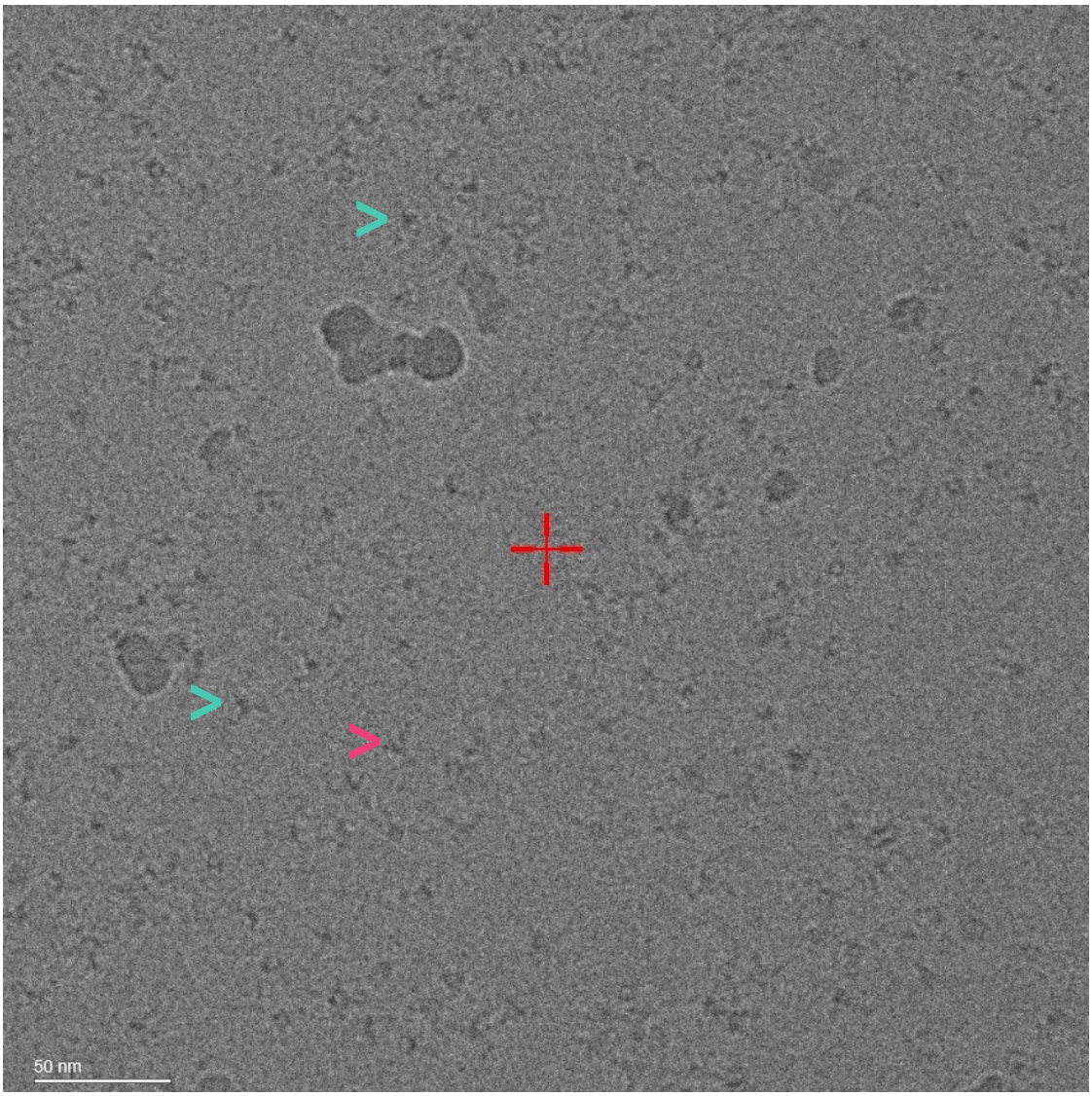
Table 5.8: SEC-MALS characterisation of purified bRSVF/B4 complex fractions.

	Peak #1	Peak #2	Peak #3
		Fraction #1	
Masses			
Total (μg)	13.380	2.315	1.420
Protein (μg)	13.830	2.315	1.306
Conjugate (μg)	-	-	-
Molar Mass Moments			
Total (g/mol)	2.933×10^5	6.836×10^5	1.460×10^6
	($\pm$ 8.791 %)	($\pm$ 0.891 %)	($\pm$ 1.343 %)
Protein (g/mol)	2.933×10^5	6.836×10^5	1.249×10^6
	($\pm$ 1.306 %)	($\pm$ 2.520 %)	($\pm$ 1.466 %)
Conjugate (g/mol)	n/a	n/a	1.102×10^5
			($\pm$ 9.826 %)
Fractional Moments			
Protein	1.088	1.116	0.918
	($\pm$ 0.112 %)	($\pm$ 0.433 %)	($\pm$ 0.535 %)
Polydispersity			
Total (Mw/Mn)	1.001	1.011	1.073
	($\pm$ 1.875 %)	($\pm$ 3.592 %)	($\pm$ 1.930 %)
Protein (Mw/Mn)	1.001	1.012	1.008
	($\pm$ 1.852 %)	($\pm$ 3.649 %)	($\pm$ 2.099 %)
Conjugate (Mw/Mn)	n/a	n/a	n/a
		Fraction #2	
Masses			
Total (μg)	14.890	1.555	1.929
Protein (μg)	14.890	1.555	1.929
Conjugate (μg)	-	-	-
Molar Mass Moments			
Total (g/mol)	3.510×10^5	1.065×10^6	1.880×10^6
	($\pm$ 0.818 %)	($\pm$ 0.886 %)	($\pm$ 0.645 %)
Protein (g/mol)	3.510×10^5	1.065×10^6	1.880×10^6
	($\pm$ 0.818 %)	($\pm$ 0.886 %)	($\pm$ 0.645 %)
Conjugate (g/mol)	n/a	n/a	n/a
Fractional Moments			
Protein	1.174	1.300	1.244
	($\pm$ 0.079 %)	($\pm$ 0.286 %)	($\pm$ 0.193 %)
Polydispersity			
Total (Mw/Mn)	1.000	1.015	1.007
	($\pm$ 1.150 %)	($\pm$ 1.196 %)	($\pm$ 0.870 %)
Protein (Mw/Mn)	1.001	1.012	1.005
	($\pm$ 1.156 %)	($\pm$ 1.263 %)	($\pm$ 0.912 %)
Conjugate (Mw/Mn)	n/a	n/a	n/a

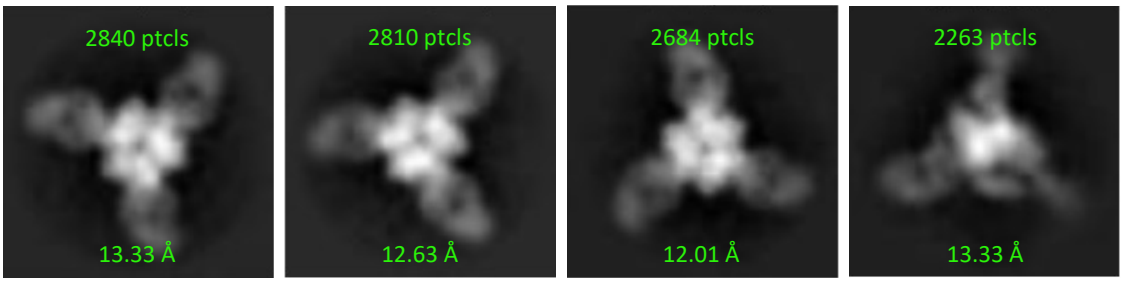
5.5.4.2 cryo-EM specimen preparation

Purification of the bRSVF/B4 complex by SEC should aid the observed cryo-EM specimen heterogeneity (Figs. 5.17 and 5.18) as elution peaks constituting just the bRSVF/B4 complex can be pooled and concentrated (Table 5.7). Screening the purified complexes by cryo-EM is therefore a useful step toward high-resolution cryo-EM data collection, and eventual structural determination. Furthermore, observing the extent of higher-order species formation of each fraction by cryo-EM would be informative to the concentration-dependent multimer self assembly theory for bRSVF/B4.

The purified bRSVF/B4 complex fractions were applied without dilution to GO-coated 300 Mesh Self-Wicking Copper grids (Section 2.8.4.5) using the Chameleon vitrification robot as single 100 nL stripes. Given the different sample concentrations, grid preparation parameters were optimised for each fraction separately. The best grid was achieved for fraction #1 following a glow discharge of 30 s, and a plunging delay of 180 ms, whereas the best parameters for fraction #2 were a glow discharge of 40 s, followed by a plunging delay of 130 ms. Specimens were screened using a 200 keV Glacios electron cryomicroscope within eBIC at DLS (Table 2.13).

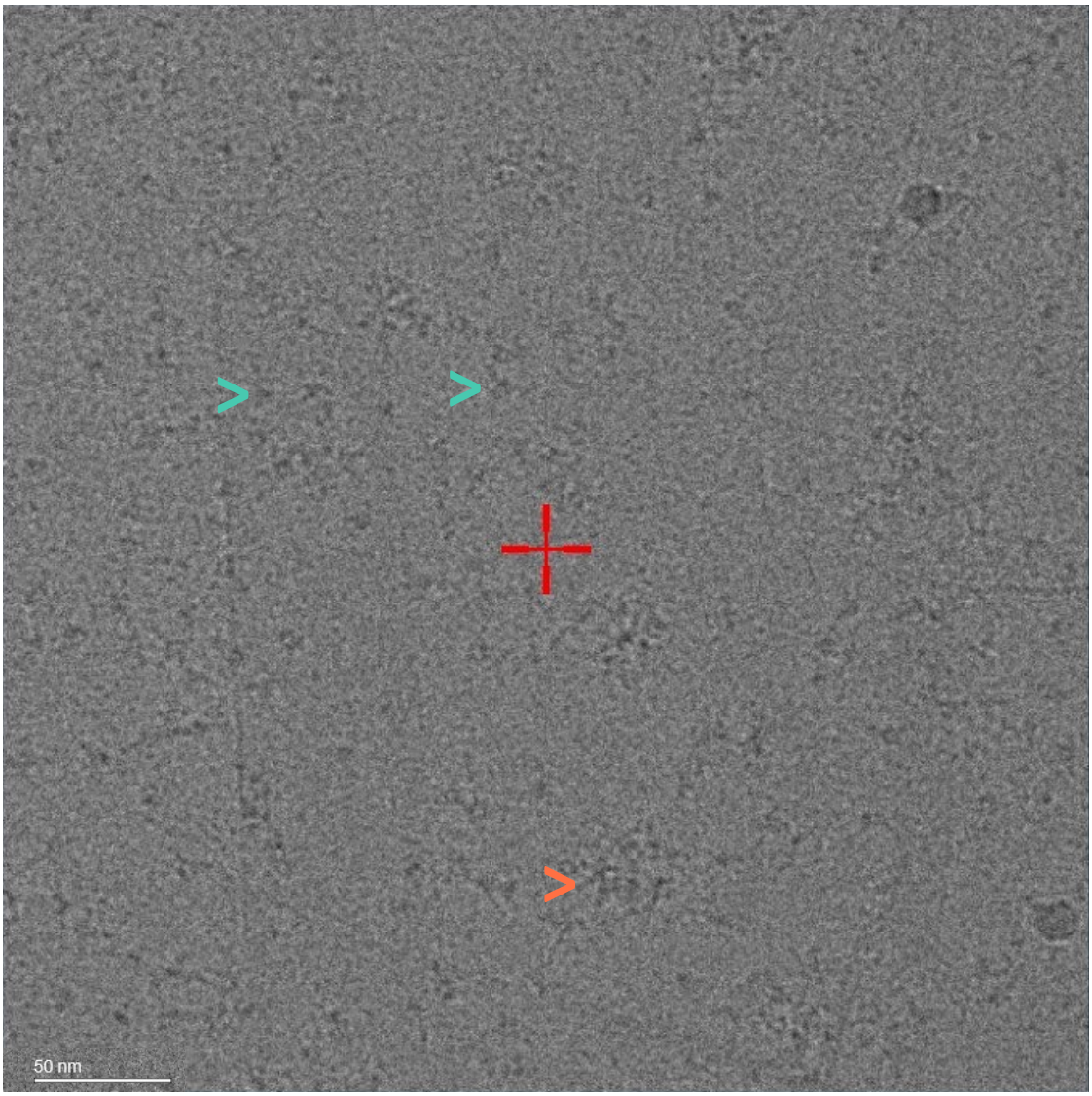


(a)

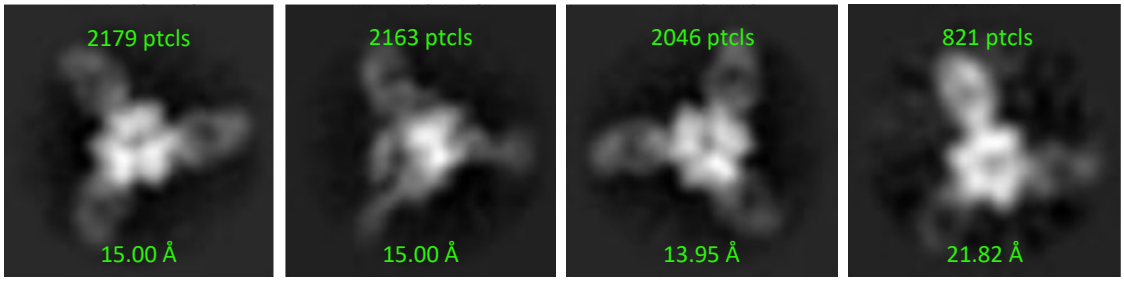


(b)

Figure 5.21: Electron cryomicroscopy of Fab B4 in complex with bRSVF homotrimers. Specimen plunge frozen by Chameleon after purification as a complex; first fraction. (5.21a) Example micrograph, with example top-view particles marked by green arrowheads, and suspected side-view indicated by pink arrowhead. (5.21b) Representative 2D class averages with indicated population and estimated resolution.



(a)



(b)

Figure 5.22: Electron cryomicroscopy of Fab B4 in complex with bRSVF homotrimers. Specimen plunge frozen by Chameleon after purification as a complex; second fraction. (5.22a) Example micrograph, with top-view particles marked by green arrowheads, and suspected aggregation indicated by orange arrowhead. (5.22b) Representative 2D class averages with indicated population and estimated resolution.

5.5.4.3 Results

In micrographs collected for both purified bRSVF/B4 complex fractions #1 (Fig. 5.21a) and #2 (Fig. 5.22a), there was a much reduced background, and the sample appeared much more homogeneous than cryo-EM grids prepared prior to purification as a complex (Fig. 5.18a). This is supported by the more successful particle picking. For both fractions, the particle picking algorithm picked proportionally many more target particles than junk particles, reflected in strong representation of 2D class averages representing the bRSVF/B4 complex (Figs. 5.21b and 5.22b). The fraction #1 grid appeared uniform at different sites of the grid, but the fraction #2 grid appeared to show signs of aggregation in places (Fig. 5.22a; orange arrowhead), however this may just be artefactual of the higher sample concentration, as discussed previously (Section 5.5.4.1).

Top-down views of the bRSVF/B4 complex were the most highly populated in both datasets, but side views were also observed with strong particle counts. All top-down views of fraction #1 showed bRSVF decorated with 3 Fabs. Intriguingly, several top-down views of fraction #2 appeared to show a reduced occupancy of Fab in one or two binding sites (Fig. 5.22b 2D class average #4). This may be further evidence of the bRSVF/B4 complex existing as a dynamic, diverse population with varying levels of Fab binding sites occupied. Given the inverse correlation with multimer self assembly or aggregation, it may be true that Fab binding to bRSVF protects the F protein from higher order species formation.

Whilst concentration-dependent multimer self assembly of bRSVF/B4 complexes is one suggestion, the correlation of multimer self assembly with lower Fab binding occupancies suggests a more complicated situation. This correlation is consistent with the disappearance of high order species in AUC sedimentation coefficient peaks (Fig. 5.12d) upon addition of Fab (Fig. 5.13), and the reduced populations of higher order species observed during SEC-MALS when comparing lone bRSVF (Fig. 5.9 and Table 5.2), to complexed bRSVF (Fig. 5.20 and Table 5.8). Heterogeneous refinement could be performed on the *ab-initio* 3D reconstruction to separate species demonstrating varying degrees of Fab decoration, and may reveal clues to

the mechanism behind bRSVF multimer self assembly; whether Fab binding occurs at a site to protect against self-association, or whether the stabilised prefusion bRSVF monomer demonstrates a conformational shift that governs the potential for Fab binding or multimeric formation.

5.5.5 High resolution bRSVF/B4 data collection

The strong representation of both top-down and side views, as well as improved particle density warranted a high resolution cryo-EM data collection for bRSVF in complex with Fab B4.

Purified bRSVF/B4 complex from the first fraction was once again applied to GO-coated self-wicking cryo-EM grids using the chameleon robot as a single 100 nL stripe. A glow discharge of 35 s, and a wicking delay of 130 ms produced the best grid. Data were collected at 300 keV with a nominal magnification of 105,000x, resulting in a pixel size of 0.83 Å/px using M07; one of the Titan Krios electron cryomicroscopes of eBIC, equipped with a Gatan GIF Quantum K3 detector. 50 frames were collected over 2 s, with an administered dose rate of 17.5 e^-/Å/s. A defoci range of −3.3 µm to −0.9 µm in 0.3 µm increments was used to sample all spatial frequencies (Section 2.8.4.10). An objective aperture of 100 µm was used with a C2 aperture of 50 µm and EV slit of 20 eV.

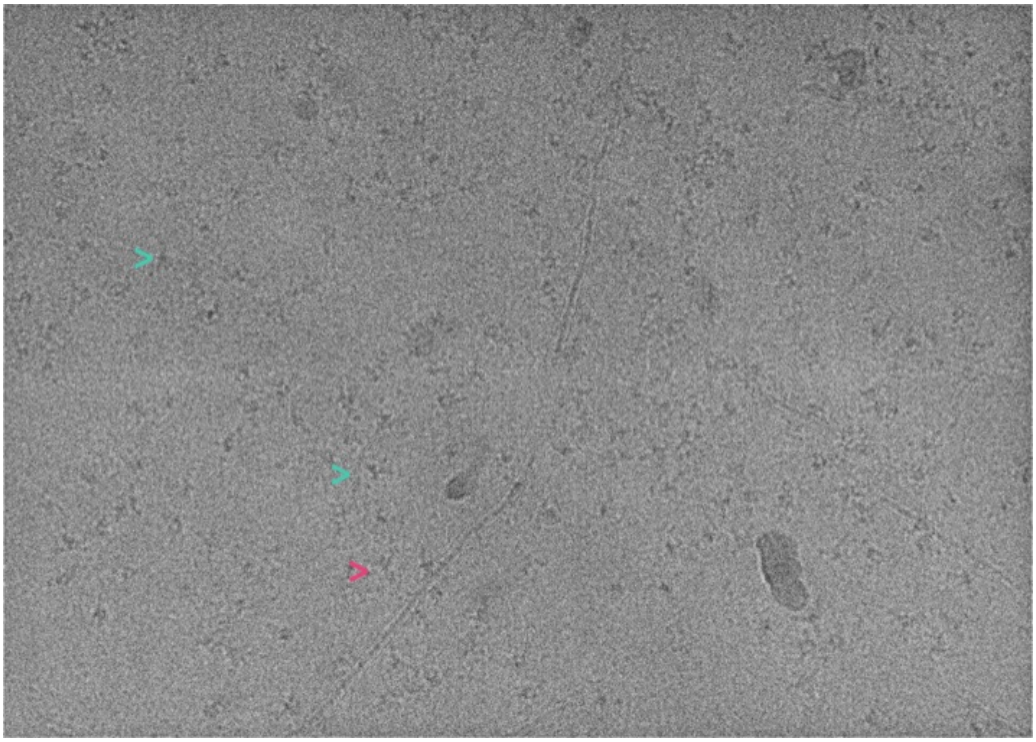
A total of 18,211 exposures were captured, which were manually curated to 15,080 by application of filters (Table 5.9). 933,990 particles were selected for downstream processing (Table 5.10) following template particle picking.

Table 5.9: High resolution bRSVF/B4 cryo-EM exposure curation.

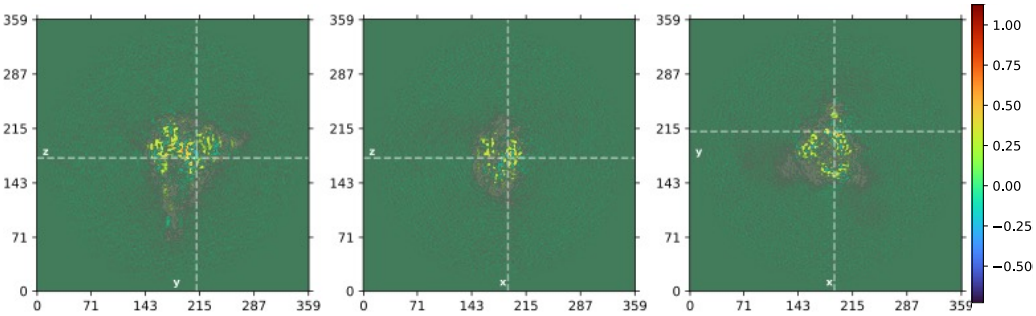
Initial Exp.	Defocus average (Å)	CTF fit (Å)	Rel. ice thickness	Max. in-frame motion	Total motion (px)	Defocus range (Å)	Final Exp.
18,211	5048 - 39046	2.047 - 8.640	1.043 - 1.142	0.121 - 2.609	4.65 - 4.05	0 - 9771	15,080

Table 5.10: High resolution bRSVF/B4 cryo-EM particle picking.

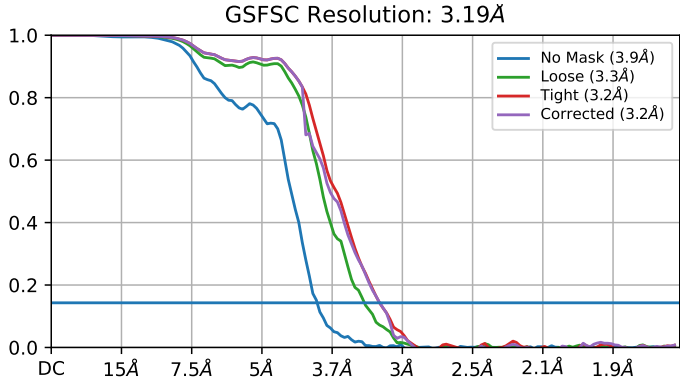
Template diameter (Å)	Box (px)	Bin (px)	Observed particles	Picked particles
250	256	64	4,495,036	933,990



(a)



(b)

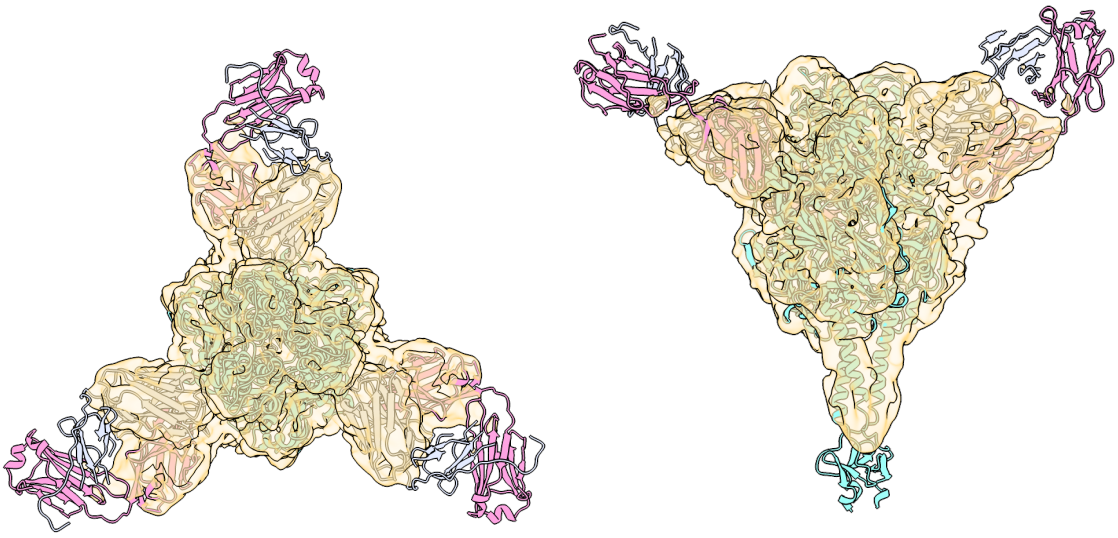


(c)

Figure 5.23: High resolution electron cryomicroscopy of Fab B4 in complex with bRSVF homotrimers. (5.23a) Example micrograph, with particles marked by green arrowheads, Suspected side view indicated by pink arrowhead. (5.23b) Central slices from density maps after final round of refinement. (5.23c) Autotightened GSFSC plot indicating determined resolution.

5.5.5.1 Preliminary structural analysis of the bRSVF/B4 complex

Iterative non-uniform and local refinements of the observed density using 699,888 particles in C3 symmetry resulted in an indicated 3.19 Å map of the bRSVF/B4 complex, into which deposited crystal structures for B4 (PDB ID: 6QN9 [160]) and bRSVF (PDB ID: 5TDG [175]) were fitted (Fig. 5.24).



(a) Top-down view **(b)** Side view

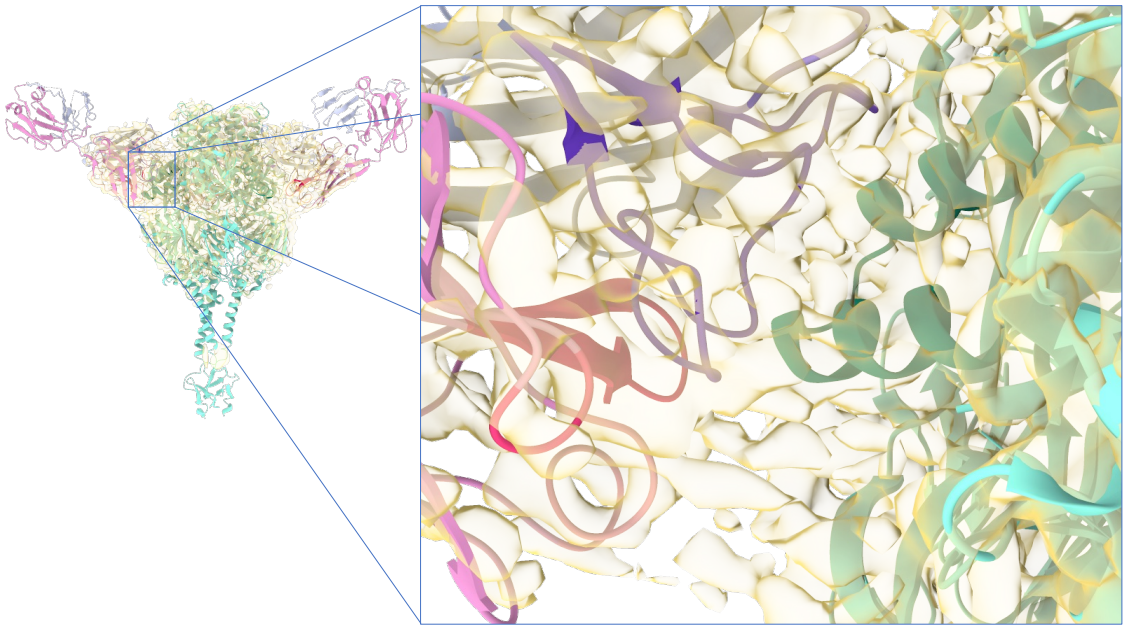
Figure 5.24: B4 and bRSVF crystal structures fitted to cryo-EM density map. Crystal structure of Fab B4 (pink and white cartoon; PDB ID: 6QN9 [160]) and bRSVF (cyan cartoon; PDB ID: 5TDG [175]) fitted into observed cryo-EM density map (translucent orange) Gaussian filtered to a B-factor of 500 for clarity. Fig. 5.24a and Fig. 5.24b are related by a 90° rotation perpendicular to the plane of the figure. Figures prepared using ChimeraX [257].

As anticipated (Section 5.1.5), additional density was observed near antigenic site II [358], comprising the α-6 and α-7 helices of the F1 subunit, which may be attributable to Fab B4. However, this additional density extends to the prefusion specific antigenic site Ø [358] of the adjacent F protein monomer. The significance of Fab B4's proximity to antigenic site Ø is unclear without atomic resolution of the mechanism of interaction.

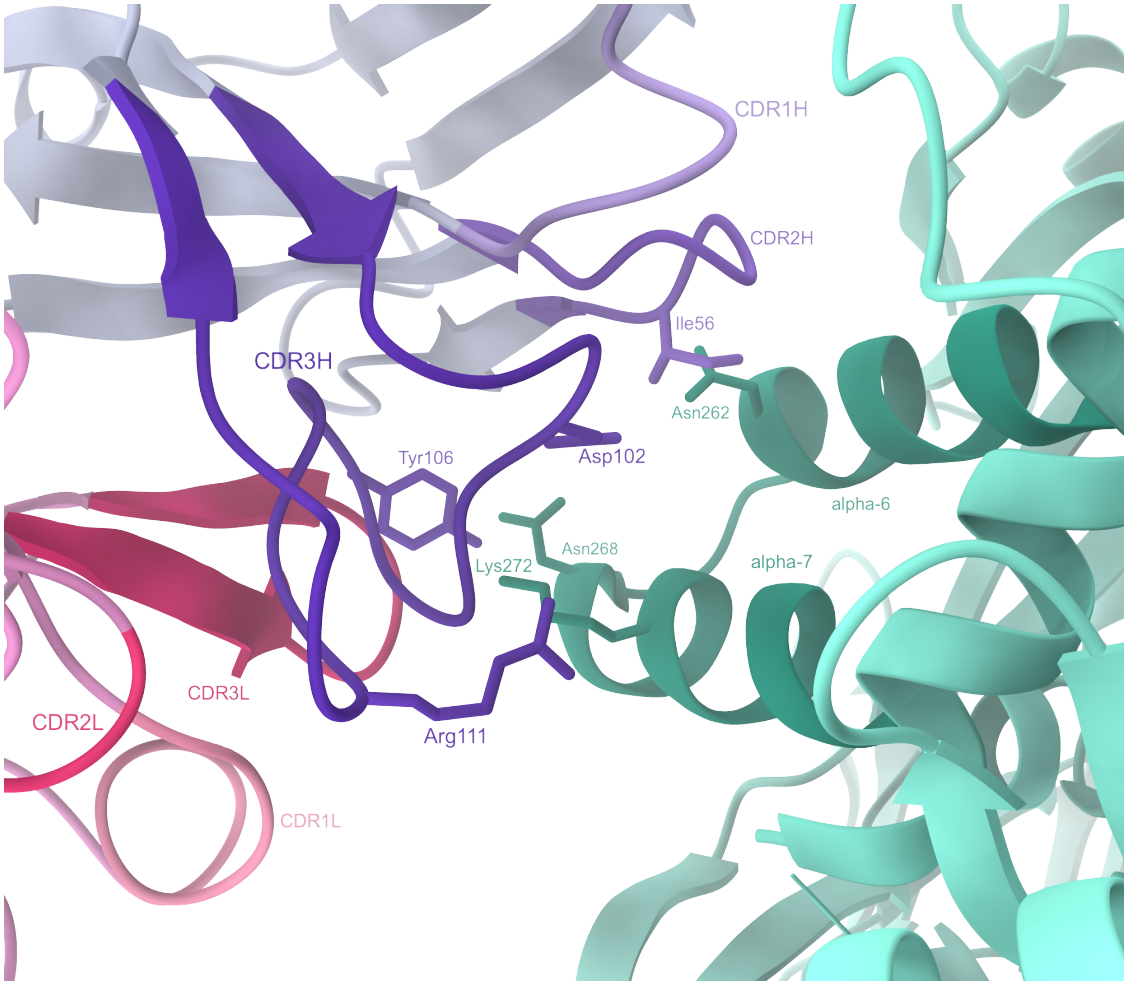
Within the time frame for this project, near-atomic resolution of the bRSVF/B4 interface could not be determined. Interestingly, upon inspection of the fitted crystal structures (Fig. 5.25), the B4 element in closest proximity to bRSVF (7.2 Å

minimum backbone separation) is the C-terminus of the CDR2H loop, close to the F protein α-6 helix (Fig. 5.25b). This is interesting, as the CDR2H was not observed to change conformation upon chain-exchange of the light chain, unlike the CDR3H. Indeed, the CDR2H of B4 (PDB ID: 6QN9 [160]) and B4* (PDB ID: 6QN7 [160]) can be superimposed with an RMSD of 0.359 Å with the only differences being side chain flips of Ile56 and Ile57, and a slight rotation of Asn53.

The CDR3H was modelled into density above the C-terminus of bRSV F protein α-7 (Fig. 5.25b), with a minimum backbone separation of 8.5 Å, and no evidence of any conformational changes upon binding. In the bRSVF crystal structure (PDB ID: 5TDG [175]), the side chains of Asn268 and Lys272 of α-7 protrude directly towards the fitted CDR3H, and may therefore be involved in the interaction (Fig. 5.25b).



(a)



(b)

Figure 5.25: Interacting residues of B4 with bRSVF. *Figure legend is continued on the next page...*

Figure 5.25: Interacting residues of B4 with bRSVF. *Figure legend is continued from the previous page.* Molecular overview of the bRSVF interaction with Fab B4 by cryo-EM. With (5.25a), and without (5.25b) electron density map thresholded to 0.190 in transparent yellow overlaid the Fab B4 (PDB ID: 6QN9 [160]) and bRSVF (PDB ID: 5TDG [175]) crystal structures fitted to the observed cryo-EM electron density map. CDR loops coloured as follows; CDR1H - light purple, CDR2H - purple, CDR3H - deep purple, CDR1L - light pink, CDR2L - hot pink, CDR3L - fuchsia. F protein α-6 and α-7 helices are indicated in forest green. Residues orientated directly towards each interacting species are shown in stick notation. Figure prepared using ChimeraX [257].

5.6 Conclusions and Discussion

5.6.1 Recalcitrant protein purification

Bovine respiratory syncytial virus F protein could not be produced efficiently by transient transfection of Expi293™ cells. For reference, the initial paper describing the development of stabilised prefusion conformation bRSV F protein achieved a yield of 2.80 mg/L by transient transfection [175]. Due to the significant material demand for structural study, alternative systems were explored. A stable cell line was created, and purification by IMAC-SEC gave a yield of 4.12 mg/L.

5.6.2 bRSVF exists in a dynamic concentration-dependent equilibrium

Subsequent to spin-concentration of bRSVF, the presence of higher order multimeric species was observed by SEC-MALS, AUC, and cryo-EM experiments. These multimeric species appear to dissociate once concentrated bRSVF is diluted, or Fab B4 is added. In all analyses, bRSVF was observed as a homotrimeric species by molecular weight, however these experiments have assumed bRSVF was present as a stable trimeric species.

Native F protomers associate with a weak intermolecular interaction to form the trimer, however these are insufficient to drive trimerisation alone. Instead, engineered F protein antigens usually include a trimerisation domain; in this thesis, the engineered DS2-v1 variant of bRSVF included the Foldon trimerisation domain, and is further aided by two additional interprotomer disulphide bonds to form bRSVF homotrimers [175]. At high concentrations, these F trimers may be

dissociating to three protomers tethered by the Foldon trimerisation domain (as observed upon binding of CR9501 to stabilised prefusion conformation human RSV F variant DS-Cav1, which also included Foldon and two additional interprotomer disulphides [361]), and this 'splayed-out' structure may be more prone to multimeric formation than the bRSVF homotrimer.

Negative stain EM of lone bRSVF may provide evidence to support this theory, but at present; splayed-out, tethered bRSVF protomers has not been observed by single particle cryo-EM. Stabilised prefusion hRSV F antigen also exhibits a small degree of flexibility within the 'closed' conformation [361], reminiscent of HIV Env 'breathing' [362]. This more subtle flexibility may also contribute to multimer formation.

Molecular dynamics simulations using the crystal structure of the stabilised prefusion bRSV F protein antigen (PDB ID: 5TDL [175]) may provide an explanation to bRSVF multimer self assembly, and it may be possible to engineer a solution to prevent multimer formation without impacting the bRSVF epitopes, again contributing to veterinary vaccine development. Furthermore, the B4 epitope is near an interprotomeric interface, and binding of Fab B4 may therefore contribute a trimer stabilising effect.

5.6.3 Variable degrees of Fab binding

An interesting observation is the reduced occupancy of Fab B4 binding. The molecular weight of bRSVF/B4 as a complex was characterised by SEC-MALS as a value intermediate between the expected molecular weights for bRSVF decorated by 2 and 3 Fabs. As a lone observation, one could dismiss this theory as variance, however the observation of 2D class averages showing reduced occupancy at Fab B4 binding sites adds weight to the theory that the B4 interaction with bRSVF is a dynamic one, despite a reported affinity of 86 nM [160].

5.6.4 Opportunity for structural determination of bRSVF in complex with other Fabs

Previous study [160] has provided that chain-exchange results in perturbed CDR3H architectures (Section 5.1.6) and affinities for bRSVF (Section 5.1.5), yet structural understanding of this effect is lacking. Examination of such would be informative to understanding how big an impact heavy-light chain pairing makes to antigen recognition, and may therefore inform on the observed preferential pairing of bovine antibodies (Section 1.4).

This chapter set out work optimising the specimen preparation for structural determination of the bRSVF in complex with bovine Fabs B4, B13, and B13*; in order to observe whether chain-exchange results in idiotypic differences in the epitope being targeted by these antibodies. The low reported affinity (12 µM) for Fab B4* implies structural determination of bRSVF in complex with B4* would likely be more difficult as complexes would be harder to co-purify, and may dissociate more easily than a complex comprising Fab B4 and bRSVF.

In this study, the preparation of cryo-EM grids by Chameleon vitrification robot has alleviated the preferential orientation issue, whilst graphene oxide coating the cryo-EM grids and purification as a complex has improved the sample homogeneity, and reduced background noise, improving the signal contrast to aid particle picking algorithms. Together, these efforts offer a protocol for the structural determination of the interaction of bovine RSV F protein in complex with bovine antibodies B4, B13, and B13* as Fabs. A high-resolution data collection for bRSVF in complex with B4 was performed on the 21st October 2022, however due to timing constraints, a high resolution structure could not be fully incorporated in to this thesis, but the ongoing data processing, and examination of bRSVF in complex with B13 and B13* will be continued by the RFI Protein Production group.

The saddest aspect to life right now, is that science gathers knowledge faster than society gathers wisdom.

— Isaac Asimov, biochemist and fiction author

6 General Discussion and Future Work

Contents

6.1 Thesis purpose 289
6.2 Conclusions from chain-exchanged ultralong antibodies chapter 290
6.3 Conclusions from FMDV-specific chapter 293
6.4 Conclusions from the preliminary bRSVF cryo-EM structural chapter 295
6.5 Summary and final remarks 296

Tables

6.1 Summary of presented structural antibody work 290

6.1 Thesis purpose

In contrast to other mammals, cattle do not rely upon recombinational diversity of V(D)J gene segments to establish a diverse functional immune repertoire.

The *Bos taurus*, and *Bos indicus* genomes contain limited germline libraries of IgH V, D, and J [37, 38] (Section 1.2.1), Igλ [32, 61, 153] and Igκ [152] V gene segments (Section 1.3.1). Diversity is further restricted by the high sequence homology of IgHV genes [34, 54, 64], IgHD genes [53] (Section 1.2.3), IgLV genes [153] (Section 1.3.2) and the ubiquitous expression of IgHJ2-4 [54, 58] (Section 1.2.2) and a single Jλ-Cλ junction [51] (Section 1.3.2).

Cattle employ two key strategies in escaping these germline constraints; somatic hypermutation in the absence of antigen exposure [39, 77] (Sections 1.2.4 and 1.2.5.1), and the evolution of an 'ultralong' CDR3H with lengths exceeding 48 residues [31] (Sections 1.2.5, 1.2.5.3 and 1.2.5.4). The ultralong CDR3H has attracted much scientific attention, owing to its novelty in the paratopic space, wherein a small cysteine knob is presented atop a β-strand ribbon [78, 82, 83, 85], that can interact with antigen independently of the rest of the antibody [138, 139], sharing a structural and functional likeness with inhibitory cysteine knot (ICK) proteins [90, 91, 94–99, 101, 103, 104].

The ultralong CDR3H scaffold offers a hypervariable canvas to build an antibody paratope, with remarkable accommodation of inter-cysteine loops varied by length and composition [54]. On the flip side, ultralong CDR3H are exclusively accommodated by light chains encoded by IgLV1-47, or IgLV1-67 [31, 61] rearranged with IgLJ4 (Section 1.4). Such restrained combinatorial diversity is observed in absence of antigen [61], favouring a structural explanation over antigen-mediated light chain maturation, which is further corroborated by high light chain sequence conservation [153] (Section 1.3.2).

This project sought to clarify the tight pairing requirements of ultralong antibodies, by observation of the structural and functional impacts of chain-exchange, and to determine whether chain-exchange alters the accessed epitope

by bovine antibodies ultralong and otherwise. For reference, structural antibody work presented in this thesis is summarised in Table 6.1.

Table 6.1: Summary of antibody structural work presented in this thesis. †Paired - Single cell paired-chain sequencing experiment. Hybrid - synthetic antibody pairing achieved through chain-exchange. Name refers to heavy chain. FMDV - Serial FMDV immunisation experiment. ‡XRD - X-ray crystallographic diffraction, AF2 - Alphafold 2, cryo-EM - cryogenic electron microscopy.

Chapter	Antibody	Source†	Antigen	Method‡	Figure
3	D08	Paired	Unknown	XRD, AF2	3.2, 3.9 and 3.33
3	D08*	Hybrid	Unknown	XRD	3.19
3	4K3E	[78]	Unknown	XRD	[78]
3	4K3E*	Hybrid	Unknown	XRD	3.26
4	#30	FMDV	O1M	XRD, cryo-EM	4.21
4	#117	FMDV	O1M, SAT2	AF2, cryo-EM	4.12
4	#126	FMDV	O1M, SAT2	AF2, cryo-EM	4.15
5	B4	[160]	bRSV F	XRD, cryo-EM	5.1
5	B13	[160]	bRSV F	XRD	5.1
5	B13*	[160]	bRSV F	XRD	5.1

6.2 Conclusions from chain-exchanged ultralong antibodies chapter

Ultralong fragment antibodies can be readily expressed by co-transfection of Expi293™ cells with plasmids encoding each of the Fab heavy and light chain respectively, and have recently been shown to express stably as recombinant single-chain antibody variable fragments (scFv) [85]. Interestingly, different pairing combinations appeared to show varying levels of expression efficiency (Figs. 3.3 and C.2), possibly indicating varying levels of stability conferred by alternative pairings, though further work would be needed to verify this finding. Quantification of expression yields and thermodynamic measurement of chain association would be useful to explain these observations.

The antigen-unknown ultralong antibody D08, and chain-exchanged hybrid D08*, share the common ultralong CDR3H fold comprising a three-pass cysteine knob flanked by β-strand stalks [78, 82, 83] (Figs. 3.9 and 3.19), though with a

novel (1-4, 2-3, 5-6) arrangement of disulphide bonds within the CDR3H knob mini-domain. The addition of (1-4, 2-3, 5-6) to the possible disulphide networks that produce a stable knob mini-domain demonstrates the remarkable structural accommodation of the cysteine knot fold, and increases the potential for structural diversity of the ultralong paratope. Should a CDR3H disulphide reductase and oxidase system be identified [78], there exists a great opportunity to further diversify the ultralong paratope with trivial genetic and material costs compared to the biosynthesis of novel ultralong antibody transcripts. However such a system would lack specificity and reproducibility.

The stalks of ultralong CDR3H have previously been shown as crucially important for the stability of the knob mini-domain, where extension or deletion impairs proper folding of the knob, and subsequent reduction of antigen binding capability [132]. Determination of the chain-exchanged structures of ultralong antibody D08 revealed the stalk domain is affected by residues of the light chain, particularly of the CDR3L (Fig. 3.23). A tight CDR3L conformation packs closely with the CDR3H ascending strand, filling a pocket otherwise occupied by a tyrosine of the descending strand 'YxYxY' motif. A tight CDR3L therefore preserves the aromatic stacking of YxYxY, which becomes destabilised if the tyrosine most proximal to the Fv regions packs between the CDR3H and CDR3L (Figs. 3.11 and 3.22). Targeted mutagenesis to the CDR3L and the base of the CDR3H stalks would be helpful in fully delineating the relationship between these loops.

Despite this, the knob mini-domain backbone was largely unaffected by light chain substitution (Fig. 3.21), with only minor accommodations affecting side chain positioning (Fig. 3.24). Given the initial evidence implying ultralong antibodies interact with antigen principally via the knob domain [78, 84, 85, 132, 138, 139, 319] (and further, the interaction of structurally similar inhibitory cysteine knot (ICK) proteins with their respective target receptors [90, 91, 94–99, 101, 103, 104]), it remains unclear whether chain-exchange is sufficient to affect ultralong antibody performance in antigenic recognition. But chain-exchange within ultralong antibodies is sufficient to twist the knob mini-domain relative to the Fv region.

Drawing conclusions on the impact of chain-exchange is complicated by observations of missing electron density for the ultralong CDR3H region of the chain-exchanged 4K3E [78] structure ('4K3E*'). Curiously, 4K3E* crystallised in spacegroup C 1 2 1, where the others were P 2_1 2_1 2_1, and 4K3E* demonstrated a wider elbow angle (as calculated by the PyMOL Elbow Angle script [267]) when compared with D08, D08*, and 4K3E. No crystals of 4K3E* were characterised as having different spacegroups, which may indicate the broader elbow angle is consequent to the combination of the 4K3E heavy chain with the D08 light chain, rather than being artefactual of crystal packing.

As a side note, antibody D08 provided a promising use case example for the VMXi *in situ* beamline at DLS, wherein the structures determined at room temperature (at VMXi) and cryogenic temperature (at I04) are very similar (Section 3.5.5), with only major differences found at the Cβ-Dβ loop of the VL domain. This loop protrudes into the solvent void between the four immunoglobulin domains. Very few solvent molecules are observed in Fab crystal structures within this void, implying disorder, and hence an insignificant contribution to the overall Fab structure, so the observed flexibility of the VL Cβ-Dβ loop at room temperature is unlikely to bear biological significance. On the contrary, the preservation of the ultralong positioning at biologically-relevant temperatures corroborates the conclusions drawn that the light chain affects the stalk geometry, however the possibility this is artefactual of crystallographic packing contacts cannot completely be excluded. It is a shame the 4K3E* CDR3H loop did not crystallise, as the positioning of an ultralong CDR3H within a crystal of alternative packing would be informative on whether ultralong CDR3H positions are inherent or crystallographic. Indeed AlphaFold *in silico* predictions of ultralong CDR3H conformations are informed by crystal structures deposited within the PDB [162], and may therefore be subject to crystallographic bias.

6.3 Conclusions from FMDV-specific chapter

To answer whether chain-exchange can alter the antigenic recognition of ultralong antibodies, novel antibodies specific for Foot-and-mouth disease virus (FMDV) were identified in collaboration with The Pirbright Institute and their external collaborator (Section 4.2).

Recent advances in the isolation and sorting of bovine B cells [30], in combination with sequencing pipelines optimised for long read lengths have made possible discovery of ultralong antibody sequences [78, 84, 85, 133–137]. The prevalence of ultralong antibodies observed in this study (15.4 % of all detected antibodies; Section 4.2.3) is slightly raised relative to the 5 - 12 % prevalence reported previously [31]. This difference could be consequential of the small sample size for repertoire analysis, or linked to antigen specificity.

Curiously, predicted structures for antibodies #117 and #126 also demonstrated an axial twist to the ultralong stalk domain when paired with differing light chains (Section 4.4.3). Though caution is advised in interpreting predicted structures, especially given the limitations of predicting ultralong loops, this prediction reflects the twist introduced to ultralong antibody D08, but it is unclear whether the #117 / #126 knob mini-domain is affected as a result. An entirely different conformation was predicted for the antibody #126 CDR3L (Fig. 4.18), possibly owing to the use of IgLV1-67 rather than IgLV1-47. Without experimental determination of the IgLV1-67 CDR3L structure, no solid conclusions dependent on the CDR3L may be drawn. Given the conservation of the IgLV1-47 CDR3L loop conformation in all previously reported structures [78, 82, 83], and the intimate contact between CDR3L and CDR3H, it is of pressing interest to resolve the IgLV1-67 CDR3L structure, and determine whether any resulting perturbation is imparted to ultralong CDR3H loops.

Unfortunately for this project, neither antibody #117 or #126 demonstrated a strong interaction with any tested serotype of intact FMD VLPs, whether expressed as Fabs (Fig. 4.47), or as whole antibodies (The Immunological Toolbox, personal communication). Consequently, observed differences by BLI constitute a trivial, if any, affect of chain-exchange on ultralong affinity. A good next step would be to

take a known ultralong antibody that binds to an antigen with strong affinity, such as antibody #30 (Section 4.7.2), NC-Cow1 [84, 137], R50 [85], or H12 [78], and perform targeted mutagenesis to the CDR3L. There is no guarantee these synthetic chain-exchanged antibodies would form viable, thermostable antibody molecules, but any effects on affinity would be clearer to spot, especially by BLI.

Cryogenic electron microscopy of ultralong antibodies complexed with FMD VLPs corroborates the flexibility notion of ultralong CDR3H stalks, as density was only observed for the distal knob mini-domain, and a few distal stalk residues. The structural analysis is not yet complete but from the preliminary data presented it appears that chain-exchange is insufficient to alter the epitopic sites being recognised, though it remains uncertain whether the precise mechanisms of interaction are precisely conserved. This finding corroborates the current understanding that ultralong knob mini-domains interact with antigen autonomously [78, 132, 138, 139].

Initial analyses implied Fab #117 and #126 each interact with a site above VP2 of O1M, and an internal site of SAT2. However, subsequent analysis of the generated difference maps for the O1M datasets showed density commensurate with the VP1 GH loop in the down conformation rather than any evidence of Fab binding at this location. Together with the lack of binding evidence from BLI study indicates the initial predicted O1M specificity for antibody #126 not to be true; at least that neither antibody #117 or #126 makes any specific interaction with O1M serotype VLPs.

Side chain resolution was only achieved during the course of this work for the interaction of #117 with dissociated SAT2 VLPs, in which the knob mini-domain fills a deep canyon between VP3 and VP2. Reconstruction of the SAT2 12S pentamers in complex with Fab #126 needs further work. Nevertheless, from the preliminary 2D classes, it is evident that the engagement of antibody #126 is, at this level, indistinguishable from #117. These antibodies will have utility as a diagnostic tool, as antibody #117 provides an antibody specific for dissociated SAT/2/Zimbabwe/7/83 and SAT/2/Saudi Arabia/6/200 capsids, and can therefore be used in antigen quality assessment.

In virus neutralisation titre assays performed by ITB, antibody #30 was found to neutralise O1M serotype FMD VLPs (personal communication). Antibody #30 appears to engage O1M via VP2 near the threefold axis (Fig. 4.44). Model building may reveal a conformational change to explain this neutralisation, but with the current data, one can only speculate, not forgetting the multitude of ways to neutralise foreign agents.

6.4 Conclusions from the preliminary bRSVF cryo-EM structural chapter

The affinities for bRSVF, and crystallographic structures of apo non-ultralong bovine antibodies B4 and B13, as well as their chain-exchanged progeny has been determined previously [160]. Work to structurally characterise the interaction of these antibodies with bRSVF is set out in Chapter 5.

Stabilised engineered mutants of the F fusion protein of bovine respiratory syncytial virus (bRSV) have historically been difficult to produce [175]. Given the DS2-v1 variant, comprising a Foldon trimerisation domain, and two added cysteine bonds for added thermal stability, is a candidate vaccine against bRSVF [175], there is great benefit to increasing the expression efficiency of this antigen. In this thesis, a stable cell lineage was created to establish a constant supply of bRSVF, incorporating purification optimisations to maximise yield. In combination, these modifications nearly doubled the yield of bRSVF compared to the transient transfection system used by the Kwong group [175].

bRSVF continues to prove a recalcitrant antigen. bRSVF demonstrates a concentration-dependent self-association effect (Sections 5.4.1 and 5.5.4.1), thereby preventing high concentration preparations typical of crystallographic study; and bRSVF rapidly associates to solution-air boundaries in a preferred directional manner (Section 5.5.1), introducing challenges to typical electron cryomicroscopy preparations. Several strategies were employed to surmount these challenges, including purification of the Fab/antigen complex, use of the Chameleon cryo-EM sample preparation robot, and coating of cryo-EM grids with graphene oxide. The

application of all of these methods was required to allow the collection of high quality single particle cryo-EM data.

This method was optimised for B4 to reduce the number of variables, with the intention of extending to B13 and B13*, however the timing of these developments precluded expansion to study of the remaining antibodies. Solving and comparing the immunostructural mechanisms behind antigenic recognition in these chain-exchanged Fabs would conclude this line of enquiry. Investigation could even include attempting to resolve the hybrid B4* in complex with bRSVF, though this would require significant investment, as the reduced affinity would likely lead to much reduced yields of intact, purified B4*/bRSVF complex, and which may dissociate more easily, resulting in bRSVF particles with low B4* occupancy upon collected micrographs. It would be interesting to model these interactions, either by molecular dynamics simulations, docking simulations, or via the multimer mode of AlphaFold 2, and comparing against experimental structural determination of antigenic recognition in non-ultralong bovine antibodies.

6.5 Summary and final remarks

The results presented in this thesis provide structural and functional insight to the nature of bovine ultralong antibodies.

The β-strands of the ultralong CDR3H stalks are subject to perturbation dependent on the light chain CDR3L. Additionally, further evidence of antigenic recognition solely achieved via the knob mini-domain of ultralong antibodies was presented. At the achieved resolution, it remains inconclusive whether alternative heavy-light chain pairings significantly modify the knob mini-domain paratope, with consequent epitope alterations. Nevertheless, these results imply a structural role for light chains paired with ultralong heavy chains, and demonstrate that the geometry of the β-stalk can be tuned without altering the β-strand sequence, or affecting the function of the distal knob mini-domain.

Bio-engineering of ultralong CDR3H loops has focused on the knob domain as an autonomous antigen-binding unit with little consideration of the long β-strand stalks.

The fact that the function of ultralong knob mini-domains appears preserved despite chain-exchange of alternative light chains, is promising for bio-engineering efforts as it negates the additional complexity of light chain design. That said, the deep recognition of HIV Env by VRC26.25 provides evidence for recognition mediated by a long β-strand that penetrates deep within the Env structure [136]. As a result, the β-strand separation of the knob mini-domain from the ultralong Fv region may provide the ability to design paratopes specific for deep cavernous epitopes. In such a circumstance, where specific geometric constraints would be known, the ability to fine tune the axial twist of ultralong β strands would be advantageous.

A cornerstone of immunological principles is the ability to produce a vast array of macromolecular machines to combat the potentially infinite array of foreign invading bodies. The majority of cattle immunoglobulins are structurally not too dissimilar from our own. Sequence invariance at the genomic level, and limited recombinations at the junctional level, hamper bovine immune repertoire diversity, however foetal, and neonatal somatic hypermutation, in absence of antigen stimulation, somewhat alleviates the genomic limitations. Even so, the most remarkable feature of cattle immunology is the evolution of ultralong CDR3H loops. The phenomenon of ultralong antibodies is unique to bovids, and offers an unorthodox route for achieving a functional humoral immune system compared to the well understood concepts of recombinatorial diversity and affinity maturation utilised in other Vertebrata. The ultralong stalk region, which appears to demonstrate some flexibility, could allow ultralong antibodies to 'bind' far more closely together whilst avoiding steric clashes for allowing increased antibody load during opsonisation or agglutination events. As well as unprecedented variety in CDR3H length, ultralong antibodies offer an entirely new paradigm of antigenic recognition. The knob mini-domains, which appear with increasing confidence to be solely responsible for binding to antigen, provide a very small convex paratope, sufficient to bind crevices and deep cavernous epitopes that remain invisible to other species. Undoubtedly, ultralong antibodies will have tremendous value as therapeutics and diagnostic tools.

Appendices

A
Sample Details

Table A.1: **Antibody variable domain sequences used in this project.** †wt - observed nucleotide sequence. opt. - codon optimised nucleotide sequence for expression in *Homo sapiens* cells. Cysteines of ultralong knob mini-domains are coloured orange.

Identifier	Type†	Sequence		
D08-H	amino	1	QVQLRESGPS LVKPSQTLSL TCTASGSSLS DEAVGWVRQA PGKSLEWLGS	50
		51	IDTGGNTGYN PNLKTRLSIT KDNSKSQVSL SMSSVTPEDS ATYFCATVHQ	100
		101	ETHQTCPDGY NSGDDCGRGN WDCGTLDCWR CDWGGFCRAS TDYRSVTATY	150
		151	TYEWYIDTWG QGLLVTVSS	169
D08-H	wt.	1	CAGGTGCAGC TGCGGGAGTC GGGCCCCAGC CTGGTGAAGC CGTCGCAGAC	50
		51	CCTCTCGCTC ACCTGCACGG CCTCTGGATC GTCATTAAGC GACGAGGCTG	100
		101	TAGGCTGGGT CCGCCAGGCT CCAGGAAAGT CGCTGGAGTG GCTCGGTAGT	150
		151	ATAGACACTG GTGGAAACAC AGGCTATAAC CCAAACCTGA AAACCCGACT	200
		201	CAGTATCACC AAGGACAACT CCAAGAGCCA AGTCTCTCTG TCGATGAGCA	250
		251	GCGTAACGCC TGAAGACTCG GCCACATACT TCTGTGCCAC TGTGCACCAG	300
		301	GAAACACACC AAACTTGTCC TGATGGTTAT AATAGTGGTG ATGATTGTGG	350
		351	AAGGGGTAAT TGGGATTGTG GTACTTTGGA TTGTTGGCGT TGTGATTGGG	400
		401	GAGGTTTTTG TCGCGCTTCC ACTGATTATC GTAGTGTTAC GGCCACTTAC	450
		451	ACTTACGAAT GGTACATCGA TACCTGGGGC CAAGGACTCC TGGTCACCGT	500
		501	CTCGTCA	507

Continued on next page

Table A.1: **Antibody variable domain sequences used in this project.** †wt - observed nucleotide sequence. opt. - codon optimised nucleotide sequence for expression in *Homo sapiens* cells. Cysteines of ultralong knob mini-domains are coloured orange. (Continued)

Identifier	Type†	Sequence	
D08-H	opt.	1 CAGGTCCAGC TCCGGGAAAG CGGCCCTTCC CTGGTTAAAC CCTCTCAGAC 50 51 GTTGAGTCTG ACCTGTACGG CAAGCGGTAG CAGTCTCTCC GATGAAGCTG 100 101 TTGGTTGGGT TAGGCAGGCA CCGGGGAAGA GCTTGGAATG GCTTGGATCA 150 151 ATCGACACGG GAGGGAACAC TGGGTACAAT CCAAACTTGA AAACAAGACT 200 201 GTCCATCACC AAAGATAATT CCAAGTCCCA GGTATCACTC AGCATGTCAT 250 251 CAGTGACTCC CGAGGATAGT GCCACGTACT TCTGCGCGAC CGTCCATCAA 300 301 GAAACTCATC AGACCTGTCC AGATGGCTAC AATTCAGGTG ATGATTGCGG 350 351 ACGAGGTAAT TGGGATTGCG GAACCTTGGA TTGCTGGAGG TGCGATTGGG 400 401 GCGGGTTCTG TCGGGCTTCT ACCGATTATC GGAGCGTCAC GGCCACTTAT 450 451 ACATACGAAT GGTATATAGA CACATGGGGG CAGGGGCTCC TTGTGACCGT 500 501 CTCATCA 507
D08-L	amino	1 QGVLTQPSSV SGSLGQRVSI TCSGSSSNVG RGYVSWYQMT PGSAPRTLIY 50 51 GDTNRASGVP DRFSASRSGN TATLTISSLQ AEDEADYFCA SAEGSSSNAV 100 101 FGSGTTLTVL 110
D08-L	wt.	1 CAGGGTGTGC TGACTCAGCC ATCATCCGTG TCCGGGTCCC TGGGCCAGAG 50 51 GGTCTCCATC ACCTGCTCTG GAAGCAGCAG CAATGTTGGA CGTGGATATG 100 101 TGAGCTGGTA CCAAATGACC CCAGGATCGG CCCCCAGAAC CCTCATCTAT 150 151 GGTGACACCA ATCGAGCCTC GGGGGTCCCC GACCGATTCT CCGCCTCCAG 200 201 GTCTGGGAAC ACAGCCACCC TGACCATCAG CTCGCTCCAG GCTGAGGACG 250 251 AGGCAGATTA TTTCTGTGCA TCTGCTGAGG GTAGTAGCAG TAATGCTGTT 300 301 TTCGGCAGCG GGACCACACT GACCGTCCTG 330
D08-L	opt.	1 CAGGGAGTTC TGACCCAACC TTCTTCTGTG AGCGGGTCCT TGGGCCAGAG 50 51 AGTAAGTATC ACGTGTTCAG GATCTTCTAG TAATGTAGGG CGGGGTTACG 100 101 TATCCTGGTA TCAGATGACA CCTGGATCTG CCCCGAGAAC GCTTATCTAT 150 151 GGTGATACGA ATCGAGCGTC AGGAGTTCCA GATCGGTTTA GTGCATCCAG 200 201 GTCTGGAAAC ACAGCAACCT TGACTATTAG CTCACTCCAG GCAGAGGATG 250 251 AGGCAGATTA TTTTTGTGCA TCCGCTGAAG GATCTAGTTC AAATGCTGTT 300 301 TTTGGATCTG GCACCACATT GACAGTGCTG 330
4K3E-H	amino	1 QVQLRESGPS LVQPSQTLSL TCTASGFSLS DKAVGWVRQA PGKALEWLGS 50 51 IDTGGSTGYN PGLKSRLSIT KDNSKSQVSL SVSSVTTEDS ATYYCTTVHQ 100 101 ETRKTCSDGY IAVDSCGRGQ SDGCVNDCNS CYYGWRNCRR QPAIHSYEFH 150 151 VDAWGRGLLV TVSS 164
4K3E-H	opt.	1 CAGGTGCAGC TGCGGGAGTC GGGCCCCAGC CTGGTGCAGC CCTCACAGAC 50 51 CCTCTCCCTC ACCTGCACGG CCTCTGGATT CTCATTGAGC GACAAGGCTG 100 101 TAGGCTGGGT CCGCCAGGCT CCAGGGAAGG CGCTGGAGTG GCTCGGTAGT 150 151 ATAGACACTG GTGGAAGCAC AGGCTATAAC CCAGGCCTGA AATCCCGGCT 200 201 CAGCATCACC AAGGACAACT CCAAGAGCCA AGTCTCTCTG TCAGTGAGCA 250 251 GCGTGACGAC GGAGGACTCG GCCACATACT ACTGTACTAC TGTGCACCAA 300 301 GAAACCAGAA AAACCTGTTC TGATGGTTAT ATTGCTGTAG ATAGTTGTGG 350 351 TCGTGGTCAG AGTGATGGTT GTGTCAATGA TTGCAATTCA TGTTATTATG 400 401 GTTGGCGGAA CTGTCGCAGG CAGCCTGCAA TTCACAGTTA CGAATTTCAC 450 451 GTCGATGCCT GGGGCCGAGG ACTCCTGGTC ACCGTCTCCT CA 492

Continued on next page

Table A.1: **Antibody variable domain sequences used in this project.** †wt - observed nucleotide sequence. opt. - codon optimised nucleotide sequence for expression in *Homo sapiens* cells. Cysteines of ultralong knob mini-domains are coloured orange. (Continued)

Identifier	Type†	Sequence					
4K3E-L	amino	1	QGVLTQPSSV	SGSLGQRVSI	TCSGSSSNVG	NGYVSWYQLI	PGSAPRTLIY	50
		51	GDTSRASGVP	DRFSGSRSGN	TATLTISSLQ	AEDEADYFCA	SAEDSSSNAV	100
		101	FGSGTTLTVL					110
4K3E-L	opt.	1	CAGGGTGTGC	TGACTCAGCC	ATCATCCGTG	TCCGGGTCCC	TGGGCCAGAG	50
		51	GGTCTCCATC	ACCTGCTCTG	GAAGCAGCAG	CAATGTTGGA	AATGGATATG	100
		101	TGAGCTGGTA	CCAACTGATC	CCAGGATCGG	CCCCCAGAAC	CCTCATCTAT	150
		151	GGTGACACCA	GTCGAGCCTC	GGGGGTCCCC	GACCGATTCT	CCGGCTCCAG	200
		201	GTCTGGGAAC	ACAGCCACCC	TGACCATCAG	CTCGCTCCAG	GCTGAGGACG	250
		251	AGGCAGATTA	TTTCTGTGCA	TCTGCTGAGG	ATAGTAGCAG	TAATGCTGTT	300
		301	TTCGGCAGCG	GGACCACACT	GACCGTCCTG			330
#117-H	amino	1	QVQLRESGPS	LVKPSQTLSL	TCTASGFSLS	DKAVGWVRQA	PGKALEWLGS	50
		52	VDSGGNTGYN	PGLKSRLSIT	KDNSKSQVSL	SVSSVTTEDS	ATYYCTTVLQ	100
		101	QTTKKENNCP	DGYSCGYRCR	SGWGCSGDEC	CGRRGGGWGS	IELIACCSST	150
		151	YIHEFHVDAW	GQGLLVTVSS				170
#117-H	wt.	1	CAGGTGCAGC	TGCGCGAGTC	GGGCCCCAGC	CTGGTGAAGC	CGTCACAGAC	50
		51	CCTCTCGCTC	ACCTGCACGG	CCTCTGGATT	CTCATTGAGC	GACAAGGCTG	100
		101	TAGGCTGGGT	CCGCCAGGCT	CCAGGGAAGG	CGCTGGAGTG	GCTCGGTAGT	150
		151	GTAGACAGTG	GTGGAAACAC	AGGCTATAAC	CCAGGCCTGA	AATCCCGGCT	200
		201	CAGTATTACC	AAGGACAACT	CCAAGAGCCA	AGTCTCTCTG	TCAGTGAGCA	250
		251	GCGTGACAAC	TGAGGACTCG	GCCACATACT	ACTGTACTAC	TGTGCTCCAA	300
		301	CAGACAACAA	AAAAAGAAAA	CAATTGTCCT	GATGGTTATA	GTTGTGGGTA	350
		351	TCGTTGTCGT	AGTGGTTGGG	GTTGTAGTGG	TGATGAGTGT	TGTGGTAGGA	400
		401	GAGGTGGAGG	TTGGGGTAGT	ATTGAGTTAA	TTGCTTGTTG	TAGTAGTACT	450
		451	TACATTCACG	AATTCCACGT	CGATGCCTGG	GGCCAAGGAC	TCCTGGTCAC	500
		501	CGTCTCCTCA					510
#117-H	opt.	1	CAGGTTCAAC	TGCGAGAGTC	TGGCCCCAGT	CTCGTGAAGC	CGAGCCAGAC	50
		51	CCTTAGCCTG	ACTTGTACTG	CATCCGGGTT	TTCCCTGAGC	GATAAAGCTG	100
		101	TGGGGTGGGT	GCGACAGGCC	CCAGGGAAGG	CACTGGAGTG	GTTGGGAAGC	150
		151	GTCGATTCCG	GGGGGAACAC	CGGCTATAAC	CCTGGCCTTA	AGTCCCGATT	200
		201	GAGCATTACG	AAGGACAACT	CAAAGAGTCA	GGTATCCCTT	TCAGTGAGTA	250
		251	GCGTAACGAC	CGAAGATTCC	GCGACCTACT	ACTGTACTAC	TGTCCTCCAG	300
		301	CAAACGACAA	AGAAAGAGAA	CAACTGCCCA	GACGGCTATA	GTTGTGGCTA	350
		351	CCGATGCCGG	TCAGGTTGGG	GCTGTAGTGG	CGACGAATGT	TGCGGGCGGA	400
		401	GAGGAGGTGG	TTGGGGCTCA	ATCGAATTGA	TAGCCTGTTG	CTCCTCTACT	450
		451	TATATACACG	AATTTCATGT	GGACGCATGG	GGTCAGGGAC	TCCTCGTCAC	500
		501	TGTTTCTTCA					510
#117-L	amino	1	QAVLTQPSSV	SGSLGQRVSI	TCSGSSSNIG	NSNVGWYQQV	PGSGLTTIIY	50
		51	SSFSRPSGVP	DRFSGSKSGN	TATLTISSLQ	AEDEADYFCA	TYDSGVSIGV	100
		101	FGSGTTLTVL					110

Continued on next page

Table A.1: **Antibody variable domain sequences used in this project.** †wt - observed nucleotide sequence. opt. - codon optimised nucleotide sequence for expression in *Homo sapiens* cells. Cysteines of ultralong knob mini-domains are coloured orange. (Continued)

Identifier	Type†	Sequence					
#117-L	wt.	1	CAGGCTGTGC	TGACTCAGCC	GTCCTCCGTG	TCCGGCTCCC	TGGGCCAGAG	50
		51	GGTCTCCATC	ACCTGCTCTG	GAAGCAGCAG	CAACATCGGT	AATAGTAATG	100
		101	TGGGCTGGTA	CCAACAGGTC	CCAGGATCGG	GCCTCACAAC	CATCATCTAT	150
		151	AGTAGTTTCA	GTCGACCCTC	GGGGGTCCCC	GACCGATTCT	CCGGCTCCAA	200
		201	GTCTGGCAAC	ACAGCCACCC	TGACCATCAG	CTCCCTCCAG	GCTGAGGACG	250
		251	AGGCGGATTA	TTTCTGTGCA	ACTTATGACA	GTGGTGTCAG	TATTGGTGTT	300
		301	TTCGGCAGCG	GGACCACACT	GACCGTCCTG			330
#117-L	opt.	1	CAGGCAGTGC	TGACGCAACC	CTCCTCTGTC	TCTGGAAGTC	TTGGTCAAAG	50
		51	AGTGAGTATT	ACATGCAGTG	GCTCTTCCAG	TAACATCGGT	AACAGCAATG	100
		101	TCGGGTGGTA	TCAACAAGTA	CCGGGAAGTG	GACTTACGAC	CATAATTTAC	150
		151	AGCTCCTTTA	GCCGACCTTC	AGGCGTTCCA	GATCGCTTCT	CTGGCTCTAA	200
		201	ATCTGGCAAT	ACGGCTACAC	TTACTATCTC	CTCTCTGCAA	GCCGAGGATG	250
		251	AAGCGGATTA	CTTTTGTGCC	ACATATGACA	GCGGCGTGAG	CATTGGCGTT	300
		301	TTCGGGTCTG	GGACTACTCT	TACTGTCCTG			330
#126-H	amino	1	QVQLRESGPS	LVKPSQTLSL	TCTASGFSLS	DKAVGWVRQA	PGKALEWLGS	50
		52	VDSGGNTGYN	PGLKSRLSIT	KDNSKSQVSL	SVSSVTTEDS	ATYYCTTVLQ	100
		101	QTTKKENNCP	DGYSCGYRCR	SGWGCSGDEC	CGRRGGGWGS	IELIACCSST	150
		151	YIHEFHVDAW	GQGLLVTVSS				170
#126-H	wt.	1	CAGGTGCAGC	TGCGCGAGTC	GGGCCCCAGC	CTGGTGAAGC	CGTCACAGAC	50
		51	CCTCTCGCTC	ACCTGCACGG	CCTCTGGATT	CTCATTGAGC	GACAAGGCTG	100
		101	TAGGCTGGGT	CCGCCAGGCT	CCAGGGAAGG	CGCTGGAGTG	GCTCGGTAGT	150
		151	GTAGACAGTG	GTGGAAACAC	AGGCTATAAC	CCAGGCCTGA	AATCCCGGCT	200
		201	CAGTATTACC	AAGGACAACT	CCAAGAGCCA	AGTCTCTCTG	TCAGTGAGCA	250
		251	GCGTGACAAC	TGAGGACTCG	GCCACATACT	ACTGTACTAC	TGTGCTCCAA	300
		301	CAGACAACAA	AAAAAGAAAA	CAATTGTCCT	GATGGTTATA	GTTGTGGGTA	350
		351	TCGTTGTCGT	AGTGGTTGGG	GTTGTAGTGG	TGATGAGTGT	TGTGGTAGGA	400
		401	GAGGTGGAGG	TTGGGGTAGT	ATTGAGTTAA	TTGCTTGTTG	TAGTAGTACT	450
		451	TACATTCACG	AATTCCACGT	CGATGCCTGG	GGCCAAGGAC	TCCTGGTCAC	500
		501	CGTCTCCTCA					510
#126-H	opt.	1	CAGGTTCAAC	TGCGAGAGTC	TGGCCCCAGT	CTCGTGAAGC	CGAGCCAGAC	50
		51	CCTTAGCCTG	ACTTGTACTG	CATCCGGGTT	TTCCCTGAGC	GATAAAGCTG	100
		101	TGGGGTGGGT	GCGACAGGCC	CCAGGGAAGG	CACTGGAGTG	GTTGGGAAGC	150
		151	GTCGATTCCG	GGGGGAACAC	CGGCTATAAC	CCTGGCCTTA	AGTCCCGATT	200
		201	GAGCATTACG	AAGGACAACT	CAAAGAGTCA	GGTATCCCTT	TCAGTGAGTA	250
		251	GCGTAACGAC	CGAAGATTCC	GCGACCTACT	ACTGTACTAC	TGTCCTCCAG	300
		301	CAAACGACAA	AGAAAGAGAA	CAACTGCCCA	GACGGCTATA	GTTGTGGCTA	350
		351	CCGATGCCGG	TCAGGTTGGG	GCTGTAGTGG	CGACGAATGT	TGCGGGCGGA	400
		401	GAGGAGGTGG	TTGGGGCTCA	ATCGAATTGA	TAGCCTGTTG	CTCCTCTACT	450
		451	TATATACACG	AATTTCATGT	GGACGCATGG	GGTCAGGGAC	TCCTCGTCAC	500
		501	TGTTTCTTCA					510
#126-L	amino	1	QAVLTQPSSV	SGSLGQSVSI	TCSGSSSNVG	NGYVSWYQMT	PGSAPRTLIY	50
		51	GPTSRASGVP	DRFSGSRSGN	TATLTISSLQ	AEDEADYFCA	SAEDSSTYAV	100
		101	FGSGTTLTVL					110

Continued on next page

Table A.1: **Antibody variable domain sequences used in this project.** †wt - observed nucleotide sequence. opt. - codon optimised nucleotide sequence for expression in *Homo sapiens* cells. Cysteines of ultralong knob mini-domains are coloured orange. (Continued)

Identifier	Type†	Sequence					
#126-L	wt.	1	CAGGCTGTAC	TGACTCAGCC	ATCATCCGTG	TCCGGGTCCC	TGGGCCAGAG	50
		51	TGTCTCCATC	ACCTGCTCTG	GAAGCAGCAG	CAATGTTGGA	AATGGATATG	100
		101	TGAGCTGGTA	CCAAATGACC	CCAGGATCGG	CCCCCAGAAC	CCTCATCTAT	150
		151	GGTCCCACCA	GTCGAGCCTC	GGGGGTCCCC	GACCGATTCT	CCGGCTCCAG	200
		201	GTCTGGCAAC	ACAGCCACCC	TGACCATCAG	TTCGCTCCAG	GCTGAGGACG	250
		251	AGGCAGATTA	TTTCTGTGCA	TCTGCTGAGG	ATAGTAGCAC	CTATGCTGTT	300
		301	TTCGGCAGCG	GGACCACACT	GACCGTCCTC			330
#126-L	opt.	1	CAGGCTGTGT	TGACCCAACC	ATCATCAGTG	TCTGGAAGTC	TCGGGCAGTC	50
		51	TGTGAGTATA	ACATGCAGTG	GAAGTTCCTC	CAATGTCGGA	AATGGGTATG	100
		101	TTAGTTGGTA	CCAGATGACC	CCAGGAAGCG	CTCCTAGAAC	TCTCATTTAC	150
		151	GGGCCAACGA	GCCGAGCGTC	TGGTGTACCA	GACCGATTCT	CCGGCAGTCG	200
		201	CAGCGGAAAC	ACAGCTACGC	TTACCATATC	CTCTCTTCAA	GCAGAAGACG	250
		251	AGGCAGACTA	TTTTTGTGCC	TCAGCTGAGG	ATTCATCAAC	CTATGCTGTC	300
		301	TTTGGAAGTG	GTACGACACT	TACCGTACTC			330
#30-H	amino	1	QVQLRESGPS	LVKPSQTLSL	TCTASGFSWS	DKAVGWVRQA	PGKALEWLGS	50
		51	VDAGGSTGYN	QGLKSRLSIT	KDNSRSQVSL	SVSSVVTEDS	ATYYCATVYQ	100
		101	KTTRG**C**PDGY	YYEEE**C**GGDW	R**C**SRRRDR**C**S	YPYTNKYEWY	VDAWGQGLLV	150
		151	TVSS					154
#30-H	wt.	1	CAGGTGCAGC	TGCGCGAGTC	GGGCCCCAGC	CTAGTGAAGC	CGTCACAGAC	50
		51	CCTCTCGCTC	ACCTGCACGG	CCTCTGGATT	CTCATGGAGC	GACAAGGCTG	100
		101	TAGGCTGGGT	CCGCCAGGCT	CCAGGGAAGG	CGCTGGAGTG	GCTCGGTAGT	150
		151	GTAGACGCTG	GTGGAAGTAC	TGGCTATAAC	CAAGGCCTGA	AATCCCGGCT	200
		201	CAGCATCACC	AAGGACAACT	CCAGGAGCCA	AGTTTCTCTG	TCAGTGAGCA	250
		251	GCGTGGTCAC	TGAGGACTCG	GCCACGTATT	ACTGTGCTAC	TGTGTACCAA	300
		301	AAAACAACGC	GTGGTTGCCC	TGATGGTTAT	TATTATGAAG	AGGAATGTGG	350
		351	TGGAGACTGG	CGTTGCTCTC	GTAGGCGTGA	CCGTTGTAGT	TATCCGTACA	400
		401	CTAATAAATA	CGAATGGTAC	GTCGATGCCT	GGGGCCAAGG	ACTCCTGGTC	450
		451	ACCGTCTCCT	CA				462
#30-H	opt.	1	CAGGTGCAGC	TCCGCGAAAG	CGGCCCATCT	CTTGTCAAGC	CCAGTCAGAC	50
		51	ATTGAGCCTG	ACCTGTACCG	CGTCCGGCTT	TTCCTGGAGC	GATAAGGCTG	100
		101	TGGGCTGGGT	CCGACAAGCA	CCTGGGAAAG	CACTGGAATG	GCTGGGCTCT	150
		151	GTGGATGCGG	GAGGCTCTAC	GGGCTACAAC	CAGGGTTTGA	AGTCACGGCT	200
		201	GAGCATAACT	AAGGACAATT	CCCGCTCTCA	GGTTTCTCTC	TCCGTTAGCA	250
		251	GCGTGGTAAC	AGAGGATTCA	GCAACCTATT	ACTGTGCTAC	GGTGTACCAA	300
		301	AAAACGACCC	GGGGATGTCC	CGATGGGTAT	TACTACGAAG	AAGAGTGCGG	350
		351	TGGGGACTGG	CGATGCAGCA	GGAGAAGAGA	TAGGTGTTCC	TATCCCTATA	400
		401	CGAATAAATA	TGAATGGTAC	GTAGATGCTT	GGGGACAGGG	ACTGCTCGTG	450
		451	ACTGTGTCAT	CA				462
#30-L	amino	1	QAVLTQPSSV	SGSLGQRVSI	TCSGSSSNVG	NGYVSWYQLI	PGSAPRTLIY	50
		51	GDTSRASGVP	DRFSGSRSGN	TATLTIGSLQ	AEDEADYFCA	SAEDSSSNGS	100
		101	FGSGTTLTVL					110

Continued on next page

Table A.1: **Antibody variable domain sequences used in this project.** †wt - observed nucleotide sequence. opt. - codon optimised nucleotide sequence for expression in *Homo sapiens* cells. Cysteines of ultralong knob mini-domains are coloured orange. (Continued)

Identifier	Type†	Sequence	
#30-L	wt.	1 CAGGCTGTGC TGACTCAGCC ATCATCCGTG TCCGGGTCCC TGGGCCAGAG 50 51 GGTCTCCATC ACCTGCTCTG GAAGCAGCAG CAATGTTGGA AATGGATATG 100 101 TGAGCTGGTA CCAACTGATC CCAGGATCGG CCCCCAGAAC CCTCATCTAT 150 151 GGTGACACCA GTCGAGCCTC GGGGGTCCCC GACCGATTCT CCGGCTCCAG 200 201 GTCTGGGAAC ACAGCCACCC TGACCATCGG CTCGCTCCAG GCTGAGGACG 250 251 AGGCAGATTA TTTCTGTGCA TCCGCTGAGG ATAGTAGCAG TAATGGTTCT 300 301 TTCGGCAGCG GGACCACACT GACCGTCCTG 330
#30-L	opt.	1 CAGGCGGTGC TGACGCAACC TAGTAGCGTG TCTGGTTCAT TGGGTCAACG 50 51 AGTCTCTATA ACTTGTAGCG GTTCAAGTTC CAACGTGGGC AATGGTTACG 100 101 TCTCATGGTA TCAACTGATT CCGGGGTCCG CTCCGCGAAC CCTCATTTAC 150 151 GGAGATACCT CCAGGGCGTC AGGCGTTCCC GACAGATTTA GTGGATCTCG 200 201 CTCCGGTAAC ACCGCCACAC TTACCATAGG ATCACTGCAA GCGGAAGATG 250 251 AAGCAGACTA TTTTTGCGCG AGCGCAGAGG ATAGCTCATC AAACGGCTCC 300 301 TTCGGTTCCG GAACAACACT CACCGTCCTG 330
B4-H	amino	1 QVQLQESGPS LVKPSQTLSL TCTVSGFSLS SYSVSWVRQA PGKTLEWLGD 50 51 ASNGGIIYYN PALKSRLGIT RDNSKSQVSL SLNTITPEDT ATYYCAK**C**SV 100 101 GDSGSYA**C**TG RKGEYVDAWG QGLLVTVSS 129
B4-H	opt.	1 CAGGTGCAGC TGCAGGAGTC GGGCCCCAGC CTGGTGAAGC CCTCACAGAC 50 51 CCTCTCCCTC ACCTGCACGG TGTCTGGATT CTCATTGAGC AGCTATTCTG 100 101 TAAGCTGGGT ACGCCAGGCT CCAGGGAAGA CGCTGGAGTG GCTTGGTGAC 150 151 GCAAGCAATG GCGGAATCAT ATATTATAAC CCAGCCCTGA AATCCCGGCT 200 201 CGGCATCACC AGGGACAACT CCAAGAGCCA AGTCTCTCTG TCACTGAACA 250 251 CGATCACACC TGAGGATACG GCGACATACT ACTGTGCGAA GTGTTCTGTT 300 301 GGTGATAGTG GAAGTTATGC TTGTACTGGA CGCAAGGGCG AGTATGTCGA 350 351 CGCCTGGGGC CAAGGACTCC TGGTCACCGT CTCCTCA 387
B4-L	amino	1 QAVLTQPPSV SGSLGQRVSI TCSGSSSNIG RWGVNWYQQV PGSGLRTIIY 50 51 YESSRPSGVP DRFSGSKSGN TATLTISSLQ AEDEADYFCA TGDYNIAVFG 100 101 SGTTLIVM 108
B4-L	opt.	1 CAGGCTGTGC TGACTCAGCC GCCCTCCGTG TCCGGCTCCC TGGGCCAGAG 50 51 GGTCTCCATC ACCTGCTCTG GAAGCAGCAG TAACATCGGT AGATGGGGTG 100 101 TGAACTGGTA CCAACAGGTC CCAGGATCGG GCCTCAGAAC CATCATTTAT 150 151 TATGAGAGTA GTCGACCCTC GGGGGTCCCC GACCGATTCT CCGGCTCCAA 200 201 GTCTGGCAAC ACAGCCACCC TGACCATCAG CTCGCTCCAG GCTGAGGACG 250 251 AGGCGGATTA TTTCTGTGCA ACTGGTGACT ACAATATTGC TGTTTTCGGC 300 301 AGCGGGACCA CACTGATCGT TATG 324
B13-H	amino	1 QVQLQESGPS LVKPSQTLSL TCTVSGLSLS DHNVGWIRQA PGKALEWLGV 50 51 IYKEGDKDYN PALKSRLSIT KDNSKSQVSL SLSSVTTEDT ATYYCATLG**C** 100 101 YFVEGVGYD**C** TYGLQHTTFH DAWGQGLLVT VSS 133

Continued on next page

Table A.1: **Antibody variable domain sequences used in this project.** †wt - observed nucleotide sequence. opt. - codon optimised nucleotide sequence for expression in *Homo sapiens* cells. Cysteines of ultralong knob mini-domains are coloured orange. (Continued)

Identifier	Type†	Sequence					
B13-H	opt.	1	CAGGTGCAGC	TGCAGGAGTC	GGGCCCCAGT	CTGGTGAAGC	CCTCACAGAC 50
		51	CCTCTCCCTC	ACCTGCACGG	TCTCCGGCCT	CTCATTAAGC	GACCATAATG 100
		101	TAGGCTGGAT	CCGCCAGGCT	CCAGGAAAGG	CGCTGGAGTG	GCTCGGTGTA 150
		151	ATATATAAGG	AAGGGGACAA	AGACTATAAC	CCAGCCCTGA	AATCCCGGCT 200
		201	CAGCATCACC	AAGGACAACT	CCAAGAGCCA	AGTCTCTCTA	TCACTGAGCA 250
		251	GCGTGACAAC	TGAGGACACG	GCCACATACT	ACTGTGCAAC	CTTGGGGTGT 300
		301	TATTTTGTTG	AAGGTGTGGG	TTATGATTGT	ACTTATGGTC	TTCAACATAC 350
		351	TACTTTCCAC	GATGCCTGGG	GCCAAGGACT	CCTGGTCACC	GTCTCCTCA 399
B13-L	amino	1	QAVLTQPPSV	SGSLGQRVSI	TCSGSSDNIG	IFAVGWYQQV	PGSGLRTIIY 50
		51	GNTKRPSGVP	DRFSGSKSGN	TATLTINSLQ	AEDEADYFCV	CGESKSATPV 100
		101	FGGGTTLTVL				110
B13-L	opt.	1	CAGGCTGTGC	TGACTCAGCC	GCCCTCCGTG	TCCGGCTCCC	TGGGCCAGAG 50
		51	AGTCTCCATC	ACCTGCTCTG	GAAGCAGCGA	CAATATCGGT	ATATTTGCGG 100
		101	TGGGCTGGTA	CCAACAGGTC	CCAGGATCGG	GCCTCAGAAC	CATCATCTAT 150
		151	GGTAATACCA	AACGACCCTC	GGGGGTCCCC	GACCGATTCT	CCGGCTCCAA 200
		201	GTCTGGCAAC	ACTGCCACCC	TGACCATCAA	CTCGCTCCAG	GCTGAGGACG 250
		251	AGGCGGATTA	TTTCTGTGTT	TGTGGTGAGA	GTAAGAGCGC	AACGCCTGTA 300
		301	TTTGGCGGCG	GGACCACACT	AACCGTCCTG		330

All antibody variable domain sequences were submitted for glycosylation site prediction using NetNGlyc [187, 189] and NetOGlyc [188, 190]. No N-glycosylation sites were predicted for any sequence, however, several possible O-glycosylation sites were identified; two in the D08 heavy chain (Ser140; 0.722935, Thr141; 0.68651), one in the Fab #30 heavy chain (Ser123; 0.722987), one in the Fab #126 light chain (Thr53; 0.5), and one in the B13 light chain (Ser57; 0.587508).

Table A.2: Predicted protein bio-characteristics for bRSVF. Characteristics predicted using ExPASy ProtParam tool [180].

Construct	Monomer	Homotrimer
Mw (Da)	57496.98	172490.94
Ext. coef. (M-1 cm-1)	52925	88570
# Amino Acids	522	1566
# Atoms	8090	24264
Theoretical pI	7.42	7.61
Abs 0.1% (=1g/L)	0.920	0.921

Table A.3: Protein bio-characteristics for Fabs under study within this project predicted using ExPASy ProtParam tool [180]**.** VH and VL ID refers to sequence ID within the OPPF sample database.

Fab	D08	D08*	4K3E*	#117	#126	#30	B4	B13	B13*
VH ID	21996	21996	22106	22404	22404	22447	18771	18773	18773
VL ID	21997	22105	21997	22405	22406	22448	18766	18767	18766
CDR3H Length (AA)	65	65	54	64	64	48	19	23	23
Mw (Da)	57823.72	57294.06	57134.16	57471.71	57463.62	56419.29	53280.26	53940.13	53982.08
Ext. coef. (M-1 cm-1)	88570	88570	72070	76205	77695	88905	73310	64830	73310
# Amino Acids	551	546	546	552	552	536	509	515	513
# Atoms	7986	7911	7918	7976	7962	7820	7431	7535	7529
Theoretical pI	6.59	6.26	8.30	8.13	7.96	8.19	8.38	7.81	7.50
Abs 0.1% (=1g/L)	1.532	1.546	1.261	1.326	1.352	1.576	1.376	1.202	1.358

Table A.4: Antibody gene designations. Calculated by submission to the IMGT/V-Quest tool [80, 202].

Antibody	Chain	V Gene	D Gene	J Gene
D08	Heavy	IGHV1-7*02	IGHD8-2*01	IGHJ2-4*01
D08	Light	IGLV1-47*01	-	IGLJ4*01
4K3E	Heavy	IGHV1-7*02	IGHD8-2*01	IGHJ2-4*01
4K3E	Light	IGLV1-47*01	-	IGLJ4*01
117	Heavy	IGHV1-7*02	IGHD8-2*01	IGHJ2-4*01
117	Light	IGLV1-67*02	-	IGLJ4*01
126	Heavy	IGHV1-7*02	IGHD8-2*01	IGHJ2-4*01
126	Light	IGLV1-47*01	-	IGLJ4*01
30	Heavy	IGHV1-7*02	IGHD8-2*01	IGHJ2-4*01
30	Light	IGLV1-47*01	-	IGLJ4*01
B4	Heavy	IGHV1-17*01	IGHD3-1*01	IGHJ2-4*01
B4	Light	IGLV1-67*01	-	IGLJ4*01
B13	Heavy	IGHV1-14*02	IGHD4-1*01	IGHJ2-4*01
B13	Light	IGLV1-67*02	-	IGLJ4*01

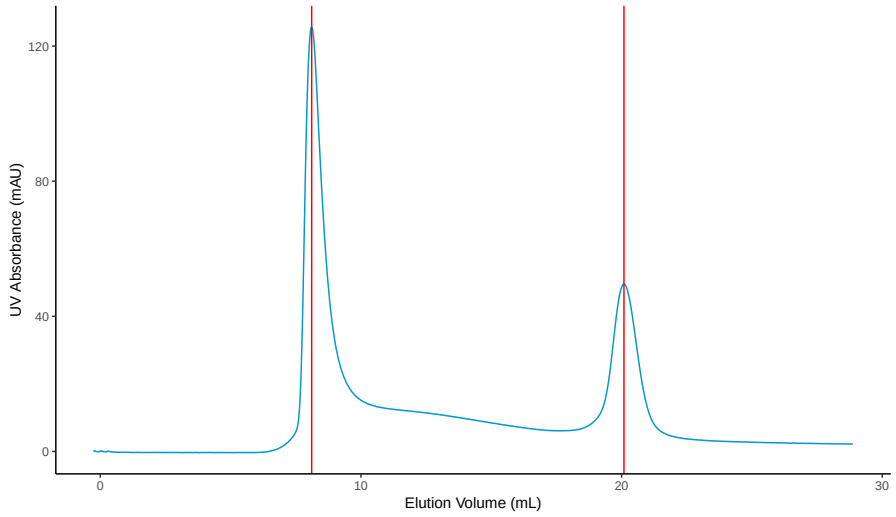
B

Chapter 2: Methodology

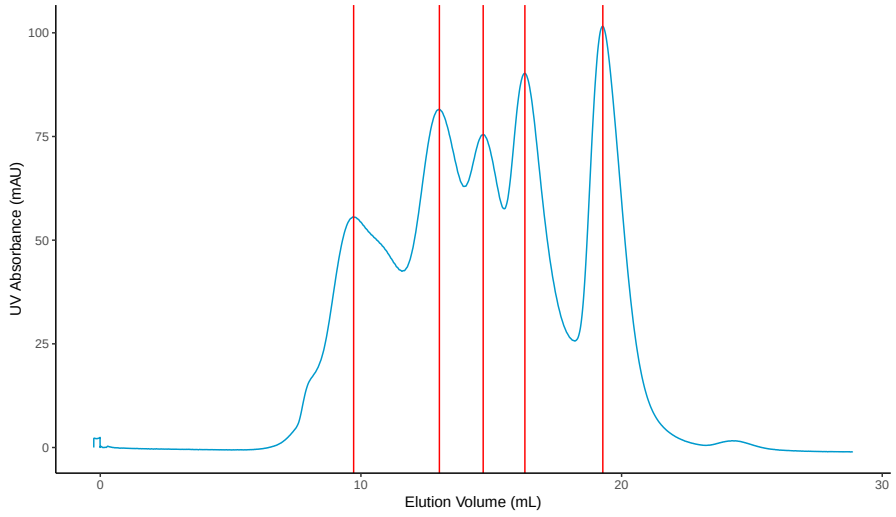
B.1 Protein gel filtration standards

Table B.1: Average distribution constants for gel filtration protein standards, calculated using peak elution volumes (Fig. B.1)**.**

Species	Molecular Weight (Da)	Component	Elution volume (mL)	K_{av}
Dextran Blue	2,000,000	V_0	8.11	0.00
Thyroglobin	670,000	V_e	9.72	0.13
Y-Globulin	158,000	V_e	13.01	0.41
Ovalbumin	44,000	V_e	14.69	0.55
Myoglobin	17,000	V_e	16.29	0.68
Vitamin B12	1,350	V_e	19.28	0.93
Acetone	58	V_C	20.09	1.00



(a)



(b)

Figure B.1: Measurement of size exclusion chromatography standard elution volumes. UV (280 nm) absorbance peaks are recorded in Table B.1.

C

Chapter 3: Chain exchange of ultralong antigen-unknown antibodies

C.1 Antigen-unknown antibody sequence discovery

The novel low-throughput sequencing to achieve paired heavy and light chain antibody sequences was applied to measure the kinetics of ASCs producing cross-reactive antibodies following serial homologous and heterologous immunisation with different serotypes of inactivated FMDV vaccines [304]. PMBCs were isolated from blood samples taken on day 25, corresponding to 4 days post boost, from an animal receiving a homologous prime and boost vaccination regime with inactivated FMDV serotype A/22/Iraq/24/64 'A22', and sorted as described in Section 3.3. The gating strategy used here to sort FMDV-specific ASCs performed less stringently than anticipated, and henceforth, these antibodies were described as 'FMDV-immunised', rather than 'FMDV-specific'. Sorted FMDV-immunised ASCs were sequenced as described in Section 3.3. In both the A22 immunised, and non-immunised animal, Sanger-sequencing revealed several observations of enriched chain usage. The frequency of achieving a single ASC per well, producing 1 functional heavy chain sequence and 1 functional light chain sequence was 23 %. After removal of duplicates, 64 unique heavy, and 54 unique light chain sequences were identified in

the first sequencing plate, and 52 unique heavy, and 69 unique light chains were identified from the second sequencing plate of sorted PBMCs from the adjuvant-only animal. The disparity between the unique heavy and light chain counts might be explained by pairing promiscuity, where a single heavy chain is observed pairing with multiple light chains and vice versa, and hence were of particular interest to this project. The antibody repertoires of the A22-immunised and control animal included 3 sets of promiscuous paired antibodies, as summarised in Fig. C.1.

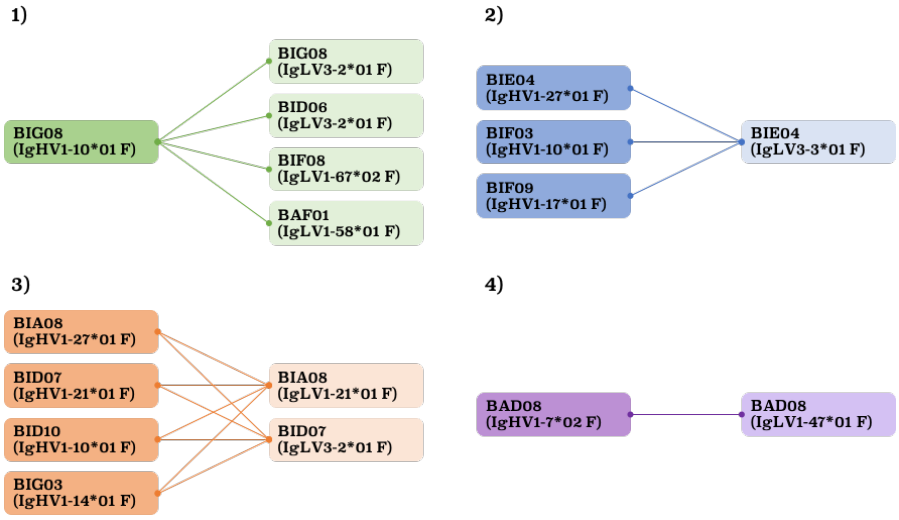


Figure C.1: Sequence clustering of naturally observed antibody pairing combinations from the A22-immunised and adjuvant-only control groups of animals produced sets of pairing promiscuity. Set 1) 4 light chains naturally paired with a common heavy chain. Set 2) 3 distinct heavy chains naturally paired with a common light chain. Set 3) 2 light chains naturally paired with 4 distinct heavy chains, each with varying length CDR3H. Set 4) An antigen-unknown ultralong Fab from the adjuvant group. 'BI' prefix indicates antibody discovered from A22-immunised animal. 'BA' indicates antibody discovered from animal receiving adjuvant only.

C.2 Transient transfection of antigen-unknown ultralong Fabs

C.2.1 Antigen-unknown Fab expression screening

The observation of pairing promiscuity within natural antibody repertoires provides grounds that these alternative pairing combinations should produce viable Fabs *in vitro*, and hence offer a strong starting position for chain-exchanged structural biology experiments.

The expression of Fabs by transient transfection was screened in 1 mL Expi29F™ cultures, co-transfected with 0.5 µg/mL of each heavy/light chain encoding plasmid (Section 2.2.3). Recombinant Fabs were detected in crude media samples by western blot, targeting the polyhistidine tail at the C terminus of the recombinant heavy chain (Fig. C.2). In this reducing conditions blot, the relative length of the CDR3H is reflected in the migration of the bands. Set 1 (lanes 2-5), with it's common heavy chain shows bands at the same migration pattern, but slightly different levels of expression when combined with differing light chains. Similarly, expression levels appear to be impacted for Set 2 (lanes 6-8), in which a common light chain is paired with distinct heavy chains. Significantly, a systemic reduction in expression efficiency within Set 3 (lanes 9-16), where pairing with the light chain encoded by IgLV1-21*01 significantly reduced recombinant Fab expression of all heavy-light combinations, compared to those paired with IgLV3-2*01. Set 3 also provides evidence that expression efficiency appears to be affected by both heavy and light chain contributions, with a broadly similar pattern of expression efficiency within these 2 distinct clades of expression efficiency. Finally, lane 17 shows strong expression efficiency for the recombinant ultralong, antigen-unknown, Fab D08. The varied expression efficiency observed despite all constructs being codon optimised and transfected in equal molar quantities could be explained by steric clashes at the VH:VL interface, particularly around the CDR3H, or other differences affecting combinatorial stability.

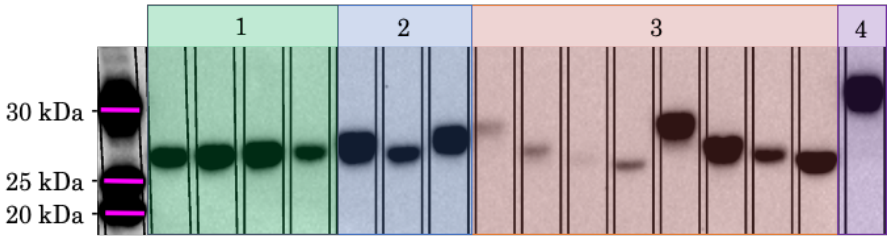


Figure C.2: Reducing anti-polyhistidine immunoblot to show relative recombinant expression efficiency of these four identified sets of naturally paired antibodies. Samples reduced using (10 % βME). Selected antibodies (Fig. C.1) expressed in Expi293F™ cells, ordered as follows: set 1; 4 antibodies with identical heavy chain but differing light chains, set 2; 3 antibodies with identical light chain but differing heavy chains, set 3; 4 distinct heavy chains each paired with 2 distinct light chains, set 4; antigen-unknown ultralong antibody.

C.2.2 Heavy-light chain recombination affects recombinant expression efficiency

The expression of recombinant Fabs in this study was unexpectedly inequal for alternative pairings of heavy and light chains, despite codon optimisation for expression in Expi293F™ cells, and equimolar concentrations used for transfection. The fact that immunoblot signal was affected only in cases of alternative recombination pairings indicates this is consequent on chain recombination, rather than any variation in specific construct expression efficiency. At the protein level, this may be explained by increased energetic cost of recombination, due to steric clashes or electrochemical repulsions at the VH:VL interface.

D

Chapter 4: Functional and structural characterisation of ultralong antibodies specific for Foot-and-mouth disease virus

D.1 VHH ELISA confirms non-binding of Fabs #117 and #126 to intact SAT2

To compare the topotype specificity of antibodies #117 and #126, a homologous ELISA was performed by Kay Humphries of Pirbright, using vaccinia virus prepared SAT2/Zim ('220516-SAT2m'), baculovirus prepared SAT2/Eth ('Ethiopia-28D22'), and baculovirus prepared SAT2/Egypt ('Egypt Classic'), and IgG #117 prepared by the Immunological Toolbox.

Upon heating the SAT2/Egypt VLPs, IgG #117 is able to recognise only the disrupted VLPs, indicating IgG #117 is binding to an internal epitope of SAT2 (Fig. D.1). Due to it's inherent instability, SAT2/Zim VLPs gave a similar signal at both temperatures, hence indicating that SAT2/Zim was already disrupted prior to the heat treatment step, and hence IgG #117 gives a similar signal in both temperatures (Fig. D.1). SAT2/Ethiopia on the other hand, gave no signal with or without heat treatment of 75 °C for 15 min, either demonstrating the greatly enhanced thermostability of SAT2/Eth over SAT2/Zim, or that SAT2/Eth is

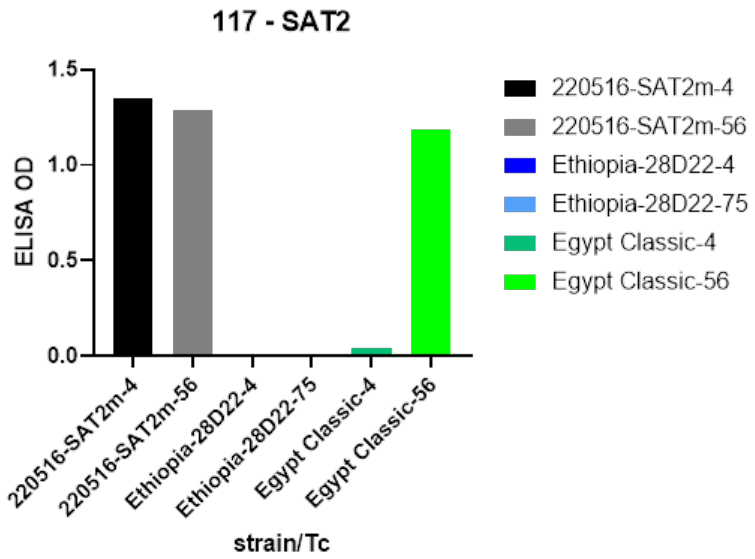


Figure D.1: Homologous ELISA to assess topospecificity of IgG #117 within intact and disrupted serotype SAT2 VLPs.

sufficiently distinct from SAT2/Zim that even when disrupted, IgG #117 finds no complimentary epitope.

To verify this observation, 2 batches of SAT2/Eth, and the SAT2 standard SAT2/Egypt were treated with a variety of temperatures before detection by homologous ELISA using both IgG #117 and M311, a SAT2-binding antibody that preferentially recognises disrupted SAT2, but can also bind to intact SAT2.

Once again, IgG #117 makes no response to SAT2/Eth, following any heat treatment and across two separate production batches, but does bind to dissociated SAT2/Egypt (those heated above 56 °C), but does not recognise intact SAT2/Egypt (4 °C) (Fig. D.2). The M311 control antibody (tested only on SAT2 VLPs heated to 65 °C and 75 °C, shows binding to both SAT2/Eth and SAT2/Egypt, indicating it is recognising dissociated Ethiopia, and therefore it the epitopic differences between SAT2/Eth and SAT2/Egypt (and to SAT2/Zim) must be distinct enough to prevent binding of IgG #117. Note that the loss of all signal for the 75 °C treated VLPs may show irreversible denaturation of either IgG, or VLP, or both. Also note that a background measurement was taken and subtracted for each assay, resulting in some negative optical density (OD) values.

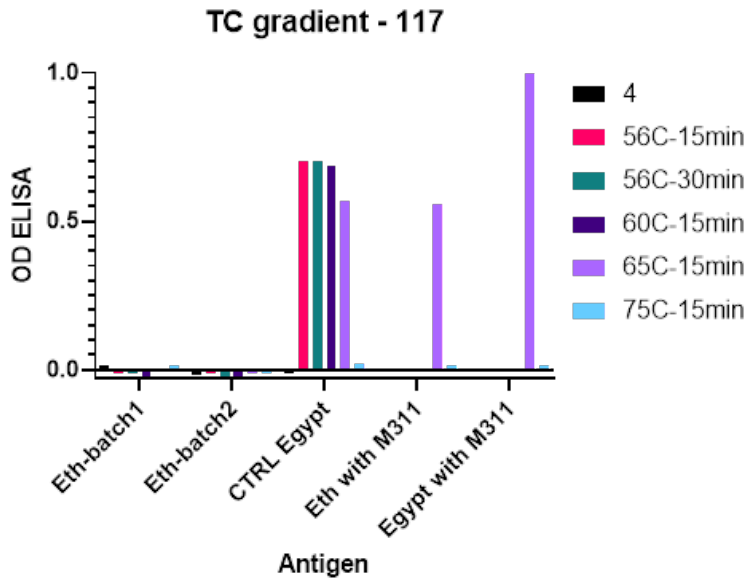


Figure D.2: Homologous ELISA to assess topospecificity of IgG #117 serotype SAT2 VLPs variably heat-treated to induce dissociation.

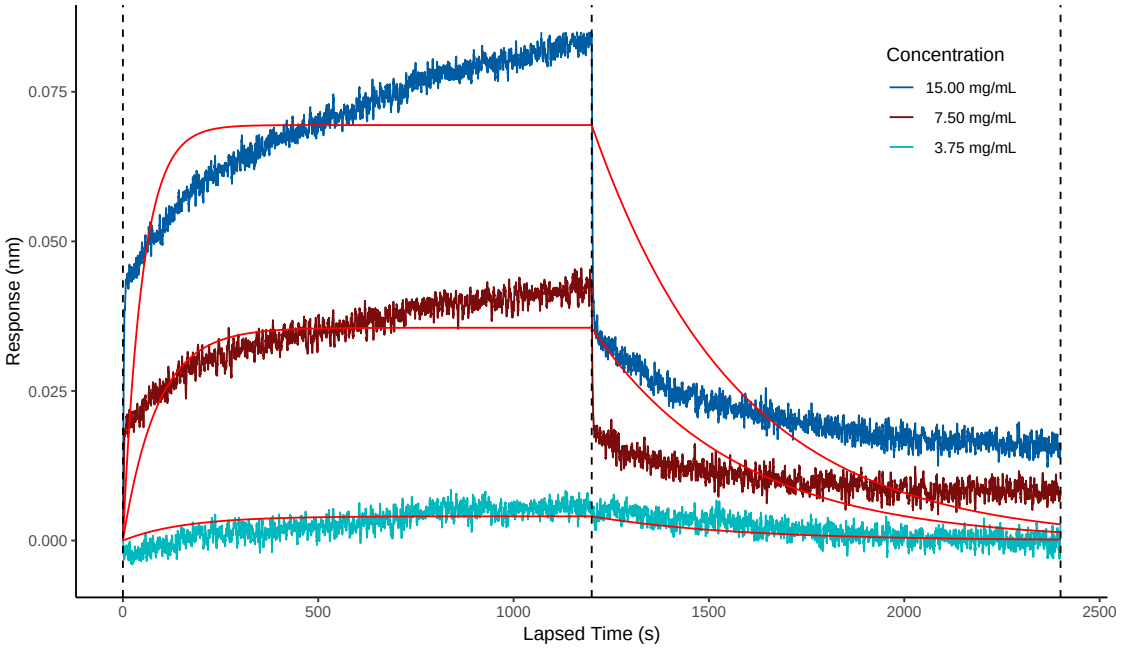
D.2 BLI experimental optimisation

D.2.1 On Fab purification buffer compatibility

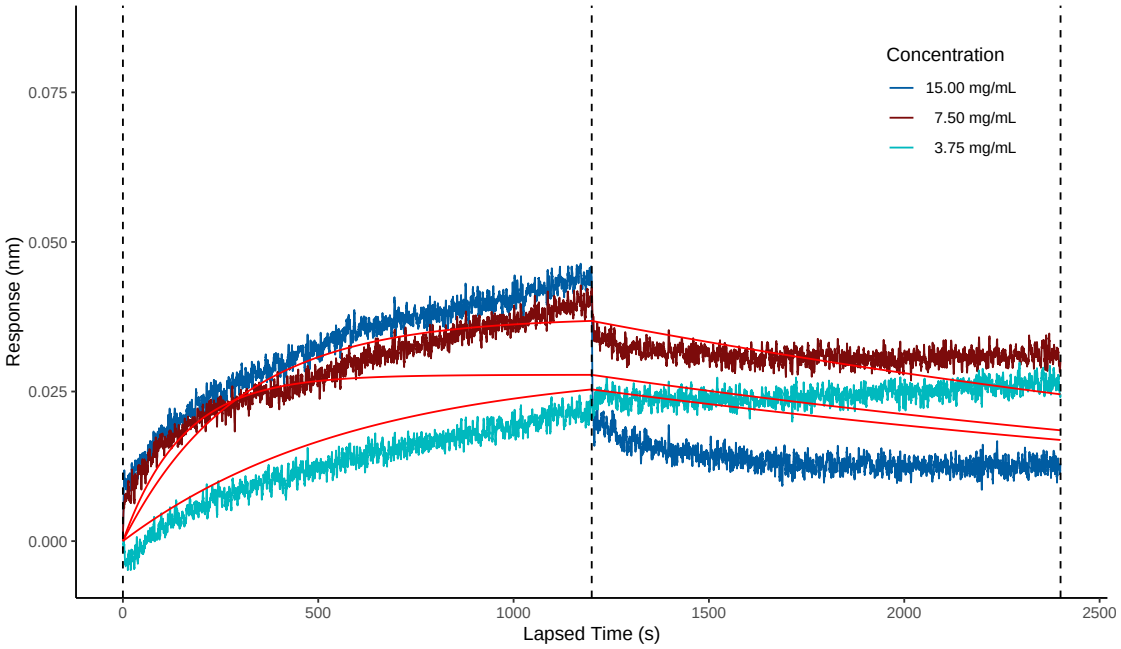
To begin, the Fab buffer composition was assessed for producing BLI traces. An aliquot of Fab #30 was re-purified using NAb™ Protein G spin columns (Section 2.6.6), with subsequent desalting using using a Zeba™ spin column (Section 2.6.7).

The BLI response of Protein G re-purified Fab #30 was greater than that of the parent stock, and produced traces closer to the model curves (red traces) (Fig. D.3), so Fabs #117 and #126 were similarly re-purified and desalted. The association curves should plateau once all binding sites on the immobilised antigen (O1M VLPs) are saturated. The fact none of these curves plateau indicates a more complex interaction than the 1:1 bimolecular model being applied. To remedy this, the association incubation was extended in subsequent studies to enable closer fitting against modelled Michaelis-Menten kinetics to provide accurate quantitative assessment of Fab-antigen interaction, though some systemic error will persist through applying the 1:1 bimolecular model, given each VLP is composed of

60-copy of the VLP protomer.



(a)



(b)

Figure D.3: Optimising BLI assay for experimentation with Fabs. Biotinylated O1M VLPs were immobilised to streptavidin biosensor probes (Sartorius) at 5 µg/mL for 350 s before being introduced to Fab #30 at the concentrations shown. Fab #30 prepared with (Fig. D.3a), and without (Fig. D.3b) Protein G re-purification and subsequent desalting step. 1:1 Michaelis-Menten kinetics modelled in red. Association and dissociation steps were each performed for 1200 s. All assays run in HBS-EP buffer at 30 °C on an Octet R8 (Sartorius), as described in Section 2.7.11.

D.2.2 On the improvement of Fab BLI responses

To ensure appropriate antigen loading to the BLI probes to generate a sufficient Fab response, a VLP loading scout assay was performed with loading concentrations of O1M VLP varying from 1.0 µg/mL up to 10 µg/mL, each introduced to a constant Fab #30 concentration of 16.9 µg/mL, or IgG #30 at 7.5 µg/mL (Fig. D.4).

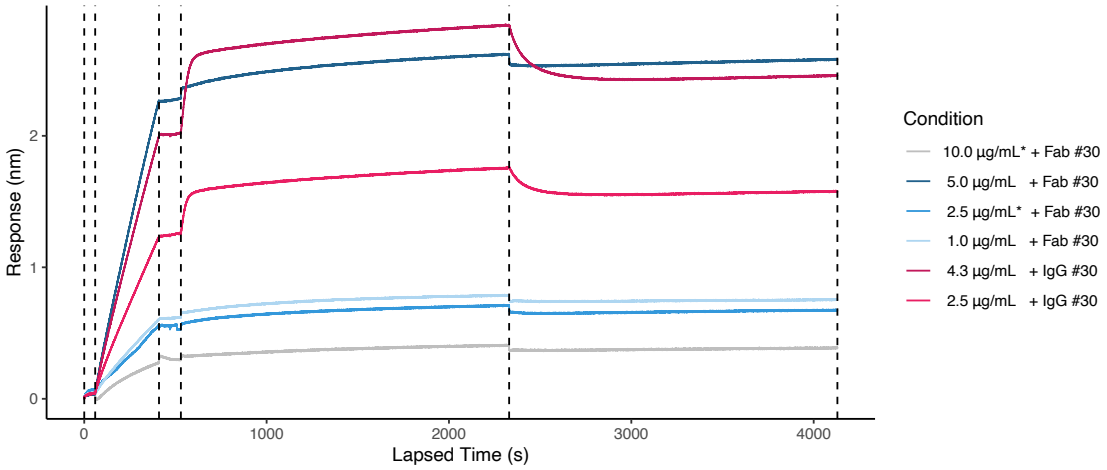


Figure D.4: Optimising VLP loading concentration to produce strong Fab response. Biotinylated O1M VLPs were immobilised to streptavidin biosensor probes (Sartorius) at the concentrations indicated above for 350 s before being introduced to Fab #30 at 300 nM (16.9 µg/mL) or IgG #30 at 50 nM (7.50 µg/mL). Both Fab #30 and IgG #30 prepared with Protein G re-purification and subsequent desalting step. *- indicates VLPs produced from different batches. Association and dissociation steps were each performed for 1800 s. All assays run in HBS-EP buffer at 30 °C on an Octet R8 (Sartorius), as described in Section 2.7.11.

Table D.1: BLI association responses for Fab #30 and IgG #30 upon introduction to O1M VLPs at varied concentrations.

O1M	Isoform	Conc (nM)	Response (nm)
4.3	IgG #30	50	0.6809
5.0	Fab #30	300	0.1715
2.5	IgG #30	50	0.3580
2.5	Fab #30	300	0.0134

Intuitively, greater VLP loading concentrations increased the BLI responses of both Fab and IgG #30. The 10 µg/mL VLP was sourced from an earlier batch, and therefore the extent of degradation may be greater than that of the other VLPs used.

The 5 µg/mL loading resulted in strong association responses for IgG #30 without reaching saturation, whilst Fab #30 gave a modest association response (Table D.1). The response of Fab #30 is only small; with a maximal observed response (R_{max}) of 0.1715 nm at 300 nM) (Fig. D.4 and Table D.1), compared to an R_{max} of 0.6809 nm at 50 nM for IgG #30 (Fig. D.4 and Table D.1), though the true disparity is even greater due to the unequal loading of 5 µg/mL VLP for Fab #30, and 4.3 µg/mL for IgG #30, due to a pipetting error. Given the time and resource consuming VLP production pipeline (Section 2.2.6), and satisfactory association responses without hitting sensor saturation, a loading O1M VLP concentration of 5 µg/mL was selected for subsequent experiments. To improve the Fab response amplitude, greater concentrations of re-purified, desalted Fab were assayed (Fig. D.5).

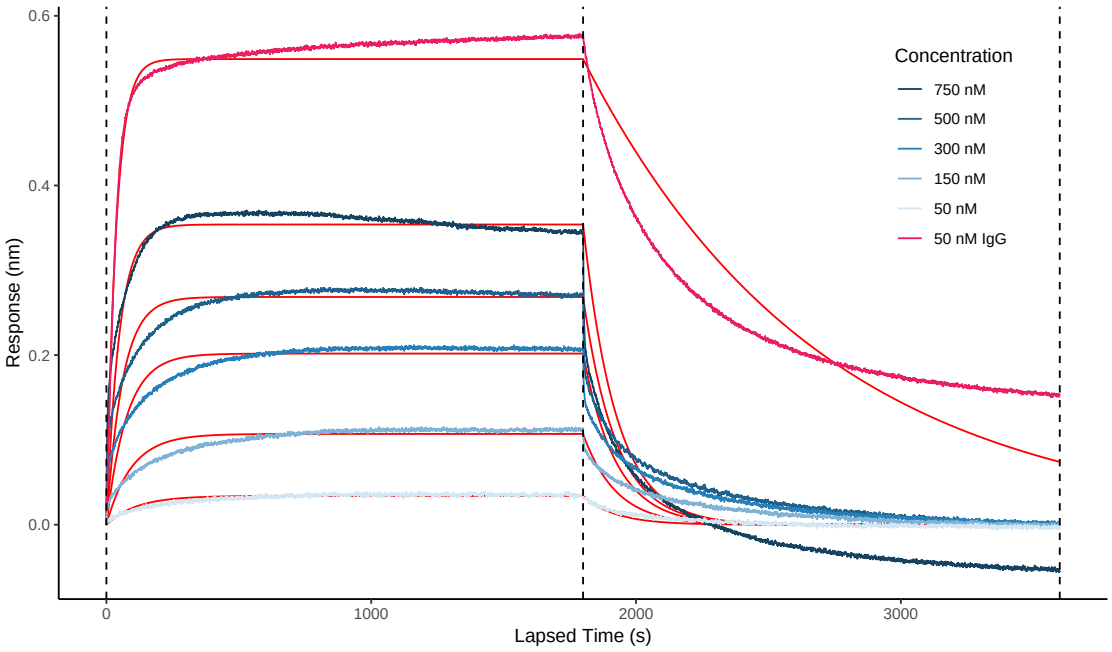


Figure D.5: Optimising Fab concentration for BLI assay. Biotinylated O1M VLPs were immobilised to streptavidin biosensor probes (Sartorius) at 5 µg/mL for 350 s before being introduced to Fab #30 at the concentrations shown. Both Fab #30 and IgG #30 prepared with Protein G re-purification and subsequent desalting step. 1:1 Michaelis-Menten kinetics modelled in red. Association and dissociation steps were each performed for 1800 s. All assays run in HBS-EP buffer at 25 °C on an Octet R8 (Sartorius), as described in Section 2.7.11.

Figure D.5 more clearly illustrates the disparity between the amplitude of Fab and IgG responses when the same concentration (50 nM) is introduced to the same

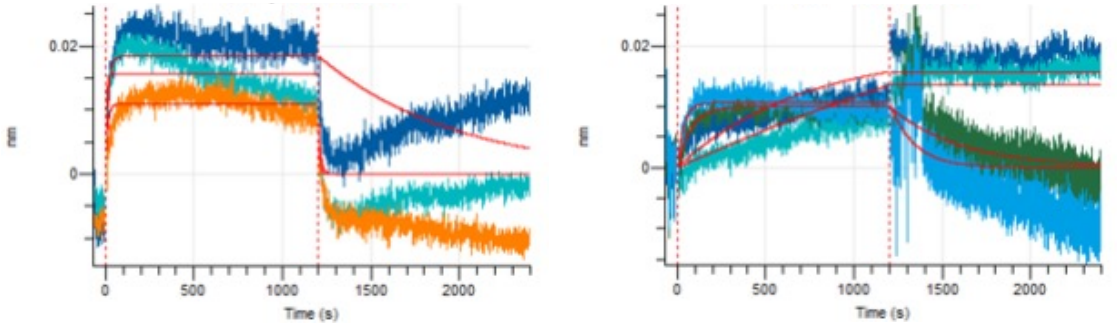
antigen. This effect may be due to the size difference between Fab and IgG, given the BLI measures path length differences of light at the sensor probe tip. In the previously presented BLI assays (Figs. D.3 to D.5), an unloaded sensor was used as a sample reference, to measure the NSB response. Additionally, one sensor in each assay is loaded with VLP, but never introduced to Fab nor IgG to act as a sensor reference, accounting for any baseline drift over the course of the experiment. Both references were subtracted from all sample observations equally. Reference subtraction, particularly of the non-specific binding (NSB component, may also result in lower response amplitudes for Fab compared to IgG, as NSB appeared to be sample-dependent, even varying with concentration.

D.2.3 On the elimination of non-specific binding

The combination of both blocking against NSB and using additional sample reference sensor probes was used to eliminate background NSB. Background NSB was initially sufficiently high that the majority of detected signal was subtracted during the sensor reference subtraction. Several blocking agents were trialled to resolve non-specific binding of Fabs to BLI sensor probes, including bovine serum albumin, biocytin and milk powder. Of these, biocytin performed the best at reducing non-specific binding. Whilst these did reduce NSB, NSB was still unacceptably high (>10 % of total signal), and varied at each concentration point. Consequently, characterisation experiments were run with a reference sensor (without antigen binding to the sensor probes) for each concentration point in addition to blocking with biocytin. The combination of these strategies sufficiently reduced NSB to negligible amplitudes.

D.3 Whole antibody BLI assessment of #117 an d#126 interaction with O1M

Antibody #126 appears to show a weak binding response to O1M VLPs, whereas IgG #117 gave no distinct binding pattern (Fig. D.6). Antibodies were tested and recognition of O1M VLPs assayed by ITB.



(a) IgG #126 **(b)** IgG #117

Figure D.6: BLI characterisation of the O1M VLP interaction with IgG #126 (Fig. D.6a), and with IgG #117 (Fig. D.6b), performed by ITB. Biotinylated O1M VLPs were immobilised to streptavidin biosensor probes (Sartorius) at 5 µg/mL for 650 s before being introduced to D.6a) IgG #126 or (D.6b) IgG #117 at the concentrations varying 25 - 100 nM. Both IgG #126 and #117 were prepared with Protein G purification and subsequent desalting step. 1:1 Michaelis-Menten kinetics modelled in red. Association and dissociation steps were each performed for 1200 s. All assays run in HBS-EP buffer at 25 °C on an Octet R8 (Sartorius), as described in Section 2.7.11.

E

Chapter 5: Preliminary work towards structural characterisation of chain-exchanged antibodies with the bRSV F protein

E.1 Inference of bRSVF/B4 complex MW from SEC elution volume

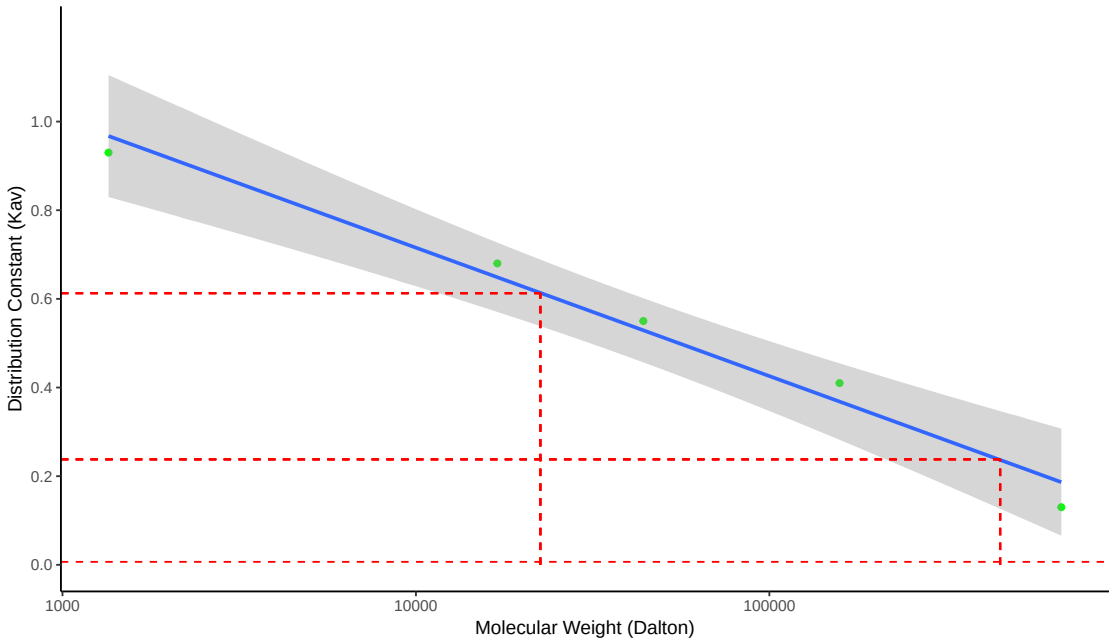


Figure E.1: Relationship between average distribution constant (K_{av}) **and molecular weight.** Markers in red indicate K_{av} values for SEC peaks of complexed bRSVF/B4 (Fig. 5.19)

I like to learn. That's an art, and a science.

— Katherine Johnson, NASA mathematician

Bibliography

References for Chapter 1

[1] Emil von Behring and Shibasaburo Kitasato. 'Ueber das Zustandekommen der Diphtherie-Immunität und der Tetanus-Immunität bei Thieren'. In: *Deutsche medizinische Wochenschrift* 16 (1890), pp. 113–114 (cit. on p. 2).

[2] Gerald M. Edelman. 'Antibody Structure and Molecular Immunology'. In: *Science* 180.4088 (1973), pp. 830–840. DOI: 10.1126/science.180.4088.830 (cit. on p. 2).

[3] Gerald M. Edelman and M. D. Poulik. 'Studies on structural units of the γ-globulins'. In: *Journal of Experimental Medicine* 113.5 (1961), pp. 861–884. DOI: 10.1084/jem.113.5.861 (cit. on p. 2).

[4] Julius M. Cruse and Robert E. Lewis. *Atlas of Immunology*. 3rd ed. Boca Raton: CRC Press, 2010, pp. 1–958. DOI: 10.1201/EBK1439802687 (cit. on pp. 2, 3).

[5] E. E. Max. *Immunoglobulins: molecular genetics*. Ed. by William E. Paul. 3rd ed. New York: Raven Press, 1993, pp. 315–382 (cit. on pp. 2, 3, 11).

[6] Charles A. Janeway Jr, Paul Travers, Mark Walport and Mark J. Shlomchik. *Immunobiology*. Ed. by Penelope Austin, Eleanor Lawrence and Sarah Gibbs. 5th ed. New York: Garland Publishing, 2001 (cit. on pp. 2, 3, 6).

[7] Lisa J. Harris, Eileen Skaletsky and Alexander McPherson. 'Crystallographic structure of an intact IgG1 monoclonal antibody.' In: *Journal of Molecular Biology* 275.5 (1998), pp. 861–872. DOI: 10.1006/jmbi.1997.1508 (cit. on p. 3).

[8] Jéssica Borghesi, Lara Carolina Mario, Marcio Nogueira Rodrigues, Phelipe Oliveira Favaron and Maria Angelica Miglino. 'Immunoglobulin Transport during Gestation in Domestic Animals and Humans—A Review'. In: *Open Journal of Animal Sciences* 04.05 (2014), pp. 323–336. DOI: 10.4236/ojas.2014.45041 (cit. on pp. 4, 14, 22).

[9] Melanie Wilson, Eva Bengtén, Norman W. Miller, L. William Clem, Louis Du Pasquier and Gregory W. Warr. 'A novel chimeric Ig heavy chain from a teleost fish shares similarities to IgD'. In: *Proceedings of the National Academy of Sciences* 94.9 (1997), pp. 4593–4597. DOI: 10.1073/pnas.94.9.4593 (cit. on p. 4).

[10] Kobayashi Kunihiko, Tomonaga Susumu and Kajii Tadashi. 'A second class of immunoglobulin other than IgM present in the serum of a cartilaginous fish, the skate, Raja kenojei: Isolation and characterization'. In: *Molecular Immunology* 21.5 (1984), pp. 397–404. DOI: 10.1016/0161-5890(84)90037-3 (cit. on p. 4).

[11] Fiona A. Harding, Chris T. Amemiya, Ronda T. Litman, Nicholas Cohen and Gary W. Litman. 'Two distinct immunoglobulin heavy chain isotypes in a primitive, cartilaginous fish,Raja erinacea'. In: *Nucleic Acids Research* 18.21 (1990), pp. 6369–6376. DOI: 10.1093/nar/18.21.6369 (cit. on p. 4).

[12] C. Hamers-Casterman, T. Atarchouch, S. Muyldermans, G. Robinson, C. Hamers, E. Bajyana, N. Bendahman and R. Hamilton. 'Naturally occurring antibodies devoid of light chains'. In: *Nature* 363.June (1993), pp. 446–448 (cit. on p. 4).

[13] Andrew S. Greenberg, Austin L. Hughes, Jian Guo, David Avila, E. Churchill McKinney and Martin F. Flajnik. 'A novel "chimeric" antibody class in cartilaginous fish: IgM may not be the primordial immunoglobulin'. In: *European Journal of Immunology* 26.5 (1996), pp. 1123–1129. DOI: 10.1002/eji.1830260525 (cit. on p. 4).

[14] Martin F. Flajnik, Nick Deschacht and Serge Muyldermans. 'A Case Of Convergence: Why Did a Simple Alternative to Canonical Antibodies Arise in Sharks and Camels?' In: *PLoS Biology* 9.8 (2011), e1001120. DOI: 10.1371/journal.pbio.1001120 (cit. on p. 4).

[15] Susumu Tonegawa. 'Somatic generation of antibody diversity'. In: *Nature* 302.5909 (1983), pp. 575–581. DOI: 10.1038/302575a0 (cit. on pp. 5, 26).

[16] W. J. Dreyer and J. C. Bennett. 'The molecular basis of antibody formation: a paradox.' In: *Proceedings of the National Academy of Sciences* 54.3 (1965), pp. 864–869. DOI: 10.1073/pnas.54.3.864 (cit. on pp. 5, 26).

[17] P. Early, H. Huang, M. Davis, K. Calame and L. Hood. 'An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH'. In: *Cell* 19.4 (1980), pp. 981–992. DOI: 10.1016/0092-8674(80)90089-6 (cit. on pp. 5, 11, 26).

[18] Hitoshi Sakano, Richard Maki, Yoshikazu Kurosawa, William Roeder and Susumu Tonegawa. 'Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes'. In: *Nature* 286.5774 (1980), pp. 676–683. DOI: 10.1038/286676a0 (cit. on pp. 5, 11, 26).

[19] Nobumichi Hozumi and Susumu Tonegawa. 'Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions.' In: *Proceedings of the National Academy of Sciences* 73.10 (1976), pp. 3628–3632. DOI: 10.1073/pnas.73.10.3628 (cit. on pp. 5, 11, 26).

[20] Christine Brack, Minoru Hirama, Rita Lenhard-Schuller and Susumu Tonegawa. 'A complete immunoglobulin gene is created by somatic recombination'. In: *Cell* 15.1 (1978), pp. 1–14. DOI: 10.1016/0092-8674(78)90078-8 (cit. on pp. 5, 11, 26).

[21] Ora Bernard, Nobumichi Hozumi and Susumu Tonegawa. 'Sequences of mouse immunoglobulin light chain genes before and after somatic changes'. In: *Cell* 15.4 (1978), pp. 1133–1144. DOI: 10.1016/0092-8674(78)90041-7 (cit. on pp. 5, 26).

[22] M. C. Raff, M. Sternberg and R. B. Taylor. 'Immunoglobulin Determinants on the Surface of Mouse Lymphoid Cells'. In: *Nature* 225.5232 (1970), pp. 553–554. DOI: 10.1038/225553a0 (cit. on pp. 6, 26).

[23] Klaus Rajewsky. 'Clonal selection and learning in the antibody system'. In: *Nature* 381 (1996), pp. 751–758. DOI: 10.1038/381751a0 (cit. on pp. 6, 14, 26).

[24] Masamichi Muramatsu, V. S. Sankaranand, Shrikant Anant, Manabu Sugai, Kazuo Kinoshita, Nicholas O. Davidson and Tasuku Honjo. 'Specific Expression of Activation-induced Cytidine Deaminase (AID), a Novel Member of the RNA-editing Deaminase Family in Germinal Center B Cells'. In: *Journal of Biological Chemistry* 274.26 (1999), pp. 18470–18476. DOI: 10.1074/jbc.274.26.18470 (cit. on pp. 6, 14, 26).

[25] Phuong Pham, Ronda Bransteitter, John Petruska and Myron F. Goodman. 'Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation'. In: *Nature* 424.6944 (2003), pp. 103–107. DOI: 10.1038/nature01760 (cit. on p. 6).

[26] Nancy S. Green, Mark M. Lin and Matthew D. Scharff. 'Somatic hypermutation of antibody genes: a hot spot warms up'. In: *BioEssays* 20.3 (1998), pp. 227–234. DOI: 10.1002/(SICI)1521-1878(199803)20:3<227::AID-BIES6>3.0.CO;2-O (cit. on p. 6).

[27] Javier M. Di Noia and Michael S. Neuberger. 'Molecular Mechanisms of Antibody Somatic Hypermutation'. In: *Annual Review of Biochemistry* 76.1 (2007), pp. 1–22. DOI: 10.1146/annurev.biochem.76.061705.090740 (cit. on pp. 6, 26).

[28] Ruben Barroso, W. Ivan Morrison and Liam J. Morrison. 'Molecular Dissection of the Antibody Response: Opportunities and Needs for Application in Cattle'. In: *Frontiers in Immunology* 11.1 (2020), p. 1175. DOI: 10.3389/fimmu.2020.01175 (cit. on p. 6).

[29] Eduard O. Roos, William N. Mwangi, Wilhelm Gerner, Ryan Waters and John A. Hammond. 'Optimized Multicolour Immunofluorescence Panel for Cattle T Cell Phenotyping by an 8-Colour, 10-Parameter Panel'. In: *bioRxiv* (2021), pp. 1–12. DOI: 10.1101/2022.01.12.476027 (cit. on p. 6).

[30] Eduard O. Roos, Marie Bonnet-Di Placido, William N. Mwangi, Katy Moffat, Lindsay M. Fry, Ryan Waters and John A. Hammond. 'OMIP-085: Cattle B-cell phenotyping by an 8-color panel'. In: *Cytometry* (2022). DOI: 10.1002/cyto.a.24683 (cit. on pp. 6, 293).

[31] Surinder S. Saini, Brian Allore, Robert M. Jacobs and Azad K. Kaushik. 'Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM antibodies'. In: *European Journal of Immunology* 29.8 (1999), pp. 2420–2426. DOI: 10.1002/(SICI)1521-4141(199908)29:08<2420::AID-IMMU2420>3.0.CO;2-A (cit. on pp. 6, 10, 12–14, 18, 23, 92, 98, 140, 289, 293).

[32] Yfke Pasman, Surinder S. Saini, Elspeth Smith and Azad K. Kaushik. 'Organization and genomic complexity of bovine λ-light chain gene locus'. In: *Veterinary Immunology and Immunopathology* 135.3-4 (2010), pp. 306–313. DOI: 10.1016/j.vetimm.2009.12.012 (cit. on pp. 6, 20, 289).

[33] Marie-Paule Lefranc. 'Immunoglobulin and T Cell Receptor Genes: IMGT® and the Birth and Rise of Immunoinformatics'. In: *Frontiers in Immunology* 5.FEB (2014), pp. 1–22. DOI: 10.3389/fimmu.2014.00022 (cit. on p. 7).

[34] Li Ma, Tong Qin, Dan Chu, Xueqian Cheng, Jing Wang, Xifeng Wang, Peng Wang, Haitang Han, Liming Ren, Robert Aitken, Lennart Hammarström, Ning Li and Yaofeng Zhao. 'Internal Duplications of DH, JH, and C Region Genes Create an Unusual IgH Gene Locus in Cattle'. In: *The Journal of Immunology* 196.10 (2016), pp. 4358–4366. DOI: 10.4049/jimmunol.1600158 (cit. on pp. 7–11, 13, 96, 289).

[35] Marie-Paule Lefranc and Gérard Lefranc. *The Immunoglobulin FactsBook*. Harcourt: Academic Press, 2001 (cit. on p. 7).

[36] Marie-Paule Lefranc. 'Nomenclature of the Human Immunoglobulin Heavy (IGH) Genes'. In: *Experimental and Clinical Immunogenetics* 18.2 (2001), pp. 100–116. DOI: 10.1159/000049189 (cit. on p. 7).

[37] M. C. Sinclair, J. Gilchrist and Robert Aitken. 'Bovine IgG repertoire is dominated by a single diversified VH gene family'. In: *The Journal of Immunology* 159.8 (1997), pp. 3883–3889 (cit. on pp. 7, 8, 11, 12, 289).

[38] Mikael Niku, Jenni Liljavirta, Keith Durkin, Eero Schroderus and Antti Iivanainen. 'The bovine genomic DNA sequence data reveal three IGHV subgroups, only one of which is functionally expressed'. In: *Developmental and Comparative Immunology* 37.3-4 (2012), pp. 457–461. DOI: 10.1016/j.dci.2012.02.006 (cit. on pp. 7, 11, 289).

[39] Shawn J. Berens, Dwane E. Wylie and Osvaldo J. Lopez. 'Use of a single V(H) family and long CDR3s in the variable region of cattle Ig heavy chains'. In: *International Immunology* 9.1 (1997), pp. 189–199. DOI: 10.1093/intimm/9.1.189 (cit. on pp. 7, 12, 14, 289).

[40] Stefanie Walther, Manfred Tietze, Claus-Peter Czerny, Sven König and Ulrike S. Diesterbeck. 'Development of a bioinformatics framework for the detection of gene conversion and the analysis of combinatorial diversity in immunoglobulin heavy chains in four cattle breeds'. In: *PLoS ONE* 11.11 (2016), pp. 1–31. DOI: 10.1371/journal.pone.0164567 (cit. on pp. 7, 10, 12).

[41] Jenni Liljavirta, Mikael Niku, Tiina Pessa-Morikawa, Anna Ekman and Antti Iivanainen. 'Expansion of the Preimmune Antibody Repertoire by Junctional Diversity in Bos taurus'. In: *PLoS ONE* 9.6 (2014). Ed. by Sebastian D. Fugmann, e99808. DOI: 10.1371/journal.pone.0099808 (cit. on p. 7).

[42] Claude-Agnès Reynaud, Viviane Anquez, Hélène Grimal and Jean-Claude Weill. 'A hyperconversion mechanism generates the chicken light chain preimmune repertoire'. In: *Cell* 48.3 (1987), pp. 379–388. DOI: 10.1016/0092-8674(87)90189-9 (cit. on pp. 7, 12, 26).

[43] Craig B. Thompson and Paul E. Neiman. 'Somatic diversification of the chicken immunoglobulin light chain gene is limited to the rearranged variable gene segment'. In: *Cell* 48.3 (1987), pp. 369–378. DOI: 10.1016/0092-8674(87)90188-7 (cit. on pp. 7, 12, 26).

[44] M. Mousavi, Hodjattallah Rabbani, L. Pilström and Lennart Hammarström. 'Characterization of the gene for the membrane and secretory form of the IgM heavy-chain constant region gene (Cμ) of the cow (Bos taurus)'. In: *Immunology* 93.4 (1998), pp. 581–588. DOI: 10.1046/j.1365-2567.1998.00444.x (cit. on p. 8).

[45] Imre Kacskovics and John E. Butler. 'The heterogeneity of bovine IgG2—VIII. The complete cDNA sequence of bovine IgG2a (A2) and an IgG1'. In: *Molecular Immunology* 33.2 (1996), pp. 189–195. DOI: 10.1016/0161-5890(95)00107-7 (cit. on p. 8).

[46] Hodjattallah Rabbani, William R. Brown, John E. Butler and Lennart Hammarström. 'Genetic polymorphism of the IGHG3 gene in cattle'. In: *Immunogenetics* 46.4 (1997), pp. 326–331. DOI: 10.1007/s002510050279 (cit. on p. 8).

[47] D. B. A. Symons, C. A. Clarkson and D. Beale. 'Structure of bovine immunoglobulin constant region heavy chain gamma 1 and gamma 2 genes'. In: *Molecular Immunology* 26.9 (1989), pp. 841–850. DOI: 10.1016/0161-5890(89)90140-5 (cit. on p. 8).

[48] M. Mousavi, Hodjattallah Rabbani and Lennart Hammarström. 'Characterization of the bovine ϵ gene'. In: *Immunology* 92.3 (1997), pp. 369–373. DOI: 10.1046/j.1365-2567.1997.00352.x (cit. on p. 8).

[49] William R. Brown, Hodjattallah Rabbani, John E. Butler and Lennart Hammarström. 'Characterization of the bovine C α gene'. In: *Immunology* 91.1 (1997), pp. 1–6. DOI: 10.1046/j.1365-2567.1997.00227.x (cit. on p. 8).

[50] Yaofeng Zhao, Imre Kacskovics, Qiang Pan, David A. Liberles, Janos Geli, Scott K. Davis, Hodjattallah Rabbani and Lennart Hammarstrom. 'Artiodactyl IgD: The Missing Link'. In: *The Journal of Immunology* 169.8 (2002), pp. 4408–4416. DOI: 10.4049/jimmunol.169.8.4408 (cit. on p. 8).

[51] C. L. Parng, Susan Hansal, Richard A Goldsby and Barbara A. Osborne. 'Gene conversion contributes to Ig light chain diversity in cattle.' In: *The Journal of Immunology* 157.12 (1996), pp. 5478–5486 (cit. on pp. 8, 11, 12, 22, 289).

[52] Madhuri Koti, Galina Kataeva and Azad K. Kaushik. 'Organization of DH-Gene Locus is Distinct in Cattle'. In: *Animal Genomics for Animal Health* 132 (2008), pp. 307–313. DOI: 10.1159/000317176 (cit. on p. 8).

[53] Brevin A. Smider and Vaughn V. Smider. 'Formation of ultralong DH regions through genomic rearrangement'. In: *BMC Immunology* 21.1 (2020), p. 30. DOI: 10.1186/s12865-020-00359-8 (cit. on pp. 9, 11, 13, 289).

[54] Thaddeus C. Deiss, Melissa L. Vadnais, Feng Wang, Patricia L. Chen, Ali Torkamani, Waithaka Mwangi, Marie-Paule Lefranc, Michael F. Criscitiello and Vaughn V. Smider. 'Immunogenetic factors driving formation of ultralong VH CDR3 in Bos taurus antibodies'. In: *Cellular and Molecular Immunology* 16.1 (2019), pp. 53–64. DOI: 10.1038/cmi.2017.117 (cit. on pp. 10, 11, 13–16, 18, 289).

[55] Véronique Giudicelli, Patrice Duroux, Chantal Ginestoux, Géraldine Folch, Joumana Jabado-Michaloud, Denys Chaume and Marie-Paule Lefranc. 'IMGT/LIGM-DB, the IMGT(R) comprehensive database of immunoglobulin and T cell receptor nucleotide sequences'. In: *Nucleic Acids Research* 34.90001 (2006), pp. D781–D784. DOI: 10.1093/nar/gkj088 (cit. on p. 10).

[56] Fumihiko Matsuda, Kazuo Ishii, Patrice Bourvagnet, Kei-ichi Kuma, Hidenori Hayashida, Takashi Miyata and Tasuku Honjo. 'The Complete Nucleotide Sequence of the Human Immunoglobulin Heavy Chain Variable Region Locus'. In: *Journal of Experimental Medicine* 188.11 (1998), pp. 2151–2162. DOI: 10.1084/jem.188.11.2151 (cit. on p. 10).

[57] Manuel Ruiz, Nathalie Pallarès, Valérie Contet, Valérie Barbié and Marie-Paule Lefranc. 'The Human Immunoglobulin Heavy Diversity (IGHD) and Joining (IGHJ) Segments'. In: *Experimental and Clinical Immunogenetics* 16.3 (1999), pp. 173–184. DOI: 10.1159/000019109 (cit. on p. 10).

[58] Samuel O. Oyola, Sonal P. Henson, Benjamin Nzau, Elizabeth Kibwana and Vishvanath Nene. 'Access to ultra-long IgG CDRH3 bovine antibody sequences using short read sequencing technology'. In: *Molecular Immunology* 139 (2021), pp. 97–105. DOI: 10.1016/j.molimm.2021.08.017 (cit. on pp. 10, 15, 289).

[59] Azad K. Kaushik, Farbod Shojaei and Surinder S. Saini. 'Novel insight into antibody diversification from cattle'. In: *Veterinary Immunology and Immunopathology* 87.3-4 (2002), pp. 347–350. DOI: 10.1016/S0165-2427(02)00063-6 (cit. on pp. 10, 12–14).

[60] Surinder S. Saini and Azad K. Kaushik. 'Extensive CDR3H length heterogeneity exists in bovine foetal VDJ rearrangements.' In: *Scandinavian Journal of Immunology* 55.2 (2002), pp. 140–8 (cit. on pp. 10, 12, 13).

[61] Surinder S. Saini, William Farrugia, Paul A. Ramsland and Azad K. Kaushik. 'Bovine IgM antibodies with exceptionally long complementarity-determining region 3 of the heavy chain share unique structural properties conferring restricted V H + Vλ pairings'. In: *International Immunology* 15.7 (2003), pp. 845–853. DOI: 10.1093/intimm/dxg083 (cit. on pp. 10, 13, 20, 23, 27, 92, 157, 163, 289).

[62] Cyrus Chothia and Arthur M. Lesk. 'Canonical Structures for the Hypervariable Regions of Immunoglobulins'. In: *Journal of Molecular Biology* 196 (1987), pp. 901–917 (cit. on pp. 11, 22).

[63] Ian A. Wilson and Robyn L. Stanfield. 'Antibody-antigen interactions'. In: *Current Opinion in Structural Biology* 3 (1993), pp. 113–118 (cit. on pp. 11, 18, 22).

[64] Osvaldo J. Lopez, Carlos Perez and Dwane E. Wylie. 'A single V(H) family and long CDR3s are the targets for hypermutation in bovine immunoglobulin heavy chains'. In: *Immunological Reviews* 162.1 (1998), pp. 55–66. DOI: 10.1111/j.1600-065X.1998.tb01429.x (cit. on pp. 11, 13, 14, 289).

[65] M. R. Lucier, R. E. Thompson, James Waire, A. W. Lin, B. A. Osborne and R. A. Goldsby. 'Multiple sites of Vλ diversification in cattle'. In: *Journal of Immunology* 161.10 (1998), pp. 5438–5444. DOI: 10.4049/jimmunol.161.10.5438 (cit. on pp. 11, 12, 22).

[66] Frederick W. Alt, T. Keith Blackwell and George D. Yancopoulos. 'Development of the Primary Antibody Repertoire'. In: *Science* 238.4830 (1987), pp. 1079–1087. DOI: 10.1126/science.3317825 (cit. on pp. 11, 13).

[67] Wayne T. McCormack, Larry W. Tjoelker and Craig B. Thompson. 'Avian B-Cell Development: Generation of an Immunoglobulin Repertoire by Gene Conversion'. In: *Annual Review of Immunology* 9.1 (1991), pp. 219–241. DOI: 10.1146/annurev.iy.09.040191.001251 (cit. on pp. 12, 26).

[68] Robert S. Becker and Katherine L. Knight. 'Somatic diversification of immunoglobulin heavy chain VDJ genes: Evidence for somatic gene conversion in rabbits'. In: *Cell* 63.5 (1990), pp. 987–997. DOI: 10.1016/0092-8674(90)90502-6 (cit. on pp. 12, 26).

[69] Katherine L. Knight and Candace R. Winstead. 'Generation of antibody diversity in rabbits'. In: *Current Opinion in Immunology* 9.2 (1997), pp. 228–232. DOI: 10.1016/S0952-7915(97)80140-9 (cit. on pp. 12, 26).

[70] Peter D. Weinstein, Arthur O. Anderson and Rose G. Mage. 'Rabbit IgH sequences in appendix germinal centers: VH diversification by gene conversion-like and hypermutation mechanisms'. In: *Immunity* 1.8 (1994), pp. 647–659. DOI: 10.1016/1074-7613(94)90036-1 (cit. on pp. 12, 26).

[71] Katherine L. Knight and Robert S. Becker. 'Molecular basis of the allelic inheritance of rabbit immunoglobulin VH allotypes: Implications for the generation of antibody diversity'. In: *Cell* 60.6 (1990), pp. 963–970. DOI: 10.1016/0092-8674(90)90344-E (cit. on p. 12).

[72] Claude-Agnès Reynaud, Charles R. Mackay, Rolf G. Müller and Jean-Claude Weill. 'Somatic generation of diversity in a mammalian primary lymphoid organ: The sheep ileal Peyer's patches'. In: *Cell* 64.5 (1991), pp. 995–1005. DOI: 10.1016/0092-8674(91)90323-Q (cit. on pp. 12, 22).

[73] Claude-Agnès Reynaud, Corinne Garcia, Wayne R. Hein and Jean-Claude Weill. 'Hypermutation Generating the Sheep Immunoglobulin Repertoire Is an Antigen-Independent Process'. In: *Cell* 80.1 (1995), pp. 115–125. DOI: 10.1016/0092-8674(95)90456-5 (cit. on p. 12).

[74] Vinciane Dufour, Sophie Malinge and François Nau. 'The Sheep Ig Variable Region Repertoire Consists of a Single VH Family'. In: *Journal of Immunology* 156.6 (1996), pp. 2163–2170 (cit. on p. 12).

[75] Azad K. Kaushik, Marcus E. Kehrli, A. Kurtz, S. Ng, Madhuri Koti, Farbod Shojaei and Surinder S. Saini. 'Somatic hypermutations and isotype restricted exceptionally long CDR3H contribute to antibody diversification in cattle'. In: *Veterinary Immunology and Immunopathology* 127.1-2 (2009), pp. 106–113. DOI: 10.1016/j.vetimm.2008.09.024 (cit. on pp. 12, 14).

[76] Subhash Verma and Robert Aitken. 'Somatic hypermutation leads to diversification of the heavy chain immunoglobulin repertoire in cattle'. In: *Veterinary Immunology and Immunopathology* 145.1-2 (2012), pp. 14–22. DOI: 10.1016/j.vetimm.2011.10.001 (cit. on pp. 12–14).

[77] Madhuri Koti, Galina Kataeva and Azad K. Kaushik. 'Novel atypical nucleotide insertions specifically at VH-DH junction generate exceptionally long CDR3H in cattle antibodies'. In: *Molecular Immunology* 47.11-12 (2010), pp. 2119–2128. DOI: 10.1016/j.molimm.2010.02.014 (cit. on pp. 12–14, 289).

[78] Feng Wang, Damian C. Ekiert, Insha Ahmad, Wenli Yu, Yong Zhang, Omar A. Bazirgan, Ali Torkamani, Terje Raudsepp, Waithaka Mwangi, Michael F. Criscitiello, Ian A. Wilson, Peter G. Schultz and Vaughn V. Smider. 'Reshaping Antibody Diversity'. In: *Cell* 153.6 (2013), pp. 1379–1393. DOI: 10.1016/j.cell.2013.04.049 (cit. on pp. 12, 13, 15–19, 23–25, 94, 98, 99, 105, 108, 109, 116, 128, 131, 132, 140–142, 162, 163, 180, 181, 206, 232, 289–294).

[79] Yaofeng Zhao, Stephen M. Jackson and Robert Aitken. 'The bovine antibody repertoire'. In: *Developmental and Comparative Immunology* 30.1-2 (2006), pp. 175–186. DOI: 10.1016/j.dci.2005.06.012 (cit. on p. 12).

[80] Bin Shi, Long Ma, Xiaoyan He, Xiaomei Wang, Peng Wang, Li Zhou and Xinsheng Yao. 'Comparative analysis of human and mouse immunoglobulin variable heavy regions from IMGT/LIGM-DB with IMGT/HighV-QUEST'. In: *Theoretical Biology and Medical Modelling* 11.1 (2014), p. 30. DOI: 10.1186/1742-4682-11-30 (cit. on pp. 12, 307).

[81] Farbod Shojaei, Surinder S. Saini and Azad K. Kaushik. 'Unusually long germline DH genes contribute to large sized CDR3H in bovine antibodies'. In: *Molecular Immunology* 40.1 (2003), pp. 61–67. DOI: 10.1016/S0161-5890(03)00098-1 (cit. on p. 13).

[82] Robyn L. Stanfield, Ian A. Wilson and Vaughn V. Smider. 'Conservation and diversity in the ultralong third heavy-chain complementarity-determining region of bovine antibodies'. In: *Science Immunology* 1.1 (2016), aaf7962. DOI: 10.1126/sciimmunol.aaf7962 (cit. on pp. 13, 15, 16, 23–25, 94, 98, 108, 109, 116, 140, 141, 162, 163, 180–182, 289, 290, 293).

[83] Jinhui Dong, Jessica A. Finn, Peter A. Larsen, Timothy P. L. Smith and James E. Crowe. 'Structural Diversity of Ultralong CDRH3s in Seven Bovine Antibody Heavy Chains'. In: *Frontiers in Immunology* 10.March (2019), pp. 1–9. DOI: 10.3389/fimmu.2019.00558 (cit. on pp. 13, 15, 16, 19, 23–25, 94, 97, 98, 108, 109, 116, 140, 141, 162, 163, 180, 181, 289, 290, 293).

[84] Robyn L. Stanfield, Zachary T. Berndsen, Ruiqi Huang, Devin Sok, Gabrielle Warner, Jonathan L. Torres, Dennis R. Burton, Andrew B. Ward, Ian A. Wilson and Vaughn V. Smider. 'Structural basis of broad HIV neutralization by a vaccine-induced cow antibody'. In: *Science Advances* 6.22 (2020), eaba0468. DOI: 10.1126/sciadv.aba0468 (cit. on pp. 13, 15, 16, 18, 23–25, 92, 94, 140, 141, 162, 163, 206, 232, 291, 293, 294).

[85] Yong He, Kun Li, Yimei Cao, Zixian Sun, Pinghua Li, Huifang Bao, Sheng Wang, Guoqiang Zhu, Xingwen Bai, Pu Sun, Xuerong Liu, Cheng Yang, Zaixin Liu, Zengjun Lu, Zihe Rao and Zhiyong Lou. 'Structures of Foot-and-mouth Disease Virus with neutralizing antibodies derived from recovered natural host reveal a mechanism for cross-serotype neutralization'. In: *PLOS Pathogens* 17.4 (2021). Ed. by Richard J. Kuhn, e1009507. DOI: 10.1371/journal.ppat.1009507 (cit. on pp. 13, 15, 16, 18, 23–25, 92, 94, 140, 141, 162, 163, 206, 232, 289–291, 293, 294).

[86] Jenni Liljavirta, Anna Ekman, J. S. Knight, A. Pernthaner, A. Iivanainen and Mikael Niku. 'Activation-induced cytidine deaminase (AID) is strongly expressed in the fetal bovine ileal Peyer's patch and spleen and is associated with expansion of the primary antibody repertoire in the absence of exogenous antigens'. In:

Mucosal Immunology 6.5 (2013), pp. 942–949. DOI: 10.1038/mi.2012.132 (cit. on p. 14).

[87] Igor B. Rogozin and Marilyn Diaz. 'Cutting Edge: DGYW/WRCH Is a Better Predictor of Mutability at G:C Bases in Ig Hypermutation Than the Widely Accepted RGYW/WRCY Motif and Probably Reflects a Two-Step Activation-Induced Cytidine Deaminase-Triggered Process'. In: *The Journal of Immunology* 172.6 (2004), pp. 3382–3384. DOI: 10.4049/jimmunol.172.6.3382 (cit. on p. 14).

[88] Veronica Morea, Anna Tramontano, Mauro Rustici, Cyrus Chothia and Arthur M. Lesk. 'Conformations of the third hypervariable region in the VH domain of immunoglobulins 1 1Edited by I. A. Wilson'. In: *Journal of Molecular Biology* 275.2 (1998), pp. 269–294. DOI: 10.1006/jmbi.1997.1442 (cit. on p. 15).

[89] Norelle L. Daly and David J. Craik. 'Bioactive cystine knot proteins'. In: *Current Opinion in Chemical Biology* 15.3 (2011), pp. 362–368. DOI: 10.1016/j.cbpa.2011.02.008 (cit. on p. 17).

[90] G. Lippens, J. Najib, Shoshana J. Wodak and A. Tartar. 'NMR Sequential Assignments and Solution Structure of Chlorotoxin, a Small Scorpion Toxin That Blocks Chloride Channels'. In: *Biochemistry* 34.1 (1995), pp. 13–21. DOI: 10.1021/bi00001a003 (cit. on pp. 17, 19, 94, 289, 291).

[91] Jamie I. Fletcher, Bogdan E. Chapman, Joel P. Mackay, Merlin E. H. Howden and Glenn F. King. 'The structure of versutoxin (δ-atracotoxin-Hv1) provides insights into the binding of site 3 neurotoxins to the voltage-gated sodium channel'. In: *Structure* 5.11 (1997), pp. 1525–1535. DOI: 10.1016/S0969-2126(97)00301-8 (cit. on pp. 17, 19, 94, 289, 291).

[92] Svend Ludvigsen, Lars Thim, Ane M. Blom and Birgitte S. Wulff. 'Solution Structure of the Satiety Factor, CART, Reveals New Functionality of a Well-Known Fold'. In: *Biochemistry* 40.31 (2001), pp. 9082–9088. DOI: 10.1021/bi010433u (cit. on pp. 17, 94).

[93] José C. Martins, Mohammed Enassar, Rudolph Willem, Jean-michel Wieruzeski, Guy Lippens and Shoshana J. Wodak. 'Solution structure of the main α-amylase inhibitor from amaranth seeds'. In: *European Journal of Biochemistry* 268.8 (2001), pp. 2379–2389. DOI: 10.1046/j.1432-1327.2001.02118.x (cit. on pp. 17, 94).

[94] Yuliya V. Korolkova, Eduard V. Bocharov, Kamilla Angelo, Innokenty V. Maslennikov, Olga V. Grinenko, Aleksey V. Lipkin, Elena D. Nosyreva, Kirill A. Pluzhnikov, Søren-peter Olesen, Alexander S. Arseniev and Eugene V. Grishin. 'New Binding Site on Common Molecular Scaffold Provides HERG Channel Specificity of Scorpion Toxin BeKm-1'. In: *Journal of Biological Chemistry* 277.45 (2002), pp. 43104–43109. DOI: 10.1074/jbc.M204083200 (cit. on pp. 17, 19, 94, 289, 291).

[95] Robert E. Oswald, Thomas M. Suchyna, Robert McFeeters, Philip Gottlieb and Frederick Sachs. 'Solution Structure of Peptide Toxins That Block Mechanosensitive Ion Channels'. In: *Journal of Biological Chemistry* 277.37 (2002), pp. 34443–34450. DOI: 10.1074/jbc.M202715200 (cit. on pp. 17, 19, 94, 289, 291).

[96] Katherine J. Nielsen, Michael Watson, David J. Adams, Anna K. Hammarström, Peter W. Gage, Justine M. Hill, David J. Craik, Linda Thomas, Denise Adams, Paul F. Alewood and Richard J. Lewis. 'Solution Structure of μ-Conotoxin PIIIA, a Preferential Inhibitor of Persistent Tetrodotoxin-sensitive Sodium Channels'. In: *Journal of Biological Chemistry* 277.30 (2002), pp. 27247–27255. DOI: 10.1074/jbc.M201611200 (cit. on pp. 17, 19, 94, 289, 291).

[97] David J. Adams, Amanda B. Smith, Christina I. Schroeder, Takahiro Yasuda and Richard J. Lewis. 'ω-Conotoxin CVID Inhibits a Pharmacologically Distinct Voltage-sensitive Calcium Channel Associated with Transmitter Release from Preganglionic Nerve Terminals'. In: *Journal of Biological Chemistry* 278.6 (2003), pp. 4057–4062. DOI: 10.1074/jbc.M209969200 (cit. on pp. 17, 19, 94, 289, 291).

[98] Jorgen Mould, Takahiro Yasuda, Christina I. Schroeder, Aaron M. Beedle, Clinton J. Doering, Gerald W. Zamponi, David J. Adams and Richard J. Lewis. 'The $\alpha 2\delta$ Auxiliary Subunit Reduces Affinity of ω-Conotoxins for Recombinant N-type (Cav2.2) Calcium Channels'. In: *Journal of Biological Chemistry* 279.33 (2004), pp. 34705–34714. DOI: 10.1074/jbc.M310848200 (cit. on pp. 17, 19, 94, 289, 291).

[99] Siddhartha P. Sarma, G. Senthil Kumar, S. Sudarslal, Prathima Iengar, P. Ramasamy, Sujit K. Sikdar, K. S. Krishnan and Padmanabhan Balaram. 'Solution Structure of δ-Am2766: A Highly Hydrophobic δ-Conotoxin fromConus amadis That Inhibits Inactivation of Neuronal Voltage-Gated Sodium Channels'. In: *Chemistry and Biodiversity* 2.4 (2005), pp. 535–556. DOI: 10.1002/cbdv.200590035 (cit. on pp. 17, 19, 94, 289, 291).

[100] Jennifer J. Smith, Justine M. Hill, Michelle J. Little, Graham M. Nicholson, Glenn F. King and Paul F. Alewood. 'Unique scorpion toxin with a putative ancestral fold provides insight into evolution of the inhibitor cystine knot motif'. In: *Proceedings of the National Academy of Sciences* 108.26 (2011), pp. 10478–10483. DOI: 10.1073/pnas.1103501108 (cit. on pp. 17, 94).

[101] Roger J. P. Dawson, Jörg Benz, Peter Stohler, Tim Tetaz, Catherine Joseph, Sylwia Huber, Georg Schmid, Daniela Hügin, Pascal Pflimlin, Gerd Trube, Markus G. Rudolph, Michael Hennig and Armin Ruf. 'Structure of the Acid-sensing ion channel 1 in complex with the gating modifier Psalmotoxin 1'. In: *Nature Communications* 3.1 (2012), p. 936. DOI: 10.1038/ncomms1917 (cit. on pp. 17, 19, 94, 289, 291).

[102] Andrew J. Wommack, Scott A. Robson, Yoshitha A. Wanniarachchi, Andrea Wan, Christopher J. Turner, Gerhard Wagner and Elizabeth M. Nolan. 'NMR Solution Structure and Condition-Dependent Oligomerization of the Antimicrobial Peptide Human Defensin 5'. In: *Biochemistry* 51.48 (2012), pp. 9624–9637. DOI: 10.1021/bi301255u (cit. on pp. 17, 94).

[103] Brad R. Green, Joanna Gajewiak, Sandeep Chhabra, Jack J. Skalicky, Min-Min Zhang, Jean E. Rivier, Grzegorz Bulaj, Baldomero M. Olivera, Doju Yoshikami and Raymond S. Norton. 'Structural Basis for the Inhibition of Voltage-gated Sodium Channels by Conotoxin μO§-GVIIJ'. In: *Journal of Biological Chemistry* 291.13 (2016), pp. 7205–7220. DOI: 10.1074/jbc.M115.697672 (cit. on pp. 17, 19, 94, 289, 291).

[104] Bin Wu, Justin K. Murray, Kristin L. Andrews, Kelvin Sham, Jason Long, Jennifer Aral, Joseph Ligutti, Shanti Amagasu, Dong Liu, Anruo Zou, Xiaoshan Min, Zhulun Wang, Christopher P. Ilch, Thomas J. Kornecook, Min-Hwa Jasmine Lin, Xuhai Be, Les P. Miranda, Bryan D. Moyer and Kaustav Biswas. 'Discovery of Tarantula Venom-Derived Na V 1.7-Inhibitory JzTx-V Peptide 5-Br-Trp24 Analogue AM-6120 with Systemic Block of Histamine-Induced Pruritis'. In: *Journal of Medicinal Chemistry* 61.21 (2018), pp. 9500–9512. DOI: 10.1021/acs.jmedchem.8b00736 (cit. on pp. 17, 19, 94, 289, 291).

[105] Andrew J. Wommack, J. J. Ziarek, Gerhard Wagner and Elizabeth M. Nolan. 'Solution structure of oxidized dimeric form of human defensin 5'. DOI: http://doi.org/10.2210/pdb2MIT/pdb (cit. on pp. 17, 94).

[106] Xinying Jia, Soohyun Kwon, Ching-I. Anderson Wang, Yen-Hua Huang, Lai Y. Chan, Chia Chia Tan, K. Johan Rosengren, Jason P. Mulvenna, Christina I. Schroeder and David J. Craik. 'Semienzymatic Cyclization of Disulfide-rich Peptides Using Sortase A'. In: *Journal of Biological Chemistry* 289.10 (2014), pp. 6627–6638. DOI: 10.1074/jbc.M113.539262 (cit. on pp. 17, 94).

[107] Tien T. Dang, Peta J. Harvey, Lai Yue Chan, Yen-Hua Huang, Quentin Kaas and David J. Craik. 'Mutagenesis of cyclotide Cter 27 exemplifies a robust folding strategy for bracelet cyclotides'. In: *Peptide Science* 114.6 (2022), pp. 2–7. DOI: 10.1002/pep2.24284 (cit. on pp. 17, 94).

[108] Neil Q. McDonald, Risto Lapatto, Judith Murray Rust, Jennifer Gunning, Alexander Wlodawer and Tom L. Blundell. 'New protein fold revealed by a 2.3-Å resolution crystal structure of nerve growth factor'. In: *Nature* 354.6352 (1991), pp. 411–414. DOI: 10.1038/354411a0 (cit. on pp. 17, 94).

[109] Debra R. Holland, Lawrence S. Cousens, Wei Meng and Brian W. Matthews. 'Nerve Growth Factor in Different Crystal Forms Displays Structural Flexibility and Reveals Zinc Binding Sites'. In: *Journal of Molecular Biology* 239.3 (1994), pp. 385–400. DOI: 10.1006/jmbi.1994.1380 (cit. on pp. 17, 94).

[110] Christian Wiesmann, Mark H. Ultsch, Steven H. Bass and Abraham M. de Vos. 'Crystal structure of nerve growth factor in complex with the ligand-binding domain of the TrkA receptor'. In: *Nature* 401.6749 (1999), pp. 184–188. DOI: 10.1038/43705 (cit. on pp. 17, 94).

[111] Shalini Iyer, Demetres D. Leonidas, G. Jawahar Swaminathan, Domenico Maglione, Mauro Battisti, Marina Tucci, M. Graziella Persico and K. Ravi Acharya. 'The Crystal Structure of Human Placenta Growth Factor-1 (PlGF-1), an Angiogenic Protein, at 2.0 Å Resolution'. In: *Journal of Biological Chemistry* 276.15 (2001), pp. 12153–12161. DOI: 10.1074/jbc.M008055200 (cit. on pp. 17, 94).

[112] Kristin M. Fox, James A. Dias and Patrick Van Roey. 'Three-Dimensional Structure of Human Follicle-Stimulating Hormone'. In: *Molecular Endocrinology* 15.3 (2001), pp. 378–389. DOI: 10.1210/mend.15.3.0603 (cit. on pp. 17, 94).

[113] Christopher A. Love, Karl Harlos, Nasim Mavaddat, Simon J. Davis, David I. Stuart, E. Yvonne Jones and Robert M. Esnouf. 'The ligand-binding face of the semaphorins revealed by the high-resolution crystal structure of SEMA4D'. In: *Nature Structural and Molecular Biology* 10.10 (2003), pp. 843–848. DOI: 10.1038/nsb977 (cit. on pp. 17, 94).

[114] Jay Groppe, Jason Greenwald, Ezra Wiater, Joaquin Rodrigues-Leon, Aris N. Economides, Witek Kwiatkowski, Kandan Baban, Markus Affolter, Wylie W. Vale, Juan Carlos Izpisua Belmonte and Senyon Choe. 'Structural basis of BMP signalling inhibition by noggin, a novel twelve-membered cystine knot protien'. In: *The Journal of Bone and Joint Surgery-American Volume* 85.SUPPL. 3 (2003), pp. 52–58. DOI: 10.2106/00004623-200300003-00010 (cit. on pp. 17, 94).

[115] Guennadi Kozlov, Audrey Perreault, Joseph D. Schrag, Morag Park, Miroslaw Cygler, Kalle Gehring and Irena Ekiel. 'Insights into function of PSI domains from structure of the Met receptor PSI domain'. In: *Biochemical and Biophysical Research Communications* 321.1 (2004), pp. 234–240. DOI: 10.1016/j.bbrc.2004.06.132 (cit. on pp. 17, 94).

[116] Qing R. Fan and Wayne A. Hendrickson. 'Structure of human follicle-stimulating hormone in complex with its receptor'. In: *Nature* 433.7023 (2005), pp. 269–277. DOI: 10.1038/nature03206 (cit. on pp. 17, 94).

[117] Michel Pieren, Andrea E. Prota, Claudia Ruch, Dirk Kostrewa, Armin Wagner, Katrin Biedermann, Fritz K. Winkler and Kurt Ballmer-Hofer. 'Crystal Structure of the Orf Virus NZ2 Variant of Vascular Endothelial Growth Factor-E'. In: *Journal of Biological Chemistry* 281.28 (2006), pp. 19578–19587. DOI: 10.1074/jbc.M601842200 (cit. on pp. 17, 94).

[118] Shalini Iyer, Pierre D. Scotney, Andrew D. Nash and K. Ravi Acharya. 'Crystal Structure of Human Vascular Endothelial Growth Factor-B: Identification of Amino Acids Important for Receptor Binding'. In: *Journal of Molecular Biology* 359.1 (2006), pp. 76–85. DOI: 10.1016/j.jmb.2006.03.002 (cit. on pp. 17, 94).

[119] Philip Leonard, Pierre D. Scotney, Talat Jabeen, Shalini Iyer, Louis J. Fabri, Andrew D. Nash and K. Ravi Acharya. 'Crystal Structure of Vascular Endothelial Growth Factor-B in Complex with a Neutralising Antibody Fab Fragment'. In: *Journal of Molecular Biology* 384.5 (2008), pp. 1203–1217. DOI: 10.1016/j.jmb.2008.09.076 (cit. on pp. 17, 94).

[120] Shalini Iyer, Paula I. Darley and K. Ravi Acharya. 'Structural Insights into the Binding of Vascular Endothelial Growth Factor-B by VEGFR-1D2'. In: *Journal of Biological Chemistry* 285.31 (2010), pp. 23779–23789. DOI: 10.1074/jbc.M110.130658 (cit. on pp. 17, 94).

[121] Kalyaneswar Mandal and Stephen B. H. Kent. 'Total Chemical Synthesis of Biologically Active Vascular Endothelial Growth Factor'. In: *Angewandte Chemie International Edition* 50.35 (2011), pp. 8029–8033. DOI: 10.1002/anie.201103237 (cit. on pp. 17, 94).

[122] Marcelo H. Ratto, Yvonne A. Leduc, Ximena P. Valderrama, Karin E. van Straaten, Louis T. J. Delbaere, Roger A. Pierson and Gregg P. Adams. 'The nerve of ovulation-inducing factor in semen'. In: *Proceedings of the National Academy of Sciences* 109.37 (2012), pp. 15042–15047. DOI: 10.1073/pnas.1206273109 (cit. on pp. 17, 94).

[123] Kalyaneswar Mandal, Maruti Uppalapati, Dana Ault-Riché, John Kenney, Joshua Lowitz, Sachdev S. Sidhu and Stephen B. H. Kent. 'Chemical synthesis and X-ray structure of a heterochiral {D-protein antagonist plus vascular endothelial growth factor} protein complex by racemic crystallography'. In: *Proceedings of the National Academy of Sciences* 109.37 (2012), pp. 14779–14784. DOI: 10.1073/pnas.1210483109 (cit. on pp. 17, 94).

[124] Kristof Nolan, Chandramohan Kattamuri, David M. Luedeke, Xiaodi Deng, Amrita Jagpal, Fuming Zhang, Robert J. Linhardt, Alan P. Kenny, Aaron M. Zorn and Thomas B. Thompson. 'Structure of Protein Related to Dan and Cerberus: Insights into the Mechanism of Bone Morphogenetic Protein Antagonism'. In: *Structure* 21.8 (2013), pp. 1417–1429. DOI: 10.1016/j.str.2013.06.005 (cit. on pp. 17, 94).

[125] Byung-Hak Yoon, Luis Esquivies, Chihoon Ahn, Peter C. Gray, Sang-kyu Ye, Witek Kwiatkowski and Senyon Choe. 'An Activin A/BMP2 Chimera, AB204, Displays Bone-Healing Properties Superior to Those of BMP2'. In: *Journal of Bone and Mineral Research* 29.9 (2014), pp. 1950–1959. DOI: 10.1002/jbmr.2238 (cit. on pp. 17, 94).

[126] Aaron Moulin, Magali Mathieu, Catherine Lawrence, Russell Bigelow, Mark Levine, Christine Hamel, Jean-Piere Marquette, Josiane Le Parc, Christophe Loux, Paul Ferrari, Cecile Capdevila, Jacques Dumas, Bruno Dumas, Alexey Rak, Julie Bird, Huawei Qiu, Clark Q. Pan, Tim Edmunds and Ronnie R. Wei. 'Structures of a pan-specific antagonist antibody complexed to different isoforms of TGFβ reveal structural plasticity of antibody-antigen interactions'. In: *Protein Science* 23.12 (2014), pp. 1698–1707. DOI: 10.1002/pro.2548 (cit. on pp. 17, 94).

[127] Kristof Nolan, Chandramohan Kattamuri, Scott A. Rankin, Randy J. Read, Aaron M. Zorn and Thomas B. Thompson. 'Structure of Gremlin-2 in Complex with GDF5 Gives Insight into DAN-Family-Mediated BMP Antagonism'. In: *Cell Reports* 16.8 (2016), pp. 2077–2086. DOI: 10.1016/j.celrep.2016.07.046 (cit. on pp. 17, 94).

[128] Miglė Kišonaitė, Xuelu Wang and Marko Hyvönen. 'Structure of Gremlin-1 and analysis of its interaction with BMP-2'. In: *Biochemical Journal* 473.11 (2016), pp. 1593–1604. DOI: 10.1042/BCJ20160254 (cit. on pp. 17, 94).

[129] Steven B. Larson and Alexander McPherson. 'The crystal structure of the β subunit of luteinizing hormone and a model for the intact hormone'. In: *Current Research in Structural Biology* 1.July (2019), pp. 1–5. DOI: 10.1016/j.crstbi.2019.07.001 (cit. on pp. 17, 94).

[130] Laura del Amo-Maestro, Laura Marino-Puertas, Theodoros Goulas and F. Xavier Gomis-Rüth. 'Recombinant production, purification, crystallization, and structure analysis of human transforming growth factor β2 in a new conformation'. In: *Scientific Reports* 9.1 (2019), p. 8660. DOI: 10.1038/s41598-019-44943-4 (cit. on pp. 17, 94).

[131] C. Robert, F. Bruck, R. Herman, M. Vandevenne, P. Filee, F. Kerff and A. Matagne. 'Functional and structural study of Bone Morphogenetic Protein 2 and its antagonist Noggin'. DOI: http://doi.org/10.2210/pdb7AG0/pdb (cit. on pp. 17, 94).

[132] Hristo L. Svilenov, Julia Sacherl, Ulrike Protzer, Martin Zacharias and Johannes Buchner. 'Mechanistic principles of an ultra-long bovine CDR reveal strategies for antibody design'. In: *Nature Communications* 12.1 (2021), p. 6737. DOI: 10.1038/s41467-021-27103-z (cit. on pp. 17, 19, 108, 140, 162, 206, 232, 291, 294).

[133] Robert Pejchal, Laura M. Walker, Robyn L. Stanfield, Sanjay K. Phogat, Wayne C. Koff, Pascal Poignard, Dennis R. Burton and Ian A. Wilson. 'Structure and function of broadly reactive antibody PG16 reveal an H3 subdomain that mediates potent neutralization of HIV-1'. In: *Proceedings of the National Academy of Sciences* 107.25 (2010), pp. 11483–11488. DOI: 10.1073/pnas.1004600107 (cit. on pp. 18, 293).

[134] Marie Pancera, Syed Shahzad-ul-Hussan, Nicole A. Doria-Rose, Jason S. McLellan, Robert T. Bailer, Kaifan Dai, Sandra Loesgen, Mark K. Louder, Ryan P. Staupe, Yongping Yang, Baoshan Zhang, Robert Parks, Joshua Eudailey, Krissey E. Lloyd, Julie Blinn, S. Munir Alam, Barton F. Haynes, Mohammed N. Amin, Lai-Xi Wang, Dennis R. Burton, Wayne C. Koff, Gary J. Nabel, John R. Mascola, Carole A. Bewley and Peter D. Kwong. 'Structural basis for diverse N-glycan recognition by HIV-1–neutralizing V1–V2–directed antibody PG16'. In: *Nature Structural and Molecular Biology* 20.7 (2013), pp. 804–813. DOI: 10.1038/nsmb.2600 (cit. on pp. 18, 293).

[135] Nicole A. Doria-Rose, Jinal N. Bhiman, Ryan S. Roark, Chaim A. Schramm, Jason Gorman, Gwo-Yu Chuang, Marie Pancera, Evan M. Cale, Michael J. Ernandes, Mark K. Louder, Mangaiarkarasi Asokan, Robert T. Bailer, Aliaksandr Druz, Isabella R. Fraschilla, Nigel J. Garrett, Marissa Jarosinski, Rebecca M. Lynch, Krisha McKee, Sijy O'Dell, Amarendra Pegu, Stephen D. Schmidt, Ryan P. Staupe, Matthew S. Sutton, Keyun Wang, Constantinos Kurt Wibmer, Barton F. Haynes, Salim Abdool-Karim, Lawrence Shapiro, Peter D. Kwong, Penny L. Moore, Lynn Morris and John R. Mascola. 'New Member of the V1V2-Directed CAP256-VRC26 Lineage That Shows Increased Breadth and Exceptional Potency'. In: *Journal of Virology* 90.1 (2016). Ed. by G. Silvestri, pp. 76–91. DOI: 10.1128/JVI.01791-15 (cit. on pp. 18, 293).

[136] Jason Gorman, Gwo-Yu Chuang, Yen-Ting Lai, Chen-Hsiang Shen, Jeffrey C. Boyington, Aliaksandr Druz, Hui Geng, Mark K. Louder, Krisha McKee, Reda Rawi, Raffaello Verardi, Yongping Yang, Baoshan Zhang, Nicole A. Doria-Rose, Bob Lin, Penny L. Moore, Lynn Morris, Lawrence Shapiro, John R. Mascola and Peter D. Kwong. 'Structure of Super-Potent Antibody CAP256-VRC26.25 in Complex with HIV-1 Envelope Reveals a Combined Mode of Trimer-Apex Recognition'. In: *Cell Reports* 31.1 (2020), p. 107488. DOI: 10.1016/j.celrep.2020.03.052 (cit. on pp. 18, 293, 297).

[137] Devin Sok, Khoa M. Le, Melissa L. Vadnais, Karen L. Saye-Francisco, Joseph G. Jardine, Jonathan L. Torres, Zachary T. Berndsen, Leopold Kong, Robyn L. Stanfield, Jennifer Ruiz, Alejandra Ramos, Chi-Hui Liang, Patricia L. Chen, Michael F. Criscitiello, Waithaka Mwangi, Ian A. Wilson, Andrew B. Ward, Vaughn V. Smider and Dennis R. Burton. 'Rapid elicitation of broadly neutralizing antibodies to HIV by immunization in cows'. In: *Nature* 548.1 (2017), pp. 108–111. DOI: 10.1038/nature23301 (cit. on pp. 18, 206, 293, 294).

[138] Alex Macpherson, Anthony Scott-Tucker, Anastasios Spiliotopoulos, Catherine Simpson, Justin Staniforth, Adam Hold, James R. Snowden, Leah Manning, Jean M. H. van den Elsen and Alastair D. G. Lawson. 'Isolation of antigen-specific, disulphide-rich knob domain peptides from bovine antibodies'. In: *PLOS Biology* 18.9 (2020). Ed. by David Nemazee, e3000821. DOI: 10.1371/journal.pbio.3000821 (cit. on pp. 19, 140, 206, 232, 289, 291, 294).

[139] Alex Macpherson, Maisem Laabei, Zainab Ahdash, Melissa A. Graewert, James R. Birtley, Monika-Sarah E. D. Schulze, Susan Crennell, Sarah A. Robinson, Ben Holmes, Vladas Oleinikovas, Per H. Nilsson, James R. Snowden, Victoria Ellis, Tom Eirik Mollnes, Charlotte M. Deane, Dmitri Svergun, Alastair D. G. Lawson and Jean M. H. van den Elsen. 'The allosteric modulation of complement C5 by knob domain peptides'. In: *eLife* 10 (2021), pp. 1–49. DOI: 10.7554/eLife.63586 (cit. on pp. 19, 26, 140, 206, 232, 289, 291, 294).

[140] Yong Zhang, Danling Wang, Gus Welzel, Ying Wang, Peter G. Schultz and Feng Wang. 'An Antibody CDR3-Erythropoietin Fusion Protein'. In: *ACS Chem. Biol* 8.10 (2013), pp. 2117–2121. DOI: 10.1021/cb4004749 (cit. on p. 19).

[141] Yong Zhang, Danling Wang, Lorenzo de Lichtervelde, Sophie B. Sun, Vaughn V. Smider, Peter G. Schultz and Feng Wang. 'Functional Antibody CDR3 Fusion Proteins with Enhanced Pharmacological Properties'. In: *Angewandte Chemie International Edition* 52.32 (2013), pp. 8295–8298. DOI: 10.1002/anie.201303656 (cit. on p. 19).

[142] Tao Liu, Yan Liu, Ying Wang, Mitchell Hull, Peter G. Schultz and Feng Wang. 'Rational Design of CXCR4 Specific Antibodies with Elongated CDRs'. In: *Journal of the American Chemical Society* 136.30 (2014), pp. 10557–10560. DOI: 10.1021/ja5042447 (cit. on p. 19).

[143] Yong Zhang, Devrishi Goswami, Danling Wang, Tsung Shing Andrew Wang, Shiladitya Sen, Thomas J. Magliery, Patrick R. Griffin, Feng Wang and Peter G. Schultz. 'An antibody with a variable-region coiled-coil "knob" domain'. In: *Angewandte Chemie International Edition* 53.1 (2014), pp. 132–135. DOI: 10.1002/anie.201307939 (cit. on p. 19).

[144] Yong Zhang, Huafei Zou, Ying Wang, Dawna Caballero, Jose Gonzalez, Elizabeth Chao, Gus Welzel, Weijun Shen, Danling Wang, Peter G. Schultz and Feng Wang. 'Rational Design of a Humanized Glucagon-Like Peptide-1 Receptor Agonist Antibody'. In: *Angewandte Chemie International Edition* 54.7 (2015), pp. 2126–2130. DOI: 10.1002/anie.201410049 (cit. on p. 19).

[145] Yong Zhang, Yan Liu, Ying Wang, Peter G. Schultz and Feng Wang. 'Rational Design of Humanized Dual-Agonist Antibodies'. In: *Journal of the American Chemical Society* 137.1 (2015), pp. 38–41. DOI: 10.1021/ja510519u (cit. on p. 19).

[146] Yfke Pasman, Caroline Soliman, Paul A. Ramsland and Azad K. Kaushik. 'Exceptionally long CDR3H of bovine scFv antigenized with BoHV-1 B-epitope generates specific immune response against the targeted epitope'. In: *Molecular Immunology* 77.1 (2016), pp. 113–125. DOI: 10.1016/j.molimm.2016.07.014 (cit. on p. 19).

[147] Rongsheng E. Wang, Ying Wang, Yuhan Zhang, Chase Gabrelow, Yong Zhang, Victor Chi, Qiangwei Fu, Xiaozhou Luo, Danling Wang, Sean Joseph, Kristen Johnson, Arnab K. Chatterjee, Timothy M. Wright, Vân T. B. Nguyen-Tran, John Teijaro, Argyrios N. Theofilopoulos, Peter G. Schultz and Feng Wang. 'Rational design of a Kv1.3 channel-blocking antibody as a selective immunosuppressant'. In: *Proceedings of the National Academy of Sciences of the USA* 113.41 (2016), pp. 11501–11506. DOI: 10.1073/pnas.1612803113 (cit. on p. 19).

[148] Alice Hawkins, Callum Joyce, Kevin Brady, Adam Hold, Alan Smith, Michael Knight, Conor Howard, Jean M. H. van den Elsen, Alastair D. G. Lawson and Alex Macpherson. 'The proximity of the N- and C- termini of bovine knob domains enable engineering of target specificity into polypeptide chains'. In: *mAbs* 14.1 (2022), pp. 1–12. DOI: 10.1080/19420862.2022.2076295 (cit. on p. 19).

[149] Ulrike S. Diesterbeck, Dalia M. Aboelhassan, Sina K. Stein and Claus-Peter Czerny. 'Detection of new allotypic variants of bovine λ-light chain constant regions in different cattle breeds'. In: *Developmental and Comparative Immunology* 36.1 (2012), pp. 130–139. DOI: 10.1016/j.dci.2011.06.011 (cit. on p. 20).

[150] T. C. Tobin-Janzen and J. E. Womack. 'The immunoglobulin lambda light chain constant region maps to bos taurus chromosome 17'. In: *Animal Biotechnology* 7.2 (1996), pp. 163–172. DOI: 10.1080/10495399609525857 (cit. on p. 20).

[151] Limei Chen, Min Li, Qing Li, Xingyuan Yang, Xiaorong An and Yongfu Chen. 'Characterization of the bovine immunoglobulin lambda light chain constant IGLC genes'. In: *Veterinary Immunology and Immunopathology* 124.3-4 (2008), pp. 284–294. DOI: 10.1016/j.vetimm.2008.04.012 (cit. on p. 20).

[152] Anna Ekman, Mikael Niku, Jenni Liljavirta and Antti Iivanainen. 'Bos taurus genome sequence reveals the assortment of immunoglobulin and surrogate light chain genes in domestic cattle'. In: *BMC Immunology* 10.22 (2009). DOI: 10.1186/1471-2172-10-22 (cit. on pp. 20, 289).

[153] M. C. Sinclair, J. Gilchrist and Robert Aitken. 'Molecular Characterization of Bovine Vλ Regions'. In: *The Journal of Immunology* 155.4 (1995), pp. 3068–3078 (cit. on pp. 20–23, 27, 289).

[154] Sina K. Stein, Ulrike S. Diesterbeck, Dalia M. Aboelhassan and Claus-Peter Czerny. 'Comparison of joining and constant κ-light chain regions in different cattle breeds'. In: *Animal Genetics* 43.6 (2012), pp. 776–780. DOI: 10.1111/j.1365-2052.2012.02324.x (cit. on p. 21).

[155] S. S. Arun, W. Breuer and W. Hermanns. 'Immunohistochemical Examination of Light-chain Expression (Λ/k Ratio) in Canine, Feline, Equine, Bovine and Porcine Plasma Cells'. In: *Journal of Veterinary Medicine Series A* 43.1-10 (1996), pp. 573–576. DOI: 10.1111/j.1439-0442.1996.tb00489.x (cit. on p. 21).

[156] Robert Aitken, A. Hosseini and R. MacDuff. 'Structure and diversification of the bovine immunoglobulin repertoire'. In: *Veterinary Immunology and Immunopathology* 72.1-2 (1999), pp. 21–29. DOI: 10.1016/S0165-2427(99)00113-0 (cit. on p. 21).

[157] Hans-Peter Brezinschek, Sandra J. Foster, Thomas Dörner, Ruth I. Brezinschek and Peter E. Lipsky. 'Pairing of Variable Heavy and Variable κ Chains in Individual Naive and Memory B Cells'. In: *Journal of Immunology* 160.10 (1998), pp. 4762–4767 (cit. on p. 22).

[158] Ruud M. T. de Wildt, Rene M. A. Hoet, Walther J. van Venrooij, Ian M. Tomlinson and Greg Winter. 'Analysis of Heavy and Light Chain Pairings Indicates that Receptor Editing Shapes the Human Antibody Repertoire'. In: *Journal of Molecular Biology* 285.3 (1999), pp. 895–901. DOI: 10.1006/jmbi.1998.2396 (cit. on p. 22).

[159] Narayan Jayaram, Pallab Bhowmick and Andrew C. R. Martin. 'Germline VH/VL pairing in antibodies'. In: *Protein Engineering, Design and Selection* 25.10 (2012), pp. 523–529. DOI: 10.1093/protein/gzs043 (cit. on p. 22).

[160] Jingshan Ren, Joanne E. Nettleship, Gemma Harris, William N. Mwangi, Nahid Rhaman, Clare F. J. Grant, Abhay Kotecha, Elizabeth E. Fry, Bryan Charleston, David I. Stuart, John A. Hammond and Raymond J. Owens. 'The role of the light chain in the structure and binding activity of two cattle antibodies that neutralize bovine respiratory syncytial virus'. In: *Molecular Immunology* 112 (2019), pp. 123–130. DOI: 10.1016/j.molimm.2019.04.026 (cit. on pp. 25, 26, 33, 92–94, 99, 105, 119, 128, 146, 164, 237, 240–243, 253, 255–258, 260, 282, 283, 285–287, 290, 295).

[161] H. E. Kennedy, B. V. Jones, E. M. Tucker, N. J. Ford, S. W. Clarke, Julie Furze, Lewis H. Thomas and E. J. Stott. 'Production and Characterization of Bovine Monoclonal Antibodies to Respiratory Syncytial Virus'. In: *Journal of General Virology* 69.12 (1988), pp. 3023–3032. DOI: 10.1099/0022-1317-69-12-3023 (cit. on pp. 25, 93, 240).

[162] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov and Philip E. Bourne. 'The Protein Data Bank'. In: *Nucleic Acids Research* 28.1 (2000), pp. 235–242. DOI: 10.1093/nar/28.1.235 (cit. on pp. 26, 64, 72, 73, 93, 94, 105, 137, 171–173, 176, 229, 292).

[163] Isabel K. MacDonald, Maria Harkiolaki, Lawrence Hunt, Timothy Connelley, A. Victoria Carroll, Niall D. MacHugh, Simon P. Graham, E. Yvonne Jones, W. Ivan Morrison, Darren R. Flower and Shirley A. Ellis. 'MHC class I bound to an immunodominant Theileria parva epitope demonstrates unconventional presentation to T cell receptors'. In: *PLoS Pathogens* 6.10 (2010). DOI: 10.1371/journal.ppat.1001149 (cit. on p. 26).

[164] Xin Li, Jun Liu, Jianxun Qi, Feng Gao, Qirun Li, Xiaoying Li, Nianzhi Zhang, Chun Xia and George F. Gao. 'Two Distinct Conformations of a Rinderpest Virus Epitope Presented by Bovine Major Histocompatibility Complex Class I N*01801: a Host Strategy To Present Featured Peptides'. In: *Journal of Virology* 85.12 (2011), pp. 6038–6048. DOI: 10.1128/JVI.00030-11 (cit. on p. 26).

[165] Jing Wang, Joren Guillaume, Nora Pauwels, Serge Van Calenbergh, Ildiko Van Rhijn and Dirk M. Zajonc. 'Crystal Structures of Bovine CD1d Reveal Altered αGalCer Presentation and a Restricted A' Pocket Unable to Bind Long-Chain Glycolipids'. In: *PLoS ONE* 7.10 (2012). Ed. by Scheherazade Sadegh-Nasseri, e47989. DOI: 10.1371/journal.pone.0047989 (cit. on p. 26).

[166] Enrico Girardi, Jing Wang, Thien-Thi Mac, Cees Versluis, Veemal Bhowruth, Gurdyal Besra, Albert J. R. Heck, Ildiko Van Rhijn and Dirk M. Zajonc. 'Crystal Structure of Bovine CD1b3 with Endogenously Bound Ligands'. In: *The Journal of Immunology* 185.1 (2010), pp. 376–386. DOI: 10.4049/jimmunol.1000042 (cit. on p. 26).

[167] Yanjie Liu, Xin Li, Jianxun Qi, Nianzhi Zhang and Chun Xia. 'The structural basis of chicken, swine and bovine CD8$\alpha\alpha$ dimers provides insight into the co-evolution with MHC I in endotherm species'. In: *Scientific Reports* 6.1 (2016), p. 24788. DOI: 10.1038/srep24788 (cit. on p. 26).

[168] P. R. Kumar, A. Casadevall and Almo S. C. 'Crystal structure of bovine CTLA-4, PSI-NYSGRC-012704'. DOI: http://doi.org/10.2210/pdb4KKN/pdb (cit. on p. 26).

References for Chapter 2

[160] Jingshan Ren, Joanne E. Nettleship, Gemma Harris, William N. Mwangi, Nahid Rhaman, Clare F. J. Grant, Abhay Kotecha, Elizabeth E. Fry, Bryan Charleston, David I. Stuart, John A. Hammond and Raymond J. Owens. 'The role of the light chain in the structure and binding activity of two cattle antibodies that neutralize bovine respiratory syncytial virus'. In: *Molecular Immunology* 112 (2019), pp. 123–130. DOI: 10.1016/j.molimm.2019.04.026 (cit. on pp. 25, 26, 33, 92–94, 99, 105, 119, 128, 146, 164, 237, 240–243, 253, 255–258, 260, 282, 283, 285–287, 290, 295).

[162] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov and Philip E. Bourne. 'The Protein Data Bank'. In: *Nucleic Acids Research* 28.1 (2000), pp. 235–242. DOI: 10.1093/nar/28.1.235 (cit. on pp. 26, 64, 72, 73, 93, 94, 105, 137, 171–173, 176, 229, 292).

[169] F. L. Graham, J. Smiley, W. C. Russell and R. Nairn. 'Characteristics of a human cell line transformed by DNA from human adenovirus type 5'. In: *Journal of General Virology* 36.1 (1977), pp. 59–72. DOI: 10.1099/0022-1317-36-1-59 (cit. on p. 32).

[170] Nick S. Berrow, David Alderton, Sarah Sainsbury, Joanne E. Nettleship, Rene Assenberg, Nahid Rahman, David I. Stuart and Raymond J. Owens. 'A versatile ligation-independent cloning method suitable for high-throughput expression screening applications'. In: *Nucleic Acids Research* 35.6 (2007), e45–e45. DOI: 10.1093/nar/gkm047 (cit. on p. 33).

[171] Nick S. Berrow, David Alderton and Raymond J. Owens. *The Precise Engineering of Expression Vectors Using High-Throughput In-Fusion PCR Cloning*. Ed. by Sharon A. Doyle. 1st ed. Walnut Creek, CA, USA: Humana Press, 2009. Chap. 5, pp. 75–90. DOI: 10.1007/978-1-59745-196-3 (cit. on p. 33).

[172] Francisco Bolivar, Raymond L. Rodriguez, Mary C. Betlach and Herbert W. Boyer. 'Construction and characterization of new cloning vehicles I. Ampicillin-resistant derivatives of the plasmid pMB9'. In: *Gene* 2.2 (1977), pp. 75–93. DOI: 10.1016/0378-1119(77)90074-9 (cit. on p. 34).

[173] Francisco Bolivar, Raymond L. Rodriguez, Patricia J. Greene, Mary C. Betlach, Herbert L. Heyneker, Herbert W. Boyer, Jorge H. Crosa and Stanley Falkow. 'Construction and characterization of new cloning vehicle. II. A multipurpose cloning system'. In: *Gene* 2.2 (1977), pp. 95–113. DOI: 10.1016/0378-1119(77)90000-2 (cit. on p. 34).

[174] Joanne E. Nettleship, Nahid Rahman-Huq and Raymond J. Owens. *The Production of Glycoproteins by Transient Expression in Mammalian Cells*. Ed. by Sharon A. Doyle. 498th ed. Totowa: Human Press, 2009. Chap. 16, pp. 245–263. DOI: 10.1007/978-1-59745-196-3_16 (cit. on pp. 34, 48).

[175] Baoshan Zhang, Lei Chen, Chiara Silacci, Michelle Thom, Jeffrey C. Boyington, Aliaksandr Druz, M. Gordon Joyce, Efrain Guzman, Wing-Pui Kong, Yen-Ting Lai, Guillaume B. E. Stewart-Jones, Yaroslav Tsybovsky, Yongping Yang, Tongqing Zhou, Ulrich Baxa, John R. Mascola, Davide Corti, Antonio Lanzavecchia, Geraldine Taylor and Peter D. Kwong. 'Protection of calves by a prefusion-stabilized bovine RSV F vaccine'. In: *npj Vaccines* 2.7 (2017), pp. 1–11. DOI: 10.1038/s41541-017-0005-9 (cit. on pp. 34, 239, 240, 245, 251, 282, 283, 285, 286, 295).

[176] Claudine Porta, Abhay Kotecha, Alison Burman, Terry Jackson, Jingshan Ren, Silvia Loureiro, Ian M. Jones, Elizabeth E. Fry, David I. Stuart and Bryan Charleston. 'Rational Engineering of Recombinant Picornavirus Capsids to Produce Safe, Protective Vaccine Antigen'. In: *PLoS Pathogens* 9.3 (2013). Ed. by Félix A. Rey, e1003255. DOI: 10.1371/journal.ppat.1003255 (cit. on pp. 36, 37, 85, 152).

[177] Claudine Porta, Xiaodong Xu, Silvia Loureiro, Saravanan Paramasivam, Junyuan Ren, Tara Al-Khalil, Alison Burman, Terry Jackson, Graham J. Belsham, Stephen Curry, George P. Lomonossoff, Satya Parida, David J. Paton, Yanmin Li, Ginette Wilsden, Nigel Ferris, Raymond J. Owens, Abhay Kotecha, Elizabeth E. Fry, David I. Stuart, Bryan Charleston and Ian M. Jones. 'Efficient production of foot-and-mouth disease virus empty capsids in insect cells following down regulation of 3C protease activity'. In: *Journal of Virological Methods* 187.2 (2013), pp. 406–412. DOI: 10.1016/j.jviromet.2012.11.011 (cit. on pp. 36, 56, 152).

[178] New England Biolabs. *NEBCloner Restriction Enzyme Double Digestion Tool*. URL: https://nebcloner.neb.com/#!/redigest (visited on 5th Aug. 2019) (cit. on p. 39).

[179] Clare F. J. Grant, B. Veronica Carr, Abhay Kotecha, Erwin van den Born, David I. Stuart, John A. Hammond and Bryan Charleston. 'The B Cell Response to Foot-and-Mouth Disease Virus in Cattle following Sequential Vaccination with Multiple Serotypes'. In: *Journal of Virology* 91.9 (2017). Ed. by Douglas S. Lyles. DOI: 10.1128/JVI.02157-16 (cit. on p. 40).

[180] Swiss Institute of Bioinformatics. *ExPASy ProtParam*. URL: https://web.expasy.org/protparam/ (visited on 15th Nov. 2022) (cit. on pp. 55, 249, 273, 305, 306).

[181] Stewart Goodwin, Tobias J. Tuthill, Armando Arias, Richard A. Killington and David J. Rowlands. 'Foot-and-Mouth Disease Virus Assembly: Processing of Recombinant Capsid Precursor by Exogenous Protease Induces Self-Assembly of Pentamers In Vitro in a Myristoylation-Dependent Manner'. In: *Journal of Virology* 83.21 (2009), pp. 11275–11282. DOI: 10.1128/JVI.01263-09 (cit. on pp. 56, 148).

[182] Ravindra Acharya, Elizabeth E. Fry, David I. Stuart, Graham Fox, David J. Rowlands and Fred Brown. 'The three-dimensional structure of foot-and-mouth disease virus at 2.9 Å resolution'. In: *Nature* 337.6209 (1989), pp. 709–716. DOI: 10.1038/337709a0 (cit. on pp. 56, 85, 148, 149, 151, 188, 202).

[183] Abhay Kotecha, Julian Seago, Katherine A. Scott, Alison Burman, Silvia Loureiro, Jingshan Ren, Claudine Porta, Helen M. Ginn, Terry Jackson, Eva Perez-Martin, C. Alistair Siebert, Guntram Paul, Juha T. Huiskonen, Ian M. Jones, Robert M. Esnouf, Elizabeth E. Fry, Francois F. Maree, Bryan Charleston and David I. Stuart. 'Structure-based energetics of protein interfaces guides foot-and-mouth disease virus vaccine design'. In: *Nature Structural and Molecular Biology* 22.10 (2015), pp. 788–794. DOI: 10.1038/nsmb.3096 (cit. on pp. 56, 85, 193, 194, 210, 211, 216, 218, 221, 222).

[184] William N. Mwangi, Giuseppe Maccari, Jayne C. Hope, Gary Entrican and John A. Hammond. 'The UK Veterinary Immunological Toolbox Website: promoting vaccine research by facilitating communication and removing reagent barriers'. In: *Immunology* 161.1 (2020), pp. 25–27. DOI: 10.1111/imm.13227 (cit. on pp. 57, 58, 223).

[185] R. Barer and S. Joseph. 'Refractometry of Living Cells'. In: *Journal of Cell Science* s3-95.32 (1954), pp. 399–423. DOI: 10.1242/jcs.s3-95.32.399 (cit. on p. 60).

[186] Huaying Zhao, Patrick H. Brown and Peter Schuck. 'On the Distribution of Protein Refractive Index Increments'. In: *Biophysical Journal* 100.9 (2011), pp. 2309–2317. DOI: 10.1016/j.bpj.2011.03.004 (cit. on p. 60).

[187] Ramneek Gupta and Søren Brunak. 'Prediction of glycosylation across the human proteome and the correlation to protein function'. In: *Pacific Symposium on Biocomputing* 7 (2002), pp. 310–22. DOI: 10.1142/9789812799623_0029 (cit. on pp. 60, 251, 252, 305).

[188] Catharina Steentoft, Sergey Y. Vakhrushev, Hiren J. Joshi, Yun Kong, Malene B. Vester-Christensen, Katrine T-B. G. Schjoldager, Kirstine Lavrsen, Sally Dabelsteen, Nis B. Pedersen, Lara Marcos-Silva, Ramneek Gupta, Eric Paul Bennett, Ulla Mandel, Søren Brunak, Hans H. Wandall, Steven B. Levery and Henrik Clausen. 'Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology'. In: *The EMBO Journal* 32.10 (2013), pp. 1478–1488. DOI: 10.1038/emboj.2013.79 (cit. on pp. 60, 251, 305).

[189] Technical University of Denmark Department of Health Technology. *NetNGlyc*. URL: https://services.healthtech.dtu.dk/service.php?NetNGlyc-1.0 (visited on 29th Oct. 2022) (cit. on pp. 61, 251, 252, 305).

[190] Technical University of Denmark Department of Health Technology. *NetOGlyc*. URL: https://services.healthtech.dtu.dk/service.php?NetOGlyc-4.0 (visited on 29th Oct. 2022) (cit. on pp. 61, 251, 305).

[191] Peter Schuck. 'Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling'. In: *Biophysical Journal* 78.3 (2000), pp. 1606–1619. DOI: 10.1016/S0006-3495(00)76713-0 (cit. on pp. 62, 255, 257).

[192] Sumit K. Chaturvedi, Jia Ma, Patrick H. Brown, Huaying Zhao and Peter Schuck. 'Measuring macromolecular size distributions and interactions at high concentrations by sedimentation velocity'. In: *Nature Communications* 9.1 (2018), p. 4415. DOI: 10.1038/s41467-018-06902-x (cit. on pp. 62, 255, 257).

[193] Abraham Savitzky and Marcel J. E. Golay. 'Smoothing and Differentiation of Data by Simplified Least Squares Procedures.' In: *Analytical Chemistry* 36.8 (1964), pp. 1627–1639. DOI: 10.1021/ac60214a047 (cit. on p. 63).

[194] James Dunbar, Konrad Krawczyk, Jinwoo Leem, Claire Marks, Jaroslaw Nowak, Cristian Regep, Guy Georges, Sebastian Kelm, Bojana Popovic and Charlotte M. Deane. 'SAbPred: a structure-based antibody prediction server'. In: *Nucleic Acids Research* 44.W1 (2016), W474–W478. DOI: 10.1093/nar/gkw361 (cit. on pp. 64, 93–95).

[195] Juan C. Almagro, Alexey Teplyakov, Jinquan Luo, Raymond W. Sweet, Sreekumar Kodangattil, Francisco Hernandez-Guzman and Gary L. Gilliland. 'Second antibody modeling assessment (AMA-II)'. In: *Proteins: Structure, Function, and Bioinformatics* 82.8 (2014), pp. 1553–1562. DOI: 10.1002/prot.24567 (cit. on p. 64).

[196] James Dunbar and Charlotte M. Deane. 'ANARCI: Antigen receptor numbering and receptor classification'. In: *Bioinformatics* 32.2 (2016), pp. 298–300. DOI: 10.1093/bioinformatics/btv552 (cit. on pp. 64, 93).

[197] James Dunbar, Konrad Krawczyk, Jinwoo Leem, Terry Baker, Angelika Fuchs, Guy Georges, Jiye Shi and Charlotte M. Deane. 'SAbDab: the structural antibody database'. In: *Nucleic Acids Research* 42.D1 (2014), pp. D1140–D1146. DOI: 10.1093/nar/gkt1043 (cit. on pp. 64, 93, 94).

[198] James Dunbar, Angelika Fuchs, Jiye Shi and Charlotte M. Deane. 'ABangle: characterising the VH-VL orientation in antibodies'. In: *Protein Engineering Design and Selection* 26.10 (2013), pp. 611–620. DOI: 10.1093/protein/gzt020 (cit. on p. 64).

[199] Yoonjoo Choi and Charlotte M. Deane. 'Predicting antibody complementarity determining region structures without classification'. In: *Molecular BioSystems* 7.12 (2011), p. 3327. DOI: 10.1039/c1mb05223c (cit. on p. 64).

[200] Andrej Šali and Tom L. Blundell. 'Comparative Protein Modelling by Satisfaction of Spatial Restraints'. In: *Journal of Molecular Biology* 234.3 (1993), pp. 779–815. DOI: 10.1006/jmbi.1993.1626 (cit. on pp. 64, 93).

[201] Georgii G. Krivov, Maxim V. Shapovalov and Roland L. Dunbrack. 'Improved prediction of protein side-chain conformations with SCWRL4'. In: *Proteins: Structure, Function, and Bioinformatics* 77.4 (2009), pp. 778–795. DOI: 10.1002/prot.22488. arXiv: NIHMS150003 (cit. on p. 64).

[202] Marie-Paule Lefranc, Christelle Pommié, Manuel Ruiz, Véronique Giudicelli, Elodie Foulquier, Lisa Truong, Valérie Thouvenin-Contet and Gérard Lefranc. 'IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains'. In: *Developmental and Comparative Immunology* 27.1 (2003), pp. 55–77. DOI: 10.1016/S0145-305X(02)00039-3 (cit. on pp. 64, 307).

[203] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli and Demis Hassabis. 'Highly accurate protein structure prediction with AlphaFold'. In: *Nature* May (2021), pp. 1–12. DOI: 10.1038/s41586-021-03819-2 (cit. on pp. 64, 93, 95).

[204] Milot Mirdita, Sergey Ovchinnikov and Martin Steinegger. 'ColabFold - Making protein folding accessible to all C Result visualization'. In: *bioRxiv* (2021), pp. 4–9. DOI: https://doi.org/10.1101/2021.08.15.456425 (cit. on pp. 64, 93, 95, 138, 169, 171, 174, 175, 177, 178).

[205] Rebecca Page, Slawomir K. Grzechnik, Jaume M. Canaves, Glen Spraggon, Andreas Kreusch, Peter Kuhn, Raymond C. Stevens and Scott A. Lesley. 'Shotgun crystallization strategy for structural genomics: an optimized two-tiered crystallization screen against the Thermotoga maritima proteome'. In: *Acta Crystallographica Section D Biological Crystallography* 59.6 (2003), pp. 1028–1037. DOI: 10.1107/S0907444903007790 (cit. on p. 66).

[206] Fabrice Gorrec. 'The MORPHEUS protein crystallization screen'. In: *Journal of Applied Crystallography* 42.6 (2009), pp. 1035–1042. DOI: 10.1107/S0021889809042022 (cit. on p. 66).

[207] Thomas S. Walter, Jonathan M. Diprose, Chris J. Mayo, Christian Siebold, Mike G. Pickford, Lester Carter, Geoff C. Sutton, Nick S. Berrow, James Brown, Ian M. Berry, Guillaume B. E. Stewart-Jones, Jonathan M. Grimes, David K. Stammers, Robert M. Esnouf, E. Yvonne Jones, Raymond J. Owens, David I. Stuart and Karl Harlos. 'A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization'. In: *Acta Crystallographica Section D Biological Crystallography* 61.6 (2005), pp. 651–657. DOI: 10.1107/S0907444905007808 (cit. on p. 66).

[208] Juan Sanchez-Weatherby, James Sandy, Halina Mikolajek, Carina M. C. Lobley, Marco Mazzorana, Jon Kelly, Geoff Preece, Rich Littlewood and Thomas L. M. Sørensen. 'VMXi: a fully automated, fully remote, high-flux in situ macromolecular crystallography beamline'. In: *Journal of Synchrotron Radiation* 26.1 (2019), pp. 291–301. DOI: 10.1107/S1600577518015114 (cit. on pp. 67, 102).

[209] S. J. Fisher, K. E. Levik, M. A. Williams, A. W. Ashton and Katherine E. McAuley. 'SynchWeb : a modern interface for ISPyB'. In: *Journal of Applied Crystallography* 48.3 (2015), pp. 927–932. DOI: 10.1107/S1600576715004847 (cit. on pp. 69, 104).

[210] Solange Delagenière, Patrice Brenchereau, Ludovic Launer, Alun W. Ashton, Ricardo Leal, Stéphanie Veyrier, José Gabadinho, Elspeth J. Gordon, Samuel D. Jones, Karl Erik Levik, Seán M. McSweeney, Stéphanie Monaco, Max Nanao, Darren Spruce, Olof Svensson, Martin A. Walsh and Gordon A. Leonard. 'ISPyB: an information management system for synchrotron macromolecular crystallography'. In: *Bioinformatics* 27.22 (2011), pp. 3186–3192. DOI: 10.1093/bioinformatics/btr535 (cit. on p. 69).

[211] Ulrich Wolfgang Arndt and A. J. Wonacott. *The Rotation Method in X-ray crystallography*. Ed. by Ulrich Wolfgang Arndt and A. J. Wonacott. 1st ed. Amsterdam, North-Holland: Amsterdam:New York:North-Holland Publishing Company, 1977 (cit. on pp. 71, 102).

[212] Zepu Zhang, Nicholas K. Sauter, Henry van den Bedem, Gyorgy Snell and Ashley M. Deacon. 'Automated diffraction image analysis and spot searching for high-throughput crystal screening'. In: *Journal of Applied Crystallography* 39.1 (2006), pp. 112–119. DOI: 10.1107/S0021889805040677 (cit. on pp. 71, 106, 111, 112, 121, 129, 184).

[213] Graeme Winter and Katherine E. McAuley. 'Automated data collection for macromolecular crystallography'. In: *Methods* 55.1 (2011), pp. 81–93. DOI: 10.1016/j.ymeth.2011.06.010 (cit. on p. 71).

[214] Graeme Winter. 'xia2 : an expert system for macromolecular crystallography data reduction'. In: *Journal of Applied Crystallography* 43.1 (2010), pp. 186–190. DOI: 10.1107/S0021889809045701 (cit. on pp. 71, 72).

[215] Clemens Vonrhein, Claus Flensburg, Peter Keller, Andrew Sharff, Oliver Smart, Wlodek Paciorek, Thomas Womack and Gérard Bricogne. 'Data processing and analysis with the autoPROC toolbox'. In: *Acta Crystallographica Section D Biological Crystallography* 67.4 (2011), pp. 293–302. DOI: 10.1107/S0907444911007773 (cit. on pp. 71, 72).

[216] Wolfgang Kabsch. 'XDS'. In: *Acta Crystallographica Section D Biological Crystallography* 66.2 (2010), pp. 125–132. DOI: 10.1107/S0907444909047337 (cit. on pp. 71, 72).

[217] Philip R. Evans. 'Scaling and assessment of data quality'. In: *Acta Crystallographica Section D Biological Crystallography* 62.1 (2006), pp. 72–82. DOI: 10.1107/S0907444905036693 (cit. on p. 72).

[218] Philip R. Evans and Garib N. Murshudov. 'How good are my data and what is the resolution?' In: *Acta Crystallographica Section D: Biological Crystallography* 69.7 (2013), pp. 1204–1214. DOI: 10.1107/S0907444913000061 (cit. on pp. 72, 114).

[219] Ian J. Tickle, Claus Flensburg, Peter Keller, Wlodek Paciorek, Andrew Sharff, Clemens Vonrhein and Gérard Bricogne. *STARANISO*. 2018. URL: http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi (cit. on p. 72).

[220] Graeme Winter, David G. Waterman, James M. Parkhurst, Aaron S. Brewster, Richard J. Gildea, Markus Gerstel, Luis Fuentes-Montero, Melanie Vollmar, Tara Michels-Clark, Iris D. Young, Nicholas K. Sauter and Gwyndaf Evans. 'DIALS : implementation and evaluation of a new integration package'. In: *Acta Crystallographica Section D Structural Biology* 74.2 (2018), pp. 85–97. DOI: 10.1107/S2059798317017235 (cit. on pp. 72, 111, 114).

[221] Martyn D. Winn, Charles C. Ballard, Kevin D. Cowtan, Eleanor J. Dodson, Paul Emsley, Philip R. Evans, Ronan M. Keegan, Eugene B. Krissinel, Andrew G. W. Leslie, Airlie J. McCoy, Stuart J. McNicholas, Garib N. Murshudov, Navraj S. Pannu, Elizabeth A. Potterton, Harold R. Powell, Randy J. Read, Alexei A. Vagin and Keith S. Wilson. 'Overview of the CCP4 suite and current developments'. In: *Acta Crystallographica Section D Biological Crystallography* 67.4 (2011), pp. 235–242. DOI: 10.1107/S0907444910045749 (cit. on pp. 72, 74, 105, 109, 115, 119, 128, 134, 182).

[222] Brian W. Matthews. 'Comparison of the predicted and observed secondary structure of T4 phage lysozyme'. In: *Biochimica et Biophysica Acta (BBA) - Protein Structure* 405.2 (1975), pp. 442–451. DOI: 10.1016/0005-2795(75)90109-9 (cit. on p. 72).

[223] Airlie J. McCoy, Ralf W. Grosse-Kunstleve, Paul D. Adams, Martyn D. Winn, Laurent C. Storoni and Randy J. Read. 'Phaser crystallographic software'. In: *Journal of Applied Crystallography* 40.4 (2007), pp. 658–674. DOI: 10.1107/S0021889807021206 (cit. on pp. 72, 105, 114, 119, 128, 182).

[224] Gerrit Langer, Serge X. Cohen, Victor S. Lamzin and Anastassis Perrakis. 'Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7'. In: *Nature Protocols* 3.7 (2008), pp. 1171–1179. DOI: 10.1038/nprot.2008.91 (cit. on pp. 73, 114).

[225] Grzegorz Chojnowski, Joana Pereira and Victor S. Lamzin. 'Sequence assignment for low-resolution modelling of protein crystal structures'. In: *Acta Crystallographica Section D Structural Biology* 75.8 (2019), pp. 753–763. DOI: 10.1107/S2059798319009392 (cit. on p. 73).

[226] Grzegorz Chojnowski, Koushik Choudhury, Philipp Heuser, Egor Sobolev, Joana Pereira, Umut Oezugurel and Victor S. Lamzin. 'The use of local structural similarity of distant homologues for crystallographic model building from a molecular-replacement solution'. In: *Acta Crystallographica Section D Structural Biology* 76.3 (2020), pp. 248–260. DOI: 10.1107/S2059798320000455 (cit. on p. 73).

[227] Garib N. Murshudov, Pavol Skubák, Andrey A. Lebedev, Navraj S. Pannu, Roberto A. Steiner, Robert A. Nicholls, Martyn D. Winn, Fei Long and Alexei A. Vagin. 'REFMAC 5 for the refinement of macromolecular crystal structures'. In: *Acta Crystallographica Section D Biological Crystallography* 67.4 (2011), pp. 355–367. DOI: 10.1107/S0907444911001314 (cit. on pp. 73, 105, 114, 119, 128, 182).

[228] Paul Emsley and Kevin D. Cowtan. 'Coot : model-building tools for molecular graphics'. In: *Acta Crystallographica Section D Biological Crystallography* 60.12 (2004), pp. 2126–2132. DOI: 10.1107/S0907444904019158 (cit. on pp. 73, 88, 114, 218).

[229] Randy J. Read, Paul D. Adams, W. Bryan Arendall, Axel T. Brunger, Paul Emsley, Robbie P. Joosten, Gerard J. Kleywegt, Eugene B. Krissinel, Thomas Lütteke, Zbyszek Otwinowski, Anastassis Perrakis, Jane S. Richardson, William H. Sheffler, Janet L. Smith, Ian J. Tickle, Gert Vriend and Peter H. Zwart. 'A New Generation of Crystallographic Validation Tools for the Protein Data Bank'. In: *Structure* 19.10 (2011), pp. 1395–1412. DOI: 10.1016/j.str.2011.08.006 (cit. on p. 73).

[230] P. Andrew Karplus and Kay Diederichs. 'Linking Crystallographic Model and Data Quality'. In: *Science* 336.6084 (2012), pp. 1030–1033. DOI: 10.1126/science.1218231 (cit. on p. 74).

[231] Jimin Wang. 'Estimation of the quality of refined protein crystal structures'. In: *Protein Science* 24.5 (2015), pp. 661–669. DOI: 10.1002/pro.2639 (cit. on p. 74).

[232] Christopher J. Williams, Jeffrey J. Headd, Nigel W. Moriarty, Michael G. Prisant, Lizbeth L. Videau, Lindsay N. Deis, Vishal Verma, Daniel A. Keedy, Bradley J. Hintze, Vincent B. Chen, Swati Jain, Steven M. Lewis, W. Bryan Arendall, Jack Snoeyink, Paul D. Adams, Simon C. Lovell, Jane S. Richardson and David C. Richardson. 'MolProbity: More and better reference data for improved all-atom structure validation'. In: *Protein Science* 27.1 (2018), pp. 293–315. DOI: 10.1002/pro.3330 (cit. on pp. 74, 88, 105).

[233] Evgeny Krissinel and Kim Henrick. 'Inference of Macromolecular Assemblies from Crystalline State'. In: *Journal of Molecular Biology* 372.3 (2007), pp. 774–797. DOI: 10.1016/j.jmb.2007.05.022 (cit. on pp. 74, 109, 111, 134, 195).

[234] EMBL-EBI. *PDBePISA (Proteins, Interfaces, Structures and Assemblies)*. URL: https://www.ebi.ac.uk/pdbe/pisa/ (cit. on p. 74).

[235] Takanori Nakane, Abhay Kotecha, Andrija Sente, Greg McMullan, Simonas Masiulis, Patricia M. G. E. Brown, Ioana T. Grigoras, Lina Malinauskaite, Tomas Malinauskas, Jonas Miehling, Tomasz Uchański, Lingbo Yu, Dimple Karia, Evgeniya V. Pechnikova, Erwin de Jong, Jeroen Keizer, Maarten Bischoff, Jamie McCormack, Peter Tiemeijer, Steven W. Hardwick, Dimitri Y. Chirgadze, Garib Murshudov, A. Radu Aricescu and Sjors H. W. Scheres. 'Single-particle cryo-EM at atomic resolution'. In: *Nature* 587.7832 (2020), pp. 152–156. DOI: 10.1038/s41586-020-2829-0 (cit. on p. 75).

[236] Ueli Aebi and Thomas D. Pollard. 'A glow discharge unit to render electron microscope grids and other surfaces hydrophilic'. In: *Journal of Electron Microscopy Technique* 7.1 (1987), pp. 29–33. DOI: 10.1002/jemt.1060070104 (cit. on p. 75).

[237] Jacques Dubochet, Marc Adrian, Jiin-Ju Chang, Jean-Claude Homo, Jean Le Pault, Alasdair W. McDowall and Patrick Schultz. 'Cryo-electron microscopy of vitrified specimens'. In: *Quarterly Reviews of Biophysics* 21.2 (1988), pp. 129–228. DOI: 10.1017/S0033583500004297 (cit. on p. 75).

[238] Rebecca F. Thompson, Matt Walker, C. Alistair Siebert, Stephen P. Muench and Neil A. Ranson. 'An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology'. In: *Methods* 100.2016 (2016), pp. 3–15. DOI: 10.1016/j.ymeth.2016.02.017 (cit. on p. 76).

[239] Kenneth A. Taylor and Robert M. Glaeser. 'Retrospective on the early development of cryoelectron microscopy of macromolecules and a prospective on opportunities for the future'. In: *Journal of Structural Biology* 163.3 (2008), pp. 214–223. DOI: 10.1016/j.jsb.2008.06.004 (cit. on p. 76).

[240] Jacques Dubochet, Marc Adrian, J. Lepault and Alasdair W. McDowall. 'Emerging techniques: Cryo-electron microscopy of vitrified biological specimens'. In: *Trends in Biochemical Sciences* 10.4 (1985), pp. 143–146. DOI: 10.1016/0968-0004(85)90150-1 (cit. on p. 76).

[241] Alex J. Noble, Venkata P. Dandey, Hui Wei, Julia Brasch, Jillian Chase, Priyamvada Acharya, Yong Zi Tan, Zhening Zhang, Laura Y. Kim, Giovanna Scapin, Micah Rapp, Edward T. Eng, William J. Rice, Anchi Cheng, Carl J. Negro, Lawrence Shapiro, Peter D. Kwong, David Jeruzalmi, Amedee des Georges, Clinton S. Potter and Bridget Carragher. 'Routine single particle CryoEM sample and grid characterization by tomography'. In: *eLife* 7 (2018), pp. 1–42. DOI: 10.7554/eLife.34257 (cit. on pp. 76, 83).

[242] Simon Kemmerling, Jörg Ziegler, Gabriel Schweighauser, Stefan A. Arnold, Dominic Giss, Shirley A. Müller, Philippe Ringler, Kenneth N. Goldie, Nils Goedecke, Andreas Hierlemann, Henning Stahlberg, Andreas Engel and Thomas Braun. 'Connecting μ-fluidics to electron microscopy'. In: *Journal of Structural Biology* 177.1 (2012), pp. 128–134. DOI: 10.1016/j.jsb.2011.11.001 (cit. on p. 76).

[243] Giulia Weissenberger, Rene J. M. Henderikx and Peter J. Peters. 'Understanding the invisible hands of sample preparation for cryo-EM'. In: *Nature Methods* 18.5 (2021), pp. 463–471. DOI: 10.1038/s41592-021-01130-6 (cit. on pp. 77, 263, 266).

[244] Venkata P. Dandey, Hui Wei, Zhening Zhang, Yong Zi Tan, Priyamvada Acharya, Edward T. Eng, William J. Rice, Peter A. Kahn, Clinton S. Potter and Bridget Carragher. 'Spotiton: New features and applications'. In: *Journal of Structural Biology* 202.2 (2018), pp. 161–169. DOI: 10.1016/j.jsb.2018.01.002 (cit. on pp. 77, 78).

[245] Tilak Jain, Patrick Sheehan, John Crum, Bridget Carragher and Clinton S. Potter. 'Spotiton: A prototype for an integrated inkjet dispense and vitrification system for cryo-TEM'. In: *Journal of Structural Biology* 179.1 (2012), pp. 68–75. DOI: 10.1016/j.jsb.2012.04.020 (cit. on p. 77).

[246] Ivan Razinkov, Venkata P. Dandey, Hui Wei, Zhening Zhang, David Melnekoff, William J. Rice, Christoph Wigge, Clinton S. Potter and Bridget Carragher. 'A new method for vitrifying samples for cryoEM'. In: *Journal of Structural Biology* 195.2 (2016), pp. 190–198. DOI: 10.1016/j.jsb.2016.06.001 (cit. on p. 77).

[247] Stefan A. Arnold, Stefan Albiez, Andrej Bieri, Anastasia Syntychaki, Ricardo Adaixo, Robert A. McLeod, Kenneth N. Goldie, Henning Stahlberg and Thomas Braun. 'Blotting-free and lossless cryo-electron microscopy grid preparation from nanoliter-sized protein samples and single-cell extracts'. In: *Journal of Structural Biology* 197.3 (2017), pp. 220–226. DOI: 10.1016/j.jsb.2016.11.002 (cit. on p. 77).

[248] Kaiying Cheng, Martin Wilkinson, Yuriy Chaban and Dale B. Wigley. 'A conformational switch in response to Chi converts RecBCD from phage destruction to DNA repair'. In: *Nature Structural a Molecular Biology* 27.1 (2020), pp. 71–77. DOI: 10.1038/s41594-019-0355-2 (cit. on p. 78).

[249] Ali Punjani, John L. Rubinstein, David J. Fleet and Marcus A. Brubaker. 'cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination'. In: *Nature Methods* 14.3 (2017), pp. 290–296. DOI: 10.1038/nmeth.4169 (cit. on pp. 81, 84, 209, 215).

[250] Richard Henderson and Robert M. Glaeser. 'Quantitative analysis of image contrast in electron micrographs of beam-sensitive crystals'. In: *Ultramicroscopy* 16.2 (1985), pp. 139–150. DOI: 10.1016/0304-3991(85)90069-5 (cit. on p. 82).

[251] Dieter Typke, Christopher J. Gilpin, Kenneth H. Downing and Robert M. Glaeser. 'Stroboscopic image capture: Reducing the dose per frame by a factor of 30 does not prevent beam-induced specimen movement in paraffin'. In: *Ultramicroscopy* 107.2-3 (2007), pp. 106–115. DOI: 10.1016/j.ultramic.2006.06.005 (cit. on p. 82).

[252] Robert M. Glaeser, G. McMullan, A. R. Faruqi and Richard Henderson. 'Images of paraffin monolayer crystals with perfect contrast: Minimization of beam-induced specimen motion'. In: *Ultramicroscopy* 111.2 (2011), pp. 90–100. DOI: 10.1016/j.ultramic.2010.10.010 (cit. on p. 82).

[253] Xueming Li, Paul Mooney, Shawn Zheng, Christopher R. Booth, Michael B. Braunfeld, Sander Gubbens, David A. Agard and Yifan Cheng. 'Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM'. In: *Nature Methods* 10.6 (2013), pp. 584–590. DOI: 10.1038/nmeth.2472 (cit. on p. 82).

[254] Amit Singer and Fred J. Sigworth. 'Computational Methods for Single-Particle Electron Cryomicroscopy'. In: *Annual Review of Biomedical Data Science* 3.1 (2020), pp. 163–190. DOI: 10.1146/annurev-biodatasci-021020-093826 (cit. on p. 82).

[255] Tristan Bepler, Andrew Morin, Micah Rapp, Julia Brasch, Lawrence Shapiro, Alex J. Noble and Bonnie Berger. 'Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs'. In: *Nature Methods* 16.11 (2019), pp. 1153–1160. DOI: 10.1038/s41592-019-0575-8 (cit. on pp. 83, 190, 199).

[256] Giuseppe C. Calafiore and Laurent El Ghaoui. *Optimization Models*. Ed. by Giuseppe C. Calafiore and Laurent El Ghaoui. 1st ed. Cambridge: Cambridge University Press, 2014 (cit. on p. 84).

[257] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng and Thomas E. Ferrin. 'UCSF Chimera - A visualization system for exploratory research and analysis'. In: *Journal of Computational Chemistry* 25.13 (2004), pp. 1605–1612. DOI: 10.1002/jcc.20084 (cit. on pp. 85, 87, 88, 107, 122, 130, 185, 194, 196, 201, 209, 211, 215, 218, 221, 222, 282, 285).

[258] Stephen Curry, Elizabeth E. Fry, Wendy E. Blakemore, Robin Abu-Ghazaleh, Terry Jackson, Andrew M. Q. King, Susan M. Lea, John W. I. Newman, David J. Rowlands and David I. Stuart. 'Perturbations in the surface structure of A22 Iraq foot-and-mouth disease virus accompanying coupled changes in host cell specificity and antigenicity'. In: *Structure* 4.2 (1996), pp. 135–145. DOI: 10.1016/S0969-2126(96)00017-2 (cit. on p. 85).

[259] Nayab Malik, Abhay Kotecha, Sarah Gold, Amin S. Asfor, Jingshan Ren, Juha T. Huiskonen, Tobias J. Tuthill, Elizabeth E. Fry and David I. Stuart. 'Structures of foot and mouth disease virus pentamers: Insight into capsid dissociation and unexpected pentamer reassociation'. In: *PLOS Pathogens* 13.9 (2017). Ed. by Yizhi Jane Tao, e1006607. DOI: 10.1371/journal.ppat.1006607 (cit. on p. 85).

[260] Ali Punjani, Haowei Zhang and David J. Fleet. 'Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction'. In: *Nature Methods* 17.12 (2020), pp. 1214–1221. DOI: 10.1038/s41592-020-00990-8 (cit. on p. 86).

[261] Dorothee Liebschner, Pavel V. Afonine, Matthew L. Baker, Gábor Bunkóczi, Vincent B. Chen, Tristan I. Croll, Bradley Hintze, Li-Wei Hung, Swati Jain, Airlie J. McCoy, Nigel W. Moriarty, Robert D. Oeffner, Billy K. Poon, Michael G. Prisant, Randy J. Read, Jane S. Richardson, David C. Richardson, Massimo D. Sammito, Oleg V. Sobolev, Duncan H. Stockwell, Thomas C. Terwilliger, Alexandre G. Urzhumtsev, Lizbeth L. Videau, Christopher J. Williams and Paul D. Adams. 'Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix'. In: *Acta Crystallographica Section D Structural Biology* 75.10 (2019), pp. 861–877. DOI: 10.1107/S2059798319011471 (cit. on p. 88).

References for Chapter 3

[31] Surinder S. Saini, Brian Allore, Robert M. Jacobs and Azad K. Kaushik. 'Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM antibodies'. In: *European Journal of Immunology* 29.8 (1999), pp. 2420–2426. DOI: 10.1002/(SICI)1521-4141(199908)29:08<2420::AID-IMMU2420>3.0.CO;2-A (cit. on pp. 6, 10, 12–14, 18, 23, 92, 98, 140, 289, 293).

[34] Li Ma, Tong Qin, Dan Chu, Xueqian Cheng, Jing Wang, Xifeng Wang, Peng Wang, Haitang Han, Liming Ren, Robert Aitken, Lennart Hammarström, Ning Li and Yaofeng Zhao. 'Internal Duplications of DH, JH, and C Region Genes Create an Unusual IgH Gene Locus in Cattle'. In: *The Journal of Immunology* 196.10 (2016), pp. 4358–4366. DOI: 10.4049/jimmunol.1600158 (cit. on pp. 7–11, 13, 96, 289).

[61] Surinder S. Saini, William Farrugia, Paul A. Ramsland and Azad K. Kaushik. 'Bovine IgM antibodies with exceptionally long complementarity-determining region 3 of the heavy chain share unique structural properties conferring restricted V H + Vλ pairings'. In: *International Immunology* 15.7 (2003), pp. 845–853. DOI: 10.1093/intimm/dxg083 (cit. on pp. 10, 13, 20, 23, 27, 92, 157, 163, 289).

[78] Feng Wang, Damian C. Ekiert, Insha Ahmad, Wenli Yu, Yong Zhang, Omar A. Bazirgan, Ali Torkamani, Terje Raudsepp, Waithaka Mwangi, Michael F. Criscitiello, Ian A. Wilson, Peter G. Schultz and Vaughn V. Smider. 'Reshaping Antibody Diversity'. In: *Cell* 153.6 (2013), pp. 1379–1393. DOI: 10.1016/j.cell.2013.04.049 (cit. on pp. 12, 13, 15–19, 23–25, 94, 98, 99, 105, 108, 109, 116, 128, 131, 132, 140–142, 162, 163, 180, 181, 206, 232, 289–294).

[82] Robyn L. Stanfield, Ian A. Wilson and Vaughn V. Smider. 'Conservation and diversity in the ultralong third heavy-chain complementarity-determining region of bovine antibodies'. In: *Science Immunology* 1.1 (2016), aaf7962. DOI: 10.1126/sciimmunol.aaf7962 (cit. on pp. 13, 15, 16, 23–25, 94, 98, 108, 109, 116, 140, 141, 162, 163, 180–182, 289, 290, 293).

[83] Jinhui Dong, Jessica A. Finn, Peter A. Larsen, Timothy P. L. Smith and James E. Crowe. 'Structural Diversity of Ultralong CDRH3s in Seven Bovine Antibody Heavy Chains'. In: *Frontiers in Immunology* 10.March (2019), pp. 1–9. DOI: 10.3389/fimmu.2019.00558 (cit. on pp. 13, 15, 16, 19, 23–25, 94, 97, 98, 108, 109, 116, 140, 141, 162, 163, 180, 181, 289, 290, 293).

[84] Robyn L. Stanfield, Zachary T. Berndsen, Ruiqi Huang, Devin Sok, Gabrielle Warner, Jonathan L. Torres, Dennis R. Burton, Andrew B. Ward, Ian A. Wilson and Vaughn V. Smider. 'Structural basis of broad HIV neutralization by a vaccine-induced cow antibody'. In: *Science Advances* 6.22 (2020), eaba0468. DOI: 10.1126/sciadv.aba0468 (cit. on pp. 13, 15, 16, 18, 23–25, 92, 94, 140, 141, 162, 163, 206, 232, 291, 293, 294).

[85] Yong He, Kun Li, Yimei Cao, Zixian Sun, Pinghua Li, Huifang Bao, Sheng Wang, Guoqiang Zhu, Xingwen Bai, Pu Sun, Xuerong Liu, Cheng Yang, Zaixin Liu, Zengjun Lu, Zihe Rao and Zhiyong Lou. 'Structures of Foot-and-mouth Disease Virus with neutralizing antibodies derived from recovered natural host reveal a mechanism for cross-serotype neutralization'. In: *PLOS Pathogens* 17.4 (2021). Ed. by Richard J. Kuhn, e1009507. DOI: 10.1371/journal.ppat.1009507 (cit. on pp. 13, 15, 16, 18, 23–25, 92, 94, 140, 141, 162, 163, 206, 232, 289–291, 293, 294).

[90] G. Lippens, J. Najib, Shoshana J. Wodak and A. Tartar. 'NMR Sequential Assignments and Solution Structure of Chlorotoxin, a Small Scorpion Toxin That Blocks Chloride Channels'. In: *Biochemistry* 34.1 (1995), pp. 13–21. DOI: 10.1021/bi00001a003 (cit. on pp. 17, 19, 94, 289, 291).

[91] Jamie I. Fletcher, Bogdan E. Chapman, Joel P. Mackay, Merlin E. H. Howden and Glenn F. King. 'The structure of versutoxin (δ-atracotoxin-Hv1) provides insights into the binding of site 3 neurotoxins to the voltage-gated sodium channel'. In: *Structure* 5.11 (1997), pp. 1525–1535. DOI: 10.1016/S0969-2126(97)00301-8 (cit. on pp. 17, 19, 94, 289, 291).

[92] Svend Ludvigsen, Lars Thim, Ane M. Blom and Birgitte S. Wulff. 'Solution Structure of the Satiety Factor, CART, Reveals New Functionality of a Well-Known Fold'. In: *Biochemistry* 40.31 (2001), pp. 9082–9088. DOI: 10.1021/bi010433u (cit. on pp. 17, 94).

[93] José C. Martins, Mohammed Enassar, Rudolph Willem, Jean-michel Wieruzeski, Guy Lippens and Shoshana J. Wodak. 'Solution structure of the main α-amylase inhibitor from amaranth seeds'. In: *European Journal of Biochemistry* 268.8 (2001), pp. 2379–2389. DOI: 10.1046/j.1432-1327.2001.02118.x (cit. on pp. 17, 94).

[94] Yuliya V. Korolkova, Eduard V. Bocharov, Kamilla Angelo, Innokenty V. Maslennikov, Olga V. Grinenko, Aleksey V. Lipkin, Elena D. Nosyreva, Kirill A. Pluzhnikov, Søren-peter Olesen, Alexander S. Arseniev and Eugene V. Grishin. 'New Binding Site on Common Molecular Scaffold Provides HERG Channel Specificity of Scorpion Toxin BeKm-1'. In: *Journal of Biological Chemistry* 277.45 (2002), pp. 43104–43109. DOI: 10.1074/jbc.M204083200 (cit. on pp. 17, 19, 94, 289, 291).

[95] Robert E. Oswald, Thomas M. Suchyna, Robert McFeeters, Philip Gottlieb and Frederick Sachs. 'Solution Structure of Peptide Toxins That Block Mechanosensitive Ion Channels'. In: *Journal of Biological Chemistry* 277.37 (2002), pp. 34443–34450. DOI: 10.1074/jbc.M202715200 (cit. on pp. 17, 19, 94, 289, 291).

[96] Katherine J. Nielsen, Michael Watson, David J. Adams, Anna K. Hammarström, Peter W. Gage, Justine M. Hill, David J. Craik, Linda Thomas, Denise Adams, Paul F. Alewood and Richard J. Lewis. 'Solution Structure of μ-Conotoxin PIIIA, a Preferential Inhibitor of Persistent Tetrodotoxin-sensitive Sodium Channels'. In: *Journal of Biological Chemistry* 277.30 (2002), pp. 27247–27255. DOI: 10.1074/jbc.M201611200 (cit. on pp. 17, 19, 94, 289, 291).

[97] David J. Adams, Amanda B. Smith, Christina I. Schroeder, Takahiro Yasuda and Richard J. Lewis. 'ω-Conotoxin CVID Inhibits a Pharmacologically Distinct Voltage-sensitive Calcium Channel Associated with Transmitter Release from Preganglionic Nerve Terminals'. In: *Journal of Biological Chemistry* 278.6 (2003), pp. 4057–4062. DOI: 10.1074/jbc.M209969200 (cit. on pp. 17, 19, 94, 289, 291).

[98] Jorgen Mould, Takahiro Yasuda, Christina I. Schroeder, Aaron M. Beedle, Clinton J. Doering, Gerald W. Zamponi, David J. Adams and Richard J. Lewis. 'The $\alpha 2\delta$ Auxiliary Subunit Reduces Affinity of ω-Conotoxins for Recombinant N-type (Cav2.2) Calcium Channels'. In: *Journal of Biological Chemistry* 279.33 (2004), pp. 34705–34714. DOI: 10.1074/jbc.M310848200 (cit. on pp. 17, 19, 94, 289, 291).

[99] Siddhartha P. Sarma, G. Senthil Kumar, S. Sudarslal, Prathima Iengar, P. Ramasamy, Sujit K. Sikdar, K. S. Krishnan and Padmanabhan Balaram. 'Solution Structure of δ-Am2766: A Highly Hydrophobic δ-Conotoxin fromConus amadis That Inhibits Inactivation of Neuronal Voltage-Gated Sodium Channels'. In: *Chemistry and Biodiversity* 2.4 (2005), pp. 535–556. DOI: 10.1002/cbdv.200590035 (cit. on pp. 17, 19, 94, 289, 291).

[100] Jennifer J. Smith, Justine M. Hill, Michelle J. Little, Graham M. Nicholson, Glenn F. King and Paul F. Alewood. 'Unique scorpion toxin with a putative ancestral fold provides insight into evolution of the inhibitor cystine knot motif'. In: *Proceedings of the National Academy of Sciences* 108.26 (2011), pp. 10478–10483. DOI: 10.1073/pnas.1103501108 (cit. on pp. 17, 94).

[101] Roger J. P. Dawson, Jörg Benz, Peter Stohler, Tim Tetaz, Catherine Joseph, Sylwia Huber, Georg Schmid, Daniela Hügin, Pascal Pflimlin, Gerd Trube, Markus G. Rudolph, Michael Hennig and Armin Ruf. 'Structure of the Acid-sensing ion channel 1 in complex with the gating modifier Psalmotoxin 1'. In: *Nature Communications* 3.1 (2012), p. 936. DOI: 10.1038/ncomms1917 (cit. on pp. 17, 19, 94, 289, 291).

[102] Andrew J. Wommack, Scott A. Robson, Yoshitha A. Wanniarachchi, Andrea Wan, Christopher J. Turner, Gerhard Wagner and Elizabeth M. Nolan. 'NMR Solution Structure and Condition-Dependent Oligomerization of the Antimicrobial Peptide Human Defensin 5'. In: *Biochemistry* 51.48 (2012), pp. 9624–9637. DOI: 10.1021/bi301255u (cit. on pp. 17, 94).

[103] Brad R. Green, Joanna Gajewiak, Sandeep Chhabra, Jack J. Skalicky, Min-Min Zhang, Jean E. Rivier, Grzegorz Bulaj, Baldomero M. Olivera, Doju Yoshikami and Raymond S. Norton. 'Structural Basis for the Inhibition of Voltage-gated Sodium Channels by Conotoxin μO§-GVIIJ'. In: *Journal of Biological Chemistry* 291.13 (2016), pp. 7205–7220. DOI: 10.1074/jbc.M115.697672 (cit. on pp. 17, 19, 94, 289, 291).

[104] Bin Wu, Justin K. Murray, Kristin L. Andrews, Kelvin Sham, Jason Long, Jennifer Aral, Joseph Ligutti, Shanti Amagasu, Dong Liu, Anruo Zou, Xiaoshan Min, Zhulun Wang, Christopher P. Ilch, Thomas J. Kornecook, Min-Hwa Jasmine Lin, Xuhai Be, Les P. Miranda, Bryan D. Moyer and Kaustav Biswas. 'Discovery of Tarantula Venom-Derived Na V 1.7-Inhibitory JzTx-V Peptide 5-Br-Trp24 Analogue AM-6120 with Systemic Block of Histamine-Induced Pruritis'. In: *Journal of Medicinal Chemistry* 61.21 (2018), pp. 9500–9512. DOI: 10.1021/acs.jmedchem.8b00736 (cit. on pp. 17, 19, 94, 289, 291).

[105] Andrew J. Wommack, J. J. Ziarek, Gerhard Wagner and Elizabeth M. Nolan. 'Solution structure of oxidized dimeric form of human defensin 5'. DOI: http://doi.org/10.2210/pdb2MIT/pdb (cit. on pp. 17, 94).

[106] Xinying Jia, Soohyun Kwon, Ching-I. Anderson Wang, Yen-Hua Huang, Lai Y. Chan, Chia Chia Tan, K. Johan Rosengren, Jason P. Mulvenna, Christina I. Schroeder and David J. Craik. 'Semienzymatic Cyclization of Disulfide-rich Peptides Using Sortase A'. In: *Journal of Biological Chemistry* 289.10 (2014), pp. 6627–6638. DOI: 10.1074/jbc.M113.539262 (cit. on pp. 17, 94).

[107] Tien T. Dang, Peta J. Harvey, Lai Yue Chan, Yen-Hua Huang, Quentin Kaas and David J. Craik. 'Mutagenesis of cyclotide Cter 27 exemplifies a robust folding strategy for bracelet cyclotides'. In: *Peptide Science* 114.6 (2022), pp. 2–7. DOI: 10.1002/pep2.24284 (cit. on pp. 17, 94).

[108] Neil Q. McDonald, Risto Lapatto, Judith Murray Rust, Jennifer Gunning, Alexander Wlodawer and Tom L. Blundell. 'New protein fold revealed by a 2.3-Å resolution crystal structure of nerve growth factor'. In: *Nature* 354.6352 (1991), pp. 411–414. DOI: 10.1038/354411a0 (cit. on pp. 17, 94).

[109] Debra R. Holland, Lawrence S. Cousens, Wei Meng and Brian W. Matthews. 'Nerve Growth Factor in Different Crystal Forms Displays Structural Flexibility and Reveals Zinc Binding Sites'. In: *Journal of Molecular Biology* 239.3 (1994), pp. 385–400. DOI: 10.1006/jmbi.1994.1380 (cit. on pp. 17, 94).

[110] Christian Wiesmann, Mark H. Ultsch, Steven H. Bass and Abraham M. de Vos. 'Crystal structure of nerve growth factor in complex with the ligand-binding domain of the TrkA receptor'. In: *Nature* 401.6749 (1999), pp. 184–188. DOI: 10.1038/43705 (cit. on pp. 17, 94).

[111] Shalini Iyer, Demetres D. Leonidas, G. Jawahar Swaminathan, Domenico Maglione, Mauro Battisti, Marina Tucci, M. Graziella Persico and K. Ravi Acharya. 'The Crystal Structure of Human Placenta Growth Factor-1 (PlGF-1), an Angiogenic Protein, at 2.0 Å Resolution'. In: *Journal of Biological Chemistry* 276.15 (2001), pp. 12153–12161. DOI: 10.1074/jbc.M008055200 (cit. on pp. 17, 94).

[112] Kristin M. Fox, James A. Dias and Patrick Van Roey. 'Three-Dimensional Structure of Human Follicle-Stimulating Hormone'. In: *Molecular Endocrinology* 15.3 (2001), pp. 378–389. DOI: 10.1210/mend.15.3.0603 (cit. on pp. 17, 94).

[113] Christopher A. Love, Karl Harlos, Nasim Mavaddat, Simon J. Davis, David I. Stuart, E. Yvonne Jones and Robert M. Esnouf. 'The ligand-binding face of the semaphorins revealed by the high-resolution crystal structure of SEMA4D'. In: *Nature Structural and Molecular Biology* 10.10 (2003), pp. 843–848. DOI: 10.1038/nsb977 (cit. on pp. 17, 94).

[114] Jay Groppe, Jason Greenwald, Ezra Wiater, Joaquin Rodrigues-Leon, Aris N. Economides, Witek Kwiatkowski, Kandan Baban, Markus Affolter, Wylie W. Vale, Juan Carlos Izpisua Belmonte and Senyon Choe. 'Structural basis of BMP signalling inhibition by noggin, a novel twelve-membered cystine knot protien'. In: *The Journal of Bone and Joint Surgery-American Volume* 85.SUPPL. 3 (2003), pp. 52–58. DOI: 10.2106/00004623-200300003-00010 (cit. on pp. 17, 94).

[115] Guennadi Kozlov, Audrey Perreault, Joseph D. Schrag, Morag Park, Miroslaw Cygler, Kalle Gehring and Irena Ekiel. 'Insights into function of PSI domains from structure of the Met receptor PSI domain'. In: *Biochemical and Biophysical Research Communications* 321.1 (2004), pp. 234–240. DOI: 10.1016/j.bbrc.2004.06.132 (cit. on pp. 17, 94).

[116] Qing R. Fan and Wayne A. Hendrickson. 'Structure of human follicle-stimulating hormone in complex with its receptor'. In: *Nature* 433.7023 (2005), pp. 269–277. DOI: 10.1038/nature03206 (cit. on pp. 17, 94).

[117] Michel Pieren, Andrea E. Prota, Claudia Ruch, Dirk Kostrewa, Armin Wagner, Katrin Biedermann, Fritz K. Winkler and Kurt Ballmer-Hofer. 'Crystal Structure of the Orf Virus NZ2 Variant of Vascular Endothelial Growth Factor-E'. In: *Journal of Biological Chemistry* 281.28 (2006), pp. 19578–19587. DOI: 10.1074/jbc.M601842200 (cit. on pp. 17, 94).

[118] Shalini Iyer, Pierre D. Scotney, Andrew D. Nash and K. Ravi Acharya. 'Crystal Structure of Human Vascular Endothelial Growth Factor-B: Identification of Amino Acids Important for Receptor Binding'. In: *Journal of Molecular Biology* 359.1 (2006), pp. 76–85. DOI: 10.1016/j.jmb.2006.03.002 (cit. on pp. 17, 94).

[119] Philip Leonard, Pierre D. Scotney, Talat Jabeen, Shalini Iyer, Louis J. Fabri, Andrew D. Nash and K. Ravi Acharya. 'Crystal Structure of Vascular Endothelial Growth Factor-B in Complex with a Neutralising Antibody Fab Fragment'. In: *Journal of Molecular Biology* 384.5 (2008), pp. 1203–1217. DOI: 10.1016/j.jmb.2008.09.076 (cit. on pp. 17, 94).

[120] Shalini Iyer, Paula I. Darley and K. Ravi Acharya. 'Structural Insights into the Binding of Vascular Endothelial Growth Factor-B by VEGFR-1D2'. In: *Journal of Biological Chemistry* 285.31 (2010), pp. 23779–23789. DOI: 10.1074/jbc.M110.130658 (cit. on pp. 17, 94).

[121] Kalyaneswar Mandal and Stephen B. H. Kent. 'Total Chemical Synthesis of Biologically Active Vascular Endothelial Growth Factor'. In: *Angewandte Chemie International Edition* 50.35 (2011), pp. 8029–8033. DOI: 10.1002/anie.201103237 (cit. on pp. 17, 94).

[122] Marcelo H. Ratto, Yvonne A. Leduc, Ximena P. Valderrama, Karin E. van Straaten, Louis T. J. Delbaere, Roger A. Pierson and Gregg P. Adams. 'The nerve of ovulation-inducing factor in semen'. In: *Proceedings of the National Academy of Sciences* 109.37 (2012), pp. 15042–15047. DOI: 10.1073/pnas.1206273109 (cit. on pp. 17, 94).

[123] Kalyaneswar Mandal, Maruti Uppalapati, Dana Ault-Riché, John Kenney, Joshua Lowitz, Sachdev S. Sidhu and Stephen B. H. Kent. 'Chemical synthesis and X-ray structure of a heterochiral {D-protein antagonist plus vascular endothelial growth factor} protein complex by racemic crystallography'. In: *Proceedings of the National Academy of Sciences* 109.37 (2012), pp. 14779–14784. DOI: 10.1073/pnas.1210483109 (cit. on pp. 17, 94).

[124] Kristof Nolan, Chandramohan Kattamuri, David M. Luedeke, Xiaodi Deng, Amrita Jagpal, Fuming Zhang, Robert J. Linhardt, Alan P. Kenny, Aaron M. Zorn and Thomas B. Thompson. 'Structure of Protein Related to Dan and Cerberus: Insights into the Mechanism of Bone Morphogenetic Protein Antagonism'. In: *Structure* 21.8 (2013), pp. 1417–1429. DOI: 10.1016/j.str.2013.06.005 (cit. on pp. 17, 94).

[125] Byung-Hak Yoon, Luis Esquivies, Chihoon Ahn, Peter C. Gray, Sang-kyu Ye, Witek Kwiatkowski and Senyon Choe. 'An Activin A/BMP2 Chimera, AB204, Displays Bone-Healing Properties Superior to Those of BMP2'. In: *Journal of Bone and Mineral Research* 29.9 (2014), pp. 1950–1959. DOI: 10.1002/jbmr.2238 (cit. on pp. 17, 94).

[126] Aaron Moulin, Magali Mathieu, Catherine Lawrence, Russell Bigelow, Mark Levine, Christine Hamel, Jean-Piere Marquette, Josiane Le Parc, Christophe Loux, Paul Ferrari, Cecile Capdevila, Jacques Dumas, Bruno Dumas, Alexey Rak, Julie Bird, Huawei Qiu, Clark Q. Pan, Tim Edmunds and Ronnie R. Wei. 'Structures of a pan-specific antagonist antibody complexed to different isoforms of TGFβ reveal structural plasticity of antibody-antigen interactions'. In: *Protein Science* 23.12 (2014), pp. 1698–1707. DOI: 10.1002/pro.2548 (cit. on pp. 17, 94).

[127] Kristof Nolan, Chandramohan Kattamuri, Scott A. Rankin, Randy J. Read, Aaron M. Zorn and Thomas B. Thompson. 'Structure of Gremlin-2 in Complex with GDF5 Gives Insight into DAN-Family-Mediated BMP Antagonism'. In: *Cell Reports* 16.8 (2016), pp. 2077–2086. DOI: 10.1016/j.celrep.2016.07.046 (cit. on pp. 17, 94).

[128] Miglė Kišonaitė, Xuelu Wang and Marko Hyvönen. 'Structure of Gremlin-1 and analysis of its interaction with BMP-2'. In: *Biochemical Journal* 473.11 (2016), pp. 1593–1604. DOI: 10.1042/BCJ20160254 (cit. on pp. 17, 94).

[129] Steven B. Larson and Alexander McPherson. 'The crystal structure of the β subunit of luteinizing hormone and a model for the intact hormone'. In: *Current Research in Structural Biology* 1.July (2019), pp. 1–5. DOI: 10.1016/j.crstbi.2019.07.001 (cit. on pp. 17, 94).

[130] Laura del Amo-Maestro, Laura Marino-Puertas, Theodoros Goulas and F. Xavier Gomis-Rüth. 'Recombinant production, purification, crystallization, and structure analysis of human transforming growth factor β2 in a new conformation'. In: *Scientific Reports* 9.1 (2019), p. 8660. DOI: 10.1038/s41598-019-44943-4 (cit. on pp. 17, 94).

[131] C. Robert, F. Bruck, R. Herman, M. Vandevenne, P. Filee, F. Kerff and A. Matagne. 'Functional and structural study of Bone Morphogenetic Protein 2 and its antagonist Noggin'. DOI: http://doi.org/10.2210/pdb7AG0/pdb (cit. on pp. 17, 94).

[132] Hristo L. Svilenov, Julia Sacherl, Ulrike Protzer, Martin Zacharias and Johannes Buchner. 'Mechanistic principles of an ultra-long bovine CDR reveal strategies for antibody design'. In: *Nature Communications* 12.1 (2021), p. 6737. DOI: 10.1038/s41467-021-27103-z (cit. on pp. 17, 19, 108, 140, 162, 206, 232, 291, 294).

[138] Alex Macpherson, Anthony Scott-Tucker, Anastasios Spiliotopoulos, Catherine Simpson, Justin Staniforth, Adam Hold, James R. Snowden, Leah Manning, Jean M. H. van den Elsen and Alastair D. G. Lawson. 'Isolation of antigen-specific, disulphide-rich knob domain peptides from bovine antibodies'. In: *PLOS Biology* 18.9 (2020). Ed. by David Nemazee, e3000821. DOI: 10.1371/journal.pbio.3000821 (cit. on pp. 19, 140, 206, 232, 289, 291, 294).

[139] Alex Macpherson, Maisem Laabei, Zainab Ahdash, Melissa A. Graewert, James R. Birtley, Monika-Sarah E. D. Schulze, Susan Crennell, Sarah A. Robinson, Ben Holmes, Vladas Oleinikovas, Per H. Nilsson, James R. Snowden, Victoria Ellis, Tom Eirik Mollnes, Charlotte M. Deane, Dmitri Svergun, Alastair D. G. Lawson and Jean M. H. van den Elsen. 'The allosteric modulation of complement C5 by knob domain peptides'. In: *eLife* 10 (2021), pp. 1–49. DOI: 10.7554/eLife.63586 (cit. on pp. 19, 26, 140, 206, 232, 289, 291, 294).

[160] Jingshan Ren, Joanne E. Nettleship, Gemma Harris, William N. Mwangi, Nahid Rhaman, Clare F. J. Grant, Abhay Kotecha, Elizabeth E. Fry, Bryan Charleston, David I. Stuart, John A. Hammond and Raymond J. Owens. 'The role of the light chain in the structure and binding activity of two cattle antibodies that neutralize bovine respiratory syncytial virus'. In: *Molecular Immunology* 112 (2019), pp. 123–130. DOI: 10.1016/j.molimm.2019.04.026 (cit. on pp. 25, 26, 33, 92–94, 99, 105, 119, 128, 146, 164, 237, 240–243, 253, 255–258, 260, 282, 283, 285–287, 290, 295).

[161] H. E. Kennedy, B. V. Jones, E. M. Tucker, N. J. Ford, S. W. Clarke, Julie Furze, Lewis H. Thomas and E. J. Stott. 'Production and Characterization of Bovine Monoclonal Antibodies to Respiratory Syncytial Virus'. In: *Journal of General Virology* 69.12 (1988), pp. 3023–3032. DOI: 10.1099/0022-1317-69-12-3023 (cit. on pp. 25, 93, 240).

[162] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov and Philip E. Bourne. 'The Protein Data Bank'. In: *Nucleic Acids Research* 28.1 (2000), pp. 235–242. DOI: 10.1093/nar/28.1.235 (cit. on pp. 26, 64, 72, 73, 93, 94, 105, 137, 171–173, 176, 229, 292).

[194] James Dunbar, Konrad Krawczyk, Jinwoo Leem, Claire Marks, Jaroslaw Nowak, Cristian Regep, Guy Georges, Sebastian Kelm, Bojana Popovic and Charlotte M. Deane. 'SAbPred: a structure-based antibody prediction server'. In: *Nucleic Acids Research* 44.W1 (2016), W474–W478. DOI: 10.1093/nar/gkw361 (cit. on pp. 64, 93–95).

[196] James Dunbar and Charlotte M. Deane. 'ANARCI: Antigen receptor numbering and receptor classification'. In: *Bioinformatics* 32.2 (2016), pp. 298–300. DOI: 10.1093/bioinformatics/btv552 (cit. on pp. 64, 93).

[197] James Dunbar, Konrad Krawczyk, Jinwoo Leem, Terry Baker, Angelika Fuchs, Guy Georges, Jiye Shi and Charlotte M. Deane. 'SAbDab: the structural antibody database'. In: *Nucleic Acids Research* 42.D1 (2014), pp. D1140–D1146. DOI: 10.1093/nar/gkt1043 (cit. on pp. 64, 93, 94).

[200] Andrej Šali and Tom L. Blundell. 'Comparative Protein Modelling by Satisfaction of Spatial Restraints'. In: *Journal of Molecular Biology* 234.3 (1993), pp. 779–815. DOI: 10.1006/jmbi.1993.1626 (cit. on pp. 64, 93).

[203] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli and Demis Hassabis. 'Highly accurate protein structure prediction with AlphaFold'. In: *Nature* May (2021), pp. 1–12. DOI: 10.1038/s41586-021-03819-2 (cit. on pp. 64, 93, 95).

[204] Milot Mirdita, Sergey Ovchinnikov and Martin Steinegger. 'ColabFold - Making protein folding accessible to all C Result visualization'. In: *bioRxiv* (2021), pp. 4–9. DOI: https://doi.org/10.1101/2021.08.15.456425 (cit. on pp. 64, 93, 95, 138, 169, 171, 174, 175, 177, 178).

[208] Juan Sanchez-Weatherby, James Sandy, Halina Mikolajek, Carina M. C. Lobley, Marco Mazzorana, Jon Kelly, Geoff Preece, Rich Littlewood and Thomas L. M. Sørensen. 'VMXi: a fully automated, fully remote, high-flux in situ macromolecular crystallography beamline'. In: *Journal of Synchrotron Radiation* 26.1 (2019), pp. 291–301. DOI: 10.1107/S1600577518015114 (cit. on pp. 67, 102).

[209] S. J. Fisher, K. E. Levik, M. A. Williams, A. W. Ashton and Katherine E. McAuley. 'SynchWeb : a modern interface for ISPyB'. In: *Journal of Applied Crystallography* 48.3 (2015), pp. 927–932. DOI: 10.1107/S1600576715004847 (cit. on pp. 69, 104).

[211] Ulrich Wolfgang Arndt and A. J. Wonacott. *The Rotation Method in X-ray crystallography*. Ed. by Ulrich Wolfgang Arndt and A. J. Wonacott. 1st ed. Amsterdam, North-Holland: Amsterdam:New York:North-Holland Publishing Company, 1977 (cit. on pp. 71, 102).

[212] Zepu Zhang, Nicholas K. Sauter, Henry van den Bedem, Gyorgy Snell and Ashley M. Deacon. 'Automated diffraction image analysis and spot searching for high-throughput crystal screening'. In: *Journal of Applied Crystallography* 39.1 (2006), pp. 112–119. DOI: 10.1107/S0021889805040677 (cit. on pp. 71, 106, 111, 112, 121, 129, 184).

[218] Philip R. Evans and Garib N. Murshudov. 'How good are my data and what is the resolution?' In: *Acta Crystallographica Section D: Biological Crystallography* 69.7 (2013), pp. 1204–1214. DOI: 10.1107/S0907444913000061 (cit. on pp. 72, 114).

[220] Graeme Winter, David G. Waterman, James M. Parkhurst, Aaron S. Brewster, Richard J. Gildea, Markus Gerstel, Luis Fuentes-Montero, Melanie Vollmar, Tara Michels-Clark, Iris D. Young, Nicholas K. Sauter and Gwyndaf Evans. 'DIALS : implementation and evaluation of a new integration package'. In: *Acta Crystallographica Section D Structural Biology* 74.2 (2018), pp. 85–97. DOI: 10.1107/S2059798317017235 (cit. on pp. 72, 111, 114).

[221] Martyn D. Winn, Charles C. Ballard, Kevin D. Cowtan, Eleanor J. Dodson, Paul Emsley, Philip R. Evans, Ronan M. Keegan, Eugene B. Krissinel, Andrew G. W. Leslie, Airlie J. McCoy, Stuart J. McNicholas, Garib N. Murshudov, Navraj S. Pannu, Elizabeth A. Potterton, Harold R. Powell, Randy J. Read, Alexei A. Vagin and Keith S. Wilson. 'Overview of the CCP4 suite and current developments'. In: *Acta Crystallographica Section D Biological Crystallography* 67.4 (2011), pp. 235–242. DOI: 10.1107/S0907444910045749 (cit. on pp. 72, 74, 105, 109, 115, 119, 128, 134, 182).

[223] Airlie J. McCoy, Ralf W. Grosse-Kunstleve, Paul D. Adams, Martyn D. Winn, Laurent C. Storoni and Randy J. Read. 'Phaser crystallographic software'. In: *Journal of Applied Crystallography* 40.4 (2007), pp. 658–674. DOI: 10.1107/S0021889807021206 (cit. on pp. 72, 105, 114, 119, 128, 182).

[224] Gerrit Langer, Serge X. Cohen, Victor S. Lamzin and Anastassis Perrakis. 'Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7'. In: *Nature Protocols* 3.7 (2008), pp. 1171–1179. DOI: 10.1038/nprot.2008.91 (cit. on pp. 73, 114).

[227] Garib N. Murshudov, Pavol Skubák, Andrey A. Lebedev, Navraj S. Pannu, Roberto A. Steiner, Robert A. Nicholls, Martyn D. Winn, Fei Long and Alexei A. Vagin. 'REFMAC 5 for the refinement of macromolecular crystal structures'. In: *Acta Crystallographica Section D Biological Crystallography* 67.4 (2011), pp. 355–367. DOI: 10.1107/S0907444911001314 (cit. on pp. 73, 105, 114, 119, 128, 182).

[228] Paul Emsley and Kevin D. Cowtan. 'Coot : model-building tools for molecular graphics'. In: *Acta Crystallographica Section D Biological Crystallography* 60.12 (2004), pp. 2126–2132. DOI: 10.1107/S0907444904019158 (cit. on pp. 73, 88, 114, 218).

[232] Christopher J. Williams, Jeffrey J. Headd, Nigel W. Moriarty, Michael G. Prisant, Lizbeth L. Videau, Lindsay N. Deis, Vishal Verma, Daniel A. Keedy, Bradley J. Hintze, Vincent B. Chen, Swati Jain, Steven M. Lewis, W. Bryan Arendall, Jack Snoeyink, Paul D. Adams, Simon C. Lovell, Jane S. Richardson and David C. Richardson. 'MolProbity: More and better

reference data for improved all-atom structure validation'. In: *Protein Science* 27.1 (2018), pp. 293–315. DOI: 10.1002/pro.3330 (cit. on pp. 74, 88, 105).

[233] Evgeny Krissinel and Kim Henrick. 'Inference of Macromolecular Assemblies from Crystalline State'. In: *Journal of Molecular Biology* 372.3 (2007), pp. 774–797. DOI: 10.1016/j.jmb.2007.05.022 (cit. on pp. 74, 109, 111, 134, 195).

[257] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng and Thomas E. Ferrin. 'UCSF Chimera - A visualization system for exploratory research and analysis'. In: *Journal of Computational Chemistry* 25.13 (2004), pp. 1605–1612. DOI: 10.1002/jcc.20084 (cit. on pp. 85, 87, 88, 107, 122, 130, 185, 194, 196, 201, 209, 211, 215, 218, 221, 222, 282, 285).

[262] Paul A. Ramsland, Azad K. Kaushik, John J. Marchalonis and Allen B. Edmundson. 'Incorporation of long CDR3s into V domains: Implications for the structural evolution of the antibody-combining site'. In: *Experimental and Clinical Immunogenetics* 18.4 (2001), pp. 176–198. DOI: 10.1159/000049197 (cit. on p. 93).

[263] Pablo Guardado Calvo. *RotationAxis*. URL: https://pymolwiki.org/index.php/RotationAxis (visited on 13th Aug. 2022) (cit. on pp. 123, 128).

[264] Shivender Shandilya, Jason Vertrees and Thomas Holder. *ColorByRMSD*. URL: https://pymolwiki.org/index.php/ColorByRMSD (visited on 3rd May 2022) (cit. on pp. 128, 195).

[265] Cyrus Chothia, Arthur M. Lesk, Michael Levitt, Adolfo G. Amit, Roy A. Mariuzza, Simon E. V. Phillips and Roberto J. Poljak. 'The Predicted Structure of Immunoglobulin D1.3 and Its Comparison with the Crystal Structure'. In: *Science* 233.4765 (1986), pp. 755–758. DOI: 10.1126/science.3090684 (cit. on p. 131).

[266] Bissan Al-Lazikani, Arthur M. Lesk and Cyrus Chothia. 'Standard conformations for the canonical structures of immunoglobulins'. In: *Journal of Molecular Biology* 273.4 (1997), pp. 927–948. DOI: 10.1006/jmbi.1997.1354 (cit. on p. 131).

[267] Jared Sampson. *Elbow angle*. URL: https://pymolwiki.org/index.php/Elbow_angle (cit. on pp. 132, 292).

References for Chapter 4

[61] Surinder S. Saini, William Farrugia, Paul A. Ramsland and Azad K. Kaushik. 'Bovine IgM antibodies with exceptionally long complementarity-determining region 3 of the heavy chain share unique structural properties conferring restricted V H + Vλ pairings'. In: *International Immunology* 15.7 (2003), pp. 845–853. DOI: 10.1093/intimm/dxg083 (cit. on pp. 10, 13, 20, 23, 27, 92, 157, 163, 289).

[78] Feng Wang, Damian C. Ekiert, Insha Ahmad, Wenli Yu, Yong Zhang, Omar A. Bazirgan, Ali Torkamani, Terje Raudsepp, Waithaka Mwangi, Michael F. Criscitiello, Ian A. Wilson, Peter G. Schultz and Vaughn V. Smider. 'Reshaping Antibody Diversity'. In: *Cell* 153.6 (2013), pp. 1379–1393. DOI: 10.1016/j.cell.2013.04.049 (cit. on pp. 12, 13, 15–19, 23–25, 94, 98, 99, 105, 108, 109, 116, 128, 131, 132, 140–142, 162, 163, 180, 181, 206, 232, 289–294).

[82] Robyn L. Stanfield, Ian A. Wilson and Vaughn V. Smider. 'Conservation and diversity in the ultralong third heavy-chain complementarity-determining region of bovine antibodies'. In: *Science Immunology* 1.1 (2016), aaf7962. DOI: 10.1126/sciimmunol.aaf7962 (cit. on pp. 13, 15, 16, 23–25, 94, 98, 108, 109, 116, 140, 141, 162, 163, 180–182, 289, 290, 293).

[83] Jinhui Dong, Jessica A. Finn, Peter A. Larsen, Timothy P. L. Smith and James E. Crowe. 'Structural Diversity of Ultralong CDRH3s in Seven Bovine Antibody Heavy Chains'. In: *Frontiers in Immunology* 10.March (2019), pp. 1–9. DOI: 10.3389/fimmu.2019.00558 (cit. on pp. 13, 15, 16, 19, 23–25, 94, 97, 98, 108, 109, 116, 140, 141, 162, 163, 180, 181, 289, 290, 293).

[84] Robyn L. Stanfield, Zachary T. Berndsen, Ruiqi Huang, Devin Sok, Gabrielle Warner, Jonathan L. Torres, Dennis R. Burton, Andrew B. Ward, Ian A. Wilson and Vaughn V. Smider. 'Structural basis of broad HIV neutralization by a vaccine-induced cow antibody'. In: *Science Advances* 6.22 (2020), eaba0468. DOI: 10.1126/sciadv.aba0468 (cit. on pp. 13, 15, 16, 18, 23–25, 92, 94, 140, 141, 162, 163, 206, 232, 291, 293, 294).

[85] Yong He, Kun Li, Yimei Cao, Zixian Sun, Pinghua Li, Huifang Bao, Sheng Wang, Guoqiang Zhu, Xingwen Bai, Pu Sun, Xuerong Liu, Cheng Yang, Zaixin Liu, Zengjun Lu, Zihe Rao and Zhiyong Lou. 'Structures of Foot-and-mouth Disease Virus with neutralizing antibodies derived from recovered natural host reveal a mechanism for cross-serotype neutralization'. In: *PLOS Pathogens* 17.4 (2021). Ed. by Richard J. Kuhn, e1009507. DOI: 10.1371/journal.ppat.1009507 (cit. on pp. 13, 15, 16, 18, 23–25, 92, 94, 140, 141, 162, 163, 206, 232, 289–291, 293, 294).

[132] Hristo L. Svilenov, Julia Sacherl, Ulrike Protzer, Martin Zacharias and Johannes Buchner. 'Mechanistic principles of an ultra-long bovine CDR reveal strategies for antibody design'. In: *Nature Communications* 12.1 (2021), p. 6737. DOI: 10.1038/s41467-021-27103-z (cit. on pp. 17, 19, 108, 140, 162, 206, 232, 291, 294).

[137] Devin Sok, Khoa M. Le, Melissa L. Vadnais, Karen L. Saye-Francisco, Joseph G. Jardine, Jonathan L. Torres, Zachary T. Berndsen, Leopold Kong, Robyn L. Stanfield, Jennifer Ruiz, Alejandra Ramos, Chi-Hui Liang, Patricia L. Chen, Michael F. Criscitiello, Waithaka Mwangi, Ian A. Wilson, Andrew B. Ward, Vaughn V. Smider and Dennis R. Burton. 'Rapid elicitation of broadly neutralizing antibodies to HIV by immunization in cows'. In: *Nature* 548.1 (2017), pp. 108–111. DOI: 10.1038/nature23301 (cit. on pp. 18, 206, 293, 294).

[138] Alex Macpherson, Anthony Scott-Tucker, Anastasios Spiliotopoulos, Catherine Simpson, Justin Staniforth, Adam Hold, James R. Snowden, Leah Manning, Jean M. H. van den Elsen and Alastair D. G. Lawson. 'Isolation of antigen-specific, disulphide-rich knob domain peptides from bovine antibodies'. In:

PLOS Biology 18.9 (2020). Ed. by David Nemazee, e3000821. DOI: 10.1371/journal.pbio.3000821 (cit. on pp. 19, 140, 206, 232, 289, 291, 294).

[139] Alex Macpherson, Maisem Laabei, Zainab Ahdash, Melissa A. Graewert, James R. Birtley, Monika-Sarah E. D. Schulze, Susan Crennell, Sarah A. Robinson, Ben Holmes, Vladas Oleinikovas, Per H. Nilsson, James R. Snowden, Victoria Ellis, Tom Eirik Mollnes, Charlotte M. Deane, Dmitri Svergun, Alastair D. G. Lawson and Jean M. H. van den Elsen. 'The allosteric modulation of complement C5 by knob domain peptides'. In: *eLife* 10 (2021), pp. 1–49. DOI: 10.7554/eLife.63586 (cit. on pp. 19, 26, 140, 206, 232, 289, 291, 294).

[160] Jingshan Ren, Joanne E. Nettleship, Gemma Harris, William N. Mwangi, Nahid Rhaman, Clare F. J. Grant, Abhay Kotecha, Elizabeth E. Fry, Bryan Charleston, David I. Stuart, John A. Hammond and Raymond J. Owens. 'The role of the light chain in the structure and binding activity of two cattle antibodies that neutralize bovine respiratory syncytial virus'. In: *Molecular Immunology* 112 (2019), pp. 123–130. DOI: 10.1016/j.molimm.2019.04.026 (cit. on pp. 25, 26, 33, 92–94, 99, 105, 119, 128, 146, 164, 237, 240–243, 253, 255–258, 260, 282, 283, 285–287, 290, 295).

[162] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov and Philip E. Bourne. 'The Protein Data Bank'. In: *Nucleic Acids Research* 28.1 (2000), pp. 235–242. DOI: 10.1093/nar/28.1.235 (cit. on pp. 26, 64, 72, 73, 93, 94, 105, 137, 171–173, 176, 229, 292).

[176] Claudine Porta, Abhay Kotecha, Alison Burman, Terry Jackson, Jingshan Ren, Silvia Loureiro, Ian M. Jones, Elizabeth E. Fry, David I. Stuart and Bryan Charleston. 'Rational Engineering of Recombinant Picornavirus Capsids to Produce Safe, Protective Vaccine Antigen'. In: *PLoS Pathogens* 9.3 (2013). Ed. by Félix A. Rey, e1003255. DOI: 10.1371/journal.ppat.1003255 (cit. on pp. 36, 37, 85, 152).

[177] Claudine Porta, Xiaodong Xu, Silvia Loureiro, Saravanan Paramasivam, Junyuan Ren, Tara Al-Khalil, Alison Burman, Terry Jackson, Graham J. Belsham, Stephen Curry, George P. Lomonossoff, Satya Parida, David J. Paton, Yanmin Li, Ginette Wilsden, Nigel Ferris, Raymond J. Owens, Abhay Kotecha, Elizabeth E. Fry, David I. Stuart, Bryan Charleston and Ian M. Jones. 'Efficient production of foot-and-mouth disease virus empty capsids in insect cells following down regulation of 3C protease activity'. In: *Journal of Virological Methods* 187.2 (2013), pp. 406–412. DOI: 10.1016/j.jviromet.2012.11.011 (cit. on pp. 36, 56, 152).

[181] Stewart Goodwin, Tobias J. Tuthill, Armando Arias, Richard A. Killington and David J. Rowlands. 'Foot-and-Mouth Disease Virus Assembly: Processing of Recombinant Capsid Precursor by Exogenous Protease Induces Self-Assembly of Pentamers In Vitro in a Myristoylation-Dependent Manner'. In: *Journal of Virology* 83.21 (2009), pp. 11275–11282. DOI: 10.1128/JVI.01263-09 (cit. on pp. 56, 148).

[182] Ravindra Acharya, Elizabeth E. Fry, David I. Stuart, Graham Fox, David J. Rowlands and Fred Brown. 'The three-dimensional structure of foot-and-mouth disease virus at 2.9 Å resolution'. In: *Nature* 337.6209 (1989), pp. 709–716. DOI: 10.1038/337709a0 (cit. on pp. 56, 85, 148, 149, 151, 188, 202).

[183] Abhay Kotecha, Julian Seago, Katherine A. Scott, Alison Burman, Silvia Loureiro, Jingshan Ren, Claudine Porta, Helen M. Ginn, Terry Jackson, Eva Perez-Martin, C. Alistair Siebert, Guntram Paul, Juha T. Huiskonen, Ian M. Jones, Robert M. Esnouf, Elizabeth E. Fry, Francois F. Maree, Bryan Charleston and David I. Stuart. 'Structure-based energetics of protein interfaces guides foot-and-mouth disease virus vaccine design'. In: *Nature Structural and Molecular Biology* 22.10 (2015), pp. 788–794. DOI: 10.1038/nsmb.3096 (cit. on pp. 56, 85, 193, 194, 210, 211, 216, 218, 221, 222).

[184] William N. Mwangi, Giuseppe Maccari, Jayne C. Hope, Gary Entrican and John A. Hammond. 'The UK Veterinary Immunological Toolbox Website: promoting vaccine research by facilitating communication and removing reagent barriers'. In: *Immunology* 161.1 (2020), pp. 25–27. DOI: 10.1111/imm.13227 (cit. on pp. 57, 58, 223).

[204] Milot Mirdita, Sergey Ovchinnikov and Martin Steinegger. 'ColabFold - Making protein folding accessible to all C Result visualization'. In: *bioRxiv* (2021), pp. 4–9. DOI: https://doi.org/10.1101/2021.08.15.456425 (cit. on pp. 64, 93, 95, 138, 169, 171, 174, 175, 177, 178).

[212] Zepu Zhang, Nicholas K. Sauter, Henry van den Bedem, Gyorgy Snell and Ashley M. Deacon. 'Automated diffraction image analysis and spot searching for high-throughput crystal screening'. In: *Journal of Applied Crystallography* 39.1 (2006), pp. 112–119. DOI: 10.1107/S0021889805040677 (cit. on pp. 71, 106, 111, 112, 121, 129, 184).

[221] Martyn D. Winn, Charles C. Ballard, Kevin D. Cowtan, Eleanor J. Dodson, Paul Emsley, Philip R. Evans, Ronan M. Keegan, Eugene B. Krissinel, Andrew G. W. Leslie, Airlie J. McCoy, Stuart J. McNicholas, Garib N. Murshudov, Navraj S. Pannu, Elizabeth A. Potterton, Harold R. Powell, Randy J. Read, Alexei A. Vagin and Keith S. Wilson. 'Overview of the CCP4 suite and current developments'. In: *Acta Crystallographica Section D Biological Crystallography* 67.4 (2011), pp. 235–242. DOI: 10.1107/S0907444910045749 (cit. on pp. 72, 74, 105, 109, 115, 119, 128, 134, 182).

[223] Airlie J. McCoy, Ralf W. Grosse-Kunstleve, Paul D. Adams, Martyn D. Winn, Laurent C. Storoni and Randy J. Read. 'Phaser crystallographic software'. In: *Journal of Applied Crystallography* 40.4 (2007), pp. 658–674. DOI: 10.1107/S0021889807021206 (cit. on pp. 72, 105, 114, 119, 128, 182).

[227] Garib N. Murshudov, Pavol Skubák, Andrey A. Lebedev, Navraj S. Pannu, Roberto A. Steiner, Robert A. Nicholls, Martyn D. Winn, Fei Long and Alexei A. Vagin. 'REFMAC 5 for the refinement of macromolecular crystal structures'. In: *Acta Crystallographica Section D Biological Crystallography* 67.4 (2011), pp. 355–367. DOI: 10.1107/S0907444911001314 (cit. on pp. 73, 105, 114, 119, 128, 182).

[228] Paul Emsley and Kevin D. Cowtan. 'Coot : model-building tools for molecular graphics'. In: *Acta Crystallographica Section D Biological Crystallography* 60.12 (2004), pp. 2126–2132. DOI: 10.1107/S0907444904019158 (cit. on pp. 73, 88, 114, 218).

[233] Evgeny Krissinel and Kim Henrick. 'Inference of Macromolecular Assemblies from Crystalline State'. In: *Journal of Molecular Biology* 372.3 (2007), pp. 774–797. DOI: 10.1016/j.jmb.2007.05.022 (cit. on pp. 74, 109, 111, 134, 195).

[249] Ali Punjani, John L. Rubinstein, David J. Fleet and Marcus A. Brubaker. 'cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination'. In: *Nature Methods* 14.3 (2017), pp. 290–296. DOI: 10.1038/nmeth.4169 (cit. on pp. 81, 84, 209, 215).

[255] Tristan Bepler, Andrew Morin, Micah Rapp, Julia Brasch, Lawrence Shapiro, Alex J. Noble and Bonnie Berger. 'Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs'. In: *Nature Methods* 16.11 (2019), pp. 1153–1160. DOI: 10.1038/s41592-019-0575-8 (cit. on pp. 83, 190, 199).

[257] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng and Thomas E. Ferrin. 'UCSF Chimera - A visualization system for exploratory research and analysis'. In: *Journal of Computational Chemistry* 25.13 (2004), pp. 1605–1612. DOI: 10.1002/jcc.20084 (cit. on pp. 85, 87, 88, 107, 122, 130, 185, 194, 196, 201, 209, 211, 215, 218, 221, 222, 282, 285).

[264] Shivender Shandilya, Jason Vertrees and Thomas Holder. *ColorByRMSD*. URL: https://pymolwiki.org/index.php/ColorByRMSD (visited on 3rd May 2022) (cit. on pp. 128, 195).

[268] Carolina Stenfeldt, Juan M. Pacheco, Manuel V. Borca, Luis L. Rodriguez and Jonathan Arzt. 'Morphologic and phenotypic characteristics of myocarditis in two pigs infected by foot-and mouth disease virus strains of serotypes O or A'. In: *Acta Veterinaria Scandinavica* 56.1 (2014), p. 42. DOI: 10.1186/s13028-014-0042-6 (cit. on p. 147).

[269] G. Ferrari, L. Tasciotti, E. Khan and A. Kiani. 'Foot-and-Mouth Disease and Its Effect on Milk Yield: An Economic Analysis on Livestock Holders in Pakistan'. In: *Transboundary and Emerging Diseases* 61.6 (2014), e52–e59. DOI: 10.1111/tbed.12072 (cit. on p. 147).

[270] Jonathan Arzt, Graham J. Belsham, Louise Lohse, Anette Bøtner and Carolina Stenfeldt. 'Transmission of Foot-and-Mouth Disease from Persistently Infected Carrier Cattle to Naive Cattle via Transfer of Oropharyngeal Fluid'. In: *mSphere* 3.5 (2018). Ed. by Matthew B. Frieman, pp. 1–12. DOI: 10.1128/mSphere.00365-18 (cit. on p. 147).

[271] T. J. D. Knight-Jones and J. Rushton. 'The economic impacts of foot and mouth disease – What are they, how big are they and where do they occur?' In: *Preventive Veterinary Medicine* 112.3-4 (2013), pp. 161–173. DOI: 10.1016/j.prevetmed.2013.07.013 (cit. on p. 148).

[272] H. Vallée and H. Carré. 'Sur la pluralité du virus aphteux.' In: *Comptes rendus de l'Académie des sciences Paris* 174 (1922), pp. 1498–1500 (cit. on p. 148).

[273] J. B. Brooksby. 'The Virus of Foot-and-Mouth Disease'. In: *Advances in Virus Research*. Vol. 5. 1958, pp. 1–37. DOI: 10.1016/S0065-3527(08)60670-3 (cit. on p. 148).

[274] M. R. Dhanda, V. R. Gopalakrishnan and H. S. Dhillon. 'Note on the occurrence of atypical strains of foot-and-mouth diseases virus in India.' In: *Indian Journal of Veterinary Science and Animal Husbandry* 26.95 (1957), pp. 1956–1957 (cit. on p. 148).

[275] O. Waldmann and K. Trautwein. 'Experimentelle Untersuchungen über die Pluralität des Maul-und Klauenseuchevirus.' In: *Berliner tierärztliche Wochenschrift,* 42 (1926), pp. 569–571 (cit. on p. 148).

[276] E. Correa Melo. 'Outbreaks of FMD Type C in Amazonas, Brazil'. In: *Organisation for European Economic Co-operation. EMPRES Transboundary Animal Disease Bulletin* 26 (2004), pp. 31–34 (cit. on p. 148).

[277] David J. Paton, Antonello Di Nardo, Nick J. Knowles, Jemma Wadsworth, Edviges M. Pituco, Ottorino Cosivi, Alejandro M. Rivera, Labib Bakkali Kassimi, Emiliana Brocchi, Kris de Clercq, Consuelo Carrillo, Francois F. Maree, Raj Kumar Singh, Wilna Vosloo, Min-Kyung Park, Keith J. Sumption, Anna B. Ludi and Donald P. King. 'The history of foot-and-mouth disease virus serotype C: the first known extinct serotype?' In: *Virus Evolution* 7.1 (2021), pp. 1–12. DOI: 10.1093/ve/veab009 (cit. on p. 148).

[278] Nick J. Knowles and Alan R. Samuel. 'Molecular epidemiology of foot-and-mouth disease virus'. In: *Virus Research* 91.1 (2003), pp. 65–80. DOI: 10.1016/S0168-1702(02)00260-5 (cit. on p. 148).

[279] Yuan Gao, Shi-Qi Sun and Hui-Chen Guo. 'Biological function of Foot-and-mouth disease virus non-structural proteins and non-coding elements'. In: *Virology Journal* 13.1 (2016), p. 107. DOI: 10.1186/s12985-016-0561-z (cit. on p. 148).

[280] Daniel T. Haydon, Alan R. Samuel and Nick J. Knowles. 'The generation and persistence of genetic variation in foot-and-mouth disease virus'. In: *Preventive Veterinary Medicine* 51.1-2 (2001), pp. 111–124. DOI: 10.1016/S0167-5877(01)00210-0 (cit. on p. 148).

[281] Esteban Domingo, Carmen M. Ruiz-Jarabo, Saleta Sierra, Armando Arias, Nonia Pariente, Eric Baranowski and Cristina Escarmís. 'Emergence and selection of RNA virus variants: memory and extinction'. In: *Virus Research* 82.1-2 (2002), pp. 39–44. DOI: 10.1016/S0168-1702(01)00385-9 (cit. on p. 148).

[282] Michael G. Rossmann, Edward Arnold, John W. Erickson, Elizabeth A. Frankenberger, James P. Griffith, Hans-Jürgen Hecht, John E. Johnson, Greg Kamer, Ming Luo, Anne G. Mosser, Roland R. Rueckert, Barbara Sherry and Gerrit Vriend. 'Structure of a human common cold virus and functional relationship to other picornaviruses'. In: *Nature* 317.6033 (1985), pp. 145–153. DOI: 10.1038/317145a0 (cit. on p. 149).

[283] James M. Hogle, M. Chow and David J. Filman. 'Three-Dimensional Structure of Poliovirus at 2.9 Å Resolution'. In: *Science* 229.4720 (1985), pp. 1358–1365. DOI: 10.1126/science.2994218 (cit. on p. 149).

[284] Ming Luo, Gerrit Vriend, Greg Kamer, Iwona Minor, Edward Arnold, Michael G. Rossmann, Ulrike Boege, Douglas G. Scraba, Greg M. Duke and Ann C. Palmenberg. 'The Atomic Structure of Mengo Virus at 3.0 Å Resolution'. In: *Science* 235.4785 (1987), pp. 182–191. DOI: 10.1126/science.3026048 (cit. on p. 149).

[285] Abhay Kotecha, Quan Wang, Xianchi Dong, Serban L. Ilca, Marina Ondiviela, Rao Zihe, Julian Seago, Bryan Charleston, Elizabeth E. Fry, Nicola G. A. Abrescia, Timothy A. Springer, Juha T. Huiskonen and David I. Stuart. 'Rules of engagement between $\alpha v\beta 6$ integrin and foot-and-mouth disease virus'. In: *Nature Communications* 8 (2017), p. 15408. DOI: 10.1038/ncomms15408 (cit. on p. 149).

[286] John Bourn. *The 2001 Outbreak of Foot and Mouth Disease*. Tech. rep. London: The National Audit Office, 2002, pp. 1–133 (cit. on p. 150).

[287] Maggie Mort, Ian Convery, Josephine Baxter and Cathy Bailey. 'Psychosocial effects of the 2001 UK foot and mouth disease epidemic in a rural population: qualitative diary based study'. In: *BMJ* 331.7527 (2005), p. 1234. DOI: 10.1136/bmj.38603.375856.68 (cit. on p. 150).

[288] John Collinge, Katie C. L. Sidle, Julie Meads, James Ironside and Andrew F. Hill. 'Molecular analysis of prion strain variation and the aetiology of 'new variant' CJD'. In: *Nature* 383.6602 (1996), pp. 685–690. DOI: 10.1038/383685a0 (cit. on p. 150).

[289] Charles Arthur. *Agonising decline that led to first diagnosis of new illness*. 1997 (cit. on p. 150).

[290] Marvin J. Grubman and Barry Baxt. 'Foot-and-Mouth Disease'. In: *Clinical Microbiology Reviews* 17.2 (2004), pp. 465–493. DOI: 10.1128/CMR.17.2.465-493.2004 (cit. on pp. 151, 153).

[291] John Beale. 'Foot and Mouth Disease: Why not Vaccinate?' In: *Journal of the Royal Society of Medicine* 94.6 (2001), pp. 263–264. DOI: 10.1177/014107680109400602 (cit. on pp. 151, 152).

[292] T. R. Doel, L. Williams and Paul V. Barnett. 'Emergency vaccination against foot-and-mouth disease: Rate of development of immunity and its implications for the carrier state'. In: *Vaccine* 12.7 (1994), pp. 592–600. DOI: 10.1016/0264-410X(94)90262-3 (cit. on pp. 151, 152).

[293] Tatjana Sitt, Mary Kenney, José Barrera, Mital Pandya, Korin Eckstrom, Megan Warner, Juan M. Pacheco, Michael LaRocco, Javier Palarea-Albaladejo, David Brake, Elizabeth Rieder, Jonathan Arzt, John W. Barlow and William T. Golde. 'Duration of protection and humoral immunity induced by an adenovirus-vectored subunit vaccine for foot-and-mouth disease (FMD) in Holstein steers'. In: *Vaccine* 37.42 (2019), pp. 6221–6231. DOI: 10.1016/j.vaccine.2019.08.017 (cit. on p. 151).

[294] José Naranjo and Ottorino Cosivi. 'Elimination of foot-and-mouth disease in South America: lessons and challenges'. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 368.1623 (2013), p. 20120381. DOI: 10.1098/rstb.2012.0381 (cit. on p. 152).

[295] David C. Ansardi, Donna C. Porter and Casey D. Morrow. 'Coinfection with recombinant vaccinia viruses expressing poliovirus P1 and P3 proteins results in polyprotein processing and formation of empty capsid structures'. In: *Journal of Virology* 65.4 (1991), pp. 2088–2092. DOI: 10.1128/jvi.65.4.2088-2092.1991 (cit. on p. 152).

[296] Jan Roosien, Graham J. Belsham, Martin D. Ryan, Andrew M. Q. King and Just M. Vlak. 'Synthesis of Foot-and-mouth Disease Virus Capsid Proteins in Insect Cells Using Baculovirus Expression vectors'. In: *Journal of General Virology* 71.8 (1990), pp. 1703–1711. DOI: 10.1099/0022-1317-71-8-1703 (cit. on pp. 152, 153).

[297] Charles C. Abrams, Andrew M. Q. King and Graham J. Belsham. 'Assembly of foot-and-mouth disease virus empty capsids synthesized by a vaccinia virus expression system'. In: *Journal of General Virology* 76.12 (1995), pp. 3089–3098. DOI: 10.1099/0022-1317-76-12-3089 (cit. on p. 152).

[298] Stephen Curry, Elizabeth E. Fry, Wendy E. Blakemore, Robin Abu-Ghazaleh, Terry Jackson, Andrew M. Q. King, Susan M. Lea, John W. I. Newman and David I. Stuart. 'Dissecting the roles of VP0 cleavage and RNA packaging in picornavirus capsid stabilization: the structure of empty capsids of foot-and-mouth disease virus'. In: *Journal of Virology* 71.12 (1997), pp. 9743–9752. DOI: 10.1128/jvi.71.12.9743-9752.1997 (cit. on p. 152).

[299] Thomas R. Fuerst, Edward G. Niles, F. William Studier and Bernard Moss. 'Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase.' In: *Proceedings of the National Academy of Sciences* 83.21 (1986), pp. 8122–8126. DOI: 10.1073/pnas.83.21.8122 (cit. on p. 152).

[300] Graham J. Belsham, Charles C. Abrams, Andrew M. Q. King, Jan Roosien and Just M. Vlak. 'Myristoylation of foot-and-mouth disease virus capsid protein precursors is independent of other viral proteins and occurs in both mammalian and insect cells'. In: *Journal of General Virology* 72.3 (1991), pp. 747–751. DOI: 10.1099/0022-1317-72-3-747 (cit. on p. 153).

[301] Marvin J. Grubman, Samuel A. Lewis and Donald O. Morgan. 'Protection of swine against foot-and-mouth disease with viral capsid proteins expressed in heterologous systems'. In: *Vaccine* 11.8 (1993), pp. 825–829. DOI: 10.1016/0264-410X(93)90357-4 (cit. on p. 153).

[302] Samuel A. Lewis, Donald O. Morgan and Marvin J. Grubman. 'Expression, processing, and assembly of foot-and-mouth disease virus capsid structures in heterologous systems: induction of a neutralizing antibody response in guinea pigs'. In: *Journal of Virology* 65.12 (1991), pp. 6572–6580. DOI: 10.1128/jvi.65.12.6572-6580.1991 (cit. on p. 153).

[303] T. R. Doel. 'Natural and vaccine-induced immunity to foot and mouth disease: the prospects for improved vaccines'. In: *Revue Scientifique et Technique de l'OIE* 15.3 (1996), pp. 883–911. DOI: 10.20506/rst.15.3.955 (cit. on p. 153).

[304] Clare F. J. Grant. 'Investigating primary and secondary B cell responses in cattle after immunisation with existing and novel vaccines'. PhD thesis. University of Oxford, 2013 (cit. on pp. 153, 310).

[305] G. E. Cottral and P. Gailiunas. 'Experimental multiple infection of animals with foot-and-mouth disease viruses'. In: *Proceedings of the Annual Meeting of the US Animal Health Association* 75.1 (1972), pp. 441–465 (cit. on p. 153).

[306] Daniel Dorey-Robinson, Giuseppe Maccari, Richard Borne and John A. Hammond. 'IgMAT: Immunoglobulin sequence Multi-species Annotation Tool for any species including those with incomplete antibody annotation or unusual characteristics'. In: *bioRxiv* (2021), p. 2021.2009.2022.461368. DOI: 10.1101/2021.09.22.461368 (cit. on p. 161).

[307] Robert C. Edgar. 'Search and clustering orders of magnitude faster than BLAST'. In: *Bioinformatics* 26.19 (2010), pp. 2460–2461. DOI: 10.1093/bioinformatics/btq461 (cit. on p. 161).

[308] Kun Li, Sheng Wang, Yimei Cao, Huifang Bao, Pinghua Li, Pu Sun, Xingwen Bai, Yuanfang Fu, Xueqing Ma, Jing Zhang, Dong Li, Yingli Chen, Xuerong Liu, Fanglan An, Faju Wu, Zengjun Lu and Zaixin Liu. 'Development of Foot-and-Mouth Disease Virus-Neutralizing Monoclonal Antibodies Derived From Plasmablasts of Infected Cattle and Their Germline Gene Usage'. In: *Frontiers in Immunology* 10 (2019), p. 2870. DOI: 10.3389/fimmu.2019.02870 (cit. on p. 164).

[309] Hao-Bo Guo, Alexander Perminov, Selemon Bekele, Gary Kedziora, Sanaz Farajollahi, Vanessa Varaljay, Kevin Hinkle, Valeria Molinero, Konrad Meister, Chia Hung, Patrick Dennis, Nancy Kelley-Loughnane and Rajiv Berry. 'AlphaFold2 models indicate that protein sequence determines both structure and dynamics'. In: *Scientific Reports* 12.1 (2022), p. 10696. DOI: 10.1038/s41598-022-14382-9 (cit. on p. 178).

[310] Helen M. E. Duyvesteyn. 'Integrative structural biology of viruses'. PhD thesis. University of Oxford, 2019, p. 251 (cit. on p. 216).

[311] Rangana Warshamanage, Keitaro Yamashita and Garib . Murshudov. 'EMDA: A Python package for Electron Microscopy Data Analysis'. In: *Journal of Structural Biology* 214.1 (2022), p. 107826. DOI: 10.1016/j.jsb.2021.107826 (cit. on p. 216).

[312] Derek Logan, Robin Abu-Ghazaleh, Wendy E. Blakemore, Stephen Curry, Terry Jackson, Andrew M. Q. King, Susan M. Lea, Richard Lewis, John W. I. Newman, Nigel R. Parry, David J. Rowlands, David I. Stuart and Elizabeth E. Fry. 'Structure of a major immunogenic site on foot-and-mouth disease virus'. In: *Nature* 362.6420 (1993), pp. 566–568. DOI: 10.1038/362566a0 (cit. on p. 218).

[313] Marcela Torres, Narcis Fernaéndez-Fuentes, Andraés Fiser and Arturo Casadevall. 'The Immunoglobulin Heavy Chain Constant Region Affects Kinetic and Thermodynamic Parameters of Antibody Variable Region Interactions with Antigen'. In: *Journal of Biological Chemistry* 282.18 (2007), pp. 13917–13927. DOI: 10.1074/jbc.M700661200 (cit. on p. 232).

[314] Gary R. McLean, Marcela Torres, Natalia Elguezabal, Antonio Nakouzi and Arturo Casadevall. 'Isotype Can Affect the Fine Specificity of an Antibody for a Polysaccharide Antigen'. In: *The Journal of Immunology* 169.3 (2002), pp. 1379–1386. DOI: 10.4049/jimmunol.169.3.1379 (cit. on p. 232).

[315] L. J. Cooper, a R. Shikhman, D. D. Glass, D. Kangisser, M. W. Cunningham and N. S. Greenspan. 'Role of heavy chain constant domains in antibody-antigen interaction. Apparent specificity differences among streptococcal IgG antibodies expressing identical variable domains.' In: *Journal of immunology (Baltimore, Md. : 1950)* 150.6 (1993), pp. 2231–2242 (cit. on p. 232).

[316] K. Schoppel, E. Haßfurther, W. Britt, M. Ohlin, C. A. K. Borrebaeck and M. Mach. 'Antibodies Specific for the Antigenic Domain 1 of Glycoprotein B (gpUL55) of Human Cytomegalovirus Bind to Different Substructures'. In: *Virology* 216.1 (1996), pp. 133–145. DOI: 10.1006/viro.1996.0040 (cit. on p. 232).

[317] John-Erik Stig Hansen, Anne Marie Sørensen, Maiken Arendrup, Sigvard Olofsson, Jens Ole Nielsen, Evelyne Janzek, Claus Nielsen and Hans Loibner. 'Enhancement of retroviral infection in vitro by anti-Le y IgG: reversal by humanization of monoclonal mouse antibody'. In: *APMIS* 101.7-12 (1993), pp. 711–718. DOI: 10.1111/j.1699-0463.1993.tb00170.x (cit. on p. 232).

[318] Otto Pritsch, Gilbert Hudry-Clergeon, Malcolm Buckle, Yves Pétillot, Jean P. Bouvet, Jean Gagnon and Guillaume Dighiero. 'Can immunoglobulin C(H)1 constant region domain modulate antigen binding affinity of antibodies?' In: *Journal of Clinical Investigation* 98.10 (1996), pp. 2235–2243. DOI: 10.1172/JCI119033 (cit. on p. 232).

[319] Behnaz Heydarchi, Rob J. Center, Jonathan Bebbington, Jack Cuthbertson, Christopher Gonelli, Georges Khoury, Charlene Mackenzie, Marit Lichtfuss, Grant Rawlin, Brian Muller and Damian F. J. Purcell. 'Trimeric gp120-specific bovine monoclonal antibodies require cysteine and aromatic residues in CDRH3 for high affinity binding to HIV Env'. In: 9.3 (2017), pp. 550–566. DOI: 10.1080/19420862.2016.1270491 (cit. on pp. 232, 291).

[320] Hui Ma, Ciarán Ó'Fágáin and Richard O'Kennedy. 'Antibody stability: A key to performance - Analysis, influences and improvement'. In: *Biochimie* 177 (2020), pp. 213–225. DOI: 10.1016/j.biochi.2020.08.019 (cit. on p. 233).

References for Chapter 5

[160] Jingshan Ren, Joanne E. Nettleship, Gemma Harris, William N. Mwangi, Nahid Rhaman, Clare F. J. Grant, Abhay Kotecha, Elizabeth E. Fry, Bryan Charleston, David I. Stuart, John A. Hammond and Raymond J. Owens. 'The role of the light chain in the structure and binding activity of two cattle antibodies that neutralize bovine respiratory syncytial virus'. In: *Molecular Immunology* 112 (2019), pp. 123–130. DOI: 10.1016/j.molimm.2019.04.026 (cit. on pp. 25, 26, 33, 92–94, 99, 105, 119, 128, 146, 164, 237, 240–243, 253, 255–258, 260, 282, 283, 285–287, 290, 295).

[161] H. E. Kennedy, B. V. Jones, E. M. Tucker, N. J. Ford, S. W. Clarke, Julie Furze, Lewis H. Thomas and E. J. Stott. 'Production and Characterization of Bovine Monoclonal Antibodies to Respiratory Syncytial Virus'. In: *Journal of General Virology* 69.12 (1988), pp. 3023–3032. DOI: 10.1099/0022-1317-69-12-3023 (cit. on pp. 25, 93, 240).

[175] Baoshan Zhang, Lei Chen, Chiara Silacci, Michelle Thom, Jeffrey C. Boyington, Aliaksandr Druz, M. Gordon Joyce, Efrain Guzman, Wing-Pui Kong, Yen-Ting Lai, Guillaume B. E. Stewart-Jones, Yaroslav Tsybovsky, Yongping Yang, Tongqing Zhou, Ulrich Baxa, John R. Mascola, Davide Corti, Antonio Lanzavecchia, Geraldine Taylor and Peter D. Kwong. 'Protection of calves by a prefusion-stabilized bovine RSV F vaccine'. In: *npj Vaccines* 2.7

(2017), pp. 1–11. DOI: 10.1038/s41541-017-0005-9 (cit. on pp. 34, 239, 240, 245, 251, 282, 283, 285, 286, 295).

[180] Swiss Institute of Bioinformatics. *ExPASy ProtParam*. URL: https://web.expasy.org/protparam/ (visited on 15th Nov. 2022) (cit. on pp. 55, 249, 273, 305, 306).

[187] Ramneek Gupta and Søren Brunak. 'Prediction of glycosylation across the human proteome and the correlation to protein function'. In: *Pacific Symposium on Biocomputing* 7 (2002), pp. 310–22. DOI: 10.1142/9789812799623_0029 (cit. on pp. 60, 251, 252, 305).

[188] Catharina Steentoft, Sergey Y. Vakhrushev, Hiren J. Joshi, Yun Kong, Malene B. Vester-Christensen, Katrine T-B. G. Schjoldager, Kirstine Lavrsen, Sally Dabelsteen, Nis B. Pedersen, Lara Marcos-Silva, Ramneek Gupta, Eric Paul Bennett, Ulla Mandel, Søren Brunak, Hans H. Wandall, Steven B. Levery and Henrik Clausen. 'Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology'. In: *The EMBO Journal* 32.10 (2013), pp. 1478–1488. DOI: 10.1038/emboj.2013.79 (cit. on pp. 60, 251, 305).

[189] Technical University of Denmark Department of Health Technology. *NetNGlyc*. URL: https://services.healthtech.dtu.dk/service.php?NetNGlyc-1.0 (visited on 29th Oct. 2022) (cit. on pp. 61, 251, 252, 305).

[190] Technical University of Denmark Department of Health Technology. *NetOGlyc*. URL: https://services.healthtech.dtu.dk/service.php?NetOGlyc-4.0 (visited on 29th Oct. 2022) (cit. on pp. 61, 251, 305).

[191] Peter Schuck. 'Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling'. In: *Biophysical Journal* 78.3 (2000), pp. 1606–1619. DOI: 10.1016/S0006-3495(00)76713-0 (cit. on pp. 62, 255, 257).

[192] Sumit K. Chaturvedi, Jia Ma, Patrick H. Brown, Huaying Zhao and Peter Schuck. 'Measuring macromolecular size distributions and interactions at high concentrations by sedimentation velocity'. In: *Nature Communications* 9.1 (2018), p. 4415. DOI: 10.1038/s41467-018-06902-x (cit. on pp. 62, 255, 257).

[243] Giulia Weissenberger, Rene J. M. Henderikx and Peter J. Peters. 'Understanding the invisible hands of sample preparation for cryo-EM'. In: *Nature Methods* 18.5 (2021), pp. 463–471. DOI: 10.1038/s41592-021-01130-6 (cit. on pp. 77, 263, 266).

[257] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng and Thomas E. Ferrin. 'UCSF Chimera - A visualization system for exploratory research and analysis'. In: *Journal of Computational Chemistry* 25.13 (2004), pp. 1605–1612. DOI: 10.1002/jcc.20084 (cit. on pp. 85, 87, 88, 107, 122, 130, 185, 194, 196, 201, 209, 211, 215, 218, 221, 222, 282, 285).

[321] A. J. Masot, C. L. Kelling, O. López, J. H. Sur and E. Redondo. 'In Situ Hybridization Detection of Bovine Respiratory Syncytial Virus in the Lung of Experimentally Infected Lambs'. In: *Veterinary Pathology* 37.6 (2000), pp. 618–625. DOI: 10.1354/vp.37-6-618 (cit. on p. 237).

[322] Steven T. Grubbs, Stephen A. Kania and L. N. D. Potgieter. 'Prevalence of Ovine and Bovine Respiratory Syncytial Virus Infections in Cattle Determined with a Synthetic Peptide-Based Immunoassay'. In: *Journal of Veterinary Diagnostic Investigation* 13.2 (2001), pp. 128–132. DOI: 10.1177/104063870101300206 (cit. on p. 237).

[323] M. F. Paccaud and C. Jacquier. 'A Respiratory syncytial virus of bovine origin'. In: *Archiv für die gesamte Virusforschung* 30.4 (1970), pp. 327–342. DOI: 10.1007/BF01258363 (cit. on p. 237).

[324] Geraldine Taylor. 'Bovine Model of Respiratory Syncytial Virus Infection'. In: *Challenges and Opportunities for Respiratory Syncytial Virus Vaccines*. Ed. by Larry J. Anderson, Barney S. Graham and Richard W. Compans. New York: Springer, 2013. Chap. Animal and, pp. 327–345. DOI: 10.1007/978-3-642-38919-1_16 (cit. on pp. 237, 238).

[325] Douglas C. Hodgins, Jennifer A. Conlon and Patricia E. Shewen. 'Respiratory Viruses and Bacteria in Cattle'. In: *Polymicrobial Diseases*. Ed. by Kim A. Brogden and Janet M. Guthmiller. Washington D.C.: ASM Press, 2002. Chap. 12, pp. 213–229 (cit. on p. 237).

[326] Rosa Sarmiento-Silva, Yuko Nakamura-Lopez and Gilberto Vaughan. 'Epidemiology, Molecular Epidemiology and Evolution of Bovine Respiratory Syncytial Virus'. In: *Viruses* 4.12 (2012), pp. 3452–3467. DOI: 10.3390/v4123452 (cit. on p. 238).

[327] T. R. Ames. 'The Epidemiology of BRSV infection'. In: *Veterinary Medicine* 88.9 (1993), pp. 881–885 (cit. on p. 238).

[328] Robert A. Smith. 'Effects of Feedlot Disease on Economics, Production and Carcass Value'. In: *American Association of Bovine Practitioners* 33 (2000), pp. 125–128 (cit. on p. 238).

[329] R. E. Sacco, J. L. McGill, A. E. Pillatzki, M. V. Palmer and M. R. Ackermann. 'Respiratory Syncytial Virus Infection in Cattle'. In: *Veterinary Pathology* 51.2 (2014), pp. 427–436. DOI: 10.1177/0300985813501341 (cit. on p. 238).

[330] W. H. M. Van der Poel, A. Brand, J. A. Kramps and J. T. Van Oirschot. 'Respiratory syncytial virus infections in human beings and in cattle'. In: *Journal of Infection* 29.2 (1994), pp. 215–228. DOI: 10.1016/S0163-4453(94)90866-4 (cit. on p. 238).

[331] Tiffany Turner, Benjamin Kopp, Grace Paul, Don Hayes Jr, Rohan Thompson and Lindsay Landgrave. 'Respiratory syncytial virus: current and emerging treatment options'. In: *ClinicoEconomics and Outcomes Research* 6.1 (2014), p. 217. DOI: 10.2147/CEOR.S60710 (cit. on p. 238).

[332] Gerard M. Murray, Rónan G. O'Neill, Simon J. More, Máire C. McElroy, Bernadette Earley and Joseph P. Cassidy. 'Evolving views on bovine respiratory disease: An appraisal of selected control measures – Part 2'. In: *The Veterinary Journal* 217 (2016), pp. 78–82. DOI: 10.1016/j.tvjl.2016.09.013 (cit. on p. 238).

[333] Robert A. Lamb and G. D. Parks. 'Paramyxoviridae: The Viruses and their Replication'. In: *Fields Virology Vol. 2*. Ed. by D. M. Knipe and P. M. Howley. 5th ed. Philadelphia, USA: Lippincott Williams and Williams, 2007, pp. 1449–1496 (cit. on pp. 238, 239).

[334] T. G. Kimman, J. Sol, F. Westenbrink and P. J. Strayer. 'A severe outbreak of respiratory tract disease associated with bovine respiratory syncytial virus probably enhanced by vaccination with modified live vaccine'. In: *Veterinary Quarterly* 11.4 (1989), pp. 250–253. DOI: 10.1080/01652176.1989.9694231 (cit. on p. 239).

[335] P. Schreiber, J. P. Matheise, F. Dessy, M. Heimann, J. J. Letesson, P. Coppe and A. Collard. 'High Mortality Rate Associated with Bovine Respiratory Syncytial Virus (BRSV) Infection in Belgian White Blue Calves Previously Vaccinated with an Inactivated BRSV Vaccine'. In: *Journal of Veterinary Medicine Series B* 47.7 (2000), pp. 535–550. DOI: 10.1046/j.1439-0450.2000.00380.x (cit. on p. 239).

[336] Adriaan F. G. Antonis, Remco S. Schrijver, Franz Daus, Paul J. G. M. Steverink, Norbert Stockhofe, Evert J. Hensen, Johannes P. M. Langedijk and Robbert G. van der Most. 'Vaccine-Induced Immunopathology during Bovine Respiratory Syncytial Virus Infection: Exploring the Parameters of Pathogenesis'. In: *Journal of Virology* 77.22 (2003), pp. 12067–12073. DOI: 10.1128/JVI.77.22.12067-12073.2003 (cit. on p. 239).

[337] Heather M. Costello, William C. Ray, Supranee Chaiwatpongsakorn and Mark E. Peeples. 'Targeting RSV with Vaccines and Small Molecule Drugs'. In: *Infectious Disorders - Drug Targets* 12.2 (2012), pp. 110–128. DOI: 10.2174/187152612800100143 (cit. on p. 239).

[338] Sheng-Jiun Wu, Albert Schmidt, Eric J. Beil, Nicole D. Day, Patrick J. Branigan, Changbao Liu, Lester L. Gutshall, Concepción Palomo, Julie Furze, Geraldine Taylor, José A. Melero, Ping Tsui, Alfred M. Del Vecchio and Marian Kruszynski. 'Characterization of the epitope for anti-human respiratory syncytial virus F protein monoclonal antibody 101F using synthetic peptides and genetic approaches'. In: *Journal of General Virology* 88.10 (2007), pp. 2719–2723. DOI: 10.1099/vir.0.82753-0 (cit. on p. 239).

[339] IMpact-RSV Study Group. 'Palivizumab, a Humanized Respiratory Syncytial Virus Monoclonal Antibody, Reduces Hospitalization From Respiratory Syncytial Virus Infection in High-risk Infants'. In: *Pediatrics* 102.3 (1998), pp. 531–537. DOI: 10.1542/peds.102.3.531 (cit. on p. 239).

[340] Robert A. Lerch, E. James Stott and Gail W. Wertz. 'Characterization of bovine respiratory syncytial virus proteins and mRNAs and generation of cDNA clones to the viral mRNAs'. In: *Journal of Virology* 63.2 (1989), pp. 833–840. DOI: 10.1128/jvi.63.2.833-840.1989 (cit. on p. 239).

[341] John C. Baker, E. Greg Wilson, George L. McKay, Ronald J. Stanek, Wendy J. Underwood, Leland F. Velicer and Maurice A. Mufson. 'Identification of subgroups of bovine respiratory syncytial virus'. In: *Journal of Clinical Microbiology* 30.5 (1992), pp. 1120–1126. DOI: 10.1128/jcm.30.5.1120-1126.1992 (cit. on p. 239).

[342] Julie Furze, Gail W. Wertz, Robert A. Lerch and Geraldine Taylor. 'Antigenic heterogeneity of the attachment protein of bovine respiratory syncytial virus'. In: *Journal of General Virology* 75.2 (1994), pp. 363–370. DOI: 10.1099/0022-1317-75-2-363 (cit. on p. 239).

[343] Robert A. Lerch, Kevin Anderson and Gail W. Wertz. 'Nucleotide sequence analysis and expression from recombinant vectors demonstrate that the attachment protein G of bovine respiratory syncytial virus is distinct from that of human respiratory syncytial virus'. In: *Journal of Virology* 64.11 (1990), pp. 5559–5569. DOI: 10.1128/jvi.64.11.5559-5569.1990 (cit. on p. 239).

[344] Jason S. McLellan, William C. Ray and Mark E. Peeples. 'Structure and function of respiratory syncytial virus surface glycoproteins'. In: *Current Topics in Microbiology and Immunology* 372 (2013), pp. 83–104. DOI: 10.1007/978-3-642-38919-1_4 (cit. on p. 239).

[345] Joan O. Ngwuta, Man Chen, Kayvon Modjarrad, M. Gordon Joyce, Masaru Kanekiyo, Azad Kumar, Hadi M. Yassine, Syed M. Moin, April M. Killikelly, Gwo-Yu Chuang, Aliaksandr Druz, Ivelin S. Georgiev, Emily J. Rundlet, Mallika Sastry, Guillaume B. E. Stewart-Jones, Yongping Yang, Baoshan Zhang, Martha C. Nason, Cristina Capella, Mark E. Peeples, Julie E. Ledgerwood, Jason S. McLellan, Peter D. Kwong and Barney S. Graham. 'Prefusion F–specific antibodies determine the magnitude of RSV neutralizing activity in human sera'. In: *Science Translational Medicine* 7.309 (2015), pp. 1–19. DOI: 10.1126/scitranslmed.aac4241 (cit. on p. 239).

[346] Margarita Magro, Vicente Mas, Keith Chappell, Mónica Vázquez, Olga Cano, Daniel Luque, María C. Terrón, José A. Melero and Concepción Palomo. 'Neutralizing antibodies against the preactive form of respiratory syncytial virus fusion protein offer unique possibilities for clinical intervention'. In: *Proceedings of the National Academy of Sciences* 109.8 (2012), pp. 3089–3094. DOI: 10.1073/pnas.1115941109 (cit. on p. 239).

[347] Gert Bolt, Lars Østergaard Pedersen and Helle Harder Birkeslund. 'Cleavage of the respiratory syncytial virus fusion protein is required for its surface expression: role of furin'. In: *Virus Research* 68.1 (2000), pp. 25–33. DOI: 10.1016/S0168-1702(00)00149-0 (cit. on p. 239).

[348] Peter L. Collins and Geneviève Mottet. 'Post-translational Processing and Oligomerization of the Fusion Glycoprotein of Human Respiratory Syncytial Virus'. In: *Journal of General Virology* 72.12 (1991), pp. 3095–3101. DOI: 10.1099/0022-1317-72-12-3095 (cit. on p. 239).

[349] Luis González-Reyes, M. Begoña Ruiz-Argüello, Blanca García-Barreno, Leslie Calder, Juan A. López, Juan P. Albar, John J. Skehel, Don C. Wiley and José A. Melero. 'Cleavage of the human respiratory syncytial virus fusion protein at two distinct sites is required for activation of membrane fusion'. In: *Proceedings of the National Academy of Sciences* 98.17 (2001), pp. 9859–9864. DOI: 10.1073/pnas.151098198 (cit. on p. 239).

[350] Gert Zimmer, Linda Budz and Georg Herrler. 'Proteolytic Activation of Respiratory Syncytial Virus Fusion Protein'. In: *Journal of Biological Chemistry* 276.34 (2001), pp. 31642–31650. DOI: 10.1074/jbc.M102633200 (cit. on p. 239).

[351] C. Gruber and S. Levine. 'Respiratory Syncytial Virus Polypeptides. III. The Envelope-associated Proteins'. In: *Journal of General Virology* 64.4 (1983), pp. 825–832. DOI: 10.1099/0022-1317-64-4-825 (cit. on p. 239).

[352] Nicole D. Day, Patrick J. Branigan, Changbao Liu, Lester L. Gutshall, Jianquan Luo, José A. Melero, Robert T. Sarisky and Alfred M. Del Vecchio. 'Contribution of cysteine residues in the extracellular domain of the F protein of human respiratory syncytial virus to its function'. In: *Virology Journal* 3.1 (2006), p. 34. DOI: 10.1186/1743-422X-3-34 (cit. on p. 239).

[353] Jason S. McLellan, Man Chen, Sherman Leung, Kevin W. Graepel, Xiulian Du, Yongping Yang, Tongqing Zhou, Ulrich Baxa, Etsuko Yasuda, Tim Beaumont, Azad Kumar, Kayvon Modjarrad, Zizheng Zheng, Min Zhao, Ningshao Xia, Peter D. Kwong and Barney S. Graham. 'Structure of RSV Fusion Glycoprotein Trimer Bound to a Prefusion-Specific Neutralizing Antibody'. In: *Science* 340.6136 (2013), pp. 1113–1117. DOI: 10.1126/science.1234914 (cit. on pp. 239, 240).

[354] Xun Zhao, Mona Singh, Vladimir N. Malashkevich and Peter S. Kim. 'Structural characterization of the human respiratory syncytial virus fusion protein core'. In: *Proceedings of the National Academy of Sciences* 97.26 (2000), pp. 14172–14177. DOI: 10.1073/pnas.260499197 (cit. on p. 240).

[355] Jason S. McLellan, Man Chen, M. Gordon Joyce, Mallika Sastry, Guillaume B. E. Stewart-Jones, Yongping Yang, Baoshan Zhang, Lei Chen, Sanjay Srivatsan, Anqi Zheng, Tongqing Zhou, Kevin W. Graepel, Azad Kumar, Syed Moin, Jeffrey C. Boyington, Gwo-Yu Chuang, Cinque Soto, Ulrich Baxa, Arjen Q. Bakker, Hergen Spits, Tim Beaumont, Zizheng Zheng, Ningshao Xia, Sung-Youl Ko, John-Paul Todd, Srinivas Rao, Barney S. Graham and Peter D. Kwong. 'Structure-Based Design of a Fusion Glycoprotein Vaccine for Respiratory Syncytial Virus'. In: *Science* 342.6158 (2013), pp. 592–598. DOI: 10.1126/science.1243283 (cit. on p. 240).

[356] J. Arbiza, Geraldine Taylor, J. A. Lopez, J. Furze, S. Wyld, P. Whyte, E. J. Stott, G. Wertz, W. Sullender, M. Trudel and J. A. Melero. 'Characterization of two antigenic sites recognized by neutralizing monoclonal antibodies directed against the fusion glycoprotein of human respiratory syncytial virus'. In: *Journal of General Virology* 73.9 (1992), pp. 2225–2234. DOI: 10.1099/0022-1317-73-9-2225 (cit. on pp. 240, 242, 260, 262).

[357] Geraldine Taylor, E. J. Stott, J. Furze, J. Ford and Paul Sopp. 'Protective epitopes on the fusion protein of respiratory syncytial virus recognized by murine and bovine monoclonal antibodies'. In: *Journal of General Virology* 73.9 (1992), pp. 2217–2223. DOI: 10.1099/0022-1317-73-9-2217 (cit. on pp. 240, 242, 260, 262).

[358] Morgan S. A. Gilman, Carlos A. Castellanos, Man Chen, Joan O. Ngwuta, Eileen Goodwin, Syed M. Moin, Vicente Mas, José A. Melero, Peter F. Wright, Barney S. Graham, Jason S. McLellan and Laura M. Walker. 'Rapid profiling of RSV antibody repertoires from the memory B cells of naturally infected adult donors'. In: *Science Immunology* 1.6 (2016), pp. 1–12. DOI: 10.1126/sciimmunol.aaj1879 (cit. on pp. 240, 282).

[359] Paul Whyte. 'Identification and characterisation of protective B cell epitopes on the fusion protein of respiratory syncytial virus.' PhD thesis. University of Reading, 1994 (cit. on pp. 240, 242, 260, 262).

[360] M. Gordon Joyce, Baoshan Zhang, Li Ou, Man Chen, Gwo-Yu Chuang, Aliaksandr Druz, Wing-Pui Kong, Yen-Ting Lai, Emily J. Rundlet, Yaroslav Tsybovsky, Yongping Yang, Ivelin S. Georgiev, Miklos Guttman, Christopher R. Lees, Marie Pancera, Mallika Sastry, Cinque Soto, Guillaume B. E. Stewart-Jones, Paul V. Thomas, Joseph G. Van Galen, Ulrich Baxa, Kelly K. Lee, John R. Mascola, Barney S. Graham and Peter D. Kwong. 'Iterative structure-based improvement of a fusion-glycoprotein vaccine against RSV'. In: *Nature Structural and Molecular Biology* 23.9 (2016), pp. 811–820. DOI: 10.1038/nsmb.3267 (cit. on pp. 249, 256).

[361] Morgan S. A. Gilman, Polina Furmanova-Hollenstein, Gabriel Pascual, Angélique B. van 't Wout, Johannes P. M. Langedijk and Jason S. McLellan. 'Transient opening of trimeric prefusion RSV F proteins'. In: *Nature Communications* 10.1 (2019), p. 2105. DOI: 10.1038/s41467-019-09807-5 (cit. on pp. 262, 286).

[362] James B. Munro, Jason Gorman, Xiaochu Ma, Zhou Zhou, James Arthos, Dennis R. Burton, Wayne C. Koff, Joel R. Courter, Amos B. Smith, Peter D. Kwong, Scott C. Blanchard and Walther Mothes. 'Conformational dynamics of single HIV-1 envelope trimers on the surface of native virions'. In: *Science* 346.6210 (2014), pp. 759–763. DOI: 10.1126/science.1254426 (cit. on p. 286).

References for Chapter 6

[30] Eduard O. Roos, Marie Bonnet-Di Placido, William N. Mwangi, Katy Moffat, Lindsay M. Fry, Ryan Waters and John A. Hammond. 'OMIP-085: Cattle B-cell phenotyping by an 8-color panel'. In: *Cytometry* (2022). DOI: 10.1002/cyto.a.24683 (cit. on pp. 6, 293).

[31] Surinder S. Saini, Brian Allore, Robert M. Jacobs and Azad K. Kaushik. 'Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM antibodies'. In: *European Journal of Immunology* 29.8 (1999), pp. 2420–2426. DOI: 10.1002/(SICI)1521-4141(199908)29:08<2420::AID-IMMU2420>3.0.CO;2-A (cit. on pp. 6, 10, 12–14, 18, 23, 92, 98, 140, 289, 293).

[32] Yfke Pasman, Surinder S. Saini, Elspeth Smith and Azad K. Kaushik. 'Organization and genomic complexity of bovine λ-light chain gene locus'. In: *Veterinary Immunology and Immunopathology* 135.3-4 (2010), pp. 306–313. DOI: 10.1016/j.vetimm.2009.12.012 (cit. on pp. 6, 20, 289).

[34] Li Ma, Tong Qin, Dan Chu, Xueqian Cheng, Jing Wang, Xifeng Wang, Peng Wang, Haitang Han, Liming Ren, Robert Aitken, Lennart Hammarström, Ning Li and Yaofeng Zhao. 'Internal Duplications of DH, JH, and C Region Genes Create an Unusual IgH Gene Locus in Cattle'. In: *The Journal of Immunology* 196.10 (2016), pp. 4358–4366. DOI: 10.4049/jimmunol.1600158 (cit. on pp. 7–11, 13, 96, 289).

[37] M. C. Sinclair, J. Gilchrist and Robert Aitken. 'Bovine IgG repertoire is dominated by a single diversified VH gene family'. In: *The Journal of Immunology* 159.8 (1997), pp. 3883–3889 (cit. on pp. 7, 8, 11, 12, 289).

[38] Mikael Niku, Jenni Liljavirta, Keith Durkin, Eero Schroderus and Antti Iivanainen. 'The bovine genomic DNA sequence data reveal three IGHV subgroups, only one of which is functionally expressed'. In: *Developmental and Comparative Immunology* 37.3-4 (2012), pp. 457–461. DOI: 10.1016/j.dci.2012.02.006 (cit. on pp. 7, 11, 289).

[39] Shawn J. Berens, Dwane E. Wylie and Osvaldo J. Lopez. 'Use of a single V(H) family and long CDR3s in the variable region of cattle Ig heavy chains'. In: *International Immunology* 9.1 (1997), pp. 189–199. DOI: 10.1093/intimm/9.1.189 (cit. on pp. 7, 12, 14, 289).

[51] C. L. Parng, Susan Hansal, Richard A Goldsby and Barbara A. Osborne. 'Gene conversion contributes to Ig light chain diversity in cattle.' In: *The Journal of Immunology* 157.12 (1996), pp. 5478–5486 (cit. on pp. 8, 11, 12, 22, 289).

[53] Brevin A. Smider and Vaughn V. Smider. 'Formation of ultralong DH regions through genomic rearrangement'. In: *BMC Immunology* 21.1 (2020), p. 30. DOI: 10.1186/s12865-020-00359-8 (cit. on pp. 9, 11, 13, 289).

[54] Thaddeus C. Deiss, Melissa L. Vadnais, Feng Wang, Patricia L. Chen, Ali Torkamani, Waithaka Mwangi, Marie-Paule Lefranc, Michael F. Criscitiello and Vaughn V. Smider. 'Immunogenetic factors driving formation of ultralong VH CDR3 in Bos taurus antibodies'. In: *Cellular and Molecular Immunology* 16.1 (2019), pp. 53–64. DOI: 10.1038/cmi.2017.117 (cit. on pp. 10, 11, 13–16, 18, 289).

[58] Samuel O. Oyola, Sonal P. Henson, Benjamin Nzau, Elizabeth Kibwana and Vishvanath Nene. 'Access to ultra-long IgG CDRH3 bovine antibody sequences using short read sequencing technology'. In: *Molecular Immunology* 139 (2021), pp. 97–105. DOI: 10.1016/j.molimm.2021.08.017 (cit. on pp. 10, 15, 289).

[61] Surinder S. Saini, William Farrugia, Paul A. Ramsland and Azad K. Kaushik. 'Bovine IgM antibodies with exceptionally long complementarity-determining region 3 of the heavy chain share unique structural properties conferring restricted V H + Vλ pairings'. In: *International Immunology* 15.7 (2003), pp. 845–853. DOI: 10.1093/intimm/dxg083 (cit. on pp. 10, 13, 20, 23, 27, 92, 157, 163, 289).

[64] Osvaldo J. Lopez, Carlos Perez and Dwane E. Wylie. 'A single V(H) family and long CDR3s are the targets for hypermutation in bovine immunoglobulin heavy chains'. In: *Immunological Reviews* 162.1 (1998), pp. 55–66. DOI: 10.1111/j.1600-065X.1998.tb01429.x (cit. on pp. 11, 13, 14, 289).

[77] Madhuri Koti, Galina Kataeva and Azad K. Kaushik. 'Novel atypical nucleotide insertions specifically at VH-DH junction generate exceptionally long CDR3H in cattle antibodies'. In: *Molecular Immunology* 47.11-12 (2010), pp. 2119–2128. DOI: 10.1016/j.molimm.2010.02.014 (cit. on pp. 12–14, 289).

[78] Feng Wang, Damian C. Ekiert, Insha Ahmad, Wenli Yu, Yong Zhang, Omar A. Bazirgan, Ali Torkamani, Terje Raudsepp, Waithaka Mwangi, Michael F. Criscitiello, Ian A. Wilson, Peter G. Schultz and Vaughn V. Smider. 'Reshaping Antibody Diversity'. In: *Cell* 153.6 (2013), pp. 1379–1393. DOI: 10.1016/j.cell.2013.04.049 (cit. on pp. 12, 13, 15–19, 23–25, 94, 98, 99, 105, 108, 109, 116, 128, 131, 132, 140–142, 162, 163, 180, 181, 206, 232, 289–294).

[82] Robyn L. Stanfield, Ian A. Wilson and Vaughn V. Smider. 'Conservation and diversity in the ultralong third heavy-chain complementarity-determining region of bovine antibodies'. In: *Science Immunology* 1.1 (2016), aaf7962. DOI: 10.1126/sciimmunol.aaf7962 (cit. on pp. 13, 15, 16, 23–25, 94, 98, 108, 109, 116, 140, 141, 162, 163, 180–182, 289, 290, 293).

[83] Jinhui Dong, Jessica A. Finn, Peter A. Larsen, Timothy P. L. Smith and James E. Crowe. 'Structural Diversity of Ultralong CDRH3s in Seven Bovine Antibody Heavy Chains'. In: *Frontiers in Immunology* 10.March (2019), pp. 1–9. DOI: 10.3389/fimmu.2019.00558 (cit. on pp. 13, 15, 16, 19, 23–25, 94, 97, 98, 108, 109, 116, 140, 141, 162, 163, 180, 181, 289, 290, 293).

[84] Robyn L. Stanfield, Zachary T. Berndsen, Ruiqi Huang, Devin Sok, Gabrielle Warner, Jonathan L. Torres, Dennis R. Burton, Andrew B. Ward, Ian A. Wilson and Vaughn V. Smider. 'Structural basis of broad HIV neutralization by a vaccine-induced cow antibody'. In: *Science Advances* 6.22 (2020), eaba0468. DOI: 10.1126/sciadv.aba0468 (cit. on pp. 13, 15, 16, 18, 23–25, 92, 94, 140, 141, 162, 163, 206, 232, 291, 293, 294).

[85] Yong He, Kun Li, Yimei Cao, Zixian Sun, Pinghua Li, Huifang Bao, Sheng Wang, Guoqiang Zhu, Xingwen Bai, Pu Sun, Xuerong Liu, Cheng Yang, Zaixin Liu, Zengjun Lu, Zihe Rao and Zhiyong Lou. 'Structures of Foot-and-mouth Disease Virus with neutralizing antibodies derived from recovered natural host reveal a mechanism for cross-serotype neutralization'. In: *PLOS Pathogens* 17.4 (2021). Ed. by Richard J. Kuhn, e1009507. DOI: 10.1371/journal.ppat.1009507 (cit. on pp. 13, 15, 16, 18, 23–25, 92, 94, 140, 141, 162, 163, 206, 232, 289–291, 293, 294).

[90] G. Lippens, J. Najib, Shoshana J. Wodak and A. Tartar. 'NMR Sequential Assignments and Solution Structure of Chlorotoxin, a Small Scorpion Toxin That Blocks Chloride Channels'. In: *Biochemistry* 34.1 (1995), pp. 13–21. DOI: 10.1021/bi00001a003 (cit. on pp. 17, 19, 94, 289, 291).

[91] Jamie I. Fletcher, Bogdan E. Chapman, Joel P. Mackay, Merlin E. H. Howden and Glenn F. King. 'The structure of versutoxin (δ-atracotoxin-Hv1) provides insights into the binding of site 3 neurotoxins to the voltage-gated sodium channel'. In: *Structure* 5.11 (1997), pp. 1525–1535. DOI: 10.1016/S0969-2126(97)00301-8 (cit. on pp. 17, 19, 94, 289, 291).

[94] Yuliya V. Korolkova, Eduard V. Bocharov, Kamilla Angelo, Innokenty V. Maslennikov, Olga V. Grinenko, Aleksey V. Lipkin, Elena D. Nosyreva, Kirill A. Pluzhnikov, Søren-peter Olesen, Alexander S. Arseniev and Eugene V. Grishin. 'New Binding Site on Common Molecular Scaffold Provides HERG Channel Specificity of Scorpion Toxin BeKm-1'. In: *Journal of Biological Chemistry* 277.45 (2002), pp. 43104–43109. DOI: 10.1074/jbc.M204083200 (cit. on pp. 17, 19, 94, 289, 291).

[95] Robert E. Oswald, Thomas M. Suchyna, Robert McFeeters, Philip Gottlieb and Frederick Sachs. 'Solution Structure of Peptide Toxins That Block Mechanosensitive Ion Channels'. In: *Journal of Biological Chemistry* 277.37 (2002), pp. 34443–34450. DOI: 10.1074/jbc.M202715200 (cit. on pp. 17, 19, 94, 289, 291).

[96] Katherine J. Nielsen, Michael Watson, David J. Adams, Anna K. Hammarström, Peter W. Gage, Justine M. Hill, David J. Craik, Linda Thomas, Denise Adams, Paul F. Alewood and Richard J. Lewis. 'Solution Structure of μ-Conotoxin PIIIA, a Preferential Inhibitor of Persistent Tetrodotoxin-sensitive Sodium Channels'. In: *Journal of Biological Chemistry* 277.30 (2002), pp. 27247–27255. DOI: 10.1074/jbc.M201611200 (cit. on pp. 17, 19, 94, 289, 291).

[97] David J. Adams, Amanda B. Smith, Christina I. Schroeder, Takahiro Yasuda and Richard J. Lewis. 'ω-Conotoxin CVID Inhibits a Pharmacologically Distinct Voltage-sensitive Calcium Channel Associated with Transmitter Release from Preganglionic Nerve Terminals'. In: *Journal of Biological Chemistry* 278.6 (2003), pp. 4057–4062. DOI: 10.1074/jbc.M209969200 (cit. on pp. 17, 19, 94, 289, 291).

[98] Jorgen Mould, Takahiro Yasuda, Christina I. Schroeder, Aaron M. Beedle, Clinton J. Doering, Gerald W. Zamponi, David J. Adams and Richard J. Lewis. 'The $\alpha 2\delta$ Auxiliary Subunit Reduces Affinity of ω-Conotoxins for Recombinant N-type (Cav2.2) Calcium Channels'. In: *Journal of Biological Chemistry* 279.33 (2004), pp. 34705–34714. DOI: 10.1074/jbc.M310848200 (cit. on pp. 17, 19, 94, 289, 291).

[99] Siddhartha P. Sarma, G. Senthil Kumar, S. Sudarslal, Prathima Iengar, P. Ramasamy, Sujit K. Sikdar, K. S. Krishnan and Padmanabhan Balaram. 'Solution Structure of δ-Am2766: A Highly Hydrophobic δ-Conotoxin fromConus amadis That Inhibits Inactivation of Neuronal Voltage-Gated Sodium Channels'. In: *Chemistry and Biodiversity* 2.4 (2005), pp. 535–556. DOI: 10.1002/cbdv.200590035 (cit. on pp. 17, 19, 94, 289, 291).

[101] Roger J. P. Dawson, Jörg Benz, Peter Stohler, Tim Tetaz, Catherine Joseph, Sylwia Huber, Georg Schmid, Daniela Hügin, Pascal Pflimlin, Gerd Trube, Markus G. Rudolph, Michael Hennig and Armin Ruf. 'Structure of the Acid-sensing ion channel 1 in complex with the gating modifier Psalmotoxin 1'. In: *Nature Communications* 3.1 (2012), p. 936. DOI: 10.1038/ncomms1917 (cit. on pp. 17, 19, 94, 289, 291).

[103] Brad R. Green, Joanna Gajewiak, Sandeep Chhabra, Jack J. Skalicky, Min-Min Zhang, Jean E. Rivier, Grzegorz Bulaj, Baldomero M. Olivera, Doju Yoshikami and Raymond S. Norton. 'Structural Basis for the Inhibition of Voltage-gated Sodium Channels by Conotoxin μO§-GVIIJ'. In: *Journal of Biological Chemistry* 291.13 (2016), pp. 7205–7220. DOI: 10.1074/jbc.M115.697672 (cit. on pp. 17, 19, 94, 289, 291).

[104] Bin Wu, Justin K. Murray, Kristin L. Andrews, Kelvin Sham, Jason Long, Jennifer Aral, Joseph Ligutti, Shanti Amagasu, Dong Liu, Anruo Zou, Xiaoshan Min, Zhulun Wang, Christopher P. Ilch, Thomas J. Kornecook, Min-Hwa Jasmine Lin, Xuhai Be, Les P. Miranda, Bryan D. Moyer and Kaustav Biswas. 'Discovery of Tarantula Venom-Derived Na V 1.7-Inhibitory JzTx-V Peptide 5-Br-Trp24 Analogue AM-6120 with Systemic Block of Histamine-Induced Pruritis'. In: *Journal of Medicinal Chemistry* 61.21 (2018), pp. 9500–9512. DOI: 10.1021/acs.jmedchem.8b00736 (cit. on pp. 17, 19, 94, 289, 291).

[132] Hristo L. Svilenov, Julia Sacherl, Ulrike Protzer, Martin Zacharias and Johannes Buchner. 'Mechanistic principles of an ultra-long bovine CDR reveal strategies for antibody design'. In: *Nature Communications* 12.1 (2021), p. 6737. DOI: 10.1038/s41467-021-27103-z (cit. on pp. 17, 19, 108, 140, 162, 206, 232, 291, 294).

[133] Robert Pejchal, Laura M. Walker, Robyn L. Stanfield, Sanjay K. Phogat, Wayne C. Koff, Pascal Poignard, Dennis R. Burton and Ian A. Wilson. 'Structure and function of broadly reactive antibody PG16 reveal an H3 subdomain that mediates potent neutralization of HIV-1'. In: *Proceedings of the National Academy of Sciences* 107.25 (2010), pp. 11483–11488. DOI: 10.1073/pnas.1004600107 (cit. on pp. 18, 293).

[134] Marie Pancera, Syed Shahzad-ul-Hussan, Nicole A. Doria-Rose, Jason S. McLellan, Robert T. Bailer, Kaifan Dai, Sandra Loesgen, Mark K. Louder, Ryan P. Staupe, Yongping Yang, Baoshan Zhang, Robert Parks, Joshua Eudailey, Krissey E. Lloyd, Julie Blinn, S. Munir Alam, Barton F. Haynes, Mohammed N. Amin, Lai-Xi Wang, Dennis R. Burton, Wayne C. Koff, Gary J. Nabel, John R. Mascola, Carole A. Bewley and Peter D. Kwong. 'Structural basis for diverse N-glycan recognition by HIV-1–neutralizing V1–V2–directed antibody PG16'. In: *Nature Structural and Molecular Biology* 20.7 (2013), pp. 804–813. DOI: 10.1038/nsmb.2600 (cit. on pp. 18, 293).

[135] Nicole A. Doria-Rose, Jinal N. Bhiman, Ryan S. Roark, Chaim A. Schramm, Jason Gorman, Gwo-Yu Chuang, Marie Pancera, Evan M. Cale, Michael J. Ernandes, Mark K. Louder, Mangaiarkarasi Asokan, Robert T. Bailer, Aliaksandr Druz, Isabella R. Fraschilla, Nigel J. Garrett, Marissa Jarosinski, Rebecca M. Lynch, Krisha McKee, Sijy O'Dell, Amarendra Pegu, Stephen D. Schmidt, Ryan P. Staupe, Matthew S. Sutton, Keyun Wang, Constantinos Kurt Wibmer, Barton F. Haynes, Salim Abdool-Karim, Lawrence Shapiro, Peter D. Kwong, Penny L. Moore, Lynn Morris and John R. Mascola. 'New Member of the V1V2-Directed CAP256-VRC26 Lineage That Shows Increased Breadth and Exceptional Potency'. In: *Journal of Virology* 90.1 (2016). Ed. by G. Silvestri, pp. 76–91. DOI: 10.1128/JVI.01791-15 (cit. on pp. 18, 293).

[136] Jason Gorman, Gwo-Yu Chuang, Yen-Ting Lai, Chen-Hsiang Shen, Jeffrey C. Boyington, Aliaksandr Druz, Hui Geng, Mark K. Louder, Krisha McKee, Reda Rawi, Raffaello Verardi, Yongping Yang, Baoshan Zhang, Nicole A. Doria-Rose, Bob Lin, Penny L. Moore, Lynn Morris, Lawrence Shapiro, John R. Mascola and Peter D. Kwong. 'Structure of Super-Potent Antibody CAP256-VRC26.25 in Complex with HIV-1 Envelope Reveals a Combined Mode of Trimer-Apex Recognition'. In: *Cell Reports* 31.1 (2020), p. 107488. DOI: 10.1016/j.celrep.2020.03.052 (cit. on pp. 18, 293, 297).

[137] Devin Sok, Khoa M. Le, Melissa L. Vadnais, Karen L. Saye-Francisco, Joseph G. Jardine, Jonathan L. Torres, Zachary T. Berndsen, Leopold Kong, Robyn L. Stanfield, Jennifer Ruiz, Alejandra Ramos, Chi-Hui Liang, Patricia L. Chen, Michael F. Criscitiello, Waithaka Mwangi, Ian A. Wilson, Andrew B. Ward, Vaughn V. Smider and Dennis R. Burton. 'Rapid elicitation of broadly neutralizing antibodies to HIV by immunization in cows'. In: *Nature* 548.1 (2017), pp. 108–111. DOI: 10.1038/nature23301 (cit. on pp. 18, 206, 293, 294).

[138] Alex Macpherson, Anthony Scott-Tucker, Anastasios Spiliotopoulos, Catherine Simpson, Justin Staniforth, Adam Hold, James R. Snowden, Leah Manning, Jean M. H. van den Elsen and Alastair D. G. Lawson. 'Isolation of antigen-specific, disulphide-rich knob domain peptides from bovine antibodies'. In: *PLOS Biology* 18.9 (2020). Ed. by David Nemazee, e3000821. DOI: 10.1371/journal.pbio.3000821 (cit. on pp. 19, 140, 206, 232, 289, 291, 294).

[139] Alex Macpherson, Maisem Laabei, Zainab Ahdash, Melissa A. Graewert, James R. Birtley, Monika-Sarah E. D. Schulze, Susan Crennell, Sarah A. Robinson, Ben Holmes, Vladas Oleinikovas, Per H. Nilsson, James R. Snowden, Victoria Ellis, Tom Eirik Mollnes, Charlotte M. Deane, Dmitri Svergun, Alastair D. G. Lawson and Jean M. H. van den Elsen. 'The allosteric modulation of complement C5 by knob domain peptides'. In: *eLife* 10 (2021), pp. 1–49. DOI: 10.7554/eLife.63586 (cit. on pp. 19, 26, 140, 206, 232, 289, 291, 294).

[152] Anna Ekman, Mikael Niku, Jenni Liljavirta and Antti Iivanainen. 'Bos taurus genome sequence reveals the assortment of immunoglobulin and surrogate light chain genes in domestic cattle'. In: *BMC Immunology* 10.22 (2009). DOI: 10.1186/1471-2172-10-22 (cit. on pp. 20, 289).

[153] M. C. Sinclair, J. Gilchrist and Robert Aitken. 'Molecular Characterization of Bovine Vλ Regions'. In: *The Journal of Immunology* 155.4 (1995), pp. 3068–3078 (cit. on pp. 20–23, 27, 289).

[160] Jingshan Ren, Joanne E. Nettleship, Gemma Harris, William N. Mwangi, Nahid Rhaman, Clare F. J. Grant, Abhay Kotecha, Elizabeth E. Fry, Bryan Charleston, David I. Stuart, John A. Hammond and Raymond J. Owens. 'The role of the light chain in the structure and binding activity of two cattle antibodies that neutralize bovine respiratory syncytial virus'. In: *Molecular Immunology* 112 (2019), pp. 123–130. DOI: 10.1016/j.molimm.2019.04.026 (cit. on pp. 25, 26, 33, 92–94, 99, 105, 119, 128, 146, 164, 237, 240–243, 253, 255–258, 260, 282, 283, 285–287, 290, 295).

[162] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov and Philip E. Bourne. 'The Protein Data Bank'. In: *Nucleic Acids Research* 28.1 (2000), pp. 235–242. DOI: 10.1093/nar/28.1.235 (cit. on pp. 26, 64, 72, 73, 93, 94, 105, 137, 171–173, 176, 229, 292).

[175] Baoshan Zhang, Lei Chen, Chiara Silacci, Michelle Thom, Jeffrey C. Boyington, Aliaksandr Druz, M. Gordon Joyce, Efrain Guzman, Wing-Pui Kong, Yen-Ting Lai, Guillaume B. E. Stewart-Jones, Yaroslav Tsybovsky, Yongping Yang, Tongqing Zhou, Ulrich Baxa, John R. Mascola, Davide Corti, Antonio Lanzavecchia, Geraldine Taylor and Peter D. Kwong. 'Protection of calves by a prefusion-stabilized bovine RSV F vaccine'. In: *npj Vaccines* 2.7 (2017), pp. 1–11. DOI: 10.1038/s41541-017-0005-9 (cit. on pp. 34, 239, 240, 245, 251, 282, 283, 285, 286, 295).

[267] Jared Sampson. *Elbow angle*. URL: https://pymolwiki.org/index.php/Elbow_angle (cit. on pp. 132, 292).

[319] Behnaz Heydarchi, Rob J. Center, Jonathan Bebbington, Jack Cuthbertson, Christopher Gonelli, Georges Khoury, Charlene Mackenzie, Marit Lichtfuss, Grant Rawlin, Brian Muller and Damian F. J. Purcell. 'Trimeric gp120-specific bovine monoclonal antibodies require cysteine and aromatic residues in CDRH3 for high affinity binding to HIV Env'. In: 9.3 (2017), pp. 550–566. DOI: 10.1080/19420862.2016.1270491 (cit. on pp. 232, 291).

References for Chapter A

[80] Bin Shi, Long Ma, Xiaoyan He, Xiaomei Wang, Peng Wang, Li Zhou and Xinsheng Yao. 'Comparative analysis of human and mouse immunoglobulin variable heavy regions from IMGT/LIGM-DB with IMGT/HighV-QUEST'. In: *Theoretical Biology and Medical Modelling* 11.1 (2014), p. 30. DOI: 10.1186/1742-4682-11-30 (cit. on pp. 12, 307).

[180] Swiss Institute of Bioinformatics. *ExPASy ProtParam*. URL: https://web.expasy.org/protparam/ (visited on 15th Nov. 2022) (cit. on pp. 55, 249, 273, 305, 306).

[187] Ramneek Gupta and Søren Brunak. 'Prediction of glycosylation across the human proteome and the correlation to protein function'. In: *Pacific Symposium on Biocomputing* 7 (2002), pp. 310–22. DOI: 10.1142/9789812799623_0029 (cit. on pp. 60, 251, 252, 305).

[188] Catharina Steentoft, Sergey Y. Vakhrushev, Hiren J. Joshi, Yun Kong, Malene B. Vester-Christensen, Katrine T-B. G. Schjoldager, Kirstine Lavrsen, Sally Dabelsteen, Nis B. Pedersen, Lara Marcos-Silva, Ramneek Gupta, Eric Paul Bennett, Ulla Mandel, Søren Brunak, Hans H. Wandall, Steven B. Levery and Henrik Clausen. 'Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology'. In: *The EMBO Journal* 32.10 (2013), pp. 1478–1488. DOI: 10.1038/emboj.2013.79 (cit. on pp. 60, 251, 305).

[189] Technical University of Denmark Department of Health Technology. *NetNGlyc*. URL: https://services.healthtech.dtu.dk/service.php?NetNGlyc-1.0 (visited on 29th Oct. 2022) (cit. on pp. 61, 251, 252, 305).

[190] Technical University of Denmark Department of Health Technology. *NetOGlyc*. URL: https://services.healthtech.dtu.dk/service.php?NetOGlyc-4.0 (visited on 29th Oct. 2022) (cit. on pp. 61, 251, 305).

[202] Marie-Paule Lefranc, Christelle Pommié, Manuel Ruiz, Véronique Giudicelli, Elodie Foulquier, Lisa Truong, Valérie Thouvenin-Contet and Gérard Lefranc. 'IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains'. In: *Developmental and Comparative Immunology* 27.1 (2003), pp. 55–77. DOI: 10.1016/S0145-305X(02)00039-3 (cit. on pp. 64, 307).

References for Chapter C

[304] Clare F. J. Grant. 'Investigating primary and secondary B cell responses in cattle after immunisation with existing and novel vaccines'. PhD thesis. University of Oxford, 2013 (cit. on pp. 153, 310).