Structural Studies of Pestivirus Envelope Glycoproteins



Hazel Aitkenhead

Balliol

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2021

Acknowledgements

I would first like to thank my two supervisors Sir Stuart and Dr El Omari for providing support, encouragement, and direction in this project. Kamel especially for his great tuition on data collection and structure solution.

Within the Stuart group at Strubi a special thank you goes to Margaret for help with all thing's lab related, Helen for help with electron microscopy, and Tom for assisting in DLS experiment. The Siebold group was kind enough to provide me a space to work for the first 2 years of this project. Benjamin particularly for help getting started with cell culture.

At Diamond a special mention goes to the MX group and especially Claire for her help in trouble shooting the many, many cloning failures. The tutors at the joint Diamond CCP4 workshop really improved my knowledge of X-ray crystallography and were extremely valuable in solution of the bovCD46_SCR1-4 structure. Nathan from B21 was extremely helpful in aiding me to figure out how to process SAXS data. Finally, Jason from eBic4i provided me a great introduction to electron microscopy and gave me to courage to try it on some of my proteins.

Thanks to the BBSRC and Diamond light source for funding the project through the Interdisciplinary biosciences DTP.

A final thanks goes to my partner Philip for providing endless cups of adequately sized tea and copious quantities of chocolate.

Abstract

Pestiviruses belong to the family *Flaviviridae* with the most well-known being *Bovine Viral Diarrhoea Virus, Border Disease Virus*, and *Classical Swine Fever Virus* affecting cows, sheep, and pigs respectively. These viruses cause a large economic impact especially in the dairy and pig industries. As a result of being enveloped viruses, to enter the cytoplasm and replicate, they must first fuse their membrane with that of the cell's endosome, in a process which requires a fusion protein. Pestiviruses require two envelope glycoproteins for this step: E2, likely involved with receptor interaction, and E1, likely the fusion protein.

Reported within is the design of over 80 expression constructs, expression testing of 70, large scale production of 16 and the crystallisation of 7. As a result, two x-ray crystallography structures of the ectodomain of a putative pestivirus receptor bovine CD46 were solved. These structures revealed an interesting dimer, like that found in human CD46, although not reported previously. This dimer presents an interesting insight to the potential orientation of CD46 in the cell surface and could potentially play a role in anchoring the virus particles at sites of endocytosis.

This thesis also presents the structure of two novel pestivirus E2 ectodomains from APPV and NRPV. These structures highlight the variety of pestivirus E2 structures and support the finding that E2 is not the fusion protein. Investigations into N-terminal disorder at low pH, such as in the late endosome, revealed that this was not the case for all pestiviruses. The more typical E2 proteins from BVDV, BDV, and CSFV all display evidence of structural changes at low pH while the more atypical pestiviruses APPV and NRPV do not. In addition, an interesting β-hairpin was revealed towards the N-terminal region of all five E2 proteins which may play a role in receptor interaction.

Contents

List of Figures **x**

List of Tables **xii**

List of Abbreviations **xiii**

1 Introduction **1**

1.1 Pestiviruses 1

1.1.1 Virus Taxonomy 1

1.1.2 Virus History 3

1.1.3 Virus Genome and Overall Structure 3

1.1.4 Economic Impact of Pestivirus Infection 5

1.1.5 Current Methods of Pestivirus Control 5

1.2 Cell Entry Through Membrane Fusion 7

1.2.1 Virus Life Cycle 7

1.2.2 Membrane Fusion Mechanism 7

1.2.3 Fusion Proteins 9

1.3 Pestivirus Glycoproteins 11

1.3.1 Erns 11

1.3.2 E1 13

1.3.3 E2 15

1.3.4 Dimerization of Pestivirus Glycoproteins 18

1.3.5 Glycosylation of Erns, E1, and E2 Proteins 19

1.3.6 Occupancy of Glycoproteins on the Viral Surface 21

1.4 Pestivirus Cell Entry 21

1.4.1 How Might Pestivirus Membrane Fusion be Different 21

1.4.2 Method of Cell Entry 22

1.4.3 Cellular Receptors for Pestiviruses 23

1.4.4 CD46 as a Cellular Receptor for Bovine Viral Diarrhoea Virus 25

1.4.5 Cellular Receptors for Porcine Pestiviruses 27

1.4.6 Cell-Cell Transmission 28

1.5 Project Aims and Overview 30

2 Materials and Methods **32**
2.1 Cloning . . . 32
2.1.1 Expression Plasmids . . . 32
2.1.2 Gene Design . . . 33
2.1.3 Primer Design . . . 33
2.1.4 Polymerase Chain Reaction Conditions . . . 34
2.1.5 Restriction Cloning . . . 35
2.1.6 Single Step Cloning . . . 35
2.1.7 Movement of DNA Inserts from One Plasmid To Another . . . 36
2.1.8 Site Directed Mutagenesis . . . 36
2.1.9 Colony Screening . . . 38
2.2 Mammalian Cell Work . . . 38
2.2.1 Cell Culture . . . 38
2.2.2 HEK293S, LentiX . . . 39
2.2.3 Freezing and Thawing Cells . . . 40
2.3 Bacterial Cell Work . . . 40
2.3.1 Preparation of Competent Cells . . . 40
2.3.2 Transformation of Competent Cells . . . 41
2.4 Protein Expression . . . 42
2.4.1 Lentivirus Production . . . 42
2.4.2 Mammalian Expression (Secretion) . . . 42
2.4.3 Small Scale Expression Testing . . . 43
2.5 Protein Purification . . . 44
2.5.1 Sample Preparation . . . 44
2.5.2 Affinity Purification . . . 45
2.5.3 Size Exclusion Chromatography . . . 45
2.6 Gel Electrophoresis . . . 46
2.6.1 Agarose Gels for DNA . . . 46
2.6.2 SDS-PAGE . . . 47
2.6.3 Western Blotting . . . 47
2.7 Crystallography . . . 48
2.7.1 Crystallisation . . . 48
2.7.2 Data Collection . . . 48
2.7.3 Data Processing and Structure Solution . . . 49
2.8 SAXS . . . 51
2.8.1 Sample Preparation . . . 51
2.8.2 Data Collection . . . 51
2.8.3 Data Analysis . . . 51
2.9 Dynamic Light Scattering . . . 52

3 Protein Production **53**
3.1 Construct Design . . . 53
3.1.1 Bioinformatic Studies . . . 53
3.1.2 Hepacivirus and Host Receptor Proteins . . . 59
3.1.3 Protein Engineering . . . 60
3.1.4 Results . . . 63
3.2 Cloning . . . 64
3.2.1 Overall Cloning Strategy . . . 64
3.2.2 Restriction Cloning . . . 66
3.2.3 Single Step Cloning . . . 70
3.2.4 Point Mutations . . . 72
3.2.5 Results . . . 74
3.3 Expression . . . 74
3.3.1 Lentivirus . . . 74
3.3.2 Cell Lines . . . 75
3.3.3 Small Scale Test Expression . . . 77
3.3.4 Large Scale Expression . . . 79
3.3.5 Results . . . 81
3.4 Purification . . . 81
3.4.1 Sample Preparation . . . 81
3.4.2 Affinity Chromatography . . . 84
3.4.3 Size Exclusion Chromatography . . . 86
3.4.4 Results . . . 88
3.5 Validation Techniques . . . 91
3.6 Summary . . . 92

4 Cellular Receptors for Pestiviruses **95**
4.1 Introduction . . . 95
4.1.1 Heparan Sulphate and Laminin Receptors . . . 96
4.1.2 ADAM17 Receptor . . . 97
4.2 CD46 Receptor . . . 99
4.3 Structure Solution of Bovine CD46 Domains 1-2 . . . 101
4.3.1 Crystallisation and Optimisation . . . 101
4.3.2 Data Collection and Processing . . . 103
4.3.3 Structure Solution and Refinement . . . 104
4.4 Structure Solution of Bovine CD46 domains 1-4 . . . 106
4.4.1 Crystallisation and Optimisation . . . 106
4.4.2 Data Collection . . . 107
4.4.3 Structure Solution and Refinement . . . 108

4.5 Structural Analysis of CD46 extracellular domain 110
4.5.1 Overall fold 110
4.5.2 SCR Domain Fold 114
4.6 Structure Solution of Porcine CD46 Domains 1-4 116
4.6.1 Crystallisation and Optimisation 116
4.6.2 Data Collection 117
4.6.3 Structure Solution Failure 118
4.7 Computational Modelling of Protein Structure 119
4.7.1 Introduction 119
4.7.2 Structural Modelling of Bovine CD46 Domains 1-4 120
4.7.3 Structural Modelling of Porcine CD46 Domains 1-4 122
4.8 Structural Comparison of CD46 Ectodomain Structures 125
4.9 CD46 Dimerization 126
4.9.1 Crystallographic Dimer 126
4.9.2 Size Exclusion Chromatography of CD46 130
4.9.3 Dynamic Light Scattering for Molecular Weight Determination 130
4.10 Putative BVDV Binding Interaction 133
4.10.1 Binding Site 133
4.10.2 Binding Studies of BVDV E2 and CD46 Ectodomains 135
4.11 CD46 Orientation on Cellular Surface 136
4.12 Conclusion 137

5 Pestivirus E2 Glycoproteins 139
5.1 Introduction 139
5.2 Crystallography of Typical Pestivirus E2 Glycoproteins 141
5.2.1 Crystallography and Optimisation of CSFV E2 141
5.2.2 Crystallography of BDV E2 143
5.3 Structure of APPV E2 144
5.3.1 The Phase Problem and Sulphur SAD Phasing 144
5.3.2 Crystallography and Optimisation 147
5.3.3 Crystal Harvesting and Data Collection 148
5.3.4 Structure Solution and Refinement 150
5.3.5 Structure Analysis 150
5.4 Structure of NRPV E2 156
5.4.1 Crystallography and Optimisation 156
5.4.2 Crystal Harvesting and Data Collection 157
5.4.3 Structure Solution and Refinement 158
5.4.4 Structure Analysis 160
5.5 SAXS of Pestivirus E2 Glycoproteins 163

5.5.1 Small Angle X-ray Scattering . . . 163
5.5.2 Experimental Conditions and Data Processing . . . 164
5.5.3 Results . . . 164
5.6 Computationally Generated Models of Typical Pestivirus E2 Glycoproteins . . . 168
5.6.1 Model of BDV E2 Ectodomain . . . 168
5.6.2 Model of CSFV E2 Ectodomain . . . 170
5.7 Structural Comparison of Pestivirus E2 Glycoproteins . . . 173
5.7.1 Comparison of Overall Structure . . . 173
5.7.2 N-terminal Disorder . . . 178
5.7.3 Solvent Exposed β-hairpin . . . 179
5.8 Conclusion . . . 181

6 Conclusions and Future Directions **183**
6.1 Conclusion . . . 183
6.1.1 Importance . . . 183
6.1.2 Summary of Findings . . . 184
6.1.3 Implications of Receptor Structure . . . 185
6.1.4 Implications of New Pestivirus E2 Structures . . . 186
6.2 Future Directions . . . 187
6.2.1 Receptor Interaction . . . 187
6.2.2 Membrane Fusion . . . 189

Appendices

A GenBank Accession Codes **192**

B Gene Sequences **193**
B.1 BVDV_E1-E2 . . . 193
B.2 BDV_E1-E2 . . . 195
B.3 CSFV_E1-E2 . . . 197
B.4 NRPV_E1-E2 . . . 199
B.5 APPV_E1+E2 . . . 201
B.6 bovCD46 . . . 202
B.7 porCD46 . . . 203
B.8 ADAM17-Ecto . . . 204
B.9 HCV_E1-Ecto . . . 206
B.10 HCV_E2-Ecto . . . 207
B.11 HCV_E1-Link-E2 . . . 208

C List of Primers **210**

D Domain Boundaries **212**

E Construct Details **213**

F Additional SAXS Results **216**

G Cryo-EM of CSFV E2 with Neutralising Antibody **222**
G.1 Introduction 222
G.2 Grid Preparation 222
G.3 Data Collection and Analysis 223
G.4 Materials and Methods 224
G.4.1 Grid Preparation 224
G.4.2 Data Collection 224
G.4.3 Data Processing 224

References **225**

List of Figures

1.1 PhylogeneticTree. 2
1.2 Polyprotein. 4
1.3 Viral Particle. 4
1.4 Virus Life Cycle 8
1.5 Viral Membrane Fusion Proteins. 10
1.6 Erns Sequence Alignment. 12
1.7 CSFV E1 Transmembrane Prediction. 14
1.8 BVDV E2 Structure. 16
1.9 E2 Sequence Alignment. 17
1.10 Human CD46 Structure. 26

3.1 Multiple Sequence Alignment of Pestivirus E1-E2 region 55
3.2 TM Prediction for E1-E2 Region 58
3.3 Construct Boundaries 64
3.4 Cloning Comparison 69
3.5 Primer Extension PCR Mutagenesis 73
3.6 Test Expression 78
3.7 Protein Production Schematic 80
3.8 Protein Production Results Overview 83
3.9 Example SEC Traces part 1 89
3.10 Example SEC Traces part 2 90

4.1 CD46 Sequence Conservation 100
4.2 Crystal of bovCD46_SCR1-2 102
4.3 Electron Density Maps for bovCD46_SCR1-2 105
4.4 Crystal of bovCD46_SCR1-4 107
4.5 Electron Density Maps for bovCD46_SCR1-4 109
4.6 Overall structure of the bovine CD46 extracellular domain 111
4.7 Superposition of bovine CD46 structures 113
4.8 Structure Comparison of Individual SCR Domains 116
4.9 Crystals of porCD46_SCR1-4 117
4.10 Computationally Generated Models of bovCD46_SCR1-4 121

4.11 Computationally Generated Models of PorCD46_SCR1-4 123
4.12 Structural Comparison of CD46 Ectodomains 126
4.13 Crystallographic Dimer Found in bovCD46 127
4.14 Crystallographic Dimer Found in Human CD46 128
4.15 LigPlot$_+$ Diagram of the Dimerization Interface 129
4.16 DLS of CD46 Ectodomain 132
4.17 Structural Comparison of bovCD46 and Human CD46 134
4.18 Potential Cell Surface Orientation. 137

5.1 CSFV E2 Ectodomain Crystals 142
5.2 Crystals of BDV E2 Ectodomain 144
5.3 Crystals of APPV E2 Ectodomain 148
5.4 Electron Density Maps for APPV E2 Ectodomain 151
5.5 Structure of APPV E2 ectodomain 153
5.6 Topology Diagrams 155
5.7 Crystals of NRPV E2 Ectodomain. 157
5.8 Electron Density Maps for the NRPV E2 Ectodomain 160
5.9 Structure of the NRPV E2 Ectodomain 162
5.10 Normalised Kratky Plots 166
5.11 SAXS Models of Pestivirus E2 Ectodomains 167
5.12 Computationally Generated Models of the BDV E2 Ectodomain . . 169
5.13 Computationally Generated Models of the CSFV E2 Ectodomain . 172
5.14 Comparison of Pestivirus E2 Ectodomain Structures 174
5.15 CSFV Missing Dilsulphide 175
5.16 Superposition of NRPV and BVDV E2 Ectodomains 177
5.17 Pestivirus E2 Hairpin 180

F.1 BVDV E2 Ectodomain 217
F.2 BDV E2 Ectodomain 218
F.3 CSFV E2 Ectodomain 219
F.4 NRPV E2 Ectodomain 220
F.5 APPV E2 Ectodomain 221

G.1 Cryo-EM Micrographs and 2D Class Averages 223

List of Tables

2.1 X-ray Crystallography Data Collection Details 50
2.3 Programs Used for Structure Solution 50
2.5 SAXS Data Collection Details 52

3.1 Protein Production Results 82
3.2 Analysis of SEC results 91

4.1 Data Collection and Refinement Statistics for bovCD46_SCR1-2 . . 104
4.2 Data Collection and Refinement Statistics for bovCD46_SCR1-4 . . 108
4.3 Data Set Statistics for porCD46_SCR1-4 118
4.4 DLS of CD46 Ectodomain 131

5.1 Data Collection and Refinement statistics for APPV E2 Ectodomain 150
5.2 Data Collection and Refinement Statistics for NRPV E2 Ectodomain 158
5.3 Data Analysis Statistics from SAXS 165

G.1 Successful Cryo-EM Grid Prepatation Parameters 223
G.2 Cryo-EM Data Collection Parameters 224

List of Abbreviations

ADAM17 . . . A Disintegrin and metalloproteinase domain-containing protein 17

APPV . . . . . Atypical porcine pestivirus

BDV . . . . . Border disease virus

BMA . . . . . β-D-mannose

bovCD46 . . . Bovine CD46

BVDV . . . . . Bovine viral diarrhea virus

CAMEO . . . Continuous automated model evaluation

CASP . . . . . Critical Assessment of protein structure prediction

CHO . . . . . Chinese hamster ovary

CSFV . . . . . Classical swine fever virus

CT . . . . . Congenital tremor

Cryo-EM . . . Cryogenic Electron Microscopy

DLS . . . . . Dynamic light scattering

FBS . . . . . Fetal bovine serum

GST . . . . . Glutathione-S-transferase

HCV . . . . . Hepaitius C virus

HEK . . . . . Human embryonic kidney

IMAC . . . . . Immobilized metal affinity chromatography

LamR . . . . . Laminin receptor

LB . . . . . Luria broth

MAb . . . . . Monoclonal antibody

MAD . . . . . Multi-wavelength anomalous diffraction

MAN . . . . . Mannose

MBP . . . . . Maltose-binding protein

MIR . . . . . Multiple isomorphous replacement

Mw . . . . . Molecular weight

NAG . . . . . N-acetyl-β-D-glucosamine

NEAA . . . . . Non-essential amino acids

NRPV . . . . . Norway rat pestivirus

nt . . . . . Nucleotide

PBS . . . . . Phosphate buffered saline

PBST . . . . . Phosphate buffered saline with detergent

PCR . . . . . Polymerase chain reaction

PDI . . . . . Polydispersity index

PEI . . . . . Polyethylenimine

porCD46 . . . Porcine CD46

RCA . . . . . Regulators of complement activation

Rg . . . . . Radius of gyration

SAD . . . . . Single-wavelength anomalous diffraction

SAXS . . . . . Small angle X-ray scattering

SCR . . . . . Short consensus repeat

SFM . . . . . Serum free media

STP . . . . . Region rich in serine, threonine, and proline

TAE . . . . . Tris-acetate-edta

TetR . . . . . Repressor of the tetracycline resistance element

1 Introduction

Contents

1.1 Pestiviruses 1
1.2 Cell Entry Through Membrane Fusion 7
1.3 Pestivirus Glycoproteins 11
1.4 Pestivirus Cell Entry 21
1.5 Project Aims and Overview 30

1.1 Pestiviruses

1.1.1 Virus Taxonomy

The family *Flaviviridae* includes the genus *Pestivirus* along with the *Flavivirus, Hepacivirus,* and *Pegivirus*. The genus *Flavivirus* has its name derived from the Latin for yellow *"flavus"* and is made up of 53 species, some well-known members include yellow fever virus, Zika virus, and dengue virus[1]. The virus Hepatitus C (HCV) is a member of the genus *Hepacivirus* and causes chronic liver disease[2], helping to name the genus from the Greek for liver *"hepatos"*. The most recent addition to the *Flaviviridae* family was the genus *Pegivirus*. Viruses from this genus have been shown to cause persistent infections in many mammalian species although there is no current association to disease[1].

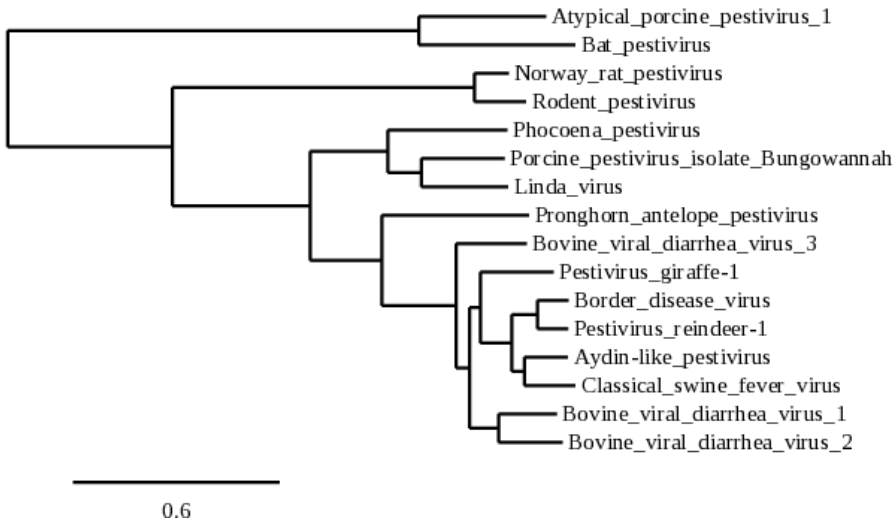


Figure 1.1: Phylogenetic tree of the entire polyprotein from the main pestivirus species generated using the phylogeny.fr[3] web service and a multiple sequence alignment generated using Clustal Omega[4]. GenBank accession numbers that were used for alignments can be found in appendix A

Up until recently the genus *Pestivirus* contained 4 species: BVDV-1 (*bovine viral diarrhea virus 1*), BVDV-2 (*bovine viral diarrhea virus 2*), BDV (*border disease virus*), and CSFV (*classical swine fever virus*). However, many other genetically related viruses have been described. These were initially known as 'unassigned pestiviruses'[5]. As can be seen in Figure 1.1 these viruses have been found in both wild and domestic hooved animals such as giraffes and pigs as well as non-hooved animals like bats[6] and rats[7]. These genetically distinct viruses, which were potentially new species in the *Pestivirus* genus, were named either after the animal they were isolated from or the disease they cause. This resulted in confusion between the taxonomic category and physical agent. Recently, 7 new species have been added to the genus and all species have been renamed as *Pestivirus A* through to *Pestivirus K* to clarify that a species is being referred to and to remove confusion over host range. The original 4 species were renamed as follows: BVDV-1 is *Pestivirus A*, BVDV-2 became *Pestivirus B*, CSFV *Pestivirus C* and BDV *Pestivirus D*. Two other species to note for the basis of this thesis are Norway Rat Pestivirus (NRPV), now *Pestivirus J* and Atypical Porcine Pestivirus (APPV) also known as *Pestivirus K*[5].

1.1.2 Virus History

The first reported incidence of a Pestivirus in the early 19th century involved CSFV (then known as hog cholera virus). During the early 20th century, the concentrated expansion in pig production made this disease one of the most important disease of pigs in the developed world[8]. BVDV was the second reported Pestivirus, first described in 1946 (diarrhoea symptoms)[9] and 1957 (mucosal symptoms)[10]. The third Pestivirus was originally identified in the border between England and Wales (hence named Border disease virus) in 1959[11]. Since then, other pestiviruses have been described in both the wild and domestic animal populations, the most recent of these were from the species *Pestivirus K*: APPV as well as the unclassified Linda virus both first discovered in 2015[12, 13].

1.1.3 Virus Genome and Overall Structure

Pestiviruses, and indeed all members of the *Flaviviridae* family, are positive-sense single stranded RNA viruses. They have relatively small 40-60nm virions with a lipid envelope and a non-segmented 9-13kb genome. The Pestivirus genome is on the larger end of the scale at 11-13kb although the virions are still in the 40-60nm range. The genome encodes 12 different proteins, 4 are structural and the other 8 non-structural. Like all other members of the *Flaviviridae* family, the pestivirus genome lacks a polyadenylated 3' terminus. However unlike the genus *Flavivirus*, the *Pestivirus* have an internal ribosome entry site rather than a 5' terminal type one cap structure[1].

The pestivirus genome is used to produce a single 4000 (roughly) amino acid polyprotein, as shown in Figure 1.2. This polyprotein is co and post-translationally processed by both cellular and viral proteases into the individual proteins. Of the 4 structural proteins 3 are envelope glycoproteins: Erns, E1, and E2. E1 and E2 are membrane anchored[14, 15] while Erns is more loosely associated via an amphipathic helix at the C-terminus[16]. The final structural protein is the single capsid protein C[17]. The 8 non-structural proteins include 2 proteases: Npro, a

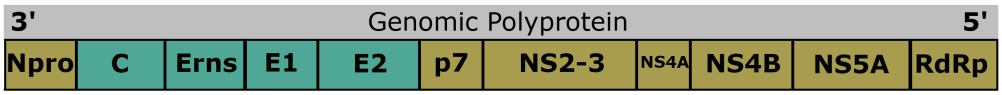


Figure 1.2: Schematic of the genome organisation of pestiviruses. Genes coding for structural proteins are coloured blue and non-structural proteins genes are coloured yellow.

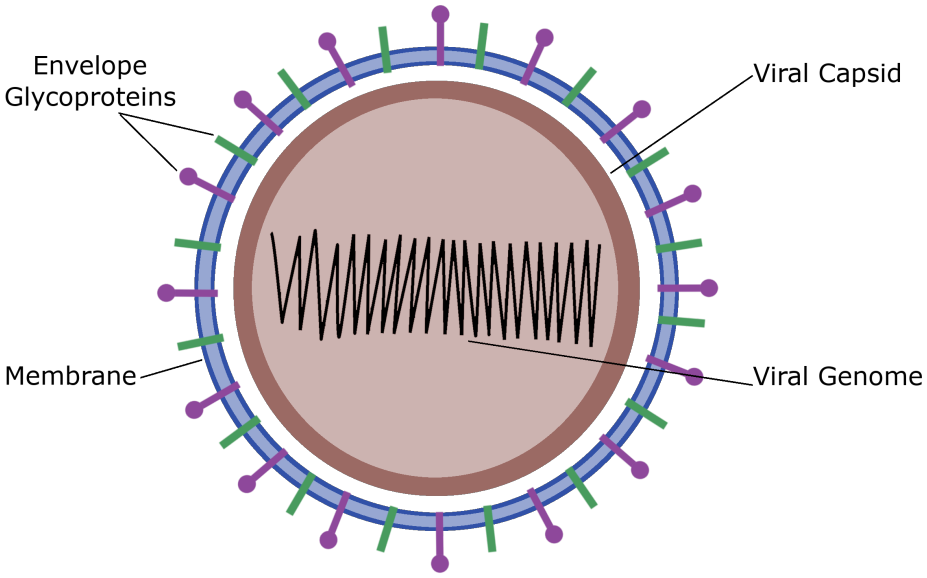


Figure 1.3: Diagram of the main components of a generic enveloped virus particle. The viral capsid is depicted in brown and the envelop in blue, while the envelope glycoprotein(s) are purple and green.

protease which auto-catalytically separates itself from the polyprotein and NS2 which processes the multifunctional protein NS2-3 into NS2 and NS3[18]. P7, NS4A, NS4B, NS5A, and NS5B need further study to fully elucidate all their roles in proteolysis, RNA replication and virion morphogenesis and maturation. X-ray crystallography structures have been solved for Npro[19], Erns[20], E2[21, 22], NS3[23], and NS5B[24].

As seen in Figure 1.3, the viral particles themselves are composed of a nucleocapsid which surrounds the RNA genome, this is built from the capsid protein C. A lipid bilayer or envelope is found surrounding the nucleocapsid, this envelope contains the glycoproteins E1, E2, and Erns[25].

1.1.4 Economic Impact of Pestivirus Infection

Different pestiviruses have had a huge economic impact in various livestock industries. Few studies have tried to calculate the exact economic impact of outbreaks due to difficulties in identifying and calculating costs not directly attributable to a pestivirus outbreak. A study by Richter *et al* in 2017[26] looking at direct monetary loss due to bovine viral diarrhea virus found the literature reports vastly varying figure for direct losses in a huge range from 0.5-687 USD per animal (looking at regional and national herds). This study looked at 44 studies from 15 countries also identified direct losses as being 24 USD higher per animal for naïve dairy cattle than beef cattle. A Pestivirus outbreak in pigs in the Netherlands in 1997 had an estimated cost of almost 2.3 billion USD[27]. In other pestiviruses the economic impact has not been estimated as the direct costs cannot currently be calculated, an example of this is APPV. APPV can cause Atypical Porcine Pestivirus associated congenital tremor (CT) in piglets[28], however this is rarely the direct cause of their death. Death is instead caused by starvation or crushing which could be attributed to CT or to a variety of other factors. Estimates indicated the mortality in CT-piglets to be between 24% and 60%[29].

In addition to the economic impact, CSFV is a threat to the food security of communities in developing countries where back yard pig rearing is common. This is due to its endemic nature in many parts of the world[30].

1.1.5 Current Methods of Pestivirus Control

For CSFV, there are effective tools for the diagnosis of infection available although the same cannot be said for vaccination. Currently modified live vaccines are being used and while these are highly efficacious, they interfere with the diagnosis tools, making it impossible to differentiate vaccinated animals from infected animals[31]. The two major methods of control used for domestic pigs are culling and disposing of all infected and potentially infected animals (known as stamping out) or vaccination. Within the wild boar population hunting or oral vaccination is used if the virus is

no longer self-limiting and has become endemic[32]. As CSFV is such a large threat to the pig industry it is reportable to the world health organisation of animal health and countries are given a CSFV-free status if appropriate[33]. In the EU, which is currently CSFV-free but has previously had outbreaks, there are very effective control procedures involving identification and recording systems, hygiene, farming, and import regulation used in combination with severe eradication procedures involving culling of all animals on affected farms. In other countries which are not virus free, prophylactic vaccination is used extensively[30]. New subunit vaccines are beginning to emerge and two have been licenced for use in the EU, these may prove valuable for use in future outbreaks as they allow differentiation between vaccinated and infected animals[31].

Control of BVDV outbreaks is difficult due in part to persistently infected animals and lengthy transient infections. Current vaccines against BVDV are modified live vaccines, which cannot be used during the first 6 months of pregnancy, or, killed vaccines which take multiple injections and several weeks to provide immunity[34]. Although CSFV can be successfully controlled with vaccination, this is not the case for BVDV mainly due to the presence of persistently infected animals which are constantly shedding virus rather than just for a couple of weeks. In this case, it would be necessary to administer constant booster vaccinations to the herd to obtain complete immunity and as such vaccination alone is not a suitable stratagem for BVDV control. There are however very cost effective BVDV detection protocols available using ELISA techniques which are enabling mass screening efforts as well as the identification of persistently infected animals[35]. Various Scandinavian countries have successfully used this method to identify and remove persistently infected animals and restrict the movement of any associated animals until the virus was eliminated leaving the area mostly free of BVDV[34]. Using this method Switzerland which initially had a very high occurrence of BVDV (>80.00% of cattle), was able to reduce this to 0.02% over only 4 years. Some countries use a combination of removal of persistently infected animals together with vaccination. Ireland used this method to reduce prevalence from 0.66% to 0.10% over 4 years[34].

1.2 Cell Entry Through Membrane Fusion

1.2.1 Virus Life Cycle

Generally, pestiviruses enter the host through the oral and nasal mucosal membranes, infect initial cells then spread throughout the host body using the lymphatic system and blood. Initially the virus interacts with receptor proteins on the cellular surface through glycoproteins on the viral envelope. These glycoproteins determine the specificity of the cell to its host and therefore the range of hosts or tropism of the virus. The attachment of the viral glycoprotein to the host cell surface receptor protein is thought to cause a conformational change in some viral protein leading to the viral envelope and host cell membrane coming into close proximity and fusing.

Once the viral and host cell membranes fuse, the viral capsid is released inside the host cell. Host cell enzymes then degrade the viral capsid releasing the viral genome to the cytoplasm of the host cell. Following its release the viral genome is replicated and can immediately serve as messenger RNA for translation into protein, via a 5' internal ribosome entry site in the case of pestiviruses[36]. When the viral proteins and genome have been produced, they are packaged into newly synthesised viral capsids to make new mature virions, the viral glycoproteins are directed to the cell membrane. These virions are then released from the host cell through a process called budding, this gives the virions their lipid envelope and the glycoproteins which the membrane now contains. The released particles are then ready to infect new cells. This entire process is illustrated in Figure 1.4.

Membrane fusion is essential for the viral life cycle and this thesis looks at the glycoproteins and cellular receptors involved in this vital process.

1.2.2 Membrane Fusion Mechanism

All currently known enveloped viruses follow the same generalised mechanism of membrane fusion to facilitate entry of the virus to the host cell. These viruses have several copies of a fusion protein, or fusions proteins, on the surface their viral

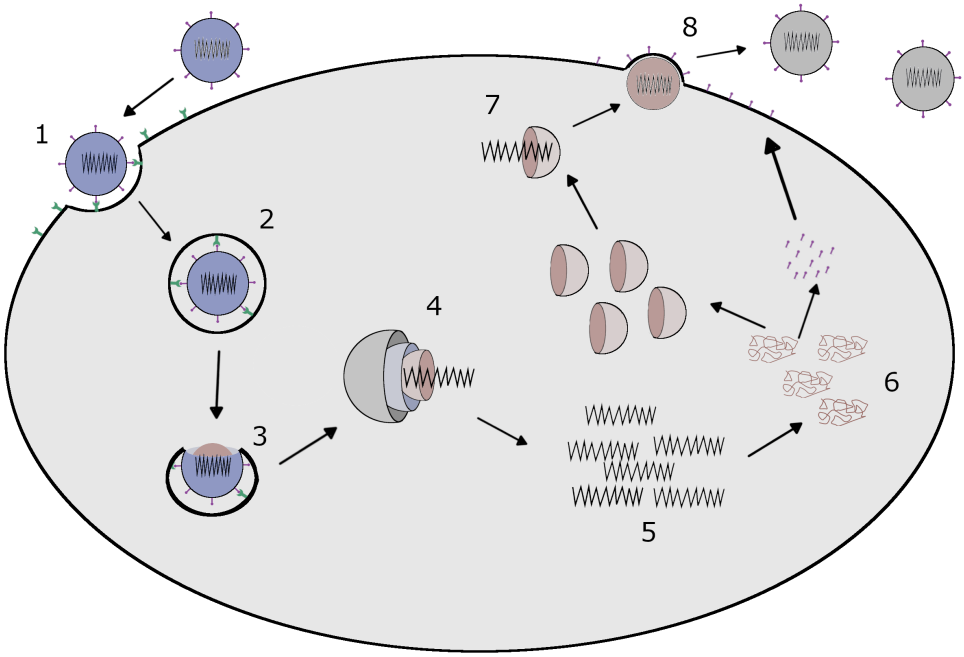


Figure 1.4: Diagram depicting a generalised virus lifecycle.1- Glycoprotein on the virus surface interact with receptors on the cell surface. 2- The virus particle is endocytosed 3- The endosomal and viral membranes fuse releasing the virus particle. 4- The virus particle is deconstructed, and the viral genome released. 5- The viral genome is replicated. 6- Viral proteins are produced. 7- New virus particles are constructed. 8- The virus buds from the cellular membrane, now displaying viral glycoproteins.

envelope. These fusion proteins are generally held in a pre-fusion conformation prior to interaction with the host cell receptor. First, there is a priming step to prepare the viral fusion protein, this is generally irreversible and can be a cleavage step. The priming step may expose a hydrophobic fusion peptide ready for insertion into the host cell phospholipid membrane. Following the priming step there is a triggering step which starts the membrane fusion process. This triggering step could be interaction with a cellular receptor or another viral protein, or a drop in pH. The priming step generally causes a conformational change in the viral fusion protein which exposes the fusion loop or peptide which is inserted into the cell membrane, the fusion protein is then triggered to fold back upon itself, causing the viral and host membrane into close proximity[37]. Membrane fusion then occurs first with hemi fusion where the membranes have merged but not fully fused, then follows the formation of a fusion pore. The fusion proteins are necessary to catalyse the membrane fusion processes due to the repulsive hydration force

between the two membranes[38]. This forms a kinetic barrier increasing as the distance between membranes decreased[37]. The energy released when the fusion protein changes conformation to its lowest energy conformation is used to overcome this repulsive hydration force as the membranes loose water from their lipid heads as the two membranes come close together[39].

1.2.3 Fusion Proteins

1.2.3.1 Class I

Class I fusion proteins have a central trimeric α-helical coiled coil (Figure 1.5), a classic example would be the hemagglutinin protein from influenza A virus[40]. In both the pre- and post-fusion states class I fusion proteins are found as non-covalently linked homotrimers. Individual monomers are type 1 membrane proteins with a single pass transmembrane anchor. There is a fusion peptide located near the N-terminal end of the longest α-helix of the coiled coil, it is located under the N-terminal subunit trimer cap. Upon interaction with receptors on the hot cell surface the N-terminal subunit cap is cleaved exposing the fusion peptide when the long α-helix springs out[39].

1.2.3.2 Class II

This class contains fusion proteins from the genus *Flavivirus* who belong to the same family as the genus *Pestivirus*. Class II fusion proteins generally have an elongated monomer conformation with 3 β-sheet domains (Figure 1.5), an example would be the E protein from Dengue type 2 virus[41]. Unlike class I proteins, class II proteins do not form prefusion trimers on the surface of the virus but instead assemble as a network encasing the entire viral particle[42]. Again, unlike class I proteins they do not contain long α-helices instead they are composed of three domains, the first and central domain is formed of a β-sandwich and contains the internal fusion loop. Domain 1 is connected to the immunoglobulin like domain 3 through its C-terminus. At the N-terminal end of domain 1 is located domain 2[41]. Post fusion the rod like structure is bent in half exposing the internal fusion loop in the central domain 1[43].

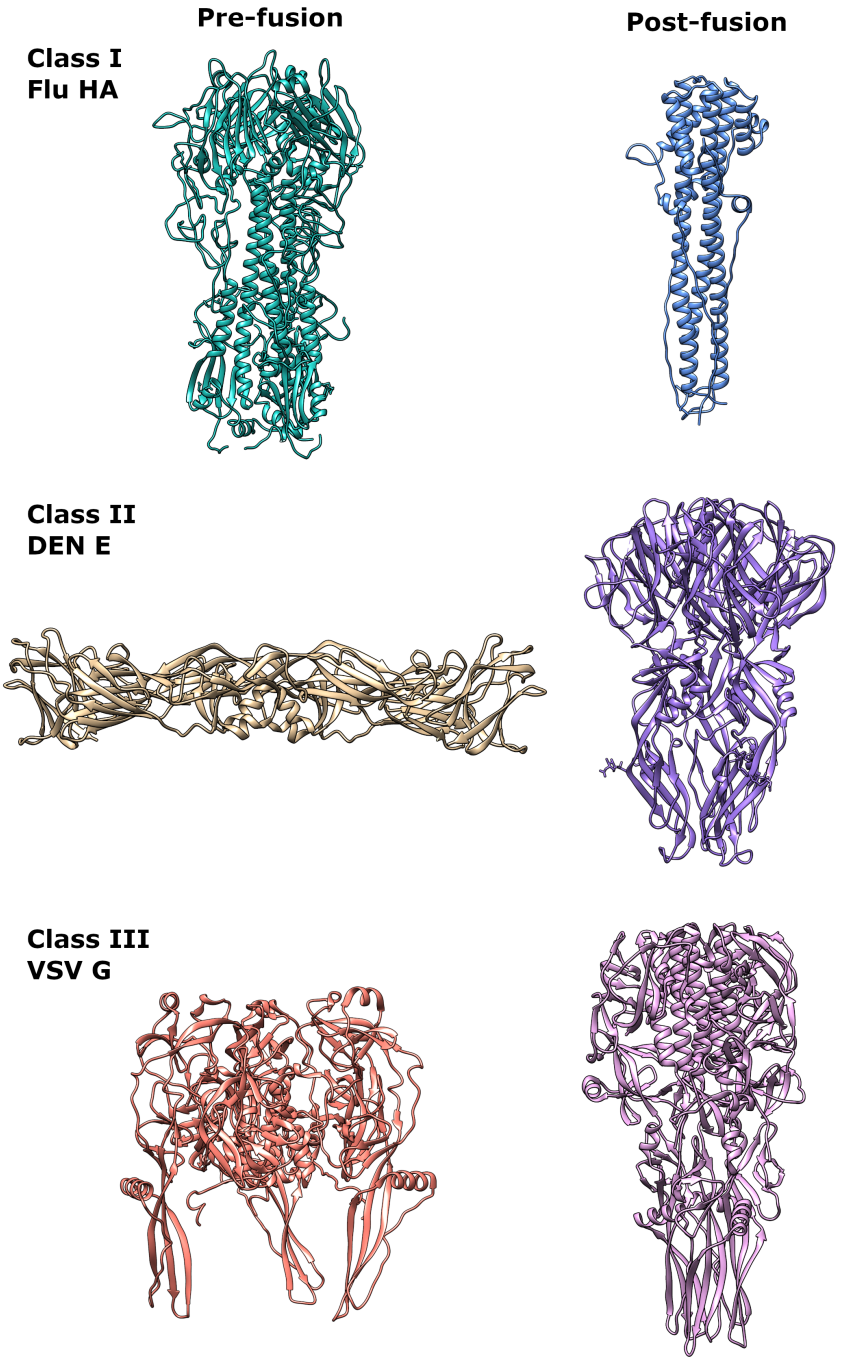


Figure 1.5: Viral fusion protein in pre- and post-fusion conformations. Pre-fusion Influenza hemagglutinin (PDB code:2HMG), Post-fusion Influenza hemagglutinin (PDB code: 1QU1), Pre-fusion Dengue virus E (PDB code: 1KOE), Post-fusion Dengue virus E (PDB code: 1KO8), Pre-fusion Vesicular stomatitis virus G (PDB code: 5I2S), Post-fusion Vesicular stomatitis virus G (PDB code: 5I2M).

1.2.3.3 Class III

The class III fusion proteins contain two exposed fusion loops along with a central α-helical coiled coil (Figure 1.5), an example is the G protein from the vesicular stomatitis virus[44]. Like class I protein, class III proteins are found as trimers on the surface of the viral particle, but like class II proteins they contain fusion loops instead of fusion peptides. Structurally they also appear to be a mixture of class I and class II proteins, they contain a central α-helical coiled coil like class I and at the N-terminus they contain a β-sheet domain[44]. However, unlike class I they do not undergo a preliminary cleavage step, in fact they do not undergo any priming step at all[45].

1.3 Pestivirus Glycoproteins

1.3.1 Erns

Erns is the second largest glycoprotein in the pestiviruses and has been found to have a mass ranging from 44kDa-48kDa depending on the species[46]. Erns stands for envelope protein RNase secreted and was first classified as belonging to the T2 ribonuclease RNase family in 1993[47]. The crystal structure of the catalytic RNase domain of the Erns protein from BVDV-1 has been solved and spans the residue 3-165 region (Figure 1.6). Erns RNase activity is necessary in animal host infection but is not needed in tissue culture, this could be due to the reported blocking of the innate immune system by the RNase activity of Erns[20].

Erns is unlike the other Pestivirus glycoproteins E1 and E2 in that it does not have a transmembrane anchor at its C-terminus. Instead it is a loosely anchored membrane protein using its multifunction C-terminus which it is folded into a 61-residue amphipathic helix[48]. It is likely that Erns binds flat to the surface of the viral membrane as the hydrophilic and hydrophobic parts of the helix are distributed throughout[49]. Also unlike E1 and E2, Erns does not have a counterpart in other members of the family *Flaviviridae*[50] and is not essential for host cell entry[51].

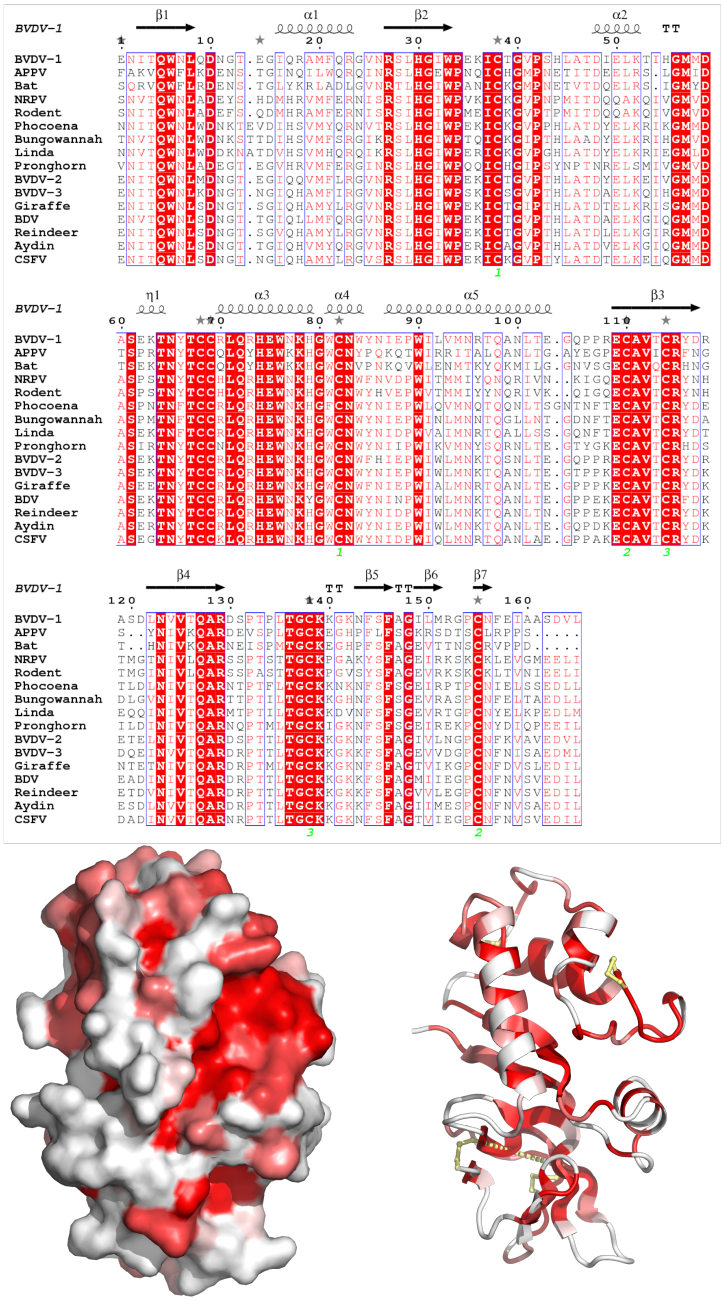


Figure 1.6: Sequence alignment of Erns protein region. **Top-** Multiple sequence alignment of the Erns position of the pestivirus polyprotein generated using Clustal Omega[4]. Sequence annotation was generated using the ESPript web service[52]. Conserved residues are highlighted in red while similar residues have a red type colour and are surrounded by a blue outline. The secondary structure assignment for BVDV-1 is shown on top of the sequence. **Bottom-** Surface (left) and cartoon (right) representation of the Erns structure (PDB code: 4DWC) coloured as in sequence. GenBank accession numbers that were used for alignments can be found in appendix A

The C-terminus of Erns has been found to have many functions in addition to binding the viral envelope. It also serves as a binding site for glycosaminoglycans on the surface of the host cell. A specific region of basic residues K213-K220 are vital for binding to the cell surface glycosaminoglycan heparin sulphate[53]. A large proportion of the Erns produced by the host cell is secreted rather than being retained on new viral particles. This feature is also due to the C-terminus lacking a transmembrane anchor. Residues L183, I190, and L208 have been shown to influence intracellular localisation of the protein[54]. This discovery gave the C-terminus the third function of a retention and secretion signal. The fourth and currently final function of the C-terminus is a cleavage site for cellular signal peptidase. This is because Erns is initially formed as a precursor protein Erns:E1 which is then cleaved by cellular signal peptidase into Erns and E1 thereby forming the C-terminus of Erns[55].

1.3.2 E1

E1 is the smallest Pestivirus envelope glycoprotein at 33kDa[46]. It is a type 1 transmembrane protein with a single span hydrophobic α-helical transmembrane anchor at the C-terminus[15]. The ectodomain is everything N-terminal of the transmembrane anchor. The E1 protein is the least well studied of the Pestivirus envelope glycoproteins, there is no structural information available, no functional information available, and the sequence has no homology to any other proteins. There is a hydrophobic region (Figure 1.7) located in the ectodomain which could potentially be a fusion peptide. This along with the fact we know that neither the E2 or Erns protein contain a fusion peptide indicate that E1 is likely to be the fusion protein[49].

E1 is significantly smaller than most other known fusion proteins, the examples shown in Figure 1.5 are around twice as big. With a protein of this size a highly elongated structure would be required to bridge the gap between two membranes to enable it to catalyse the membrane fusion process and overcome the kinetic barrier. It would be impossible for it to adopt a class II fusion protein fold like other members

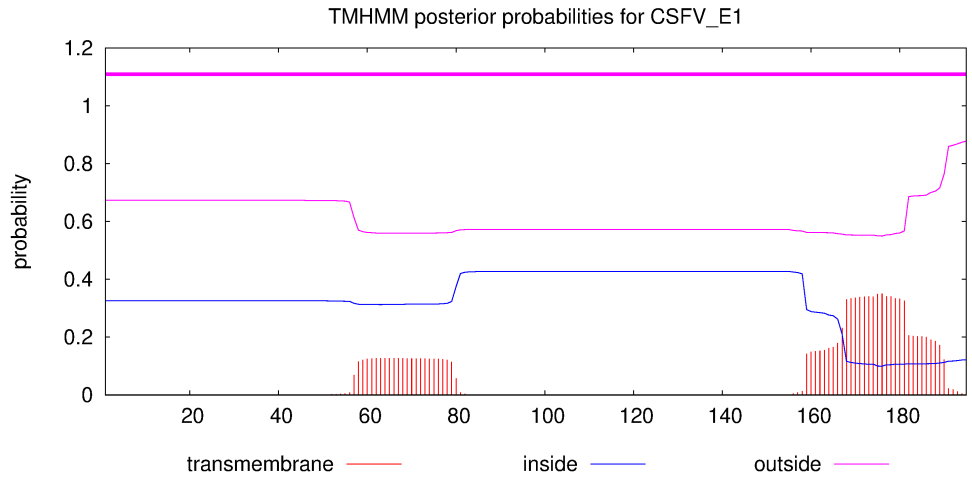


Figure 1.7: Transmembrane prediction for the CSFV E1 protein. Figure generated using TMHMM server v. 2.0[59].

of the *Flaviviridae* family and be long enough to reach the host cell membrane. It is also too small adopt a class III fusion protein fold with an extended enough conformation to bridge the 20nm[56] gap between membranes. The predicted 30% α-helical composition of E1 does not support a class II or class III fusion protein fold[57]. The central α-helical coiled coil of class I fusion proteins allows them to support more extended conformations, the α-helical content of E1 could form such a central alpha helical bundle. Additionally, initial evidence supporting the formation of BVDV E1 homotrimers was published in 2021[58], a classic feature of class I membrane fusion proteins. However, the E1 sequence does not appear to have a leucine zipper motif, a common feature of helical coiled coils. The E1 sequence also lacks evidence of an N-terminal fusion peptide and E1 has not been reported to undergo proteolytic activation, both classic features of class I fusion proteins[57].

It has been shown in the closely related HCV virus that E1 will not adapt a proper fold without the presence of E2 indicating E2 may act as a scaffold or chaperone for E1[60]. Additionally, the combination of E1 and E2 is both necessary and sufficient for membrane fusion and host cell entry so clearly both proteins play an important role[61]. As E1 does not appear to belong to any of the current membrane fusion protein classes it may represent a new class, potentially alongside HCV E1, requiring the support of a second protein for membrane fusion.

1.3.3 E2

E2 was originally thought to be the fusion protein of the pestiviruses as it is the largest of the glycoproteins at 55kDa[46]. Like E1 it has a membrane anchor, computational modelling suggests it is composed of a hydrophobic α-helix hairpin located at its C-terminus[14]. The structure of the E2 ectodomain has been solved and shows it has an elongated structure composed of four domains in the N-terminal ectodomain followed by the C-terminal membrane anchor[49]. Solving the structure confirmed that E2 did not contain a fusion peptide and could not be the fusion protein. E2 has been hypothesised to play a role in the interaction of BVDV with the putative cellular receptor bovine CD46 (bovCD46)[62]. E2 may therefore play a role in receptor binding, along with acting as a scaffold or chaperone for E1.

The crystal structures of the E2 ectodomain solved at both high and low pH (Figure 8.1) were very similar apart from the N-terminal most domain DA[21, 22]. The C-terminal domain DD located next to the viral envelope is highly conserved between the different species of the pestiviruses and is involved with homodimerization. It has a unique fold composed of β-sheets and loop regions. The next domain DC also has a β-sheet fold with no structural homologs currently solved. The N-terminal most domains DB and DA are likely to be most exposed on the surface and both have IG-like folds. This indicated they may be involved with receptor binding. A peptide on domain DB near the DA:DB domain interface has been identified as a potential receptor binding region[63].

Interestingly in the low pH structure of the ectodomain, the N-terminal most domain DA was completely disordered exposing a β-hairpin to the solvent[21]. This β-hairpin is one of the peptides described previously as potentially involved with receptor binding[63]. On entry to the endosome (the current proposed model of cell entry for pestiviruses) a drop of pH could cause protonation of a conserved histidine H70 (Figure 1.9) in the N-terminal DA domain. It has been postulated that protonation of this conserved residue could destabilise the entire domain, computational studies confirm that the DA domain is likely to unfold at low pH

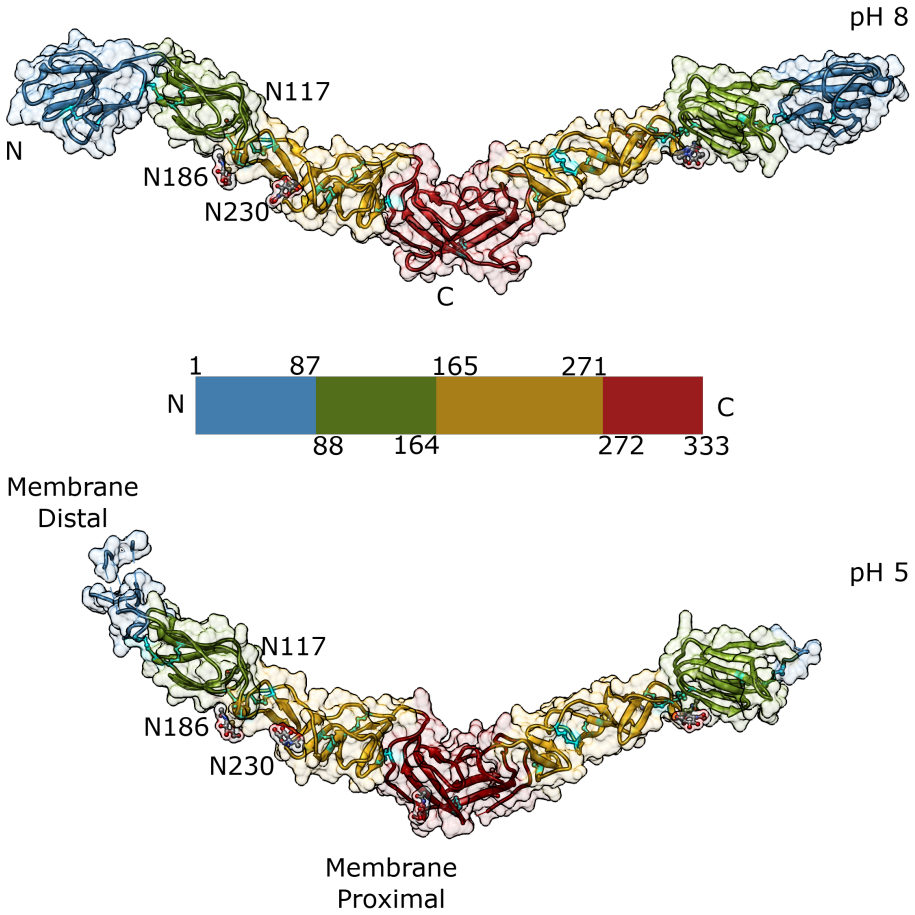


Figure 1.8: Structure of BVDV E2 (Adapted from El Omari *et al*[21]). Domain DA is shown in blue, DB in green, DC in yellow, and DD in red. High pH structure is on top (PDB code:2YQ2) and the low pH structure is on the bottom (PDB code: 2YQ3), in the middle is the domain boundaries. Glycosylated asparagines are labelled and cysteines, all involved in disulphide bonds, are shown in turquoise.

in contrast to the other domains[21, 22]. Other viruses have been shown to use a pH sensing histidine switch to expose a fusion peptide on a partner protein. An example would be the E2 protein of Sindbis virus becoming disordered and exposing the fusion peptide of E1[64].

Within the *Pestivirus* genus the sequence of the E2 envelope glycoprotein is well-conserved with the notable exception of APPV. In this virus, there is a large truncation at the N-terminus of the protein making it around 30kDa (Figure 1.9). In addition, there is much lower sequence identity in the parts which are present and several conserved disulphide bonds are missing[12]. This protein may prove

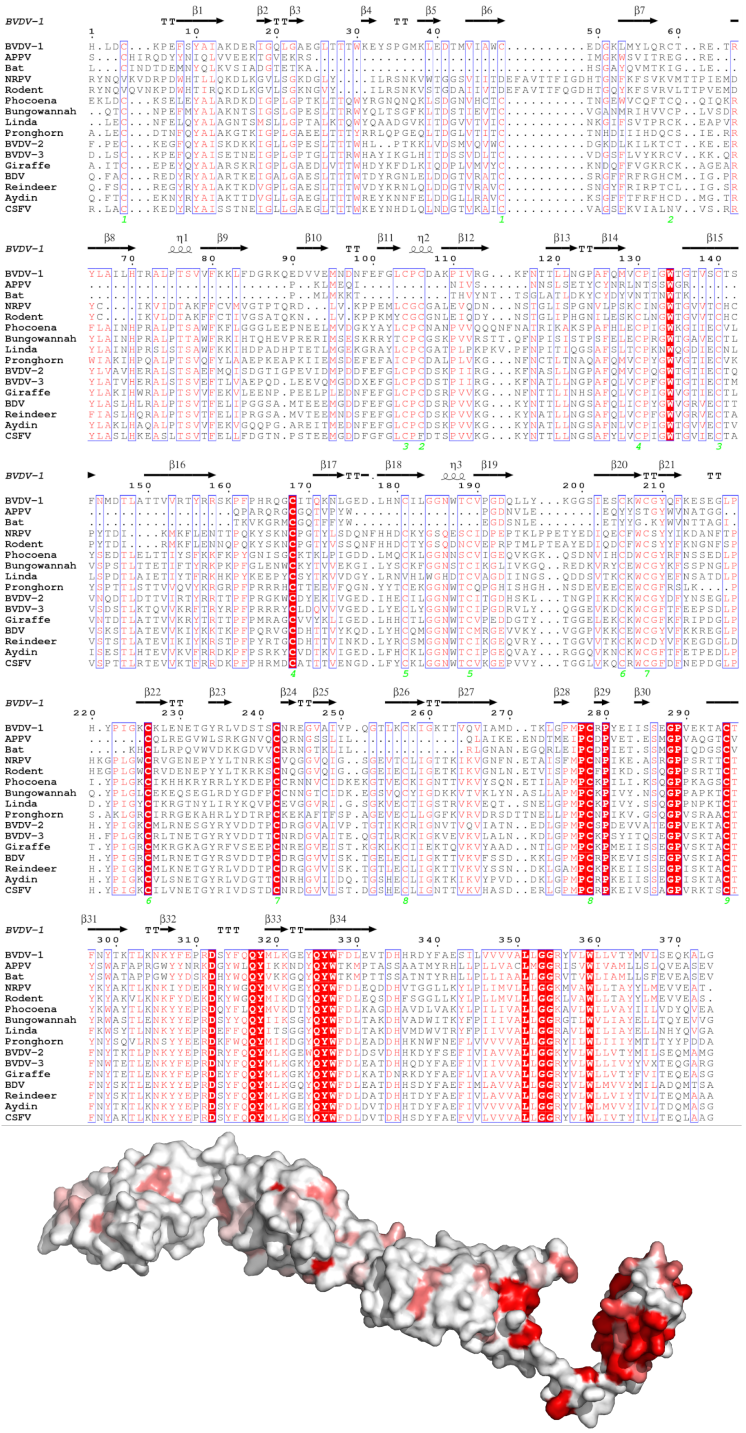


Figure 1.9: E2 Sequence Alignment. **Top-** Multiple sequence alignment of the E2 protein of the pestivirus polyprotein generated using Clustal Omega[4]. Sequence annotation was generated using the ESPript web service[52]. Conserved residues are highlighted in red while similar residues have a red type colour and are surrounded by a blue outline. The secondary structure assignment for BVDV-1 is shown on top of the sequence. **Bottom-** Surface E2 structure at pH 8 (PDB code: 2YQ2) coloured as in alignment. GenBank accession numbers that were used for alignments can be found in appendix A

extremely useful in elucidating further information on the Pestivirus mechanism of cell entry and membrane fusion as the deleted N-terminus contains the region currently though to bind the cellular receptor.

1.3.4 Dimerization of Pestivirus Glycoproteins

In 1991 when studying the composition of the virions of CSFV (then called hog cholera virus), it was discovered that all 3 envelope glycoproteins exist as multimers. Both the Erns and E2 proteins are found as homodimers of around 110kDa, however E2 is also seen as a E1-E2 heterodimer which is slightly smaller at around 75kDa[46]. The homodimers of Erns and E2 have been confirmed with crystal structures providing details of the disulphide bonds stabilising them. Information on the disulphide bonds supporting the E1-E2 heterodimer is provided through computationally generated structural models and point mutation experiments[49].

Erns homodimerization has been shown to hinge on a single cysteine C171, mutation of which in CSFV prevents the dimerization[65]. C171 in CSFV is also necessary for binding of the virus to the putative receptor heparin sulphate and has an impact on virulence. This suggests that the homodimer is necessary for interaction with the receptor. The other 8 cysteines present in the monomer form 4 disulphide bonds, all intramolecular and therefore unlikely to be involved with homodimerization. As these 8 cysteines are conserved throughout all the Pestivirus species, they are likely important for correct protein folding or protein function. Surprisingly, C171 in CSFV is not conserved in other Pestivirus species suggesting there must be a different method of homodimerization in those species missing a cysteine residue in this position[49].

E2 homodimerisation appears to occur through the C-terminal most domain DD. There are 3 cysteine residues, C256, C277, and C294, in classical swine fever virus which have been implicated in homodimerization[66]. Interestingly the bovine viral diarrhoea virus E2 structures revealed that only a single cysteine, C294 was involved in an intermolecular disulphide bond[21, 22]. The crystal structure of the

BVDVE2 ectodomain showed that in addition to the intermolecular disulphide bond between C294 the homodimer also involves a domain swap between DD domain monomers. This domain contains a β-hairpin connected to the next domain DC by an extended loop forming a hydrophobic pocket. The two domains interact through hydrophobic areas composed of tyrosines and phenylalanines[21, 22]. Although the C-terminal domain swap is seen in 2 crystal structures it has not been biologically verified and therefore it is not clear if it is biologically relevant.

The final type of multimer formed from Pestivirus envelope glycoproteins is the E1-E2 heterodimers. The current hypothesis for formation of E1-E2 heterodimers is that they mostly occur through the transmembrane anchors as well as potentially an intermolecular disulphide bond between C295 of E2 and C171 of E1. It has also been suggested that the charged residues encourage the formation of the E2 C295, E1 C171 intermolecular disulphide bond which itself stabilises the heterodimer[14]. The charged residues in the transmembrane anchor that may be involved in an interaction are R355 of E2 and K174 and R177 in E1[61]. There are a total of 6 cysteine residues in the E1 protein, in CSFV two of these C24 and C94 have been shown to affect the formation of the E1-E2 heterodimers as well as virulence. However, there is not currently a hypothesis for how this may happen and there is a very real possibility that it may just be caused by improper folding of the protein.

1.3.5 Glycosylation of Erns, E1, and E2 Proteins

As is usual with proteins found on the outside of a viral envelope, Erns, E1 and E2 are glycoproteins. The main type of glycosylation of Pestivirus glycoproteins is N-linked glycosylation, meaning asparagine residues located in a specific environment (Asn-X-Ser/Thr) are modified to have a large branched carbohydrate molecule attached. These carbohydrate molecules are around 2.5 kDa each and consist of N-Acetylglucosamine and mannose residues, sialic acid, galactose, and fructose residues can also be incorporated[67]. All 3 glycoproteins contain glycosylation sites that are highly conserved between the different species. For CSFV: Erns has 7 putative sites, E1 has 3 putative sites, and E2 has 6 putative sites[68–70].

However, it has been shown that BVDV E2 is only glycosylated at N117 (domain DB), N186 (domain DC), N230 (domain DC), and N298 (domain DD) indicating that not all putative sites are used in reality[21].

Several studies have shown that removal of glycosylation sites in Erns, E1, or E2 weakens the virus and reduces its virulence[69, 70]. This may be because the glycosylation state plays a role in virus entry or exit from the host cell or even viral attachment. The influence of glycosylation on Erns function has been studied more than for the other Pestivirus glycoproteins. Studies have implicated glycosylation state of Erns on protein synthesis, dimer formation, processing, and protein secretion[49, 68, 71]. A study also showed that the correct glycosylation is necessary to enable Erns to bind RNA and stop the induction of interferon[72].

The sugar molecules added during the glycosylation modification are large (around 2.5kDa each[67]) and can cover to a large portion of the protein they are attached to. This is important when thinking about the immune response of the host animal. It has been shown that Erns and E2 without proper glycosylation patterns are non-immunogenic proteins and therefore host animals do not develop protection to the wild type virus (although it was not confirmed if the deglycosylated proteins were properly folded)[73].

Looking at the antigenic structure and antigenic epitopes of the glycoproteins, it is possible to elucidate which regions of the protein are exposed. Inversely, it is also possible to see which regions are hidden either because they are folded on the inside of the protein or hidden by glycosylation. Currently, several antigenic peptides have been discovered for Erns, some are located between residues 55-110. The E2 protein is the most immunogenic of all the envelope glycoproteins suggesting that it may be most exposed on the surface of the virus particle. The antigenic epitopes are mostly well-conserved in domains DA and DD but not in DB and DC[49].

1.3.6 Occupancy of Glycoproteins on the Viral Surface

In a study looking at the morphology and molecular composition of the pestiviruses, there was very little E1 and E2 protein found in comparison to the amount of Erns. Supporting this was the fact that neither E1 nor E2 protein could be visualised on the surface of the particles under cryogenic electron microscopy (Cryo-EM) despite both being confirmed as present[25]. It is possible that the E1 and E2 glycoproteins lie flat on the virion surface causing them to be undetectable on the electron micrographs.

1.4 Pestivirus Cell Entry

1.4.1 How Might Pestivirus Membrane Fusion be Different

The E2 envelope glycoproteins from the *Pestivirus* and *Hepacivirus* genra do not have class II fusion protein folds as was initially expected based on the structures of envelope glycoproteins from other members of the family *Flaviviridae*[21, 22]. Not only does the E2 envelope glycoprotein not have the expected fold, it also has an entirely novel fold, as does the envelope glycoprotein E2 from HCV[74]. Neither structure contains an α-helical core, as would be necessary for a class I or III fusion protein, nor do they have a flexible multidomain structure as is the case for class II and III fusion proteins. Most importantly, these crystal structures showed that there was no fusion loop or peptide present with any resemblance to those currently discovered in any family. These results together suggest that the E1 envelope glycoprotein not E2 is the functional fusion protein[57] or perhaps the E1:E2 heterodimer.

Initially, E2 was thought to be the fusion protein as E1 is surprisingly small for a fusion protein at only around 180 amino acids long. Additionally, in the *Flavivirus* genus which is from the same family as the *Pestivirus* genus, the larger E protein rather than the smaller prM is the fusion protein[43]. However, solution of the E2 structures and their lack of fusion peptide like motif left E1 as the most likely candidate for the fusion protein. Supporting this is the discovery of

proposed fusion peptide in the E1 protein of HCV[75], a closely related virus from the *Flaviviridae* family.

Although the E1 envelope glycoprotein is currently the most well supported candidate for the Pestivirus fusion protein, E2 is also necessary for membrane fusion to take place. Multiple studies have shown that virus particles lacking an E2 glycoprotein or those in which the E1-E2 heterodimer has been prevented from forming are not fusogenic[61]. Additionally, it is possible that E1 cannot fold properly without the presence of E2 as is the case for HCV[76]. Therefore the current working theory within the field is that E1 is the functional fusion protein with E2 acting as a chaperon or scaffold for E1, this is in addition to its hypothesised role in receptor binding[57].

1.4.2 Method of Cell Entry

As stated previously when enveloped viruses enter a cell, they must fuse their membrane with that of the host cell. This membrane fusion takes place in one of two places: either it happens on the cell surface in a pH independent matter, or it can happen in the low pH environment of a late endosome following endocytosis of the virus particle. Like other members of the *Flaviviridae* family, pestiviruses are thought to enter the host cell using pH-triggered membrane fusion within an endosome after receptor facilitated endocytosis[77]. A clathrin-mediated endocytosis route for cell entry has been shown for flaviviruses, another member of the *Flaviviridae* family[78]. The E glycoprotein of flaviviruses is susceptible to acid treatment and changes from the prefusion dimeric form to the fusogenic trimeric form[79]. This acid treatment is sufficient to trigger membrane fusion. However, the story is not quite as simple with pestiviruses as two separate studies discovered that BVDV is resistant to acidic treatment, this contrasts with most other viruses which utilise pH-dependent entry[77]. It is not uncommon for different species within a family to have slightly different methods of cell entry: for the coronaviruses infectious bronchitis virus needs only low pH to trigger membrane fusion[80] while severe acute

respiratory syndrome coronavirus need to both bind to a receptor and undergo proteolysis before membrane fusion can take place[81].

When considering enveloped viruses which enter the cell through an endocytic route, there are three main methods: clathrin-mediated endocytosis, micropinocytosis, and clathrin- and caveolin-independent endocytosis. In addition to the low pH in endosomes, some viruses have specific membrane lipid requirements or other factors necessary to trigger membrane fusion. A study in 2004[82] showed that BVDV enters the host cell through clathrin-mediated endocytosis by showing BVDV co-localised with clathrin and that disruption of the clathrin mediated pathway prevented infection while disruption of the caveolin-dependent pathway did not. The clathrin-mediated endocytosis method of cell entry has been confirmed for BVDV by several other studies since[83, 84]. A similar dynein dependent clathrin-mediated endocytic route for cell entry has also been proposed for CSFV and likely also the case for all pestiviruses[85].

During the 2005 study[83] on the cell entry mechanism, it was confirmed that BVDV infection was sensitive to low pH, as is the case in late endosomes. BVDV E2 glycoprotein has been predicted to also change conformation at low pH with its N-terminal most domain becoming disordered[21]. However, low pH is not sufficient to triggered membrane fusion. This apparent acid resistance of BVDV may be due to a missing receptor being required to initiate membrane fusion after the pH trigger[83]. It may also be due to some other mechanism which destabilises the glycoprotein perhaps through alteration of the disulphide bonds making it fusogenic at low pH. This is supported by the discovery that treatment of BVDV with a reducing agent to disrupt the disulphide bonds in combination with low pH allowed a small amount of fusion at the cell membrane[77].

1.4.3 Cellular Receptors for Pestiviruses

There is currently little known about the cellular receptors for different *Pestivirus* species. Only CSFV and BVDV having proposed receptors and even for these the

whole story is not yet apparent. Both the E2 and the Erns envelope glycoproteins interact with cell surface molecules such as cellular receptors and glycosaminoglycans. Current knowledge suggests that Erns makes the first contact with the cell through interactions with cell surface heparan sulphates. Heparan sulphate is a membrane associated glycosaminoglycan, an unbranched polysaccharide chain linked to a protein core, which forms a network covering the surface of most cells and is utilised as initial attachment factors by many viruses[86]. This interaction has been shown for both CSFV Erns and BVDV Erns. The next step thought to take place is the E2 glycoprotein interacting with either porcine CD46 or bovine CD46. This is then followed by clathrin-dependent endocytosis and eventually membrane fusion[87].

There are several issues with the current scientific knowledge about Pestivirus receptors which have not been addressed over the last 21 years since the first studies came out. The first and arguable most worrying of which is the fact that for none of the proposed receptors, other than the recently discovered A disintegrin and metalloprotease 17 (ADAM17), has direct binding been shown, this is especially troubling for the proposed interaction with CD46[62, 87–89]. It is frequently assumed in the literature that the E2 glycoprotein binds to cellular CD46, however a direct interaction had never been shown and so future studies may be designed over a false assumption. Secondly, the state of receptor knowledge is particularly cloudy for CSFV. In 2015, two studies were published about CSFV receptors, one presents evidence to support the hypothesis that porcine CD46 is a cellular receptor for CSFV[87] while the other suggests that CD46 is not a receptor and proposes the laminin receptor instead[90]. Most recently in 2021, a study has been published presenting evidence that porcine CD46 is a receptor for APPV but not CSFV[88]. A second possible receptor, A disintegrin and metalloproteinase 17 (ADAM17), has been identified for CSFV[91]. The final issue present is that all studies presenting putative receptors are unable to completely abolish susceptibility to virus by removing their proposed receptor.

1.4.4 CD46 as a Cellular Receptor for Bovine Viral Diarrhoea Virus

As discussed above bovine CD46 has been suggested as a cellular receptor for BVDV, indeed several studies have demonstrated that functional bovine CD46 is necessary for BVDV infection to take place[62, 89]. CD46 is a regulator of the complement cascade and is also known as membrane cofactor protein. It is a glycoprotein found on the surface of many eukaryotic cells with a type 1 transmembrane anchor. Along with serine protease I, it mediates the inactivation of C3b and C4b[92], many disease associated mutations in this protein have been found[93]. The structure of the extracellular domain of human CD46 has been solved[94], and given the high sequence identity of the human and bovine proteins bovine CD46 is likely to have a highly similar structure. The approximately 50kDa protein forms an elongated antenna-like structure with an unstructured N-terminal portion, 4 short consensus repeats (SCR) termed SCR1-SCR4 followed by a C-terminal membrane anchor (Figure 1.10). The overall structure of the extracellular portion has been compared to a hockey stick with the first 3 short consensus repeats SCR1-SCR3 arranged linearly and the last SCR4 at an angle to the others, due to a large bend in the region between SCR3 and SCR4[94].

In 2004, Maurer et al[89] purified and identified bovCD46 from the surface of MDBK cells as being recognised by monoclonal antibodies(MAb) which inhibit BVDV infection. However, expression of bovCD46 does not confer infectivity to completely unsusceptible cell lines. It did however increase the susceptibility of porcine cells to BVDV 90-fold, almost making them as susceptible as bovine cells. They also managed to show that BVDV binds to cells which express bovCD46 even if those cells were not susceptible to BVDV infection. Further study[62] revealed that the N-terminal domain SCR1 (the domain furthest from the membrane) is essential in BVDV binding and subsequent infection. Two peptide sequences within this SCR which map to a single β-sheet were shown to be critical. By exchanging these peptides between the porcine and bovine versions the authors were able to

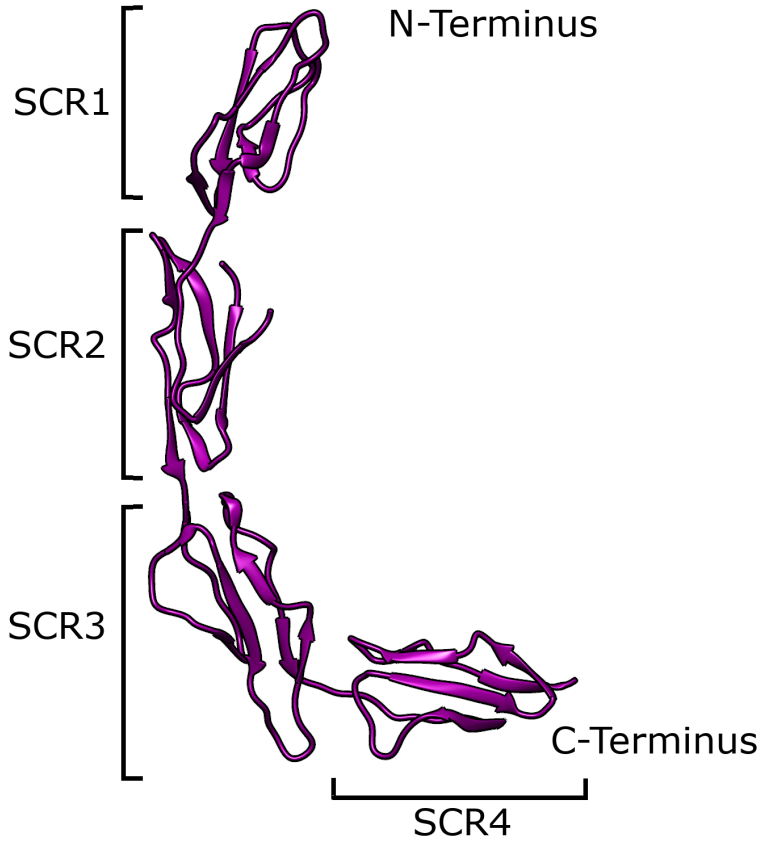


Figure 1.10: Cartoon depiction of the structure of the human CD46 ectodomain (PDB code: 3O8E) with N-, and C-termini marked, and domains labelled

reduce infection in bovine cells and increase infection in porcine cells. Finally, the number of SCRs present was shown to be important for infection, insertion of up to 6 SCRs (extending the distance between the putative binding site and the cell membrane) had very little effect on BVDV infectivity. However, removal of any SCRs (reducing the distance between the putative binding site and the cell membrane) massively reduced BVDV susceptibility[62].

Recently a group has been able to fluorescently label BVDV E2 envelope glycoprotein and the CD46 cellular receptor and therefore use this to study the BVDV virus particles in real time during the early stages of infection[95], this has enabled several interesting discoveries. Firstly, over half of all entry events were associated with a fluorescently labelled CD46 receptor vesicular structure, and the fluorescently labelled BVDV E2 remained associated with the vesicle

after endocytosis in most cases. Secondly, removal of the putative BVDV binding domain on CD46 did not change the movement of the virus on the surface of the cell but did dramatically decrease the number of entry events and the particles had an increased velocity. Furthermore, it was seen that 4 times more virus particles moved away from the cell if CD46 lacked the putative binding domain. Finally, it was shown that BVDV entry does not happen quickly with most entry events happening more than 20 minutes after addition of the virus and only 3% of tracked virus particles had a recorded entry event. Overall, the study suggested the involvement of CD46 in reducing the speed of virus particles on the cell surface and potentially the anchoring of particles at sites of potential endocytosis. They also reiterate that BVDV can infect cells in the absence of CD46 so it is likely another cellular receptor is involved[95].

1.4.5 Cellular Receptors for Porcine Pestiviruses

Four cellular receptors/attachment factors have been proposed for porcine pestiviruses: heparin sulphate[87, 96], the laminin receptor (LamR)[90], CD46[87, 88], and ADAM17[91]. The first three of these molecules have been shown as receptors for many different viruses while ADAM17 is only implicated as an attachment factor for CSFV. Both heparin sulphate and the laminin receptor have been proposed as initial attachment factors as they have never been shown to be able to mediate virus entry to a host cell. Both heparin sulphate and the laminin receptor have been suggested to interact with the loosely associated Erns envelope glycoprotein, in keeping with them being attachment factors rather than cellular receptors which mediate cell entry[90, 96].

Expression of porcine LamR did not confer infectivity to non-susceptible cell lines but did increase infectivity in cell lines where heparin sulphate had been removed. Removal of both heparin sulphate and LamR CSFV resulted in almost completely abolishment of CSFV replication. Heparin sulphate has been shown to have a higher affinity for CSFV than LamR and removal of LamR does not abolish CSFV infectivity suggesting that LamR is an alternative attachment factor

to heparin sulphate[90]. This is supported by a study which showed that CSFV was able to propagate in the absence of heparin sulphate, and could adapt to the absence of heparin sulphate in a limited number of passages[97]. Although both LamR and heparin sulphate are involved at the beginning of cell entry neither one is the cellular receptor necessary to initiate the cell entry process.

In early 2021, ADAM17 was identified as an essential attachment factor for CSFV[91]. This study demonstrated that knockout of ADAM17 in a CSFV susceptible cell line eliminated viral entry, but this could be restored with the addition of ADAM17 cDNA. What was particularly exciting was the discovery of direct binding between CSFV E2 protein and the metalloproteinase domain of ADAM17. A 45 amino-acid surface exposed region of the metalloproteinase domain of ADAM17 was identified as being essential for CSFV entry. This is the first-time direct binding has been identified between a Pestivirus E2 protein and a cellular receptor.

A proposed cellular receptor for both CSFV and APPV is porcine CD46, this makes sense as bovine CD46 is the proposed receptor for BVDV[89]. Blocking of porcine CD46 has been shown to reduce infectivity of CSFV and infectivity is further reduced when heparin sulphate is also removed. However, infectivity is not abolished so other factors are likely involved[87]. It is worth noting that LamR has been proposed as an alternative pathway to heparin sulphate and the study did not test infectivity of CSFV on a cell line without functioning heparin sulphate, LamR, and CD46 so LamR may be the missing link[90]. The matter was further complicated by a study released early in 2021 which identified porcine CD46 as a receptor for APPV but not for CSFV[88].

1.4.6 Cell-Cell Transmission

Although the clathrin mediated endocytic method of cell entry has been verified as the initial method of infection it may not be the only route later on in the infection process[83, 85]. After the initial infection some viruses spread throughout the host

organism using another method of cell entry, direct cell-to-cell transmission. This has the advantage of hiding the virus from the host immune system and neutralising antibodies as the virus no-longer has to diffuse through the extra cellular space to find its next host cell. Direct cell-to-cell transmission is also much faster and more efficient as three of the longest steps in the virus life cycle, viral particle release, diffusion, and cell entry, can happen simultaneously at cell-cell contact sites. This method of direct cell-to-cell transmission has been confirmed for many animal viruses, in fact some even induce the formation of cell-cell contact sites[98].

In 2007 it was discovered that HCV, another member of the *Flaviviridae* family, could spread through direct cell-cell transmission[99]. However, it was not until over 10 years later that the first reports on Pestivirus direct cell-to-cell transmission were published with the discovery that BVDV could propagate through this route in cell culture even in the presence of neutralising antibodies[100]. In fact, it was shown that antibody resistant (and therefore likely cell-to-cell) transmission was the main mechanism of infection spread in cell culture.

The overall mechanism suggested for cell-to-cell transmission of pestiviruses, particularly BVDV, involves the virus particles leaving the cell in hidden exocytic vesicles. They then accumulate in the extracellular space next to cell-to-cell contact sites before entering the new cell through receptor mediated clathrin-dependent endocytosis as in the initial infection model. Interestingly, the putative BVDV receptor CD46 was unnecessary for the cell-to-cell transmission, a discovery supported by the fact that attachment receptors for HCV were superfluous for cell-to-cell transmission in this closely related virus as well[101] (although a more recent study contests this[102]). In HCV, the post attachment receptors were needed for cell-to-cell transmission and cell-free transmission, so there may be a yet undiscovered post attachment or co receptor for BVDV which is used in cell-to-cell transmission[102]. The BVDV envelope glycoprotein E2 is required for cell-to-cell transmission as it is in cell-free transmission[100].

BVDV along with other pestiviruses can infect hosts *in utero* which requires the virus to cross the placental barrier to reach the foetus. The identification of BVDV in cells on either side of the placental barrier, PBMC and trophoblasts, suggest that Pestiviruses could use cell-to-cell transmission to cross to the foetus unnoticed by passing directly from infected peripheral blood mononuclear cells to nearby trophoblasts[100].

1.5 Project Aims and Overview

Species within the *Pestivirus* genus cause a variety of disease in many animals including several livestock species. Pestivirus infections have a great economic impact especially in the dairy and pig industries as well as being a source of waste in the food chain. Currently, very little is known about the method of cell entry used by pestiviruses. However, they don't appear to follow the same method of cell entry as other species from their family, *Flaviviridae*, and could use a completely novel method of membrane fusion. The current lack of knowledge in the method of cell entry is a source of difficulty in the development of drugs and vaccines. This project combines protein engineering and structural techniques to provide further insight into the mechanism of host cell entry of different species within the *Pestivirus* genus.

Chapter 2 describes the methods used to design of over 80 protein constructs and the cloning techniques used to create them. The expression and purification methods used to produce recombinant proteins and to crystallise them are described, as is the collection of data at different beamlines at Diamond Light Source and the methods used to solve and refine the protein structures, and finally, small angle X-ray scattering (SAXS) and dynamic light scattering (DLS) experiments used to elucidate further information on the overall mechanism of cell entry of pestiviruses.

Chapter 3 (Protein Production) presents the use of medium throughput cloning and expression testing techniques to test many protein constructs and identify those which could be efficiently produced. Also described is the use of the HEK293S cell line to help with the glycosylation issue and the lentivirus system of cell transduction

to quickly produce stable cell lines. Finally, a simple protein purification procedure is described which enabled further studies to be started as soon as 24 hours after protein harvest.

Chapter 4 (Structural Studies of Putative Receptors) describes the production of protein crystals of two putative pestivirus receptors bovine, and porcine CD46 and the collection of high-resolution datasets. Two structures of the CD46 cellular receptor extracellular domain solved by molecular replacement are presented alongside investigations into a potential biological dimer.

Chapter 5 (Structural Studies of Pestivirus E2 Glycoproteins) presents the generation of protein crystals for four Pestivirus glycoproteins and the use of various crystal optimisation techniques. Multiple beamlines at Diamond Light Source were utilised to collect both high-resolution and long-wavelength datasets on the above-mentioned crystals. The structures of APPV E2 and NRPV E2 extracellular domains solved by sulphur SAD phasing are described. Finally, I present small angle X-ray scattering (SAXS) data on several pestivirus E2 ectodomains with a discussion on pH based structural changes.

Overall, this thesis presents for the first time the structures of the bovCD46 extracellular domain, the structures of APPV E2 extracellular domain at both high and low pH and the structure of NRPV E2.

2
Materials and Methods

Contents

2.1 Cloning **32**
2.2 Mammalian Cell Work **38**
2.3 Bacterial Cell Work **40**
2.4 Protein Expression **42**
2.5 Protein Purification **44**
2.6 Gel Electrophoresis **46**
2.7 Crystallography **48**
2.8 SAXS **51**
2.9 Dynamic Light Scattering **52**

2.1 Cloning

2.1.1 Expression Plasmids

The transfer plasmids: pHR-CMV-TetO$_2$-3C-Avi-His6, pHR-CMV-TetO$_2$-Avi-His6 _IRES-EmGFP, pHR-CMV-TetO$_2$-Avi-His6_IRES-mRuby2, and pHR-CMV-TetO2-Avi-His6 were donated by Dr Jonathan Elegheert. The envelope plasmid (pMD2.G) and packaging plasmid (psPAX2) were donated by Dr Benjamin Bishop. Together the packaging, envelope, and transfer plasmids were used to generate lentiviruses for expression of protein constructs in mammalian cells.

2.1.2 Gene Design

Sequences of the pestiviruses BVDV (*bovine viral diarrhea virus* NP_040937.1), BDV (*border disease virus* NP_620062.1), CSFV (*classical swine fever virus* NP_075354.1), APPV (*atypical porcine Pestivirus* YP_009268709.1) and NRPV (*Norway rat Pestivirus* YP_009109567) were aligned using the program Clustal Omega1[103]. The regions spanning the E1 and E2 proteins were identified either using the literature or estimated based on the alignment and position of the proteins in other pestiviruses. These E1-E2 regions were synthesised by Gene Universal for BVDV, BDV, CSFV, and APPV and by GeneArt (Thermo Fisher Scientific) for NRPV. The entire protein was ordered from GeneArt for Bovine CD46 and Porcine CD46 as well as the ectodomain of porcine ADAM17. The full gene sequences for all synthesised genes can be found in appendix B.

2.1.3 Primer Design

The Benchling platform[104] was used to virtually digest and ligate insert and vector sequences to generate a plasmid map of the finished product. The overall primer structure for each primer type is described below and all primer sequences used are shown in appendix C.

For restriction cloning forward primers were designed with a 6 nucleotides (nt) handle (sequence GGATCT) followed by the *AgeI* restriction site (sequence ACCGGT) finally there is a 18nt sequence identical to the first 18nt of the insert sequence. These 18nt were extended to the first G or C if the primer sequence did not end in one originally. The reverse primers were designed with a 6nt handle (sequence GGATCT) followed by the *KpnI* restriction site (sequence GGTACC). Finally, there is a 18nt sequence identical to the reverse complement of the final 18nt of the insert sequence. These 18nt were extended to the first G or C if the primer sequence did not end in one originally. Primers were ordered from either Eurofins Genomics as lipolyzed powder or subsequently from Sigma-Aldrich at 100µM in water. Details of each primer and the production company can be found in appendix C.

For single step cloning primers were designed around the restriction sites denoting either end of the insert. Using the plasmid map generated from Benchling[104] the forward primers were designed to have an identical sequence to the last 15nt of the vector before the insert, followed by the 6nt restriction site, followed by the first 15nt of the insert sequence. The reverse primers have a sequence identical to the reverse complement of the first 15nt of the vector following the insert, followed by the 6nt restriction sequence, followed by the reverse complement of the last 15nt of the insert sequence. Primers were ordered from Sigma-Aldrich at 100µM in water.

Primers for mutagenesis were designed to centre on the point of mutation. The forward mutagenesis primer had a sequence identical to the target sequence (after mutagenesis) of the 15nt before the change, the bases being changed, and the 15nt after them. The reverse primer had a sequence identical to the reverse complement of the forward primer. Other primers used during the mutagenesis process were designed previously to make the unmutated version of the construct.

Screening primers were designed to be used for both colony screening and sequencing. The forward primer has a sequence identical to an 18nt region at the end of the promoter. The reverse primer has a sequence identical to the reverse complement of an 18nt region downstream of the multiple cloning site.

2.1.4 Polymerase Chain Reaction Conditions

Inserts were generated using PCR and Hot-start KOD DNA polymerase (Novagen). Each tube was prepared with 0.5µl each of forward and reverse primer at 10pmol/µl and 0.5µl of template DNA a 10ng/µl. A PCR master mix was made up containing the following for each reaction: 1x KOD hot start polymerase buffer (Novagen), 1mM $MgCl_2$ (Novagen), 0.2mM of all four dNTPs (Invitrogen), 0.5µl KOD hot start DNA polymerase (Novagen), and water to a total volume of 23.5µl. 23.5µl of master mix was added to each tube and mixed before being placed in a GS4 thermocycler (G-storm) and the following program run: 5 minutes at 95°C degrees followed by 25 cycles of: 15 second 95°C, 5 seconds 52°C, and 10 seconds 72°C.

PCR products were stored in the thermocycler at 10°C before being visualised on an agarose gel (section 2.6.1). Positive hits were cleaned up using a Promega PCR cleanup kit and stored at -20°C.

2.1.5 Restriction Cloning

10 units of the enzymes *AgeI*-HF, *KpnI*-HF, and *DpnI* (all NEB) were used to digest the purified PCR products for 1 hour at 37°C in a heat block. 20 units of *AgeI*-HF and *KpnI*-HF (NEB) were used to digest 22µg of vector overnight at 37°C in a heat block. 5 units of Antarctic Phosphatase (NEB) was added to the digested vector and incubated at 37°C in a heat block for 1 hour. The digested vector was run on an agarose gel (section 2.6.1). Promega PCR clean-up kit was used to clean the digested PCR products and gel extract the cleaved vector. 200 units of T4 DNA ligase (Thermo Fisher Scientific) was used to insert 1µl of the PCR product into 0.5µl of the vector for 20 minutes at room temperature. DNA was transfected into Mach1 cells (Invitrogen) with a 30-minute incubation on ice followed by a 46 second heat shock in a 42°C water bath. Cells were incubated back on ice for 2 minutes before 100µl of Luria Broth (LB) was added and tubes were shaken at 37°C for 1 hour in a heat block. All 100µl was plated on an LB agar plate containing 100µg/ml carbenicillin (Melford) which were subsequently incubated overnight at 37°C.

2.1.6 Single Step Cloning

Single step cloning was used for constructs which were difficult to obtain through regular restriction cloning or those created after the procurement of a more affordable single step cloning kit, ClonExpress II, from Vazyme. 20 units of *AgeI*-HF and *KpnI*-HF (NEB) were used to digest 22µg of vector overnight at 37°C. The digested vector was run on a 1.5% agarose gel (section 2.6.1). A Promega PCR clean-up kit was used to gel extract the cleaved vector. The DNA recombination step was carried out in a 5µl reactions containing 1x CE II Buffer, 0.5µl Exnase II (both Vazyme), 200ng of linearized vector, and 40ng of PCR product. This mixture was well mixed, incubated at 37°C for 30 minutes then transferred to ice before

transformation. The whole 5µl of DNA mixture was transfected into Mach1 cells with a 30-minute incubation on ice followed by a 46 second heat shock in a 42°C water bath. Cells were incubated back on ice for 2 minutes before 100µl of LB was added and tubes were shaken at 37°C for 1 hour in a heat block. 100µl was plated on an LB agar plate containing 100µg/ml carbenicillin (Melford) which was subsequently incubated overnight at 37°C.

2.1.7 Movement of DNA Inserts from One Plasmid To Another

The original DNA construct was digested overnight with 10 units of *AgeI*-HF and *KpnI*-HF (NEB) in 1x cut smart buffer (NEB) at 37°C. The new vector was also digested overnight with 10 units of *AgeI*-HF and *KpnI*-HF (NEB) in 1x cutsmart buffer (NEB) at 37°C. The digested constructs were run on an agarose gel (section 2.6.1) and the bands corresponding to the insert were cut out and gel purified using a Promega PCR cleanup kit. The digested vector was treated with Antarctic phosphatase (NEB) for 1 hour at 37°C before being cleaned up with a Promega PCR cleanup kit. 200 units of T4 DNA ligase (Thermo Fisher Scientific) was used to insert 1µl of the PCR product into 0.5µl of the vector for 20 minutes at room temperature. DNA was transfected into Mach1 cells (Invitrogen) with a 30-minute incubation on ice followed by a 46 second heat shock in a 42°C water bath. Cells were incubated back on ice for 2 minutes before 100µl of LB was added and tubes were shaken at 37 °C for 1 hour. 100µl was plated on an LB agar plate containing 100µg/ml carbenicillin which was subsequently incubated overnight at 37°C.

2.1.8 Site Directed Mutagenesis

A primer extension PCR method was used to carry out point mutations. The method used is described in detail in Methods in Molecular Biology[105]. Briefly: forward and reverse primers were designed to cover the mutated region. These along with primers for the entire fragment were used to create 2 PCR products both containing

the mutation. These PCR products were then used as a template in a second short round of PCR with the entire fragment primers to create the final product.

For the first round of PCR two reactions were set up: one included the forward construct primer and the reverse mutagenic primer, the other contained the forward mutagenic primer and the reverse construct primer, both used the synthetic gene as template DNA. The reactions each contained 0.26ng/ml of template DNA, 0.3µM of each primer, 1X Herculase II buffer (Agilent), 0.32mM dNTP mixture (Invitrogen), 0.26µl of Herculase II fusion polymerase (Agilent) with a total reaction volume of 24µl. The tubes were placed in a GS-4 thermocycler (G-storm) and the following program run: 10 minutes at 95°C followed by 25 cycles of: 30 seconds 95°C, 30 seconds 62°C, 1 minute at 68°C, after cycles 68°C for 5 min. The products were visualised on an agarose gel (section 2.6.1), positive hits were then digested individually with 10 units of *DpnI* (NEB) in 1x cutsmart buffer (NEB) at 37°C for 1 hour.

The second round of PCR used the construct forward and reverse primers and used the digested PCR products as the DNA template. The reaction contained 1.5µl of each of the digested PCR products, 0.3µM of each primer, 1X Herculase II buffer (Agilent), 0.32mM dNTP mixture (Invitrogen), 0.26µl of Herculase II fusion polymerase (Agilent) with a total reaction volume of 24µl. The tubes were placed in a GS4 thermocycler (G-Storm) and the following program was run: 10 minutes at 95°C followed by 10 cycles of: 30 seconds 95°C, 30 seconds 62°C, and 1 minute at 68°C, after cycles 68°C for 5 minutes.

200 units of T4 DNA ligase (Thermo Fisher Scientific) was used to insert 1µl of the PCR product into 0.5µl of the vector for 20 minutes at room temperature. DNA was transfected into Mach1 cells (Invitrogen) with a 30-minute incubation on ice followed by a 46 second heat shock in a 42°C water bath. Cells were incubated back on ice for 2 minutes before 100µl of LB was added and tubes were shaken at 37°C for 1 hour. 100µl was plated on an LB agar plate containing 100µg/ml carbenicillin (Melford) which was subsequently incubated overnight at 37°C.

2.1.9 Colony Screening

Single colonies were screened for the presence of the correct insert. A master mix was made containing the following per colony to be screened: 10.0µl of SapphireAmp fast PCR - hot-start master mix (Takara), 0.5µl of both forward and reverse screening primers at 10µM, 0.6µl DMSO, and 8.4µl water. 20.0µl of this mix was placed into PCR tubes, an equal number of wells in a 2ml deep well block were filled with 1ml of LB containing 100µg/ml of carbenicillin (Melford). A small 1µl loop was used to harvest a single colony, the loop was then placed into the appropriate PCR reaction tube and twirled against the bottom of the tube 10 times. The loop was then placed into the appropriate well of the 2ml deep well block. The PCR tubes were sealed and placed into a GS4 thermo cycler (G-Storm) and the following program was used: 95°C for 5 minutes, 25 cycles of (95°C for 30 seconds, 52°C for 30 seconds, and 72°C for 1 minute) finally 72°C for 5 minutes. The deep well block was incubated overnight at 37°C. The PCR products were visualised on an agarose gel (section 2.6.1) and positive hits were cleaned up using a Promega PCR purification kit. These pure PCR products were sent for Sanger sequencing with Source Bioscience using the forward screening primer to confirm correct sequence. If sequencing was not complete it was repeated using the reverse screening primer. Positive colonies which had been sequence verified were grown overnight by placing 100µl of the culture from the appropriate well of the 2ml deep well block into 5ml of LB containing 100µg/ml cabenicillin (Melford) in a 50ml Facon tube. These were shaken overnight in an innova 44 incubator (New Brunswick) at 200rpm 37°C. The cultures were then subsequently miniprepped using a Qiagen QIAprep Spin Miniprep Kit.

2.2 Mammalian Cell Work

2.2.1 Cell Culture

All cells were split 1 in 6 once the reached approximately 80% confluency (determined by eye under 10x magnification), all work was carried out in a laminar flow hood using sterile technique. To split 25ml for T175 flask (12ml for T75) of old media was

removed and disposed of. The cells were washed with 8ml(4ml) of phosphate buffered saline (PBS) which was also subsequently removed and disposed of. 3ml(1.5ml) of trypsin EDTA(Gibco) was added and the flasks rocked to ensure all cells were covered, the flask was then incubated in the hood for 5 minutes. The cells were then detached from the flask by vigorous tapping and rocking of the flask. Once all cells are detached 26ml(12ml) of 10% media (DMEM (sigma) with 1% non-essential amino acids (NEAA), 1% L-glutamine, and 10% foetal bovine serum (FBS), all from Gibco) was added to the flask and pipetted up and down until no further cell clumps could be seen. New flasks were prepared with 21ml(10ml) of fresh 10% media and 5ml(2ml) of the resuspended cell culture was added. Flask was swirled to disperse cells evenly and ensure the entire floor of flask was covered.

While cells were not actively in use, they were frozen (see section 2.2.3) and stored in the nitrogen vapor phase.

To prepare 6 well plates an 80% confluent T175 flask was split into 6 six well plates inside a laminar flow hood using sterile technique. The process follows the above to the end of the cell detachment stage after which the resuspended cells are topped up to 73ml with 10% media. 2ml of the diluted cell stock was added to each well of the 6 plates. The plates were rocked to ensure the entire surface of each well was covered; the plates were placed in a 37°C incubator. Approximately 80% confluency was reached after around 24 hours.

2.2.2 HEK293S, LentiX

Initial HEK293S, HEK293T, and LentiX cells were kindly donated by Dr Benjamin Bishop and Weixian Lu.

Cell were maintained at 37°C and 5% CO2, HEK293S and T cells were maintained in T175 flasks and LentiX cells were maintained in T75 flasks. Growth media (10%) was prepared as follows: 500ml of DMEM from Sigma with 5ml of non-essential amino acids (NEAA), 5ml L-glutamine, and 50ml of FBS all from Gibco.

Every 10 passages the HEK293S cells were cultured in the presence of 2µg/ml blasticidin.

2.2.3 Freezing and Thawing Cells

Cells were prepared for vapor phase nitrogen storage following detachment from flask using trypsin EDTA. 10ml of 2% media (as 10% but with 2% FBS) was added and cells pipetted up and down until no further cellular clumps were observed. Cells were transferred to a 15ml tube and centrifuged at 1000rpm for 5 minutes at room temperature. After centrifugation the media was poured off and the cells resuspended to a total volume of 3ml in 2% media supplemented with 10% DMSO. 1ml was placed in labelled cryo vials which were sealed and placed in a coolcell freezing system (Corning). The coolcell freezing system was placed in a -80°C freezer overnight before the cryo vials were transferred to vapour phase nitrogen for long term storage.

To revive cells from nitrogen storage vials were quickly defrosted in room temperature water. As soon as the cells were defrosted, they were transferred to a 15ml tube containing 14ml of 10% media and centrifuged at 1000rpm for 5 minutes at room temperature. The media was poured off and the cells resuspended in 5ml of 10% media. New T175 flasks were prepared with 20ml of fresh 10% media and all 5ml of the resuspended cells was added. The flask was swirled to disperse cells evenly and ensure the entire floor of flask was covered. Once cells reached approximately 80% confluency (4-5days) normal cell passage (as described in section 2.2.1) was started.

2.3 Bacterial Cell Work

2.3.1 Preparation of Competent Cells

Mach1 cells (Invitrogen) were streak plated on to a LB agar plate containing 100g/ml carbenicillin (Melford) and incubated overnight at 37°C. Five 50ml falcon tubes were prepared with 5ml of LB containing 100µg/ml of carbenicillin (Melford), each was inoculated with a single colony from the overnight growth. These tubes were

shaken at 200rpm in a Innova 44 incubator (New Brunswick) at 37°C for 5 hours. Tubes were removed from the incubator and chilled on ice for 10 minutes while an Eppendorf 5810 R benchtop centrifuge was precooled to 4°C. Meanwhile 10ml of 0.1M $CaCl_2$ and 5ml of 0.1M $CaCl_2$ with 15% glycerol were filter sterilised and placed on ice to chill. The tubes were subsequently centrifuged in the precooled benchtop centrifuge at 4000rpm for 3 minutes. After centrifugation the media was poured off and each tube of cells was resuspended by pipetting in (extremely gently) 2ml of the ice cold 0.1M $CaCl_2$. All 5 tubes were combined, and the resuspended cells placed on ice for 20 minutes. The cells were again pelleted by centrifugation for 3 minutes at 4000rpm and 4°C. The supernatant was removed and the cells gently resuspended (by pipetting) in the 5ml of 0.1M CaCl2 with 15% glycerol. The resuspended cells were aliquoted into single tubes (50µl/tube) and flash frozen in liquid nitrogen. They were then stored in a -80°C freezer.

After preparation cells were checked for susceptibility to carbenicillin and for transformation efficiency.

2.3.2 Transformation of Competent Cells

A single tube of competent cells per transformation were removed from the freezer and placed on ice for 5-10 minutes to defrost. 1µl of plasmid or 5µl of a ligation or infusion reaction were placed in a 1.5ml microtube with 50µl of competent cells and incubated on ice for 20 minutes. The tubes were placed in a 42°C water bath for 46 seconds to heat shock, and then placed back on ice immediately. After a 2-minute incubation on ice 100µl of antibiotic free LB was added to each tube and they were placed in a 37°C heat block shaking at 600rpm for 1 hour. Meanwhile the appropriate number of LB agar plates containing 100µg/ml of carbenicillin (Melford) were prepared. After the 1 hour was up, 100µl of the transformation mixture was placed onto each plate and spread evenly across the surface using a fresh disposable spreader for each plate. The plates were left at room temperature for 10 minutes to dry out then they were covered and placed upside down in a static 37°C incubator overnight.

2.4 Protein Expression

2.4.1 Lentivirus Production

First DNA was prepared for transfection, all work was carried out in a laminar flow hood using sterile technique. For each construct a 15ml tube containing 1.5ml of Serum free media (SFM: as 10% media but lacking FBS) had 10µg each of the packing plasmid, envelope plasmid, and the appropriate insert plasmid added. Each tube then had 1.5ml of SFM containing 60µg of polyethylenimine (PEI) added and was mixed well. The tubes were incubated at room temperature for 10 minutes. A T75 flask containing lentiX cells at around 80% confluency had the media removed and replaced with 9ml of 2% media. After the 10 minutes incubation had finished the 3ml of DNA/SFM/PEI mixture was gently added to the flask so the cells are not disturbed. The cells were then placed in a 5% CO_2 37°C incubator for 4-5 days on the designated virus shelf.

After this time the virus was harvested, all work was carried out in a laminar flow hood using sterile technique and ensuring no virus was released. The media from the T75 of transfected LentiX cells was removed with a 10ml stripette and transferred to a leur lock syringe fitted with a 0.45µm filter. The stripette was resheathed and placed into the biocontamination bin. The media was filtered into a 50ml falcon tube and the syringe/filter unit was placed into a sharps bin inside the hood. 15ml of 2% media was added to the filtered virus solution in the 50ml tube along with polybrene to 1x concentration.

2.4.2 Mammalian Expression (Secretion)

The media was removed from a T175 flask of the appropriate cell line at approximately 80% confluence. This was replaced with the 25ml of virus in 2% media with 1x polybrene. The flask was placed in a 37°C 5% CO_2 incubator for 3-4 days. The media was pipetted off and placed in 2x Virkon solution in the hood. The cells were then split using the technique described in section 2.2.1 until the desired number of cells are produced.

Once the desired number of flasks had been produced and grown to around 80% confluency, they were split into roller bottles. The cells were washed, detached from the flask and resuspended as described in section 2.2.1. Roller bottles were prepared with approximately 225ml of 10% FBS media, all 25ml of resuspended cells were then added: 1 flask of cells per roller bottle. The roller bottles were then transferred to a roller incubator at 37°C and 5% CO_2 for 3-4 days.

After 3-4 days it was assumed that cells were at around 80% confluency, so protein expression was induced. The old 10% media was poured out of the roller bottles and replaced with around 250ml of 2% FBS media containing 1X doxycycline. The roller bottles were then transferred to a roller incubator at 37°C and 5% CO_2 for 1 week, checking every day for signs of cell death or infection.

After the 1st week expression, the media was harvested and replaced with fresh 2% FBS media and expression allowed to continue for another week before a 2nd harvest was taken. Cells and roller bottles were disposed of through autoclaving at this point as expressed protein is located in the harvested media.

2.4.3 Small Scale Expression Testing

A 6-well plate of lenti-X cells was prepared 1 day in advance for use after 24 hours. On transfection day DNA was prepared in 1.5ml tubes. Each tube had 1.3µg each of the insert plasmid, packaging plasmid, and envelope plasmid. 250µl of serum free media was added to each tube and mixed with the DNA. 250µl of SFM containing 8µg of PEI was then added and mixed. The DNA/PEI mixture was left to incubate for 10 minutes. Meanwhile the 2ml of media was removed from each well of the 6-well plate and replaced with 1.5ml of 2% media. Once the 10 minutes incubation was up 500µl of the DNA/PEI mixture was added to each well and the plate returned to the 37°C incubator.

A 6-well plate of HEK293S cells was prepared 24 hours before transduction day.

3-4 days after transfection the virus was harvested and used to transduce HEK293S cells. Media from each well of the lenti-X plate was transferred to a syringe and filtered through a 0.4µm filter into a 15ml tube. 1.5ml of 2% media was added to each 15ml tube bringing to total volume to 3ml. The 2ml of media was removed from each well of the 6-well plate of HEK293S cells. 2ml of filtered virus was added to each well of the HEK293S plate before it was placed into the 37°C incubator.

3-4 days later protein expression was induced. Induction media was prepared by adding doxicyclin to 2% media to a final concentration of 10µg/ml. The media was removed from each well of the test expression plate and replaced with 2ml of induction media. The plate was placed in the 37°C incubator.

3-4 days later the media was harvested, and a western blot was used to visualise expression levels. An anti-His antibody was used and 15µl of media was loaded onto the gel.

2.5 Protein Purification

2.5.1 Sample Preparation

Media harvested from roller bottles was combined, diluted to the nearest litre and placed into 1L floor standing centrifuge pots. These were then centrifuged at 5000g for 20 minutes at 18°C using a JLA9.1 rotor (Beckman Coulter). The supernatant was filtered through a 0.2µm filter using a vacuum pump. Dialysis tubing (50.8mm, limit 12-14kDa) was prepared by placing in 1.6L of hot 1.1mM EDTA pH8 and 0.26M sodium bicarbonate for 10 minutes (heated by microwaving on full power for 10 minutes). The EDTA/sodium bicarbonate solution was discarded and replaced by ultra-pure water; the tubing was microwaved in the water for another 10 minutes. The hot water was then replaced with tap temperature ultra-pure water to cool the dialysis tubing down again. The clarified media was placed into the dialysis tubing and dialysed overnight at room temperature against 15L of 20mM Tris pH8.0 and 250mM NaCl.

2.5.2 Affinity Purification

A 5ml bed volume gravity flow drip column was packed and washed with 10 column volumes of fresh base buffer. The resin used was chelating sepharose charged with Ni^{2+} (GE Healthcare) and base buffer contained 20mM Tris pH8.0 and 250mM NaCl. The clarified media was passed over this column at an approximately 2 drips/sec flow rate. The resin was then washed with 5 column volumes of base buffers containing 10mM imidazole. Protein was eluted with 3 times 2 column volumes of base buffer containing 250mM imidazole. All collected fractions were analysed on an SDS-PAGE gel.

2.5.3 Size Exclusion Chromatography

Further purification and imidazole removal used size exclusion chromatography with a Superdex 200 column (GE Healthcare).

Fractions from the affinity purification containing clean protein were combined and concentrated using Amicon spin concentrators (Merck) to a volume of less than 500µl. Concentration took place at 25°C in an Eppendorf 5810 R bench top centrifuge at 4000G. The centrifuge was stopped every 10 minutes to mix and prevent a large concentration gradient. If the protein was slow to concentrate, then an A280 reading was taken on a nano drop (Thermo Fisher Scientific) periodically to check protein was concentrating.

A Superdex 200 16/600 column (GE Healthcare), or occasionally 200 10/300(GE Healthcare), and Akta pure (GE Healthcare) fitted with a 500µl loop were equilibrated in to 10mM Tris pH8.0 and 125mM NaCl at 4°C. The concentrated protein was loaded into to 500µl loop manually using a syringe. The loop was emptied to the column with 3ml of buffer, 1 column volume of buffer was passed through the column while 2ml fraction (0.5ml in the case of 10/300) were collected. Fractions corresponding to the peak were visualised on an SDS-PAGE gel.

Fractions containing pure protein were combine and concentrated using the above protocol to above 10mg/ml. Protein was aliquoted, flash frozen in liquid nitrogen, and stored at -80°C.

2.6 Gel Electrophoresis

2.6.1 Agarose Gels for DNA

Agarose gels for DNA electrophoresis were hand poured. 50x Tris Acetate EDTA (TAE) buffer was made up by mixing 121g Tris base, 29ml of glacial acetic acid, and 50ml of 0.5M EDTA (pH8.0) and adjusting to a total volume of 500ml with ultra-pure water. TAE running buffer was made up using 50x TAE stock and ultrapure water to make a 1x solution. 1.5% agarose was prepared by mixing 3g of agarose powder with 200ml of TAE running buffer and microwaving on full power until the powder was fully dissolved. If the 1.5% agarose had been previously made up, then it was melted in the microwave using full power. A Bio-Rad horizontal electrophoresis agarose gel caster was used to cast the gel in a Mini-Sub cell UV-transparent gel tray with a 15 well comb (or two for larger trays). 10ml of melted agarose was used for a 7x7cm try and 15ml for a 7x10cm tray. Prior to casting SYBR safe stain (Invitrogen) was added to 1x concentration to the appropriate amount of melted agarose. The agarose was then poured into the assembled casting device and allowed to set at room temperature for 30 minutes. The cast gel was placed in a mini-sub cell horizontal electrophoresis system (Bio-Rad) and covered with TAE running buffer to the fill line.

Samples were prepared by adding 1µl of Blue Juice loading dye (Invitrogen) to 5µl of DNA. The gels were loaded with 5µl of 1kb+ DNA ladder (Invitrogen) and 5µl of each of the pre-prepared samples. The gels were run at 120 volts for 30 minutes. Gels were imaged on the blue tray in a BioRad imager under UV light.

2.6.2 SDS-PAGE

Invitrogen NuPAGE 4-12% Bis-Tris protein gels were used in either a Thermo Fisher Scientific mini gel tank or xCell sure lock Mini-cell tank with 1x MES buffer (Invitogen). Samples were prepared by mixing 5µl of SDS-PAGE sample loading buffer (Novex) with 15µl of sample. Protein sample was diluted prior to mixing if protein concentration was suspected to be high. Mixed samples were incubated at 95°C for 3 minutes before being loaded. 5µl of benchmark protein ladder (Novex) was loaded and 15µl of the pre-prepared samples. The gel was run at 200 volts for 40 minutes. The gel was stained with instant blue gel stain (Abcam) for 2 hours then destained in water before being imaged on the white tray in a Bio Rad imager.

2.6.3 Western Blotting

The initial gel was loaded and run as per section 2.6.2 with the following change to the ladder volume. 1µl of unstained benchmark protein ladder (Novex) was loaded alongside 7µl of pre-stained benchmark protein ladder (Novex).

While the initial gel was running milk was prepared by adding 2.5g of fat free milk powder to 50ml of PBS. The PBS was prepared using PBS tablets from Sigma and ultra-pure water. Transfer sponges, 2 pieces of filter paper, and 1 piece of transfer membrane were soaked in 1x transfer buffer. Once the gel had finished running, the top wells and bottom protrusion were removed. The gel was then assembled into the Thermo Fisher Scientific mini blot module topped up with 1x transfer buffer and run at 20 volts for 60 minutes. After the transfer was complete the transfer membrane was placed protein side up in a small tub and covered with 25ml of milk. This was incubated for 10 minutes while shaking. The milk was then poured off and replaced with 12ml of milk containing 2.5µl of penta-His antibody. This was incubated with shaking for 60 minutes. The membrane was then washed with PBS twice then PBST (PBS with 0.1% Tween), shaking for 2 minutes between washes. The membrane was then incubated with 12ml of milk containing 2.5µl of secondary antibody for 30 minutes while shaking. The membrane was then

washed with PBS twice then PBST, shaking for 2 minutes between washes. During the final wash imaging solution was prepared by mixing 1.5ml of both solutions together. Once clean the membrane was placed into a white weigh boat and covered with imaging solution. The membrane was imaged immediately using a Bio-Rad imager and the Chemi high sensitivity program.

2.7 Crystallography

2.7.1 Crystallisation

Mosquito HTS liquid handler (TTP Labtech) was used to set up sitting drop vapour diffusion in 96 well swiss ci 3 or crystal quickx plates. Drop ratios of protein:well solution of 2:1, 1:1, and 1:2 were used for swiss ci 3 plates and the crystal quickx plates used 2:1 and 1:1. Plates were stored in a Rock Imager 1000 (Formulatrix) at 20°C and imaged with visible light in a Fibonacci sequence. Proteins were screened in JCSG+[106], Morpheus, BCS[107], PACT , (all Molecular Dimensions) and index (Hampton Research) crystallography screens, for porCD46 the index screen was replaced by Hampton 1+2 (also from Hampton Research). Plates were set up with the native protein as well as plates with deglycosylated proteins. For deglycosylated proteins, 200 units of endoHf (NEB) per 0.6mg of protein was added 30 minutes before the plates were set up so deglycosylation mainly happened in the drop. In some cases, UV imaging was used to confirm that positive hits were protein crystals. Hits were optimized using a combination of grid screening (using stock options pH and salt kits from Hampton Research), microseed screening (using seed bead kit from Hampton Research), additive screening with the GRAS and Silver bullets additive screens (Hampton Research), 4°C incubation, altering protein concentrations, and addition of glycerol.

2.7.2 Data Collection

Crystals were cryoprotected by flooding the drop with either glycerol or ethylene glycol diluted in mother liquor. Subsequently they were harvested and flash-cooled in liquid nitrogen. Data collection took place at Diamond Light Source beamlines

I03 and I23. Data sets were recorded for APPV E2, NRPV E2, bovCD46_SCR1-2, bovCD46_SCR1-4, porCD46_SCR1-4. Crystal harvest and data collection parameters are described in table 2.1.

2.7.3 Data Processing and Structure Solution

Datasets were processed with xia2_3dii-XDS/[115, 120], autoPROC[108], or autoPROC-STARANISO[114] via the diamond auto processing pipeline, NRPV E2 was manually reprocessed into the correct space group. The bovCD46 SCR1-2, bovCD46_SCR1-4, porCD46_SCR1-4, and APPV E2 (high resolution) were all solved via molecular replacement. APPV E2 (low resolution) and NRPV E2 were both solved using sulphur SAD phasing. The exact programs used are detailed in table 2.3.

Table 2.1: X-ray Crystallography Data Collection Details

Protein	Cryo	Xtal size (µm)	Beamline	Wavelength	Data Sets	Beam size (µm)	Resolution (Å)	Transmission	flux	Exposure (sec)	Omega Rotation	Kappa	No Of Images
bovCD46_SCR1-2	20% Glycerol	150x75	I03	0.976	1	80x20	2	25%	9.32e11	0.025	0.1	0	3600
bovCD46_SCR1-4	20% Glycerol	300	I03	0.976	1	50x20	3	100%	1.26e12	0.2	0.1	0	3600
porCD46_SCR1-4	25% Glycerol	50x150	I23	2.755	1	200x200	2	100%	0	0.1	0.1	0	3600
APPV E2 S-SAD	20% Glycerol	75x150	I23	2.755	3	200x350	2	26%	1.62e11, 8.05e10, 8.16e10	0.1	0.1	0, -10, -20	3x3600
APPV E2 pH8	20% EtGly	75x150	I03	0.976	1	80x20	1.4	100%	3.65e12	0.005	0.1	0	1800
NRPV E2	25% Glycerol	400x100	I23	2.755	3	100x200	2	26%	1.21e11, 1.21e11, 1.22e11	0.1	0.1	0, 10, 20	3x3600

Table 2.3: Programs Used for Structure Solution

	Data Processing	MR Model Preparation	Structure Solution	Model Building	Refinement
bovCD46_SCR1-2	autoPROC[108]	Sculptor[109]	PHASER[110]	AutoSol[111] COOT[112]	REFMAC[113]
bovCD46_SCR1-4	autoPROC[108]/Staraniso[114]		PHASER[110]	COOT[112]	REFMAC[113]
porCD46_SCR1-4	Xia2_3dii-XDS[115] autoPROC[108] autoPROC[108]/Staraniso[114]	Ensembler Sculptor[109] AlphaFold[116] RoseTTAFold[117]	MOLREP[118] PHASER[110]		
APPV E2 S-SAD	Xia2_3dii-XDS XSCALE[119, 120]		HKL2MAP[121] Autosol[111]	COOT[112]	Phenix refine[122]
APPV E2 high res	Xia2_3dii-XDS[115, 120]		Phaser[110]	COOT[112]	REFMAC[113]
NRPV E2	XDS[115] XSCALE[119]		HKL2map[121] Autosol[111]	COOT[112]	Phenix refine[122]

2.8 SAXS

2.8.1 Sample Preparation

Samples of the E2 ectodomain constructs BVDV_9, BDV_10, CDFV_9, APPV_9, and NRPV_9 were diluted to 7mg/ml in 10mM Tris pH 7.5, 150mM NaCl, 1% glycerol. A high pH buffer (10mM Tris pH 7.5, 150mM NaCl, 1% glycerol) and low pH buffer (10mM Citrate pH 5, 150mM NaCl, 1% glycerol) were also prepared.

2.8.2 Data Collection

Data collection took place at Diamond Light Source beamline B21 using the mail in service. The HPLC mode was used with a KW-403 column (Shodex) at 0.16 ml/min. All samples were run in both the high and low pH buffers described above. Further data collection details can be found in table 2.5.

2.8.3 Data Analysis

Data was processed in ScÅtter[123] then analysed in both ScÅtter[123] and AT-SAS[124]. Buffer subtraction was carried out in ScÅtter as was peak selection, scaling, and merging. Guinier analysis and Guinier peak analysis were carried out in ScÅtter by manually removing points obtain the best fit to calculate the radius of gyration. ScÅtter was used to carry out a Kratky, normalised Kratky analysis, and porod. The P(r) function was calculated in GNOM[125] then DAMMIF[126] was used for *ab initio* modelling (10 models per sample) followed by average model selection by DAMAVER[127] all from the ATSAS[124] suit of programs.

Wavelength	0.95 Å
Energy	13.00 keV
Beam centre X	129.37 mm
Beam centre Y	20.03 mm
Sample to detector distance	3698.80 mm
Detector	EigerX 4M (Dectris)
Source	bending magnet
Flux	2x104 photons.s-1
Q range	0.00-0.34 Å-1
Q definition	$4\pi * sin(\theta)/\lambda$
Intensity units	cm^{-1}
Capillary diameter	1.50 mm
Default exposure temperature	15 °C
Beam size at sample	1.00 x 0.25 mm
Beam size at detector (focus)	0.05 x 0.05 mm (FWHM)

Table 2.5: SAXS Data Collection Details

2.9 Dynamic Light Scattering

The sizing and polydispersity program was used on an Uncle machine from Nn-chained Labs. 10µl of protein at 1mg/ml was loaded into each sample capillary, each sample was run in triplicate. For the deglycosylated samples approximately 200 units of endoHf (NEB) per 0.6mg of protein was added to the samples prior to dilution. This was incubated at room temperature for 24 hours before being diluted to 1mg/ml for sample acquisition. Four 20 second exposures were taken from each sample capillary. The program defaults for globular proteins were used to analyse the data and calculate estimated molecular weights and polydispersity.

3
Protein Production

Contents

3.1 Construct Design 53
3.2 Cloning 64
3.3 Expression 74
3.4 Purification 81
3.5 Validation Techniques 91
3.6 Summary 92

3.1 Construct Design

3.1.1 Bioinformatic Studies

The first step in designing this project involved choosing which viruses within the *Pestivirus* genus to focus on. There were the obvious choices of the well-known livestock pathogens BVDV and CSFV, in addition BDV is less well known but is still important in sheep farming[8–10]. However, the project aimed to provide insight into the role of the envelope glycoproteins in membrane fusion so studying pestiviruses which are slightly different may be useful as well. As such, a multiple sequence alignment was carried out to investigate the sequence identity between different pestiviruses to identify which targets may provide insight into pestivirus membrane fusion. Additionally, as the pestivirus genome initially produces a

single polyprotein, for the less well known pestiviruses the domain boundaries for individual proteins have not yet been reported. The multiple sequence alignment was used to hypothesise the domain boundaries for the target proteins given those known for some other pestiviruses (detailed in appendix D).

A literature search was performed to identify the main pestiviruses focusing only on those where the entire genetic sequence was available. Clustal Omega[103] was then used to align the sequences of the E1-E2 region of the 16 identified pestiviruses (Figure 3.1). On examination of the alignment output, the 16 pestivirus E1-E2 sequences can be easily classifed into 3 different groups based on their similarity to BVDV. The first group with highest sequence identity, 65-70%, is also the largest. It contains all the most well-known pestiviruses: BVDV1, BVDV2, BVDV3, CSFV, and BDV along with three less well known pestiviruses: *giraffe pestivirus, reindeer pestivirus*, and *Aydin pestivirus*. The second group has less sequence identity at 41-44% and contains only 3 pestiviruses: *pronghorn pestivirus, bungowannah pestivirus*, and *LINDA pestivirus*. All the pestiviruses in these two groups have proteins of very similar length and have well conserved regions features despite the differing sequence identity. This contrasts with the final group which has much lower sequence identity and is most different from the rest of the pestiviruses. This group contains NRPV and the *rodent pestivirus* which have 30% sequence identity, and APPV and *bat pestivirus* with only 19-20% sequence identity. The most different region of these pestiviruses is in the E2 proteins, specifically the N-terminus. In these atypical E2 proteins a large portion of the N-terminus has little to no sequence identity and is very difficult, if not impossible to align. The N-terminus of APPV and *bat pestivirus* is the most unusual, it is missing entirely. This makes the E2 proteins from APPV and *bat pestivirus* two thirds of the size of all other currently sequenced pestiviruses.

Although APPV was only recently discovered[12] and has not yet been associated with a large economic cost, the lack of N-terminus makes it a very interesting target. Firstly, as mentioned previously the N-terminal domain of BVDV E2 becomes disordered at endosomal pH[21] and is therefore thought to play some role in

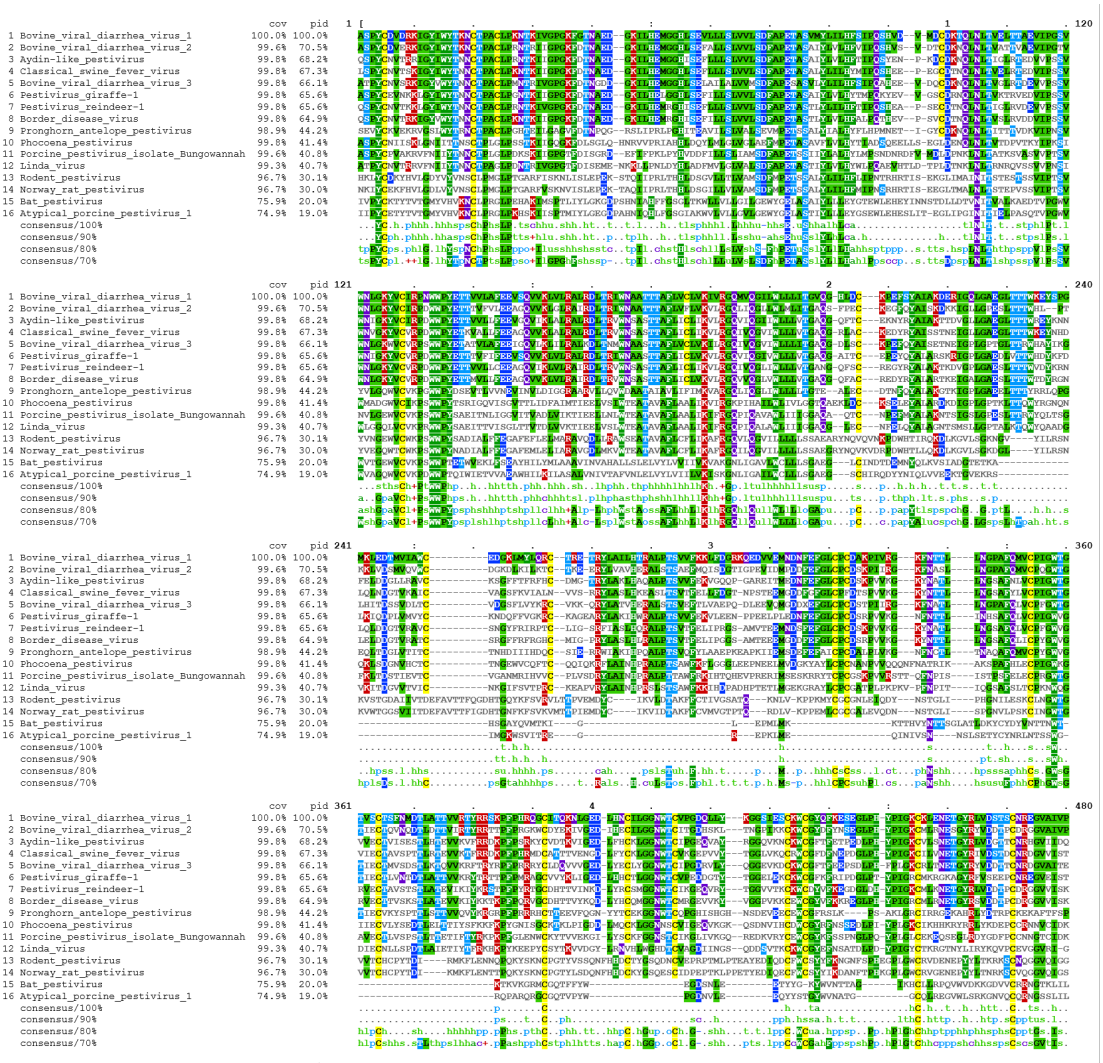


Figure 3.1: Multiple sequence alignment of pestivirus E1-E2 region ordered by sequence identity to BVDV1, generated by Clustal Omega[103] and displayed with MView[128]. Sequences are coloured according to chemical nature of sidechain, with the predominant amino acid coloured, cysteines are highlighted in yellow. The final 7 residues are omitted for figure clarity. GenBank accession numbers are used for alignments can be found in appendix A.

membrane fusion. The missing region of APPV E2 entirely covers this N-terminal domain of BVDV providing an interesting potential insight into the role of E2 in membrane fusion. One can wonder how the E2 function is preserved with its N-terminal domain missing. Secondly, in CSFV a peptide has been suggested to play a role in membrane fusion[63], potentially as a binding site for a receptor. This peptide is located at the join between the two domains at the N-terminus of the E2 protein, another region missing in the APPV E2 protein. Finally, E2 is the largest envelope glycoprotein and as such was initially thought to be the fusion protein. Once this was ruled out it was hypothesised that the large E2 protein facilitates the small E1 protein in membrane fusions. Interestingly in APPV the E2 protein is almost as small as the E1 protein.

Given the above deductions BVDV1, CSFV, BDV, and APPV were chosen as the initial targets for structural studies. After initial results were obtained the NRPV from the species *Pestivirus J* was added to the list. As the E2 protein of NRPV is the same length as the other pestivirus E2 proteins but the N-terminus lacks sequence identity. It was hoped that NRPV may be an intermediate between APPV and the rest of the pestiviruses and provide detail on the role E2 plays in viral membrane fusion.

When comparing the E1 and E2 glycoproteins sequences overall, the E1 glycoprotein has higher sequence conservation between species than E2. This is also the case for cysteines and glycosylation sites, the former are important as they are involved in disulphide bonds used to stabilise the protein fold. The higher sequence conservation of E1 is expected because it is thought to be the fusion protein[21]. E1 is a small protein, smaller than most other fusion proteins so is likely to by highly specialised for its function with little room for sequence variation. E2 is hypothesised to be involved with receptor binding[89], hence a lower sequence identity between pestiviruses would reflect the differing specificity for host target receptors. Within the E2 protein, there are differences in sequence conservation, the N-terminus is much less well conserved than the C-terminus. This is in line with the fact that

the C-terminus is involved in homo- and hetero-dimerisation[14, 21], a process critical for membrane fusion. However, the N-terminus is the most exposed part and involved with the host cell interaction described above, so for the reason previously described, it will be able to support a lower degree of sequence conservation.

The E1 and E2 proteins are class 1 membrane proteins and have a single pass transmembrane anchor to fix them in the viral envelope[14]. While it would be ideal to solve the structure of the entire protein there are several problems with this approach. Firstly, membrane proteins are much harder and more expensive to produce[129]. They are less likely to express and when they do the yields are vastly less than for soluble proteins. Additional expensive reagents and detergents are required to keep the protein soluble during the purification process. In addition to this there are additional optimisation steps required for every protein construct to identify the most suitable detergent. Secondly, it is significantly harder to produce crystals of membrane proteins and the process to make and optimise them is also harder; they often require specific crystallisation methods such as Lipidic Cubic Phase[130]. When crystals are produced, they often diffract to a lower resolution than an average soluble protein. With this in mind, in addition to the full-length proteins truncated extracellular domains missing the transmembrane anchor were studied as most of the protein function takes place in these regions. The TM server[59] was used to identify these transmembrane regions with the results shown in Figure 3.2. As expected, all proteins contain putative transmembrane regions and although the sequence varied the transmembrane regions largely aligned. This includes the interesting APPV and NRPV E2 proteins.

Solving such crystal structures can be a difficult task with the two rate limiting steps usually being the ability to produce pure well folded recombinant protein and the ability to grow well diffracting protein crystals. It is frequently the case that neither the full-length protein or the full extracellular domain can be produced and crystallised. If this is the case, the most common course of action is to try and produce/crystallise smaller parts of the protein individually. To design these,



	E1 Ecto End	E2 Ecto End
BVDV	161	343
BDV	161	341
CSFV	161	341
APPV	146	211
NRPV	142	348

Figure 3.2: Transmembrane prediction for E1-E2 region of the 5 target pestiviruses. The hydrophobic region prior to residue 200 is the E1 membrane anchor with E2 being approximately residue 200 onwards. Figure generated using TMHMM server v2.0[59]. The top horizontal line is coloured as follows: red is a transmembrane region, pink indicates extracellular, and blue represents intracellular. The lower horizontal lines indicate the probability of a residue to be intracellular or extracellular. The vertical red lines show the probability of being transmembrane (y axis) for each residue (x axis). The residue numbers are coloured green for E1 and blue for E2. The table at the bottom right displays the final ectodomain residue for each protein.

there are two additional pieces of information which are particularly useful: the predicted secondary structure of the protein and regions of predicted disorder which were taken into consideration when designing constructs.

The final criteria which is important to consider when designing recombinant protein constructs, particularly extracellular proteins, is the disulphide bonding pattern. When making a truncated recombinant protein it must fold in the same way as the wild-type protein does. If a disulphide bond is broken by removing one of the cysteines there is a possibility that the protein will misfold[131], so it is important to speculate what the disulphide bonding pattern is likely to be and try to maintain

it. Luckily cysteines involved in disulphide bonds are generally well conserved and if the protein structure of a homolog is available then the disulphide bonding pattern can be easily deduced. For the E2 proteins in group 1, all cysteines are conserved apart from 2 in CSFV whose absence indicate the loss of a single disulphide bond. The second group have 2 lost cysteines indicating a loss of 1, possibly 2 disulphide bonds. The structure of BVDV E2[21, 22]) was also known so it was straightforward to design constructs which do not disrupt the disulphide bonding pattern. For E1 there is no structural information available so this method cannot be utilised.

3.1.2 Hepacivirus and Host Receptor Proteins

When designing protein constructs for structural studies one of the strategies that is most likely to produce crystals is to define the construct boundaries based on structures of homologs already solved. If it is a struggle to even produce proteins, then a search of literature to find both structural and non-structural studies which used recombinant proteins can provide even more options to try. In this study, the aim was to produce structures using three different proteins: E1, E2, and the CD46 receptor. Ideally, these structures would be obtained in complexes (E1/E2, E2/receptor, or E1/E2/receptor) but for the basis of designing recombinant protein constructs, those are the three units needed.

E2 was likely to be the easiest protein to produce. There are two different E2 structures available one from El Omari *et al*[21] and one from Li *et al*[22]. The construct boundaries for these are very similar, encompass almost the entire extracellular domain and so provide a very good basis for construct design. It is worth noting that the El Omari *et al* protein construct has a glycosylation site removed which improved protein crystallisation. If that is not successful, then there is another genus within the *Flaviviridae* family which, like pestiviruses, has E1 and E2 glycoproteins. This genus the *Hepacivirus*, of which HCV is an example, is thought to be the closely related to the *Pestivirus* genus. There is a structure available of a large portion of the HCV E2 glycoprotein[74]. Although it does not have the same fold as the BVDV E2 structure it could be tried as a smaller core subunit to express.

For Pestivirus E1 structures there are none currently reported, moreover in the literature there is no report of anybody being able to make recombinant E1 proteins. HCV is the likely the closest potential structural homologue with protein structures available according to BLAST[132]. The E1 proteins of HCV and the pestiviruses are both thought to be the fusion proteins, they are a similar size, and they are unlikely to belong to any current class of fusion proteins. As such it is reasonable to hypothesise that they may share some structural features and overall mechanism of catalysing membrane fusion. On the structural front, there is only a very small portion of the HCV E1 protein currently solved, and this is from the N-terminus[133]. However, solving this portion of a pestivirus E1 could prove useful information on the mechanism of action as the N-terminus is thought to be involved in the fusion. Additionally, if this region were solved it might finally answer the question of whether HCV and pestiviruses share a similar and novel mechanism of actions. However, it would be preferable to solve a larger portion of the protein especially one including the fusion peptide which the above region does not.

The final protein that was produced in this study was the cellular receptor CD46 and looking at previous reports it may be the easiest. Although the bovine version of the receptor has not been crystallised, it shares a high sequence identity with Human CD46, 48% across the longest available structure. Additionally the human CD46 has been crystallised extensively in both large[94] and smaller fragments[134]. From previously reported structures, it is clear that constructs are best designed to exclude the N-terminal approximately 41 residues before the first SCR. And as expected all current structures contain between 1 and 4 complete SCR domains, definition of a set of constructs will be easy to accomplish as it is easy to decipher the domain boundaries of each SCR.

3.1.3 Protein Engineering

To aid in recombinant protein expression and purification several modifications can be implemented at the protein construct level. This can be to improve protein expression so to make a protein more soluble, to stop it being degraded or to

send it to a certain location after production. Construct design can also help with the purification step, again to keep it soluble until the buffer can be adjusted to maintain solubility, or to aid in separating the target protein from any cellular contaminants. A final reason to alter the protein construct would be to help with some sort of interaction after production such as the addition of an Avi-tag to facilitate easy biotinylation for use in downstream assays.

For the recombinant proteins in this study, initially only one addition was made to help with expression and that was a secretion signal. The presence of this signal on a protein directs it to be exported out of the cell and into the extracellular environment which is ideal for the proteins in this study. Firstly, the proteins are extracellular in nature themselves possessing multiple cysteines (6 in E1 proteins and 8-17 in E2 proteins, Figure 3.1). In the non-reducing extracellular environment, these form intermolecular and intramolecular disulphide bonds which are essential for protein function. Secondly, by secreting the protein from the cell into the media the cell does not become overloaded with protein and die. This allows expression to be carried out for long periods of time: up to two weeks, using the set up in this study, with the same set of cells. As a bonus the secretion signal is cleaved during the export process meaning that it does not interfere with the final protein product. A final advantage of this expression system is the ability to make use of the calnexin cycle, a chaperone system located in the endoplasmic reticulum. Here the chaperones calnexin and calreticulin ensure the correct folding of the protein[135].

The second modification was to add an affinity tag to aid protein purification. This tag interacts with a purification matrix and allows contaminants to be washed off leaving pure protein[136]. The tag is typically added at the end of the protein, either the N- or C-terminus. Choosing the location of the tag is important as it must not be hidden by the protein, but it must also not get in the way of protein folding or protein function. With all three of the target proteins in this study the N-terminus is involved with protein functions, while the C-terminus has already been truncated. The tag was therefore added to the C-terminus where the

membrane anchor would normally reside as it was least likely to interfere with the protein there. In this study, a relatively small simple 6H tag was used which would interact with either nickel or talon resin (cobalt) through six histidine residues. The tag also contained a 15 residue Avi-tag sequence to give the option of a biotin streptavidin interaction if needed. This tag will not be automatically removed by the cell so a cleavage site must be added. This is a site which is recognised by a protease, 3C in this case, and can be used to remove the tag once it is no longer needed leaving only a couple of residues after cleavage. The 3C site was eight residues in length, after cleavage six residues remain from the 3C site and an additional two from the *KpnI* restriction enzyme site.

After initial efforts to produce any of the desired protein complexes it was decided to make some protein modifications to try and drive complex formation. One of these strategies was the addition of a long flexible linker to join E1 and E2 proteins and hopefully produce a heterodimer. It has been suggested that E2 may be necessary for the production and correct folding of E1 envelope glycoprotein as is the case in the closely related HCV virus[60], it was hoped that by expressing them in close proximity E1 could be produced. Due to the direction of the proteins in the genome the linker was designed to be long enough to span the entire distance from the C-terminus of the E1 protein to the N-terminus of the E2 protein. This was necessary as the C-termini are located inside the cell whereas the N-termini are exposed on the cell surface. As the linker goes from the bottom of one protein to the top of the other, it will only work with the extracellular domains as it cannot pass through the plasma membrane. Another advantage of this method would be that both proteins could be obtained in equal quantities and the complex could be easily identified on an SDS-PAGE gel.

The final modification tried was the addition of portions of dimerization proteins Jun and Fos to the C-terminus of the E1 and E2 glycoproteins proteins respectively. Jun and Fos belong to the basic Helix-Loop-helix (bHLH) family and can form a very strong heterodimer through a leucine zipper to form a coiled coil[137]. It was

hoped that these proteins could force and stabilise the heterodimerisation of E1 and E2. As it is likely that E1 needs the presence of E2 to fold and that the interaction between E1 and E2 happens mostly in the transmembrane regions[14]. Replacing the difficult to work with hydrophobic transmembrane regions with a different soluble protein fragments that still interact may enable E2 to chaperone the folding of E1 without the difficulty of full-length proteins. Additionally, the interaction between Jun and Fos is very strong, therefore a heterodimer is more likely to form even if the proteins are not produced in the correct ratio. The resulting heterodimer may also be easier to work with as it may be more stable, and the presence of a coiled coil could make crystallisation and structure solution easier.

3.1.4 Results

Overall, 88 constructs were designed (Figure 3.3). Successfully expressing constructs are detailed in Table 3.1 later in the chapter. A full list of constructs can be found in appendix E.

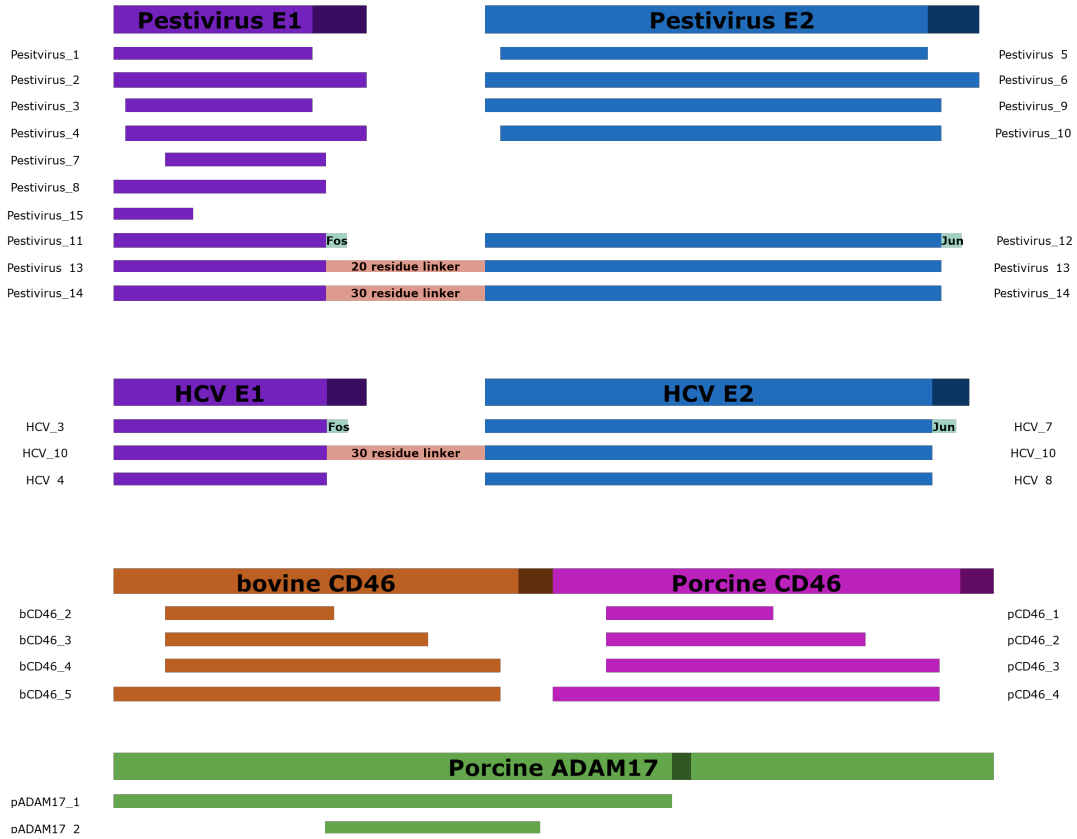


Figure 3.3: Schematic representation of construct boundaries. The full-length protein is shown as a wide line at the top with the transmembrane region in a darker grey colour. The thin lines underneath represent the boundaries of each construct with the construct names to the left or right. Full construct boundaries can be found in appendix E.

3.2 Cloning

3.2.1 Overall Cloning Strategy

The main use for the recombinant proteins produced in this study was for protein crystallography. Success rates for any protein construct in crystallography are very low with less than 3% of protein sequences producing well diffracting crystals. To increase the chances of successfully solving multiple protein structures, multiple proteins constructs were designed as described in the previous section and tested for each protein. If each individual protein construct had been tested on its own, one after another, it would have taken years to reach the first structure let alone any further structures. Therefore, constructs were designed, cloned, and expression

tested in batches of generally between 18 and 24 at a time.

The lentivirus transduction system was used during the protein expression step, so the vectors used were compatible with this[138]. These vectors were kindly donated by the Siebold group from the Department of Structural Biology. The vectors were based on the pHL-sec vector[139] system which is expertly designed to work with a restriction digest method of cloning with different restriction sites present to allow easy addition and removal of features at both the 5' and 3' of the multiple cloning site for the protein sequence. This was ideal because the restriction sites were the same between vectors streamlining the cloning process by allowing a single digested PCR fragment to be place in multiple different vectors and allowing a single digested vector stock to be used for multiple PCR fragments. An additional benefit of these vectors is the ability to be used for both lentivirus and transient protein expression in mammalian cells. Finally, they are high copy number so easy to produce in large quantities.

To make so many clones at once some medium throughput strategies used by the Structural Genomics Consortium at Oxford were employed. Firstly, primers were stored in plate format and a second plate was made with diluted primers in the same plate location as the primer stock. This prevented any mistakes as when working with up to 48 primers as it would be easy to pick up the wrong tube. Secondly, a generic overall PCR program was developed using a low annealing temperature of 52°C which should work for most primers and did in this case. Thirdly, restriction enzymes were chosen so that all digestion reactions in a cloning session could be done in the same buffer. This meant that on some occasions enzymes from NEB were used and on others Thermo Fisher Scientific enzymes were utilised instead. By using a single buffer, digestion reactions could be carried out simultaneously on both ends of the DNA resulting in faster cloning, less room for error by removing two steps, and a higher yield as the DNA clean-up/purification step is removed. Fourth, to minimise negative clones the PCR products were treated with *DpnI* to remove any template DNA, as the DNA templates used had the same antibiotic

resistance as the target vectors. The vectors were dephosphorylated to prevent any empty vector re-ligating and showing through. This meant that there was a much greater chance any clone growing contained the correct insert rather than template DNA or empty vector. Finally, colony screening was used to screen for positive colonies as growing up, mini-prepping, and digesting 24 colonies is a time consuming, expensive, and frustrating task especially if it needs to be repeated if not all colonies are positive. By utilising colony PCR instead of restriction digest, it was known if there were any positive clones within 2 hours rather than approximately 24 hours. Additionally, the PCR fragments produced were present in a large enough quantity they could be cleaned up and used for sequencing. Using PCR fragments instead of entire plasmids generally produces a better sequencing read and meant clones could be sent for sequencing the same day. An overnight sequencing provider was used, and any clones sent for sequencing were grown up overnight, that way by the next morning the sequencing results had been returned and it was known whether miniprepping the DNA was necessary or not.

These strategies together allowed, in ideal circumstances, the ability to go from primers through to verified plasmid in as little as 3 days (Figure 3.4). Detailed protocols can be found in the materials and methods chapter.

3.2.2 Restriction Cloning

Initially restriction digest was chosen as the method of cloning because the vectors were designed specifically for restriction cloning. Additionally, this technique which is used by several groups at Oxford Universities' Department of Structural Biology is well tested. Finally, when compared to some other methods of cloning it is cheaper and the reagent are much easier to obtain. Briefly, restriction cloning involves producing PCR fragments containing a restriction site at either end, the chosen target vector will also have these same restriction sites present at a multicloning site. Both the PCR fragments and the target vector are then digested with restriction enzymes specific to those sites. The restriction enzymes that were used created sticky ends, whereby a few single stranded nucleotides are left at each end, which

were complementary between the PCR insert and the target vector. Another enzyme is then used to ligate the insert and vector together thereby producing a circular plasmid. As the enzymes used leave sticky ends, the insert can only be ligated in one direction as the DNA will not be complementary if the insert is reversed. This process is summarised in Figure 3.4.

It is worth noting that restriction cloning is not the be all and end all of cloning techniques, in fact part way through the project this was changed to an alternative technique. The main issue, certainly the one that plagued this study the most, is that restriction enzymes can be extremely temperamental and will not cut efficiently if some parameters are sub-optimal. If only one parameter is causing problems, it is relatively easily to troubleshoot. However, if there are multiple issues at the same time, as was initially the case in this project, troubleshooting becomes an almost insurmountable task. Other disadvantages include that fact that there must be a restriction site at either end of the insert. This remains in place after cloning which might pose a problem depending on the desired protein product. Care must also be taken when designing the vector or insert that the desired restrictions sites are not present anywhere else. Finally, incomplete digestion can be an issue as it is impossible to tell if a PCR product has been digested as generally only a few nucleotides are removed. This can be an issue for vector double digestion (digesting with two enzymes simultaneously) as well when it is difficult to differentiate between singly and double cut vector as both appear the same on a gel.

Inserts were generated using the polymerase chain reaction (PCR) where primers and template DNA are mixed in a buffer with free nucleotides and a polymerase enzyme before being placed in a thermocycler. After several temperature change cycles to anneal, elongate, and denature the DNA, PCR products are produced. The template DNA used was synthetically generated DNA genes from various sources listed in section 2.1.2. The PCR primers were initially designed to have a random sequence equal to the length required for the restriction enzyme being used followed by the restriction site sequence and finally the first 18 nucleotides of the insert

sequence or longer until the first C or G nucleotide, the reverse primer contained the reverse complement of the last 18 nucleotides of the gene insert.

The inserts and vectors were digested, cleaned up, ligated, and transformed as described in section 2.1 with results verified after PCR (and digestion for vectors) on an agarose gel and colonies screened using colony PCR with positive hits being verified with Sanger sequencing. The primers used for colony screening were designed to work with all vectors and to produce a small product in the absence of an insert, acting as a positive control themselves. The same primers were used for sequencing.

Initially, there were major issued with the cloning process, screening of 96 colonies yielded no positives. In order to troubleshoot the vector digestion, a vector was obtained which contain an insert already which would be visible on an agarose gel if both restriction enzymes worked. Using this vector and single enzyme digests it was discovered one of the enzymes was not digesting the vector. After multiple attempts at optimising the digestion reaction and experimenting with enzyme from a different company it was discovered that the miniprep kit used to produce the vector DNA was the problem. By changing the miniprep kit from Promega kit used by the Macromolecular Crystallography group at Diamond Light Source, to the Quiagen kit used at The Department of Structural Biology (Oxford University) digestion was restored and verified by presence of the small insert on an agarose gel. This did not initially lead to any positive colonies being obtained, only by changing the PCR primers to include a 6nt handle before the restriction was the cloning successful. Unfortunately, when changing to using the vector containing an insert the wrong vector was handed over leading to proteins produced during the first 2.5 years of the project having a non-cleavable tag.

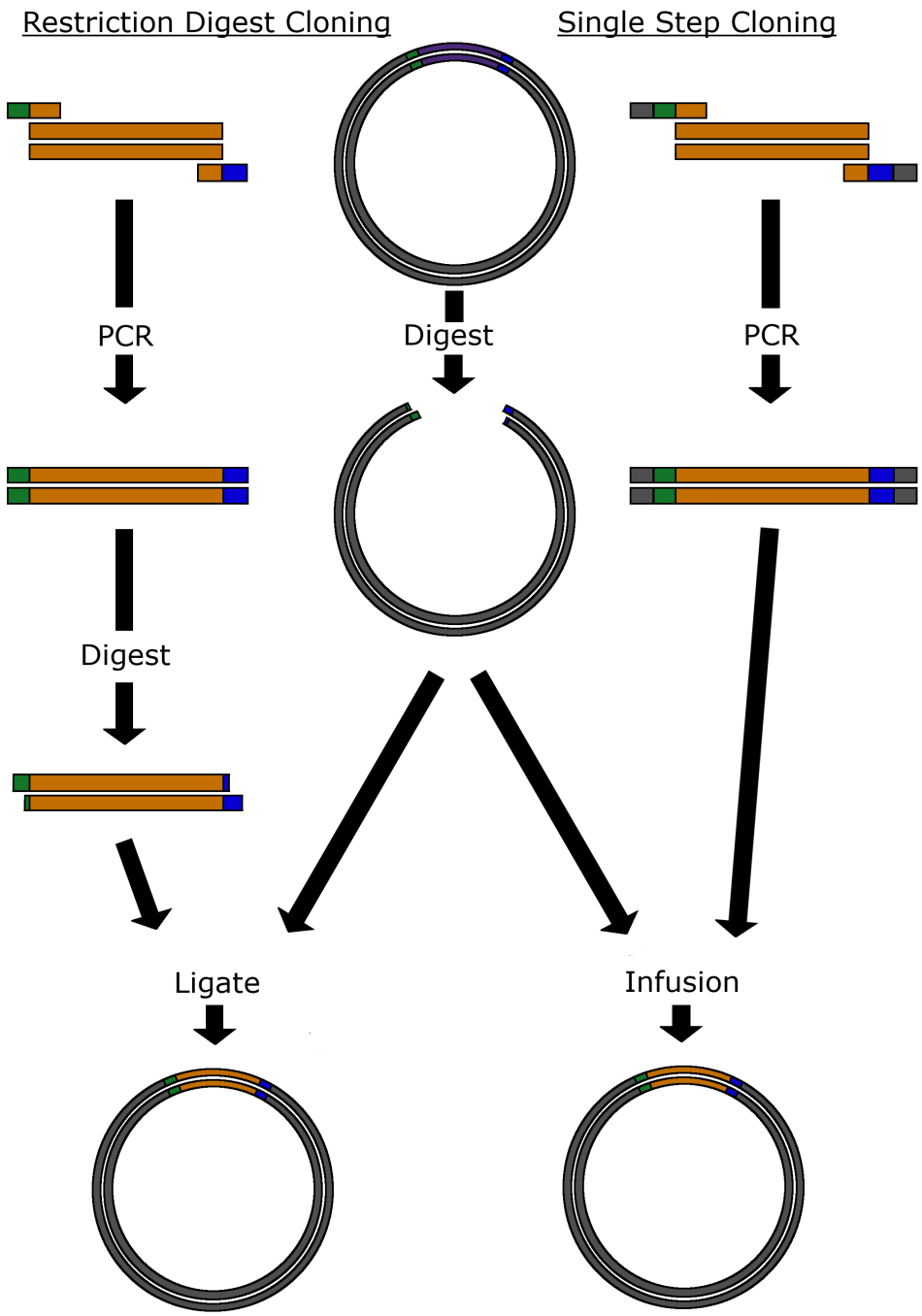


Figure 3.4: Comparison of restriction digest cloning and single step cloning methods. Vector backbone in grey, insert in target orange, restriction sites in blue and green.

3.2.3 Single Step Cloning

Midway through the project there were several constructs which were proving very difficult to clone taking several attempts to give positive clones or not working at all. These constructs were used to test a different cloning technique known as single step cloning, introduced as In-fusion cloning from Takara. Here a proprietary enzyme is used to anneal an insert into a linearized vector, it is a ligation independent cloning method and relies on complementary ends on insert and vector (Figure 3.4).

There are several advantages to this technique with the first and foremost being the high efficiency, claims are currently of over 95%[140]. Additionally, it is also an easy single step process which takes the place of the insert digestion, and subsequent clean-up, and the ligation steps in the restriction digest method above. The complementary ends are added during the PCR and sticky ends are not needed so there is no sequence or restriction site needed at either end of the insert. Again because of the complementary ends the insert is always added in the correct direction and can be added anywhere in the vector. Although it seems like the ideal cloning technique it is extremely expensive leading to it being used only in very difficult cases.

A different company, Vazyme, recently came out with their own reagent, ClonExpress II, which is significantly cheaper. When combined with the savings from not having to repeat any steps (as most constructs worked first time) it ends up cheaper then restriction cloning. With this technique the PCR primers do end up being a little longer to add the vector complementarity to the insert, however primers have become so cheap this is not really a consideration. As such, about 2.5 years into the project the main cloning method was switched to single step for all constructs.

For single step cloning the inserts were again generated using PCR in the same fashion as for the restriction cloning although the primers were a little different. Primers for this technique started with the 15 nucleotides in the vector located up stream of the insert followed by the restriction site sequence and finally the first 18 nucleotides of the insert sequence or longer until the first C or G nucleotide. For the

reverse primer it is the same design but with 15nt of the reverse complement of the region immediately downstream from where the insert is added, followed then by the reverse complement of the restriction site and the final 18nt of the insert. As the restriction sites were retained for ease of and future cloning efforts, such as moving an insert from one vector to another, following this vector complementary region there is the appropriate 6nt restriction site sequence. Finally, there is 15nt identical to the first 15nt of the insert or 15nt of the reverse complement of the last 15nt of the insert. Although the 6nt restriction site could have formed part of the 15nt region complementary to the vector here it was chosen to have it as an additional region. This was because it did not add significantly to the primer cost and having a set primer design outline minimises mistakes. When designing the primers, initially a tool on the TaKaRa site[141] was used however with this tool it was easy to accidently put the insert in out of frame. So, subsequently when construct maps were generated it was verified that everything was in frame before just copying the region where the insert and vector meet, the online service Benchling[104] was used for this.

The next two steps in the cloning process, digestion and ligation, are almost merged in the single step technique. Although the vector still needs to be linearized this can be done in advance and a linearized high concentration vector stock stored for future use. While technically only a single enzyme is required for linearization of a vector, here one was used either side of the insert region. This allowed confirmations of digestion through visualisation of then separated insert on an agarose gel as well as a band shift for the now linearised vector. The band shift of the digested vector enabled separation of the fully digested vector from that which was only partially digested or completely undigested through gel extraction. Once the vector stock has been linearized the single step cloning process becomes extremely easy, as the vector is incubated with PCR product, single step cloning reagent, and buffer before transformation. It was seen that dephosphorylation of the vector was unnecessary, as was *DpnI* treatment of the PCR products. Digested vectors are normally dephosphorylated to prevent relegation without a new insert, *DpnI* is a restriction enzyme that digests methylated DNA so is used to remove any

remaining template from the PCR product. Even using PCR products straight from the thermo cycler without any clean-up worked well. In addition, with the new cheaper reagent from Vazyme Biotech, a deactivation step is not needed after the single step cloning reaction, so that transformation can be carried out immediately after the incubation period.

3.2.4 Point Mutations

There were some points in this project where it was necessary to produce a protein with an amino acid sequence slightly different from that found in the synthetic gene. One reason was to remove a glycosylation site. Glycosylation adds large, branched carbohydrates to the surface of a protein. These can prevent crystallisation by blocking potential crystal contacts or by introducing structural heterogeneity and flexibility. By mutating an asparagine to a glutamine, the protein is prevented from being glycosylated at a certain site. El Omari *et al* had to mutate N298 in order to crystallise and solve the structure of the ectodomain of BVDV E2[21]. In order to change the amino acid sequence of a protein, first the DNA sequence that is used to produce said protein must be changed.

The most well used method of introducing a mutation into a DNA is to use a commercial kit. These kits are quick and easy to use and only require two primers. However, these kits are generally based on whole plasmid amplification which is not ideal in our case. This is because of the size of our vector at over 10kb makes it highly likely that undesired mutations will be introduced into the vector backbone. So instead, a primer extension PCR method was used, this method required more primers but as only the short insert is being amplified undesired mutations are much less likely.

Primer extension PCR mutagenesis is a two-step process as shown in Figure 3.5[105]. The first round of PCR is used to generate a forward and reverse fragment containing the desired mutation, then a second round of PCR created the final product. This final PCR product can then be cloned back into the desired vector.

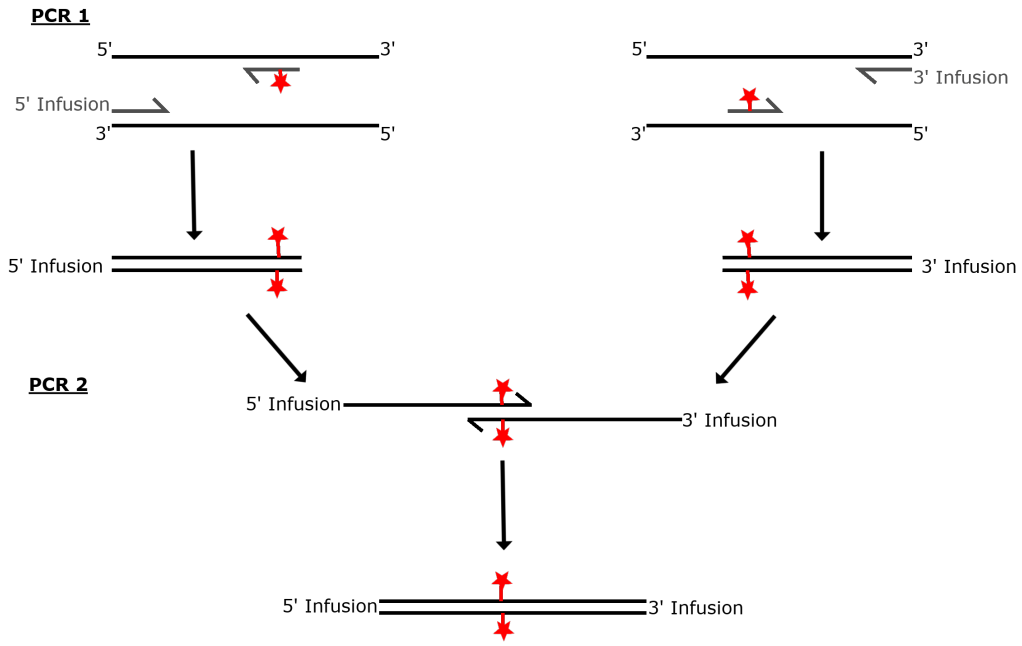


Figure 3.5: Schematic of primer extension PCR mutagenesis adapted from Methods in Molecular Biology[105] The point mutation is depicted as a red star.

The four primers required for this method of mutagenesis included forward and reverse primers for the entire insert, these should also contain any features required to clone the final mutated insert back into the vector. This is alongside forward and reverse mutagenic primers which cover the region where the mutation is to be added and contain the new sequence. For the entire insert primers, the original cloning primers for our construct were used. The mutagenic primers were centred on the region being mutated and included 12bp on either side of it.

As seen in Figure 3.5, the first PCR step involved two reactions both using the original construct or gene as template DNA: one with the forward cloning primer and the reverse mutagenic primer and the other with the forward mutagenic primer and the reverse cloning primer. In-between the first and second PCR, is an important *DpnI* treatment step to remove any remaining methylated template DNA which could accidentally be amplified in the second PCR step. In the second PCR is a single reaction, the PCR products from the first step are both used as template DNA along with the forward and reverse cloning primers. Both rounds of PCR have fewer cycles then as is customary to minimise the risk of introducing

undesired mutations. After the second round of PCR is complete then the mutated insert has been obtained and it can be cloned back into the vector as described above in the single step and restriction cloning sections.

3.2.5 Results

Overall 77 constructs were successfully cloned. Constructs which expressed protein are detailed in Table 3.1 later in the chapter. A full list of constructs can be found in appendix E.

3.3 Expression

3.3.1 Lentivirus

As all three types of proteins produced in this study (E1, E2, and the receptor CD46) contain the post translational modification N-linked glycosylation, and the viruses infect mammalian hosts, recombinant expression in mammalian cells should maximise the possibility of obtaining correctly folded and post translationally modified protein. One of the most commonly used methods for expressing protein in mammalian cells is transient protein expression[139]. With this technique mammalian cells are grown until the required quantity is obtained then a transfection reagent is used to introduce DNA to all the cells. This can be an efficient procedure; indeed, at present it is probably the most common way of expressing protein in mammalian cells, however there are some disadvantages. In order to transfect the required number of cells a lot of plasmid DNA is required, which is expensive and time-consuming to produce. Additionally, the transfection efficiency at this scale has been reported to be as low as 60-75%[138] so a significant proportion of cells are not expressing recombinant protein, another source of wasted resources. Generating a stable cell line would overcome the need for so much DNA in addition to the large-scale cell wastage, but it can take up to 10 weeks to generate using selection and expansion. However, there is now an alternative, Lentiviruses, which can be used to develop polyclonal cell lines in only 7 days and protein expression within 3 weeks, only required a single miniprep of DNA, and can have up to 100%

of cells expressing protein[138]. Furthermore, it has been suggested that stable gene expression leads to higher quantities of recombinant protein being obtained when compared with transient expression[138].

The lentiviruses used in recombinant protein expression are derived from genetic elements of HIV virus and can be used to deliver genetic information into a cell via membrane fusion where it is then incorporated into the host cell genome[142]. Initially, the desired insert is cloned into the lentivirus transfer plasmid and the resulting plasmid is mixed with a packaging plasmid and an envelope plasmid. This mixture is transfected into a lentivirus producing cell line, HEK293 Lenti-X, using PEI. After some time, the supernatant, now containing virus, is then used to transduce the expression cell line straight away without concentrating. The resulting polyclonal cell line can be expanded for protein expression and/or put into storage for future use[138].

There are some safety concerns with this method of mammalian protein expression namely potentially infecting yourself instead of the cell line and potentially generating replication competent Lentiviruses. To prevent accidents training is carried out, PPE is worn, class II safety hoods are utilised, generation of aerosols are avoided, and Virkon is used for disinfection. Additionally, the inserts used for this procedure are carefully studied for potential oncogenes. Furthemore the whole system has been developed to avoid the accidental generation of replication competent virus particles. The viral DNA has been split into three plasmids with only the transfer plasmid (this one contains the insert) being incorporated into the host cell genome. This plasmid has been designed to not include all the necessary information to produce a virus particle without the other plasmids (packaging and envelope) present[138].

3.3.2 Cell Lines

When expressing recombinant proteins there are four main expression systems: *E. coli*, insect cell, yeast, and mammalian cells. Each has advantages and disadvantages

but for this project insect cells and mammalian cells are the only likely viable options because of the previously mentioned glycosylation and disulphide bonds. Therefore mammalian expression is the most appropriate for this project for the reasons given above.

Two main types of cell lines are used for the large-scale mammalian expression of recombinant proteins: Chinese hamster ovary (CHO) cells[143] and human embryonic kidney (HEK) 293 cells[144]. Both of these can be transfected with commonly available PEI. However, HEK293 cells have been reported to have the highest level of PEI-mediated transfection and support the formation of stable cell lines, indeed the lentiviral system used here was developed for use with this cell line[138].

The specific cell line chosen for this project was the HEK293S GnTI$^-$ TetR which is suitable for both adherent and suspension-based cell culture[145]. The TetR refers to the presence of the pcDNA6/TR vector leading to expression of Tet repressor protein (TetR) which can repress the transcription of the transgene. Expression of the protein from the transgene can be induced (have the repression removed) by the addition of doxycyclin which binds the TetR preventing it from repressing transcription[146]. The GnTI$^-$ means the cell lined lacks N-acetylglucosaminyltransferase I, an enzyme required to process high mannose N-linked glycans to hybrid or complex glycans. Due to the lack of N-acetylglucosaminyltransferase I the cell is unable to transfer N-Acetylglucosamine onto a glycan chain terminating in a mannose preventing the formation of complex N-linked sugars[145]. Moreover, the high mannose type glycosylation is easily cleaved by endoglycosidase H or F1 to aid in crystallisation. A final advantage of this cell line is that it expresses c-Myc-tagged blasticidin deaminase (BSD) through the pcDNA6/TR vector. This means is it possible to select for the retention of the pcdNA6/TR plasmid and therefore the induction capability using blasticidin[145]. Initially commercial HEK29S GntI$^-$ TetR cell line was used but after the first expression test, this was switched to a similar but

inhouse version of the HEK293S GntI$^-$ TetR cell line which ensures the TetR and BSD are expressed as separate proteins and increases the nuclear localisation of the TetR.

3.3.3 Small Scale Test Expression

As the strategy for this project is based on trying as many protein constructs as possible to see which express and crystallise, it was not practical to do a large-scale expression for each construct. Not only would this take more time than was available, but it would also be a massive waste of resources as most construct will not produce protein. These wasted resources are generally very costly both in monetary and environmental terms, and this is exacerbated with mammalian expression as it involves vast amounts of non-recyclable disposable plasticware and small bottles of premixed reagents. Ideally expression screening would triage constructs prior to large-scale expression so only well expression positive constructs are taken forward, for this process a small-scale expression testing method was used.

The easiest and most common method of testing expression on a small scale would be to use a transient transfection method. This would only require a small amount of DNA as only a small number of cells are used, additionally protein expression would start soon after DNA transfection. This method of expression testing would be very fast with results possible only 4 days after DNA transfection. However, to ensure the small-scale results would have the best likelihood of transferring over to the larger scale then it is necessary to use the same DNA delivery method, same cell line, and the same method of initiating protein expression. Therefore, lentivirus was used to deliver the DNA to the cells, and then to have an inducible system where protein is not produced until after an initiation signal is added.

Both the transient and lentivirus method of DNA delivery were tested when developing the test expression method. Surprisingly there were several cases where protein was expressed with the lentivirus method and not with transient expression (Figure 3.6). It is unclear whether this is due to the DNA being stably incorporated into the genome or because the protein expression was inducible, this was not

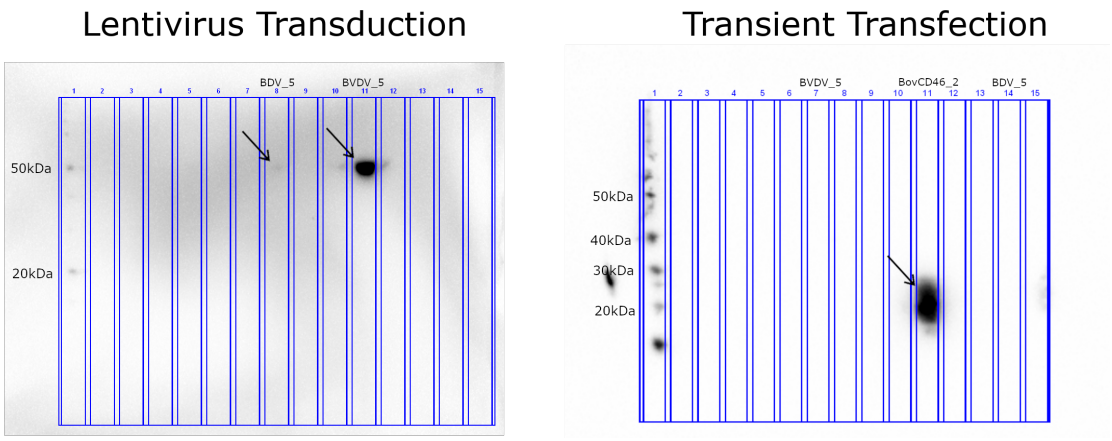


Figure 3.6: Two test expressions with DNA delivered to the cells via transient transfection or lentivirus transduction. Western blot of a reducing gel with protein detected using an anti-His antibody against the protein tag. Positive hits are highlighted with a black arrow. The E2 extracellular domains BDV_5 and BVDV_5 both express with lentivirus transduction but not with transient transfection.

investigated further as the test expression system is just a means to an end. Nonetheless, it was decided to make the extra effort to use the lentivirus method of delivery even in the small-scale system.

In the final version of the small scale system that I utilized (described in detail in section 2.4.3), 6 well plates were used as these could be seeded one day and ready to transfection the next. The cells in this 2ml of media produced more than enough protein over 2-3 days to be detectable on a western blot with only a few ul of media run, additionally the wells are large enough to be easily manipulated. 4 plates can be easily maintained at once so the expression of 24 protein constructs can be tested simultaneously. 2 days after transfecting the lentivirus cells their media is harvested directly via a syringe; a filter was then added, and the media filtered into a 15ml tube. Fresh media, containing 2x polybrene, was added to the tube to dilute the virus in half and then this was applied to the cells. After 2 days, the media was changed and protein expression was induced, a further 3 days later the media and cells were tested for protein expression via western blot. Although this method has a few more steps, it only takes an extra week, and a much better representation of the protein expression level at large scale is obtained.

3.3.4 Large Scale Expression

A large quantity of recombinant protein is required for crystallisation trials. After a well-expressing protein construct has been identified a large-scale protein expression procedure needs to be put into place (Figure 3.7). As stated above, mammalian cells (HEK293S) were used to ensure proper folding and post translational modifications whilst truncating N-linked glycosylation to the high mannose level to aid in crystallography. The lentivirus system was used to introduce the DNA to the cells in order to speed up generation of stable cell lines, increase efficiency of expression, and remove the need for large amounts of DNA. This had the added advantage discovered in the expression test stage of increasing expression levels for many protein constructs. However, all tests were carried out in 6 well plates, so it was necessary to scale up to reach the desired levels of protein expression. Because the protein is secreted, the cells were cultured in an adherent fashion. This is particularly easy these days as disposable cell culture flasks and bottles are available to buy which are already coated and prepared for adherent cells. Rather than have a large number of flasks, roller bottles are utilised after the final split. These are ribbed bottles which are laid in their side in a roller incubator where they are slowly and continually rolled. This provides adequate gas exchange and nutrient supply while the ribs doubled the surface area on which the cells can grow.

Protein production was not induced until the cells have reached 80-90% confluency in the roller bottles, this way a maximum amount of protein is obtained from each roller bottle position in the incubator. Additionally, a double harvest was utilised to obtain maximum protein from each cell batch. After the initial harvest fresh media was added and the cells allowed to continue producing protein for another week. This double harvest technique almost doubled the amount of protein obtained without any discernible problems with the final protein product.

By increasing the efficiency of protein production, more space was available in the incubators enabling multiple constructs to be expressed simultaneously. Additionally, the amount of protein obtained was increased for the resources used

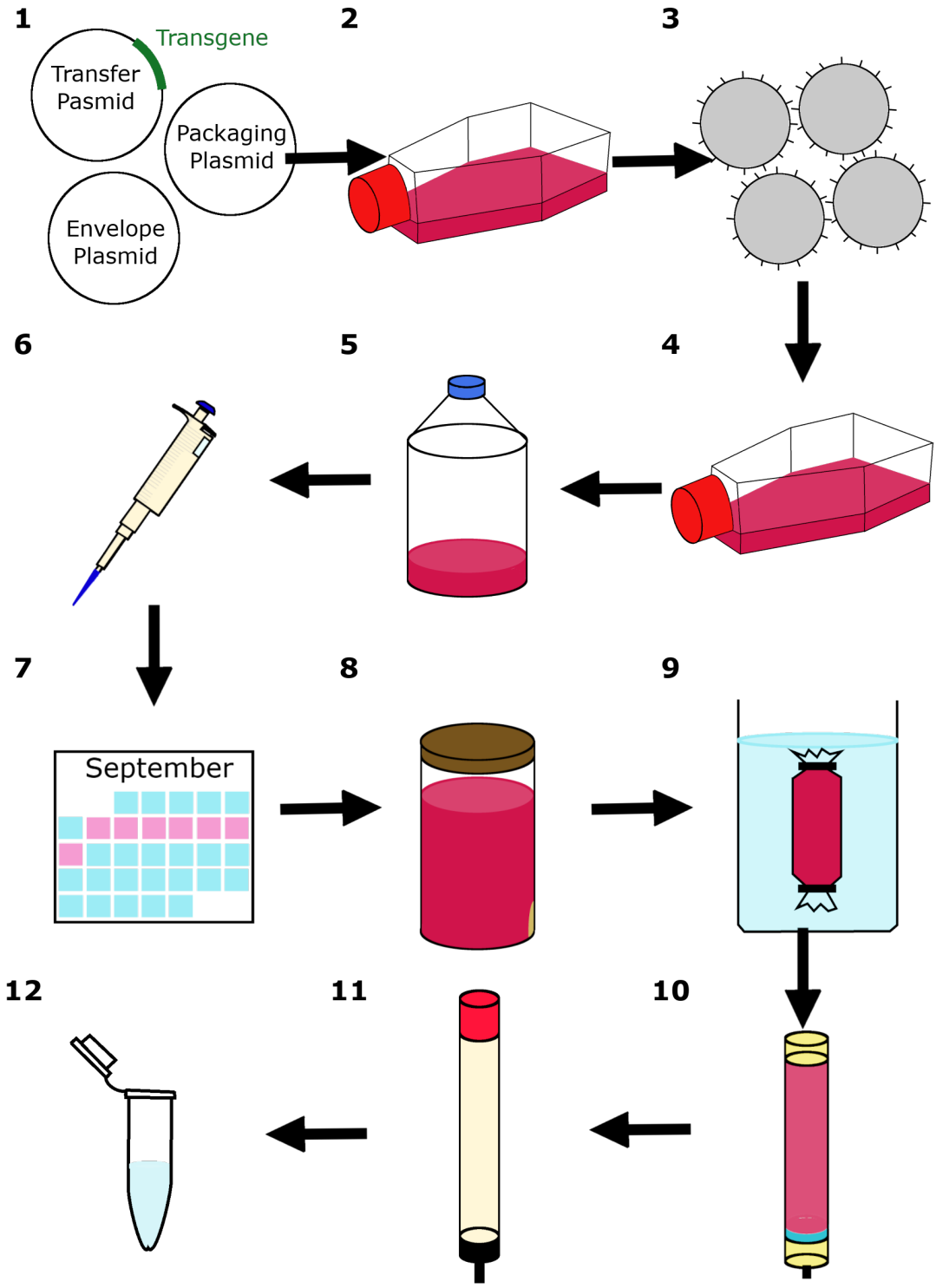


Figure 3.7: Schematic of the overall large scale expression and purification process. **1-** Transfer, packaging, and envelope plasmids are transfected into Lenti X cell. **2-** Cells produce virus particles. **3-** Virus particles are harvested. **4-** Particles used to transduce HEK293S cell. **5-** Cells are then scaled up to roller bottles. **6-** Protein expression was then induced with doxycycline. **7-** The cells are left to express for 1 week. **8-** The media was harvested using centrifugation. **9-** Prior to purification the harvested media was dialysed **10-** Initial purification on an IMAC column **11-** Further purification and buffer exchange using a SEC column.**12-** Finally protein was concentrated.

as resource use is largely dependent on number roller bottles rather than number of cells. Finally, as the protein is secreted by increasing the concentration of protein in the harvested media the total volume of media is decreased making the next stage of protein production much easier.

3.3.5 Results

Overall E2 ectodomain protein was successfully expressed for all 5 of the pestiviruses targeted in this study, a total of nine constructs (Figure 3.8). The ectodomain of both the Bovine and Porcine CD46 receptor was also produced, seven constructs. Finally, both the E1 and E2 ectodomain was obtained for HCV, a close relative to the pestiviruses, four constructs. Successfully expressed protein constructs are detailed in Table 3.1 below. All attempted constructs of pestivirus E1 failed (including the small N-terminal section equivalent to the HCV E1 structure) as did all constructs of full-length proteins (Figure 3.8). Some co-expressions of E1 and E2 protein were also carried out but ultimately failed. The failure of E1 proteins to express and the E1:E2 co-expressions was likely due to E1 not being correctly folded. Initial constructs of the putative CSFV cellular receptor ADAM17 also failed, although this receptor was only identified 6 months before the end of this thesis so the expression constructs were not optimised. A full list of tested constructs and expression results can be found in appendix E.

3.4 Purification

3.4.1 Sample Preparation

After expression the protein is secreted into the media where it is incubated for up to a week before it is harvested. However, the media is not the best environment to purify protein from as it contains many reagents which could impede future purification steps. Additionally, the media is full of cellular debris from cells which have died and detached over the 1-week expression period. This debris falls into two categories: larger visible pieces and clumps of cells as well as smaller pieces which are not visible both of which will easily block affinity purification columns.

Table 3.1: Protein Production Results

Construct	Description	Test Expression Level	Pure Protein Yield/L (mg)	Storage Concentration (mg/ml)
BVDV_5	BVDV E2 ectodomain from PDB:2YQ2	High	2.4	7.1
BVDV_9	Full BVDV E2 ectodomain	High	N/A	N/A
BVDV_17	BVDV_9 with 3C cleavage site	High	28.7	11.4
CSFV_9	Full CSFV E2 ectodomain	High	20.0	29.0
CSFV_16	CSFV_9 with N292Q mutation	High	21.1	15.0
CSFV_17	CSFV_9 with 3C cleavage site	High	29.4	16.7
BDV_10	Full BDV E2 ectodomain	Low	1.0	11.1
APPV_9	Full APPV E2 ectodomain	High	4.3	31.0
NRPV_9	Full NRPV E2 ectodomain	High	8.6	6.3
BovCD46_2	bovCD46_SCR1- 2	High	10.3	26.0
BovCD46_4	bovCD46_SCR1- 4	High	24.4	20.4
BovCD46_5	Full bovine CD46 ectodomain	High	N/A	N/A
BovCD46_6	BovCD46_4 with 3C cleavage site	High	N/A	N/A
BovCD46_7	BovCD46_5 with 3C cleavage site	High	N/A	N/A
PigCD46_1	porCD46_SCR1-2	High	3.1	7.1
PigCD46_2	porCD46_SCR1-3	Medium	1.0	7.7
PigCD46_3	porCD46_SCR1-4	High	8.2	9.2
PigCD46_4	Full PigCD46 ectodomain	High	2.2	6.9
HCV_3	Full HCV E1 ectodomain with C-terminal Fos	Low	0.1	4.2
HCV-4	Full HCV E1 ectodomain	Low	N/A	N/A
HCV_7	Full HCV E2 ectodomain with C-terminal Jun	Medium	1.5	4.0
HCV_8	Full HCV E2 ectodomain	Medium	N/A	N/A

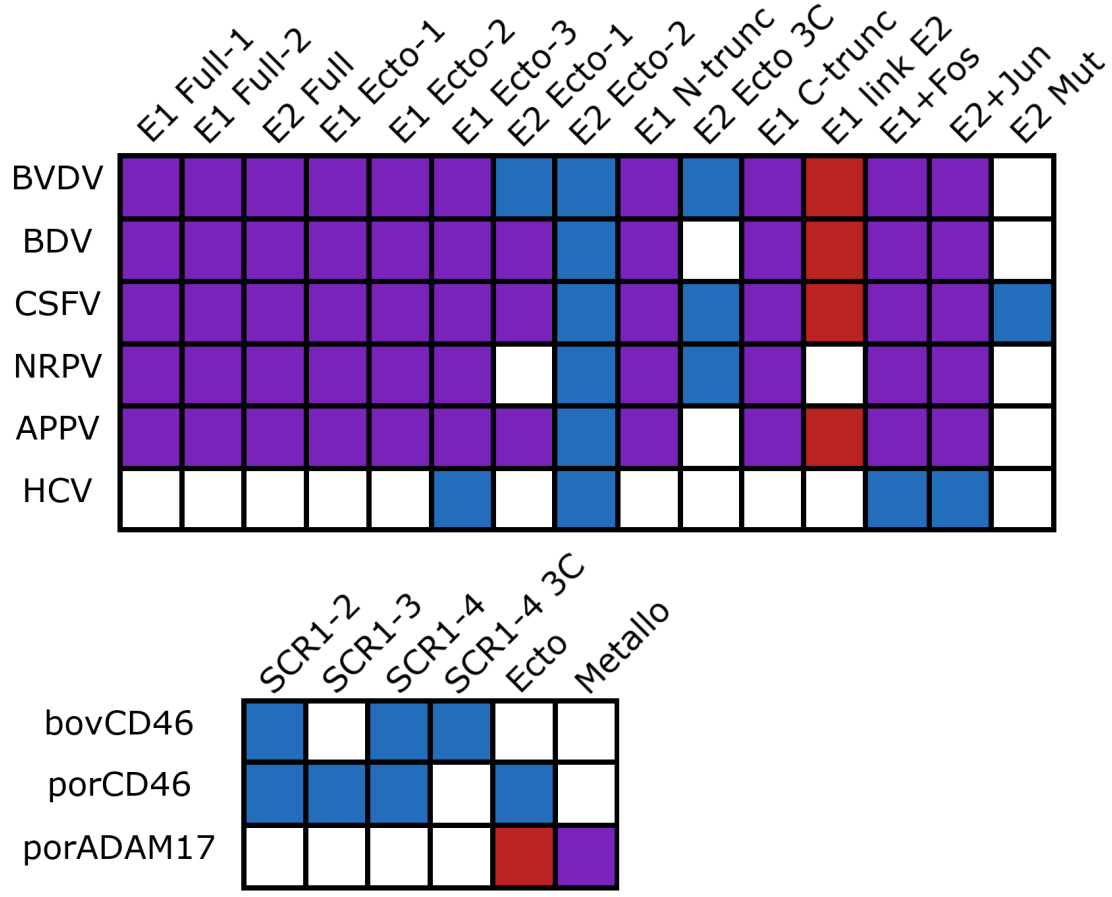


Figure 3.8: Summary table of cloning, expression and purification results. Red represents cloning failure, purple represents expression failure, blue represents large scale pure protein obtained, and white represents untested constructs.

This debris and media components must be removed before purification starts to ensure maximum protein yield. The first step is centrifugation of the harvested media at 5000G to pellet the large pieces of cellular debris and clumps of cells (Figure 3.7). The supernatant is then filtered through a 0.2µm filter using a vacuum pump to draw the media through. This removes smaller pieces of debris which would otherwise block the filters on the affinity purification column. It also sterilises the media so that it does not support a bacterial infection during the next media preparation step.

Cell culture media does not present the ideal conditions for immobilised metal affinity chromatography (IMAC) purification. Ideally both the pH and salt concentration would be higher, additionally the media contains some components which reduce the efficacy of column binding. As such the media underwent a buffer exchange process, an overnight dialysis step. The average media harvest was around 1.5 litres, from 6 roller bottles, therefore a large volume of buffer was needed to dialyse against. Large 50.8mm 12-14kDa dialysis tubing was used in a 20-litre brewing bucket filled with 15 litres of dialysis buffer overnight at room temperature with constant mixing. This process did not completely remove the media components (around 3 rounds of serial dialysis would be needed to achieve this) but did reduce them to such a level that they no longer interfere with affinity tag binding. After dialysis, the sample was ready to proceed to the next step, affinity purification. A QuixStand benchtop system could have been used instead of dialysis with the added benefit of reducing the sample volume at the same time, however using this machine takes a lot of active time and only one protein can be processed at a time. With dialysis multiple proteins can be processed simultaneously (normally between 2 and 4 in this study) and it occurs overnight without any active supervision. Finally, dialysis can take place over multiple days without having to book out any machines further increasing productivity.

3.4.2 Affinity Chromatography

The easiest way to remove a target protein from a large volume of solution containing many contaminants is to attach to the protein a tag which will bind a purification

column. The contaminating proteins in the media/buffer should not bind the column and can be washed away before the target protein is eluted. Initially tags made of large proteins were used, for example staphylococcal protein A[147], however newer and smaller tags have since been developed and while the original protein tags are still used the newer tags generally have improved binding. The some commonly used tags include the tiny polyhistidine tag[136], the larger glutathione S-transferase (GST)[148] and maltose binding protein (MBP)[149] tags, and those based on the biotin/streptavidin[150]) interaction. A histidine tag will bind a chelating sepharose column, which may be charged with a range of metals, the most common being Ni^{2+} and Co^{2+} (Talon resin), this is called immobilize metal affinity chromatography (IMAC). The GST tag from the N-terminus of the glutathione S-transferase molecule binds resin charged with glutathione. Proteins are sometimes expressed better and/or are more likely to be soluble with large soluble tags attached such as the GST and MBP tags[151]. The streptavidin biotin interaction is an extremely strong interaction, hindering elution, so special resin is used to weaken the interaction. Other tags are just starting to be used on a larger scale, these can have additional benefits such as the sumo tag[152] which has been seen to improve protein solubility and not require a recombinant linker region.

Advantages of the polyhistidine tag include its small size, ability to work in many different systems, and relative cheapness of reagents. A polyhistidine tag is generally a short sequence added to a protein containing around 6 -10 histidine residues next to each other. Prior to the invention of the histidine tag, this histidine/metal ion interaction was used with proteins containing naturally occurring histidine rich regions. Two histidine residues bind a metal ion and the additional histidine residues are there to increase affinity by binding more than one nickel ion[136]. In order to remove the protein from the column, once all the contaminants have been removed imidazole is added. Imidazole has a similar structure to the aromatic ring of histidine that binds the nickel, an excess of imidazole therefore competes with the histidine for binding to the nickel and the target protein is eluted.

The small size of the tag along with its lack of charge at physiological pH means it rarely affects protein function so in many cases does not need to be removed, furthermore although it does not tend to be ordered it tends not to interfere with crystal formation. The interaction works in the presence of denaturing agents, detergents, and even small amounts of reducing agents, in fact just about the only thing that must be avoid is large amounts of reducing or chelating agents[136]. Chelating sepharose is a little high in price but not as expensive as some alternatives and it can be reused many times with different metals if desired. The imidazole used for elution is not expensive and resources are not wasted producing large protein tags. When choosing which resin to use it is useful to know that the cobalt (Talon) resin will bind the histidine tag with greater selectivity but reduced affinity than the nickel resin meaning contaminants are less likely to bind but overall protein yield will be lower as well. With nickel resin contaminants may bind to some degree but the overall protein yield will be improved[153]. After testing both nickel and cobalt resin and discovering little difference in protein purity as the proteins purified here are secreted (and so have reduced contaminants), it was decided to use the cheaper and improved yield of the nickel resin.

3.4.3 Size Exclusion Chromatography

After affinity purification the samples were generally reasonably pure, but they were in a high salt and high imidazole buffer. Neither high salt nor high imidazole are good for crystallography as they can inhibit crystal formation. They are also present at such a high concentration that they overwhelm most of the crystallization conditions. To rectify this the buffer must be exchanged for one that is more suited to onward application of the protein. The three main options for this are dialysis, a desalting column, and size exclusion chromatography.

There is the potential that the elution contains multiple multimeric states of the protein which cannot be seen on an SDS-PAGE gel or potentially unusable aggregated protein. As such size exclusion chromatography was used for buffer exchange as it had the added benefit of separating proteins based on size and

shape. This means an additional purification step was added if any contaminants remain as well as separating different multimeric states and removing aggregates. Size exclusion chromatography works by pumping the sample through a column containing beads with channels of different sizes. When the sample passed through, the small molecules flow through the channels within and the large molecule go straight around the beads, therefore the small molecules have a longer journey and elute after the larger molecules. Unfortunately, this process will dilute the protein and in order to get the sharpest peaks possible with no overlap between them the sample applied to the column must be as low a volume as possible, column length and flow rate are also important factors to consider.

Spin concentrators were used to concentrate the sample to less than 500µl before application to the column. Some constructs are not as stable as others and some aggregation took place as the protein is concentrated. In order to minimise this aggregation, the concentration process was carried out at room temperature. For a particularly troublesome construct BVDV_5, the gel filtration was also carried out at room temperature rather than the standard 4°C. As these proteins are generally extremely stable and are functionally optimised for the mammalian extracellular environment at 37°C this is probably why moderate temperatures give the least aggregation. Normally protein purification is carried out at 4°C to minimise degradation and proteolysis, however, in this study room temperature was used instead. The proteins are stabilised by disulphide bonds and covered in N-linked glycosylation; both of these features may protect the protein from degradation. A superdex 200 column was chosen due to the majority of the proteins being studied having a suspected molecular weight (Mw) of around 80kDa, too large for the smaller s75 column. Initially the 10/300 size of gel filtration column was used as the expression scale was only 0.5 litres but subsequent purification utilised a longer 16/600 column as they expression scale was increased to 1.5 litres.

3.4.4 Results

Protein was successfully expressed and purified for all 16 constructs detailed in Table 3.1. This included the E2 ectodomain of all 5 pestiviruses as well as HCV and the E1 ectodomain of HCV. Pestivirus E1 proteins did not express which is in line with the hypothesis that E2 is a protein chaperone for E1. Also obtained were several constructs encompassing the ectodomains of both bovine and porcine CD46 protein. Figures 3.9 and 3.10 depict example SEC traces and related SDS-PAGE gels for each protein showing their final purity and SEC elution volume. Table 3.2 contains the SEC analysis results including Mw estimates.

The Mw estimates were higher than expected due to the elongated structure of these proteins and the presence of N-linked glycans. Most constructs were purified as putative dimers according to their SEC elution volumes and none were aggregated. This was expected for the E2 constructs as the literature reports the formation of homodimers and both papers reporting the BVDV E2 ectodomain crystal structure[21, 22] describe a homodimer. All proteins were stored at the highest concentration possible before aggregation started to occur. The exact concentration for each construct is detailed in Table 3.1 alongside the yield per litre of media. The proteins were stored at -80°C after flash freezing in liquid nitrogen. The buffer used for SEC for all proteins was also the final storage buffer: 10mM Tris pH8.0, 150mM NaCl.

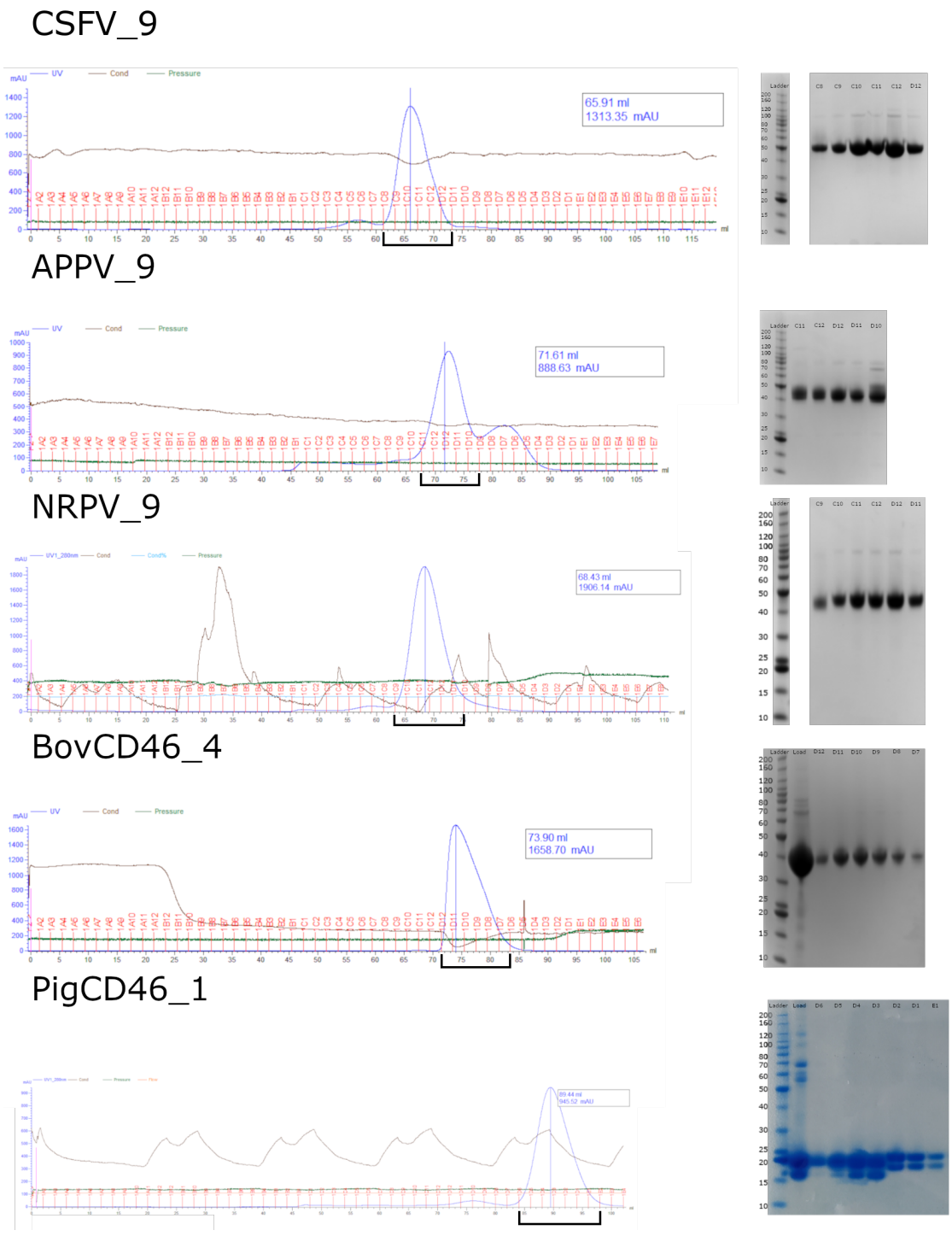


Figure 3.9: Example SEC traces and associated reducing SDS-PAGE gels for proteins purified on a 16/600 s200 gel filtration column. On the y axis of the SEC trace is an A280 measurement and on the x axis is the elution volume with 2ml fractions marked. Lanes on the gel are marked with the appropriate fraction and the region run on the gel is highlighted on the SEC trace with a black box.

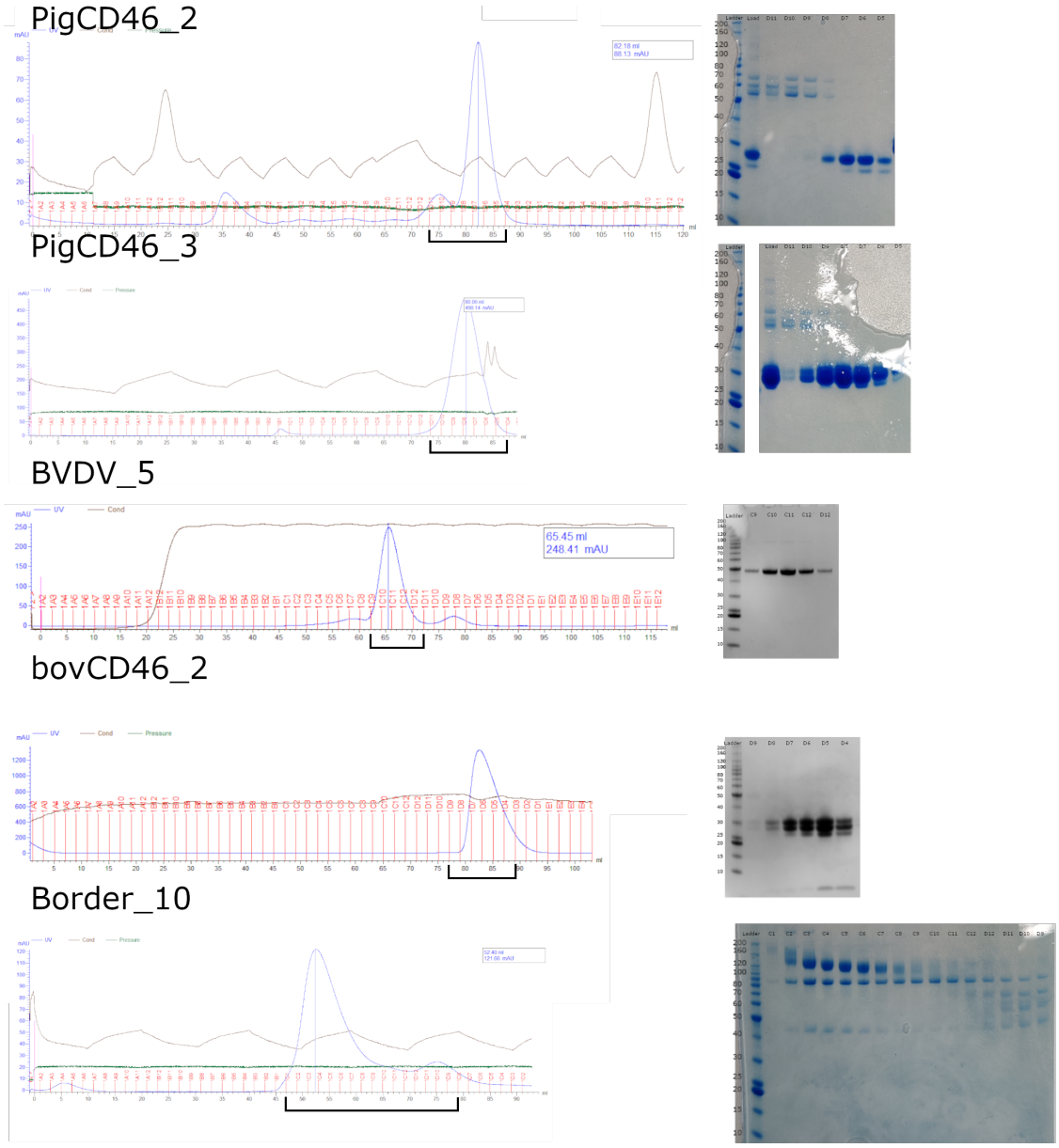


Figure 3.10: Example SEC traces and associated reducing SDS-PAGE gels for proteins purified on a 16/600 s200 gel filtration column. On the y axis of the SEC trace is an A280 measurement and on the x axis is the elution volume with 2ml fractions marked. Lanes on the gel are marked with the appropriate fraction and the region run on the gel is highlighted on the SEC trace with a black box.

Protein	Elution Volume (ml)	Estimated Mw (kDa)	Putative Multimeric State	Aggregated	Peak A280 (AU)	Peak Conc (mg/ml)
CSFV_9	65.91	218	Dimer	No	1.31	4.46
APPV_9	71.61	132	Dimer	No	0.89	1.60
NRPV_9	68.43	174	Dimer	No	1.91	5.70
bovCD46_4	73.90	108	Dimer	No	1.66	5.39
porCD46_1	89.44	27	Monomer	No	0.95	3.41
porCD46_2	82.18	52	Monomer	No	0.09	0.32
porCD46_3	80.06	62	Monomer	No	0.49	1.73
BVDV_5	65.45	227	Dimer	No	0.28	0.93
bovCD46_2	82.58	50	Dimer	No	1.33	4.65
BDV_10	52.40	>440	-	No	0.12	0.40

Table 3.2: Analysis of SEC results with estimated Mw calculated from the SEC elution volume

3.5 Validation Techniques

During the cloning step agarose gels are used to visualise the DNA and verify the size throughout the cloning process. There is an antibiotic selection region in the vector DNA so only clones containing the circular vector were able to grow on the antibiotic containing plate, meaning most of the clones will have a properly ligated insert present. After the final clone was obtained, the sequence was verified using Sanger sequencing of the region spanning the insert.

After expression a western blot with an anti-his antibody was used to verify that the protein was been produced and had the correct size. This was always done during the expression test but was also used if there were any questions as to whether a large-scale expression had worked or not.

The main method of verifying protein purity was direct visualisation of the final protein sample on an SDS-PAGE gel to see if there are any extra bands. If extra bands were seen, then these were sent for mass spectroscopy analysis to see if it is a stable protein multimer or a contaminating protein. The size exclusion chromatography step can also be a verification step as the volume at

which the protein elutes give information on the protein size and therefore if it is truncated, a multimer, or aggregated. As shown in Table 3.3 all E2 proteins elute at over double their approximate monomeric molecular weight (40kDa) so are likely dimeric, bovine CD46 is also likely dimeric although interestingly porcine CD46 appears to be monomeric. None of the proteins eluted at the void volume of approximately 40ml so are unlikely to be aggregated, although BDV_10 (BDV E2 ectodomain) did elute at 52ml indicating >440kDa size therefore it is clearly forming large multimers (Figure 3.8).

3.6 Summary

After an extensive literature search and multiple sequence alignment of all pestivirus species five were chosen to work on simultaneously to increase chances of success. The three most well-known and economically important species: BVDV, BDV, and CSFV affecting cows, sheep, and pigs respectively were chosen due to their impact to global food production. A further two species: APPV, and NRPV affecting pigs and rats were chosen due to interesting differences in their E2 glycoprotein sequence. APPV has an N-terminal truncation as well as little sequence similarity other than at the C-terminus. While NRPV has a full length E2 protein but has little sequence similarity in the N-terminal region. Of the three envelope glycoproteins in pestiviruses: E1, E2, and Erns only E1 and E2 are required for membrane fusion. As such only these two proteins were chosen for the focus of this thesis alongside two putative cellular receptors.

Some of the more recently discovered species do not yet have individual protein domain boundaries identified. The multiple sequence alignment and domain boundaries from other proteins were used to estimate the missing boundaries. The newly identified sequences were then used to identify possible transmembrane regions and possible disulphide bonds. 88 protein constructs were then designed of both the extracellular region and the entire proteins. Structures of the E2 protein of BVDV and of the similar HCV virus were used to inform the design of some

constructs. A secretion signal was added to the protein constructs as well as a polyhistidine tag for purification. Some other more radical constructs were designed to try and improve protein expression. These included flexible linkers, C-terminal additions, and removal of glycosylation sites (Figure 3.3).

The protein constructs were cloned into an expression vector suitable for use in both a transient and lentivirus expression system. The vector was based on the well-known pHL-sec vector[139] so inserts could be easily switched between different vectors and the N- and C-terminal regions easily swapped out for different tags for example. Initially a restriction digest cloning protocol was used as this was what the vector were designed to be used with. However, after multiple cloning issues and low efficacy a single step cloning reagent was used instead. After switching to single step cloning, efficacy increased to such a degree that dephosphorylation of digested vector (to prevent ligation of empty vector), *dpnI* treatment of PCR products (to remove template DNA) and clean-up of PCR products was no longer necessary. Colony screening was used to identify positive clones and the resultant PCR product sent for Sanger sequencing. This resulted in the obtainment of mini-prepped sequence verified plasmids after less than 3 days. When cloning point mutation constructs a two-step primer extension PCR method was used instead of a commercial point mutation kit because of the large vector length.

A mammalian expression system was used due to the extracellular location of the proteins as well their extensive glycosylation. The HEK293S cell line was chosen as it truncates glycosylation to a high mannose level and has the additional advantage of being able to control when protein expression starts. Both lentivirus and transient transfection were tested as a method of delivering DNA to the cells. Surprisingly in a small test 2 constructs expressed with the lentivirus method which did not with transient transfection. This had the upside of the easy creation of stable cell lines for future protein expression, and a much-reduced DNA requirement but the downside of a longer timescale to first protein expression. Expression was first tested in a 6 well plate format to identify well-expressing constructs. Ideally, for

maximum speed, this would have been carried out using transient transfection but since this method was not a good surrogate for the lentivirus expression, lentivirus was used also at small scale. Large scale expression was carried out in two litre roller bottles over 7 days with a double media harvest. Overall, nine pestivirus E2 protein constructs were expressed and purified along with seven receptor constructs.

After media harvest high speed centrifugation and filtration was used to remove cellular debris and any large contaminants. Overnight dialysis was carried out to remove any metal ions and histidine present in the grown media which might interfere with the initial IMAC column. Following dialysis an initial IMAC column was used to remove the protein from the media. Both cobalt and nickel resin were tested for use with the histidine tag however nickel was ultimately chosen as it has higher affinity for the histidine tag. Size exclusion chromatography was used as a final purification step to remove any remaining contaminants, separate different multimeric states, and remove protein aggregates. It also had the advantage of being an easy way to move the protein into a more physiological buffer. As these proteins were secreted for a week at 37°C and are normally exposed to the harsh extracellular environment the entire purification was carried out at room temperature. In fact, the BVDV constructs precipitated when exposed to low temperatures after purification. Overall, out of the 88 constructs designed 77 were successfully cloned, just under 30% resulted in protein expression and 15 were taken through to large scale expression and purification.

4

Cellular Receptors for Pestiviruses

Contents

4.1 Introduction 95
4.2 CD46 Receptor 99
4.3 Structure Solution of Bovine CD46 Domains 1-2 . . . 101
4.4 Structure Solution of Bovine CD46 domains 1-4 106
4.5 Structural Analysis of CD46 extracellular domain . . . 110
4.6 Structure Solution of Porcine CD46 Domains 1-4 . . . 116
4.7 Computational Modelling of Protein Structure 119
4.8 Structural Comparison of CD46 Ectodomain Structures 125
4.9 CD46 Dimerization 126
4.10 Putative BVDV Binding Interaction 133
4.11 CD46 Orientation on Cellular Surface 136
4.12 Conclusion 137

4.1 Introduction

To enter a cell, pestiviruses must first interact with a cellular receptor. This interaction is one of the main regulatory steps dictating the tissue tropism and host range. Following binding to a cellular receptor, or multiple cellular receptors, an enveloped virus is then able to enter the cell by fusing its membrane to the plasma membrane. The plasma membrane is the host cells main defence mechanism, and once breached the virus can access the necessary cellular resources to carry

out its life cycle. Information on virus receptor interactions can play a key role in deciphering the mechanism by which a virus may enter a cell. This is in addition to providing insight into potential vaccines or therapies.

Current knowledge on pestivirus cellular receptors is limited although it is of particular interest as pestiviruses have been hypothesised to have a novel fusion protein structure. Some cell surface components such as heparin sulphate and the LamR which are commonly used as cellular receptors have been implicated in pestivirus cell entry, however, they are thought to be attachment factors rather than cellular receptors and seem to interact with the Erns protein rather than E2[87, 90, 154]. A recent study by Yuan and Li *et al*[91] presented the first evidence that a different receptor, ADAM17, may be a cellular receptor for CSFV. The most researched putative receptor is the complement control protein CD46 which is a receptor for multiple human viruses. Published results indicate an interaction of bovCD46 with BVDV, and porcine CD46 (porCD46) with CSFV, and APPV[62, 87–89], although the research on CSFV is currently a little contentious[88].

4.1.1 Heparan Sulphate and Laminin Receptors

Heparan sulphate is a membrane associated glycosaminoglycan. These are unbranched polysaccharide chains, which have a net negative charge, linked to a protein core. The chains form a network covering the surface of most cell types. Many viruses use heparan sulphate as a cellular receptor as well as other pathogens. Erns of BVDV has been seen to bind heparan sulphate along with some other glycosaminoglycans, but not all[155]. Culture of CSFV in porcine kidney cells has been shown to select for CSFV variants which bind the host cells through heparan sulphate[154]. This could be through the acquisition of a S476R mutation in the Erns protein, although some variants lacking this mutation were also able to make the switch to use of heparin sulphate.

LamR has also been implicated in pestivirus cell entry. This protein is found both in the cytoplasm of the cell as well as in the cellular membrane and acts as

a cell surface receptor for laminin although there is little known about how this happens. Just like heparan sulphate and CD46 LamR acts as a cellular receptor for many viruses. LamR was shown to co-localised with CSFV viral particles on the surface of the cell[90]. Coimmunoprecipitation suggests that the interaction is likely to happen through the Erns protein. However, addition of laminin, anti-laminin receptor antibodies, or laminin receptor protein does not eliminate CSFV infection, so this receptor is unlikely to be necessary or sufficient for cell entry.

Interestingly, there appears to be a relationship between the use of heparan sulphate and LamR for cell entry by CSFV[90]. Although removal of either heparan sulphate or LamR does not prevent CSFV infection, removal of both receptors does drastically reduce infection. Additionally, increased expression of the laminin receptor does not increase the infection rate if there is ample heparan sulphate present on the surface of the cell.

Current indications are that pestiviruses interact with both heparin sulphate and LamR through the Erns protein, these interactions are not essential for cellular entry, but might help recruiting viruses on the cell surface to increase the probability of binding to other specific receptors. As this study does not focus on Erns due to the fact it is not essential for pestivirus cell entry, neither heparin sulphate nor LamR were further studied.

4.1.2 ADAM17 Receptor

ADAM17 is also known as tumour necrosis factor-α-converting enzyme. The A disintegrin and metalloproteinase (ADAM) family of proteins are involved with multiple cellular processes such as cell adhesion, cell fusion, proteolysis, and cell signalling[156]. These proteins are extracellular membrane anchored enzymes, however 8 of the 21 members do not have any proteolytic activity[157]. Proteolysis of some cell surface proteins to release the soluble active form is an important post-transcriptional modification. ADAM17 was first discovered in 1997[158, 159] when it was identified as catalysing the proteolysis of TNF-α, however it catalyses

the proteolysis of far more than just tumour necrosis factor alpha (TNF-α), at last count it was implicated in the processing of over 80 substrates[160]. The protein is composed of several domains, arranged in N- to C-terminal order these are: a cytoplasmic domain, a transmembrane domain, a membrane proximal domain, a disintegrin domain, a metalloproteinase domain, and finally a prodomain[157].

A recent study by Yuan and Li *et al*[91] identified ADAM17 as an essential attachment factor for CSFV. Initially ADAM17 was identified as a result of a pull-down experiment utilising the CSFV E2 protein and CSFV permissive PK15 cells. No binding was detected with non-permissive BHK-21 cells. Additionally, the binding could be prevented with the addition of the CSFV neutralizing antibody V3. This is particularly interesting as it is the first-time direct binding between a pestivirus E2 and a cellular receptor has been described. The study also described the knockout of ADAM17 in permissive porcine cells resulting in prevention of E2 binding. This binding was restored by stable expression of porcine ADAM17 cDNA. In contrast to CD46, ADAM17 does not appear to affect CSFV host tropism as expression of ADAM17 from mice and humans restored E2 binding in porcine ADAM17 knockout cells to a similar level as expression of porcine ADAM17. The study also identified the metalloproteinase domain of ADAM17 as being the site of E2 binding through introducing a H405D point mutation to prevent zinc binding. Expression of this mutant ADAM17 failed to restore E2 binding in knock-out cell lines. Additionally, SPR experiments with soluble E2 and ADAM17 metalloproteinase domain showed an increased binding affinity with the addition of zinc.

This study was published only six months before thesis submission so there was little time to carry out structural studies. As described in chapter 3, a gene for the porcine ADAM17 ectodomain was designed and ordered. Two protein constructs were designed and one, encompassing the metalloproteinase domain, was successfully cloned. Unfortunately, the protein did not express and there was no time remaining to test further constructs.

4.2 CD46 Receptor

The most well-known putative pestivirus receptor is CD46 which is also known as membrane cofactor protein (MCP). This type 1 transmembrane protein was first discovered in 1986 and is identified as a complement regulatory protein, or CRP[161]. The protein is encoded in a gene also called CD46 which is located in the regulators of complement activation (RCA) gene cluster on chromosome 1q32 alongside other members of the RCA family. Upon activation of the complement cascade CD46 responds by binding the two complement activation fragments deposited on the host cell, C3b and C4b and acting as a cofactor for serine protease factor-1[161]. In addition to its role in complement regulation, CD46 has been linked with other cellular processes such as T cell activation, cell metabolism, reproduction, and autophagy[162].

An ever-rising number of pathogens use CD46 as a cellular receptor leading to it being labelled a pathogen magnet[163]. As of 2021, these include 11 human pathogens (6 types of bacteria and 5 different viruses) as well as members of the Pestivirus genus for bovine CD46 and porcine CD46 and 2 bacterial pathogens in the case of teleost CD46[93]. The reason behind CD46 being utilised by so many pathogens to facilitate cell entry is because it is expressed on all cell types (with a few exceptions). Another reason behind the heavy use of CD46 is its other cellular functions such as its complement regulatory activity, immune modulating signalling functions, and internalization mechanisms[164]. Interestingly, different parts of the CD46 extracellular region are targeted by different pathogens[93]. Pathogen interaction with CD46 can result in changes to various cellular processes potentially through changes to the CD46 cytoplasmic tail such as cleavage. This may potentially change the immune microenvironment or help break through the host immune defences via changes in protein expression, cytoskeletal rearrangement, apoptosis, and immunosuppression[164].

As indicated above bovine and porcine CD46 may act as a cellular receptor for BVDV, and APPV respectively. Pestiviruses are species specific, and this has

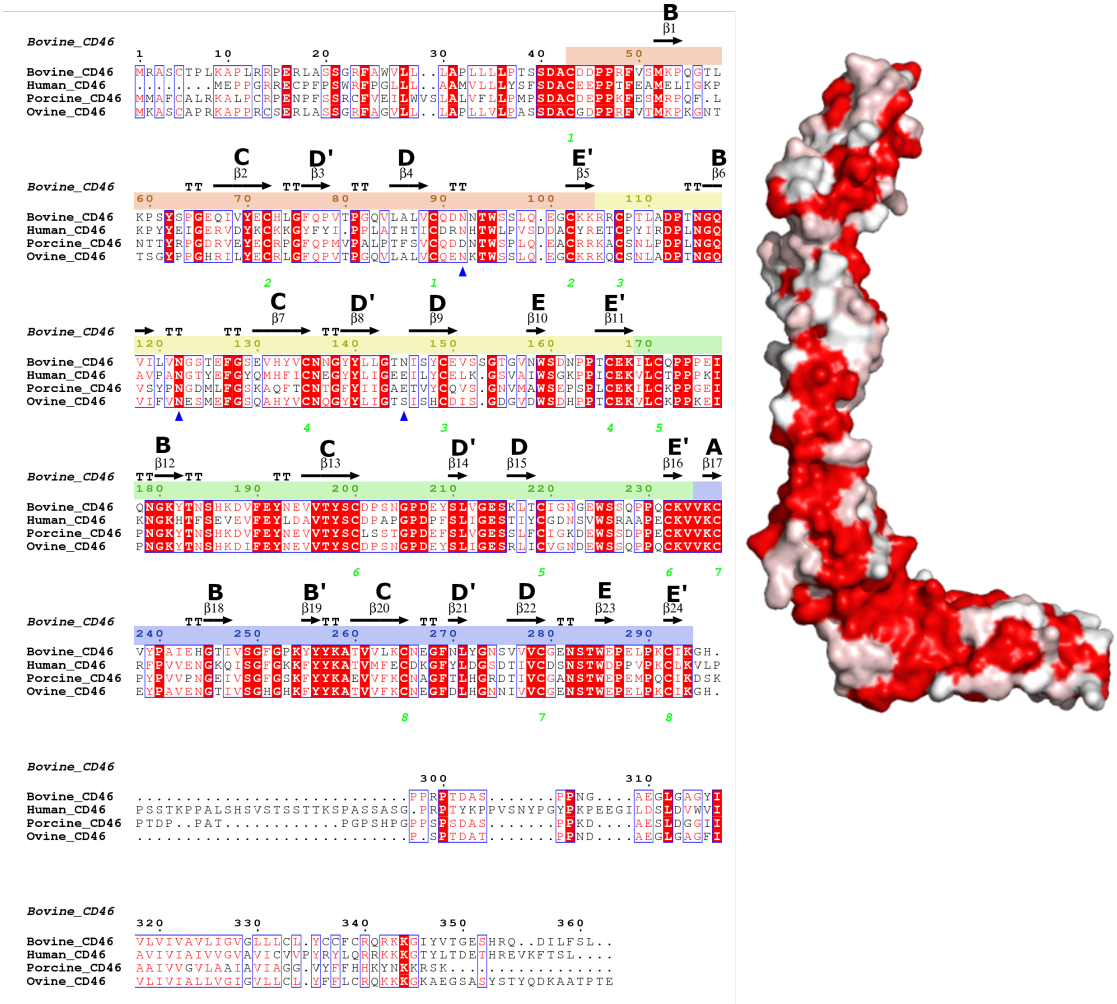


Figure 4.1: Left- Sequence alignment of CD46bov with CD46 from humans, pigs, and sheep generated using Clustal Omega[103] and ESPript web service[165]. Conserved residues are highlighted in red and similar residues are in red text. Secondary structure from CD46bov is shown above and SCR beta sheet labelling has been added. The sequence numbering for each SCR has been highlighted as: SCR1-orange, SCR2- yellow, SCR3-green, SCR4-blue. Green numbering below shows cysteines involved in disulphide bonds and glycosylated asparagines are marked with a blue triangle. The genbank accession numbers: Bovine- Q6VE48, Human- P15529, Porcine- O02839, Ovine- W5PWS0. **Right**- Structure of human CD46_SCR1-4 with the surface representing sequence conservation on a scale of non-conserved white residues to conserved red residues.

been shown to extend to CD46 with bovCD46 conferring increased susceptibility of porcine cells to BVDV[62]. Deletion of the BVDV binding site on bovCD46 results in a vast reduction in BVDV cell entry events but has little effect on cell surface motility[95]. Two peptides in bovine CD46, $E_{66}QIV_{69}$ and $G_{82}QVLAL_{87}$, were shown to be important in BVDV interaction. When modelled on the human CD46 structure these peptides form a single binding patch. Porcine CD46 has also been shown to be a cellular receptor for APPV utilising porCD46 knockout cell lines. APPV was unable to infect CD46 knockout porcine cell lines but was able to infect wildtype cell lines[88]. A study by Dräger *et al* suggested porcine CD46 may also play a role CSFV interaction as antibody blocking of porcine CD46 reduced CSFV infection[87]. However, it has been demonstrated that both CSFV and bungowannah pestivirus were able to infect two CD46 knockout porcine cell lines. Additionally, CSFV was able to infect a porCD46 deficient porcine lymphoma cell line while APPV was unable to[88]. Clearly CD46 is important in pestivirus infections but does not encompass the entire story, and very likely other cellular receptors are involved.

High-resolution X-ray crystallography structures are available for the human CD46_SCR1-2 (PDB: 1CKL)[134] and SCR1-4 (PDB: 3O8E)[94] which have 41% and 48% sequence identity respectively to bovine CD46 respectively (Figure 4.1). Since pestiviruses are species specific an X-ray crystallography structure of a relevant animal CD46 could provide valuable insight into future control stratagems. Insight into which region of the virus interacts with the cellular receptor could inform targeted vaccine design and blocking of the virus interaction region on the receptor could provide a potential treatment option.

4.3 Structure Solution of Bovine CD46 Domains 1-2

4.3.1 Crystallisation and Optimisation

In order to solve the structure of bovine CD46, several protein constructs covering both the entire protein sequence and smaller domains of the protein were cloned,



Figure 4.2: Crystal of bovCD46_SCR1-2

and their expression tested (Chapter3). Although the full-length ectodomain protein construct did not express, two constructs covering regions of the ectodomain were successful. One of these, bovCD46_SCR1-2, encompassed the ectodomain region SCR1-SCR2 covering residues C43 to I169 (Figure 3.3, bovCD46_2). After large scale expression and purification 15.4mg of pure, stable, soluble protein was obtained (Table 3.1). The bovCD46_SCR1-2 construct was put into crystallisation trials with BVDV E2 ectodomain construct BVDV_5 in the hope of crystallising the E2/receptor complex. The two proteins were mixed in a 1:1.1 molar ration (slight excess of bovCD46_SCR1-2) with a total protein concentration of around 13mg/ml. The JCSG+[106] screen from molecular dimensions was used and the plate incubated at 20°C.

Initial crystals appeared after 8 days in multiple conditions. 7 crystals were harvested, and diffraction quality was tested. One crystal grown in the 1:1 protein to well solution drop in 0.5M Tris pH 8.5, 20% PEG 8k, and 0.2M MgCl2, diffracted well and a dataset was collected. This crystal is shown in Figure 4.2, it is a jagged cluster of approximately 150nm by 75nm.

4.3.2 Data Collection and Processing

Crystals were cryoprotected by flooding the drop with a mixture of well solution with 25% ethylene glycol cryoprotectant. Crystals were harvested from the drop with a litholoop and flash cooled in liquid nitrogen. During data collection, the crystal remained under a gaseous nitrogen stream at 100K to prevent devitrification. A single dataset was collected at the I03 beamline at Diamond Light Source Ltd. 25% transmission was used for 0.025 second exposures with a wavelength of 0.976Å and flux of 9.23×10^{11}, 3600 images were collected with an omega rotation of 0.1°. The data were processed with autoPROC[108] implemented as part of the Diamond processing pipeline[120] in space group $P4_12_12$, to a resolution of 2.33Å with completeness 99.4 and I/sig(I) of 10.19, full statistics are shown in Table 4.1.

Wavelength	0.98
Resolution range	44.12 - 2.33 (2.41 - 2.33)
Space group	*P* 41 21 2
Unit cell	91.09 91.09 121.12 90.00 90.00 90.00
Total reflections	570316 (55868)
Unique reflections	22417 (2168)
Multiplicity	25.40 (25.80)
Completeness (%)	99.40 (99.45)
Mean I/sigma(I)	10.19 (1.12)
Wilson B-factor	52.03
R-merge	0.23 (3.60)
R-meas	0.24 (3.67)
R-pim	0.05 (0.72)
CC1/2	1.00 (0.74)
CC*	1.00 (0.92)
Reflections used in refinement	22295 (2163)
Reflections used for R-free	1123 (112)
R-work	0.23 (0.35)
R-free	0.26 (0.38)
CC(work)	0.95 (0.81)
CC(free)	0.91 (0.81)
Number of non-hydrogen atoms	2129
macromolecules	1956
ligands	111
solvent	62
Protein residues	262
RMS(bonds)	0.01

RMS(angles)	0.90
Ramachandran favored (%)	97.64
Ramachandran allowed (%)	2.36
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	1.77
Clashscore	2.49
Average B-factor	79.12
macromolecules	78.91
ligands	94.22
solvent	58.84
Number of TLS groups	4

Table 4.1: Data Collection and Refinement Statistics for bovCD46_SCR1-2

4.3.3 Structure Solution and Refinement

The bovCD46_SCR1-2 structure was solved with molecular replacement using SCR1-2 from the human CD46_SCR1-4 protein structure (PDB: 3O8E). From the calculation of Matthews coefficient, 1.26% and 2.31% solvent for a BVDV E2 dimer and a monomer of bovCD46_SCR1-2, it was clear that there was only a single protein in the asymmetric unit and that a protein complex (bovCD46-BVDVE2) had not been obtained. PHASER[110] from the phenix[166] suite of programs was used to solve the structure.

Initially, the molecular replacement solution was calculated by PHASER[110] with a TFZ score of 16.8 and a LLG of 241.6. The solution placed the SCR2 domain well while the SCR1 domain was poorly resolved, this was surprising given the reasonable sequence similarity of the model, 39.1% and percentage of the sequence covered, 99.2% (as calculated by clustal omega[103]). In order to place SCR1 MR-SAD was run through AutoSol[111] which also built some of the structure to an Rwork/Rfree of 37.50/39.83%. A combination of REFMAC[113] refinement and manual building in COOT[112] was used to complete and refine the model leading to a final Rwork and Rfree of 26.25/25.55% at 2.33Å resolution respectively with good stereochemistry. Refinement statistics are shown in Table 4.1 and example electron density in Figure 4.3. There were two molecules in the asymmetric unit

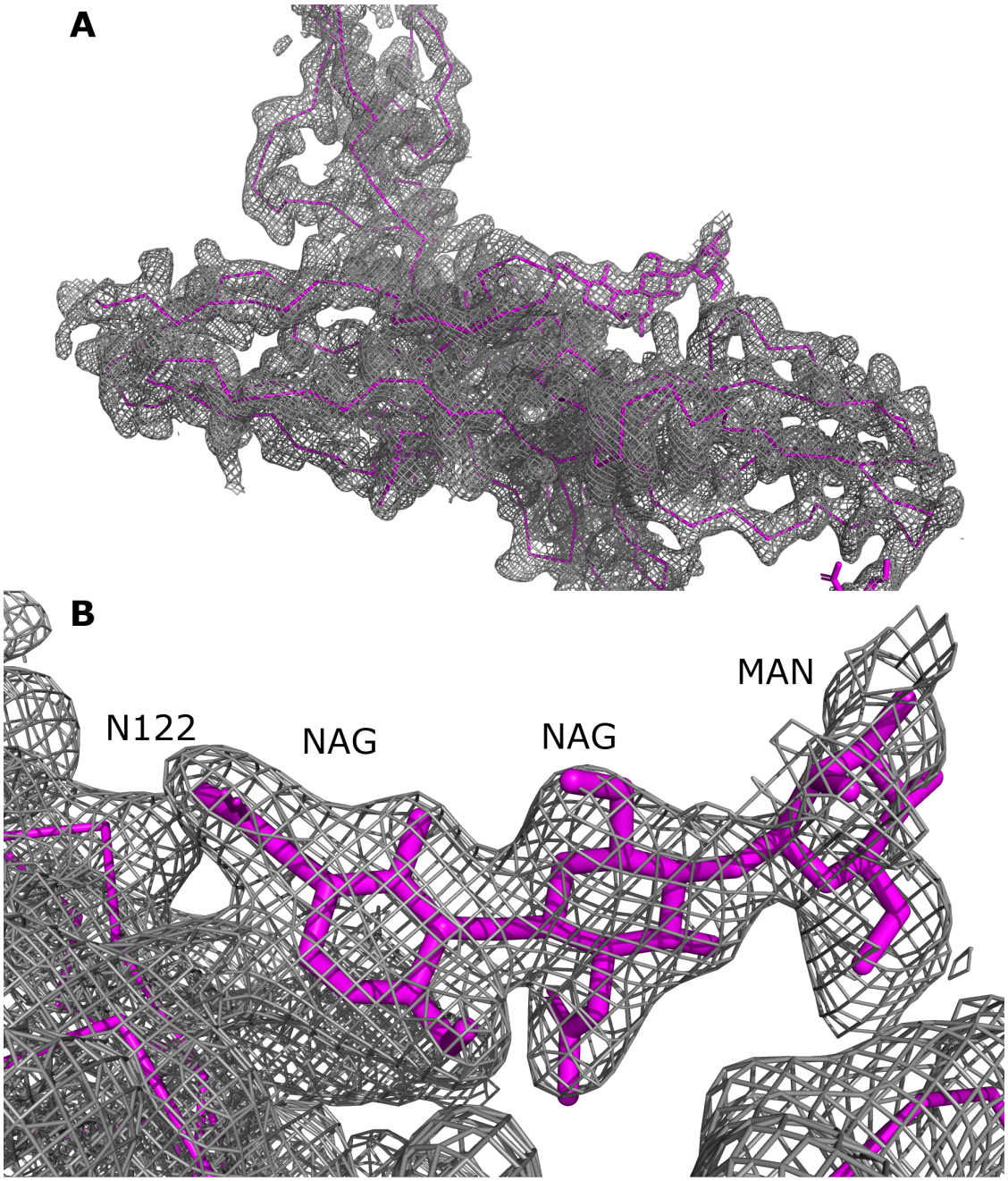


Figure 4.3: - Electron Density Maps (2mFo-DFc generated using phenix map creation tool[167]) for bovCD46_SCR1-2 with sigma level 1. **A-** Entire structure, two monomers in magenta carbon alpha trace. **B-** Glycosylation at N122 shown as sticks.

and the final structure encompassed the full protein construct, residues C43 to I169. An additional threonine and glycine were also resolved at the N-terminus, these are from the AgeI cloning restriction site.

4.4 Structure Solution of Bovine CD46 domains 1-4

4.4.1 Crystallisation and Optimisation

The second protein construct of bovine CD46 ectodomain which produced crystals was bovCD46_SCR1-4. This construct encompassed the ectodomain region SCR1-SCR4 covering residues C43 to K294 (Figure 3.3, bovCD46_4). After large scale expression and purification, 36.6mg of pure, stable and soluble protein were obtained as described in chapter 3. The protein was put into crystallisation trials at 20mg/ml at 20°C in a sitting drop vapour diffusion with approximately 480 different conditions tested. Crystallisation trials were carried out with both glycosylated and deglycosylated protein. Further details on the crystallography and deglycosylation process can be found in section 2.7.1.

After testing of the initial crystals a condition from the BCS screen[107] (well D6), was chosen for optimisation. Optimisation techniques used included multiple grid screens around the two crystal conditions, use of additive screen GRAS and micro seeding into fine grid screens, and deglycosylation of protein. Eventually a brute force approach was successful, after testing of 70 different crystals, two diffracted to better than 3Å resolution. Both crystals came from a wide grid screen of the BCS D6 condition with deglycosylation protein. The crystal used to solve the structure was grown in 14% PEG smear broad, 0.1M sodium cacodylate tetrahydrate pH 6.0, 5% glycerol, and 0.2M calcium chloride. This crystal can be seen in Figure 4.4 it was a long chunky needle of more than 300µm in length.



Figure 4.4: Crystal of bovCD46_SCR1-4 visualised under UV light

4.4.2 Data Collection

Crystals were cryoprotected by flooding the drop with a mixture of well solution with well solution containing 20% glycerol, harvested from the drop and flash cooled in liquid nitrogen. Data were collected on the Diamond beamline I03 and where the crystal remained under a gaseous nitrogen stream at 100K to prevent devitrification.

A data set was collected with a wavelength of 0.976Å, 100% transmission and flux of 1.26x10^{12} with 0.2 second exposures, 3600 images were collected with 0.1° omega rotation. Data were processed with autoPROC[108]/Staraniso[114] implemented as part of the Diamond processing pipeline[120] in space group $P4_12_12$, with completeness 43.91, I/sig(I) of 7.10, and extending to a resolution of 4.28Å along the a and b axes, and 2.73.Å along the c axis. Full statistics are shown in Table 4.2.

Wavelength	0.98
Resolution range	91.73-2.84(3.24-2.84)
Space group	P 41 21 2
Unit cell	129.73 129.73 120.19 90.00 90.00 90.00
Total reflections	278989 (11165)
Unique reflections	10884 (544)
Multiplicity	25.60 (20.50)
Completeness (spherical	44.10 (7.00)
Completeness (ellipsoidal)	93.30 (73.30)
Mean I/sigma(I)	7.10 (1.80)

Wilson B-factor	50.37
R-merge	0.49 (2.23)
R-meas	0.50 (2.28)
R-pim	1.00 (0.47)
CC1/2	1.00 (0.63)
Reflections used in refinement	10871 (39)
Reflections used for R-free	528 (1)
R-work	0.25 (0.39)
R-free	0.29 (0.28)
Number of non-hydrogen atoms	3505
macromolecules	3435
ligands	70
Protein residues	453
RMS(bonds)	0.02
RMS(angles)	1.84
Ramachandran favored (%)	95.50
Ramachandran allowed (%)	4.50
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.25
Clashscore	14.55
Average B-factor	87.48
macromolecules	88.24
ligands	50.18
Number of TLS groups	7

Table 4.2: Data Collection and Refinement Statistics for bovCD46_SCR1-4

4.4.3 Structure Solution and Refinement

The bovCD46_SCR1-4 structure was solved by molecular replacement using the human CD46_SCR1-4 protein structure (PDB: 3O8E). The CCP4[168] suite of programs was used to solve the structure, Phaser[110] was used for molecular replacement, then combination of REFMAC[113] and COOT[112] was used to build and refine the model to Rwork and Rfree of 25.28/28.58% respectively Data collection and refinement statistics can be found in Table 4.2. There are two molecules in the crystallographic asymmetric unit. The final structure encompassed the full protein structure, residues C43 to G295. The final glycine was not resolved, most likely due to flexibility, whilst an additional threonine and glycine were resolved at the N-terminus, from the *AgeI* cloning restriction site. Due to flexibility and the

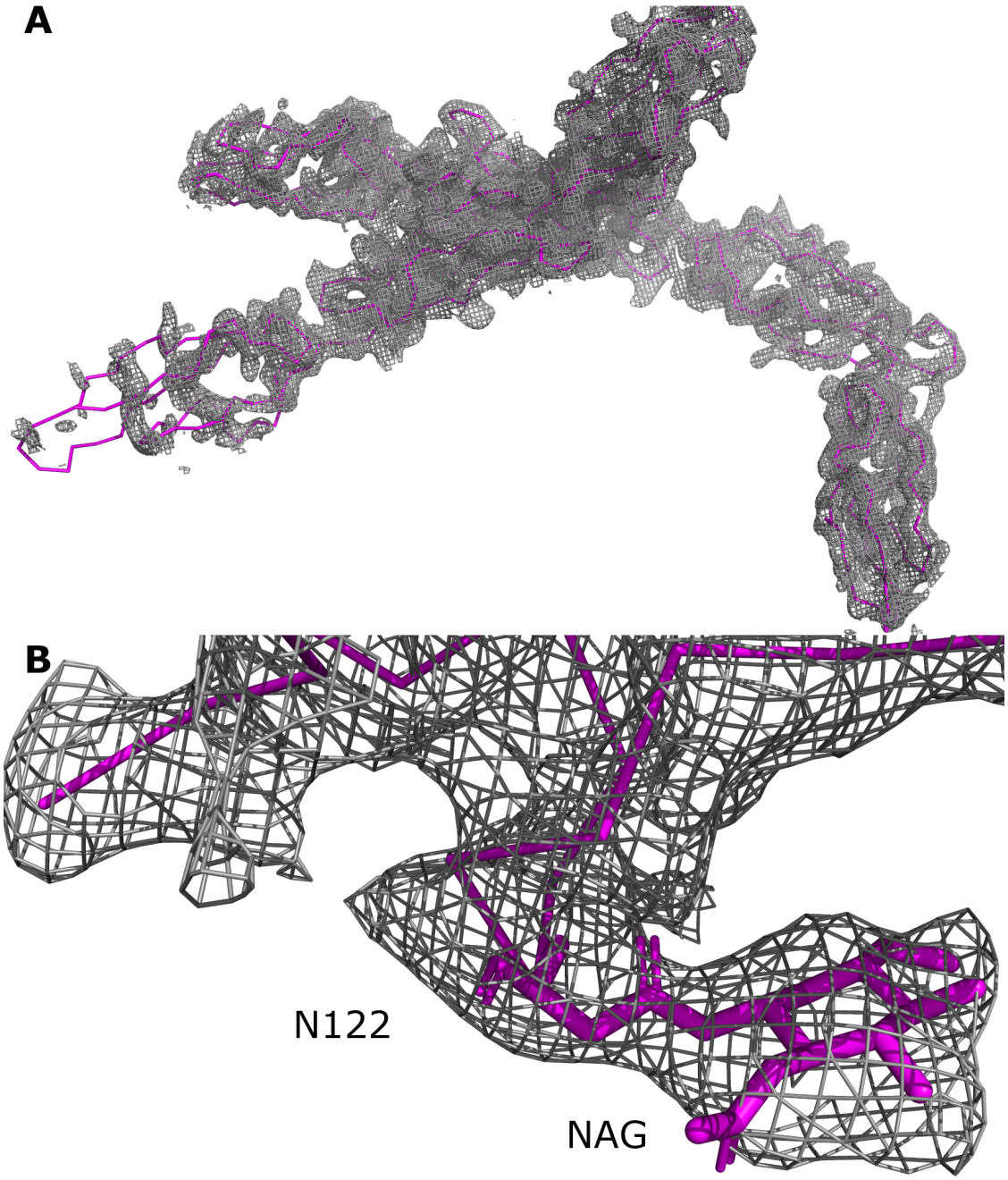


Figure 4.5: - Electron Density Maps (2mFo-DFc generated using phenix map creation tool[167]) for bovCD46_SCR1-4 with sigma level 1. **A-** Entire structure, two monomers in magenta carbon alpha trace. **B-** Glycosylation at N122 shown as sicks.

low resolution (also likely due to interdomain flexibility) there was little electron density for the SCR4 domain in the second molecule in the asymmetric unit so it could not be built (Figure 4.5).

4.5 Structural Analysis of CD46 extracellular domain

4.5.1 Overall fold

As can be seen in Figure 4.6, the overall fold of the bovine CD46 is a highly elongated structure, mainly composed of β-strands and loops, extending to around 112Å. All RCA family members contain a varying number of SCRs. CD46 has four SCR domains (Figure 1) although the number found in other RCA family members varies from 4 to 30[52]. In CD46, the SCR domains are found at the N-terminus of the protein and make up the majority of the protein, this region is located in the extracellular environment. Following the SCR domains, there is a 11 residue linker region abundant in serine, proline, and threonine residues (STP), a membrane spanning anchor, and finally a cytoplasmic tail at the C-terminus of the protein. Splicing generated several distinct isoforms with changes in the STP region resulting in varying O-glycosylation as well as different cytoplasmic tails[169].

The bovCD46_SCR1-4 structures contain SCR1 through to SCR4 arranged head to tail in an almost linear fashion (Figure 4.6). However, there is a bend in the C-terminal section of the structure caused by a much-reduced angle between SCR3 and SCR4 in both structures. This bend gives the structure an unusual appearance which has been compared to the shape of a hockey stick (Figure 4.6). In this analogy, the N-terminal domains SCR1-SCR3 form the shaft and are located distal to the membrane and the final domain SCR4 forms the blade at the C-terminus of the protein, proximal to the membrane. As the electron density is better resolved at the centre of the bovCD46_SCR1-4 structure, it is likely that the overall ectodomain is reasonable flexible leading to the limited electron density at either end of the monomer and better-defined resolution near the crystal contacts in the centre of the protein. There are two molecules in the asymmetric unit of bovCD46_SCR1-4, one has all four domains visible but in the other there is only density for SCR1-3. Despite this difference both molecules are otherwise alike with a low RMSD of 0.60Å across the 193 residues in SCR1-3 (Figure 4.7).

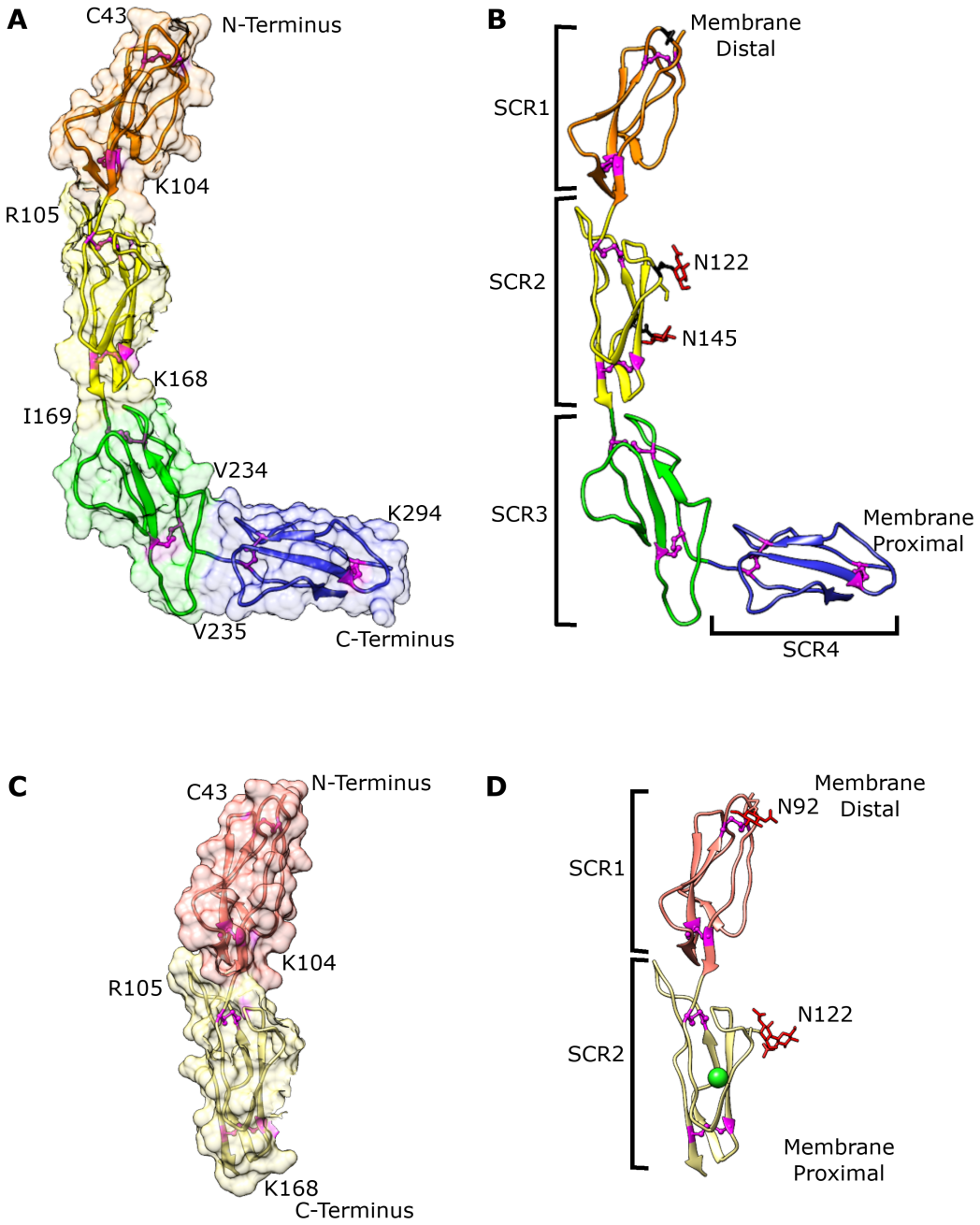


Figure 4.6: **A-** Structure of SCR1-SCR4 monomer in a cartoon representation with surface shown. SCR1 at the N-terminus is orange, SCR2 is yellow, SCR3 is green, and SCR4 at the C-terminus is blue. Cysteines and therefore disulphide bridges are shown in magenta balls and sticks. Domain boundaries are marked with N-terminal boundary at the top left of the domain and the C-terminal at the bottom right. **B-** Cartoon representation of SCR1-SCR4 monomer without surface with position in relation to membrane marked and SCRs labelled. The glycan residues are shown in red sticks and the asparagines which are glycosylated in at least one monomer are labelled and shown in black sticks. **C-** As in A but for the SCR1-SCR2 monomer. **D-** As in B but for the SCR1-SCR2 monomer.

The bovCD46_SCR1-2 structure consists of SCR1 and SCR2 arranged head to tail in a linear fashion. Again there are two molecules in the asymmetric unit, both could be fully built and were near identical with an RMSD of 0.28Å across all 129 residues. They adopt the same conformation as in the bovCD46_SCR1-4 structure with and RMSD of 0.50Å across the 129 residues when superposed (Figure 4.7). The slightly elevated RMDS is due to differences in some loops, a similar RMSD (0.60Å across the 194 residues in SCR1-3) is seen between the two molecules in the asymmetric unit of bovCD24_SCR1-4 and is caused by the lower resolution resulting in poor electron density in these loop regions (Figure 4.5).

The SCR region of the protein exposed on the cell surface contains N-linked glycosylation. There are five potential glycosylation sites in the SCR1-SCR4 region of the bovine CD46 ectodomain. According to the NetNGly server[170] there are two high probability sites, N122, and N145 both located in SCR2. There is also a medium probability site at N158, again in SCR2, and two lower probability sites at N92 and N282, located in SCR1 and SCR4 respectively. In the bovCD45_SCR1-2 two sites can clearly be seen at N92 and N122 (Figure 4.6) with density present for two N-acetyl-β-D-glucosamine (NAG) residues at both sites. A single mannose (MAN) can also be visualised at N122 in one monomer. In the bovCD46_SCR1-4 structure, evidence of glycosylation can also be seen at N122, although in this case only a single NAG residue is present due to the deglycosylation treatment with Endo Hf. Interestingly, a third glycosylation site can be seen at N145 in the bovCD46_SCR1-4 structure which was not resolved in the bovCD46_SCR1-2 structure. The final predicted glycosylation site in the C-terminal domains, SCR4, could not be seen in the bovCD46_SCR1-4 structure. However, SCR4 is the least well resolved part of the structure, indeed it is only resolved on one of the two the monomers in the asymmetric unit, likely due to its flexibility and the fact that it is not stabilised by crystal contacts (Figure 4.5 and 4.7). Nevertheless, this glycosylation site is conserved throughout different species (Figure 4.1), so is likely to be present but not resolved due to flexibility.

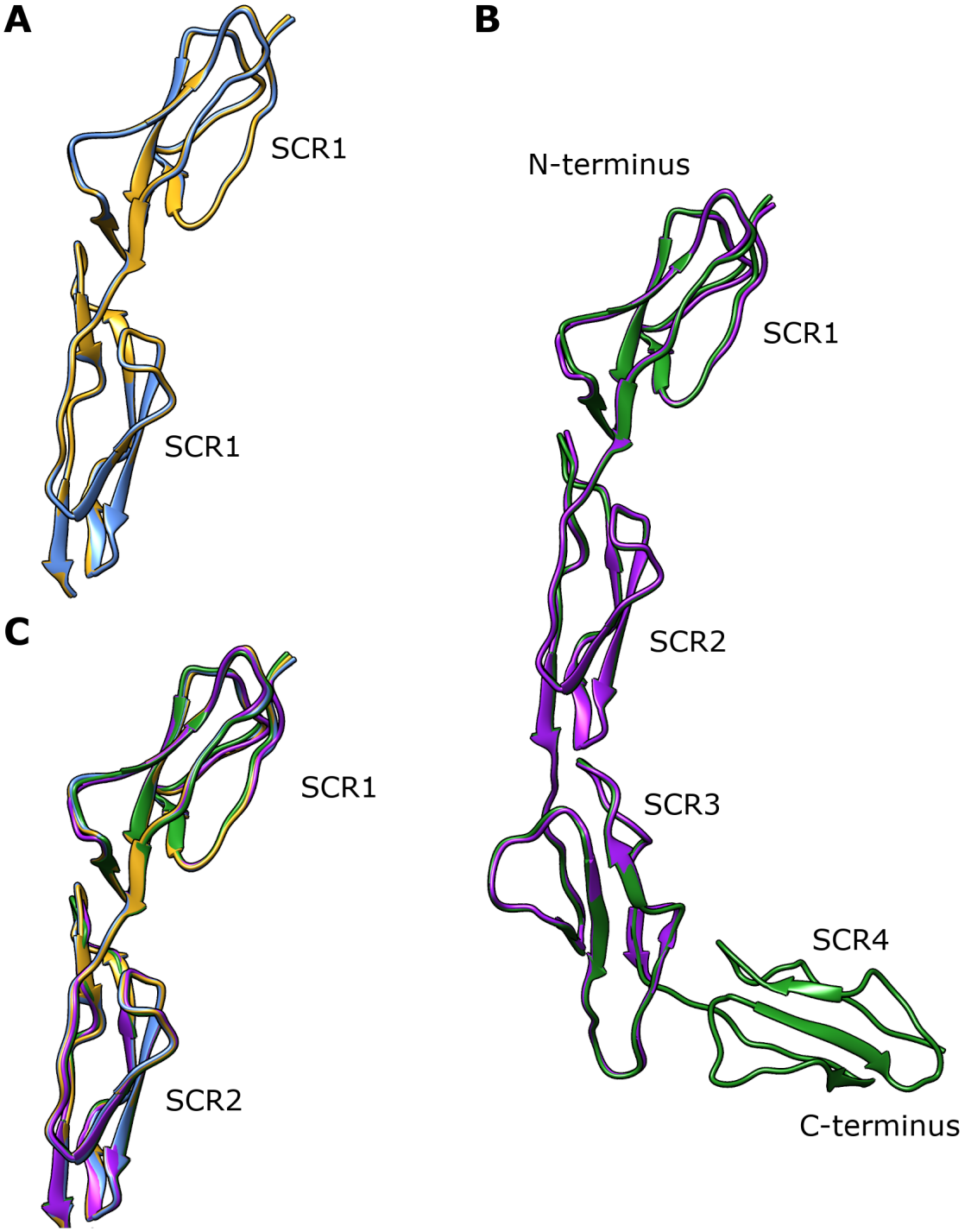


Figure 4.7: Superposition of bovine CD46 structures. **A-** Two molecules in the asymmetric unit of bovCD46_SCR1-2 structure, RMSD: (0.28Å across all 129 residues). **B-** Two molecules in the asymmetric unit of bovCD46_sSCR1-4 structure, RMSD: (0.60Å across all 193 residues). **C-** SCR1 andSCR2 from the bovCD46_SCR1-2 and bovCD46_SCR1-4 structures, RMSD: (0.50Å across all 129 residues), coloured as in A and B.

Intriguingly, all three glycans are arranged along the concave face of the protein leaving the convex side bare to the solvent. The fourth predicted site at N282 may also follow this trend if it is glycosylated due to the location of N282 on the top side of SCR4 facing the concave side of the protein.

The previously published human CD46_SCR1-2 structure showed a bend between the two domains and some interdomain flexibility[134]. Here they postulated that the DE-loop of SCR2 adopted an extended conformation causing steric interference with residues from SCR1 and therefore the bent conformation[134]. This contrasts with the bovCD46_SCR1-2 structure presented here (Figure 4.6) which has a linear arrangement without any bends. However, a later study presented a ligand bound structure of humanCD46_SCR1-2 which also has a linear arrangement, this study hypothesised that the protein was straightened upon binding of the ligand[171]. The bovCD46_SCR1-2 structure presented here has a linear arrangement despite not having ligand bound. Therefore, it is more likely that this domain interface is inherently flexible irrespective of ligand bound state.

The overall structure of the bovCD46_SCR1-4 adopts a similar conformation to the humanCD46_SCR1-4 structure of an elongated protein of over 100Å with a hockey stick like shape. The expected linear arrangement is changed by the presence of a bend between SCR3 and SCR4. The bend is due to the insertion of five hydrophobic residues in the CD' loop of SCR3 (Figure 4.8) which first identified in the humanCD46_SCR1-4 structure. This hydrophobic insertion is discussed in more detail in section 4.5.2. Upon examination of sequence alignments (Figure 4.1) it is clear that this is also present bovine as well as both porcine and ovine CD46[94]. Other members of the RCA family do not have this hydrophobic insertion so its presence in all these mammalian CD46 points to an important functional role.

4.5.2 SCR Domain Fold

RCA family proteins are composed of several SCR domains of around 60 amino acids which consist of a central four stranded β-barrel surrounded by a few small

β-strands and flexible linker loops. In the SCR nomenclature, the central β-strands are named B, C, D, E and the smaller β-strands A, B', D', and E' (Figure 4.1). Not every SCR domain contain all eight β-strands, as is the case for the structures presented in this thesis, although some semblance of a central β-barrel is present in all. Upon superposition of the different CD46 SCRs (Figure 4.8), the core of the SCR domain is structurally conserved. The entire domains have RMSD of between 3.70Å and 4.60Å over the 60-64 residues, the higher RMSD are due to the number of flexible loops joining the β-strands.

As seen in Figure 4.8, there are two interesting differences in the loop conformations which are seen in SCR1 and SCR3. Firstly, a loop shift can be seen in SCR1 in the D'D loop, this encompasses the putative BVDV binding site which will be described in section 4.10. The second difference involves an extension of the CD' loop in SCR3 (Figure 4.8) which affects the SCR3:SCR4 domain interface. An insertion of five hydrophobic residues in this loop causes it to adopt an extended conformation which protrudes from the bottom of the domain. The protrusion perturbs the head to tail domain interface and causes the bend between SCR3 and SCR4 domains, resulting in the hockey stick like appearance.

As can be seen in Figure 4.6 and 4.8, each SCR contains four cysteine residues which form two disulphide bonds, located at the top and at the bottom of the SCR domain. For clarity, the four cysteines found in each SCR are named: Cysteine 1 (C43 in SCR1) bonds to cysteine 3 (C89 in SCR1) and cysteine 2 (C72 in SCR1) bonds to cysteine 4 (C102 in SCR1) (Figure 4.1). In the high-resolution structure of bovine CD46_SCR1-2, one of these disulphide bonds it is not fully oxidised. There is electron density supporting two different cysteine conformations in one of which the disulphide bond is broken, this is likely due to radiation damage.

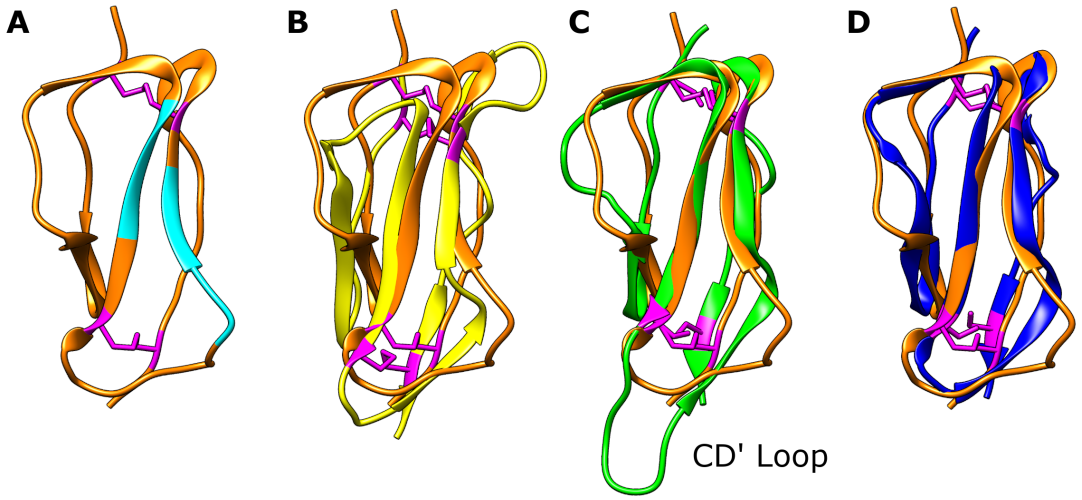


Figure 4.8: Structure comparison of individual SCR domains, SCR domain from bovCD46_SCR1-4 shown as cartoon style with cysteines coloured magenta. **A-** Cartoon representation of SCR1 of bovCD46_SCR1-4 coloured in orange with the BVDV interaction site highlighted in cyan. **B-** Cartoon representation of SCR1 of bovCD46_SCR1-4 coloured in orange superposed with SCR2 in yellow, RMSD 4.11Å over 60 residues. **C-** Cartoon representation of SCR1 of bovCD46_SCR1-4 coloured in orange superposed with SCR3 in green, RMSD 4.62Å over 64 residues. **D-** Cartoon representation of SCR1 of bovCD46_SCR1-4 coloured in orange superposed with SCR4 in blue, RMSD 3.65Å over 60 residues.

4.6 Structure Solution of Porcine CD46 Domains 1-4

4.6.1 Crystallisation and Optimisation

Four protein constructs covering regions of the extracellular domain porcine CD46 were successfully cloned and expressed. One of these, porCD46_3, encompassed the ectodomain region SCR1-SCR4 covering residues C45 to K294 (Figure 3.3, porCD46_3). After large scale expression 4.1mg of pure, stable and soluble protein was obtained (from 2 bottles/0.5l), further details can be found in Chapter 3. Initial crystallisation trials were carried out with 9.2mg/ml protein at 20°C in a sitting drop vapour diffusion experiment with a total of approximately 480 different conditions tested. Both glycosylated and deglycosylated protein was tested, further details on crystallisation and deglycosylation can be found in section 2.7.1.

Initial crystals appeared after 12 hours in 0.1M Hepes pH 7.5, 4.3M sodium

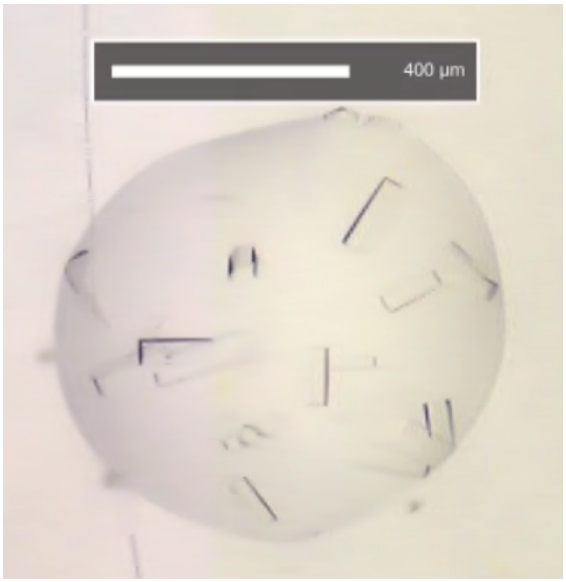


Figure 4.9: Crystals of porCD46_SCR1-4

chloride and were fully grown after only two days. Crystals were present for both glycosylated and deglycosylated protein with the deglycosylated crystals in the 2:1 ration drops and the glycosylated crystals in both the 2:1 and 1:1 ratio drops. These crystals displayed poor diffraction quality, however a crystals appeared in a second condition after 1 day and was fully grown after 3 days. These crystals were grown in 0.1M Hepes pH7.5, 1.5M lithium sulphate with glycosylated protein in the 2:1 ratio drop. These cubic rods of around 50µm by 150µm, as seen in Figure 4.9, diffracted a little further to around 6Å.

4.6.2 Data Collection

Crystals were cryoprotected by flooding the drop with a mixture of well solution with 25% glycerol added for cryoprotection. Crystals were then harvested from the drop with a Litholoop and flash cooled in liquid nitrogen. Crystals were immediately loaded into the I23 in-vacuum beamline vessel for data collection the following day. During data collection the crystal were cooled by heat conduction via the goniometer to around 50°K.

Diffraction was tested at the Diamond light source I23 beamline. Most crystals produced very poor diffraction and smeary spots, although one crystal from the

lithium sulphate condition, diffracted to around 6Å resolution. A single dataset was collected with a wavelength of 2.76Å at 100% transmission for 0.1 second exposures, 3600 images were collected with an omega rotation of 0.1°. Data were processed with autoPROC[108], autoPROC[108]/Staraniso[114] (although the data was not very anisotropic), and xia2[120]_3dii-XDS[119] implemented as part of the Diamond processing pipeline details can be found in Table 4.3. It is possible that there are two molecules in the asymmetric unit with a Matthews coefficient of 2.01 and a solvent content of 38.81%, although the poor resolution is likely have been caused by a higher solvent of 69.40% (Matthews coefficient of 4.02) if there were only one molecule in the asymmetric unit. High calculated solvent contents are often seen with glycosylated proteins as the flexible glycan molecules reduce the number of allowable packing options.

	autoPROC	STARANISO	Xia2_3Dii-XDS
Wavelength	2.76	2.76	2.76
Resolution range	59.02 - 6.03 (6.03 - 6.14)	68.16 - 4.73 (6.11 - 4.73)	68.22 -4.63 (4.71 -4.63)
Space group	*P* 6 2 2	*P* 6 2 2	*P* 6 2 2
Unit cell	136.31 136.31 57.36 90.00 90.00 120.00	136.31 136.31 57.36 90.00 90.00 120.00	136.44 136.44 57.40 90.00 90.00 120.00
Total reflections	28913 (1739)	32495 (6681)	60088 (2866)
Unique reflections	923 (48)	1052 (211)	1936 (93)
Multiplicity	31.30 (36.20)	30.90 (31.70)	31.00 (30.80)
Completeness (%)	100.00 (100.00)	62.30 (25.80)	100.00 (100.00)
Mean I/sigma(I)	7.20 (0.70)	6.60 (1.80)	3.60 (0.50)
R-meas	0.57 (8.30)	0.72 (3.36)	1.17 (6.86)
CC1/2	1.00 (0.60)	1.00 (0.50)	1.00 (0.30)

Table 4.3: Data set statistics for porCD46_SCR1-4 processed using three different programs

4.6.3 Structure Solution Failure

Molecular replacement was attempted using the human CD46_SCR1-4 protein structure (PDB: 3O8E), the previously solved bovCD46_SCR1-4 structure, as well as a RoseTTAFold models[117]) (generated using the robetta.bakerlab server[172] of the porCD46_SCR1-4 (described in section 4.7.3). These all failed to generate

phases using both MOLREP[118] and PHASER[110] in the CCP4[168] suite and phaser-MR[110] from the PHENIX[166] suite of programs. This is likely due to the low 6Å resolution. There was not enough anomalous signal from sulphur atoms to carry out MR-SAD.

4.7 Computational Modelling of Protein Structure

4.7.1 Introduction

A very recent advance in the field of structural biology is the development of a new type of computational protein structure modelling using deep learning. Traditionally computational modelling has been very limited and relied heavily on a structural homolog already being available. Two structure prediction programs utilising deep learning neural networks were made available in 2021: RoseTTAFold[117] and AlphaFold2[116]. Both programs work by using multiple tracks and jumping between them thereby simultaneously considering patterns in different information streams. AlphaFold2 uses two tracks while RoseTTAFold uses three.

AlphaFold by DeepMind was first showcased in 2018 at the 13th CASP (Critical Assessment of Structure Prediction) assessment which uses recently discovered but not deposited protein structures as a blind test for participating structure prediction programs. It performed exceptionally well successfully predicting high accuracy structures for 24 out of the 43 while the next best program only predicted high accuracy for 14 structures[173]. At the 14th CASP assessment in 2020, DeepMind presented a new program called AlphaFold2 which performed even better obtaining a median backbone accuracy RMSD of 0.96Å and an overall atom accuracy RMSD of 1.50Å[116]. This is an incredible feat and proves the program is beginning to present an alternative method for structure generation.

RoseTTAFold was developed in response to AlphaFold2s results at the 14th CASP assessment therefore it has not been entered into a CASP assessment itself as these only run every other year. The CAMEO (Continuous Automated

Model EvaluatiOn) experiment provides an alternative blind test using protein structures as they are submitted to the pdb. In this test, it outperformed every other server evaluated using 69 medium and hard targets. When the program was tested on the CASP14 targets it achieved a $C_{\alpha\,\mathrm{RMSD}}$ of 2.60Å[117]. The reduction in accuracy when compared to AlphaFold2 is though to be due to reduced computational resources as RoseTTAFold runs much faster than the version of AlphaFold2 submitted to CASP14. It is also possible that the reduced accuracy could also be due to DeepMind having access to greater deep learning expertise.

4.7.2 Structural Modelling of Bovine CD46 Domains 1-4

The initial step in the prediction process for these new structure prediction programs is a multiple sequence alignment they work particularly well with subjects which have high sequence similarly with proteins already deposited in the protein data bank. The bovCD46_SCR1-4 structure has not been deposited in PDB but has high sequence identity (47.2% as calculated by clustal omega[103]) to the human CD46_SCR1-4 structure covering the same region. Therefore, it is a perfect subject to test how accurate the predictions may be for CD46 protein from other species. As such the Robetta.bakerlab server[117] and AlphaFold2[116] were used to generate five models of the SCR1-SCR4 region of the bovine CD46 with the results shown in Figure 4.10.

As can be seen in Figure 4.10 both programs are very precise in their structure predictions for these proteins. The RoseTTAFold models have an RMSD of 0.68Å over all 253 residues while the AlphaFold2 models were even more similar with an RMSD of 0.30Å over the 253 residues. The higher RMSD for the RosettAFold structures is caused by slight differences in loop regions. The confidence in the reported structures in shown in Figure 4.10 and is high throughout the entirety of both sets of predictions, >80/100 for the AlphaFold2 structures and RoseTTAFold structures. When comparing the two sets of predictions there the main differences are in the positioning of the C-terminal domain and the structure of various loops in the N-terminal domain, SCR1. On comparison with the experimental structure,

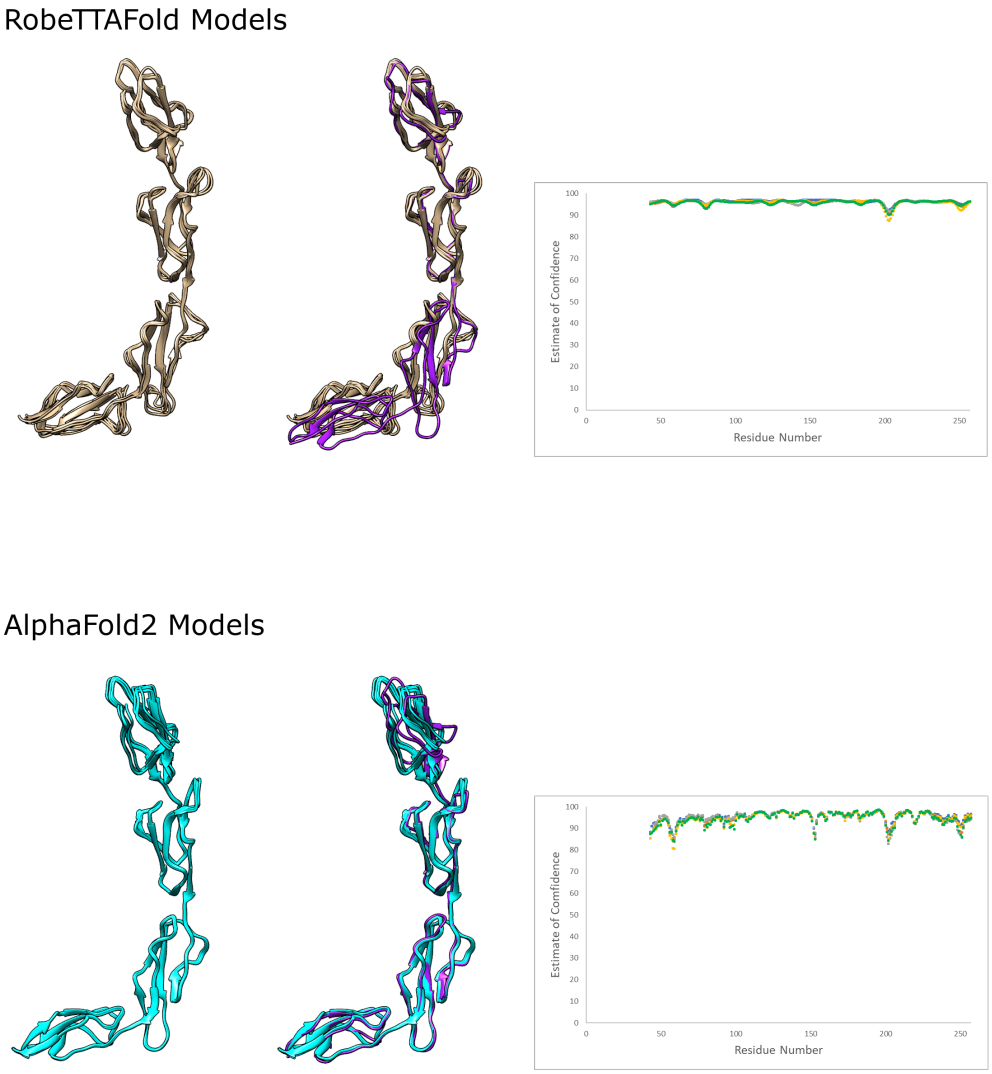


Figure 4.10: Computationally generated models of bovCD46_SCR1-4 generated by the AlphaFold2 and RoseTTAFold programs. For each program on the top are the five individual models generated and underneath the five models are superposed on each other (RMSD for RoseTTAFold models: 0.68Å over 253 residues, RMSD for AlphaFold2 models: 0.30Å over 253 residues) and then on one monomer of the crystallographic structure which is shown in purple (RMSD for RoseTTAFold models: 13.79Å over 252 residues, RMSD for AlphaFold2 models: 1.88Å over 252 residues). At the right are the confidence out of 100 per residue for the RoseTTAFold and AlphaFold2 models. The RoseTTAFold confidence estimate was generated by assuming an estimated error of 2Å was equivalent to 90/100 confidence so 100-(5*estimated error (Å)) was used to generate the confidence estimates per residue.

it becomes clear that the AlphaFold2 are more accurate than the RoseTTAFold structures. The RoseTTAFold structures have an RMSD of 13.79Å over the 252 residues while the AlphaFold2 models have a much lower RMSD of 1.88Å over the 252 residues. This is especially true when considering the positioning of the C-terminal domain, SCR4, and the SCR3:SCR4 interface angle. Here RoseTTAFold predicts an increased angle which is not supported by the crystallographic structure, while the AlphaFold2 predictions match the crystallographic structure.

Given these results it is clear both programs generate very acceptable structural predictions. However, the AlphaFold2 models were more accurate overall, especially in the C-terminal region.

4.7.3 Structural Modelling of Porcine CD46 Domains 1-4

As the human CD46_SCR1-4 structure has 50% sequence identity with the porcine sequence covering the same region and the accurate models generate for bovine CD46 it is likely that the model generated for porcine CD46 will be accurate as well. Therefore, both the Robetta.bakerlab server[117] and AlphaFold2[116] were used to generate five models of the SCR1-SCR4 region of the porcine CD46.

Both programs predicted a highly elongated structure composed of four SCR domains. SCR1-SCR3 are arranged linearly in a head to tail arrangement but there is a decreased angle between SCR3 and SCR4 causing a bend in the structure (Figure 4.11). This bend is due to a 5-residue hydrophobic insertion in the CD' loop of SCR3[94] and causes the structure to resemble a hockey stick. As with the bovine CD46 prediction both programs gave very precise modelling prediction. The RMSD for the AlphaFold2 models was just 0.51Å over the 250 residues and the RoseTTAFold RMSD was a little higher at 0.80Å over the 250 residues. Most difference were slight shifts in loop regions. Both modelling programs had good certainty with RobeTTAFold and AlphaFold2 having >80/100 certainty (Figure 4.11). Like the bovine CD46 modelling the main difference between the AlphaFold2 and RoseTTAFold models was the positioning of the C terminal domain SCR4 with

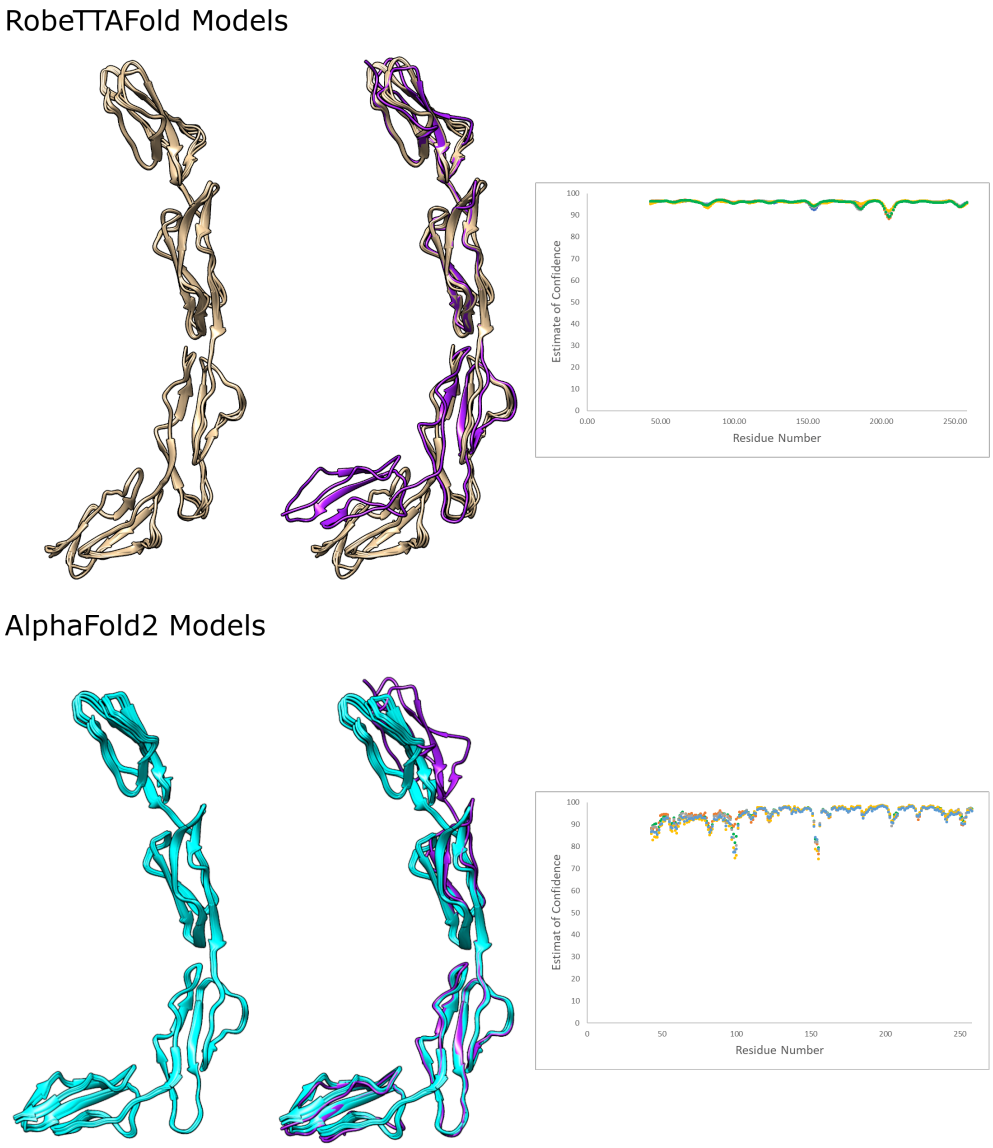


Figure 4.11: Computationally generated models of porCD46_SCR1-4 generated by the AlphaFold2 and RoseTTAFold programs. For each program on the top are the five individual models generated and underneath the five models are superposed on each other (RMSD for RoseTTAFold models: 0.80Å over 250 residues, RMSD for AlphaFold2 models: 0.51Å over 250 residues) and then on one monomer of the crystallographic bovCD46_SCR1-4 structure which is shown in purple. At the right are the confidence out of 100 per residue for the RoseTTAFold and AlphaFold2 models. The RoseTTAFold confidence estimate was generated by assuming an estimated error of 2Å was equivalent to 90/100 confidence so 100-(5*estimated error (Å)) was used to generate the confidence estimates per residue.

RoseTTAFold predicting a shallower angle between SCR3 and SCR4 as compared to AlphaFold2. As the bovine CD46 crystallographic has the sharper angle between SCR3 and SCR4 and because RoseTTAFold made a similar error when predicting the bovine CD46 structure, it is most likely that the AlphaFold2 model is more accurate.

Porcine CD46 is has a slightly different predicted glycosylation profile to bovine CD46. There is one predicted site at N61 in SCR1 and a second with much lower potential at N283 according to the NetNGlyc server[174]. It is likely the N59 is an alternate site due to the lack of a site at the N92 equivalent position as in the BovCD46_SCR1-2 structure K59 and N92 are only 16Å apart and N92 is conserved in other pestiviruses (Figure 4.1). The N122 equivalent is conserved but not predicted to be glycosylated and the N145 and N158 are only found in bovine CD46. As the crystal structure has not been solved these cannot be confirmed.

4.8 Structural Comparison of CD46 Ectodomain Structures

As can be seen in Figure 4.12, human, bovine, and most likely porcine CD46 ectodomains, all have a very similar structure. They are all highly elongated with 4 SCR domains arranged head to tail. As expected from the multiple sequence alignment (Figure 4.1), they all contain a hydrophobic insertion to SCR3 causing an elongated CD' loop (Figure 4.8). This has the result of blocking the head to tail interface of SCR3 and SCR4 causing a bend in the structure and the resulting hockey stick like appearance. Unsurprisingly, the bovine and porcine CD46 structures are most similar while the human CD46 structure has some slight differences. These are mainly in the C-terminus at the SCR3 to SCR4 interface which has a slightly reduced angle in the human CD46.

The RobettaFold models of the procine CD46 had one glaring difference again in the C-terminus. In these models, there was an increased angle at the SCR3 SCR4 interface causing the molecular to appear straighter overall. This is particularly interesting as the CD' loop which was initially thought to cause the bend is very similar in all three structures. The SCR4 domain in the RobeTTAFold porcine CD46 is rotated in relation to the rest of the molecule. The B'B loop adopts a different conformation which would cause steric interference with the CD' if SCR4 were to orient as it does in the other structures. However, it is likely this region has been incorrectly modelled by RobettaFold as it modelled the bovine CD46 with a similar overall structure which was incorrect when compared to the crystallographic model.

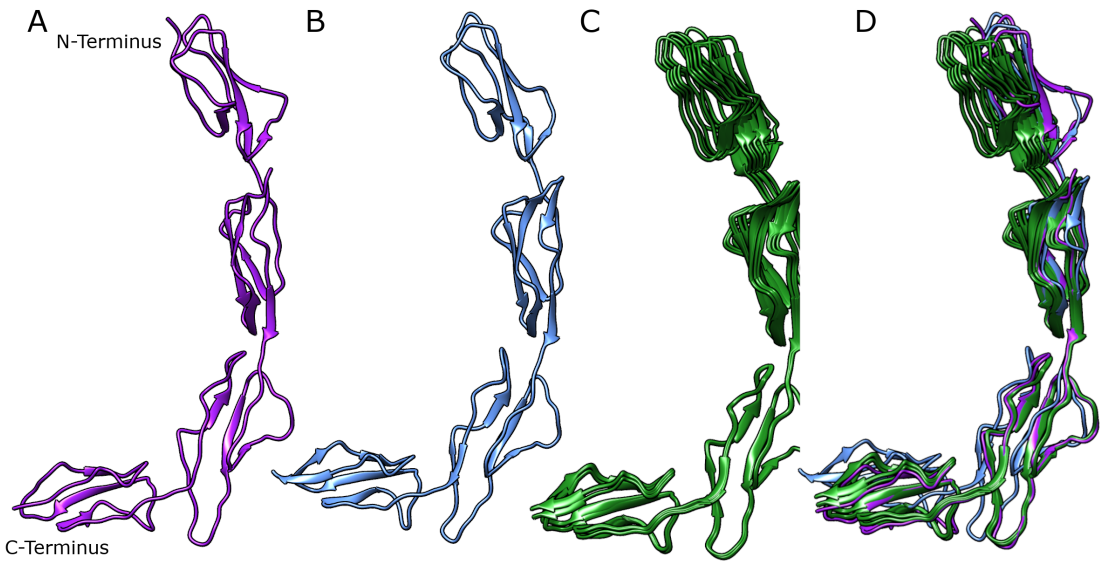


Figure 4.12: Structural comparison of CD46 ectodomains. **A**- Cartoon model of CD46bov1-4 coloured in purple. **B**- Cartoon model of human CD46 1-4 (PDB: 3O8E) coloured in blue. **C**- Cartoon model of all 5 AlphaFold2 computationally generated models of Porcine CD46 1-4 superposed and coloured in green. **D**- Superposition of A, B, and C with colours as previous.

4.9 CD46 Dimerization

4.9.1 Crystallographic Dimer

In both bovCD46 crystal structures, there are two monomers in the asymmetric unit, interestingly these appear to form a dimer. This crystallographic dimer is particularly interesting as it is identical in both protein constructs despite them varying in size, unit cell, and space group (Figure 4.13). The dimer interaction occurs at the top of SCR2 in one monomer as well as at the SCR1-2 interface for both copies of the protein (Figure 4.14).

PISA server[175] results indicate an area of interaction for both constructs of 577Å for bovCD46_SCR1-2 and 581Å for bovCD46_SCR1-4. This is within the range of interaction areas seen for antibody:antigen interactions. Furthermore, the PISA server suggests that there are multiple stabilising interactions, 14 hydrogen bonds and salt bridges in total at the interface in the high-resolution bovCD46_SCR1-2 structure, these hydrogen bonds are shown in Figure 4.13.

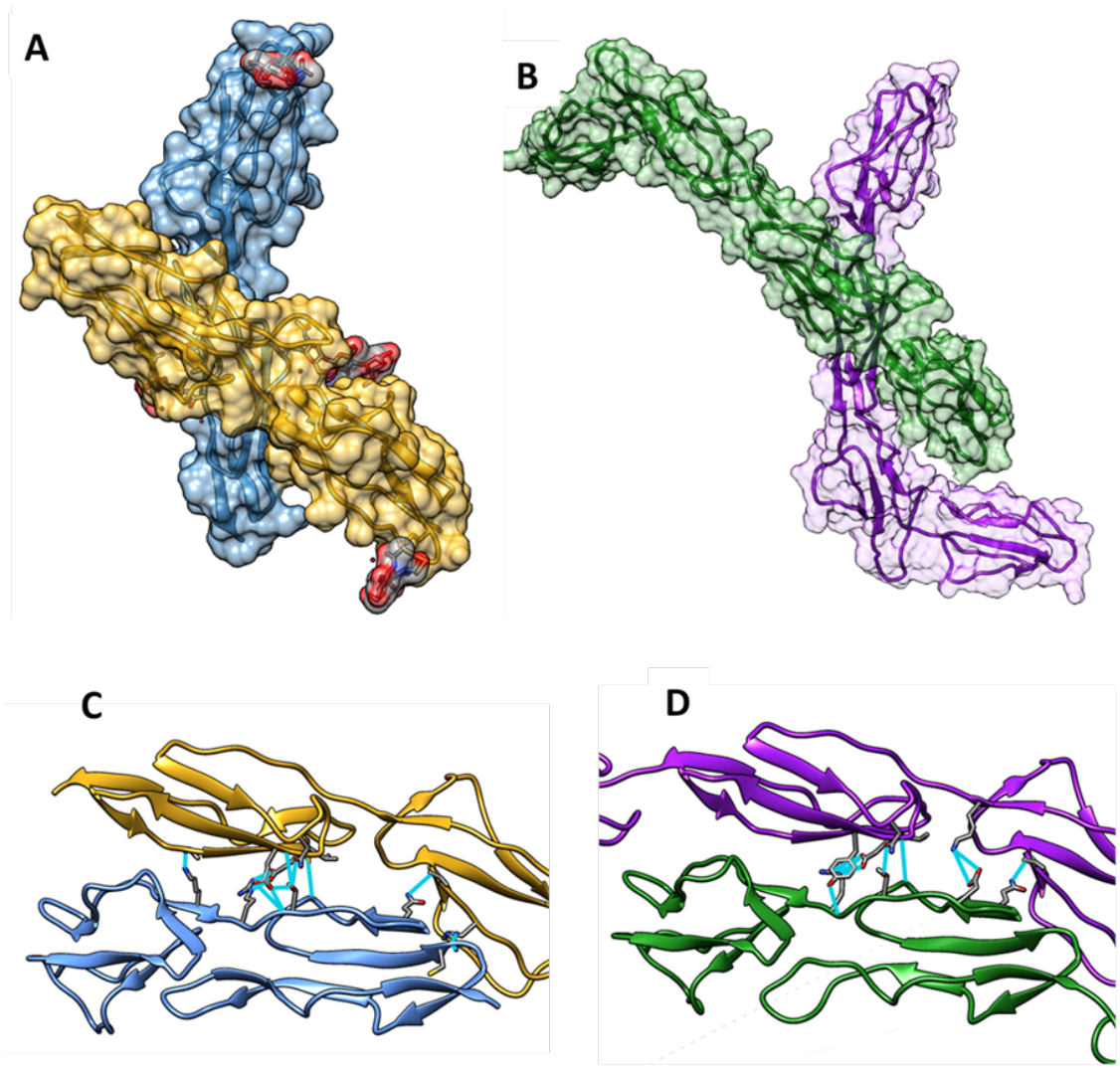


Figure 4.13: Crystallographic dimer found in bovCD46.**A**- Cartoon model of bovCD46_SCR1-2 with both monomers from the asymmetric unit shown with a surface representation in blue and yellow. **B**- Cartoon model of bovCD46_SCR1-4 with both monomers from the asymmetric unit shown with a surface representation in purple and green (the green monomer is a copy of the purple monomer superposed in the correct position as SCR4 was not resolved in this chain). **C**- Hydrogen bonding patch at the dimer interface of bovCD46_SCR1-2 with residues in hydrogen bonding distance shown in cyan sticks and the potential hydrogen bonds also in cyan. **D**- Dimer interface of CD46bov_SCR1-4 with residues in hydrogen bonding distance shown in cyan sticks and the potential hydrogen bonds also in cyan.

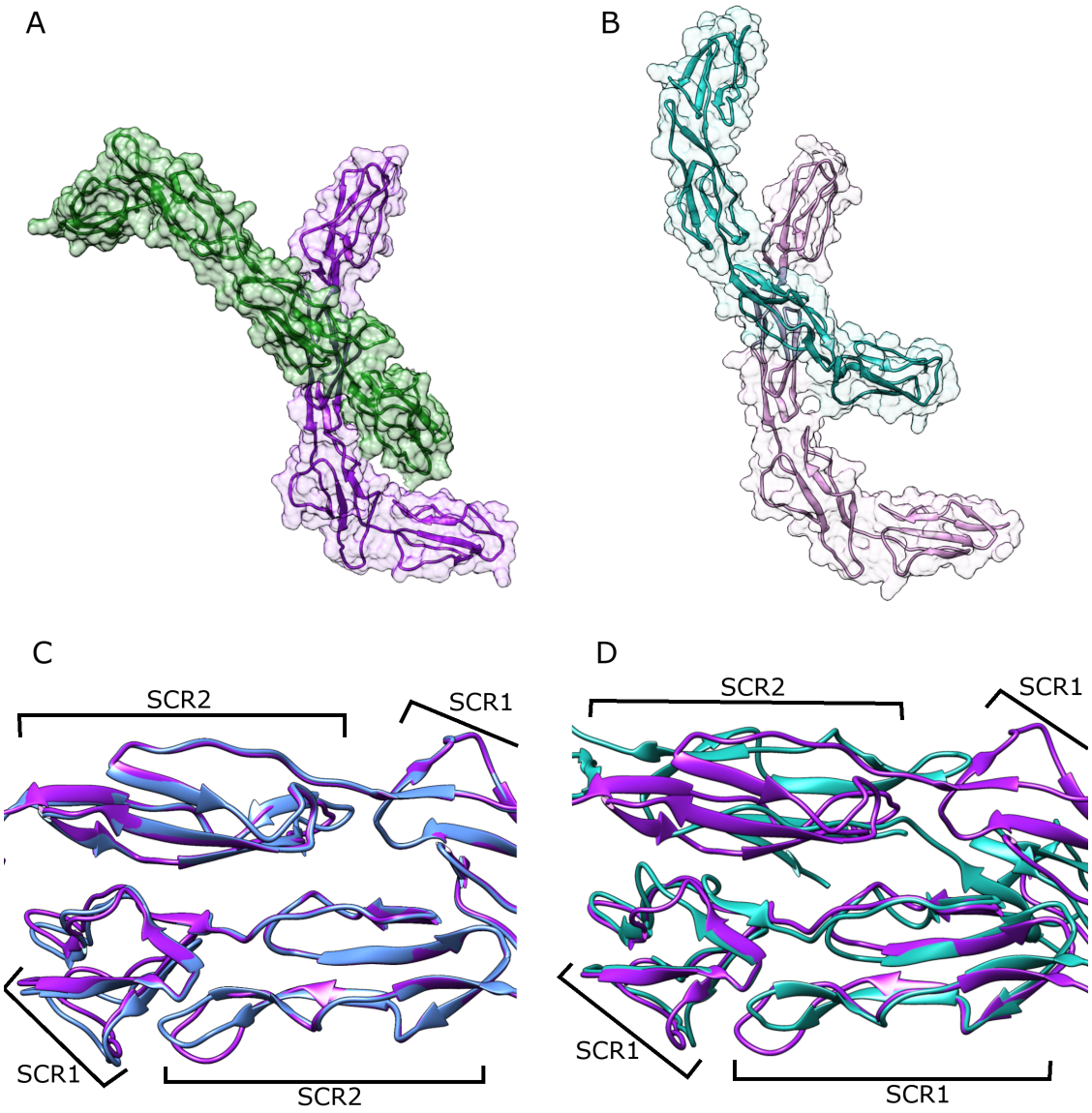


Figure 4.14: Crystallographic dimer found in human CD46.**A**- Cartoon model of bovCD46_SCR1-4 with both monomers from the asymmetric unit shown with a surface representation in purple and green (the green monomer is a copy of the purple monomer superposed in the correct position as SCR4 was not resolved in this chain). **B**- Cartoon model of human CD46_SCR1-4 with two monomers shown with a surface representation in pink and cyan. **C**- Dimer interface of bovCD46 with the bovCD_SCR1-2 and bovCD46_SCR1-4 superposed. **D**- Dimer interface of bovCD46 and human CD46 with the bovCD_SCR1-4 and humans CD46_SCR1-4 superposed.

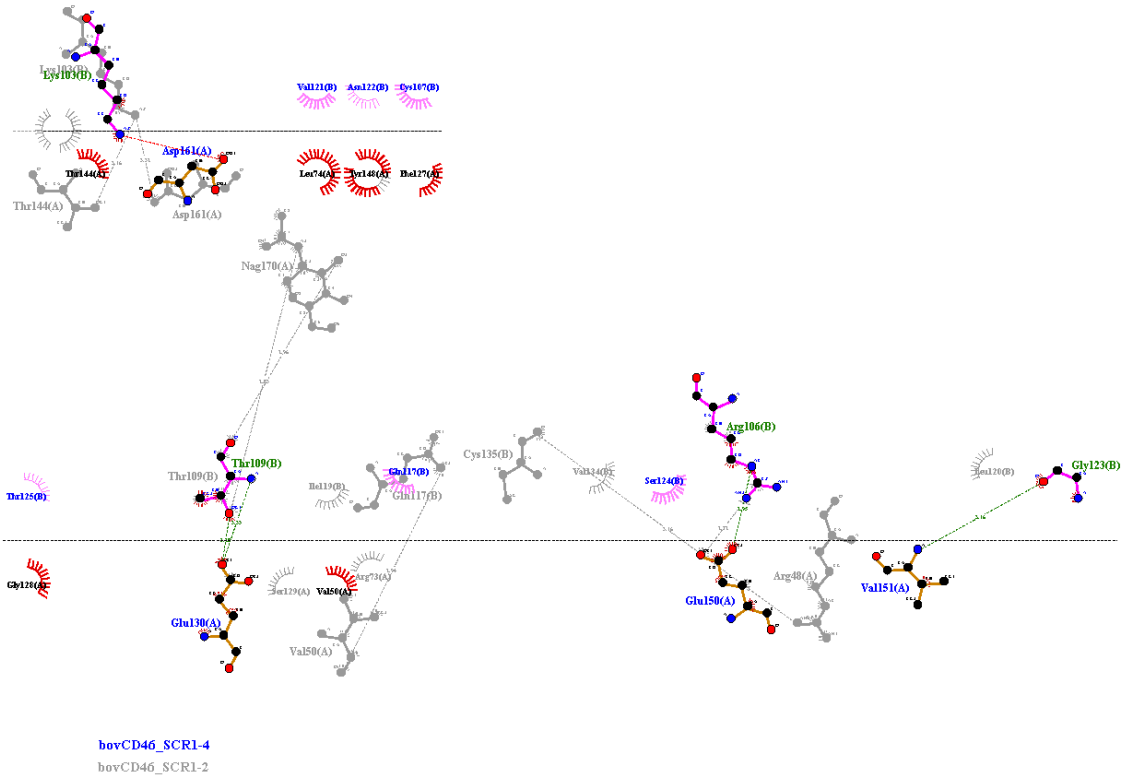


Figure 4.15: LigPlot$_+$[176] diagram of the dimerization interface in bovCD46. In colour are the interactions seen in the bovCD46_SCR1-4 structure and in grey are the additional interactions seen in the bovCD46_SCR1-2 structure. Hydrogen bonds are shown in green (grey) dotted lines and hydrophobic interactions as fringed semicircles.

Two hydrogen bonds are between SCR1 of chain A and SCR2 of chain B. The remaining seven hydrogen bonds are between SCR2 of chain A and SCR2 of chain B. The five salt bridges are located between SCR2 of both chains. Only five hydrogen bonds and the five salt bridges were predicted by the PISA server in the bovCD46_SCR1-4 structure, hydrogen bonds shown in Figure 4.13. All interactions occurring between SCR2 domains of both chains, these are however broadly a subset of those described above.

Figure 4.15 shows a LigPlot$_+$[176] analysis of the dimerization interface. Here the hydrogen bonds match those indicated by the PISA server but it shows a many hydrophobic interactions rather than salt bridges. Interestingly, a similar crystallographic intermonomer interface is seen in the human CD46_SCR1-4 Structure (PDB:3O8E) although it is not mentioned in the paper (Figure 4.14).

4.9.2 Size Exclusion Chromatography of CD46

All bovine CD46 constructs eluted from the size exclusion column at a much higher volume than expected for a monomer, and they eluted at a higher volume than the equivalent porcine CD46 construct (Table 3.2). It is likely they are all purified as dimers, but size exclusion results are not conclusive as both protein constructs are glycosylated (two sites identified in each construct) which interferes with expected elution volumes. Due to the fact that the bovCD46 recombinant proteins appear to be dimeric and that the crystallographic dimers are identical; it is worth exploring the possibility that this is a biological dimer rather than solely crystallographic dimer. There are two glycosylated asparagine's N122 and N145 located close to this interface in the bovCD46_SCR1-4 structure (only N122 in the bovCD46_SCR1-2 structure). These may have a stabilising effect on the dimer formation. The recombinant protein was produced with high mannose type glycosylation rather than the complex/hybrid glycosylation which would be found in the wild-type protein. The effect of physiological glycosylation on dimer formation would need to be tested.

As the porcine CD46 crystal structure was not solved experimentally, there is no information on the potential of a crystallographic oligomer.

4.9.3 Dynamic Light Scattering for Molecular Weight Determination

The Uncle all-in-one biological stability screening platform was used to test the molecular weight of both the porcine and bovine SCR1-4 protein constructs. The sizing and polydispersity program was used to measure the hydrodynamic size of the proteins utilizing dynamic light scattering which was used to estimate a molecular weight. The results can be found in Table 4.4 but briefly the estimated molecular weight for the bovine CD46_SCR1-4 construct (glycosylated) was 125.75 ± 14.47kDa while the estimated molecular weight for the porcine CD46_SCR1-4 construct (glycosylated) was 88.72 ± 0.00kDa.

Upon initial examination it would appear that bovine CD46 is likely to be a tetramer and porcine a dimer. The theoretical molecular weight for the bovCD46_SCR1-4 tetramer with two high mannose type glycosylation's per monomer is 138.8kDa while the theoretical molecular weight for the porCD46_SCR1-4 dimer with one high mannose type glycosylation per monomer is 66.2kDa. However, it was anticipated that this technique would result in an artificially high estimated molecular weight due to the highly elongated structure of these proteins. Therefore, CSFV E2 ectodomain was also run as a control, this protein is also highly elongated, heavily glycosylated, and known to be a dimer in solution (section 5.5.3). The CSFV E2 ectodomain EndoHf treated dimer has a calculated molecular weight of 83.88kDa and an experimentally estimated molecular weight of 252.75 ± 51.98kDa. As it is clear that this technique massively overestimated the molecular weight of elongated proteins it is more likely that bovCD46_SCR1-4 is a dimer (as seen in the crystal structure) and porCD46_SCR1-4 is a monomer.

	EndoHf treated	Est. MW (kDa)	PID	Z-Ave. Dia. (nm)
porCD46_SCR1-4	No	88.72 ± 0.00	0.21 ± 0.03	7.27 ± 0.24
porCD46_SCR1-4	Yes	72.14 ± 0.02	0.19 ± 0.04	7.28 ± 0.45
bovCD46_SCR1-4	No	125.75 ± 14.47	0.61 ± 0.04	9.02 ± 0.42
bovCD46_SCR1-4	Yes	144.38 ± 17.71	0.11 ± 0.02	9.55 ± 0.14
CSFV E2Ectodomain	No	218.23 ± 26.86	0.13 ± 0.03	11.55 ± 0.14
CSFV E2Ectodomain	Yes	252.75 ±51.98	0.15 ± 0.04	11.61 ± 0.13

Table 4.4: Uncle results shown as mean of three replicates with standard deviation. CSFV E2 ectodomain is shown as a control for an elongated dimer of approximately 80 kDa.

As can be seen in Figure 4.16 both proteins gave a single peak with the bovine protein having a larger hydrodynamic diameter than the porcine construct, implying a single multimeric state. The measured polydispersity index (PDI) for each protein was above 0.1 suggesting a non-monodispersed sample. However, they were also

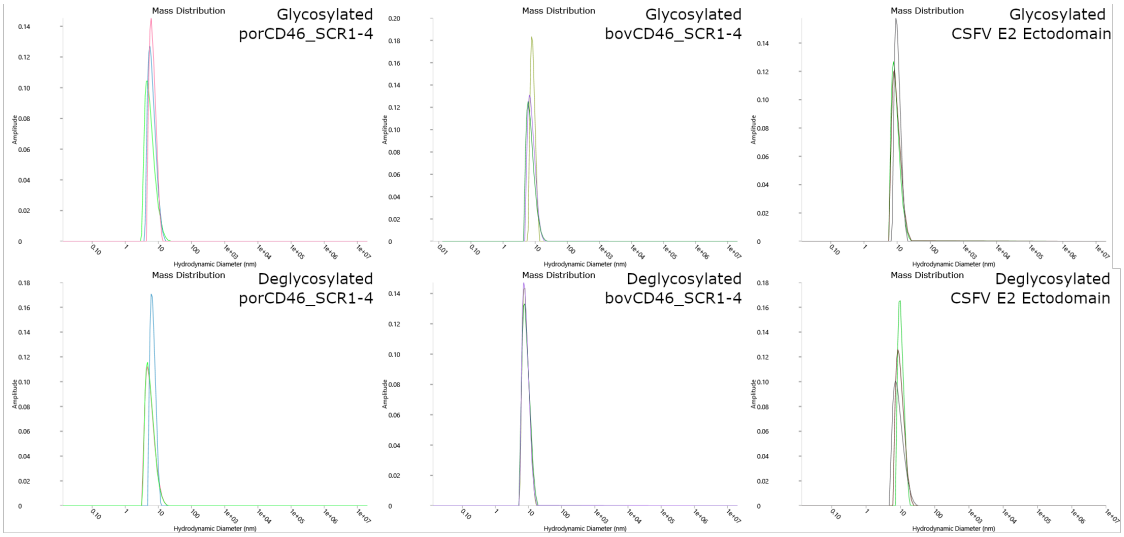


Figure 4.16: Mass Distribution of porCD46_SCR1-4, bovCD46_SCR1-4, and CSFV E2 Ectodomain constructs both glycosylated and deglycosylated, triplicates shown.

below 0.5 indicating that there is not a significant aggregate content. The proteins were measured in both a glycosylated and deglycosylated state in order to see if the presence of glycosylated had a large effect on the estimated molecular weight. As can be seen in Table 4.4 removal of glycosylation can be seen in the smaller porCD46_SCR14 but not in the larger two samples. There are two likely reasons for this: firstly, as the molecular weight estimation is based on the hydrodynamic diameter of the sample, the extended conformation of the dimer may drown out the impact of glycosylation on the hydrodynamic diameter. Secondly, the standard deviation between the estimated molecular weight measurements for replicates of the larger two proteins are large enough to cover the 1.5kDa of a high mannose glycosylation. Although the estimated molecular weights are hugely over estimated, they do show that bovCD46_SCR1-4 is some sort of multimer as it is estimated to have a molecular weight twice that of porCD46_SCR1-4 (both proteins have a theoretical molecular weight for a monomer without glycosylation of 31.5kDa). Combine this with the fact CSFV E2 ectodomain had an estimated molecular weight of over twice the theoretical molecular weight, it is likely that bovCD46_SCR1-4 is a dimer and porCD46_SCR1-4 is a monomer. Further experiments such as analytical ultracentrifugation or small angle X-ray scattering will be needed to confirm this.

4.10 Putative BVDV Binding Interaction

4.10.1 Binding Site

Bovine CD46 has been identified as a receptor for BVDV due to increased permissibility on overexpression of bovine CD46 in previously non-infectable cells[89]. A putative binding site was located in SCR1 which was shown to encompass two peptides E66-V69 and G82-L87[62]. In the protein structure, these two peptides are found on the C and D β-strands as well as the D'D loop. As expected, these two peptides form a single binding site in the structure due to the antiparallel configuration of the β-strands (Figure 4.17). The first peptide has sequence $EQIV_{69}$ in bovine CD46 and other species carry a similar sequence. Usually, this peptide starts with a negatively charged amino acid such as aspartic acid or glutamic acid. The second residue is generally a polar uncharged glutamine although in the bovine sequence it is a positively charged arginine. The third residue is normally a hydrophobic valine or isoleucine which are very similar in structure and only differ in a single methyl group. There is little conservation in the final residue in the peptide. Reasonable sequence conservation with similar residues makes it easy to hypothesise that the structure of this peptide would be similar in both the human and bovine CD46 structures, which is indeed the case (Figure 4.17). The second peptide displays some sequence conservation in the first three residues and none in the final three, it has sequence $GQVLAL_{87}$ in bovine CD46. The first residue is generally small non-polar residues such as glycine, alanine, valine, or proline. While the second residue being a much bulkier leucine or glutamine. Finally, there is again generally a small non-polar residue located in position three. The final three residues have no sequence conservation when looking at human, bovine, porcine, or sheep CD46 (Figure 4.17). The pattern of sequence conservation is reflected in the protein structure when comparing the human and bovine structures. The first three residues have a similar conformation in both structures but there is a loop shift covering the final three residues which comprise part of the D'D loop (Figure 4.17). It has been shown that changing the sequence of interaction site residues

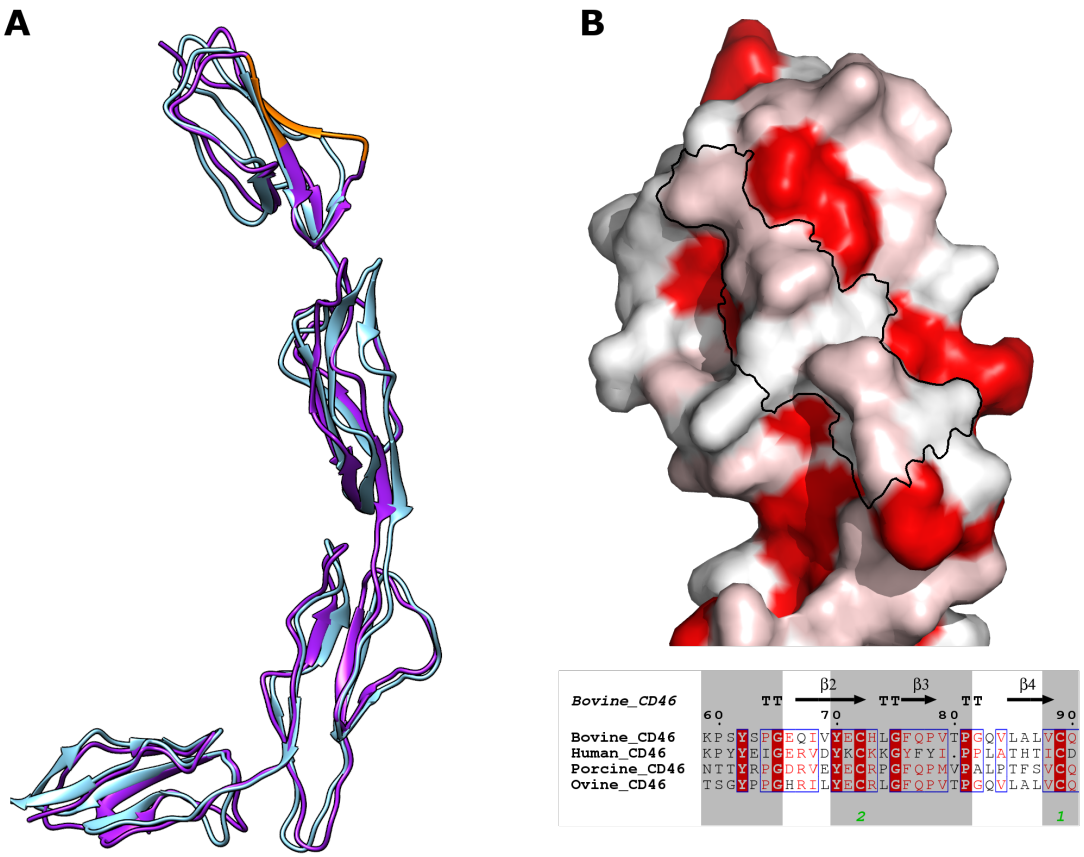


Figure 4.17: Structural comparison of bovCD46 and human CD46. **A-** Cartoon representation of bovCD46_SCR1-4 coloured purple with the BVDV interaction site highlighted in orange, a cartoon representation of the human CD46 is superposed in blue (PDB:3O8E). **B-** Surface representation of bovCD46_SCR1-4 SCR 1 with the putative BVDV interaction site outlined in black, below is excerpt from CD46 alignment focusing on BVDV interaction site peptides, both generated using ESPript web service[165]

to those of porcine CD46 caused a gain of function in porcine cells and a loss of function in bovine cells[62]. This difference in interaction site conformation is likely to confer the Pestivirus specificity for their target host CD46. Unfortunately, due to the lack of a porcine CD46 crystal structure, it is impossible to know the loop conformation in this protein as this is one of the regions with lower confidence in the computationally generated model (Figure 4.12).

It has been shown that all four bovine CD46 SCRs are necessary for BVDV entry. However, upon addition of up to six additional SCRs there is little effect on BVDV susceptibility[62]. With this in mind, it could be hypothesised that the first N-terminal SCR, SCR1, must be a certain distance (i.e. 4 SCRs) from the

plasma membrane to interact with BVDV. However, an equally valid hypothesis is that each individual SCR may play a biologically important role. As discussed previously, SCR1 is involved with BVDV interaction and is necessary for infectivity. The extended CD' loop in SCR3 causes an increased angle at the SCR3 SCR4 interface and is important in the overall protein structure. Finally, at first glance SCR2 does not seem to have a particular role. However, if the crystallographic dimer presented in this thesis is a biological dimer, then SCR2 would be necessary for dimer formation. If these individual SCR roles were all necessary for protein function, then the distance from the plasma membrane may not be a factor. Even so it is also possible that both the distance from the plasma membrane and the individual domain functions are important. A final theory as to the reason behind this result is based on the location and conformation of the protein as related to the plasma membrane. If the SCR region of the protein was lying on the plasma membrane as in the first and third view of the potential CD46 cell surface conformations (Figure 4.18), then addition of SCRs would have little effect on BVDV binding as they would not change the distance of SCR1 from the membrane.

4.10.2 Binding Studies of BVDV E2 and CD46 Ectodomains

I carried out various experiments to try and demonstrate direct binding between pestivirus E2 ectodomain and CD46 ectodomain without success. Co-crystallisation trials were carried out with BVDV E2 and bovCD46_SCR1-2 and bovCD46_SCR1-4 as well as APPV and CSFV E2 ectodomains with porCD46_SCR1-4, porCD46_SCR1-3, and porCD46_SCR1-2. No crystals of the complex were grown. Size exclusion chromatography was carried out with a 1:1 molar ration of BVDV E2 ectodomain and bovCD46_SCR1-2. The proteins did not appear to interact and eluted as two separate peaks. Finally pull-down experiments were carried out with BVDV E2 ectodomain and bovCD46_SCR1-2 and bovCD46_SCR1-4. These also failed to demonstrate any interaction between the two proteins. It is possible that the membrane anchor portions of the proteins are required to facilitate the interaction. All current data supporting the BVDV CD46 interaction has been carried out in

the context of cells so it is also possible that an additional factor is involved. The final possibility, although less likely, is that the interaction between BVDV and CD46 happens through some protein other than E2.

4.11 CD46 Orientation on Cellular Surface

The initial study which presented the structure of the human CD46_SCR1-4 by Persson *et al*[94] postulated that the bent conformation offered two main options for cell surface positioning. The extracellular portion of the protein could be positioned such that SCR4 points straight up from the STP in line with the membrane anchor (Figure 4.18). This would force the rest of the protein to lie almost parallel to the membrane. The other hypothesis was that there could be a bend in the STP region, located between the membrane anchor and the SCRs causing SCR4 to lie almost parallel to the membrane and the rest of the protein to stick straight out like an antenna[94] (Figure 4.18). As there are multiple different CD46 isoforms which differ in the STP region is has been suggested that the protein may adopt both conformations through separate isoforms with different STP regions. If the crystallographic dimer presented in this thesis is a biological dimer, then this provides an interesting constraint on cell surface positioning. In this case the only orientation possible would be one not considered before as the two options presented above would not allow formation of a dimer. Here the proteins might lie on their side close to the membrane with one crossed over the other (Figure 4.18). A final possibility is that all three orientations may be adopted by different isoforms.

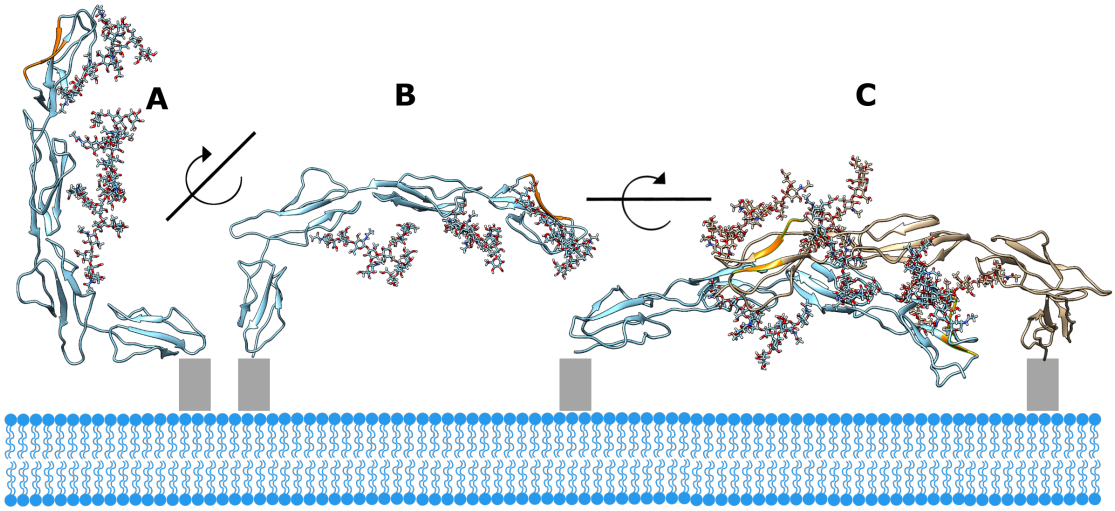


Figure 4.18: Hypothesised orientations of CD46 ectodomain on the cell surface. The lipid bilayer is shown in blue and the STP region of CD46 depicted as a grey box. The glycans are modelled as hybrid style in the locations glycosylated in the crystal structures. The bovCD46_SCR1-4 structure is depicted in cartoon style in pale blue (and pale yellow for the second monomer in panel C) with the putative BVDV binding site highlighted in orange. Rotation between different orientations shown as black arrows.

4.12 Conclusion

CD46 is a type one transmembrane protein belonging to the RCA family. In humans CD46 has been implicated in a wide range of cellular processes in addition to the regulation of complement activation. These include T-cell activation, cellular metabolism, reproduction, and autophagy. Additionally, CD46 acts as a cellular receptor for 11 human pathogens, both viral and bacterial[164] as well as the pestiviruses BVDV and APPV[88, 89]. As cellular receptors are necessary for cell entry by viruses structural information on pestivirus receptors could provide valuable insight into pestivirus membrane fusion. This thesis presents the X-ray crystallography structure of the bovine CD46 SCR1-2 and SCR1-4 extracellular regions.

The bovCD46_SCR1-4 region has been shown to adopt essentially the same conformation as the human CD46_SCR1-4 structure[94]. It is composed of four SCR domains arranged head to tail in an antenna-like structure. There is an increased angle at the membrane proximal SCR3:SCR4 domain interface resulting in a hockey

stick like appearance. The bend in the structure is a result of a hydrophobic insertion in the CD' loop of SCR3, multiple sequence alignments revealed the presence of this insertion in both ovine and porcine CD46 as well. However, computationally generated structural models of porcine CD46_SCR1-4 predict a reduced angle despite the presence of the hydrophobic insertion. This is likely due to the fact the bend in the structure is only seen in CD46 proteins and not in other CRP proteins which are composed of SCR domains, which may have biased the deep learning.

Both bovCD46 structures also revealed a loop shift in the region thought to be involved with BVDV interaction when compared to the same region in the human structure. This could possibly play a role in host cell tropism. An interesting feature of both crystal structures was the presence of a crystallographic dimer. This interaction took place between the SCR1:SCR2 interface and SCR2 of the other monomer. It is has an interface area of nearly 600Å and involves several hydrogen bonds and salt bridges. The human CD46_SCR1-4 crystallographic structure also displays a similar crystallographic dimer with the interaction occurring at the SCR1:SCR2 domain interface and SCR of the other monomer. The dimer is supported by a size exclusion chromatography elution volume and dynamic light scattering calculated molecular weights consistent with the presence of a dimer. The potential for a biological dimer poses a new potential orientation for the protein on the cell surface. In this case instead of pointing straight up or lying parallel to the membrane the protein could lie on its side on the cell surface with one monomer crossed on top of the other.

5
Pestivirus E2 Glycoproteins

Contents

5.1 Introduction 139
5.2 Crystallography of Typical Pestivirus E2 Glycoproteins 141
5.3 Structure of APPV E2 144
5.4 Structure of NRPV E2 156
5.5 SAXS of Pestivirus E2 Glycoproteins 163
5.6 Computationally Generated Models of Typical Pestivirus E2 Glycoproteins 168
5.7 Structural Comparison of Pestivirus E2 Glycoproteins 173
5.8 Conclusion 181

5.1 Introduction

In order to enter a cell to replicate, a virus must pass through the cell membrane, the cell's first line of defence. For enveloped viruses, such as pestiviruses, this usually involves the endocytosis of the virus followed by fusion of the viral envelope with the endosomal membrane, both of which are phospholipid bilayers. In the case of pestiviruses, two envelope glycoproteins are involved with membrane fusion: E1 and E2[61]. Prior to membrane fusion, the virus will bind one or more cellular receptors, triggering the internalization of the viral particle in an endosome and possibly a conformational change in viral envelope glycoproteins priming the viral

fusion protein which then catalyses the membrane fusion process[177]. Although it is known that pestiviruses undergo pH triggered membrane fusion in late endosomes[77] there is currently no information on the fusion protein itself. In pestiviruses, E1 is thought to act as the membrane fusion protein and E2 is believed to perform the receptor binding function as well as potentially acting as a chaperone or scaffold for E1[60]. Structural information on pestivirus E2 proteins could prove useful in deciphering their function and role in membrane fusion.

Together E1 and E2 have been shown to be both necessary and sufficient for membrane fusion, ruling out an essential role by the third glycoprotein Erns[61]. E2 is larger than E1, with a large extracellular domain and a C-terminal membrane anchor. E2 is found as a homodimer as well as a heterodimer with E1[46]. E2 was initially thought to be the fusion protein but after solution of BVDV E2 ectodomain structure revealed a lack of a fusion motif it is now thought to be involved with receptor interaction[21, 22]. A recent paper presented the first reports of a direct interaction between E2 and a cellular receptor with information on the interaction between CSFV E2 and ADAM17 metalloproteinase domain[91].

The crystal structure of BVDV E2 has been solved at both physiological and the lower late endosomal pH (pH5) revealing disorder of the N-terminal domain at low pH[21]. On disorder of the N-terminal domain a β-hairpin is exposed to the solvent, this hairpin has been implicated in receptor binding[178]). The equivalent region in CSFV has also been identified as potentially having a role in receptor interaction[63]. Most pestiviruses have high sequence conservation in the E2 protein sequence although *rodent pestivirus* and NRPV have a well conserved C-terminus but much reduced sequence conservation towards the N-terminus (Figure 1.9). APPV and *bat pestivirus* differ even more as they have a truncated E2 protein missing a large N-terminal section alongside reduced sequence conservation in other regions. It is hoped that solution of the structure of these atypical E2 structures could provide insight into the cell entry mechanism. This is because the reduced sequence

conservation covers the region reported to be involved with receptor interaction. As such the E2 proteins of NRPV and APPV are a focus of this study.

As some pestivirus E2 proteins have a highly divergent sequence (Figure 1.9), it is unlikely the usual molecular replacement technique will be useful in solving the crystal structures. Therefore an alternative technique such as single wavelength anomalous diffraction (SAD) experimental phasing may have to be used. Several years ago, the E2 protein from BVDV-1 was solved by two separate groups[21, 22], but no more pestivirus E2 structures have been published since. It is possible that this is because the proteins do not crystalise readily so other techniques such as SAXS, cryo-EM, or computational modelling may have to be used instead.

This chapter details the crystallography of the E2 ectodomain from four pestiviruses: BDV, CSFV, APPV, and NRPV. X-ray crystallography structures are shown for both APPV and NRPV while computationally generated models are presented for BDV and CSFV. SAXS experiments on BVDV, BDV, CSFV, APPV, and NRPV shed insight into pH triggered conformational changes.

5.2 Crystallography of Typical Pestivirus E2 Glycoproteins

5.2.1 Crystallography and Optimisation of CSFV E2

The ectodomain of CSFV E2 was put into crystallization trials. This construct encompassed the entire ectodomain region covering residues R1 to E341 (Figure 3.3, CSFV_9). After large scale expression and purification 30mg of pure, stable, soluble protein was obtained from 6 flasks (1.5 litres) as described Chapter 3. The protein was put into crystallisation screening experiments at 27mg/ml in a fully glycosylated state (5 predicted sites according to NetNGlyc[170] so an additional 7.5kDa per monomer). Sitting drop vapour diffusion trials of total of approximately 480 different conditions were tested, further details can be found in section 2.7.1.

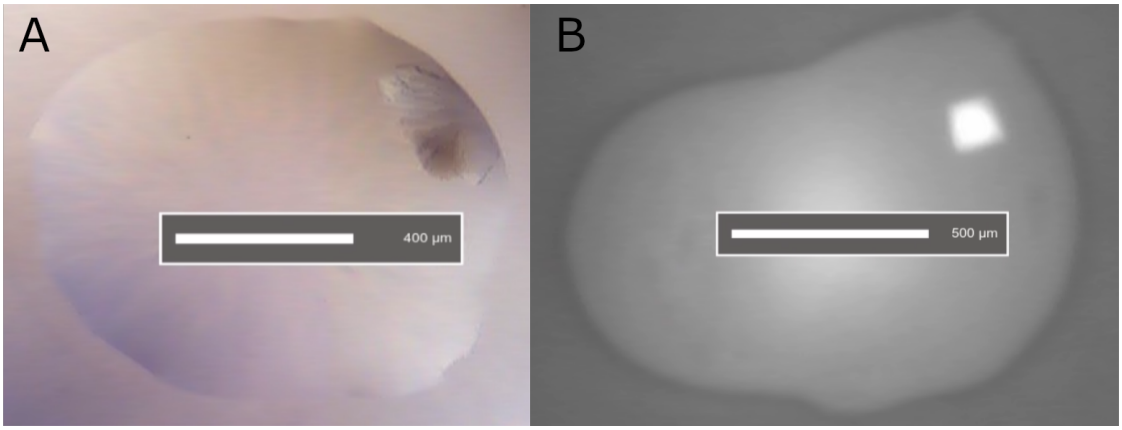


Figure 5.1: Crystals of CSFV E2 ectodomain, initial crystals shown in A and optimized crystals in B. B was imaged under UV light.

The initial trials in crystal quick X plates at 20°C gave no crystals, so a second set of screens was set up in swiss ci 3 plates. An initial crystal appeared after 2 days in 0.1M Bis Tris pH 5.5, 0.3M magnesium formate with a drop ration of 2:1 and was fully grown after 21 days. This crystal was a non-singular plate extending to over 300µm in its largest dimension (Figure 5.1). Due to a scheduled synchrotron shutdown a grid optimisation follow-up was carried out without testing diffraction. This produced many crystals in both the 1:1 and 2:1 drop ratios of which 5 were harvested and cryoprotected with the addition of 20% glycerol. Diffraction was tested but all were badly cryoprotected, and no diffraction was seen. Subsequently, a further 16 crystals were harvested and cryoprotected with the addition of between 30%-50% ethylene glycol or glycerol. Again, all had poor cryoprotection and displayed little diffraction. The following optimisation techniques were performed: micro seeding, additive screens GRAS and silver[179], 4°C incubation, and addition of glycerol to the condition. The GRAS additive screen produced some crystals, the best looking was in 360 mM magnesium formate dihydrate, 0.1M Bis-Tris pH5.5, and 50mM Caffeine in the 1:1 drop ratio. These were nice singular cubes of just over 100µm that could only be imaged under UV light (Figure 5.1) but again these failed to diffract to sufficient resolution.

Two further constructs were designed, produced, and tested: one with the addition of a 3C site at the C-terminus and one with a N292Q mutation. The

N292Q construct was designed to remove the equivalent glycosylation site to the one removed in the BVDV E2 structure 2YQ2[21]. These two new constructs were examined in in presence and absence of Endo Hf to test both glycosylated and deglycosylated protein (details on deglycosylation can be found in the materials and methods chapter) but did not produce any crystals. I also attempted to grow the crystals in crystal quick X plates to enable the use of the VMXi beamline and collect data directly in the plate, removing the need for cryoprotection. However, despite extensive efforts with all three constructs, no crystals were grown in these plates.

5.2.2 Crystallography of BDV E2

The BDV E2 ectodomain construct used for crystallisation trials encompassed the entire ectodomain, residues C4 to E341 (Figure 3.3, BDV_10). After large scale expression and purification only 0.5mg of pure, stable, soluble protein was obtained from 2 flasks (0.5 litres) as detailed in the chapter 3. Initial crystallisation trials were carried out at 20°C with 10.5mg/ml protein in a sitting drop vapour diffusion experiment. Approximately 480 different conditions were tested with fully glycosylated protein. Further crystallisation details can be found in section 2.7.1.

Initial crystals appeared in 0.2M magnesium chloride, 0.1M Hepes pH7.5, 30% PEG 400 and were only visible under UV light, these were thin needles extending to around 200µm (Figure 5.2). Crystals were harvested and cryoprotected with the addition of 25% glycerol, unfortunately diffraction testing revealed only weak diffraction. A crystal from a grid screen follow up on this condition diffracted to around 8Å resolution although the diffraction was very anisotropic extending to 6.5Å along the a and b axes, and 10.2Å along the c axis and was successfully process by autoPROC[108]/Staraniso[114] in the $I422$ space group. These crystals were gown in 24% PEG 400, 0.1M Hepes pH7.6, 0.1M $MgCl_2$ in the 1:1 protein to well solution drop ratio, they were short chunky hexagonal rods of around 100µm by 100µm (Figure 5.2). Unfortunately, the poor yield of protein and time prevented further optimisation.

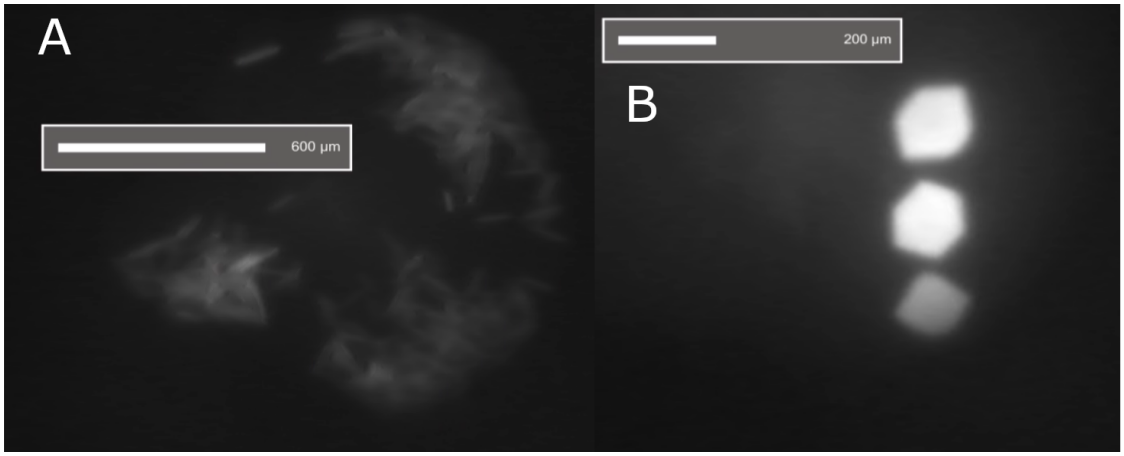


Figure 5.2: Crystals of BDV E2 ectodomain, initial crystals shown in A and optimized crystals in B; both imaged under UV light.

5.3 Structure of APPV E2

5.3.1 The Phase Problem and Sulphur SAD Phasing

One of the main challenges to solve in crystallography is the phase problem. An X-ray diffraction data set records the intensity of each reflection from which structure factor amplitudes can be calculated. However, in order to solve the crystal structure, both the amplitude and phase of the structure factors of each reflection are required. Unfortunately, the phase cannot be directly recorded. This is the phase problem. There are several methods used to recover the phases, which are then combined with the amplitudes to solve the 3-dimensional X-ray diffraction pattern via the inverse Fourier transform of the structure factors.

The most straightforward and popular method of calculating the phases is molecular replacement[180]. Here, a previously solved structure of a protein with high sequence similarity is used as a model. This structure is then manipulated within the repeating unit of the crystal so that the theoretical diffraction pattern it would produce is as close to the experimental data as possible. Practically, this is done by calculating Paterson maps which are Fourier transforms of the intensities of the structure factors which are directly derived from the diffraction pattern. Once searches have been carried out the phases from the model are combined

with the experimental amplitudes and the inverse Fourier transform is used to calculate an approximate electron density.

Another method of calculating phases is direct methods. Here the facts that the crystal is composed of similarly shaped atoms and that there are statistical relationships between sets of structure factors are combined to hypothesise possible phase values. This only works for extremely simple structures with a limited number of atoms, or for very high resolution data set (close to 1Å resolution)[181]. Therefore, this method is very rarely used to solve the phase problem in protein crystallography.

If these methods cannot be deployed, then experimental phasing options must be used. Generally, this involves finding the position of several 'heavy' atoms within the structure and using them to infer phases for the entire structure. The two ways of doing this are multiple isomorphous replacement (MIR)[182] and using anomalous dispersion either multi-wavelength (MAD) or single wavelength (SAD)[183]. With MIR heavy atoms are soaked into the crystal and contribute greatly to the overall scattering intensity as the scattering of an atom is proportional the number of electrons it contains. Since randomly dispersed lighter atoms scatter according to the average number of electrons in each atom and the square root of the number of atoms, a single gold atom of will scatter more strongly than 100 atoms of carbon, nitrogen and oxygen[184]. By comparing the scattering from a native and a derivative crystal with heavy atoms, it can be calculated what scattering is from the heavy atoms and this can be used to figure out the location of the heavy atoms in the structure (using the direct methods mentioned above, since this is now a far simpler structure than the protein), this is called substructure determination. Once the locations of the heavy atoms have been deduced, this substructure can be used to phase the rest of the structure. Anomalous scattering methods rely on certain atoms which have an absorption edge near energies available at synchrotron light sources[184]. Collecting near the absorption edge changes the scattering for these atoms and the resulting structure factors. Then just like MIR, it is possible to find the substructure for those few atoms with anomalous signal and locate them

in the repeating unit. These atoms are then used to calculate the phases for the rest of the protein. Traditionally, selenomethionine is incorporated into the protein in the place of methionine to make this technique possible[185].

Sulphur SAD phasing is a way of solving the phase problem of protein crystallography using the single-wavelength anomalous dispersion of sulphur atoms when a model is not available for molecular replacement[186]. Sulphur atoms are found natively in proteins in both cysteine and methionine amino acids, some proteins also bind sulphur atoms such a sulphate salts present in some crystallisation conditions. The advantage of this method is that it is not necessary to incorporate heavy atoms or other elements into the protein which often have a negative effect on diffraction quality. Sulphur SAD is particularly suitable for the proteins studied in this project as they contain many sulphur atoms in the cysteine residues usually involved in the disulphide bonds. Indeed, disulphide bonds are often found in extracellular proteins, they are particularly helpful in Sulphur SAD because, since they are close together they scatter together to increase the sulphur anomalous signal (especially at lower resolution). Using the native sulphur atoms in the protein overcomes the lower expression, reduced solubility, lack of methionines, and reduced tendency to crystallisation which have been seen to pose a problem with some selenomethionine experiments, especially in the case of eukaryotic expression. The I23 beamline at Diamond Light Source is designed specifically for use with longer wavelengths of light to maximise anomalous scattering[187]. It is optimised for use at a wavelength between 1.5Å and 4Å, so ideally suited for sulphur SAD utilising a wavelength of 2.755Å, the optimal wavelength to maintain a high signal to noise ratio but also obtain enough anomalous signal. This beamline has an in-vacuum end station which minimises absorption and scattering effects of air, increasing the signal to noise ratio. Braggs law states that in order to get constructive interference and therefore a reflection, the wavelength must be an integer multiple of the sine of the diffraction angle ($n\lambda=2d\sin\theta$). Therefore, longer wavelengths of light will produce larger diffractions angles. In order to capture these more widely scattered reflections, the I23 beamline has a large semi cylindrical detector (Pilatus 12M, Dectris). Due

to the in-vacuum end station the crystals are kept below 100°C via heat conduction through the goniometer and specially designed brass pins.

5.3.2 Crystallography and Optimisation

The APPV protein construct which crystalised, encompassed the ectodomain region covering residues S1 to H211 (Figure 3.3, APpv_9). After large scale expression and purification, 6.4mg of pure, stable, soluble protein was obtained from 6 flasks (1.5 litres) as described in chapter 3. Initial crystallisation trials utilised protein at 26mg/ml and were carried out at 20°C with a total of approximately 480 different conditions tested. Further details can be found in section 2.7.1. The protein expression, purification, and crystallisation were carried out in the absence of deglycosylation enzymes to maintain the glycosylation.

Initial crystals appeared after 1 days in multiple conditions. The first long wavelength data sets were collected from a crystal grown in 0.1M Bis Tris pH 5.5, 0.2M ammonium sulphate, 25% PEG 3350. The crystal was a triangular rod of approximately 75µm by 150µm, which appeared after 1 day in the 1:1 ratio drop and was fully grown after 7 days (Figure 5.3). As this crystal (cryo-protected as described below) diffracted past the detector limits on I23, another dataset was collected on a higher energy beamline (I03). This crystal was grown in 0.1M Bis Tris pH 7.8, 0.05M sodium citrate, 0.05M magnesium chloride, 25% PEG Smear High, using a 2:1 ratio drop. It appeared after 13 days and was full grown by the next scheduled image at 21 days; again, the crystal was a triangular rod of approximately 75µm by 150µm (Figure 5.3).

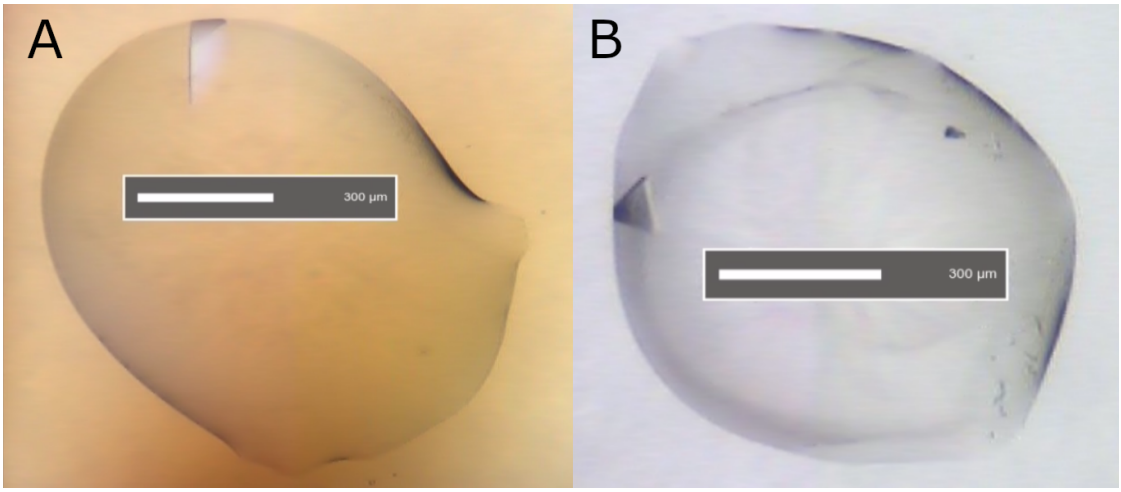


Figure 5.3: Crystals of APPV E2 ectodomain grown in pH5.5 (A) and pH7.8 (B).

5.3.3 Crystal Harvesting and Data Collection

Crystals were cryoprotected by flooding the drop with a mixture of well solution with cryoprotectant added (20% glycerol for the initial dataset and 20% ethylene glycol for the 2nd data set). Crystals were then harvested from the flooded drop with a litholoop and flash-cooled in liquid nitrogen. During data collection the crystal was kept at low temperature either by heat conduction through the goniometer to around 50°K at I23, or in a 100°K gaseous nitrogen stream at I03 to prevent devitrification.

Initially three long wavelength datasets were collected at the I23 beamline at diamond light source. A wavelength of 2.76Å with an attenuator reducing the beam to 26% of its intensity and 8.05x10^{10} flux was used for 0.1 second exposures with an omega rotation of 0.1°. Each data set contained 3600 images and was collected with different κ angles of 0°, -20°, and -10°. The advantage of offsetting the κ angle is to minimise systematic errors and obtain true multiplicity. The diffraction limits from this dataset was 1.8 Å resolution but with a mean I/sigma(I) of 2.5 in the highest resolution shell suggesting the diffraction extended past the I23 detector limits. Therefore 16 more crystals were tested on the faster beamline I03. One of these diffracted to 1.2Å and a dataset was collected. For this dataset a wavelength of 0.98Å with 100% transmission and 3.65x10^{12} flux was used for 0.005 second exposures with a 0.1° omega oscillation for a total of 1800 images.

Initial data processing for each of the lower resolution data sets was carried automatically with XIA2[120]_3dii-XDS[115] implemented as part of the Diamond processing pipeline, then Xscale[115] was used to merge the datasets. The space group was $R3$, and it extended to a resolution of 1.80Å with completeness 99.27% and I/sig(I) of 26.23. The high resolution dataset was initially processed with XIA2[120]_3dii-XDS[115] implemented as part of the Diamond processing pipeline in the R3 space group, with a resolution of 1.20Å, completeness of 91.26% and I/sig(I) of 18.37. Full statistics for both datasets can be found in Table 5.1.

	APPV-E2 Hires	APPV-E2 SSAD
Wavelength	0.98	2.76
Resolution range	26.26 - 1.20 (1.24 - 1.20)	38.68 - 1.80 (1.87 - 1.80)
Space group	R 3	R 3
Unit cell	101.12 101.12 43.12 90.00 90.00 120.00	101.33 101.33 43.09 90.00 90.00 120.00
Total reflections	219204 (6889)	257597 (9731)
Unique reflections	46937 (2865)	15203 (1454)
Multiplicity	4.70 (2.40)	16.90 (6.70)
Completeness (%)	91.26 (56.09)	99.27 (92.92)
Mean I/sigma(I)	18.37 (1.17)	26.38 (2.49)
Wilson B-factor	18.14	29.74
R-merge	0.032 (0.60)	0.06 (0.50)
R-meas	0.04 (0.75)	0.06 (0.54)
R-pim	0.02 (0.45)	0.01 (0.20)
CC1/2	1.00 (0.60)	1.00 (0.86)
CC*	1.00 (0.87)	1.00 (0.96)
Reflections used in refinement	46928 (2865)	15203 (1456)
Reflections used for R-free	2287 (131)	787 (86)
R-work	0.15 (0.24)	0.18 (0.28)
R-free	0.18 (0.27)	0.23 (0.30)
CC(work)	0.97 (0.82)	0.97 (0.82)
CC(free)	0.96 (0.67)	0.95 (0.77)
Number of non-hydrogen atoms	1527	1379
macromolecules	1311	1206
ligands	56	42
solvent	160	131
Protein residues	149	148
RMS(bonds)	0.01	0.01
RMS(angles)	1.14	0.82
Ramachandran favored (%)	96.55	97.92
Ramachandran allowed (%)	3.45	2.08

Ramachandran outliers (%)	0.00	0.00
Rotamer outliers (%)	1.36	0.00
Clashscore	4.80	2.05
Average B-factor	30.74	34.16
macromolecules	27.83	32.72
ligands	62.52	56.10
solvent	43.49	40.33

Table 5.1: Data Collection and Refinement statistics for APPV E2 Ectodomain

5.3.4 Structure Solution and Refinement

Initially the APPV ectodomain structure was solved using sulphur SAD phasing. HKL2map[121] was used to find the sulphur locations and generate an initial electron density map. Autosol[111] from the phenix[166] suite of programs was then used to improve the electron density maps with density modification, and build an initial model. Multiple rounds of phenix refine[122] and manual building in coot[112] were used to complete and refine the structure giving an Rwork and Rfree of 18.12 and 22.72% respectively at 1.80Å resolution with good stereochemistry. The higher resolution dataset was solved using molecular replacement with phaser[110] from the CCP4[168] suite of programs. A combination of refmac[113] and manual coot[112] refinement was used to finish the model giving an Rwork and Rfree of 15.07/18.00% at 1.20Å resolution respectively with good stereochemistry. Refinement statistics for both structures can be found in Table 5.1 and example of electron density in Figure 5.4. The final structure encompassed the N-terminus of protein structure with residues S1 to T153 being visible. An additional glycine was also resolved at the N-terminus which originated from the restriction digest site located immediately N-terminal to the protein encoding sequence.

5.3.5 Structure Analysis

The protein was initially solved vis sulphur SAD to 1.80Å using a crystal from a pH5.5 condition. A higher resolution structure (1.20Å) was then solved from a crystal grown at pH7.8 by molecular replacement. As the structures were extremely similar

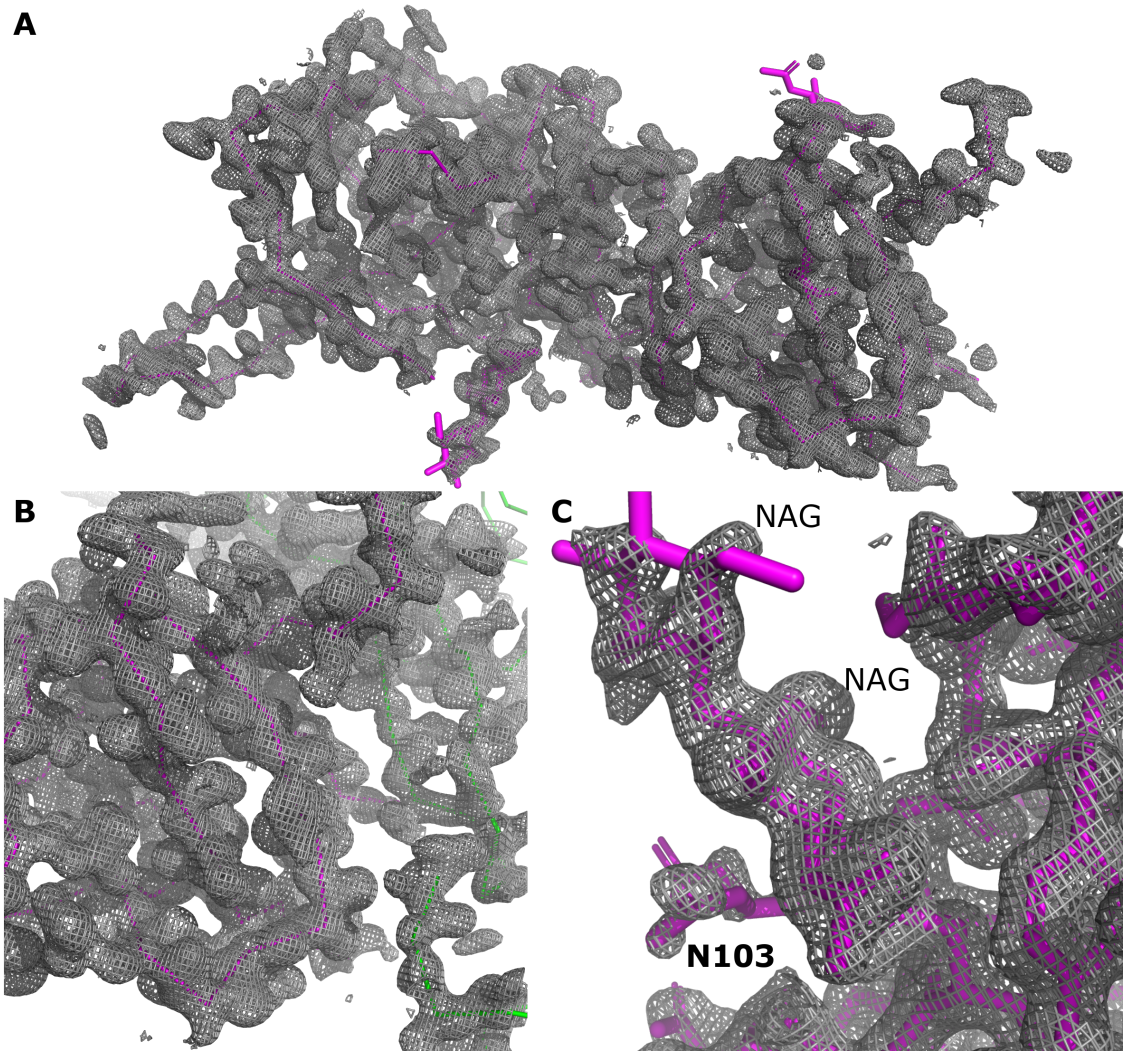


Figure 5.4: Electron Density Maps (2mFo-DFc generated using phenix map creation tool[167]) for the higher resolution APPV E2 structure with sigma level 1. **A**- Entire structure, monomer in magenta carbon alpha trace. **B**- APPV E2 ectodomain C- terminal region coloured as before with symmetry copies in green carbon alpha trace. C- APPV E2 glycosylation at residue N103 in sticks and surrounding area in carbon alpha trace.

(RMSD: 0.15Å for C_α atoms), only the high-resolution structure is discussed further. The structural model for the APPV E2 ectodomain contained density for residues S1 to T153. As the crystallized protein construct covered the region encompassing residues S1 to H211, there are 58 residues not resolved at the C-terminus. The full-length protein has a predicted solvent content for a single monomer in the asymmetric unit of only 18%. As this is very unlikely, there is a strong possibility that the protein is truncated, especially considering the fact that glycosylation sites are present. The presence of glycosylation often results in higher than expected

solvent content as the flexible sugars reduce the available packing interfaces. Indeed, on examination of the crystal packing, there is not enough space available at the C-terminus for the missing residues to fit in the asymmetric unit (Figure 5.4). Hence, it is unlikely that the residues are present but disordered, and more likely that the protein was truncated prior to crystallisation. As the protein was purified with a non-cleavable C-terminal tag and entered crystallography immediately following purification the C-terminal region was present on entering crystallisation trials. Intact mass spectroscopy confirmed that the protein used for crystallisation was not truncated even after a freeze thaw cycle (data not shown). It is plausible that a very small amount of cellular protease was copurified with APPV and was responsible of the C-terminus cleavage during the crystallisation experiment.

The N-terminal region of APPV E2 (truncated compared to BVDV E2) forms a single domain located distal to the membrane, as seen for BVDV E2. There is a small peptide at the C-terminus which is likely the start of a C-terminal membrane proximal dimerisation domain, as seen in the BVDV E2 structure. The N-terminal domain which makes up the bulk of the extracellular portion of the protein is composed entirely of β-strands and loops and adopts a completely novel fold according to the DALI server[188] (Figure 5.5). The core of the protein is composed of three, three-stranded antiparallel β-sheets arranged linearly one on top of the other surrounded by loop regions. At the N-terminal membrane distal region of the protein is a solvent exposed β-hairpin which will be discussed further in section 5.7.3. Figure 5.5 presents a superposition (RMSD of 1.100Å over 39 residues, 17.400Å over all 307 residues) of APPV E2 with BVDV E2. Note that only the small C-terminal region with high sequence similarity superposes well, but the molecular envelope resembles a truncated version of BVDV E2. Due to the high sequence similarity in the C-terminus, it is likely that the missing APPV E2 C-terminal domain is similar in structure to that of BVDV. It may also dimerise in a similar fashion although the presence of a C-terminal domain swap cannot be speculated upon.

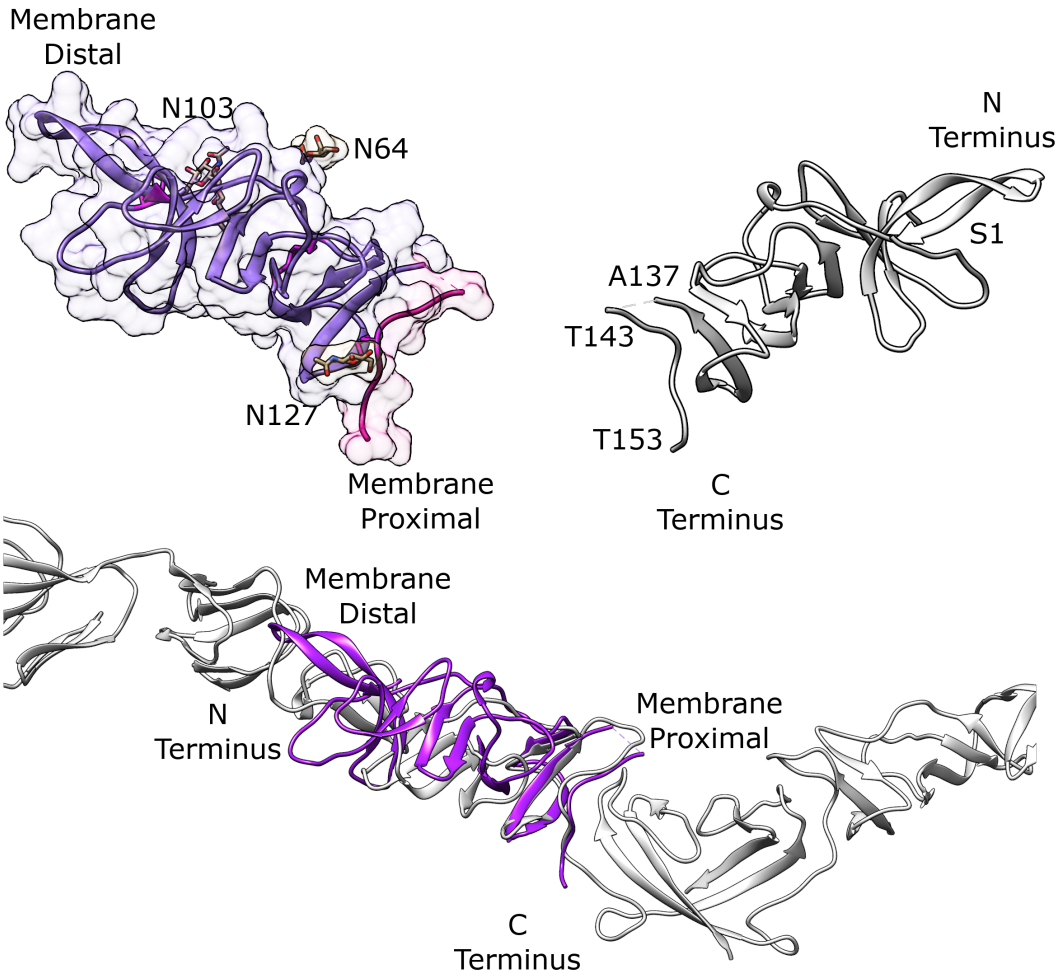


Figure 5.5: Structure of APPV E2 ectodomain. **Top-** Crystal structure of N-terminal domain of the APPV E2 ectodomain there is a cartoon representation on the left with a transparent surface shown. Glycosylated asparagines are marked and drawn as sticks and cysteines are shown in magenta. The N-terminal domain is shown in purple and the small peptide of the C-terminal domain is shown in pink. On the right, in grey is the position of the other half of the dimer if it were positioned as in the BVDV E2 dimer (PDB: 2YQ2) with the domain residue boundaries marked. **Bottom-** A cartoon representation of the APPV E2 ectodomain in purple superposed on the left monomer of the BVDV E2 dimer in grey (PDB:2YQ2).

APPV E2 residues S1 to H211 has two strongly predicted N-linked glycosylation sites at N51 and N64 and three less likely sites at N103, N127, and N141 (according to the NetNGlyc server[170]). In the crystal structure, three glycosylation sites can be visualised. A single NAG residue can be seen at N127 and N64 while density for two NAG residues is present at site N103 (Figure 5.4). The N103 and N64 sites are located on one face of the protein while the other site at N127 is located on the opposite face. None of the glycosylation sites are located such that they might prevent dimerization through the C-terminal domain. Although the density in the area surrounding N51 is not ideal, there is no density to support glycosylation. Additionally, this residue is located at the tip of the N-terminal β-hairpin which is hypothesised to be involved with receptor binding, which glycosylation may interfere with. The final predicted glycosylation site at N141 is located in a loop which does not have well-resolved density, if this location was glycosylated, it would be located on the same face at N103 and N64.

In the S1 to H153 region of the protein which crystalized, there are four cysteines which form two disulphide bonds. The first is located at the N-terminus of the protein and involved the first two cysteines and the second is located at the C-terminus of the crystallised region and encompasses the final two cysteines. Cysteine 148, involved with the disulphide bond at C-terminus, can be seen in two positions both forming and breaking a disulphide bond, this is most likely due to radiation damage. A topology diagram of APPV can be found in Figure 5.6.

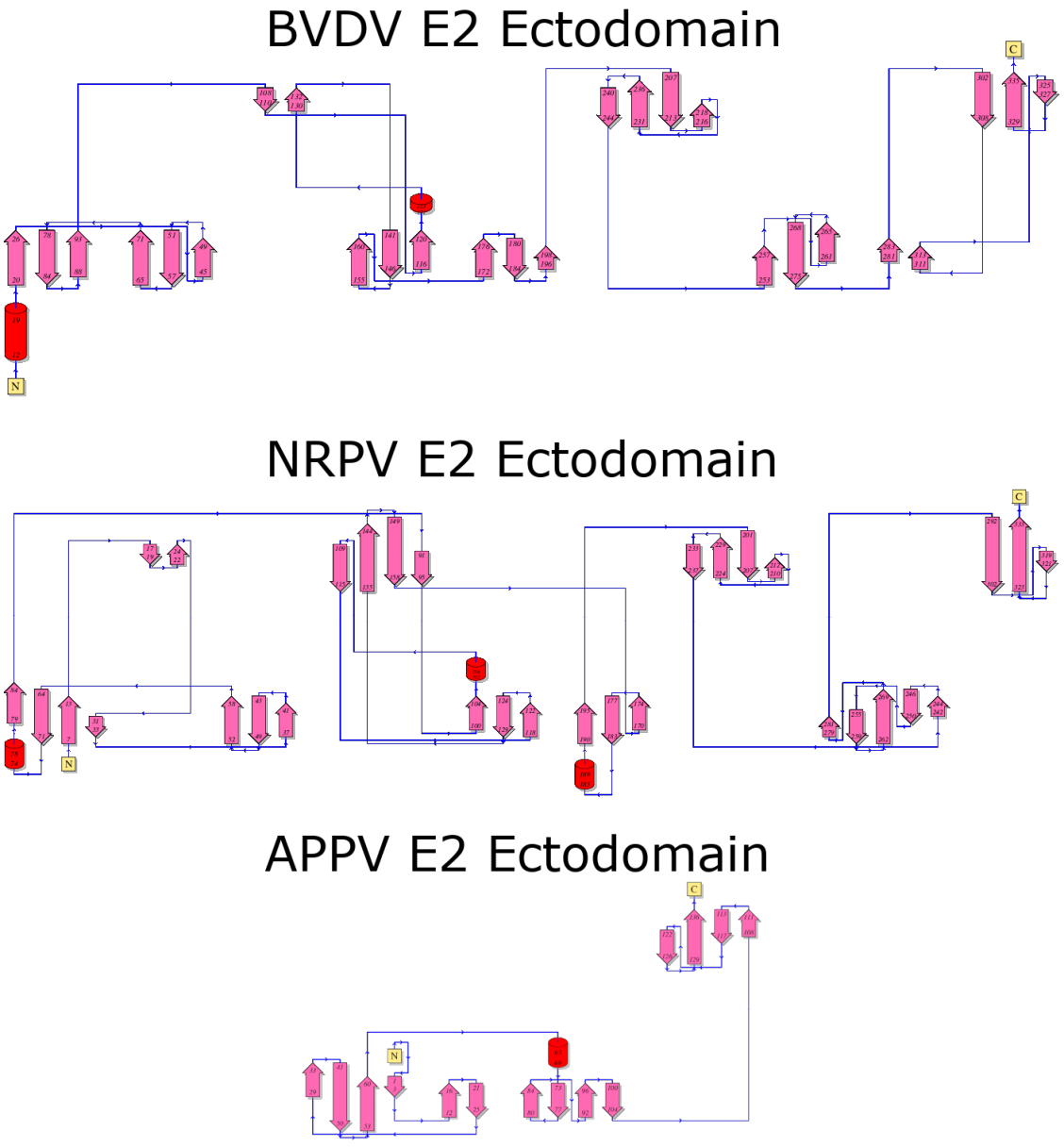


Figure 5.6: Topology diagrams of APPV, NRPV, and BVDV (PDB 2YQ2) E2 ectodomains. Diagrams generated using PDBsum[189]

5.4 Structure of NRPV E2

5.4.1 Crystallography and Optimisation

The NRPV E2 construct encompassed the entire ectodomain region covering residues G1 to K348 (Figure 3.3, NRPV_9). After large scale expression and purification, 12.9mg of pure, stable, soluble protein was obtained as detailed in chapter 3. The protein was put into crystallisation trials at 5mg/ml and 20°C in a sitting drop vapour diffusion experiment with approximately 480 different conditions tested. Crystallisation trials were carried out both a glycosylated and deglycosylated protein. Further details on the crystallisation trials and deglycosylation process can be found in section 2.7.1.

Initial crystals appeared after five days multiple different conditions. The only crystal which diffracted to useful resolution was grown in 0.2M sodium formate, 20% PEG 3350 in the 1:1 ratio drop with fully glycosylated protein. It had was an almost tooth shape of 400µ by 100µm, it initially appeared after 12 days and was fully grown by 21 days (Figure 5.7). Grid Screen follow ups were carried out for nine different conditions, but all failed to produce better diffracting crystals.

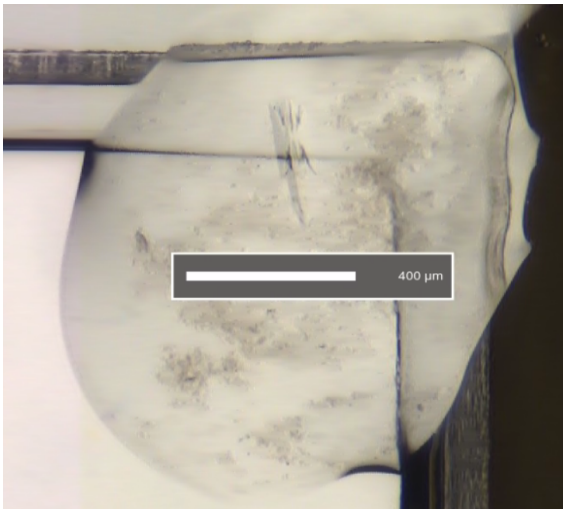


Figure 5.7: Crystals of NRPV E2 Ectodomain.

5.4.2 Crystal Harvesting and Data Collection

Crystals were cryoprotected by flooding the drop with a mixture of well solution with 25% glycerol added for cryoprotection. Crystals were then harvested from the drop with a litholoop and flash frozen in liquid nitrogen. Crystals were immediately moved to the beamline vessel at I23 for data collection and cooled by heat conduction via the goniometer to around 50°K to prevent devitrification. Three datasets were collected at a wavelength of 2.76Å with 26% transmission and 1.21x10^{11} flux. 0.1 second exposures were taken with an omega rotation of 0.1°. Each data set contained 3600 images collected at different kappa angles of 0, 10 and 20°. Data collection statistics can be found in Table 5.2.

Initial data processing of all three datasets was carried out with XIA2-XDS[115] implemented as part of the Diamond processing pipeline[120] then Xscale[115] was used to merge the datasets. This processing assigned the datasets to the $P4_122$ space group with a resolution of 3.50Å, completeness of 89.75% and I/sig(I) of 23.79. Subsequent to structure solution the data were reprocessed into the enantiomorphic space group $P4_322$ space group using XDS[115].

Wavelength	2.76
Resolution range	47.03 - 3.50 (3.63 - 3.50)
Space group	$P\ 4_3\ 2\ 2$
Unit cell	151.92 151.92 59.88 90.00 90.00 90.00
Total reflections	586490 (49357)
Unique reflections	8374 (684)
Multiplicity	70.00 (72.20)
Completeness (%)	89.75 (75.66)
Mean I/sigma(I)	23.79 (0.97)
Wilson B-factor	163.80
R-merge	0.17 (5.69)
R-meas	0.17 (5.73)
R-pim	0.02 (0.66)
CC1/2	1.00 (0.50)
CC*	1.00 (0.82)
Reflections used in refinement	8365 (684)
Reflections used for R-free	396 (38)
R-work	0.27 (0.38)
R-free	0.28 (0.46)
CC(work)	0.96 (0.57)
CC(free)	0.94 (0.45)
Number of non-hydrogen atoms	2706
macromolecules	2628
ligands	78
Protein residues	331
RMS(bonds)	0.01
RMS(angles)	1.17
Ramachandran favored (%)	93.27
Ramachandran allowed (%)	6.73
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	3.38
Clashscore	13.74
Average B-factor	179.70
macromolecules	177.77
ligands	244.56
Number of TLS groups	1

Table 5.2: Data Collection and Refinement Statistics for NRPV E2 Ectodomain

5.4.3 Structure Solution and Refinement

The NRPV E2 ectodomain could be solved using three long wavelength datasets and sulphur SAD phasing despite the low 3.50Å resolution. HKL2map[121] was used to find 8 sulphur sites and initial phases, Phenix autosol[111] was then used to generate

electron density maps and perform density modification. At this point it was clear that the data had initially been processed in the wrong enantiomorphic space group so the initial XDS[115] data processing was rerun with the correct space group. The low-resolution data had to be cut to 50Å resolution to achieve usable statistics. XSCALE[115], HKL2map[121], and phenix autosol[111]) were then used as before. The resulting electron density map was very poor quality and barely connected in places so rather than building the structure from scratch the BVDV E2 structure (2YQ2) was used as an initial model. The C-terminal two domains DC, and DD could be easily manually placed in COOT[112] with little manipulation. However, in order to place the N-terminal two domains DA and DB, they had to be separated fitted as rigid bodies. This model was then placed into phenix refine[122] which massively improved the electron density maps. At this point the N-terminal most domain DA was completely rebuilt in coot[112]. A combination of refmac[113] and manual coot[112] refinement was used to finish the model, Rwork and Rfree of 26.92/28.16% at 3.50Å resolution respectively with good stereochemistry. Refinement statistics can be found in Table 5.2 and electron density in Figure 5.8.

The final structure encompassed almost the full protein construct sequence, residues R2 to D335, there are three residues (P296-R298) missing which should be found at the tip of a loop in the C-terminal domain DD. The 13 C-terminal residues which are not resolved would be located in domain DD. Density in this region is poor and this region sits on a symmetry axis further confusing the matter due to statistical noise. However, there is space for these residues in the crystal packing, so it is likely they are present but just not well resolved (Figure 5.8). It is worth noting that the resolution, 3.50Å, of this structure is rather low for sulphur SAD phasing. Therefore, it is not unexpected that the initial maps are rather poor, and it is actually rather surprising it was solved without need for additional higher resolution data. The anomalous signal extended to 5.35Å with a d"/sig of 1.1. Initially SHELEX C[121, 190] found 8 sites (equivalent to 16 cysteines due to disulphide bonds), AutoSol[111] then successfully located the remaining

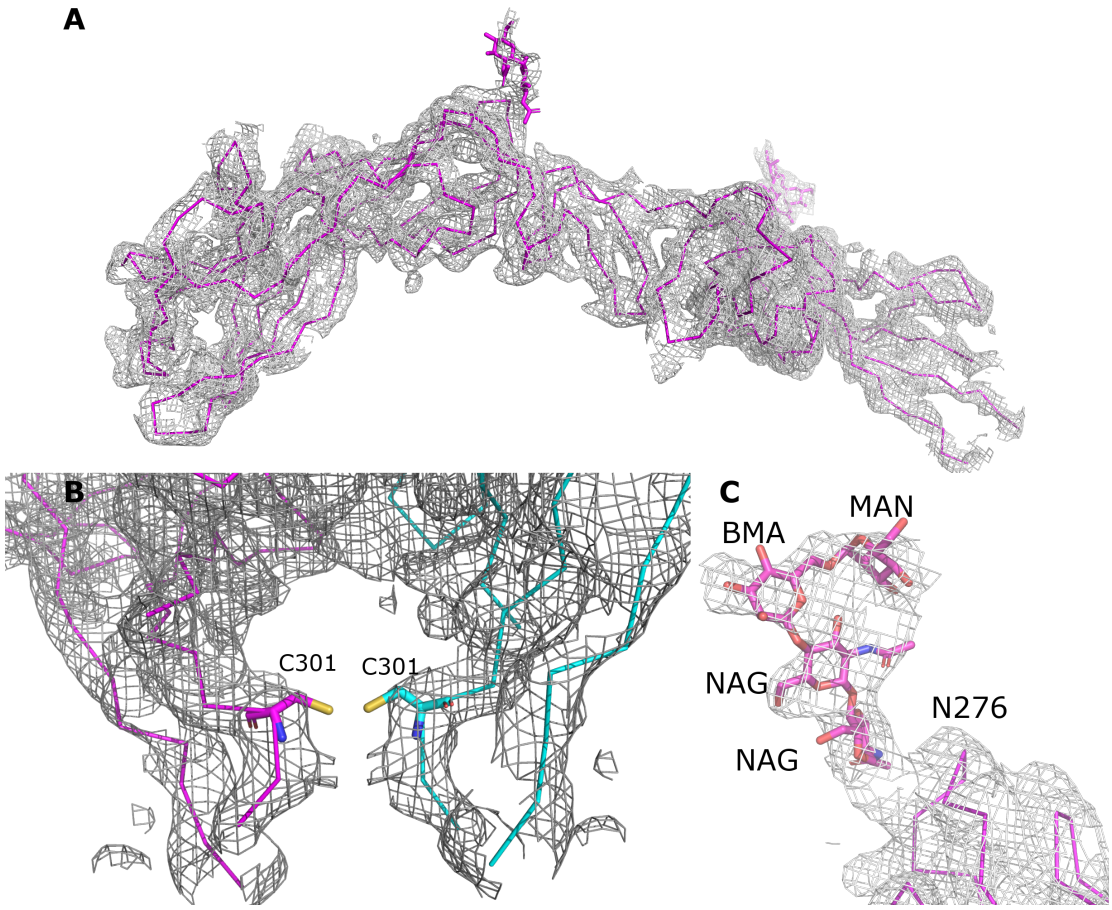


Figure 5.8: Electron Density Maps (2mFo-DFc generated using phenix map creation tool[167]) for the NRPV E2 structure. **A-** Entire structure, monomer in magenta carbon alpha trace. **B-** NRPV E2 C-terminal region highlighting intermolecular disulphide in sticks, monomer in magenta carbon alpha trace and symmetry copy in blue **C-** Glycosylation at residue N276 in sticks and surrounding area in carbon alpha trace.

five cysteines for a total of 21. The AutoSol solution had a figure of merit of 0.27 and an overall score of 27.46±17.21.

5.4.4 Structure Analysis

The final structure of the NRPV E2 ectodomain has one molecule in the asymmetric unit and is dimerized around a 2-fold symmetry axis. There was enough density to build one molecule in its entirety with only one small loop as well as the final 13 residues not resolved. The protein construct encompassed residues G1 to K348 and residues R2 to D335 (minus a small loop P296-R298) are resolved in the crystal structures. The resolution is quite poor (3.50Å) and some long flexible side chains

are not well resolved, nonetheless, the majority of the structure could be built.

As seen in Figure 5.9, the overall NRPV E2 ectodomain forms an elongated structure. The three C-terminal domains arrange head to tail in a near linear fashion with an increased interface angle for the N-terminal most domain. The C-terminal membrane proximal domain DD encompasses residues S281 to K348 and appears to be involved with dimerization as is the case with BVDV E2. It does not have well-resolved density although it is clear that it does contain a β-hairpin as well as loop regions. This is consistent with the structure of BVDV E2 DD domain and the fact that this region is the most well conserved amongst all pestivirus species. Although the electron density in this region is not complete, there was no density present to support the presence of a domain swap as was the case in the El Omari *et al* BVDV E2 structure[21]. However, the Li *et al* BVDV E2 structure[22] presented a homodimer without the domain swap. NRPV is likely to dimerize in a similar fashion stabilised by an intermolecular disulphide at C301 (Figure 5.8. The next domain DC encompasses residues Y166 to I280 and has a fold only seen before in BVDV E2. The overall structure is highly elongated and is composed of three antiparallel β-sheets stacked on top of each other in a linear fashion surrounded by loop regions.

Domain DB encompasses residues P97 to K165, has a loosely Ig-like fold and sits head to tail on top of domain DC. The fold is slightly modified as one of the β-sheets forming the β-sandwich has been perturbed and is not fully formed. There is also a β-hairpin to the N-terminus which is exposed to the solvent on one side. This is formed from an extension of the loop between the first and second β-strand of the fold. The domain interface between domains DA and DB is not head to tail as expected but rather has a bend. The membrane distal domain DA comprises residues S1 to T96 and has a more typical Ig-like fold with a central β-sandwich formed of two four stranded antiparallel β-sheets.

NRPV E2 ectodomain was expressed, purified, and crystalised in the absence of EndoHf. It is predicted to have a N-linked glycosylation site at residue N276

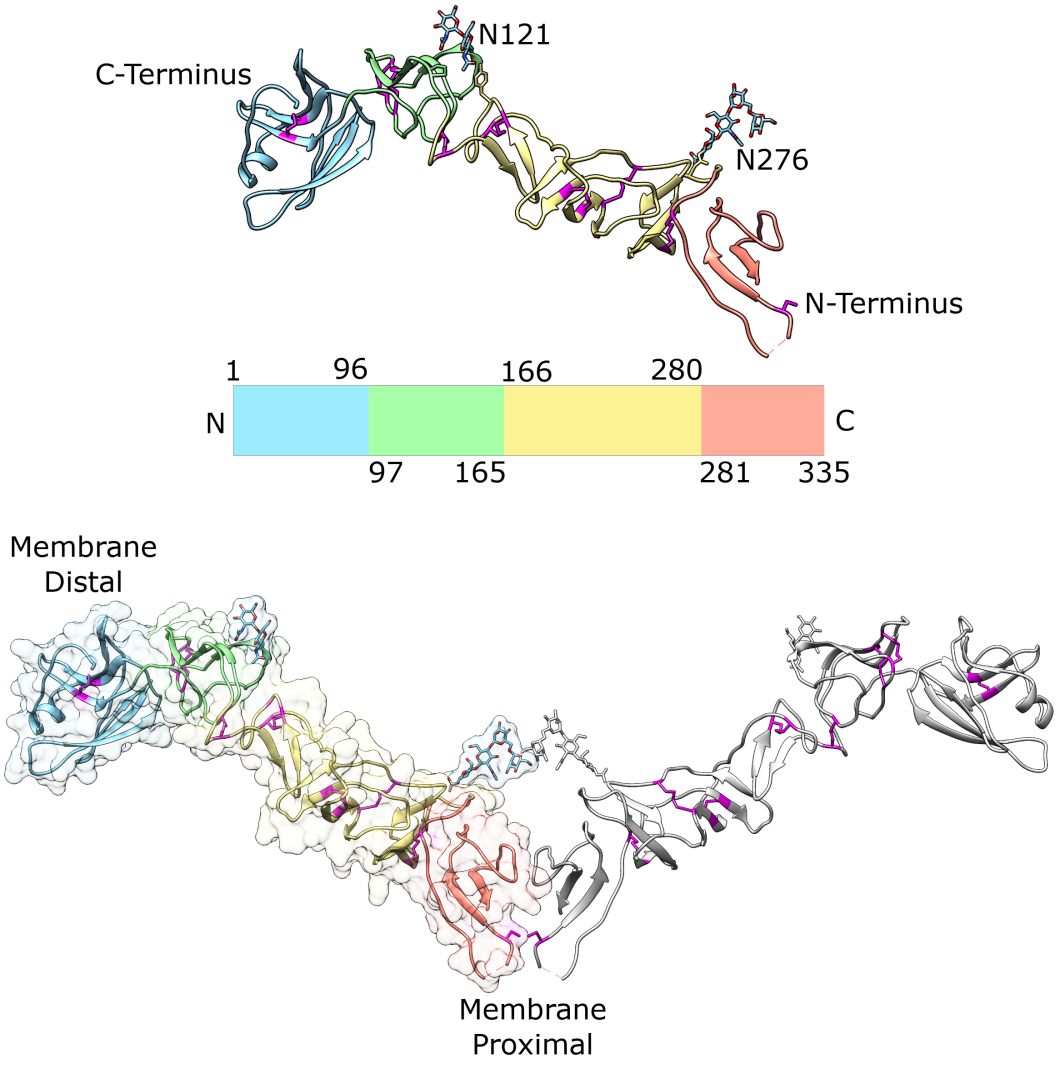


Figure 5.9: Structure of the NRPV E2 ectodomain with domains coloured as follows: DA-Blue, DB-Green, DC-Yellow, DD-Orange. **Top-** Cartoon representation of the NRPV E2 ectodomain with asparagine's marked and cysteines shown in magenta. **Middle-** Domain boundaries, coloured as per the structure cartoons. **Bottom-** On the left is a cartoon representation of the NRPV E2 ectodomain structure with surface shown, and on the right is the symmetry related copy which completes the dimer.

and three less likely glycosylation sites at N120/121, N160, and N222 (according to the NetNGlyc server[170]). In the crystal structure, glycosylation can only be seen at two sites N121 and N276. Density is particularly well resolved at residue N276 where two NAG residues, one β-D-mannose (BMA) residue, and one MAN residue have been modelled (Figure 5.8). At the other site N121, only the initial two NAG molecules are resolved. It is unlikely that N120 would be glycosylated if N121 were to be glycosylated as is the case here. N160 is located very close to N121 in the folded protein structure and points in towards the centre of the protein so there is unlikely to be room for this residue to support glycosylation. The final potential glycosylation site N222 was predicted to be the least likely, however the lack of density could be explained by the flexibility of the sugar chains. Therefore, it is impossible to tell if this site is glycosylated or not.

The protein construct used to solve this structure contained 17 cysteine residues, all of which are resolved in the electron density map. These cysteines form 9 disulphide bonds of which 8 are intramolecular located throughout the entire protein. There is one cysteine, C301, which appears to make an intermolecular disulphide bond with a molecule from the neighbouring asymmetric unit (Figure 5.8). This bond stabilises the homodimer as is the case with the related residue in the BVDV E2 ectodomain structures([21, 22].

5.5 SAXS of Pestivirus E2 Glycoproteins

5.5.1 Small Angle X-ray Scattering

SAXS is like X-ray crystallography in that it involves irradiating a sample with X-rays in order to determine structural information[191]. There are a few advantages over X-ray crystallography, a major one being that it uses a sample in solution. This means the difficult and chance-based process of growing a crystal is not necessary. Due to this, in general less material is needed as lots of protein is not used trying to grow a crystal and optimise it. Another advantage is that molecules in a crystal are static and could theoretically be stabilised in a non-physiological conformation.

Whereas with SAXS the proteins are in solution and able to move and take on different conformation and orientations. However, this is also a disadvantage as many proteins are inherently flexible and the infinite range of different conformations can cause difficulties in interpreting the data. The major disadvantage of SAXS is the fact it is much more 'information poor' and does not give high resolution atomic level data, but at best an overall shape for a molecule and information on its molecular weight[191]. As such it is usually not the first choice for a structural biology project but can provide very useful information if a crystal cannot be grown or optimised to generate a sufficient diffraction resolution. In this case an additional advantage is the ability to choose the experiment conditions, such as a pH that can reflect either the extracellular or endosomal condition.

5.5.2 Experimental Conditions and Data Processing

SAXS was carried out on all purified E2 proteins via a mail in service to the dedicated SAXS beamline B21 at Diamond light source. The samples were submitted at 7mg/ml and tested at both pH5 and pH8. Previous crystallographic structures of the BVDV E2 glycoprotein were solved at pH5.0 and pH7.5 and showed large structural differences between the two[21]. A size exclusion chromatography coupled SAXS method was used to guarantee an exact buffer match and separate any different multimers ensuring a homogeneous sample. Data were initially processed in ScÅtter[123] to generate the radius of gyration and the pair distance distribution function was generated using GNOM[125]. DAMMIF[126] was then used for Ab initio modelling and DAMAVER[127] to generate an average model both from the ATSAS[124] suite of programs.

5.5.3 Results

Table 5.3 shows the radius of gyration (Rg), Dmax used to calculate the Pr distribution, and calculated molecular weight (Mw) from DAMMIF[126]. The normalized Kratky plots can be seen in Figure 5.10 where they illustrate the flexible nature of these elongated multidomain proteins. Further plots can be seen in

appendix F. As can be seen, the Dmax is approximately 10% higher for the typical pestivirus E2 proteins (BVDV, BDV, and CSFV) at pH7.5 than it is at pH5.0. While this increase was very small it is a clear trend and is consistent with pH triggered changes in part of the protein. In the more atypical pestivirus E2 proteins (APPV, and NRPV), the Dmax is the same at both pH5.0 and pH7.5. This indicates that there is no pH triggered changes in these proteins. The effect of pH in pestivirus E2 glycoproteins is discussed more in section 5.7.2.

Protein	pH	Dmax	Rg	I(0)	MW (kDa)
			69.08	$0.57e^{-1}$	
BVDV E2 Ecto	5.0	216.0			147
			± 0.26	$\pm 0.16e^{-3}$	
			70.64	$0.86e^{-1}$	
BVDV E2 Ecto	7.5	233.0			147
			± 0.20	$0.22e^{-3}$	
			62.27	$0.16e^{-1}$	
BDV E2 Ecto	5.0	201.5			102
			± 0.14	$\pm 0.94e^{-4}$	
			72.14	$0.23e^{-1}$	
BDV E2 Ecto	7.5	224.0			135
			± 0.31	$\pm 0.88e^{-4}$	
			65.78	$0.92e^{-1}$	
CSFV E2 Ecto	5.0	204.5			134
			± 0.05	$\pm 0.10e^{-3}$	
			67.24	$0.15e^{0}$	
CSFV E2 Ecto	7.5	224.5			116
			± 0.13	$\pm 0.24e^{-3}$	
			63.81	$0.61e^{-1}$	
NRPV E2 Ecto	5.0	213.0			116
			± 0.19	$\pm 0.14e^{-3}$	
			62.62	$0.12e^{0}$	
NRPV E2 Ecto	7.5	213.5			103
			± 0.14	$\pm 0.19e^{-3}$	
			41.48	$0.83e^{-1}$	
APPV E2 Ecto	5.0	146.0			90
			± 0.06	$\pm 0.69e^{-4}$	
			41.65	$0.10e^{0}$	
APPV E2 Ecto	7.5	143.5			90
			± 0.06	$\pm 0.94e^{-4}$	

Table 5.3: Data Analysis Statistics from SAXS

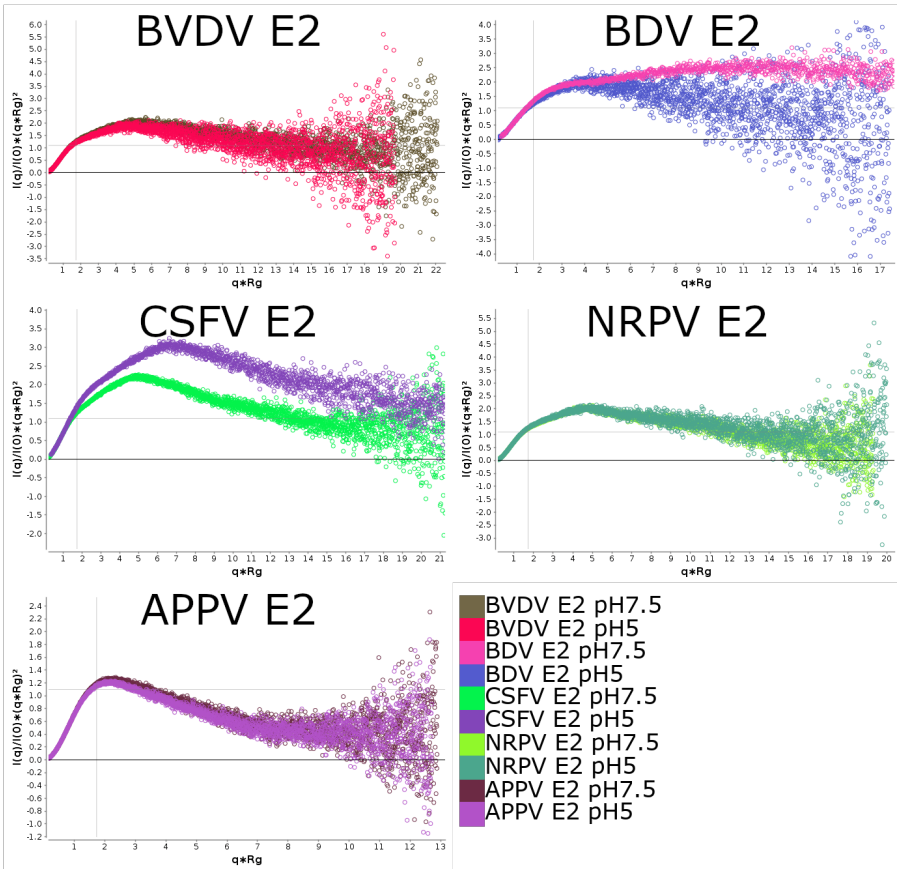


Figure 5.10: Normalised Kratky plots for each pestivirus E2 at both pH5 and pH7.5

Figure 5.11 contains averaged models for each of the five pestivirus E2 glycoproteins used in this study. These are only a rough outer envelope as SAXS is a low-resolution experiment. The data collected in this study are not optimal due to flexibility of the protein further diminishing the resolution. As can be seen, the four full length E2 proteins from BVDV, BDV, CSFV, and NRPV all have the same general elongated shape. They also all appear to be dimerised in solution at both pH5 and pH8. Unfortunately, due to the low resolution of the technique potential pH triggered changes cannot be visualised in the models. The models of APPV E2 correspond to a full-length homodimer. The crystal structure of APPV E2 ectodomain presented in section 5.3.5 is truncated with the loss of the C-terminal domain hypothesised to be involved with dimerization. For SAXS, APPV E2 was fully glycosylated and presented in solution so the full-length dimer model generated from this technique is likely to be the most biologically relevant.

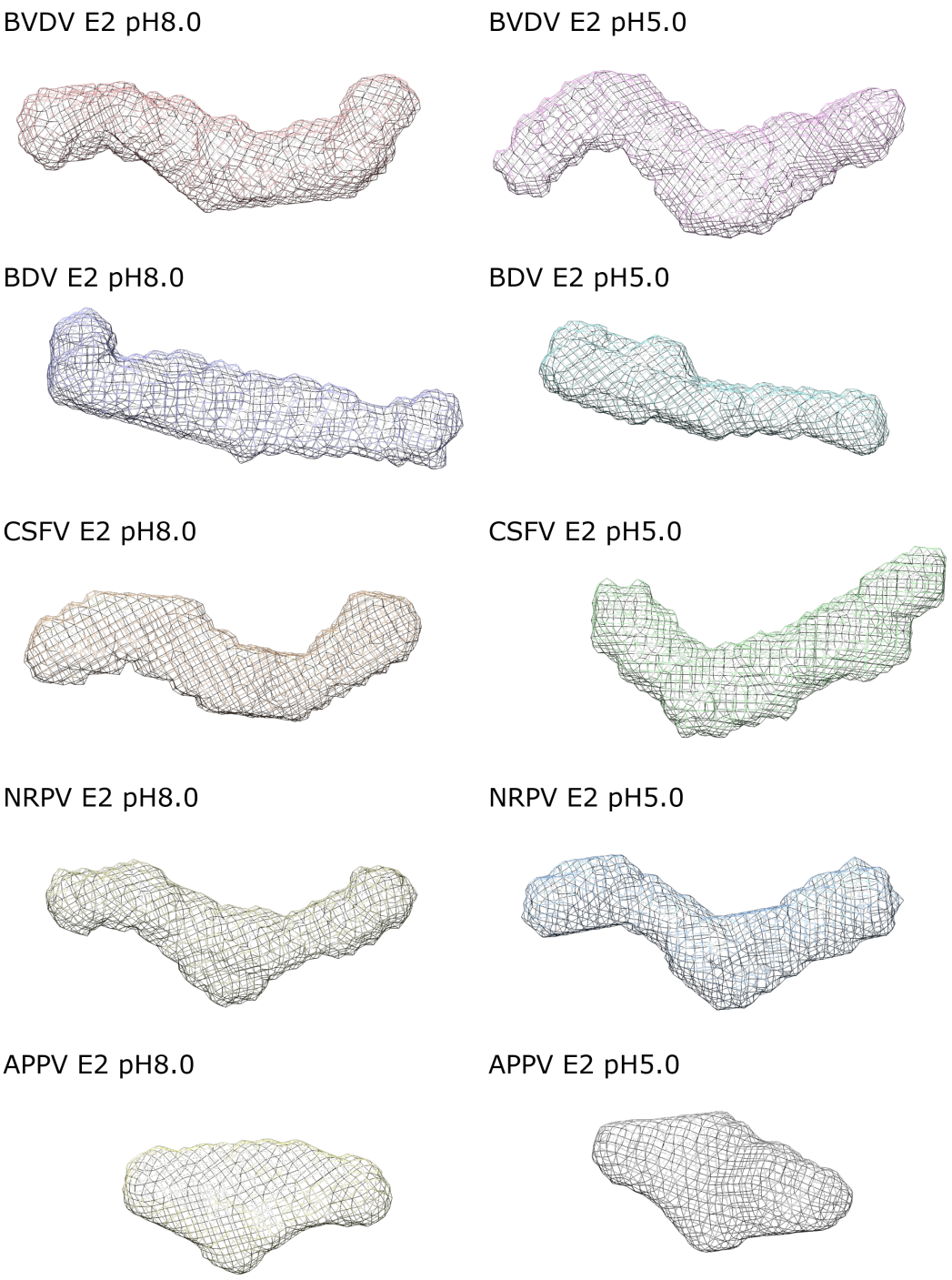


Figure 5.11: SAXS models of pestivirus E2 ectodomains at high and low pH. Shown are map representations of the average model from 10 rounds of ab initio modelling using DAMMIF.

5.6 Computationally Generated Models of Typical Pestivirus E2 Glycoproteins

5.6.1 Model of BDV E2 Ectodomain

Due to the success of the porCD46_SCR1-4 modelling (section 4.7.2) AlphaFold2[116] and RoseTTAFold[117] were used to generate computational models of the extracellular portion of the BDV E2 protein encompassing residues C4 to E341. Both AlphaFold2 and RoseTTAFold generated five models which are shown in the top and bottom of Figure 5.12 respectively. As can be seen the overall extended four domain β structure which is seen for BVDV E2 and NRPV E2 ectodomains is also predicted for the BDV E2 ectodomain. Both programs have good estimated confidence for the majority of the structure although this drops after about residue 280 for both the AlphaFold2 and roseTTAFold models. This means the majority of the C-terminal dimerization domain cannot be trusted. Interestingly this low confidence region covers almost exactly the region which was truncated at the C-terminus of the APPV E2 ectodomain structure. This C-terminal region is the most well conserved part of the protein, in fact it is the only region well conserved in both NRPV and APPV so it should be the region with the most confidence rather than the least. It is possible that the dimer is required to stabilse the fold, which would make it difficult to predict.

All ten models have the same general core structure for the two middle domain DB and DC. The C-terminal domain DD is modelled differently by the two programs with AlphaFold2 predicting a more extended conformation. This region has low confidence estimates for both programs so is unlikely to reflect the true structure. As the programs are only generating a monomer and this domain is involved with dimerization it is likely that the lack of second monomer is causing a lack of certainty in this area. Indeed, there is an intramolecular disulphide bond in the BVDV E2 structure[21] which stabilised the homodimer.

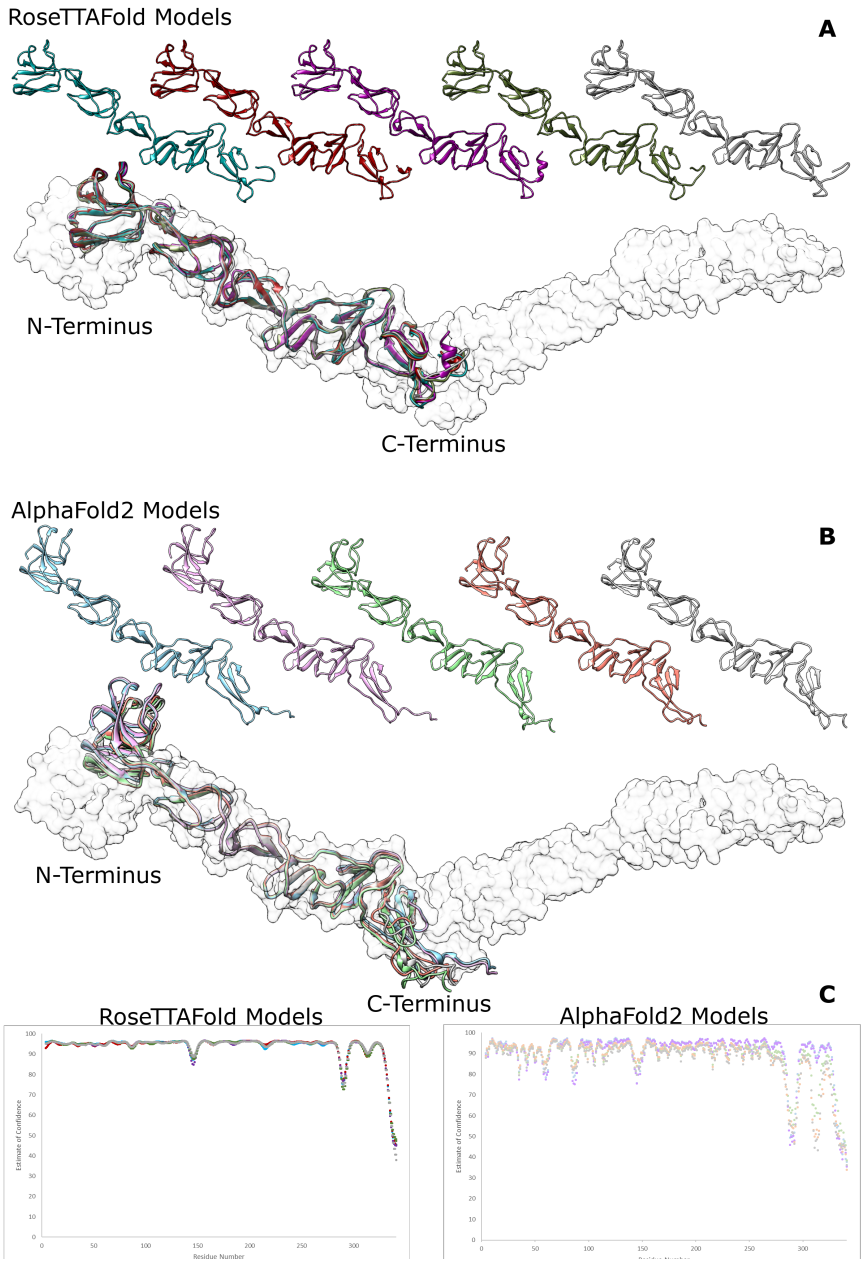


Figure 5.12: Computationally generated models of the BDV E2 ectodomain generated by the RoseTTAFold (**A**) and AlphaFold2 (**B**) programs. For each program on the top are the five individual models generated and underneath the five models are superposed on each other (RMSD for RoseTTAFold (**A**) models: 0.650Å/1.198Å over 322/338 residues, RMSD for AlphaFold2 (**B**) models: 0.760Å/4.203Å over 234/338 residues) and then on one monomer of the BVDV E2 structure (PDB: 2YQ2) which is shown as a surface representation(RMSD for RoseTTAFold (**A**) models: 0.853Å/14.080Å over 158/327 residues, RMSD for AlphaFold2 (**B**) models: 0.928Å/16.997Å over 146/327 residues). At the bottom (**C**) are the confidence out of 100 per residue for the RoseTTAFold and AlphaFold2 models. The RoseTTAFold confidence estimate was generated by assuming an estimated error of 2Å was equivalent to 90/100 confidence so 100-(5*estimated error (Å)) was used to generate the confidence estimates per residue.

The two programs also disagree as to the orientation of the N-terminal domain DA. While both programs predict the same conformation for the domain, they do predict a different orientation to that in the BVDV E2 structure. In BVDV E2, there is a bend at the DA:DB domain interface which the models predict a much straighter head to tail interface. The RoseTTAFold models all have the DA domain in the same orientation despite the straighter DA:DB domain interface with the central β-sandwich running parallel to the membrane. This differs to the AlphaFold2 models which all have the DA domain rotated slightly so the central β-sandwich is at an almost 45° angle to the membrane.

5.6.2 Model of CSFV E2 Ectodomain

As the crystal structure was also not solved for the CSFV E2 ectodomain and inital Cryo-EM experiments failed (appendix G), both AlphaFold2 and RoseTTAFold were used to generate computational models. The CSFV E2 protein construct encompassing residues R1 to E341 was used. The five models generated by both programs can be seen in Figure 5.13 with the RoseTTAFold models at the top, and the AlphaFold2 models at the bottom. As with the BDV E2 models, the overall extended four domain β-structure which is seen for BVDV E2 and NRPV E2 ectodomain is also predicted for the CSFV E2 ectodomain. Again, the overall model confidence is high but drops off after resides 280. This region of with less confidence corresponds to the region missing in the APPV E2 ectodomain X-ray structure and encompasses almost the entire C-terminal DD domain.

Once more, the two middle domains, DB and DC, both have the same general core structure in all the models as is found the BVDV E2 structure. Just as with the BDV E2 models, the C-terminal domain DD is modelled differently by the two programs. RoseTTAFold generally predicts the same conformation for all five models while AlphaFold2 predicts a more diverse set of options. Clearly the modelling of this domain is not likely to be accurate as reflected in the confidence scores. This is not surprising as BVDV has been seen to have two different conformations in this domain. In the structure solved by El Omari *et al*[21], there is a domain swap found

to stabilise the homodimer. While in the Li *et al* structure[22], a non-swapped homodimer has been modelled but the electron density is poorly defined at the domain swap. Due to high sequence identity (Figure 1.9) between pestiviruses in this region, it is likely that the actual conformation will be similar to that found the BVDV E2, and two different conformations are not mutually exclusive.

The N-terminal domain DA is again modelled in a different orientation by the two programs. As with BDV both programs predict the same domain conformation as was seen in the BVDV E2 structure. The RoseTTAFold models have a DA to DB domain interface which is more similar to the seen in the BVDV E2 structure. While in the AlphaFold2 models there is again a reduced angle at the domain interface leading to a straight head to tail arrangement, this is discussed further in section 5.7.1. The models also predict a difference in the rotation of the DA domain with regards to the rest of the protein. Interestingly, in contrast to the BDV models, AlphaFold2 predicts the β-sandwich will lie parallel to the membrane, while RoseTTAFold models place the β-sandwich at a 45° angle to the membrane.

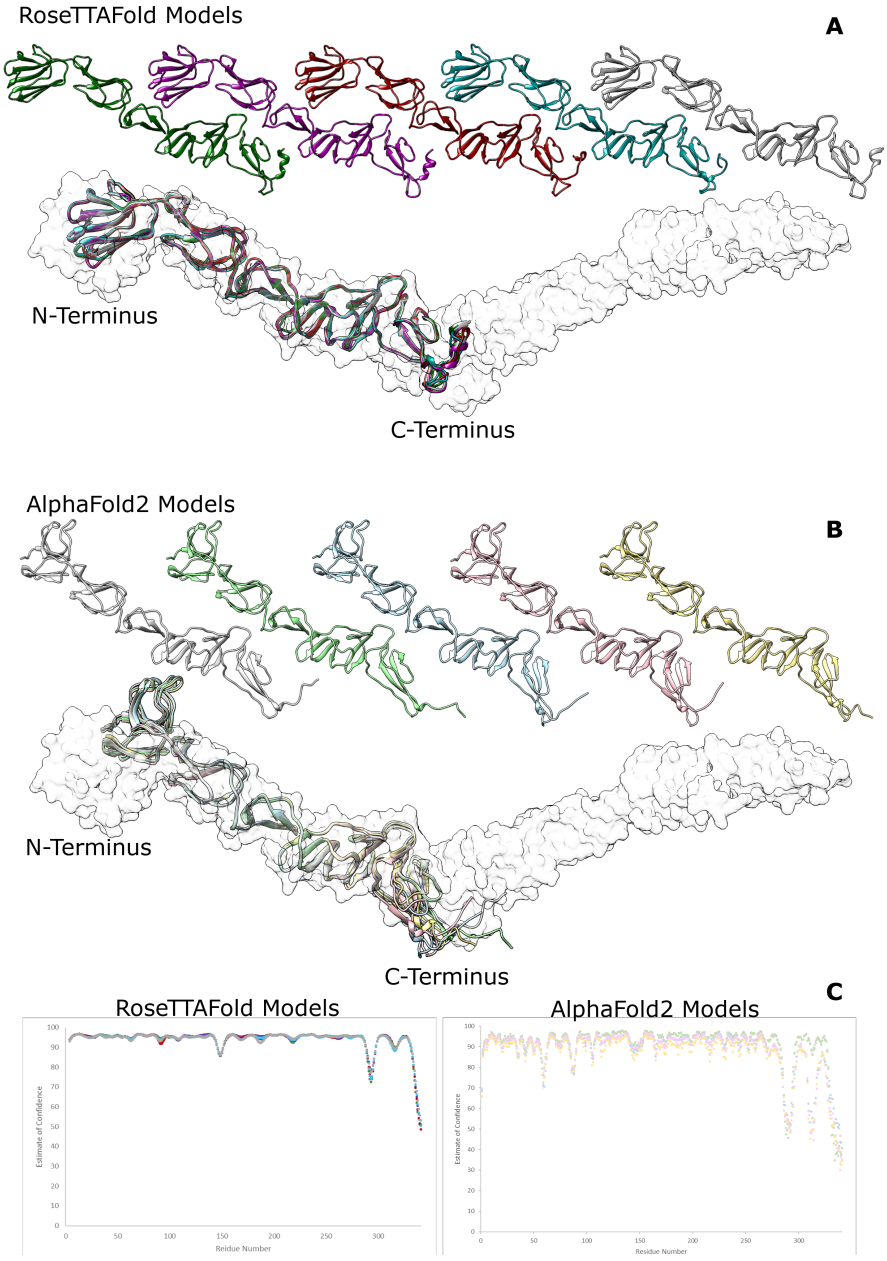


Figure 5.13: Computationally generated models of the CSFV E2 ectodomain generated by the RoseTTAFold (**A**) and AlphaFold2 (**B**) programs. For each program on the top are the five individual models generated and underneath the five models are superposed on each other (RMSD for RoseTTAFold (**A**) models: 0.600Å/1.000Å over 331/341 residues, RMSD for alphaFold2 (**B**) models:0.900Å/2.856Å over 304/341 residues) and then on one monomer of the BVDV E2 structure (PDB: 2YQ2) which is shown as a surface representation(RMSD for RoseTTAFold (**A**) models: 0.900Å/11.200Å over 163/329 residues, RMSD for AlphaFold2 (**B**) models: 1.100Å/17.488Å over 178/329 residues). At the bottom (**C**) are the confidence out of 100 per residue for the RoseTTAFold and AlphaFold2 models. The RoseTTAFold confidence estimate was generated by assuming an estimated error of 2Å was equivalent to 90/100 confidence so 100-(5*estimated error (Å)) was used to generate the confidence estimates per residue.

5.7 Structural Comparison of Pestivirus E2 Glycoproteins

5.7.1 Comparison of Overall Structure

As seen in Figure 1.9, of all the pestivirus species used in this study, APPV and NRPV have the most highly divergent E2 sequences. Sequence conservation is highest towards the C-terminus and drops off towards the N-terminus. NRPV E2 has very little sequence conservation in the N-terminal domain but it is still present, APPV however has a much-truncated E2 sequence. In APPV, there is little sequence conservation in the entire protein other than the C-terminal DD domain. Moreover, glycosylation sites are not conserved in APPV, although it is also a heavily glycosylated at sites different from those in NRPV. It is a similar story with the cysteines involved with disulphide bonds although these do become more well conserved towards the C-terminus where overall sequence similarity increases. This sequence identity or lack of is reflected in the structures of the individual proteins as illustrated in Figure 5.14.

The CSFV E2 sequence is missing two cysteine residues when compared to the other typical E2 protein from BVDV and BDV. In BVDV, these two cysteines for a single disulphide bond between the N-terminus of domain DB and the C-terminus of domain DA (Figure 5.15). The AlphaFold 2 models of the CSFV E2 ectodomain placed domain DA in a different orientation to where it is placed in the BVDV structure (Figure 5.14). Giving it a more head to tail arrangement with domain DB rather than the slightly reduced angle seen BVDV. On examination of this domain interface in an example AlphaFold2 CSFV E2 model (Figure 5.15) it can be seen this could be due to the loss of the disulphide bond. However, AlphaFold2 predicted a similar orientation for the BDV models (Figure 5.12) despite this disulphide being conserved in the BDV sequence (Figure 1.9). Therefore, it is more likely that this is a modelling artifact by AlphaFold2 rather than a structural difference. These residues and those surrounding them have lower confidence that the majority of the structure supporting the hypothesis of modelling error.

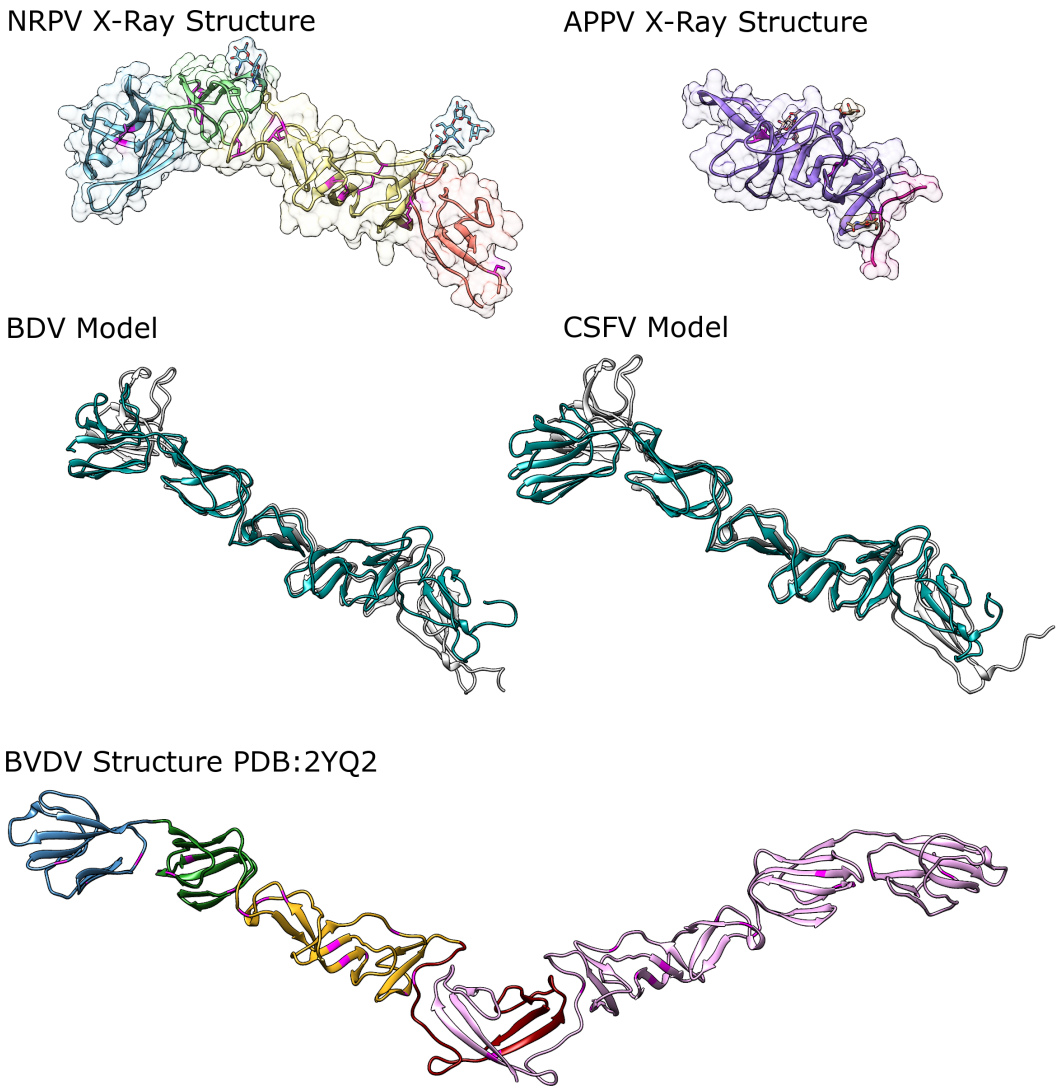


Figure 5.14: Comparison of pestivirus E2 ectodomain structures. **Top-** Cartoon and surface representation of the NRPV and APPV E2 ectodomain structures presented in this thesis. **Middle-** Superposition of computationally generated models of the CSFV (RMSD: 1.00Å over 142 residues, 10.2Å over 341 residues) and BDV E2 (RMSD: 0.96Å over 132 residues, 7.70Å over 338 residues) ectodomain with the AlphaFold2 models shown in grey and the RoseTTAFold models in turquoise, (Only one representative model shown for each modelling program) **Bottom-** Cartoon representation of the BVDV E2 ectodomain structure from PDB 2YQ2.

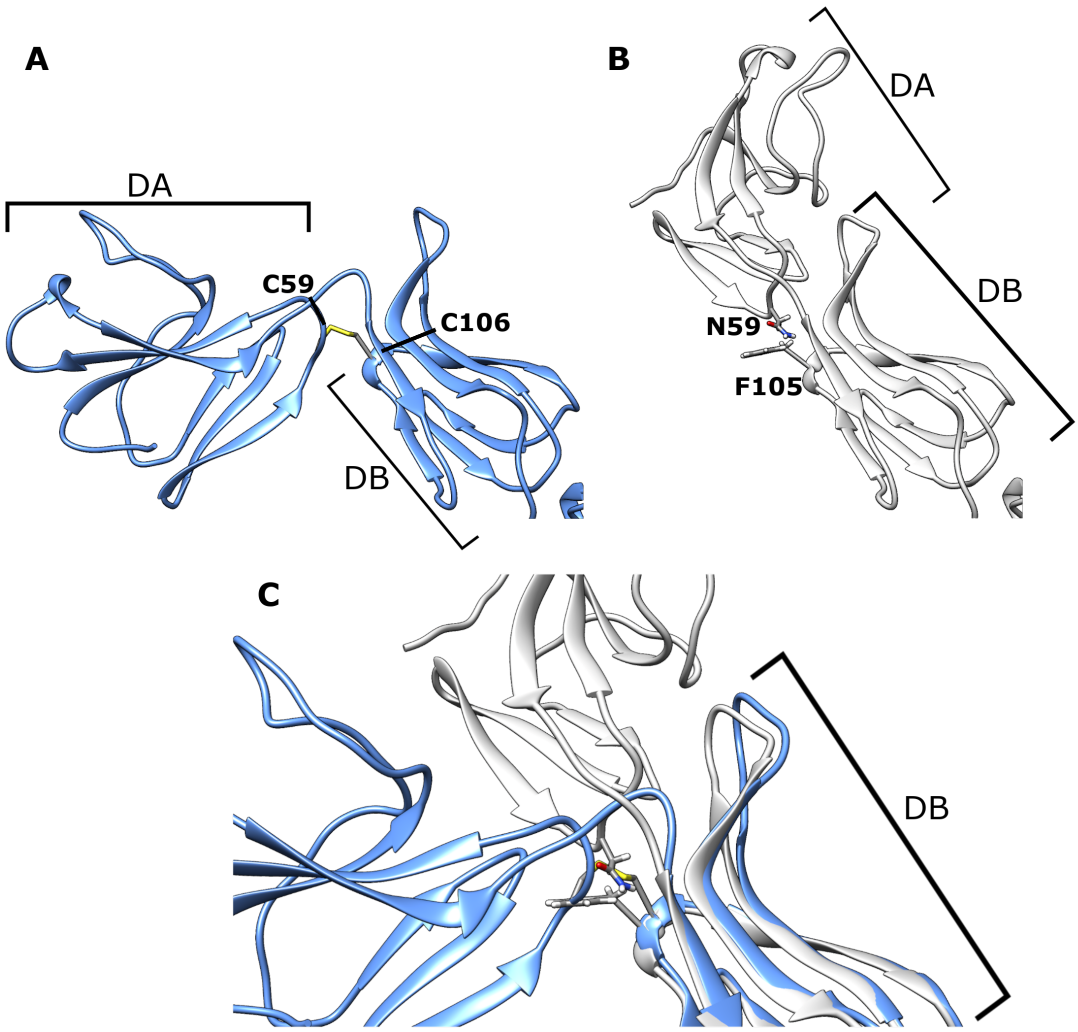


Figure 5.15: DA:DB domain interface in BVDV and CSFV E2 glycoproteins. A- Cartoon diagram of BVDV E2 domains DA and DB (PDB: 2YQ2) with the second disulphide highlighted (C59 and C106). B- Cartoon diagram of an example AlphaFold2 structure prediction of the CSFV E2 domains DA and DB with residues N59 and F105 shown in sticks. C- Super position of the BVDV(blue) and CSFV(light grey) E2 domains DA and DB

The overall shape of the NRPV E2 ectodomain is largely similar to that of the pH8 BVDV E2 structure although matching residues following the progression rule the RMSD is enormous, at 23.70Å over only 136 residues (Figure 5.16). Before the structure was solved, it was thought that the structures would be similar in the two unique C-terminal domains. On superposition with the BVDV E2 structure (PDB:2YQ2) domain DC has an RMSD of 1.17Å/3.32Å 65/106 residues and the C-terminal DD has an RMSD of 7.07Å over only 21 residues. The N-terminal

domains were hypothesised to be different due to the reduced sequence similarity. This was especially true for the N-terminal most domain DA which does not have homology to any proteins (other than rodent pestivirus E2) in the non-redundant protein sequence database. However, on examination of the protein structure the individual domains appear surprisingly structurally similar to those in the BVDV E2 ectodomain. As can be seen in Figure 5.16, they are easily superposed with RMSD of 7.30Å over all 75 residues for domain DA with the central β-sandwich being well conserved with most differences being in the loop regions (Figure 5.16). For domain DB the RMSD is 6.60Å over 68 residues, there is a β-sandwich in both structures although they do not quite align (Figure 5.16). The loop regions are more similar with the main difference being in the N-terminal β-hairpin which is shorter in NRPV. As seen in Figure 5.16 the main differences are at the domain interfaces which is likely causing the very high RMSD for the entire structure. This is especially evident at the DB:DA domain interface where there is a reduced angle between the two domains so they now sit at a 90° angle to each other rather than the almost linear 160° in the BVDV structure. In addition to the increased angle in the domain interface, NRPV domain DA is rotated around 90° when compared to BVDV domain DA.

Other than the conserved C-terminus, there is only one other structural feature common to both NRPV E2 ectodomain and APPV E2 ectodomain. This is a solvent exposed β-hairpin located at the N-terminus of domain DC in NRPV and the N-terminus of the entire APPV E2 protein (Figure 5.17), this is discussed further in section 5.7.3.

As expected from the negligeable sequence similarity of APPV to the other pestiviruses (around 20% sequence similarity to BVDV1 rather than the 60-70% of more typical pestiviruses), the structure is also completely different. A very short 39 amino acid section at the C-terminus of the structure can be used to superpose the structure with BVDV E2 ectodomain. Unfortunately, the rest of the C-terminus which has higher sequence similarity to other pestiviruses is

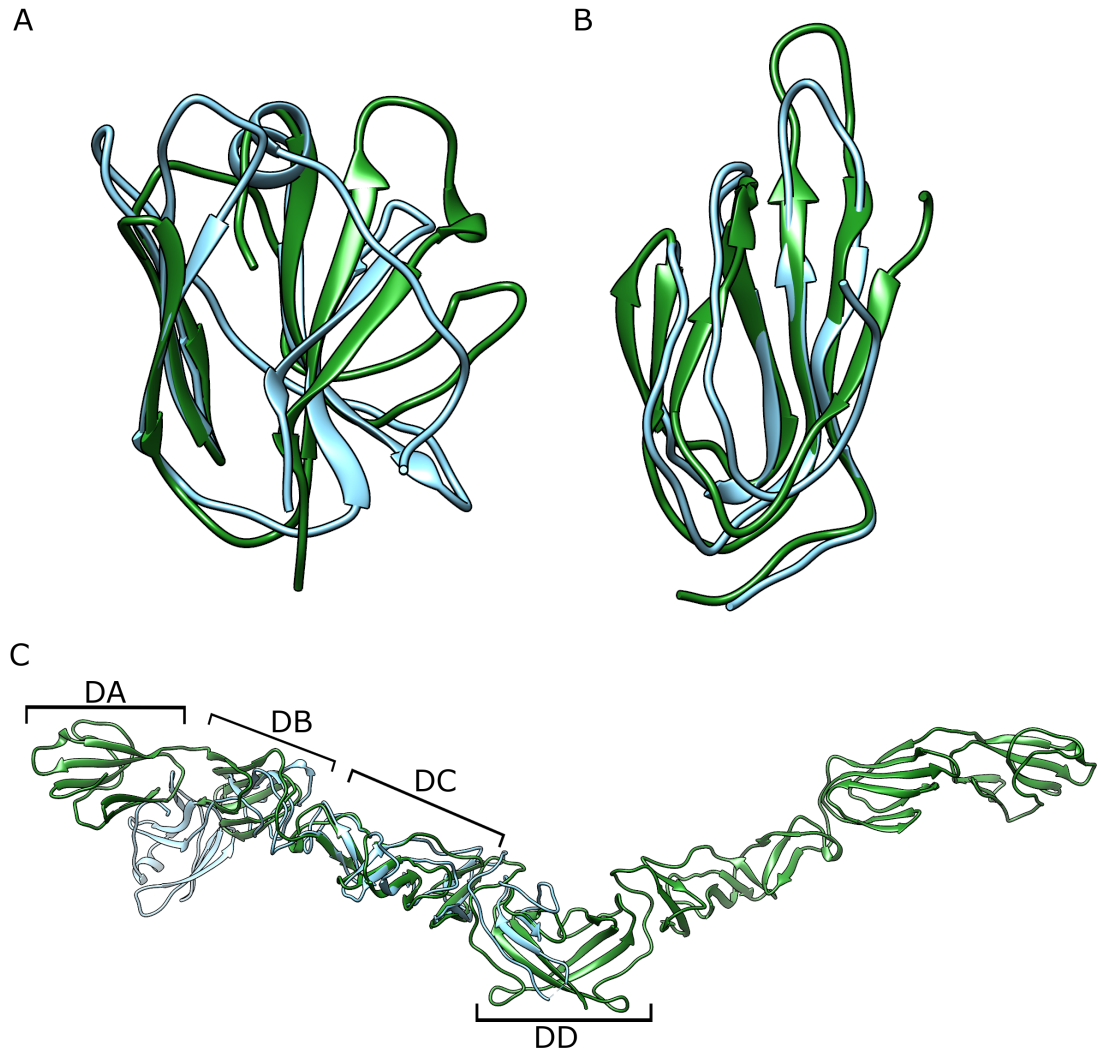


Figure 5.16: Superposition of NRPV E2 ectodomain with the BVDV E2 ectodomain (PDB:2YQ2), in all structures NRPV is in pale blue and BVDV is in green. **A-** Superposition of the N-terminal domain DA, **B-** Superposition of domain DB, **C-** Superposition of the entire ectodomain

missing in the structure. The structure also does not have homology to any other currently deposited structures according to the DALI[188] server. This is as expected due to the sequence of the N-terminal domain having negligible homology with any other proteins (other than APPV E2 sequences) in the non-redundant protein sequence database.

5.7.2 N-terminal Disorder

BVDV E2 ectodomain was solved at pH8 and pH5, in the low pH structure, the N-terminal most domain DA was entirely disordered and not visible. This was attributed to a single histidine residue H70 in domain DA acting as a pH sensor. Histidine residues have a pKa between physiological and endosomal pH hence have been implicated in pH-induced conformational changes in membrane fusion proteins. Both HCV and Sindbis virus E2 proteins also have conserved histidine residues which are sensitive to low pH[64, 192]. In Sindbis virus, low pH triggers the distal domain of the E2 protein to become disordered and expose a fusion loop on the E1 protein which it previously shielded. The proposed histidine switch in pestivirus E2 proteins is conserved in all species other than the atypical NRPV, APPV, Bat, and rodent pestiviruses. In these pestiviruses, it is either located in a truncated region for bat and APPV or replaced by a tyrosine residue in the case of NRPV and *rodent pestivirus*.

APPV was crystalised at both pH 8 and pH5, as well as various pHs in-between which are not presented. Both structures are nearly identical (RMSD: 0.20Å over all 148 residues) and as predicted by the absence of the conserved H70 there is no disorder as a result of the change in pH. NRPV was unfortunately only crystalised at pH8 but is likely to also lack disorder at pH5 due to the lack of the conserved H70.

Although the SAXS results did not present a large enough difference in Dmax to be ruled significant, there was a clear trend which agreed with the above hypothesis. The pestiviruses with typical E2 protein sequences all displayed a reduced Dmax at pH5 which the atypical pestivirus E2s all have a similar Dmax at both pH7.5 and pH5.0.

5.7.3 Solvent Exposed β-hairpin

All currently solved pestivirus E2 ectodomains contain a solvent exposed β-hairpin (Figure 5.17). In both the NRPV E2 ectodomain and the BVDV E2 ectodomain it is located at the very N-terminus of the DB domain, while in the APPV structure it is found at the N-terminus of the entire protein. These hairpins are particularly interesting for two reasons. Firstly, in both BVDV and CSFV a peptide in the β-hairpin region has been associated with receptor binding and cell entry[63, 178]. This could potentially allow this region to interact with another protein, such as a receptor, in a similar but not identical fashion for each virus. As a result, it could play a role in host cell tropism.

In the BVDV E2 protein, the hairpin is hidden within the protein at pH 8. However, due to the N-terminal disorder at pH5 the hairpin becomes more exposed to the solvent. As stated above, both APPV and NRPV are unlikely to undergo this N-terminal disorder and as a result their β-hairpins are permanently solvent exposed. In APPV, the hairpin sticks straight up from the protein exposing the entire surface to the solvent, while in the NRPV structure the hairpin lies alongside the N-terminal domain DA resulting in only one face being exposed to the solvent.

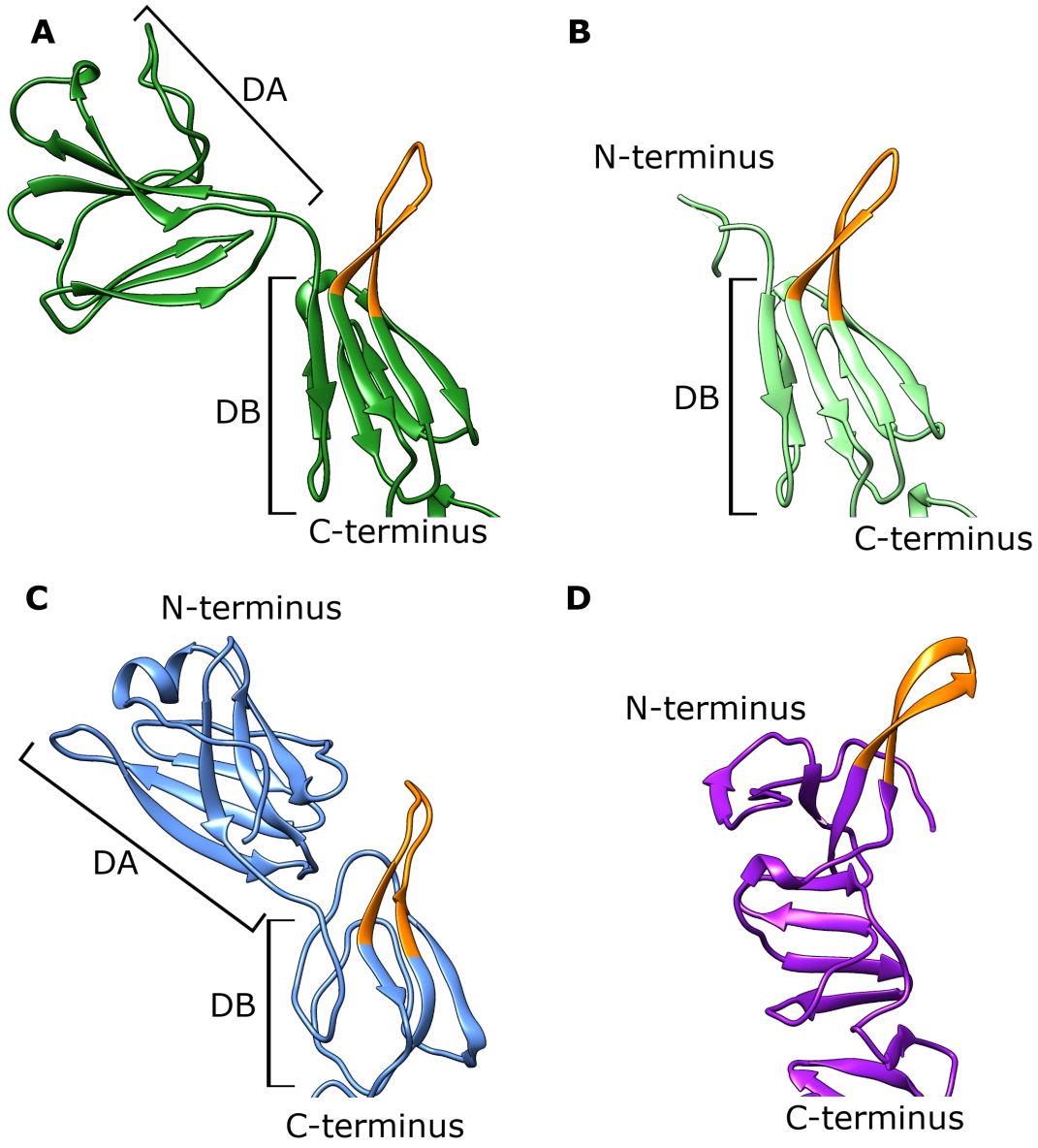


Figure 5.17: Cartoon representation of the 3 currently solved pestivirus E2 ectodomains, highlighting the N-terminal β-hairpins which are coloured orange. **A-** Domains DA and DB of BVDV E2 from pH8 (PDB:2YQ2) coloured in dark green with C140-V153 in orange. **B-** Domain DB (and peptide from DA) of BVDV E2 from pH5 (PDB:2YQ3) coloured in light green with C140-V153 in orange. **C-** Domain DA and DB of NRPV E2 coloured in light blue with V144-F157 in orange. **D-** N-terminal section of APPV E2 ectodomain coloured in purple with Q45-Y58 in orange.

5.8 Conclusion

Pestiviruses contain two membrane anchored envelope glycoproteins E1 and E2([46]. Of these two proteins E1 is thought to be the fusion protein and to contain the fusion peptide. E2 is thought to act as a chaperone or scaffold for E1 as well as playing a role in receptor interaction[21, 22, 91]. Information on these could provide valuable insight membrane fusion and cellular entry. Sequence alignments of the E2 region of different pestivirus species reveal two different types of E2 proteins (Figure 1.9). There are those with a typical sequence in the E2 region such as E2 from BVDV, BDV, and CSFV where there is high sequence similarity across the entire protein sequence. E2 proteins from NRPV and APPV however have atypical E2 protein sequences with higher sequence similarity at the C-terminus and little towards the N-terminus. At the extreme, some such as APPV have a truncated E2 protein entirely missing the N-terminal region. This thesis reveals the X-ray crystal structure of two atypical E2 ectodomain protein structures from NRPV and APPV as well as computational models of the more typical CSFV and BDV E2 ectodomains.

The NRPV E2 ectodomain structure is similar to that of the BVDV E2 ectodomain, with some differences towards the N-terminus as suggested by the reduced sequence similarity. Overall, it adopts an elongated structure composed of four domains. The two N-terminal domains both have a conformation only seen before in the BVDV E2 ectodomain structures. The two N-terminal domains both have an IG like fold but with a slightly different conformation to that seen in the BVDV structure. The APPV E2 ectodomain revealed a completely novel fold composed of β-strands and loops with a very prominent solvent exposed β-hairpin to the C-terminus. This structure appeared to have a C-terminal truncation of 58 residues which comprised almost the entire C-terminal domain when aligned with the BVDV sequence. A small section of around 39 residues at the C-terminus of the APPV structure can be aligned with the BVDV structure. This region covered a small peptide at the N-terminus of the DD domain and a β-hairpin at the N-terminus of the DC domain. Due to this and the high sequence similarity at the C-terminus it

is likely that the C-terminal domain of APPV adopts a conformation similar to the DD domain of BVDV E2. It is possible that this high sequence conservation reflects the importance of homodimerization, it is also possible that heterodimerisation with E1 also involved this domain. Although crystals were obtained for the BDV and CSFV E2 ectodomains, these failed to produce strong diffraction despite extensive optimisation attempts. As such SAXS was carried out as an alternative alongside computational modelling utilising both AlphaFold2 and RoseTTAFold.

Two interesting features were observed upon examination of the new NRPV and APPV crystal structures, both relating to the N-terminal region. The BVDV E2 ectodomain was solved at both low and high pH with the low pH structure demonstrating disorder of the N-terminal DA domain[21]. APPV was also solved at high and low pH but shows no conformational differences between the two conditions. Although NRPV was only solved at high pH, SAXS results indicated it is unlikely to suffer any pH triggered disorder. These findings are supported by the lack of sequence conservation of a histidine residue (H70 in BVDV E2) in the DA domain by the atypical E2 proteins. This histidine is thought to act as a pH switch and trigger the destabilisation of the N-terminal domain. There is an exposed β-hairpin at the N-terminus of the APPV E2 protein which appears structurally similar to the β-hairpin which is exposed upon pH triggered disorder of the BVDV DA domain. Although NRPV is not thought to undergo pH triggered conformational changes it also contains an exposed β-hairpin with similar residue characteristics. Due to the conserved structural feature and residue characteristics, it is likely this region is important in protein function. This is supported by two cellular studies which implicated this region as being involved with cellular entry in[63, 178].

6

Conclusions and Future Directions

Contents

6.1 Conclusion 183
6.2 Future Directions 187

6.1 Conclusion

6.1.1 Importance

Pestiviruses, a member of the *Flaviviridae* family, affect many animals, mainly ungulates although more recently both bats and rats have been seen to be affected[5]. These viruses have been shown to have a huge economic impact especially in the dairy and pig industry through reduced milk production as well as loss of animals[26, 27]. Due to the economic impact of outbreaks, pestivirus infections, specifically by BVDV, CSFV, and BDV are notifiable diseases[33] and in many countries there is often a test and cull program in place. Many current vaccines are not used (or outright banned in some countries) due to the inability to distinguish between those animals with an infection and those who have been vaccinated[31]. Another difficulty presented by these viruses is their endemic nature in some wild animal populations, such as wild boar[32], providing a virus reservoir for mutation and subsequent farm

animal infection. Information on the viral cell entry mechanism through investigation of the envelope glycoproteins involved, could provide insights into future directions for development of new more targeted vaccines or potentially treatment options.

The *Hepacivirus* are another member of the *Flaviviridae* family of which the important human pathogen HCV is a member[2]. The genera *Hepacivirus* and *Pestivirus* are thought to potentially share a similar membrane fusion mechanism which is likely to be different from the other members of the *Flaviviridae*. Two proteins are necessary for membrane fusion, E1 and E2[61], rather than only an E protein. The smaller of the two proteins, E1, is thought to be the fusion protein for both the Pestiviruses and Hepaciviruses. Due to its small size, it is unlikely to be able to adopt the class II fusion protein fold seen in the other members of the *Flaviviridae* family. In fact, when taking into account the small size and lack of any predicted alpha helices it is hypothesised that the pestiviruses belong to an entirely new class of fusion proteins[57]. As such, information on pestivirus membrane fusion is interesting on a basic science front as well as for informing on future research into control of pestiviruses and potentially also HCV.

6.1.2 Summary of Findings

Presented within this thesis is the design of nearly 100 protein constructs, expression testing of 70, and the successful expression of 22. Large scale expression was carried out on 16 resulting in the successful production of pure protein for the E2 ectodomain of all five pestiviruses studied. Protein was also produced for various regions of the CD46 ectodomain, both bovine and porcine.

X-ray crystallography structures of the bovCD46_SCR1-2 and bovCD46_SCR1-4 constructs were solved. Although an equivalent porcine construct was crystalized, it did not diffract to sufficient resolution to solve the structure, so instead computationally generated models are presented. The bovine CD46 ectodomain structures revealed an overall structure similar to that of human CD46. Size exclusion chromatography and dynamic light scattering experiments revealed the presence of

a homodimer in bovine CD46 in solution (as was seen in both crystal structures), although only a monomer was seen for porcine CD46.

Structures of two atypical pestivirus E2 ectodomains from APPV and NRPV were solved, again using X-ray crystallography. Here sulphur-SAD phasing was utilised due to the lack of any homologues available in the PDB. The two other pestiviruses studied in this thesis crystalised but did not diffract sufficiently well to solve the structure despite extensive optimisation efforts. The two recently released deep learning based structure prediction programs, AlphaFold2[116] and RoseTTAFold[117], were therefore used to generate models. These structures highlighted the diversity in pestivirus E2 ectodomain structure. Sequence alignments and the structures presented allowed the classification of the pestivirus E2 proteins as typical (BVDV, BDV, and CSFV) or atypical (APPV and NRPV). Finally, SAXS experiments supported pH triggered conformational changes in the typical E2 proteins but no changes in the atypical proteins from APPV, and NRPV. SAXS also confirmed that BVDV, BDV, CSFV, APPV, and NRPV E2 ectodomains are dimeric in solution.

6.1.3 Implications of Receptor Structure

The structures of bovine CD46 revealed a structure similar to that of the human proteins. A region in the N-terminal most SCR has been identified as being involved with pestivirus infection[62]. The interaction was narrowed to two peptides in SCR1 which form a single patch on a β-sheet. The first peptide has a well conserved sequence between pestiviruses while the second is less well conserved. On examination of the newly solved structures, there is a clear loop shift in the 2nd peptide region when compared with the human structure. Since pestiviruses are species specific, it is possible that host cell tropism is determined within this region.

The bovCD46_SCR1-4 structure revealed an overall shape akin to a hockey stick with a much-reduced angle between the membrane proximal SCRs 3 and 4. This shape is not found in other SCR containing proteins which normally have all SCRs

stacked head to tail. The bend between SCR3 and SCR4 is also seen in the human CD46_SCR1-4 structure and is attributed to a five-residue hydrophobic insertion in the CD' loop of SCR3[94]. Sequence alignments demonstrate the presence of this insertion in other mammalian species such as pigs and sheep, indicating they also likely share this unusual structure.

Another interesting feature revealed by the crystal structures was the presence of homodimerization through SCR2 and the SCR1:SCR2 interface. A similar crystallographic dimer is present in the human CD46_SCR1-4 structure although it is not mentioned in the publication[94]. This dimer formation was supported by size exclusion chromatography elution volumes and molecular weight estimates through dynamic light scattering.

Formation of this homodimer would require both monomers to lie on the surface of the cell. This is due to the angle between SCR3 and SCR4, if the proteins were to stick up from the cell, then one of the monomers would be unable to be membrane anchored. The STP region of CD46 is located between the membrane anchor and SCR4, different isoforms of the protein have changes in this region[161]. It is possible that these changes result in the protein sticking up from the surface or lying along it.

Fluorescent imaging of BVDV and CD46 in cells shows that BVDVs movement across the cell surface is much faster if it is unable to bind CD46[95]. Thus, it is hypothesised that CD46 may anchor the virus at points of endocytosis. It is possible the dimer formation aids in this anchoring role.

6.1.4 Implications of New Pestivirus E2 Structures

The two novel X-ray structures of the atypical pestivirus E2 proteins from APPV and NRPV highlight the variety of structure within these proteins. Despite the highly divergent structures, all contain a solvent exposed β-hairpin towards the N-terminus. Two papers have implicated the hairpin region in BVDV and CSFV as being involved with receptor binding[63, 178]. This is particularly interesting

as the hairpin in APPV is located much closer to the viral membrane in APPV than it is in the other pestiviruses.

An interesting feature of the previously solved BVDV structure is the disorder of the N-terminal domain DA at low pH which exposes the previously mentioned hairpin to the solvent[21]. SAXS experiments support these structural changes in the typical E2 proteins from BVDV, BDV, and CSFV but not in the atypical E2s from NRPV and APPV. The crystal structures of APPV at pH5 and pH8 also confirm there are no structural changes. As the pestiviruses have been shown to undergo pH triggered membrane fusion, this lack of disorder is particularly interesting[84]. Alphaviruses, from the *Togaviridae* family, have an E1:E2 heterodimer and have a membrane fusion mechanism triggered by a drop in pH[64]. With these viruses, the E2 protein covers the E1 protein protecting its fusion peptide. When triggered by low pH structural changes, E2 partially unfolds and reveals the fusion peptide in E1 thus allowing membrane fusion[64]. It was thought that a similar mechanism could take place in pestiviruses, although, the lack of structural changes in NRPV and APPV challenges this. However, given their highly divergent sequences it possible that they might have a slightly different mechanism of membrane fusion.

A secondary hypothesis for the N-terminal disorder in domain DA is to expose the β-hairpin on domain DB to the solvent[21]. As APPV does not contain a DA domain the hairpin in APPV is always solvent exposed due to it sticking straight up from the N-terminus. However, there is still the problem of NRPV which contains a DA domain but does not display any N-terminal disorder. As no receptors have been proposed for NRPV, it is possible this virus utilises a different receptor to the other pestiviruses and no longer requires this hairpin to be exposed.

6.2 Future Directions

6.2.1 Receptor Interaction

Currently, the only receptor which has been shown to directly bind pestivirus E2 is ADAM17[91]. The other main cellular receptor suggested for pestiviruses is

CD46, this has been a candidate receptor for much longer than ADAM17, however direct binding has still not been demonstrated[62]. As I was unable to demonstrate binding between BVDV E2 ectodomain and bovCD46_SCR1-4 during the course of this project, it is of particular importance to more thoroughly explore the potential of an E2 CD46 interaction. It has been clearly shown CD46 plays an important role in BVDV[89] and APPV[88] infection, however it is not clear as to whether this is through direct interaction with the virus. The cellular receptors for CSFV also need to be clarified as there are currently two studies which contradict each other as to whether CD46 is involved CSFV infection[87, 88].

The second pestivirus cellular receptor, ADAM17, has only currently been implicated in CSFV infection[91]. Hence, it would be interesting to explore which other pestivirus species may also bind ADAM17 from their host organism. Pestiviruses enter the cell through pH triggered membrane fusion after endocytosis[84]. Because pestiviruses can still infect CD46 knockout cell lines, albeit to a lesser degree, it has long been assumed that another receptor is involved[62]. It would also be interesting to investigate whether ADAM17 is an alternative receptor to CD46 or whether they are both utilized.

Pestiviruses are species specific, and it is not currently known what determines host cell tropism. Once direct binding experiments can be carried out, then it will be possible to investigate the region of pestivirus E2 which is involved in receptor binding and host cell tropism. This information could inform future control mechanisms. Clearly, a complex structure containing both E2 and a receptor would be extremely valuable in this area.

As demonstrated by the strong prediction of bovCD46_SCR1-4 clearly the new wave of structural modelling programs such as AlphaFold2[116] and RoseTTAFold[117] will play an important role in the future of structural biology. Either as a method of generating molecular replacement models to reduce the need for experimental phasing, as a tool for designing expression constructs, or as an alternative to experimental structure solution methods when they do not yield results. Hopefully, future

iterations of these programs will be able to look at interactions between proteins which would be extremely valuable in this project and other such difficult systems.

6.2.2 Membrane Fusion

The largest outstanding question in pestivirus membrane fusion is the structure of the pestivirus E1 protein. It seems the largest difficulty in solving this is the production of recombinant E1 proteins. It has been suggested that E2 is required for the correct folding of E1[60]however, in this study E1:E2 co-expression was not successful. The interaction between E1 and E2 is thought to also occur through their membrane anchors, so the full-length protein may be needed[14]. As the production of full-length proteins, replacement of membrane anchors with dimerization proteins, and expression of E1 linked to E2 was not successful in this study, it is likely this will prove difficult.

An interesting feature studied in this thesis was the pH triggered conformational changes to the N-terminal domain in typical pestivirus E2 proteins[21]. Two potential roles have been proposed for this feature: unveiling of the E1 protein and its fusion protein, or exposure of a β-hairpin thought to be involved in receptor interaction. As it is known that reduced pH triggers membrane fusion in pestiviruses[84], investigation into the role of this N-terminal disorder should provide valuable insight into the membrane fusion process. As the atypical E2 proteins do not seem to undergo this pH triggered conformation changes, it would be interesting to see if membrane fusion in the pestiviruses with atypical E2 proteins is also triggered by a drop in pH. The structures presented in this thesis highlighted an interesting β-hairpin located at the N-terminus of APPV and at the N-terminus of the DB domain in other pestiviruses. This region has been implicated in receptor binding in both CSFV and BVDV[63, 178]. Swapping the sequence of this region for that of different pestiviruses species in gain and loss of function experiments could confirm if this region determines host cell tropism. As the structure of the APPV E2 ectodomain is so different, confirming this hairpin is involved in receptor binding in this species could prove insight into membrane fusion. This would be particularly

interesting as the hairpin of APPV is located much closer to the viral membrane than the hairpins in the other pestiviruses. The hairpin in BVDV becomes more exposed to solvent at low pH due to conformational changed in the N-terminal DA domain[21]. Investigations into hairpin function in NRPV at different pH would be particularly interesting as it is not though to change conformation at low pH.

Although initial cryo-EM experiments were carried out during the course of this project they were not successful (Appendix G). Far more could be done in this area especially on those proteins which did not diffract to sufficient resolution and those which have very low expression levels. Tomography and sub tomogram averaging would also be a great way to study the fusion proteins in their membrane embedded environment. A recent study by Renner *et al*[193] managed to refine immature and mature flavivirus particles from a single grid and obtain the structure of Spondweni virus to a resolution of 3.8Å. The study of pestivirus particles is likely to be much more difficult as they do not display many E1 and E2 glycoproteins on their surface[25]. Additionally, the viral particles themselves are polymorphic and do not display any symmetry[25]. Therefore, it is likely tomography will be needed to identify the structures on the surface of the particles and then sub tomogram averaging to solve the structures. Electron microscopy technology and software are advancing rapidly so hopefully study of these difficult systems will become increasing feasible.

Appendices

A

GenBank Accession Codes

Bovine viral diarrhea virus 1 - NP_040937

Bovine viral diarrhea virus 2 - AMO65207

Bovine viral diarrhea virus 3 - YP_002967452

Border disease virus - NP_620062

Classical swine fever virus - NP_075354

Atypical porcine pestivirus - YP_009268709

Norway rat pestivirus - YP_009109567

Aydin-like pestivirus - YP_006860588

Pestivirus giraffe - NP_620053

Pronghorn antelope pestivirus - YP_009026415

Porcine pestivirus isolate bungowannah - YP_008992092

Linda virus - YP_009407716

Bat pestivirus - AYV99177

Rodent pestivirus - ATP66856

Phocoena pestivirus - QFQ60724

B
Gene Sequences

B.1 BVDV_E1-E2

ATGGCCTCTCCTTACTGCGACGTGGACCGGAAGATCGGCTACATCTGGTA
CACCAAGAACTGCACCCCTGCCTGTCTGCCCAAGAACACCAAGATCGTTG
GCCCTGGCAAGTTCGGCACCAATGCCGAGGATGGCAAGATCCTGCACGAG
ATGGGAGGCCACCTGTCTGAAGTGCTGCTGCTGTCTCTGGTGGTGCTGAG
CGATTTCGCCCCTGAGACAGCCAGCGTGATGTACCTGATTCTGCACTTCA
GCATCCCTCAGAGCCACGTGGACGTGATGGACTGCGATAAGACCCAGCTG
AACCTGACCGTGGAACTGACCACCGCCGAAGTGATCCCTGGCAGCGTTTG
GAACCTGGGCAAATACGTGTGCATCAGACCCAACTGGTGGCCCTACGAGA
CAACAGTGGTGCTGGCCTTCGAAGAGGTGTCCCAGGTTGTGAAACTGGTG
CTGCGGGCCCTGAGAGATCTGACCAGAATTTGGAACGCCGCTACCACCAC
CGCCTTTCTCGTGTGCCTGGTTAAGATCGTGCGGGGCCAGATGGTGCAGG
GAATCCTGTGGTTGCTGCTGATCACAGGCGTGCAGGGACACCTGGATTGC
AAGCCCGAGTTCAGCTACGCCATTGCCAAGGACGAGAGAATCGGACAGCT
GGGAGCCGAGGGCCTGACCACAACATGGAAAGAATACTCCCCAGGCATGA
AGCTTGAGGACACCATGGTTATCGCTTGGTGCGAGGACGGCAAGCTGATG
TATCTGCAGCGGTGCACCAGAGAGACAAGATACCTGGCCATCCTGCACAC
AAGAGCCCTGCCTACCAGCGTGGTGTTCAAGAAGCTGTTCGACGGCCGGA

AGCAAGAGGACGTGGTGGAAATGAACGACAACTTCGAGTTCGGACTGTGC
CCCTGCGACGCCAAACCTATTGTGCGGGGCAAGTTCAACACCACACTGCT
GAATGGCCCAGCCTTCCAGATGGTCTGCCCTATCGGATGGACCGGCACCG
TGTCCTGCACCAGCTTCAACATGGATACCCTGGCCACCACCGTCGTGCGG
ACCTACAGAAGAAGCAAGCCCTTTCCACACCGGCAGGGCTGCATCACCCA
GAAGAATCTGGGCGAAGATCTGCACAACTGTATCCTCGGCGGCAACTGGA
CATGCGTGCCAGGCGATCAGCTGCTGTACAAAGGCGGCAGCATCGAGAGC
TGCAAGTGGTGCGGCTACCAGTTCAAAGAGTCTGAGGGCCTGCCTCACTA
CCCCATCGGCAAGTGCAAGCTGGAAAACGAGACAGGCTACCGGCTGGTGG
ATAGCACCTCCTGCAATAGAGAAGGCGTGGCCATCGTGCCTCAGGGCACC
CTGAAATGCAAGATCGGCAAGACCACCGTGCAAGTGATCGCCATGGACAC
CAAGCTGGGCCCCATGCCTTGTCGGCCTTACGAGATCATCTCTAGCGAGG
GCCCCGTGGAAAAGACCGCCTGCACCTTCAACTACACAAAGACCCTGAAG
AACAAGTACTTCGAGCCCCGGGACAGCTACTTCCAGCAGTACATGCTGAA
GGGCGAGTACCAGTATTGGTTCGACCTGGAAGTGACCGACCACCACAGAG
ACTACTTCGCCGAGAGCATCCTGGTGGTGGTTGTGGCTTTGCTCGGCGGA
AGATACGTGCTGTGGCTGCTGGTCACCTACATGGTGCTGTCTGAGCAGAA
GGCCCTGGGCATT

B.2 BDV_E1-E2

ATGCAGAGCCCCTACTGCAACGTGACCCGGAAGATCGGCTACGTGTGGTA
CACCAACAACTGCACCCCTGCCTGTCTGCCCAAGAACACCAAGATCATCG
GCCCTGGCAAGTTCGACACCAACGCCGAGGATGGCAAGATCCTGCACGAG
ATGAGAGGCCACATCAGCGAGTTCATCCTGCTGAGCCTGGTGGTGCTGAG
CGACTTTGCCCCTGAGACAGCCAGCACACTGTACCTGGTGCTGCACTTTG
CCCTGCCTCAGACACACGAGGTGCCCAGCGTGTGCGATACCAACCAGCTG
AACCTGACCGTGTCTCTGAGAGTGGACGACGTGATCCCTAGCAGCGTGTG
GAACCTGGGCAAATACGTGTGCGTGCGGCCTGACTGGTGGCCCTACGAAA
CAACAATGGTGCTGCTGTTCGAAGAGGCCGGCCAGGTTGTGAAGCTGGTC
CTGAGAGCCATCCGGGACCTGACCAGAGTGTGGAACAGCGCCAGCACCAC
AGCCTTTCTGATCTGCCTGGTCAAGGTGCTGAGAGGACAGGTGGTGCAGG
GACTTGTGTGGCTGTTGCTTGTGACAGGCGCCCAGGGACAGTTCGCCTGC
AGAGAGGATTACAGATACGCCCTGGCCAGAACAAAAGAGATCGGAGCCCT
GGGAGCCGAGAGCCTGACAACCACATGGACCGACTACCGGGGCAACCTGG
AACTGGATGATGGAACAGTGCGGGCTACCTGCAGCCGGGGCTTCTTTAGA
TTCAGAGGCCACTGCATGATCGGCCCCAGATACCTGGCTAGCCTGCACCT
TAGAGCCCTGCCTACCAGCGTGACCTTCGAGCTTATTCCTGGCGGCAGCG
CCATGACCGAGGAAGAAATGGGCGACGACTTCGAGTTCGGACTGTGCCCC
TGTGATAGCAGACCTGTGGTCAAGGGCAAGTACAACACCACACTGCTGAA
CGGCTCCGCCTTCCAGCTGATCTGTCCTTATGGCTGGGTCGGAAGAGTGG
AATGTACCACCGTGTCCAAGAGCACCCTGGCCACAGAGGTGGTCAAGATC
TACAAAAAGACGAAGCCCTTTCCACAGAGAGTGGGCTGCGATCACACCAC
AGTGTACAAGCAGGACCTGTACCACTGCCAGATGGGCGGCAATTGGACCT
GCATGAGAGGCGAGGTCGTGAAATACGTTGGCGGCCCTGTGAAGAAATGC
GAGTGGTGCGGCTATGTGTTCAAGAAGAGAGAGGGCCTGCCTCACTACCC
CATCGGCAGATGCATGCTGAGAAACGAGACAGGCTACAGAAGCGTGGACG
ACACCCCTTGTGATAGAGGCGGCGTGGTCATCAGCAAGACCGGCGAACTG
GAATGCCTGATCGGCAAGACCACCGTGAAGGTGTTCAGCAGCGACAAGAA
ACTGGGCCCCATGCCTTGCAGACCCAAAGAAGTGATCAGCTCTGAGGGCC

CCGTGTCTAAGATCGCCTGTACCTTCAACTACAGCAAGACGCTGGAAAAC
AAGTACTACGAGCCCCGGGACAGCTACTTCCAGCAGTACATGCTGAAGGG
CCAGTACCAGTATTGGTTCGACCTGGAAGCCACCGACCACCACAGCGATT
ACTTCGCCGAGTTTATCATGCTGGCCGTGGTCGCCCTTCTCGGCGGAAGA
TATGTTCTGTGGCTGATGGTGGTGTACATGATCCTGGCCGACCAGATGAC
AAGCGCC

B.3 CSFV_E1-E2

ATGCTGAGCCCCTACTGCAACGTGACCAGCAAGATCGGCTACATCTGGTA
CACCAACAACTGCACCCCTGCCTGTCTGCCCAAGAACACCAAGATCATCG
GCCCTGGCAAGTTCGACACCAACGCCGAGGATGGCAAGATCCTGCACGAG
ATGGGAGGCCACCTGAGCGAGTTTCTGCTGCTGTCTCTGGTGGTGCTGAG
CGACTTTGCCCCTGAGACAGCCAGCGCTCTGTACCTGATTCTGCACTACA
TGATCCCTCAGAGCCACGAGGAACCCGAGGGCTGCGATACCAACCAGCTG
AACCTGACCGTGGAACTGAGAACCGAGGACGTGATCCCTAGCAGCGTGTG
GAACGTGGGCAAATACGTGTGCGTGCGGCCTGACTGGTGGCCCTATGAGA
CAAAAGTGGCCCTGCTGTTCGAAGAGGCCGGCCAGGTTGTGAAACTGGCC
CTGAGAGCACTGCGGGACCTGACCAGAGTTTGGAACAGCGCCAGCACCAC
AGCCTTTCTGATCTGCCTGATCAAGGTGCTGCGGGGCCAGATTGTGCAGG
GCGTTATCTGGCTGCTGCTGGTTACAGGTGCTCAGGGCAGACTGGCCTGC
AAAGAGGACTACAGATACGCCATCAGCAGCACCAACGAGATCGGACTGCT
GGGAGCCGAGGGCCTGACCACAACATGGAAAGAGTACAACCACGACCTGC
AGCTGAACGACGGCACCGTGAAGGCTATTTGTGTGGCCGGCAGCTTCAAA
GTGATCGCCCTGAACGTGGTGTCCAGAAGATACCTGGCCAGCCTGCACAA
AGAAGCCAGCCTGACCAGCGTGACCTTCGAGCTGCTGTTTGACGGCACAA
ACCCCAGCACCGAGGAAATGGGCGACGACTTCGGCTTTGGACTGTGCCCC
TTCGACACAAGCCCTGTGGTCAAGGGCAAGTACAACACCACACTGCTGAA
CGGCAGCGCCTTCTACCTCGTGTGTCCTATCGGATGGACCGGCGTGATCG
AGTGTACAGCCGTGTCTCCCACCACTCTGAGAACAGAGGTGGTCAAGACC
TTCCGGCGGGACAAGCCCTTTCCACACAGAATGGATTGCGCCACCACCAC
CGTGGAAAACGGCGACCTGTTCTACTGCAAGCTCGGCGGCAATTGGACCT
GCGTGAAAGGCGAGCCTGTGGTGTATACAGGCGGCCTGGTTAAGCAGTGT
CGTTGGTGTGGCTTCGACTTCAACGAGCCCGATGGCCTGCCTCACTACCC
TATCGGCAAGTGCATCCTGGTCAACGAGACAGGCTACCGGATCGTGGACA
GCACCGACTGCAATAGAGATGGCGTGGTCATCTCCACCGACGGCTCTCAC
GAGTGCCTGATCGGCAACACCACTGTGAAGGTGCACGCCAGCGACGAGAG
ACTGGGACCTATGCCTTGCAGACCCAAAGAAATCGTGTCCAGCGCCGGAC

CTGTGCGCAAGACCAGCTGCACCTTTAACTACGCCAAGACGCTGAAGAAC
AAGTACTACGAGCCCCGGGACAGCTACTTCCAGCAGTACATGCTGAAGGG
CGAGTACCAGTATTGGTTCGACCTGGACGTGACCGACCGGCACAGCGATT
ACTTCGCCGAGTTTGTGGTGCTCGTGGTGGTTGCCCTTCTCGGCGGAAGA
TATGTGCTGTGGCTGATCGTGACCTACATCGTGCTGACCGAACAGCTGGC
TGCAGGA

B.4 NRPV_E1-E2

TACTGCGAGAAGTTCCACGTGCTGGGCGACCTGGTGTACGTGAACAGCTG
TCTGCCTATGGGCCTGCCTACAGGCGCCAGATTCGTGTCCAAGAACGTGA
TCAGCCTGGAACCTGAGAAAACGGCCCAGATCATCCCCAGACTGACCCAC
CACCTGGATAGCGGAATCCTGCTGGTGCTGGTGGCCATGAGCGACTTCAT
GCCTGAGACAAGCAGCGCCCTGTACCTGATCCTGCACTTCATGATCCCCA
ACAGCCGGCACCGGACCATCTCTGAAGAGGGACTGACAATGGCCCTGAAC
CTGACCAGCACCGAGCCTGTGTCCAGCGTGATCCCTACAAGCGTGTACGT
GGAAGGCCAGTGGACCTGCTGGAAGCCTTCTTGGTGGCCCTACAACGCCG
ATATCGCCCTGTTTTTCGAGGGCGCCTTCGAGATGCTGGAACTGATTGCT
AGAGCCGTGGGCGATCTGATGAAAGTGTGGACCGAAGCTACCGCCGTGGC
CTTCCTGTGCTTTCTGATCAAGGCCTTCCGGGGCCAGATTCTGCAGGGCG
TTATCCTGCTGCTCCTGCTGTCTAGCGCCGAGGGCAGATACAATCAAGTG
AAGGTGGACAGACCCGACTGGCACACCCTGCTGCAGAAAGATCTGAAGGG
CGTGCTGAGCGGCAAGGACGGCCTGTACATCCTGCGGAGCAACAAAGTTT
GGACCGGCGGCTCCGTGATCATCACCGATGAGTTTGCCGTGACCACCTTC
ATCGGCGACCACACCGGCAACTTCAAGTTCAGCGTGAAAGTGATGACCAC
GCCAATCGAGATGGACTACTGCATCAAAGTGATCGATACCGCCAAGTTCT
TCTGCGTGATGGTCGGAACCCCTACACAGCGGGATCTCGTGAAGCCTCCT
GAGATGCTGTGTGGCTGTGGCGCTCTGGAAGTCCAGGACAACAATAGCAC
CGGCCTGATCAGCCCCGGCAATGTGCTGCCTAGCAAGTGCATCAATGGCT
GGACAGGCGTCGTGACCTGTCACTGCCCTTACACCGATATCAAGATGAAG
TTCCTGGAAAACACGACGCCCCAGAAGTACAGCAAGAACTGCCCTGGCAC
CTACCTGAGCGACCAGAACTTCCACCACGACTGCAAATACGGCAGCCAAG
AGAGCTGCATCGACCCCGAGCCTACAAAGCTGCCTCCAGAGACATACGAG
GATATCCAAGAGTGCTTCTGGTGCAGCTACTACATCAAGGACGCCAACTT
CACCCCTCACAAGGGCCCTCTCGGATGGTGTAGAGTGGGCGAGAACGAGC
CCTACTACCTGACCAACAGAAAGTCCTGCGTGCAAGGCGGCGTGCAGATC
GGATCTGGCGAAGTGACATGTCTGATCGGCACCACCAAGATCAAAGTGGG
GAACTTCAACGAGACAGCCATCAGCTTCATGCCCTGCAATCCCATCAAAG

AGGCCAGCAGAGGCCCTCCTAGCAGAACCACCTGTACCTATAAGTACGCC
AAGACGCTGAAGAACAAGATCTACGACGAGAAGGACAGATACTGGGGCCA
GTACATGGTCAAGGGCGAGTACCAGTATTGGTTCGACCTGGAACAGGACG
ACCACGTGACAGGCGGCCTGCTGAAATACCTGCCTCTGATCATGGTGCTG
CTGCTCGGCGGAAAAATGGTGGCCTGGCTGCTGACCGCCTACTATCTGAT
GGAAGTCGTCGAGGCCACC

B.5 APPV_E1+E2

ATGATCATCCCCTACTGCGAGACATACACCGTGACCGGGATGTACGTCCA
CGTGAAGAACTGCCTGCCTCGGGGACTGCCTAAGCACAGCAAGATCATCA
GCCCCACCATGATCTACCTCGGCGAGGGCGATCCCGCTCACAATATCCAG
CACCTGTTCGGCAGCGGAATCGCCAAATGGGTGCTCGTGCTGCTGGGAGT
GCTTGGAGAGTGGTACGGCGAACTGGCCAGCACCATCTACCTGCTGCTGG
AATACGGCAGCGAGTGGCTGGAACACGAGAGCCTGATTACCGAGGGACTG
ATCCCCGGCATCAACATCACCATTGAGCTGCCCGCCAGCCAGACAGTGCC
TGGATGGGTTTGGGTGGCAGGACAGTGGGTTTGCGTGAAGCCTGACTGGT
GGCCCACACAGATCTGGATCGAAACAGTGGTGGCCGAGGCCTGGCACATC
CTGAAGATTCTGGCCAGCGCTCTGGTCAACATCGTGACCGCCTTCGTGAA
CCTGGAACTGGTGTACCTGGTCATCATCCTGGTCAAGATCAGCAAGGGCA
ACCTGATCGGCGCCATCCTGTGGTGTTTGCTGCTGTCTGGCGCCGAGGGC
AGCTGTCACATCAGACAGGACTACTACAACATCCAGCTGGTGGTGGAAGA
GAAAACCGGCGTGGAAAAGCGGAGCATCATGGGCAAGTGGAGCGTGATCA
CCAGAGAGGGCAGAGAACCCAAGCTGATGGAACAGATCAATATCGTGTCC
AACAACAGCCTGAGCGAAACCTACTGCTACAACCGGCTGAACACCAGCAG
CTGGGGTAGACAGCCTGCCAGACAAAGAGGCTGCGGACAGACCGTGCCTT
ACTGGCCTGGCGACAACGTGCTGGAAGAACAGTACTACAGCACCGGCTAC
TGGGTCAACGCTACAGGCGGTTGCCAACTGAGAGAAGGCGTGTGGCTGAG
CAGAAAGGGCAACGTGCAGTGCCAGAGAAACGGCAGCTCTCTGATTCTGC
AGCTGGCCATCAAAGAAGAGAACGACACCATGGAAATCCCTTGCGACCCC
GTGGAAACCGAGTCTATGGGACCTGTGGCTCAGGGCACCTGTGTGTACTC
TTGGGCCTTTGCTCCCAGAGGCTGGTACTACAATCGGAAGGATGGCTACT
GGCTGCAGTACATCAAGAAGAACGACTACCAGTATTGGACCAAGATGCCC
ACCGCCAGCTCTGCCGCCACCATGTATAGACATCTGCTGCCTCTGCTGGT
GGCCTGCCTTATGGGCGGAAGAATCAGCGTGTGGATCGTGGCCATGCTGC
TGAGCCTGCAG

B.6 bovCD46

ACCGGTATGAGAGCCAGCTGCACCCCTCTGAAGGCCCCCCTGAGAAGGCC
TGAGAGACTGGCCAGCAGCGGCAGATTTGCCTGGGTGCTGCTGCTGGCTC
CTCTCCTGCTGCTGCCTACCAGCAGCGACGCCTGTGACGACCCCCCTAGA
TTCGTGTCCATGAAGCCCCAGGGCACCCTGAAGCCCAGCTACTCTCCTGG
CGAGCAGATCGTGTACGAGTGCAGACTGGGCTTCCAGCCCGTGACACCTG
GACAGGTGCTGGCTCTCGTGTGCCAGGACAACAACACCTGGTCCAGCCTG
CAGGAAGGCTGCAAGAAAAGACGGTGCCCCACCCTGGCCGACCCCACAAA
TGGCCAAGTGATCCTCGTGAACGGCAGCACCGCCTTCGGCAGCGAGGTGC
ACTACGTGTGCAACAACGGCTACTACCTGCTGGGCACCAACATCAGCTAC
TGCGAGGTGTCCAGCGGCACCGGAGTGAATTGGAGCGACAACCCCCCTAC
CTGCGAGAAGATCCTGTGCCAGCCCCCACCCGAGATCCAGAACGGCAAGT
ACACCAACAACCACAAGGACGTGTTCGAGTACAACGAGGTCGTGACCTAC
AGCTGCGACCCCAGCAACGGCCCCGATGAGTACTCTCTCGTGGGCGAGAG
CAAGCTGACCTGCATCGGCAATGGCGAGTGGTCCTCCCAGCCCCCTCAGT
GCAAGGTCGTGAAGTGCGTGTACCCCGCCATCGAGCACGGCACCATCGTG
TCTGGCTTCGGCCCCAAGTACTACTACAAGGCCACCGTGGTGCTGAAGTG
CAACGAGGGCTTCAACCTGTACGGCAACAGCGTGGTCGTGTGCGGCGAGA
ACAGCACCTGGGAGCCTGAGCTGCCCAAGTGCATCAAGGGCCACCCTCCC
AGACCTACCGATGCCAGCCCTCCTAATGGCGCCGAAGGACTGGGAGCCGG
CTACATCGTGCTCGTGATTGTGGCTCTGCTGGCCCTGGACTACTGCAGCG
CCTGTACAGCCGTGTTCGCCGACAGAGGCAGAAAGGGCGGTACCTAATGA
CTCGAG

B.7 porCD46

ATGATGGCCTTCTGCGCCCTGAGAAAGGCCCTGCCTTGCAGACCCGAGAA
TCCCTTCAGCAGCCGGTGCTTCGTGGAAATCCTGTGGGTTTCCCTGGCTC
TGGTGTTCCTGCTGCCTATGCCTAGCGACGCCTGTGACGAGCCTCCTAAG
TTCGAGTCTATGCGGCCCCAGTTCCTGAACACCACCTACAGACCTGGCGA
CAGAGTGGAATACGAGTGCAGGCCTGGCTTCCAGCCTATGGTTCCTGCTC
TGCCTACCTTCAGCGTGTGCCAGGACGACAACACTTGGAGCCCTCTGCAA
GAGGCCTGCAGAAGAAAGGCCTGCTCCAACCTGCCTGATCCTCTGAACGG
CCAGGTGTCATACCCCAACGGCGATATGCTGTTCGGCAGCAAGGCCCAGT
TCACCTGTAACACCGGCTTCTACATCATCGGAGCCGAGACAGTGTACTGT
CAGGTGTCCGGCAACGTGATGGCTTGGAGCGAACCTTCTCCACTGTGCGA
GAAGATCCTGTGCAAGCCTCCTGGCGAGATCCCCAATGGCAAGTACACCA
ACAGCCACAAGGACGTGTTCGAGTACAACGAGGTGGTCACCTACAGCTGC
CTGAGCAGCACAGGCCCTGATGAGTTTAGCCTCGTGGGCGAGAGCAGCCT
GTTCTGCATCGGAAAGGACGAGTGGTCCAGCGATCCTCCTGAGTGCAAGG
TGGTCAAGTGCCCCTATCCTGTGGTGCCTAACGGCGAGATCGTGTCCGGC
TTTGGCAGCAAGTTCTACTACAAGGCCGAGGTCGTGTTCAAGTGCAACGC
CGGCTTTACCCTGCACGGCAGAGATACCATTGTGTGCGGCGCCAATAGCA
CCTGGGAGCCTGAAATGCCCCAGTGCATCAAGGACAGCAAGCCCACCGAT
CCTCCAGCCACACCTGGACCTTCTCATCCCGGACCTCCATCTCCATCCGA
TGCCTCTCCACCTAAGGATGCCGAGTCTCTGGACGGCGGAATCATTGCCG
CTATCGTTGTGGGAGTGCTGGCCGCCATTGCTGTGATTGCTGGCGGCGTG
TACTTCTTCCACCACAAGTACAACAAGAAGCGGAGCAAG

B.8 ADAM17-Ecto

ATGAGACAGTGCGCCCTGTTCCTGACAAGCCTGGTGCCTATTGTGCTGGC
CCCTAGACCTCCTGATGAGCCTGGCTTTGGCAGCCCTCAGAGACTGGAAA
AGCTGGACAGCCTGCTGAGCGACTACGACATCCTGAGCCTGAGCAGCATC
AGACAGCACAGCGTGCGGAAGAGAGATCTGCAGGCCTCCACACACCTGGA
AACCCTGCTGACCTTCAGCGCCCTGAACCGGCACTTCAAGCTGTACCTGA
CCAGCAGCACCGAGCGGTTCAGCCAGAACTTCAAAGTGGTGGTGGTCGAC
GGCGAGGACGAGAGCGAGTATCCTGTGAAGTGGCAGGACTTCTTCAGCGG
CCACGTTGTGGGCGAGCCTGATTCTAGAGTGCTGGCTCACATCGGCGACG
ACGATATCACCGTGCGGATCAATACTGACGGCGCCGAGTACAACATCGAG
CCTCTGTGGCGGCTGATCAACGACACCAAGGACAAGCGCGTGCTGGTGTA
CAAGAGCGAGGACATCAAGAACGTGTCCCGGCTGCAGAGCCCTAAAGTGT
GCGGCTACATCAAGGCCGACAACGAGGAACTGCTGCCCAAAGGCCTGGTG
GACAGAGAGCCTCCTGACGAACTGGTGCACAGAGTGAAGCGGAGAGCCGA
TCCTAATCCTCTGCGGAATACCTGCAAGCTGCTGGTGGTGGCCGATCACA
GATTCTACAAGTACATGGGCAGAGGCGAGGAAAGCACCACCACAAACTAC
CTGATCGAGCTGATCGACAGAGTGGACGACATCTACCGGAACACCAGCTG
GGACAACGCCGGCTTTAAAGGCTACGGCATCCAGATCGAGCAGATCCGCA
TCCTGAAGTCCCCACAAGAAGTGAAGCCCGGCGAGCGGCACTACAACATG
GCCAAGAGCTACCCCAACGAAGAGAAGGACGCCTGGGACGTGAAGATGCT
GCTGGAACAGTTCAGCTTCGATATCGCCGAAGAGGCCAGCAAAGTGTGCC
TGGCTCACCTGTTCACCTACCAAGACTTCGACATGGGCACACTGGGCCTC
GCCTATGTGGGAAGCCCTAGAGCCAATTCTCACGGCGGCGTGTGTCCTAA
GGCCTACTACTCTCCCATCGGCAAGAAGAACATCTACCTGAATAGCGGCC
TGACCTCCACCAAGAACTACGGCAAGACCATCCTGACCAAAGAGGCCGAC
CTGGTCACCACACACGAGCTGGGCCATAATTTTGGCGCCGAGCACGATCC
TGATGGACTGGCCGAATGTGCCCCTAACGAAGATCAAGGCGGCAAATACG
TGATGTACCCTATCGCCGTGTCCGGCGACCACGAGAACAACAAGAGCATC
TACAAGACGATCGAGAGCAAGGCCCAAGAGTGCTTCCAAGAGCGGAGCAA
CAAAGTCTGCGGCAACAGCAGAGTGGATGAGGGCGAAGAGTGCGACCCCG

GCATCATGTACCTGAACAACGATACCTGCTGCAACAGCGACTGCACCCTC
AGACCTGGCGTGCAGTGCAGCGACAGAAATAGCCCTTGCTGCAAGAACTG
CCAGTTCGAGACAGCCCAGAAGAAGTGCCAAGAGGCCATCAACGCCACAT
GCAAGGGCGTGTCCTACTGCACCGGCAACTCTAGCGAATGTCCTCCACCT
GGCAACGCCGAGGATGATACCGTGTGTCTGGATCTGGGCAGATGCAAGGA
CGGGAAGTGCGTGCCCTTTTGCGAGAGAGAGCAGCGGCTGGAAAGCTGCG
CCTGCAACGAGACAGACCACAGCTGCAAAGTCTGTTGCAGGGCCCCTAGC
GGCAGATGCCTGCCTTATGTGGATGCCGAGCAGAAGAATCTGTTCCTGCG
GAAGGGCAAGCCCTGCACCGTGGGCTTTTGCGACATGAACGGCAAGTGCG
AGAAGCGAGTGCAGGACGTGATCGAGAGATTCTGGGAGTTCATCGACAAG
CTGAGCATCAATACCTTCGGCAAGTTCCTGGCCGACAAT

B.9 HCV_E1-Ecto

ACCGGTTATCAGGTCCGAAACTCAACTGGGCTGTACCATGTCACTAACGA
TTGTCCAAACTCCTCCATTGTCTATGAAGCCGCCGACGCTATTCTGCACA
CCCCAGGATGCGTGCCTTGCGTGCGGGAGGGAAACGCCAGCAGATGCTGG
GTGGCCATGACCCCAACAGTGGCAACACGGGACGGCAAGCTGCCAGCAAC
CCAGCTGCGGAGACACATCGATCTGCTGGTGGGAAGCGCCACACTGTGCT
CCGCCCTGTACGTGGGCGACCTGTGCGGCAGCGTGTTCCTGGTGGGCCAG
CTGTTCACCTTTAGCCCAAGGCGCCACTGGACCACACAGGGATGCAACTG
TTCCATCTATCCTGGCCACATCACAGGCCACAGGATGGCCTGGGATATGA
TGATGAATTGGTCTCCTACCACCGCCCTGGTCATGGCACAGCTGCTGCGC
ATCCCACAGGCCATCCTGGACATGATCGCAGGCGCCCACGGAGGAGGCAC
CGGCACAGATACCCTGCAGGCAGAGACCGACCAGCTGGAGGATGAGAAGT
CCGCCCTGCAGACAGAAATCGCCAACCTGCTGAAAGAAAAAGAGAAGCTG
GAGTTTATCGAGCTC

B.10 HCV_E2-Ecto

AGCGAAACTCATGTGACTGGGGGCTCCGCTGGGCATACCGTCTCTGGCTT
CGTCTCCCTGCTGGCTCCTGGGGCAAAACAGAATGTGCAGCTGATCAACA
CCAATGGCTCTTGGCACCTGAACAGCACAGCCCTGAACTGCAATGACAGC
CTGAATACCGGCTGGCTGGCCGGACTGTTCTACCACCACAAGTTTAACAG
CTCCGGCTGCCCCGAGCGGCTGGCCTCCTGTAGACCTCTGACAGACTTCG
ATCAGGGCTGGGGCCCCATCTCCTATGCCAATGGCTCTGGCCCAGACCAG
AGGCCCTACTGCTGGCACTATCCCCCTAAGCCCTGTGGCATCGTGCCTGC
AAAGTCCGTGTGCGGCCCCGTGTACTGTTTCACCCCCTCTCCTGTGGTGG
TGGGCACCACAGATCGGAGCGGCGCCCCAACATATTCCTGGGGCGAGAAC
GACACCGACGTGTTCGTGCTGAACAATACAAGACCACCCCTGGGCAACTG
GTTCGGCTGTACCTGGATGAATTCTACAGGCTTTACCAAGGTGTGCGGCG
CCCCTCCATGCGTGATCGGAGGAGCCGGCAACAATACACTGCACTGCCCA
ACCGACTGTTTCAGAAAGCACCCCGATGCCACCTACAGCAGATGCGGCTC
CGGCCCTTGGATCACACCAAGGTGTCTGGTGGACTACCCATATCGCCTGT
GGCACTACCCCTGCACAATCAACTATACCATCTTTAAGATCAGGATGTAT
GTGGGAGGAGTGGAGCACCGCCTGGAGGCAGCATGCAATTGGACCAGGGG
CGAGCGGTGCGATCTGGAGGACAGGGATAGATCCGAGCTGTCTCCTCTGC
TGCTGACCACAACCCAGTGGCAGGTGCTGCCTTGTTCTTTTACAACCCTG
CCAGCCCTGAGCACAGGCCTGATCCACCTGCACCAGAACATCGTGGATGT
GCAGTACCTGTATGGAGTGGGCTCTAGCATCGCCTCCTGGGCCATCAAGT
GGGAGGGAGGAGGCACCGGCATCGCAAGGCTGGAGGAGAAGGTGAAGACA
CTGAAGGCCCAGAATAGCGAACTGGCCTCAACTGCTAACATGCTGCGGGA
ACAGGTGGCACAGCTGAAACAGAAGTCCGGA

B.11 HCV_E1-Link-E2

TACCAAGTGCGGAATAGCACCGGCCTGTACCACGTGACCAACGACTGCCC
CAATAGCAGCATCGTGTACGAAGCCGCCGATGCCATCCTGCACACACCTG
GATGTGTGCCCTGTGTCAGAGAGGGCAACGCCAGCAGATGTTGGGTCGCC
ATGACACCTACCGTGGCCACCAGAGATGGAAAGCTGCCTGCTACACAGCT
GCGGAGACACATCGATCTGCTCGTGGGAAGCGCCACACTGTGTAGCGCTC
TGTATGTGGGCGATCTGTGCGGCTCCGTGTTTCTCGTGGGCCAGCTGTTC
ACATTCAGCCCCAGAAGGCACTGGACCACACAGGGCTGCAACTGCAGCAT
CTACCCTGGCCACATCACCGGCCATAGAATGGCCTGGGACATGATGATGA
ATTGGAGCCCCACCACCGCTCTGGTCATGGCTCAGCTGCTGAGAATCCCT
CAGGCCATCCTGGACATGATTGCCGGCGCTCATCTGGAAGTGCTGTTTCA
GGGACCTGCTAGCGGAGGCTCTGGTGGATCTGGCGGAAGTGGTGGAAGTG
GCGGTAGCGGAGGAAGCGGCGGTTCAGGCGGATCCGGTGGAAGCGGAGGA
TCTCTTGAGGTCCTGTTCCAGGGGCCTAGCGAGACACATGTGACTGGCGG
ATCTGCCGGCCACACAGTGTCTGGCTTTGTGTCTCTTCTGGCTCCTGGCG
CCAAGCAGAACGTGCAGCTGATCAACACCAACGGCAGCTGGCACCTGAAC
AGCACAGCCCTGAACTGCAACGACAGCCTGAATACCGGATGGCTGGCCGG
CCTGTTCTACCACCACAAGTTTAACAGCAGCGGCTGCCCCGAGAGACTGG
CCTCTTGTAGACCTCTGACCGACTTCGATCAAGGCTGGGGCCCTATCAGC
TACGCCAATGGCTCTGGACCTGACCAGAGGCCTTACTGCTGGCACTACCC
TCCAAAGCCTTGCGGAATCGTGCCTGCCAAGTCTGTGTGCGGCCCTGTGT
ACTGCTTCACCCCATCTCCAGTGGTCGTGGGCACCACCGATAGATCTGGC
GCCCCTACCTATAGCTGGGGCGAGAACGATACCGACGTGTTCGTGCTGAA
CAACACCCGGCCTCCACTCGGCAATTGGTTCGGCTGTACCTGGATGAACT
CCACCGGCTTCACCAAAGTGTGCGGAGCCCCTCCTTGTGTGATTGGCGGA
GCCGGAAACAATACCCTGCACTGCCCTACCGACTGCTTCAGAAAGCACCC
CGACGCCACCTACTCCAGATGTGGTTCTGGCCCTTGGATCACCCCTAGAT
GCCTGGTGGACTACCCCTACCGGCTGTGGCACTATCCCTGCACCATCAAC
TACACCATCTTCAAGATCCGTATGTACGTCGGCGGCGTGGAACACAGACT
GGAAGCCGCCTGTAACTGGACCAGGGGCGAGAGATGCGACCTGGAAGATA

GAGACAGAAGCGAGCTGAGCCCTCTGCTGCTGACCACCACACAATGGCAG
GTCCTGCCTTGCAGCTTCACCACACTGCCTGCTCTGAGCACAGGCCTGAT
CCATCTGCACCAGAACATCGTGGACGTGCAGTACCTGTACGGCGTGGGCA
GCTCTATTGCCAGCTGGGCCATTAAGTGGGAG

C
List of Primers

Name	Sequence	Supplier
BVDV_Fprimer1	GGATCTACCGGTGCCTCTCCTTACTGCGAC	Eurofins
BVDV_Fprimer2	GGATCTACCGGTTGCAAGCCCGAGTTCAGC	Eurofins
BVDV_Fprimer3	GGATCTACCGGTTACTGCGACGTGGACCGG	Eurofins
BVDV_Fprimer4	GGATCTACCGGTCACCTGGATTGCAAGCCC	Eurofins
BVDV_Fprimer5	GGATCTACCGGTGGCAAGATCCTGCACGAG	Eurofins
BVDV_Fprimer11	GGATCTGAGCTCATGTCTATACATGGTGGC	Sigma
BVDV_Fprimer13	CGTAGCTGAAACCGGTGCCTCTCCTTACTGCGACGT	Sigma
BVDV_Rprimer1	GGATCTGGTACCTCCCTGCACGCCTGTGATC	Eurofins
BVDV_Rprimer2	GGATCTGGTACCTCCCTGCACGCCTGTGATC	Eurofins
BVDV_Rprimer3	GGATCTGGTACCGGTCACTTCCAGGTCGAAC	Eurofins
BVDV_Rprimer4	GGATCTGGTACCGCCCAGGGCCTTCTGCTC	Eurofins
BVDV_Rprimer5	GGATCTGGTACCGCTCTCGGCGAAGTAGTC	Sigma
BVDV_Rprimer6	GGATCTGGTACCGTTCCAAATTCTGGTCAG	Eurofins
BVDV_Rprimer11	GGATCTGGTACCCAGCACTTCAGACAGGTG	Sigma
BVDV_Rprimer12	GGATCTGAGCTCGTTCCAAATTCTGGTCAG	Eurofins
BVDV_Rprimer13	GGATCTGAGCTCGCTCTCGGCGAAGTAGTC	Sigma
BVDV_Rprimer15	CCACCGGAACCTCCGGTACCTCCCTGCACGCCTGTGAT	Sigma
BDV_Fprimer1	GGATCTACCGGTCAGAGCCCCTACTGCAAC	Eurofins
BDV_Fprimer2	GGATCTACCGGTTGCAGAGAGGATTACAGATAC	Eurofins
BDV_Fprimer3	GGATCTACCGGTTACTGCAACGTGACCCGG	Eurofins
BDV_Fprimer4	GGATCTACCGGTCAGTTCGCCTGCAGAGAG	Eurofins
BDV_Fprimer5	GGATCTACCGGTGGCAAGATCCTGCACGAG	Eurofins
BDV_Fprimer11	GGATCTTCCGGACAGTTCGCCTGCAGAGAG	Eurofins
BDV_Fprimer16	CGTAGCTGAAACCGGTTGCAGAGAGGATTACAGATACGCC	Sigma
BDV_Rprimer1	GGATCTGGTACCGATGGCTCTCAGGACCAG	Eurofins
BDV_Rprimer2	GGATCTGGTACCTCCCTGGGCGCCTGTCAC	Eurofins
BDV_Rprimer3	GGATCTGGTACCGGTGGCTTCCAGGTCGAAC	Eurofins
BDV_Rprimer4	GGATCTGGTACCGGCGCTTGTCATCTGGTC	Eurofins
BDV_Rprimer5	GGATCTGGTACCCTCGGCGAAGTAATCGCTG	Eurofins
BDV_Rprimer6	GGATCTGGTACCGTTCCACACTCTGGTCAG	Eurofins
BDV_Rprimer11	GGATCTGGTACCGATGAACTCGCTGATGTG	Sigma
BDV_Rprimer12	GGATCTGAGCTCGTTCCACACTCTGGTCAG	Eurofins
BDV_Rprimer13	GGATCTGAGCTCCTCGGCGAAGTAATCGCTG	Sigma
BDV_Rprimer16	CCACCGGAACCTCCGGTACCCTCGGCGAAGTAATCGCTGTGG	Sigma
CSFV_Fprimer1	GGATCTACCGGTCTGAGCCCCTACTGCAAC	Eurofins
CSFV_Fprimer2	GGATCTACCGGTTGCAGAGAGGATTACAGATAC	Eurofins
CSFV_Fprimer3	GGATCTACCGGTTACTGCAACGTGACCAGC	Eurofins
CSFV_Fprimer4	GGATCTACCGGTAGACTGGCCTGCAAAGAG	Eurofins
CSFV_Fprimer5	GGATCTACCGGTGGCAAGATCCTGCACGAG	Eurofins
CSFV_Fprimer11	GGATCTTCCGGAAGACTGGCCTGCAAAGAG	Sigma
CSFV_Fprimer12	GGTTGCGTAGCTGAAACCGGTCTGAGCCCCTACTG	Sigma
CSFV_Fprimer16	GAAGTGGCGTTCCGGAAGACTGGCCTGCAAA	Sigma
CSFVmut_Fprimer1	CAGCTGCACCTTTCAGTACGCCAAGACG	Sigma
CSFV_Rprimer1	GGATCTGGTACCCAGTGCTCTCAGGGCCAG	Eurofins
CSFV_Rprimer2	GGATCTGGTACCGCCCTGAGCACCTGTAAC	Eurofins
CSFV_Rprimer3	GGATCTGGTACCGGTCACGTCCAGGTCGAAC	Eurofins
CSFV_Rprimer4	GGATCTGGTACCTCCTGCAGCCAGCTGTTC	Eurofins
CSFV_Rprimer5	GGATCTGGTACCCTCGGCGAAGTAATCGCTG	Eurofins
CSFV_Rprimer6	GGATCTGGTACCGTTCCAAACTCTGGTCAG	Eurofins
CSFV_Rprimer11	GGATCTGGTACCGATGAACTCGCTGATGTG	Sigma
CSFV_Rprimer12	GGATCTGAGCTCGTTCCAAACTCTGGTCAG	Eurofins
CSFV_Rprimer13	GGATCTGAGCTCGTTCCAAACTCTGGTCAG	Eurofins
CSFV_Rprimer14	TGCCTCCTCCGTTCCAAACTCTGGTCAGGTCCC	Sigma
CSFV_Rprimer17	CCGGACCCTCCGAGCTCGTTCCAAACTCTGG	Sigma
CSFV_Rprimer18	ACCACCGGAACCTCCGGTACCCTCGGCGAAGTAATC	Sigma
CSFVmut_Rprimer1	CGTCTTGGCGTACTGAAAGGTGCAGCTG	Sigma
NRPV_Fprimer1	GTTGCGTAGCTGAAACCGGTTACTGCGAGAAGTTC	Sigma
NRPV_Fprimer4	GGTTGCGTAGCTGAAACCGGTGGCAGATACAATCA	Sigma

NRPV_Fprimer5	GGTTGCGTAGCTGAAACCGGTAAAACGGCCCAGAT	Sigma
NRPV_Fprimer6	TTTTTCGAGGGCGCCGAGCTCGGAGGAGGCACCGG	Sigma
NRPV_Fprimer7	GGCGGCCTGCTGAAAGAGCTCGGAGGAGGCACCGG	Sigma
NRPV_Fprimer8	GCGGAGGGTCTGGCGGATCTGGCGGAAGTGGCGTTCCGGAGGCAGATACAATCA	Sigma
NRPV_Rprimer2	CCACCGGAACCTCCGGTACCCTCGGCGCTAGACAG	Sigma
NRPV_Rprimer4	CCACCGGAACCTCCGGTACCGGCCTCGACGACTTC	Sigma
NRPV_Rprimer5	CCACCGGAACCTCCGGTACCTTTCAGCAGGCCGCC	Sigma
NRPV_Rprimer6	CCACCGGAACCTCCGGTACCGGCGCCCTCGAAAAA	Sigma
NRPV_Rprimer11	CCACCGGAACCTCCGGTACCGATTCCGCTATCCAG	Sigma
NRPV_Rprimer12	CCGGTGCCTCCTCCGAGCTCGGCGCCCTCGAAAAA	Sigma
NRPV_Rprimer13	CCGGTGCCTCCTCCGAGCTCTTTCAGCAGGCCGCC	Sigma
NRPV_Rprimer14	CGCCAGACCCTCCGCTCCCACCGCTACCACCGGACCCTCCGAGCTCGGCGCCCTCGAAAA	Sigma
APPV_Fprimer1	GGATCTACCGGTATCATCCCCTACTGCGAG	Eurofins
APPV_Fprimer2	GGATCTACCGGTATCAGACAGGACTACTAC	Eurofins
APPV_Fprimer3	GGATCTACCGGTTACTGCGAGACATACACC	Eurofins
APPV_Fprimer4	GGATCTACCGGTAGCTGTCACATCAGACAG	Eurofins
APPV_Fprimer5	GGATCTACCGGTGATCCCGCTCACAATATC	Eurofins
APPV_Fprimer11	GGATCTTCCGGAAGCTGTCACATCAGACAG	Eurofins
APPV_Rprimer1	GGATCTGGTACCCCAGATCTGTGTGGGCCAC	Eurofins
APPV_Rprimer2	GGATCTGGTACCGCCCTCGGCGCCAGACAG	Eurofins
APPV_Rprimer3	GGATCTGGTACCGGCGGTGGGCATCTTGGTC	Eurofins
APPV_Rprimer4	GGATCTGGTACCCTGCAGGCTCAGCAGCATGGCCACGATCCAC	Eurofins
APPV_Rprimer5	GGATCTGGTACCATGTCTATACATGGTGGC	Eurofins
APPV_Rprimer6	GGATCTGGTACCGGCCTCGGCCACCACTG	Eurofins
APPV_Rprimer11	GGATCTGGTACCCACCCATTTGGCGATTCC	Sigma
APPV_Rprimer12	GGATCTGAGCTCGGCCTCGGCCACCACTGTTTC	Eurofins
APPV_Rprimer13	GGATCTGAGCTCATGTCTATACATGGTGGC	Sigma
HCV_Fprimer1	GGATCTACCGGTTATCAGGTCCGCAACTCC	Eurofins
HCV_Fprimer3	GGATCTACCGGTAGCGAAACTCATGTCAC	Eurofins
HCV_Fprimer8	CCACCGGAACCTCCGGTACCGCTCCGCTCTTGGAAGCAC	Sigma
HCV_Rprimer3	GGATCTGGTACCGATAAACTCCAGCTTCTC	Eurofins
HCV_Rprimer4	GGATCTGGTACCGTGGGCGCCTGCGATCATG	Eurofins
HCV_Rprimer5	GGATCTGGTACCCTTCTGTTTCAGCTGTGC	Eurofins
HCV_Rprimer7	GGATCTGGTACCCTCCCACTTGATGGCCCAG	Eurofins
HCV_Rprimer13	CCACCGGAACCTCCGGTACCCTCCCACTTAATGGCCCAGC	Sigma
Jun_Fprimer1	GGATCTGAGCTCGGAGGAGGCACCGGCATC	Eurofins
Fos_Fprimer1	GGATCTGAGCTCGGAGGAGGCACCGGCACAG	Eurofins
bCD46_Fprimer1	GGATCTACCGGTTGTGACGACCCCCCTAG	Eurofins
bCD46_Fprimer3	GGATCTACCGGTATGAGAGCCAGCTGCACC	Eurofins
bCD46_Rprimer1	GGATCTGGTACCGATCTTCTCGCAGGTAGG	Eurofins
bCD46_Rprimer2	GGATCTGGTACCCACGACCTTGCACTGAGG	Eurofins
bCD46_Rprimer3	GGATCTGGTACCGCCCTTGATGCACTTGGG	Eurofins
bCD46_Rprimer5	GGATCTGGTACCGGTACCGCCCTTTCTGCC	Eurofins
pCD46_Fprimer3	CCACCGGAACCTCCGGTACCCTCCCACTTAATGGCCCAGC	Sigma
pCD46_Fprimer4	CCACCGGAACCTCCGGTACCCTCCCACTTAATGGCCCAGC	Sigma
pCD46_Rprimer4	CCACCGGAACCTCCGGTACCCTCCCACTTAATGGCCCAGC	Sigma
pCD46_Rprimer5	CCACCGGAACCTCCGGTACCGACCACCTTGCACTCAGGAGG	Sigma
pCD46_Rprimer6	CCACCGGAACCTCCGGTACCGTCCTTGATGCACTGGGGC	Sigma
pADAM17_Fprimer1	CGTAGCTGAAACCGGTATGAGACAGTGCGCCCTG	Sigma
pADAM17_Fprimer2	CGTAGCTGAAACCGGTAATACCTGCAAGCTGCTGGTGG	Sigma
pADAM17_Rprimer1	CCACCGGAACCTCCGGTACCATTGTCGGCCAGGAACTTG	Sigma
pADAM17_Rprimer2	CCACCGGAACCTCCGGTACCGCTCCGCTCTTGGAAGCAC	Sigma
Screening_Fprimer	GGGAGGTCTATATAAGCAGAGC	
Screening_Rprimer	GCAGCTCGACTCAAGAGTC	

D
Domain Boundaries

Pestivirus Species	E1	E2
APPV	501-616	617-943
Bat	502-672	673-951
NRPV	605-772	773-1180
Rodent	635-802	803-2210
Phocoena	339-506	507-917
Bungowannah	498-664	665-1071
Linda	503-672	673-1077
Pronghorn	493-657	658-1058
BVDV1	498-662	663-1066
BVDV2	498-662	663-1064
BVDV3	497-661	662-1064
Giraffe	494-658	659-1062
Border	496-660	661-1063
Reindeer	496-660	661-1063
Aydin	496-660	661-1063
CSFV	495-659	660-1062

E

Construct Details

Contruct	Description	Boundaries	Template Gene	Forward Primer	Reverse Primer	Cloning Style	Cloned	Test Expression Level	Expression Scale bottles/liters	Protein Yield mg/l	Xtals	Solved
BVDV_1	E1 Ecto-1	1-152	BVDV_E1-E2	BVDV_Fprimer1	BVDV_Rprimer1	Restriction	Yes	None	N/A	N/A	N/A	N/A
BVDV_2	Full length E1-1	1-196	BVDV_E1-E2	BVDV_Fprimer1	BVDV_Rprimer2	Restriction	Yes	None	N/A	N/A	N/A	N/A
BVDV_3	E1 Ecto-2	4-152	BVDV_E1-E2	BVDV_Fprimer3	BVDV_Rprimer1	Restriction	Yes	None	N/A	N/A	N/A	N/A
BVDV_4	Full length E1-2	4-196	BVDV_E1-E2	BVDV_Fprimer3	BVDV_Rprimer2	Restriction	Yes	None	N/A	N/A	N/A	N/A
BVDV_5	E2 Ecto-1	4-333	BVDV_E1-E2	BVDV_Fprimer2	BVDV_Rprimer3	Restriction	Yes	High	2/0.5	2.4	No	N/A
BVDV_6	Full length E2	1-374	BVDV_E1-E2	BVDV_Fprimer4	BVDV_Rprimer4	Restriction	Yes	None	N/A	N/A	N/A	N/A
BVDV_7	E1 N-ter Trunc	44-161	BVDV_E1-E2	BVDV_Fprimer5	BVDV_Rprimer6	Restriction	Yes	None	N/A	N/A	N/A	N/A
BVDV_8	E1 Ecto-3	1-161	BVDV_E1-E2	BVDV_Fprimer1	BVDV_Rprimer6	Restriction	Yes	None	N/A	N/A	N/A	N/A
BVDV_9	E2 Ecto-2	1-343	BVDV_E1-E2	BVDV_Fprimer2	BVDV_Rprimer5	Restriction	Yes	High	N/A	N/A	N/A	N/A
BVDV_11	E1+Fos	1-161	BVDV_E1-E2 HCV_E1_Ecto	BVDV_Fprimer13 Fos_Fprimer1	BVDV_Rprimer15 HCV_Rprimer3	Restriction	Yes	None	N/A	N/A	N/A	N/A
BVDV_12	E2+Jun	1-343	BVDV_E1-E2 HCV_E2_Ecto	BVDV_Fprimer4 Jun_Fprimer1	BVDV_Rprimer13 HCV_Rprimer5	Restriction	Yes	None	N/A	N/A	N/A	N/A
BVDV_13	E1-link20-E2	1-161 1-343	BVDV_E1-E2	BVDV_Fprimer1 BVDV_Fprimer11	BVDV_Rpimer12 BVDV_Rprimer5	Restriction	No	N/A	N/A	N/A	N/A	N/A
BVDV_14	E1-link30-E2	1-161 1-343	BVDV_E1-E2	BVDV_Fprimer1 BVDV_Fprimer11	BVDV_Rpimer12 BVDV_Rprimer5	Restriction	No	N/A	N/A	N/A	N/A	N/A
BVDV_15	E1 C-ter Trunc	1-58	BVDV_E1-E2	BVDV_Fprimer1	BVDV_Rprimer11	Restriction	Yes	None	N/A	N/A	N/A	N/A
BVDV_17	BVDV_9 +3C	1-343	BVDV_E1-E2	N/A	N/A	Restriction	Yes	High	6/1.5	28.7	No	N/A
BDV_1	E1 Ecto-1	1-153	BDV_E1-E2	BDV_Fprimer1	BDV_Rprimer1	Restriction	Yes	None	N/A	N/A	N/A	N/A
BDV_2	Full length E1-1	1-195	BDV_E1-E2	BDV_Fprimer1	BDV_Rprimer2	Restriction	Yes	None	N/A	N/A	N/A	N/A
BDV_3	E1 Ecto-2	4-153	BDV_E1-E2	BDV_Fprimer3	BDV_Rprimer1	Restriction	Yes	None	N/A	N/A	N/A	N/A
BDV_4	Full length E1-2	4-195	BDV_E1-E2	BDV_Fprimer4	BDV_Rprimer4	Restriction	Yes	None	N/A	N/A	N/A	N/A
BDV_5	E2 Ecto-1	4-332	BDV_E1-E2	BDV_Fprimer2	BDV_Rprimer3	Restriction	Yes	None	N/A	N/A	N/A	N/A
BDV_6	Full length E2	1-373	BDV_E1-E2	BDV_Fprimer4	BDV_Rprimer4	Restriction	Yes	None	N/A	N/A	N/A	N/A
BDV_7	E1 N-ter Trunc	44-161	BDV_E1-E2	BDV_Fprimer5	BDV_Rprimer6	Restriction	Yes	None	N/A	N/A	N/A	N/A
BDV_8	E1 Ecto-3	1-161	BDV_E1-E2	BDV_Fprimer1	BDV_Rprimer6	Restriction	Yes	None	N/A	N/A	N/A	N/A
BDV_9	E2 Ecto-2	1-341	BDV_E1-E2	BDV_Fprimer16	BDV_Rprimer16	Single Step	Yes	None	N/A	N/A	N/A	N/A
BDV_10	E2 Ecto-3	4-341	BDV_E1-E2	BDV_Fprimer2	BDV_Rprimer5	Restriction	Yes	Low	6/1.5	1	Yes	No
BDV_11	E1+Fos	1-161	BDV_E1-E2 HCV_E1_Ecto	BDV_Fprimer1 Fos_Fprimer1	BDV_Rprimer12 HCV_Rprimer3	Restriction	Yes	None	N/A	N/A	N/A	N/A
BDV_12	E2+Jun	4-341	BDV_E1-E2 HCV_E1_Ecto	BDV_Fprimer4 Jun_Fprimer1	BDV_Rprimer13 HCV_Rprimer5	Restriction	Yes	None	N/A	N/A	N/A	N/A
BDV_13	E1-link20-E2	1-161 4-341	BDV_E1-E2	BDV_Fprimer1 BDV_Fprimer11	BDV_Rprimer12 BDV_Rprimer5	Restriction	No	N/A	N/A	N/A	N/A	N/A
BDV_14	E1-link30-E2	1-161 4-341	BDV_E1-E2	BDV_Fprimer1 BDV_Fprimer11	BDV_Rprimer12 BDV_Rprimer5	Restriction	No	N/A	N/A	N/A	N/A	N/A
BDV_15	E1 C-ter Trunc	1-59	BDV_E1-E2	BDV_Fprimer1	BDV_Rprimer11	Restriction	Yes	None	N/A	N/A	N/A	N/A
BDV_17	BDV_10 +3C	4-341	BDV_E1-E2	N/A	N/A	Restriction	Yes	N/A	N/A	N/A	N/A	N/A
CSFV_1	E1 Ecto-1	1-153	CSFV_E1-E2	CSFV_Fprimer1	CSFV_Rprimer1	Restriction	Yes	None	N/A	N/A	N/A	N/A
CSFV_2	Full length E1-1	1-195	CSFV_E1-E2	CSFV_Fprimer1	CSFV_Rprimer2	Restriction	Yes	None	N/A	N/A	N/A	N/A
CSFV_3	E1 Ecto-2	4-153	CSFV_E1-E2	CSFV_Fprimer3	CSFV_Rprimer1	Restriction	Yes	None	N/A	N/A	N/A	N/A
CSFV_4	Full length E1-2	4-195	CSFV_E1-E2	CSFV_Fprimer3	CSFV_Rprimer2	Restriction	Yes	None	N/A	N/A	N/A	N/A
CSFV_5	E2 Ecto-1	4-332	CSFV_E1-E2	CSFV_Fprimer2	CSFV_Rprimer3	Restriction	Yes	None	N/A	N/A	N/A	N/A
CSFV_6	Full length E2	1-373	CSFV_E1-E2	CSFV_Fprimer4	CSFV_Rprimer4	Restriction	Yes	None	N/A	N/A	N/A	N/A
CSFV_7	E1 N-ter Trunc	44-161	CSFV_E1-E2	CSFV_Fprimer5	CSFV_Rprimer6	Restriction	Yes	None	N/A	N/A	N/A	N/A
CSFV_8	E1 Ecto-3	1-161	CSFV_E1-E2	CSFV_Fprimer1	CSFV_Rprimer6	Restriction	Yes	None	N/A	N/A	N/A	N/A
CSFV_9	E2 Ecto-2	1-341	CSFV_E1-E2	CSFV_Fprimer4	CSFV_Rprimer5	Restriction	Yes	High	6/1.5	29	Yes	No
CSFV_11	E1+Fos	1-161	CSFV_E1-E2 HCV_E1_Ecto	CSFV_Fprimer12 Fos_Fprimer1	CSFV_Rprimer14 HCV_Rprimer3	Restriction	Yes	None	N/A	N/A	N/A	N/A
CSFV_12	E2+Jun	1-341	CSFV_E1-E2 HCV_E2_Ecto	CSFV_Fprimer4 Jun_Fprimer1	CSFV_Rprimer13 HCV_Rprimer5	Restriction	Yes	None	N/A	N/A	N/A	N/A
CSFV_13-1	E1-link20-E2	1-161 1-341	CSFV_E1-E2	CSFV_Fprimer1 CSFV_Fprimer11	CSFV_Rprimer12 CSFV_Rprimer5	Restriction	No	N/A	N/A	N/A	N/A	N/A

CSFV_13-2	E1-link20-E2	1-161 1-341	CSFV_E1-E2	CSFV_Fprimer12 CSFV_Fprimer16	CSFV_Rprimer17 CSFV_Rprimer18	Single Step	No	N/A	N/A	N/A	N/A	N/A
CSFV_14	E1-link30-E2	1-161 1-341	CSFV_E1-E2	CSFV_Fprimer1 CSFV_Fprimer11	CSFV_Rprimer12 CSFV_Rprimer5	Restriction	No	N/A	N/A	N/A	N/A	N/A
CSFV_15	E1 C-ter Trunc	1-58	CSFV_E1-E2	CSFV_Fprimer1 CSFVmut_Fprimer1	CSFV_Rprimer11 CSFV_Rprimer5	Restriction	Yes	None	N/A	N/A	N/A	N/A
CSFV_16	CSFV_9 N292Q	1-341	CSFV_E1-E2	CSFV_Fprimer4 CSFV_Fprimer4	CSFVmut_Rprimer1 CSFV_Rprimer5	Restriction	Yes	High	6/1.5	21.1	No	N/A
CSFV_17	CSFV_9 +3C	1-341	CSFV_E1-E2	N/A	N/A	Restriction	Yes	High	6/1.5	29.4	No	N/A
NRPV_2	Full length E1-1	1-194	NRPV_E1-E2	NRPV_Fprimer1	NRPV_Rprimer2	Single Step	Yes	None	N/A	N/A	N/A	N/A
NRPV_6	Full length E2	1-379	NRPV_E1-E2	NRPV_Fprimer4	NRPV_Rprimer4	Single Step	Yes	None	N/A	N/A	N/A	N/A
NRPV_7	E1 N-ter Trunc	40-142	NRPV_E1-E2	NRPV_Fprimer5	NRPV_Rprimer6	Single Step	Yes	None	N/A	N/A	N/A	N/A
NRPV_8	E1 Ecto-3	1-142	NRPV_E1-E2	NRPV_Fprimer1	NRPV_Rprimer6	Single Step	No	None	N/A	N/A	N/A	N/A
NRPV_9	E2 Ecto-2	1-348	NRPV_E1-E2	NRPV_Fprimer4	NRPV_Rprimer5	Single Step	Yes	Medium	6/1.6	8.6	Yes	Yes
NRPV_15	E1 C-ter Trunc	1-56	NRPV_E1-E2	NRPV_Fprimer1	NRPV_Rprimer11	Single Step	No	N/A	N/A	N/A	N/A	N/A
APPV_1	E1 Ecto-1	1-138	APPV_E1-E2	APPV_Fprimer1	APPV_Rprimer1	Restriction	Yes	None	N/A	N/A	N/A	N/A
APPV_2	Full length E1-1	1-199	APPV_E1-E2	APPV_Fprimer1	APPV_Rprimer2	Restriction	Yes	None	N/A	N/A	N/A	N/A
APPV_3	E1 Ecto-2	4-138	APPV_E1-E2	APPV_Fprimer3	APPV_Rprimer1	Restriction	Yes	None	N/A	N/A	N/A	N/A
APPV_4	Full length E2-2	4-199	APPV_E1-E2	APPV_Fprimer3	APPV_Rprimer2	Restriction	Yes	None	N/A	N/A	N/A	N/A
APPV_5	E2 Ecto-1	4-202	APPV_E1-E2	APPV_Fprimer2	APPV_Rprimer3	Restriction	Yes	None	N/A	N/A	N/A	N/A
APPV_6	Full length E2	1-237	APPV_E1-E2	APPV_Fprimer4	APPV_Rprimer4	Restriction	Yes	None	N/A	N/A	N/A	N/A
APPV_7	E1 N-ter Trunc	43-146	APPV_E1-E2	APPV_Fprimer5	APPV_Rprimer6	Restriction	Yes	None	N/A	N/A	N/A	N/A
APPV_8	E1 Ecto-3	1-146	APPV_E1-E2	APPV_Fprimer1	APPV_Rprimer6	Restriction	Yes	None	N/A	N/A	N/A	N/A
APPV_9	E2 Ecto-2	1-211	APPV_E1-E2	APPV_Fprimer4	APPV_Rprimer5	Restriction	Yes	High	6/1.5	4.3	Yes	Yes
APPV_11	E1+Fos	1-146	APPV_E1-E2 HCV_E1_Ecto	APPV_Fprimer1 Fos_Fprimer1	APPV_Rprimer12 HCV_Rprimer3	Restriction	Yes	None	N/A	N/A	N/A	N/A
APPV_12	E2+Jun	1-211	APPV_E1-E2 HCV_E2_Ecto	APPV_Fprimer4 Jun_Fprimer1	APPV_Rprimer13 HCV_Rprimer5	Restriction	Yes	None	N/A	N/A	N/A	N/A
APPV_13	E1-link20-E2	1-146 1-211	APPV_E1-E2	APPV_Fprimer1 APPV_Fprimer11	APPV_Rprimer12 APPV_Rprimer5	Restriction	No	N/A	N/A	N/A	N/A	N/A
APPV_14	E1-link30-E2	1-146 1-211	APPV_E1-E2	APPV_Fprimer1 APPV_Fprimer11	APPV_Rprimer12 APPV_Rprimer5	Restriction	No	N/A	N/A	N/A	N/A	N/A
APPV_15	E1 C-ter Trunc	1-60	APPV_E1-E2	APPV_Fprimer1	APPV_Rprimer11	Restriction	Yes	None	N/A	N/A	N/A	N/A
APPV_17	APPV_9 +3C	1-211	APPV_E1-E2	N/A	N/A	Restriction	Yes	N/A	N/A	N/A	N/A	N/A
HCV_3	E1 Ecto +Fos	1-161	HCV_E1_Ecto	HCV_Fprimer1	HCV_Rprimer3	Restriction	Yes	Low	6/1.5	0.1	No	N/A
HCV_4	E1 Ecto	1-161	HCV_E1_Ecto	HCV_Fprimer1	HCV_Rprimer4	Restriction	Yes	Low	N/A	N/A	N/A	N/A
HCV_7	E2 Ecto +Jun	1-335	HCV_E2_Ecto	HCV_Fprimer3	HCV_Rprimer5	Restriction	Yes	Medium	6/1.5	1.5	No	N/A
HCV_8	E2 Ecto	1-335	HCV_E2_Ecto	HCV_Fprimer3	HCV_Rprimer7	Restriction	Yes	Medium	N/A	N/A	N/A	N/A
HCV_10	E1-link30-E2	1-161,1-161	HCV_E1-link-E2	HCV_Fprimer8	HCV_Rprimer13	Single Step	Yes	None	N/A	N/A	N/A	N/A
bCD46_1	Full length	1-344	bovCD46	bCD46_Fprimer3	bCD46_Rprimer5	Restriction	Yes	N/A	N/A	N/A	N/A	N/A
bCD46_2	SCR1-SCR2	43-169	bovCD46	bCD46_Fprimer1	bCD46_Rprimer1	Restriction	Yes	High	2/0.5	10.3	Yes	Yes
bCD46_3	SCR1-SCR3	43-253	bovCD46	bCD46_Fprimer1	bCD46_Rprimer2	Restriction	No	N/A	N/A	N/A	N/A	N/A
bCD46_4	SCR1-SCR4	43-295	bovCD46	bCD46_Fprimer1	bCD46_Rprimer3	Restriction	Yes	High	6/1.5	24.4	Yes	Yes
bCD46_5	Entire Ecto	1-295	bovCD46	bCD46_Fprimer3	bCD46_Rprimer3	Restriction	Yes	High	N/A	N/A	N/A	N/A
bCD46_6	bCD46_4 +3C	43-295	bovCD46	N/A	N/A	Restriction	Yes	High	N/A	N/A	N/A	N/A
bCD46_7	bCD46_5 +3C	1-295	bovCD46	N/A	N/A	Restriction	Yes	High	N/A	N/A	N/A	N/A
pCD46_1	SCR1-SCR2	44-167	PorCD46	pCD46_Fprimer3	pCD46_Rprimer4	Single Step	Yes	High	6/1.5	3.1	No	N/A
pCD46_2	SCR1-SCR3	44-233	PorCD46	pCD46_Fprimer3	pCD46_Rprimer5	Single Step	Yes	Medium	2/0.5	1	No	N/A
pCD46_3	SCR1-SCR4	44-299	PorCD46	pCD46_Fprimer3	pCD46_Rprimer6	Single Step	Yes	High	2/0.5	8.2	Yes	No
pCD46_4	Entire Ecto	1-299	PorCD46	pCD46_Fprimer4	pCD46_Rprimer6	Single Step	Yes	High	2/0.5	2.2	No	N/A
pADAM17_1	Ecto Domain	1-663	ADAM17_Ecto	pADAM17_Fprimer1	pADAM17_Rprimer1	Single Step	No	N/A	N/A	N/A	N/A	N/A
pADAM17_2	Mtalloproteinase	223-466	ADAM17_Ecto	pADAM17_Fprimer2	pADAM17_Rprimer2	Single Step	Yes	None	N/A	N/A	N/A	N/A

F

Additional SAXS Results

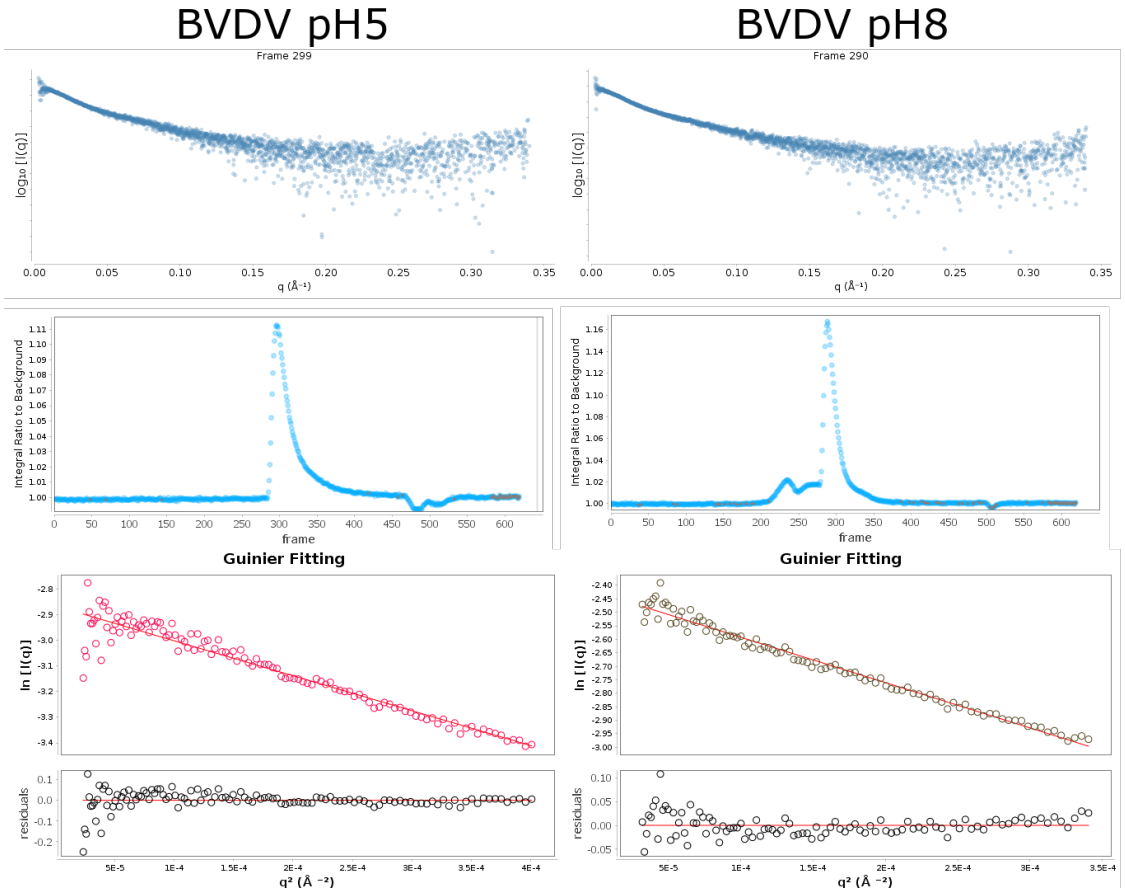


Figure F.1: BVDV E2 Ectodomain. **Top-** Scattering Intensities of $\log I_{10}(Q)$ against Q. **Middle-** Chromatogram of the HPLC run. **Bottom-** Guinier Analysis.

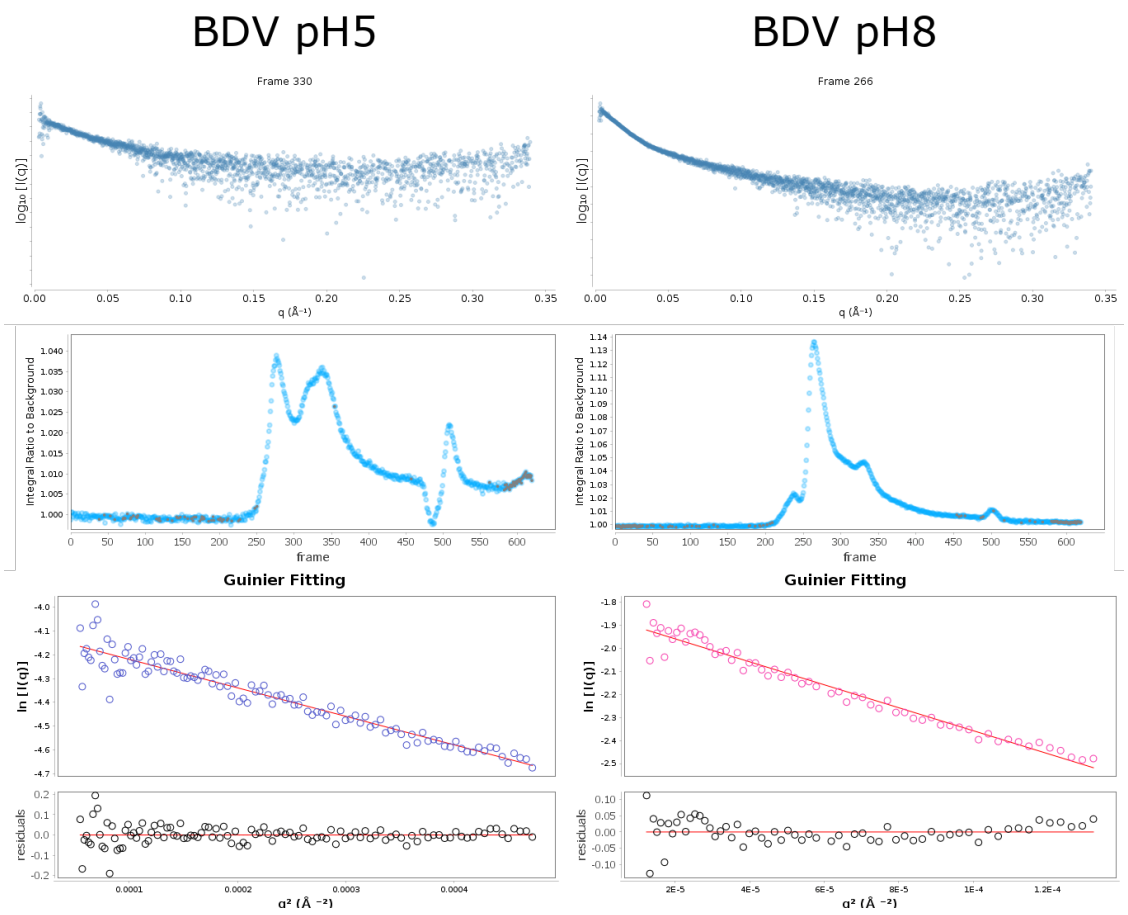


Figure F.2: BDV E2 Ectodomain. **Top-** Scattering Intensities of $\log I_{10}(Q)$ against Q. **Middle-** Chromatogram of the HPLC run. **Bottom-** Guinier Analysis.

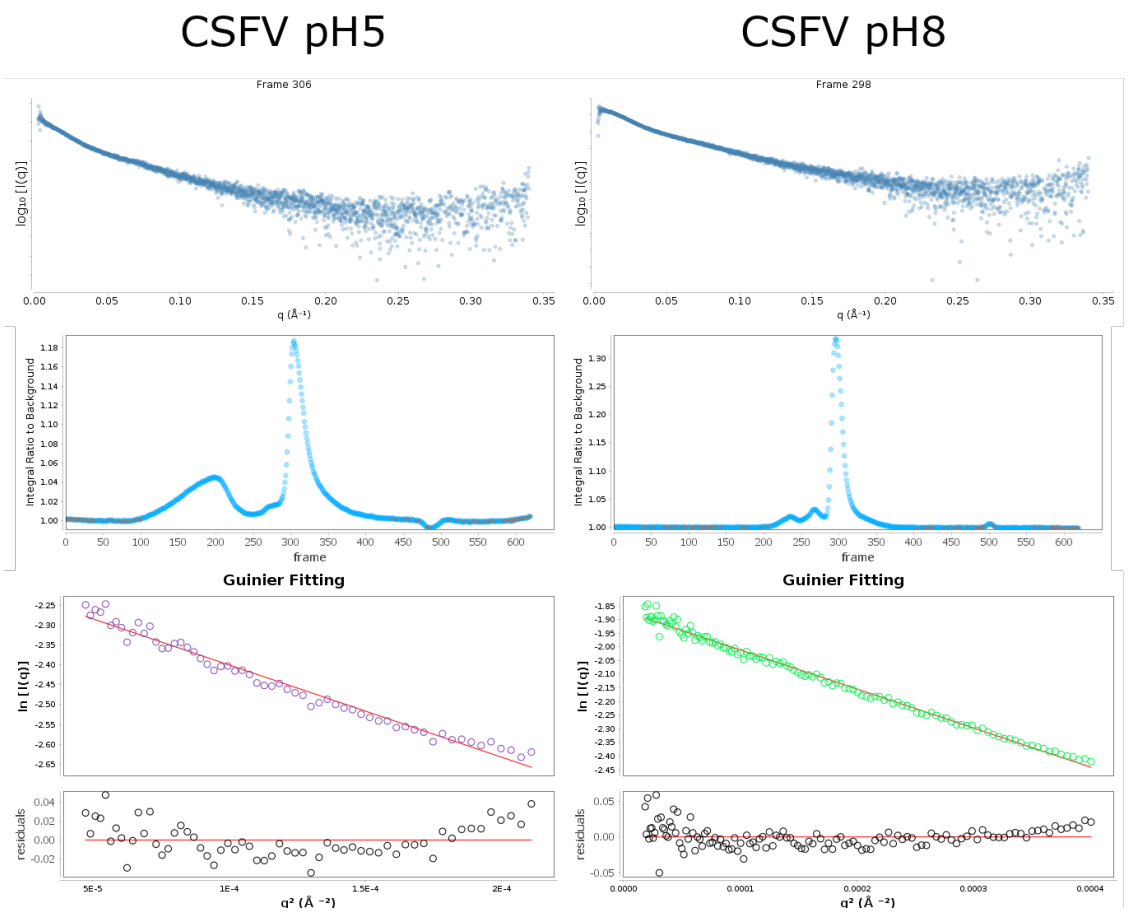


Figure F.3: CSFV E2 Ectodomain. **Top-** Scattering Intensities of $\log I_{10}(Q)$ against Q. **Middle-** Chromatogram of the HPLC run. **Bottom-** Guinier Analysis.

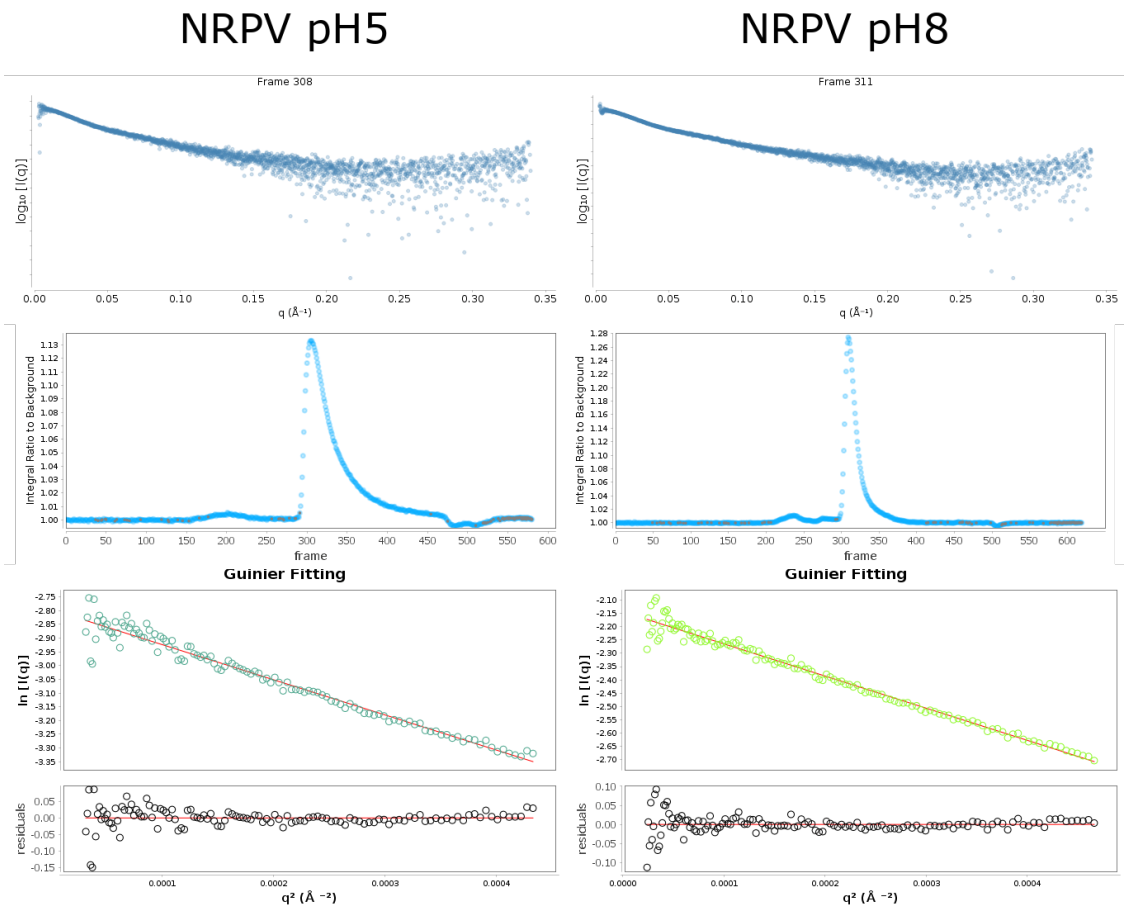


Figure F.4: NRPV E2 Ectodomain. **Top-** Scattering Intensities of $\log I_{10}(Q)$ against Q. **Middle-** Chromatogram of the HPLC run. **Bottom-** Guinier Analysis.

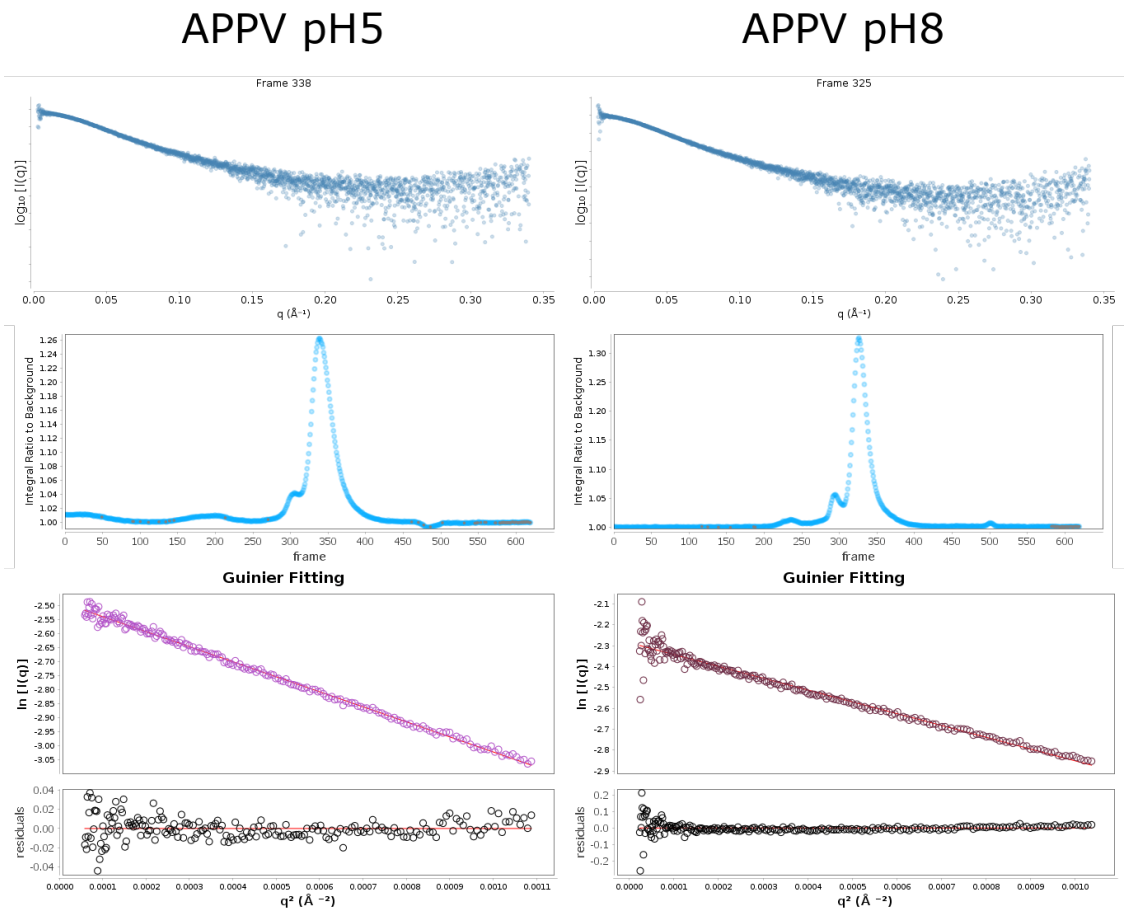


Figure F.5: APPV E2 Ectodomain. **Top-** Scattering Intensities of $\log I_{10}(Q)$ against Q. **Middle-** Chromatogram of the HPLC run. **Bottom-** Guinier Analysis.

G

Cryo-EM of CSFV E2 with Neutralising Antibody

G.1 Introduction

Cryo-EM is a technique which has been around since the 1980s but until recently generally produced low-resolution structures of large molecules[194]. Advances in software and detector technology have now allowed the acquisition of high-resolution structures for smaller protein targets. Recently, structures of proteins as small as 52kDa have been solved[195], but it is more routinely used for proteins or protein complexes of over 100kDa in size. Not only is cryo-EM now able to work with smaller protein but it can sometimes compete with X-ray crystallography in term of achievable resolution. Indeed, the structure of apoferritin has been solved with cryo-EM to a resolution of 1.2Å[196] although most structures deposited in the last couple of years have a lower resolution of around 3-4Å.

In cryo-EM, a few microlitres of concentrated protein solution is deposited on a fine metal mesh which is generally covered in a thin carbon film with holes. The excess sample is blotted away, and the grid plunged into liquid ethane cooled to near its freezing point in liquid nitrogen. This produces a grid with the protein sample, hopefully located in the holes in the carbon, surrounded with a thin layer of vitreous ice[197]. The grid is then placed in a microscope which uses electrons instead of photons and images, or movies, are collected on a direct electron detector. Software is then used to combine the 2D images of tens to hundreds of thousands of particles in multiple orientations to generate a 3D reconstruction of the particle.

Cryo-EM has multiple advantages over X-ray crystallography which all result from the lack of requirement to have a protein crystals. Many protein crystals are very difficult to grow and some proteins will not crystalise at all, so being able to use a liquid protein sample is highly advantageous. The protein is more likely to adopt its native conformation in a liquid sample and not be distorted by the crystallisation process. The protein is also free to adopt multiple conformations which again would be impossible in a crystal. Unfortunately, it is still, in general, a lower resolution technique than X-ray crystallography and requires proteins larger than the approximately 40kDa of a pestivirus E2. As such, this study attempted to use a fab fragment from a E2 neutralizing antibody in complex with the E2 dimer for a total size of around 180kDa.

G.2 Grid Preparation

Both BDV E2 ectodomain and CSFV E2 ectodomain crystalised but did not diffract to sufficient resolution for structure determination despite multiple rounds of optimisation. SAXS was carried out and demonstrated the presence of folded protein adopting a similar conformation to the structures of BVDV E2 and NRPV E2. However, this is a low-resolution technique and only provide a maximum resolution of around 15Å. As cryo-EM does not require protein crystals, it is an obvious technique to try and obtain a higher resolution structure. Both CSFV E2 ectodomain and BDV E2 ectodomain are only around 40kDa in size, 80kDa for the dimer, so I investigated finding another component to increase the size of the complex. Fortuitously, a collaborator (Christiane Riedel) donated a neutralising antibody reported to bind most typical pestivirus E2 proteins. As the amount of antibody available was not enough for the average crystallography experiment, utilising it for cryo-EM was ideal due to its lower sample requirements.

Initially, the antibody was directly mixed with a BVDV E2 sample which produced a gelatinous mass. This was likely due to the fact each antibody could bind two BVDV E2 molecules and as BVDV E2 is a dimer, it could bind two antibody molecules, so the entire mixture became cross linked. In order to avoid this, the antibodies were digested with a Pierce™ Fab Micro Preparation Kit from Thermo Fisher Scientific, and the fab fragments purified. These were then mixed with the CSFV E2 ectodomain protein in a slight molar excess and incubated for 30 minutes before grid preparation. A total protein concentration of both 2µM and 4µM was tested alongside multiple blotting and freezing parameters as well as different grid types (further details in the materials and methods section). Two small datasets were collected with the successful grid preparation parameters described in table F.1.

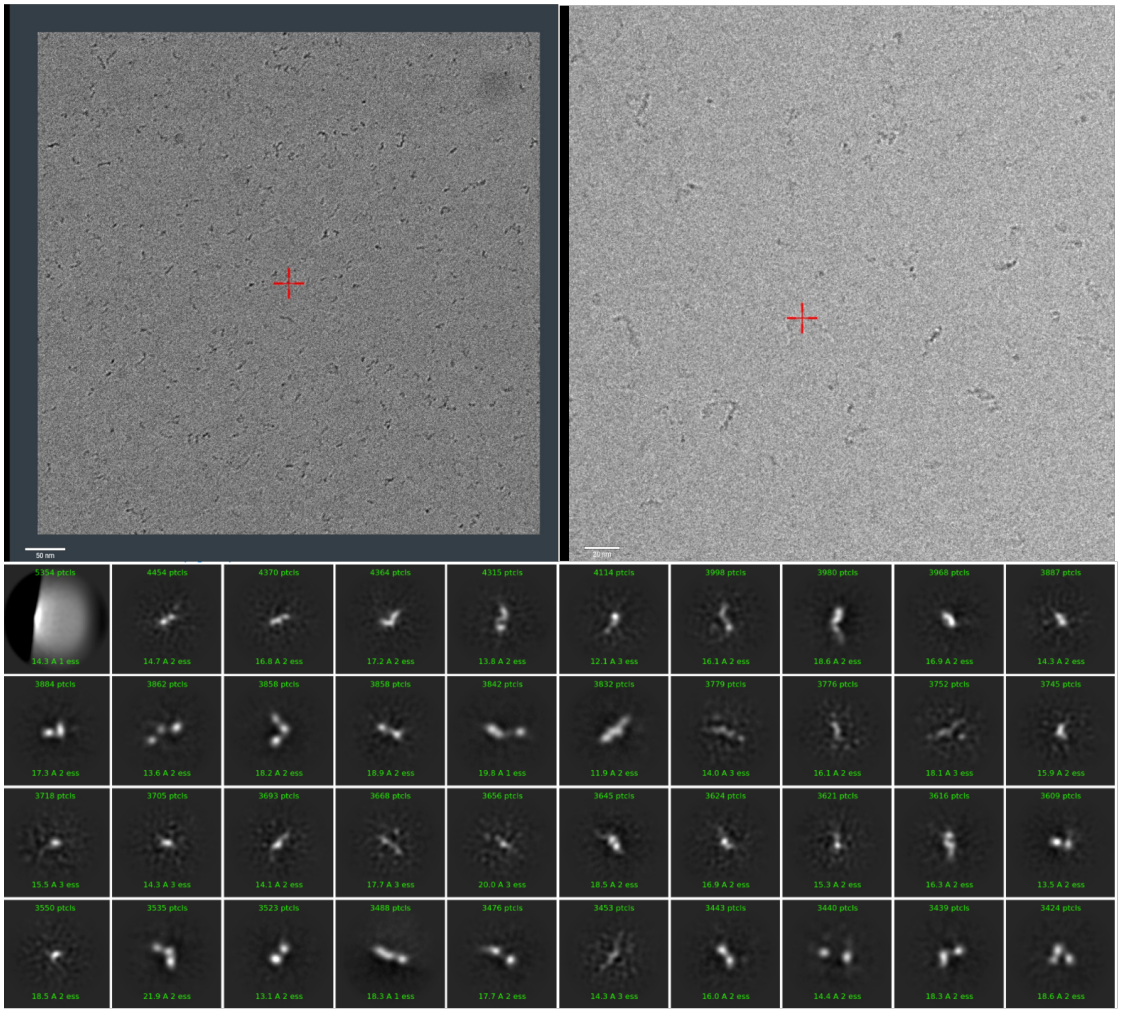


Figure G.1: Example micrographs with scale bars representing 50nm for left picture and 20nm for right picture. Below are some example 2D class averages.

	1	2
Grid Type	R2/1 Copper 200	R2/1 Copper 300
Protein Concentration	4µM	2µM
Volume Loaded	3µl	3µl
Blot Force	-1	-1
Blot Time	5 sec	6 sec
Wait before plunge	0 sec	0 sec

Table G.1: Successful Cryo-EM Grid Prepatation Parameters

G.3 Data Collection and Analysis

Data were collected as described in the materials and methods section on a Glacios with a Falcon III detector (Thermo Fisher Scientific) for a total of around 1500 exposures. Data were processed in cryoSPARK with automatic, manual, and template picking all attempted. Unfortunately, 2D class averaging never managed to align the particles properly. Due to time constraints, the project was abandoned. Figure F1 represents an example of micrograph and particles as well as a subset of the failed 2D class averages.

G.4 Materials and Methods

G.4.1 Grid Preparation

Qualtifoil holey carbon grids were glow discharged for 45 seconds at the high setting using an expanded plasma cleaner (Harrick Plasma). Both R 2/1 and R 1.2/1.3 200 copper mesh grids were used as well as 300 copper mesh R2/1. Either 3ul of 3.5ul of protein mixture at 2µM or 4µM was applied to the grids before blotting with force -1 or -25 for between 3 and 6 seconds before plunge freezing in liquid ethane. Blotting and plunge freezing were carried out with a Vitrobot System (Thermo Fisher Scientific). Frozen grids were transferred to liquid nitrogen, assembled in an autogrid cartridge (Thermo Fisher Scientific) and stored in liquid nitrogen.

G.4.2 Data Collection

Screening was carried out on a 200kV Glacios Cryo Electron microscope (Thermo Fisher Scientific) with a Falcon III detector at the Oxford Particle Imaging Centre. Two small data sets were collected with the collection parameters in table F2.

	Data Set 1	Data Set 2
Magnification	92000	92000
Pixel size (Å)	1.5	1.5
Total Dose (e/ Å^2)	60.22	59.79
Dose Rate (e/ Å^2/s)	31.19	28.76
Number of dose Fractions	50	40
Objective	70	70
Defocus Series	-1.8, -2.1, -2.4, -2.7, -3.0 -3.3	-1.8, -2.1, -2.4, -2.7, -3.0 -3.3
Exposures	1860	1484

Table G.2: Cryo-EM Data Collection Parameters

G.4.3 Data Processing

Motion correction, and CTF estimation were carried out in CryoSPARC live mode with default settings. Around 1000 particles were selected using the manual picking program before 2D classification was used to generate templates which were used for template picking. Full 2D classification, again with CryoSPARK, was carried out on the template picker results. Due to lack of well aligned 2D class averages data processing was stopped here.

References

[1] Peter Simmonds et al. “ICTV Virus Taxonomy Profile: Flaviviridae”. In: *The Journal of General Virology* 98.1 (Jan. 2017), pp. 2–3. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5370391/ (visited on 04/15/2020).

[2] Harvey J. Alter et al. “Posttransfusion Hepatitis After Exclusion of Commercial and Hepatitis-B Antigen-Positive Donors”. In: *Annals of Internal Medicine* 77.5 (Nov. 1972). Publisher: American College of Physicians, pp. 691–699. URL: https://www.acpjournals.org/doi/10.7326/0003-4819-77-5-691 (visited on 07/25/2021).

[3] A. Dereeper et al. “Phylogeny.fr: robust phylogenetic analysis for the non-specialist”. eng. In: *Nucleic Acids Research* 36.Web Server issue (July 2008), W465–469.

[4] Fábio Madeira et al. “The EMBL-EBI search and sequence analysis tools APIs in 2019”. eng. In: *Nucleic acids research* 47.W1 (July 2019), W636–W641. URL: https://europepmc.org/articles/PMC6602479 (visited on 07/29/2021).

[5] Donald B. Smith et al. “Proposed revision to the taxonomy of the genus Pestivirus, family Flaviviridae”. In: *Journal of General Virology* 98.8 (2017), pp. 2106–2112. URL: http://jgv.microbiologyresearch.org/content/journal/jgv/10.1099/jgv.0.000873 (visited on 02/14/2018).

[6] Zhiqiang Wu et al. “Virome Analysis for Identification of Novel Mammalian Viruses in Bat Species from Chinese Provinces”. In: *Journal of Virology* 86.20 (Oct. 2012), pp. 10999–11012. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3457178/ (visited on 07/24/2021).

[7] Cadhla Firth et al. “Detection of Zoonotic Pathogens and Characterization of Novel Viruses Carried by Commensal Rattus norvegicus in New York City”. In: *mBio* 5.5 (Oct. 2014), e01933–14. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4205793/ (visited on 07/24/2021).

[8] Steven Edwards et al. “Classical swine fever: the global situation”. In: *Vet Microbiol* (2000).

[9] T. Childs. “X Disease of Cattle - Saskatchewan”. eng. In: *Canadian Journal of Comparative Medicine and Veterinary Science* 10.11 (Nov. 1946), pp. 316–319.

[10] N. R. Underdahl, O. D. Grace, and A. B. Hoerlein. “Cultivation in tissue-culture of cytopathogenic agent from bovine mucosal disease”. eng. In: *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, N.Y.)* 94.4 (Apr. 1957), pp. 795–797.

[11] "Chapter 29 - Flaviviridae". en. In: *Fenner's Veterinary Virology (Fifth Edition)*. Ed. by N. James MacLachlan and Edward J. Dubovi. Boston: Academic Press, Jan. 2017, pp. 525–545. URL: http://www.sciencedirect.com/science/article/pii/B9780128009468000295 (visited on 04/15/2020).

[12] Ben M. Hause et al. "Discovery of a novel putative atypical porcine pestivirus in pigs in the USA". eng. In: *The Journal of General Virology* 96.10 (Oct. 2015), pp. 2994–2998.

[13] Benjamin Lamp et al. "Novel Pestivirus Species in Pigs, Austria, 2015". eng. In: *Emerging Infectious Diseases* 23.7 (July 2017), pp. 1176–1179.

[14] Jimin Wang, Yue Li, and Yorgo Modis. "Structural models of the membrane anchors of envelope glycoproteins E1 and E2 from pestiviruses". In: *Virology* 0 (Apr. 2014), pp. 93–101. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3986810/ (visited on 02/06/2018).

[15] Yu Mu et al. "Characterization of Membrane Topology and Retention Signal of Pestiviral Glycoprotein E1". In: *Journal of Virology* 95.15 (). Publisher: American Society for Microbiology, e00521–21. URL: https://journals.asm.org/doi/10.1128/JVI.00521-21 (visited on 09/25/2021).

[16] Daniel Aberle et al. "Structure of the Membrane Anchor of Pestivirus Glycoprotein Erns, a Long Tilted Amphipathic Helix". en. In: *PLOS Pathogens* 10.2 (Feb. 2014). Publisher: Public Library of Science, e1003973. URL: https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1003973 (visited on 09/25/2021).

[17] T. Rümenapf et al. "Structural proteins of hog cholera virus expressed by vaccinia virus: further characterization and induction of protective immunity". eng. In: *Journal of Virology* 65.2 (Feb. 1991), pp. 589–597.

[18] T. Lackner, H.-J. Thiel, and N. Tautz. "Dissection of a viral autoprotease elucidates a function of a cellular chaperone in proteolysis". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 103.5 (Jan. 2006), pp. 1510–1515.

[19] Keerthi Gottipati et al. "The Structure of Classical Swine Fever Virus Npro: A Novel Cysteine Autoprotease and Zinc-Binding Protein Involved in Subversion of Type I Interferon Induction". en. In: *PLOS Pathogens* 9.10 (Oct. 2013). Publisher: Public Library of Science, e1003704. URL: https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1003704 (visited on 09/25/2021).

[20] Thomas Krey et al. "Crystal Structure of the Pestivirus Envelope Glycoprotein Erns and Mechanistic Analysis of Its Ribonuclease Activity". English. In: *Structure* 20.5 (May 2012). Publisher: Elsevier, pp. 862–873. URL: https://www.cell.com/structure/abstract/S0969-2126(12)00136-0 (visited on 09/25/2021).

[21] Kamel El Omari et al. "Structure of a Pestivirus Envelope Glycoprotein E2 Clarifies Its Role in Cell Entry". In: *Cell Reports* 3.1 (Jan. 2013), pp. 30–35. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3607223/ (visited on 02/06/2018).

[22] Yue Li et al. "Crystal structure of glycoprotein E2 from bovine viral diarrhea virus". In: *Proceedings of the National Academy of Sciences of the United States of America* 110.17 (Apr. 2013), pp. 6805–6810. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3637714/ (visited on 02/06/2018).

[23] Fengwei Zheng et al. "Uncoupling of Protease trans-Cleavage and Helicase Activities in Pestivirus NS3". In: *Journal of Virology* 91.21 (). Publisher: American Society for Microbiology, e01094–17. URL: https://journals.asm.org/doi/10.1128/JVI.01094-17 (visited on 09/25/2021).

[24] Weiwei Li et al. "Crystal Structure of Classical Swine Fever Virus NS5B Reveals a Novel N-Terminal Domain". In: *Journal of Virology* 92.14 (). Publisher: American Society for Microbiology, e00324–18. URL: https://journals.asm.org/doi/10.1128/JVI.00324-18 (visited on 09/25/2021).

[25] Nathalie Callens et al. "Morphology and Molecular Composition of Purified Bovine Viral Diarrhea Virus Envelope". en. In: *PLOS Pathogens* 12.3 (Mar. 2016), e1005476. URL: http://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1005476 (visited on 06/05/2018).

[26] Veronika Richter et al. "A systematic worldwide review of the direct monetary losses in cattle due to bovine viral diarrhoea virus infection". In: *The Veterinary Journal* 220 (Feb. 2017), pp. 80–87. URL: http://www.sciencedirect.com/science/article/pii/S1090023317300102 (visited on 08/07/2019).

[27] Miranda P. M Meuwissen et al. "A model to estimate the financial consequences of classical swine fever outbreaks: principles and outcomes". In: *Preventive Veterinary Medicine* 42.3 (Dec. 1999), pp. 249–270. URL: http://www.sciencedirect.com/science/article/pii/S0167587799000793 (visited on 08/07/2019).

[28] Ad De Groof et al. "Atypical Porcine Pestivirus: A Possible Cause of Congenital Tremor Type A-II in Newborn Piglets". en. In: *Viruses* 8.10 (Oct. 2016). Number: 10 Publisher: Multidisciplinary Digital Publishing Institute, p. 271. URL: https://www.mdpi.com/1999-4915/8/10/271 (visited on 07/25/2021).

[29] Igor Renan Honorato Gatto, Karina Sonálio, and Luís Guilherme de Oliveira. "Atypical Porcine Pestivirus (APPV) as a New Species of Pestivirus in Pig Production". English. In: *Frontiers in Veterinary Science* 6 (2019). URL: https://www.frontiersin.org/articles/10.3389/fvets.2019.00035/full (visited on 08/15/2019).

[30] Alexander Postel et al. “Epidemiology, diagnosis and control of classical swine fever: Recent developments and future challenges”. en. In: *Transboundary and Emerging Diseases* 65.S1 (2018). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/tbed.12676, pp. 248–261. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/tbed.12676 (visited on 04/24/2020).

[31] Liani Coronado et al. “A Critical Review about Different Vaccines against Classical Swine Fever Virus and Their Repercussions in Endemic Regions”. In: *Vaccines* 9.2 (Feb. 2021), p. 154. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7918945/ (visited on 07/25/2021).

[32] Volker Moennig. “The control of classical swine fever in wild boar”. In: *Frontiers in Microbiology* 6 (Nov. 2015), p. 1211. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4635204/ (visited on 07/25/2021).

[33] *Classical swine fever*. en-GB. URL: https://www.oie.int/en/disease/classical-swine-fever/ (visited on 07/25/2021).

[34] Volker Moennig and Paul Becher. “Control of Bovine Viral Diarrhea”. In: *Pathogens* 7.1 (Mar. 2018). URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5874755/ (visited on 04/24/2020).

[35] P. J. Durham and L. E. Hassard. “An enzyme-linked immunosorbent assay (ELISA) for antibodies to bovine viral diarrhea virus”. eng. In: *Veterinary Microbiology* 22.1 (Mar. 1990), pp. 1–10.

[36] T. L. Poole et al. “Pestivirus translation initiation occurs by internal ribosome entry”. eng. In: *Virology* 206.1 (Jan. 1995), pp. 750–754.

[37] Stephen C. Harrison. “Viral membrane fusion”. en. In: *Virology* 479-480 (May 2015), pp. 498–507. URL: http://linkinghub.elsevier.com/retrieve/pii/S004268221500183X (visited on 05/04/2018).

[38] Leonid V. Chernomordik and Michael M. Kozlov. “Protein-Lipid Interplay in Fusion and Fission of Biological Membranes”. In: *Annual Review of Biochemistry* 72.1 (2003). _eprint: https://doi.org/10.1146/annurev.biochem.72.121801.161504, pp. 175–207. URL: https://doi.org/10.1146/annurev.biochem.72.121801.161504 (visited on 07/25/2021).

[39] Felix A. Rey and Shee-Mei Lok. “Common Features of Enveloped Viruses and Implications for Immunogen Design for Next-Generation Vaccines”. In: *Cell* 172.6 (Mar. 2018), pp. 1319–1334. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7112304/ (visited on 04/16/2020).

[40] James Stevens et al. “Structure of the Uncleaved Human H1 Hemagglutinin from the Extinct 1918 Influenza Virus”. en. In: *Science* 303.5665 (Mar. 2004). Publisher: American Association for the Advancement of Science Section: Report, pp. 1866–1870. URL: https://science.sciencemag.org/content/303/5665/1866 (visited on 07/25/2021).

[41] Félix A. Rey. “Dengue virus envelope glycoprotein structure: New insight into its interactions during viral entry”. In: *Proceedings of the National Academy of Sciences of the United States of America* 100.12 (June 2003), pp. 6899–6901. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC165800/ (visited on 07/25/2021).

[42] “Structures of immature flavivirus particles”. In: *The EMBO Journal* 22.11 (June 2003). Publisher: John Wiley & Sons, Ltd, pp. 2604–2613. URL: https://www.embopress.org/doi/full/10.1093/emboj/cdg270 (visited on 07/25/2021).

[43] Vinod Nayak et al. “Crystal Structure of Dengue Virus Type 1 Envelope Protein in the Postfusion Conformation and Its Implications for Membrane Fusion”. In: *Journal of Virology* 83.9 (May 2009). Publisher: American Society for Microbiology, pp. 4338–4344. URL: https://journals.asm.org/doi/full/10.1128/JVI.02574-08 (visited on 07/25/2021).

[44] Stéphane Roche et al. “Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G”. eng. In: *Science (New York, N.Y.)* 313.5784 (July 2006), pp. 187–191.

[45] Sonia Libersou et al. “Distinct structural rearrangements of the VSV glycoprotein drive membrane fusion”. In: *The Journal of Cell Biology* 191.1 (Oct. 2010), pp. 199–210. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2953448/ (visited on 07/25/2021).

[46] H J Thiel et al. “Hog cholera virus: molecular composition of virions from a pestivirus.” In: *Journal of Virology* 65.9 (Sept. 1991), pp. 4705–4712. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC248926/ (visited on 07/25/2021).

[47] R. Schneider et al. “Identification of a structural glycoprotein of an RNA virus as a ribonuclease”. en. In: *Science* 261.5125 (Aug. 1993). Publisher: American Association for the Advancement of Science Section: Reports, pp. 1169–1171. URL: https://science.sciencemag.org/content/261/5125/1169 (visited on 07/25/2021).

[48] Daniel Aberle, Kay-Marcus Oetter, and Gregor Meyers. “Lipid Binding of the Amphipathic Helix Serving as Membrane Anchor of Pestivirus Glycoprotein Erns”. en. In: *PLOS ONE* 10.8 (Aug. 2015). Publisher: Public Library of Science, e0135680. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0135680 (visited on 07/25/2021).

[49] Fun-In Wang et al. “Structures and Functions of Pestivirus Glycoproteins: Not Simply Surface Matters”. In: *Viruses* 7.7 (June 2015), pp. 3506–3529. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4517112/ (visited on 02/06/2018).

[50] Brett D Lindenbach, Heinz-Jurgen Thiel, and Charles M Rice. “Flaviviridae: The Viruses and Their Replication”. en. In: (2006), p. 52.

[51] Zai Wang et al. “Characterization of classical swine fever virus entry by using pseudotyped viruses: E1 and E2 are sufficient to mediate viral entry”. In: *Virology* 330.1 (Dec. 2004), pp. 332–341. URL: http://www.sciencedirect.com/science/article/pii/S0042682204006257 (visited on 02/06/2018).

[52] Gérard J. Arlaud et al. “Deciphering complement mechanisms: The contributions of structural biology”. en. In: *Molecular Immunology*. XIth European meeting on Complement in Human Disease 44.16 (Sept. 2007), pp. 3809–3822. URL: https://www.sciencedirect.com/science/article/pii/S0161589007002556 (visited on 06/19/2021).

[53] Munir Iqbal and John W. McCauley. “Identification of the glycosaminoglycan-binding site on the glycoprotein E(rns) of bovine viral diarrhoea virus by site-directed mutagenesis”. eng. In: *The Journal of General Virology* 83.Pt 9 (Sept. 2002), pp. 2153–2159.

[54] Sandra Burrack et al. “A new type of intracellular retention signal identified in a pestivirus structural glycoprotein”. eng. In: *FASEB journal: official publication of the Federation of American Societies for Experimental Biology* 26.8 (Aug. 2012), pp. 3292–3305.

[55] Ioana Bintintan and Gregor Meyers. “A new type of signal peptidase cleavage site identified in an RNA virus polyprotein”. eng. In: *The Journal of Biological Chemistry* 285.12 (Mar. 2010), pp. 8572–8584.

[56] Yong Ho Kim et al. “Capture and imaging of a prehairpin fusion intermediate of the paramyxovirus PIV5”. In: *Proceedings of the National Academy of Sciences of the United States of America* 108.52 (Dec. 2011), pp. 20992–20997. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3248524/ (visited on 01/03/2022).

[57] Yue Li and Yorgo Modis. “A novel membrane fusion protein family in Flaviviridae?” en. In: *Trends in Microbiology* 22.4 (Apr. 2014), pp. 176–182. URL: http://linkinghub.elsevier.com/retrieve/pii/S0966842X14000201 (visited on 06/07/2018).

[58] Yu Mu et al. “Interaction of Pestiviral E1 and E2 Sequences in Dimer Formation and Intracellular Retention”. en. In: *International Journal of Molecular Sciences* 22.14 (Jan. 2021). Number: 14 Publisher: Multidisciplinary Digital Publishing Institute, p. 7285. URL: https://www.mdpi.com/1422-0067/22/14/7285 (visited on 09/25/2021).

[59] A. Krogh et al. “Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes”. eng. In: *Journal of Molecular Biology* 305.3 (Jan. 2001), pp. 567–580.

[60] J. Patel, A. H. Patel, and J. McLauchlan. "The transmembrane domain of the hepatitis C virus E2 glycoprotein is required for correct folding of the E1 glycoprotein and native complex formation". eng. In: *Virology* 279.1 (Jan. 2001), pp. 58–68.

[61] Saskia Ronecker et al. "Formation of bovine viral diarrhea virus E1–E2 heterodimers is essential for virus entry and depends on charged residues in the transmembrane domains". In: *Journal of General Virology,* 89.9 (2008), pp. 2114–2121. URL: https://www.microbiologyresearch.org/content/journal/jgv/10.1099/vir.0.2008/001792-0 (visited on 09/04/2019).

[62] Thomas Krey et al. "Function of Bovine CD46 as a Cellular Receptor for Bovine Viral Diarrhea Virus Is Determined by Complement Control Protein 1". In: *Journal of Virology* 80.8 (Apr. 2006), pp. 3912–3922. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1440463/ (visited on 02/12/2018).

[63] Xuewu Li et al. "Identification of host cell binding peptide from an overlapping peptide library for inhibition of classical swine fever virus infection". en. In: *Virus Genes* 43.1 (Aug. 2011), pp. 33–40. URL: https://doi.org/10.1007/s11262-011-0595-7 (visited on 08/08/2019).

[64] Long Li et al. "Structural changes of envelope proteins during alphavirus fusion". eng. In: *Nature* 468.7324 (Dec. 2010), pp. 705–708.

[65] H. G. P. van Gennip et al. "Dimerisation of glycoprotein Erns of classical swine fever virus is not essential for viral replication and infection". en. In: *Archives of Virology* 150.11 (Nov. 2005), pp. 2271–2286. URL: https://doi.org/10.1007/s00705-005-0569-y (visited on 04/21/2020).

[66] Chia-Yi Chang et al. "Identification of conformational epitopes and antigen-specific residues at the D/A domains and the extramembrane C-terminal region of E2 glycoprotein of classical swine fever virus". en. In: *Virus Research* 168.1 (Sept. 2012), pp. 56–63. URL: http://www.sciencedirect.com/science/article/pii/S0168170212002109 (visited on 04/21/2020).

[67] Rosalind Kornfeld and Stuart Kornfeld. "Assembly of Asparagine-Linked Oligosaccharides". In: *Annual Review of Biochemistry* 54.1 (1985). _eprint: https://doi.org/10.1146/annurev.bi.54.070185.003215, pp. 631–664. URL: https://doi.org/10.1146/annurev.bi.54.070185.003215 (visited on 01/03/2022).

[68] Norica Branza-Nichita et al. "Role of N-glycan trimming in the folding and secretion of the pestivirus protein E(rns)". eng. In: *Biochemical and Biophysical Research Communications* 319.2 (June 2004), pp. 655–662.

[69] I. Fernandez-Sainz et al. "Alteration of the N-linked glycosylation condition in E1 glycoprotein of Classical Swine Fever Virus strain Brescia alters virulence in swine". eng. In: *Virology* 386.1 (Mar. 2009), pp. 210–216.

[70] G. R. Risatti et al. "N-linked glycosylation status of classical swine fever virus strain Brescia E2 glycoprotein influences virulence in swine". eng. In: *Journal of Virology* 81.2 (Jan. 2007), pp. 924–933.

[71] Jolanta Tyborowska, Ewa Zdunek, and Bogusław Szewczyk. "Effect of N-glycosylation inhibition on the synthesis and processing of classical swine fever virus glycoproteins". eng. In: *Acta Biochimica Polonica* 54.4 (2007), pp. 813–819.

[72] Xuelian Luo et al. "Glycosylation of classical swine fever virus E(rns) is essential for binding double-stranded RNA and preventing interferon-beta induction". eng. In: *Virus Research* 146.1-2 (Dec. 2009), pp. 135–139.

[73] Boris K. Gavrilov et al. "Effects of glycosylation on antigenicity and immunogenicity of classical swine fever virus envelope proteins". eng. In: *Virology* 420.2 (Nov. 2011), pp. 135–145.

[74] Leopold Kong et al. "Hepatitis C virus E2 envelope glycoprotein core structure". eng. In: *Science (New York, N.Y.)* 342.6162 (Nov. 2013), pp. 1090–1094.

[75] Heidi E. Drummer, Irene Boo, and PantelisYR 2007 Poumbourios. "Mutagenesis of a conserved fusion peptide-like motif and membrane-proximal heptad-repeat region of hepatitis C virus glycoprotein E1". In: *Journal of General Virology* 88.4 (). Publisher: Microbiology Society, pp. 1144–1148. URL: https://www.microbiologyresearch.org/content/journal/jgv/10.1099/vir.0.82567-0 (visited on 07/26/2021).

[76] J P Michalak et al. "Characterization of truncated forms of hepatitis C virus glycoproteins." In: *Journal of General Virology* 78.9 (). Publisher: Microbiology Society, pp. 2299–2306. URL: https://www.microbiologyresearch.org/content/journal/jgv/10.1099/0022-1317-78-9-2299 (visited on 07/26/2021).

[77] Thomas Krey, Heinz-Jürgen Thiel, and Till Rümenapf. "Acid-Resistant Bovine Pestivirus Requires Activation for pH-Triggered Fusion during Entry". en. In: *Journal of Virology* 79.7 (Apr. 2005), pp. 4191–4200. URL: http://jvi.asm.org/content/79/7/4191 (visited on 02/06/2018).

[78] J. J. H. Chu and M. L. Ng. "Infectious entry of West Nile virus occurs through a clathrin-mediated endocytic pathway". eng. In: *Journal of Virology* 78.19 (Oct. 2004), pp. 10543–10555.

[79] Richard Fritz, Karin Stiasny, and Franz X. Heinz. "Identification of specific histidines as pH sensors in flavivirus membrane fusion". In: *The Journal of Cell Biology* 183.2 (Oct. 2008), pp. 353–361. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2568029/ (visited on 07/26/2021).

[80] Victor C. Chu et al. "The Avian Coronavirus Infectious Bronchitis Virus Undergoes Direct Low-pH-Dependent Fusion Activation during Entry into Host Cells". In: *Journal of Virology* 80.7 (Apr. 2006), pp. 3180–3188. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1440383/ (visited on 07/26/2021).

[81] Donald J. Benton et al. "Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion". en. In: *Nature* 588.7837 (Dec. 2020), pp. 327–330. URL: https://www.nature.com/articles/s41586-020-2772-0 (visited on 07/26/2021).

[82] B. Grummer, S. Grotha, and I. Greiser-Wilke. "Bovine Viral Diarrhoea Virus is Internalized by Clathrin-dependent Receptor-mediated Endocytosis". en. In: *Journal of Veterinary Medicine, Series B* 51.10 (2004). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1439-0450.2004.00798.x, pp. 427–432. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1439-0450.2004.00798.x (visited on 04/24/2020).

[83] Steve Lecot et al. "Bovine Viral Diarrhea Virus Entry Is Dependent on Clathrin-Mediated Endocytosis". en. In: *Journal of Virology* 79.16 (Aug. 2005), pp. 10826–10829. URL: http://jvi.asm.org/content/79/16/10826 (visited on 05/08/2018).

[84] Basavaraj Shrishail Mathapati et al. "Entry of bovine viral diarrhea virus into ovine cells occurs through clathrin-dependent endocytosis and low pH-dependent fusion". In: *In Vitro Cellular & Developmental Biology. Animal* 46.5 (2010). Publisher: [Springer, Society for In Vitro Biology], pp. 403–407. URL: https://www.jstor.org/stable/40663748 (visited on 04/24/2020).

[85] Bao-Jun Shi et al. "Entry of Classical Swine Fever Virus into PK-15 Cells via a pH-, Dynamin-, and Cholesterol-Dependent, Clathrin-Mediated Endocytic Pathway That Requires Rab5 and Rab7". eng. In: *Journal of Virology* 90.20 (2016), pp. 9194–9208.

[86] Valeria Cagno et al. "Heparan Sulfate Proteoglycans and Viral Attachment: True Receptors or Adaptation Bias?" In: *Viruses* 11.7 (July 2019), p. 596. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6669472/ (visited on 09/25/2021).

[87] Carolin Dräger, Martin Beer, and Sandra Blome. "Porcine complement regulatory protein CD46 and heparan sulfates are the major factors for classical swine fever virus attachment in vitro". eng. In: *Archives of Virology* 160.3 (Mar. 2015), pp. 739–746.

[88] Gökce Nur Cagatay et al. "Porcine Complement Regulatory Protein CD46 Is a Major Receptor for Atypical Porcine Pestivirus but Not for Classical Swine Fever Virus". In: *Journal of Virology* 95.9 (Apr. 2021), e02186–20. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8104093/ (visited on 06/19/2021).

[89] Karin Maurer et al. "CD46 Is a Cellular Receptor for Bovine Viral Diarrhea Virus". In: *Journal of Virology* 78.4 (Feb. 2004), pp. 1792–1799. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC369467/ (visited on 02/12/2018).

[90] Jianing Chen et al. "The laminin receptor is a cellular attachment receptor for classical Swine Fever virus". eng. In: *Journal of Virology* 89.9 (May 2015), pp. 4894–4906.

[91] Fei Yuan et al. "ADAM17 is an essential attachment factor for classical swine fever virus". en. In: *PLOS Pathogens* 17.3 (Mar. 2021). Publisher: Public Library of Science, e1009393. URL: https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1009393 (visited on 07/27/2021).

[92] T. Seya, J. R. Turner, and J. P. Atkinson. "Purification and characterization of a membrane protein (gp45-70) that is a cofactor for cleavage of C3b and C4b". eng. In: *The Journal of Experimental Medicine* 163.4 (Apr. 1986), pp. 837–855.

[93] M. Kathryn Liszewski and John P. Atkinson. "Complement regulator CD46: genetic variants and disease associations". In: *Human Genomics* 9.1 (June 2015), p. 7. URL: https://doi.org/10.1186/s40246-015-0029-z (visited on 06/19/2021).

[94] B. David Persson et al. "Structure of the Extracellular Portion of CD46 Provides Insights into Its Interactions with Complement Proteins and Pathogens". en. In: *PLOS Pathogens* 6.9 (Sept. 2010). Publisher: Public Library of Science, e1001122. URL: https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1001122 (visited on 06/19/2021).

[95] Christiane Riedel et al. "Real Time Analysis of Bovine Viral Diarrhea Virus (BVDV) Infection and Its Dependence on Bovine CD46". In: *Viruses* 12.1 (Jan. 2020), p. 116. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7019258/ (visited on 06/24/2021).

[96] M. M. Hulst, H. G. van Gennip, and R. J. Moormann. "Passage of classical swine fever virus in cultured swine kidney cells selects virus variants that bind to heparan sulfate due to a single amino acid change in envelope protein E(rns)". eng. In: *Journal of Virology* 74.20 (Oct. 2000), pp. 9553–9561.

[97] Rita Eymann-Häni et al. "Propagation of classical swine fever virus in vitro circumventing heparan sulfate-adaptation". eng. In: *Journal of Virological Methods* 176.1-2 (Sept. 2011), pp. 85–95.

[98] Peng Zhong et al. "Cell-to-cell transmission of viruses". In: *Current opinion in virology* 3.1 (Feb. 2013), pp. 44–50. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3587356/ (visited on 07/27/2021).

[99] Jennifer M. Timpe et al. "Hepatitis C virus cell-cell transmission in hepatoma cells in the presence of neutralizing antibodies". en. In: *Hepatology* 47.1 (2008). _eprint: https://aasldpubs.onlinelibrary.wiley.com/doi/pdf/10.1002/hep.21959, pp. 17–24. URL: https://aasldpubs.onlinelibrary.wiley.com/doi/abs/10.1002/hep.21959 (visited on 07/27/2021).

[100] Fernando Merwaiss, Cecilia Czibener, and Diego E. Alvarez. "Cell-to-Cell Transmission Is the Main Mechanism Supporting Bovine Viral Diarrhea Virus Spread in Cell Culture". en. In: *Journal of Virology* 93.3 (Feb. 2019). Publisher: American Society for Microbiology Journals Section: Virus-Cell Interactions. URL: https://jvi.asm.org/content/93/3/e01776-18 (visited on 04/22/2020).

[101] Jeroen Witteveldt et al. "CD81 is dispensable for hepatitis C virus cell-to-cell transmission in hepatoma cells". eng. In: *The Journal of General Virology* 90.Pt 1 (Jan. 2009), pp. 48–58.

[102] Claire L. Brimacombe et al. "Neutralizing antibody-resistant hepatitis C virus cell-to-cell transmission". eng. In: *Journal of Virology* 85.1 (Jan. 2011), pp. 596–605.

[103] Fabian Sievers et al. “Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega”. eng. In: *Molecular Systems Biology* 7 (Oct. 2011), p. 539.

[104] *Benchling*. 2021.

[105] Claire Strain-Damerell and Nicola A. Burgess-Brown. “High-Throughput Site-Directed Mutagenesis”. eng. In: *Methods in Molecular Biology (Clifton, N.J.)* 2025 (2019), pp. 281–296.

[106] Janet Newman et al. “Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy”. en. In: (2005), p. 6.

[107] A. Chaikuad, S. Knapp, and F. von Delft. “Defined PEG smears as an alternative approach to enhance the search for crystallization conditions and crystal-quality improvement in reduced screens”. en. In: *Acta Crystallographica Section D: Biological Crystallography* 71.8 (Aug. 2015). Number: 8 Publisher: International Union of Crystallography, pp. 1627–1639. URL: http://scripts.iucr.org/cgi-bin/paper?nj5221 (visited on 10/05/2021).

[108] C. Vonrhein et al. “Data processing and analysis with the autoPROC toolbox”. en. In: *Acta Crystallographica Section D: Biological Crystallography* 67.4 (Apr. 2011), pp. 293–302. URL: //scripts.iucr.org/cgi-bin/paper?ba5166 (visited on 09/04/2019).

[109] Gábor Bunkóczi and Randy J. Read. “Improvement of molecular-replacement models with Sculptor”. In: *Acta Crystallographica Section D: Biological Crystallography* 67.Pt 4 (Mar. 2011), pp. 303–312. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3069745/ (visited on 10/05/2021).

[110] A. J. McCoy et al. “Phaser crystallographic software”. en. In: *Journal of Applied Crystallography* 40.4 (Aug. 2007), pp. 658–674. URL: http://scripts.iucr.org/cgi-bin/paper?he5368 (visited on 08/30/2019).

[111] Thomas C. Terwilliger et al. “Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard”. In: *Acta Crystallographica Section D: Biological Crystallography* 65.Pt 6 (June 2009), pp. 582–601. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2685735/ (visited on 09/26/2021).

[112] P. Emsley et al. “Features and development of *Coot*”. en. In: *Acta Crystallographica Section D Biological Crystallography* 66.4 (Apr. 2010), pp. 486–501. URL: http://scripts.iucr.org/cgi-bin/paper?S0907444910007493 (visited on 08/30/2019).

[113] Alexei A. Vagin et al. “REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use”. eng. In: *Acta Crystallographica. Section D, Biological Crystallography* 60.Pt 12 Pt 1 (Dec. 2004), pp. 2184–2195.

[114] I.J Tickle et al. *STARANISO*. Cambridge, United Kingdom, 2018.

[115] W. Kabsch. "XDS". en. In: *Acta Crystallographica Section D: Biological Crystallography* 66.2 (Feb. 2010). Number: 2 Publisher: International Union of Crystallography, pp. 125–132. URL: //scripts.iucr.org/cgi-bin/paper?dz5179 (visited on 06/30/2021).

[116] John Jumper et al. "Highly accurate protein structure prediction with AlphaFold". en. In: *Nature* 596.7873 (Aug. 2021), pp. 583–589. URL: https://www.nature.com/articles/s41586-021-03819-2 (visited on 09/26/2021).

[117] Minkyung Baek et al. "Accurate prediction of protein structures and interactions using a three-track neural network". en. In: *Science* 373.6557 (Aug. 2021). Publisher: American Association for the Advancement of Science Section: Research Article, pp. 871–876. URL: https://science.sciencemag.org/content/373/6557/871 (visited on 08/22/2021).

[118] A. Vagin and A. Teplyakov. "Molecular replacement with MOLREP". en. In: *Acta Crystallographica Section D: Biological Crystallography* 66.1 (Jan. 2010). Number: 1 Publisher: International Union of Crystallography, pp. 22–25. URL: http://scripts.iucr.org/cgi-bin/paper?dz5175 (visited on 09/20/2021).

[119] W. Kabsch. "Integration, scaling, space-group assignment and post-refinement". en. In: *Acta Crystallographica Section D: Biological Crystallography* 66.2 (Feb. 2010), pp. 133–144. URL: http://scripts.iucr.org/cgi-bin/paper?dz5178 (visited on 09/04/2019).

[120] G. Winter. "xia2: an expert system for macromolecular crystallography data reduction". en. In: *Journal of Applied Crystallography* 43.1 (2010). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1107/S0021889809045701, pp. 186–190. URL: https://onlinelibrary.wiley.com/doi/abs/10.1107/S0021889809045701 (visited on 09/28/2021).

[121] T. Pape and T. R. Schneider. "HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs". en. In: *Journal of Applied Crystallography* 37.5 (Oct. 2004). Number: 5 Publisher: International Union of Crystallography, pp. 843–844. URL: http://scripts.iucr.org/cgi-bin/paper?ea5021 (visited on 09/26/2021).

[122] P. V. Afonine et al. "Towards automated crystallographic structure refinement with phenix.refine". en. In: *Acta Crystallographica Section D: Biological Crystallography* 68.4 (Apr. 2012), pp. 352–367. URL: http://scripts.iucr.org/cgi-bin/paper?ba5180 (visited on 08/30/2019).

[123] *ScÅtter*. URL: https://www.bioisis.net/tutorials (visited on 09/29/2021).

[124] K. Manalastas-Cantos et al. "ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis". en. In: *Journal of Applied Crystallography* 54.1 (Feb. 2021). Number: 1 Publisher: International Union of Crystallography, pp. 343–355. URL: https://journals.iucr.org/j/issues/2021/01/00/ge5081/ (visited on 09/29/2021).

[125] D. I. Svergun. "Determination of the regularization parameter in indirect-transform methods using perceptual criteria". en. In: *Journal of Applied Crystallography* 25.4 (Aug. 1992). Number: 4 Publisher: International Union of Crystallography, pp. 495–503. URL: //scripts.iucr.org/cgi-bin/paper?wi0087 (visited on 10/05/2021).

[126] Daniel Franke and Dmitri I. Svergun. "DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering". eng. In: *Journal of Applied Crystallography* 42.Pt 2 (Apr. 2009), pp. 342–346.

[127] V. V. Volkov, D. I. Svergun, and IUCr. *Uniqueness of ab initio shape determination in small-angle scattering*. en. text. ISSN: urn:issn:0021-8898 Issue: 3 Number: 3 Pages: 860-864 Publisher: International Union of Crystallography Volume: 36. Apr. 2003. URL: http://scripts.iucr.org/cgi-bin/paper?S0021889803000268 (visited on 09/29/2021).

[128] N. P. Brown, C. Leroy, and C. Sander. "MView: a web-compatible database search or multiple alignment viewer". eng. In: *Bioinformatics (Oxford, England)* 14.4 (1998), pp. 380–381.

[129] Aditya Pandey et al. "Current strategies for protein production and purification enabling membrane protein structural biology". In: *Biochemistry and cell biology = Biochimie et biologie cellulaire* 94.6 (Dec. 2016), pp. 507–527. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5752365/ (visited on 08/09/2021).

[130] Martin Caffrey. "A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes". In: *Acta Crystallographica. Section F, Structural Biology Communications* 71.Pt 1 (Jan. 2015), pp. 3–18. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4304740/ (visited on 01/03/2022).

[131] Nigel Darby and Thomas E. Creighton. "Disulfide Bonds in Protein Folding and Stability". en. In: *Protein Stability and Folding: Theory and Practice*. Ed. by Bret A. Shirley. Methods in Molecular Biology™. Totowa, NJ: Humana Press, 1995, pp. 219–252. URL: https://doi.org/10.1385/0-89603-301-5:219 (visited on 08/09/2021).

[132] Sergey A. Shiryev et al. "Improved BLAST searches using longer words for protein seeding". eng. In: *Bioinformatics (Oxford, England)* 23.21 (Nov. 2007), pp. 2949–2951.

[133] Kamel El Omari et al. "Unexpected structure for the N-terminal domain of hepatitis C virus envelope glycoprotein E1". en. In: *Nature Communications* 5 (Sept. 2014), p. 4874. URL: https://www.nature.com/articles/ncomms5874 (visited on 02/06/2018).

[134] J M Casasnovas, M Larvie, and T Stehle. "Crystal structure of two CD46 domains reveals an extended measles virus-binding surface." In: *The EMBO Journal* 18.11 (June 1999), pp. 2911–2922. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1171373/ (visited on 06/19/2021).

[135] Guennadi Kozlov and Kalle Gehring. "Calnexin cycle – structural features of the ER chaperone system". en. In: *The FEBS Journal* 287.20 (2020). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/febs.15330, pp. 4322–4340. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/febs.15330 (visited on 01/03/2022).

[136] Joshua A. Bornhorst and Joseph J. Falke. "[16] Purification of Proteins Using Polyhistidine Affinity Tags". In: *Methods in enzymology* 326 (2000), pp. 245–254. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2909483/ (visited on 08/09/2021).

[137] L. J. Ransone et al. "Fos-Jun interaction: mutational analysis of the leucine zipper domain of both proteins". eng. In: *Genes & Development* 3.6 (June 1989), pp. 770–781.

[138] Jonathan Elegheert et al. "Lentiviral transduction of mammalian cells for fast, scalable and high-level production of soluble and membrane proteins". eng. In: *Nature Protocols* 13.12 (Dec. 2018), pp. 2991–3017.

[139] A. R. Aricescu, W. Lu, and E. Y. Jones. "A time- and cost-efficient system for high-level protein production in mammalian cells". en. In: *Acta Crystallographica Section D: Biological Crystallography* 62.10 (Oct. 2006). Number: 10 Publisher: International Union of Crystallography, pp. 1243–1250. URL: //scripts.iucr.org/cgi-bin/paper?gx5088 (visited on 09/04/2021).

[140] Malathi Raman and Karen Martin. "One solution for cloning and mutagenesis: In-Fusion® HD Cloning Plus". en. In: *Nature Methods* 11.9 (Sept. 2014), pp. iii–v. URL: https://www.nature.com/articles/nmeth.f.373 (visited on 08/09/2021).

[141] *Primer design and other tools*. URL: https://www.takarabio.com/learning-centers/cloning/primer-design-and-other-tools (visited on 08/09/2021).

[142] Luigi Naldini, Didier Trono, and Inder M. Verma. "Lentiviral vectors, two decades later". eng. In: *Science (New York, N.Y.)* 353.6304 (Sept. 2016), pp. 1101–1102.

[143] M. I. Cockett, C. R. Bebbington, and G. T. Yarranton. "High level expression of tissue inhibitor of metalloproteinases in Chinese hamster ovary cells using glutamine synthetase gene amplification". eng. In: *Bio/Technology (Nature Publishing Company)* 8.7 (July 1990), pp. 662–667.

[144] P. Meissner et al. "Transient gene expression: recombinant protein production with suspension-adapted HEK293-EBNA cells". eng. In: *Biotechnology and Bioengineering* 75.2 (Oct. 2001), pp. 197–203.

[145] Philip J. Reeves et al. "Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 99.21 (Oct. 2002), pp. 13419–13424.

[146] F. Yao et al. "Tetracycline repressor, tetR, rather than the tetR-mammalian cell transcription factor fusion derivatives, regulates inducible gene expression in mammalian cells". eng. In: *Human Gene Therapy* 9.13 (Sept. 1998), pp. 1939–1950.

[147] Douglas A. Stirling et al. "Protein A–calmodulin fusions: a novel approach for investigating calmodulin function in yeast". en. In: *Molecular Microbiology* 6.6 (1992). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2958.1992.tb01519.x, pp. 703–713. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.1992.tb01519.x (visited on 09/09/2021).

[148] Sandra Harper and David W. Speicher. "Purification of proteins fused to glutathione S-tranferase". In: *Methods in molecular biology (Clifton, N.J.)* 681 (2011), pp. 259–280. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3584333/ (visited on 08/10/2021).

[149] Krisna C. Duong-Ly and Sandra B. Gabelli. "Affinity Purification of a Recombinant Protein Expressed as a Fusion with the Maltose-Binding Protein (MBP) Tag". In: *Methods in enzymology* 559 (2015), pp. 17–26. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4612625/ (visited on 08/10/2021).

[150] Kui Lin et al. "A simple method for non-denaturing purification of biotin-tagged proteins through competitive elution with free biotin". In: *BioTechniques* 68.1 (Jan. 2020). Publisher: Future Science, pp. 41–44. URL: https://www.future-science.com/doi/10.2144/btn-2019-0088 (visited on 08/10/2021).

[151] Michelle E. Kimple, Allison L. Brill, and Renee L. Pasker. "Overview of Affinity Tags for Protein Purification". In: *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]* 73 (Sept. 2013), Unit–9.9. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4527311/ (visited on 09/09/2021).

[152] Tauseef R. Butt et al. "SUMO fusion technology for difficult-to-express proteins". In: *Protein Expression and Purification* 43.1 (Sept. 2005), pp. 1–9. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7129290/ (visited on 08/10/2021).

[153] *Obtain highest purity with cobalt resin.* URL: https://www.takarabio.com/learning-centers/protein-research/his-tag-purification/tech-note-cobalt-resin (visited on 09/09/2021).

[154] M. M. Hulst et al. "Interaction of classical swine fever virus with membrane-associated heparan sulfate: role for virus replication in vivo and virulence". eng. In: *Journal of Virology* 75.20 (Oct. 2001), pp. 9585–9595.

[155] Munir Iqbal, Helen Flick-Smith, and John W. YR 2000 McCauley. "Interactions of bovine viral diarrhoea virus glycoprotein Erns with cell surface glycosaminoglycans". In: *Journal of General Virology* 81.2 (). Publisher: Microbiology Society, pp. 451–459. URL: https://www.microbiologyresearch.org/content/journal/jgv/10.1099/0022-1317-81-2-451 (visited on 08/22/2021).

[156] Monika Gooz. "ADAM-17: The Enzyme That Does It All". In: *Critical reviews in biochemistry and molecular biology* 45.2 (Apr. 2010), pp. 146–169. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2841225/ (visited on 10/06/2021).

[157] Dylan R. Edwards, Madeleine M. Handsley, and Caroline J. Pennington. "The ADAM metalloproteinases". In: *Molecular Aspects of Medicine* 29.5 (Oct. 2008), pp. 258–289. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7112278/ (visited on 10/06/2021).

[158] R. A. Black et al. "A metalloproteinase disintegrin that releases tumour-necrosis factor-alpha from cells". eng. In: *Nature* 385.6618 (Feb. 1997), pp. 729–733.

[159] M. L. Moss et al. "Cloning of a disintegrin metalloproteinase that processes precursor tumour-necrosis factor-alpha". eng. In: *Nature* 385.6618 (Feb. 1997), pp. 733–736.

[160] Friederike Zunke and Stefan Rose-John. "The shedding protease ADAM17: Physiology and pathophysiology". eng. In: *Biochimica Et Biophysica Acta. Molecular Cell Research* 1864.11 Pt B (Nov. 2017), pp. 2059–2070.

[161] Tsukasa Seya et al. "Human membrane cofactor protein (MCP, CD46): multiple isoforms and functions". en. In: *The International Journal of Biochemistry & Cell Biology* 31.11 (Nov. 1999), pp. 1255–1260. URL: https://www.sciencedirect.com/science/article/pii/S1357272599000928 (visited on 06/19/2021).

[162] M K Liszewski, T W Post, and J P Atkinson. "Membrane Cofactor Protein (MCP or CD46): Newest Member of the Regulators of Complement Activation Gene Cluster". In: *Annual Review of Immunology* 9.1 (Apr. 1991). Publisher: Annual Reviews, pp. 431–455. URL: https://www.annualreviews.org/doi/10.1146/annurev.iy.09.040191.002243 (visited on 06/19/2021).

[163] Roberto Cattaneo. "Four Viruses, Two Bacteria, and One Receptor: Membrane Cofactor Protein (CD46) as Pathogens' Magnet". In: *Journal of Virology* 78.9 (May 2004), pp. 4385–4388. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC387720/ (visited on 10/06/2021).

[164] M Kathryn Liszewski and John P Atkinson. "Membrane cofactor protein (MCP; CD46): deficiency states and pathogen connections". en. In: *Current Opinion in Immunology* 72 (Oct. 2021), pp. 126–134. URL: https://www.sciencedirect.com/science/article/pii/S0952791521000443 (visited on 06/19/2021).

[165] Xavier Robert and Patrice Gouet. "Deciphering key features in protein structures with the new ENDscript server". In: *Nucleic Acids Research* 42.W1 (July 2014), W320–W324. URL: https://doi.org/10.1093/nar/gku316 (visited on 10/04/2021).

[166] P. D. Adams et al. “PHENIX: a comprehensive Python-based system for macromolecular structure solution”. en. In: *Acta Crystallographica Section D: Biological Crystallography* 66.2 (Feb. 2010), pp. 213–221. URL: //scripts.iucr.org/cgi-bin/paper?dz5186 (visited on 07/27/2019).

[167] J. Pražnikar et al. “Averaged kick maps: less noise, more signal... and probably less bias”. en. In: *Acta Crystallographica Section D: Biological Crystallography* 65.9 (Sept. 2009). Number: 9 Publisher: International Union of Crystallography, pp. 921–931. URL: http://scripts.iucr.org/cgi-bin/paper?dz5160 (visited on 01/06/2022).

[168] M. D. Winn et al. “Overview of the CCP4 suite and current developments”. en. In: *Acta Crystallographica Section D: Biological Crystallography* 67.4 (Apr. 2011), pp. 235–242. URL: http://scripts.iucr.org/cgi-bin/paper?dz5219 (visited on 07/27/2019).

[169] T W Post et al. “Membrane cofactor protein of the complement system: alternative splicing of serine/threonine/proline-rich exons and cytoplasmic tails produces multiple isoforms that correlate with protein phenotype.” In: *Journal of Experimental Medicine* 174.1 (July 1991), pp. 93–102. URL: https://doi.org/10.1084/jem.174.1.93 (visited on 06/19/2021).

[170] Ramneek Gupta and Søren Brunak. “Prediction of glycosylation across the human proteome and the correlation to protein function”. eng. In: *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* (2002), pp. 310–322.

[171] César Santiago et al. “Structure of the measles virus hemagglutinin bound to the CD46 receptor”. en. In: *Nature Structural & Molecular Biology* 17.1 (Jan. 2010). Bandiera_abtest: a Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group, pp. 124–129. URL: https://www.nature.com/articles/nsmb.1726 (visited on 06/19/2021).

[172] *Robetta.bakerlab*. URL: https://robetta.bakerlab.org/ (visited on 09/18/2021).

[173] Andrew W. Senior et al. “Improved protein structure prediction using potentials from deep learning”. en. In: *Nature* 577.7792 (Jan. 2020), pp. 706–710. URL: http://www.nature.com/articles/s41586-019-1923-7 (visited on 09/26/2021).

[174] Ramneek Gupta. “Prediction of glyccoorsryellaattiioonnatcoropsrsottheienhfuumncatniopnroteome and the”. en. In: *Pacific Symposium on Biocomputing* (2002), p. 13.

[175] Evgeny Krissinel and Kim Henrick. “Inference of macromolecular assemblies from crystalline state”. eng. In: *Journal of Molecular Biology* 372.3 (Sept. 2007), pp. 774–797.

[176] Roman A. Laskowski and Mark B. Swindells. “LigPlot+: multiple ligand-protein interaction diagrams for drug discovery”. eng. In: *Journal of Chemical Information and Modeling* 51.10 (Oct. 2011), pp. 2778–2786.

[177] Stephen C Harrison. “Viral membrane fusion”. In: *Nature structural & molecular biology* 15.7 (July 2008), pp. 690–698. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2517140/ (visited on 02/06/2018).

[178] Fernando Merwaiss et al. “A -Hairpin Motif in the Envelope Protein E2 Mediates Receptor Binding of Bovine Viral Diarrhea Virus”. In: *Viruses* 13.6 (June 2021), p. 1157. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8235316/ (visited on 09/26/2021).

[179] Alexander McPherson and Bob Cudney. “Searching for silver bullets: An alternative strategy for crystallizing macromolecules”. en. In: *Journal of Structural Biology* 156.3 (Dec. 2006), pp. 387–406. URL: https://www.sciencedirect.com/science/article/pii/S1047847706002899 (visited on 10/05/2021).

[180] Philip Evans and Airlie McCoy. “An introduction to molecular replacement”. In: *Acta Crystallographica Section D: Biological Crystallography* 64.Pt 1 (Jan. 2008), pp. 1–10. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2394790/ (visited on 09/27/2021).

[181] Richard J. Morris and Gérard Bricogne. “Sheldrick’s 1.2 A rule and beyond”. eng. In: *Acta Crystallographica. Section D, Biological Crystallography* 59.Pt 3 (Mar. 2003), pp. 615–617.

[182] J. C. Kendrew et al. “A three-dimensional model of the myoglobin molecule obtained by x-ray analysis”. eng. In: *Nature* 181.4610 (Mar. 1958), pp. 662–666.

[183] Wayne A. Hendrickson. “Evolution of diffraction methods for solving crystal structures”. eng. In: *Acta Crystallographica. Section A, Foundations of Crystallography* 69.Pt 1 (Jan. 2013), pp. 51–59.

[184] Garry L. Taylor. “Introduction to phasing”. In: *Acta Crystallographica Section D: Biological Crystallography* 66.Pt 4 (Apr. 2010), pp. 325–338. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2852296/ (visited on 09/27/2021).

[185] W. A. Hendrickson, J. R. Horton, and D. M. LeMaster. “Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure”. eng. In: *The EMBO journal* 9.5 (May 1990), pp. 1665–1672.

[186] Wayne A. Hendrickson and Martha M. Teeter. “Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur”. eng. In: *Nature* 290.5802 (Mar. 1981), pp. 107–113.

[187] A. Wagner et al. “In-vacuum long-wavelength macromolecular crystallography”. en. In: *Acta Crystallographica Section D: Structural Biology* 72.3 (Mar. 2016). Number: 3 Publisher: International Union of Crystallography, pp. 430–439. URL: //scripts.iucr.org/cgi-bin/paper?ba5241 (visited on 09/27/2021).

[188] Liisa Holm and Laura M. Laakso. “Dali server update”. eng. In: *Nucleic Acids Research* 44.W1 (July 2016), W351–355.

[189] Roman A. Laskowski et al. “PDBsum: Structural summaries of PDB entries”. en. In: *Protein Science* 27.1 (2018). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pro.3289, pp. 129–134. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.3289 (visited on 01/05/2022).

[190] G. M. Sheldrick. "Experimental phasing with SHELXC/D/E: combining chain tracing with density modification". en. In: *Acta Crystallographica Section D: Biological Crystallography* 66.4 (Apr. 2010). Number: 4 Publisher: International Union of Crystallography, pp. 479–485. URL: http://scripts.iucr.org/cgi-bin/paper?ba5143 (visited on 10/04/2021).

[191] Alexey G. Kikhney and Dmitri I. Svergun. "A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins". en. In: *FEBS Letters*. Dynamics, flexibility, and intrinsic disorder in protein assemblies 589.19, Part A (Sept. 2015), pp. 2570–2577. URL: https://www.sciencedirect.com/science/article/pii/S0014579315007279 (visited on 09/29/2021).

[192] Irene Boo et al. "Distinct roles in folding, CD81 receptor binding and viral entry for conserved histidine residues of hepatitis C virus glycoprotein E1 and E2". eng. In: *The Biochemical Journal* 443.1 (Apr. 2012), pp. 85–94.

[193] Max Renner et al. "Flavivirus maturation leads to the formation of an occupied lipid pocket in the surface glycoproteins". en. In: *Nature Communications* 12.1 (Feb. 2021), p. 1238. URL: https://www.nature.com/articles/s41467-021-21505-9 (visited on 10/07/2021).

[194] M. Adrian et al. "Cryo-electron microscopy of viruses". eng. In: *Nature* 308.5954 (Mar. 1984), pp. 32–36.

[195] Xiao Fan et al. "Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution". en. In: *Nature Communications* 10.1 (June 2019), p. 2386. URL: https://www.nature.com/articles/s41467-019-10368-w (visited on 09/29/2021).

[196] Ka Man Yip et al. "Atomic-resolution protein structure determination by cryo-EM". eng. In: *Nature* 587.7832 (Nov. 2020), pp. 157–161.

[197] Lori A. Passmore and Christopher J. Russo. "Specimen preparation for high-resolution cryo-EM". In: *Methods in enzymology* 579 (2016), pp. 51–86. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5140023/ (visited on 09/29/2021).