University of Glasgow

Devlin, Stephen (2026) *Length polymorphisms in the hypervariable region-1 of hepatitis C virus E2: a novel regulator of viral entry and neutralisation*. PhD thesis.

https://theses.gla.ac.uk/86281/

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

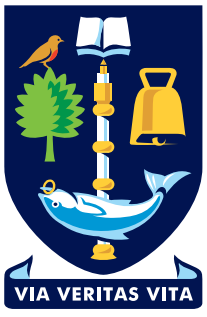
Enlighten: Theses
https://theses.gla.ac.uk/
research-enlighten@glasgow.ac.uk

Length Polymorphisms in the Hypervariable Region-1 of Hepatitis C Virus E2: A Novel Regulator of Viral Entry and Neutralisation

Stephen Devlin

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

College of Medical, Veterinary and Life Sciences
University of Glasgow



University of Glasgow

April 2026

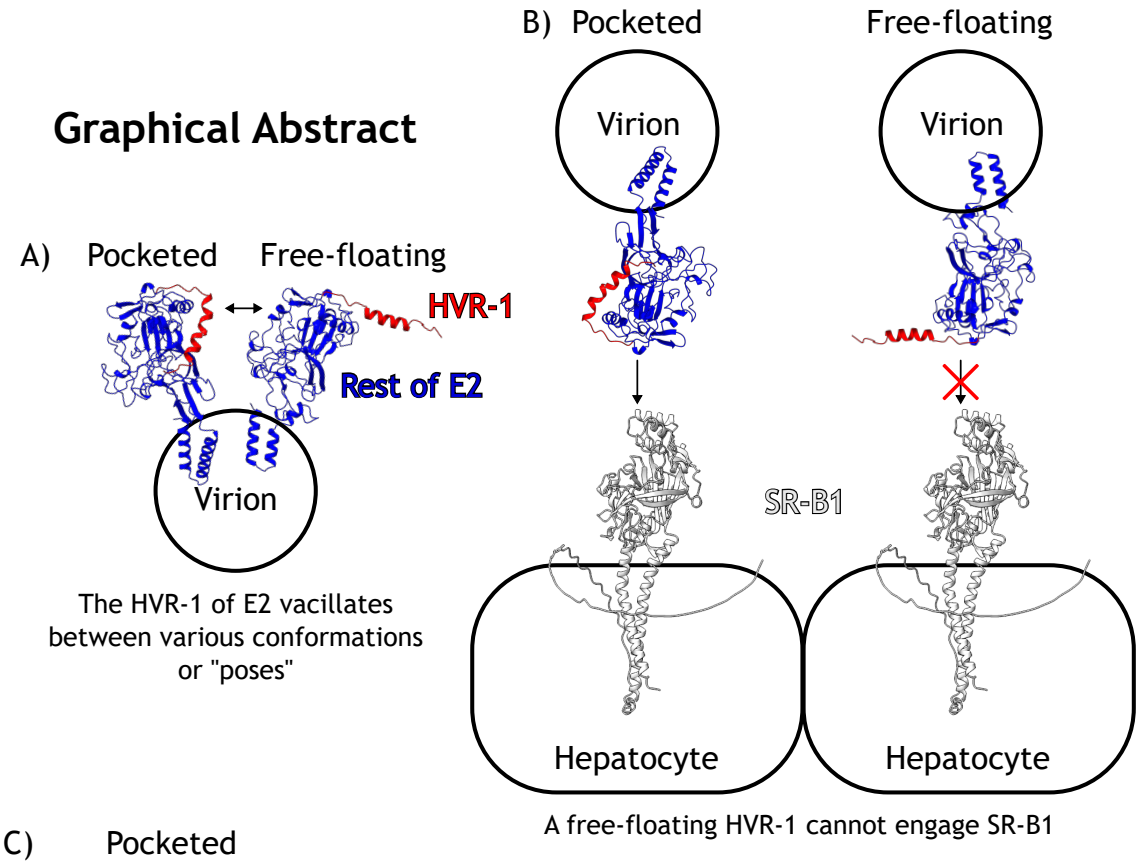
Abstract

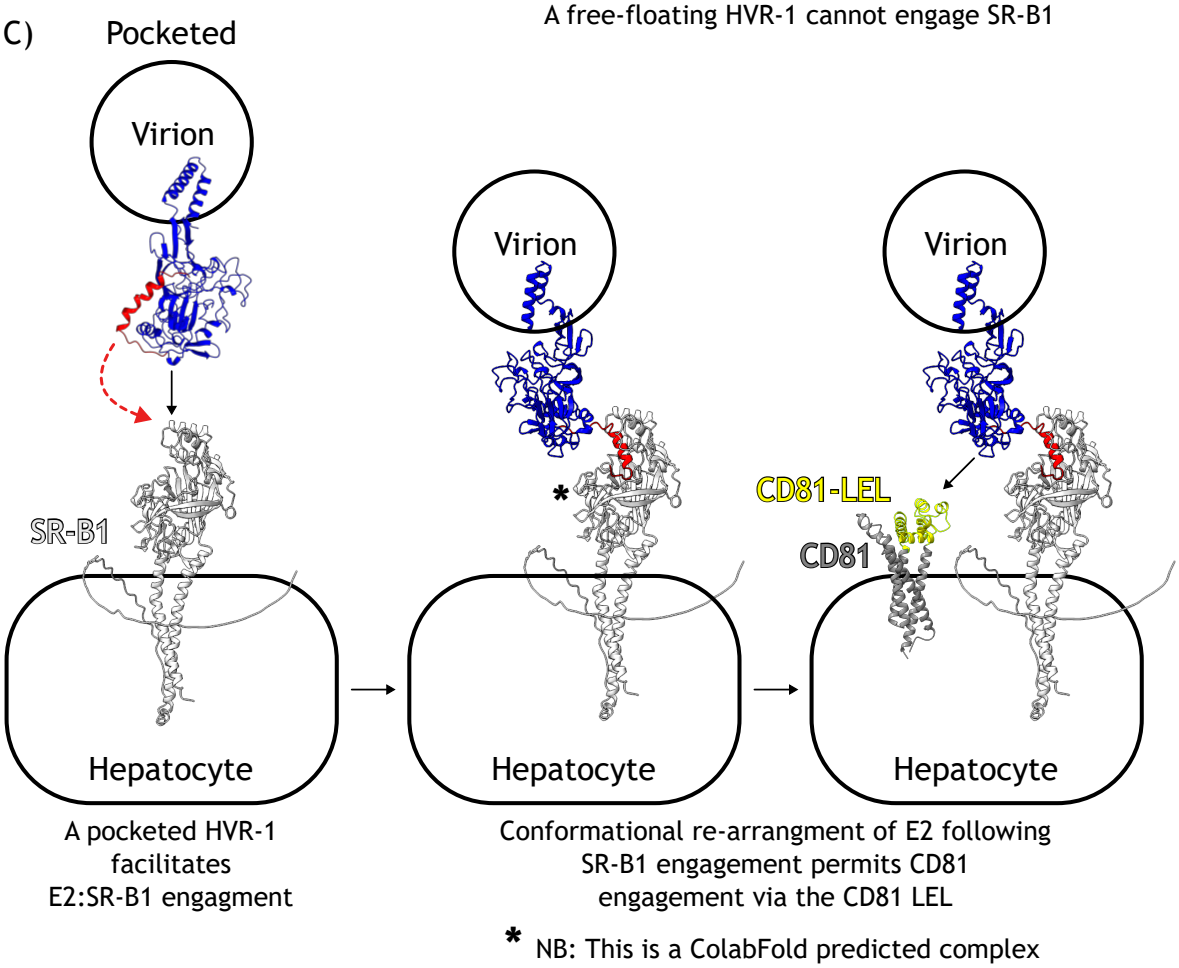
Currently 50 million people globally are living with **Hepatitis C Virus** (HCV). Whilst direct-acting antivirals offer a cure, there is no vaccine for HCV. One major barrier to the development of an effective HCV vaccine is the lack of mechanistic understanding of E1E2, the HCV viral entry proteins. Of particular importance for vaccine design is the **Hypervariable Region-1** (HVR-1) of the E2 glycoprotein. HVR-1 is an intrinsically disordered tail which mediates entry and immune evasion through an unknown mechanism. Through state-of-the-art machine learning-enabled protein structure prediction, bioinformatic analysis, and *in vitro* experiments, we investigated how length polymorphisms in the HVR-1 contribute to phenotype.

We analysed 30,289 sequences covering >25 % of the E2 protein. To examine distribution of length polymorphisms across HCV genotypes, 9,103 unique HVR-1 sequences were extracted from this dataset and assessed. To evaluate HVR-1 conformation, we modelled 105 E2 proteins, and 96 E1E2 protein complexes using different protein structure prediction software platforms. To investigate how length polymorphisms in the HVR-1 affect virus phenotype, we cloned HVR-1s of varying length into the prototypical J6 HCV to produce **HCV pseudoparticles** (HCVpp). We evaluated the infectivity, receptor dependency, and antibody sensitivity of this panel of HCVpp. Moreover, we produced sE2 proteins with the same length polymorphisms in the HVR-1 and interrogated their ability to bind SR-B1.

This study found that the HVR-1 is predicted to be positioned in distinct conformations, length polymorphisms are common across all HCV genotypes, and length polymorphisms can modulate infectivity, receptor dependency, and neutralisation. Additionally, the sE2 experiments demonstrated that SR-B1 dependency and SR-B1 binding are uncoupled. These data allowed us to pull apart some of the molecular mechanism of the HVR-1. Altogether, this work suggests that the ability of the HVR-1 to adopt certain conformational states may correlate with phenotype. Moreover, length polymorphisms represent one way for HCV to fine tune the function of the HVR-1

Graphical Abstract





Contents

1. Introduction 31

1. 1. HCV – A Global Disease 31

1. 1. 1. HCV Genotype Distribution 34

1. 1. 2. Transmission of HCV 37

1. 1. 3. Diagnosis and Prognosis 41

1. 2. The HCV Life Cycle 43

1. 2. 1. The HCV Virion 43

1. 2. 2. HCV Entry and Fusion 44

1. 2. 3. HCV Translation – Protein Synthesis 50

1. 2. 4. Polyprotein Processing 52

1. 2. 5. HCV Genome Replication – Building the Replication Organelle 54

1. 2. 6. HCV Genome Replication 57

1. 2. 7. Particle Assembly and Release 59

1. 3. Direct-acting Antivirals 61

1. 3. 1. Overview of types of DAAs 63

1. 3. 2. HCV Resistance to DAAs 66

1. 3. 3. DAAs: The Imperfect Solution 68

1. 4. Vaccine Development 69

1. 4. 1. Development of a T-cell Vaccine 71

1. 4. 2. Development of a B-cell Vaccine 73

1. 5. Viral Fusion Mechanisms 78

1. 5. 1. Class I Viral Fusion Mechanisms 80

1. 5. 2. Class II Viral Fusion Mechanisms 81

1. 5. 3. Class III Viral Fusion Mechanisms 82

1. 6. The E1E2 Glycoproteins – Evidence for a Novel Fusion Mechanism and Insights for Vaccine Design 85

1. 6. 1. E1 85

1. 6. 2. E2 90

1. 6. 2. The E1E2 Complex 96

1. 7. The HVR-1 102

1. 8. Aims and Objectives 106

2. Methods and Materials 109

2. 1. Bioinformatic Analysis of the HVR-1 of HCV 109

2. 1. 1. HCV-Glue: Retrieving Sequence Information 110

2. 1. 2. Extracting HVR-1 Sequence Information 111

2. 1. 3. Preparing E1E2/E2 Sequences for Structural Prediction 115

2. 1. 4. Protein Structure Prediction.....117

2. 1. 5. Comparing the Protein Structure Prediction Outputs.....118

2. 1. 6. Phylogenetics.....121

2. 1. 7. Sequence Logos.....122

2. 2. *In Vitro* Analysis of Length Polymorphisms in HVR-1 123

2. 2. 1. Cell Lines 123

2. 2. 2. Pseudoparticle Production.....124

2. 2. 3. Generating Length Polymorphism-containing Mutant E2 Constructs 125

2. 2. 4. Pseudoparticle Infection Assays 128

2. 2. 5. Pseudoparticle Receptor Dependency Assays.....129

2. 2. 6. Pseudoparticle Neutralisation Assays – Polyclonal Antibodies.....130

2. 2. 7. Pseudoparticle Neutralisation Assays – Monoclonal Antibodies 131

2. 2. 8. Lentiviral Vector Production 132

2. 2. 9. sE2 Production, Purification, and Buffer Exchange.....133

2. 2. 10. sE2 Binding Assay.....135

2. 2. 11. Plasmid Cloning for smFRET System Development.....137

2. 2. 12 Pseudoparticle Production for smFRET System Development.....139

2. 2. 13. Statistical Analyses.....140

2. 2. 14. Western Blots 142

3. Bioinformatic Analysis of the HCV HVR-1.....146

3. 1. The Protein Folding Problem and Protein Structure Prediction.....146

3. 1. 1. ColabFold Structure Prediction and Comparisons.....152

3. 1. 2. Comparisons Between Protein Folding Software Packages.....163

3. 2. Investigating Length Polymorphisms in the HVR-1 176

3. 2. 1. Prevalence of Length Polymorphisms in the HVR-1.....176

3. 2. 2. Universality of Length Polymorphisms in the HVR-1.....181

3. 3. Investigating HVR-1 Composition 184

3. 4. Concluding Remarks 187

4. Assessing the Phenotypic Consequences of Length Polymorphisms in the HVR-1 192

4. 1. HCV Pseudoparticles: Their History and Applications 194

4. 1. 1. The History and Development of HCV Pseudoparticles 194

4. 1. 2. Comparison of HCV Pseudoparticles with HCV Cell Culture Particles and Applications of HCV Pseudoparticles 196

4. 2. Comparing the Controls.....198

4. 2. 1. Comparing the Controls – The Results.....202

4. 2. HCVpp Infectivity.....206

4. 2. 1. J6 WT Backbone HCVpp Infectivity.....206

4. 2. 2. I438V A524T J6 Backbone HCVpp Infectivity.....210

4. 2. 3. Comparing Infectivity Across Backbones.....214

4. 3. E2 Western Blots 216

4. 4. HCVpp Receptor Dependency.....218

4. 4. 1. Non-Permissive KO Cell Lines.....223

4. 4. 2. Huh7 WT Cells 226

4. 4. 3. Huh7 LDLR KO Cells.....228

4. 4. 4. Huh7 SR-B1/LDLR DKO Cells 231

4. 4. 5. Huh7 SR-B1 KO Cells.....234

4. 5. HCVpp Neutralisation – Polyclonal Antibodies 237

4. 5. 1. J6 WT Backbone HCVpp Neutralisation – Polyclonal Antibodies 239

4. 5. 2. I438V A524 T J6 Backbone HCVpp Neutralisation – Polyclonal Antibodies 244

4. 6. HCVpp Neutralisation – Monoclonal Antibodies.....248

4. 6. 1. HCVpp Neutralisation – J6.36.....250

4. 6. 2. HCVpp Neutralisation – AR3C.....254

4. 6. 3. HCVpp Neutralisation – AP33 259

4. 6. 4. HCVpp Neutralisation – AR4A.....263

4. 7. Concluding Remarks 267

4. 7. 1. Infectivity.....268

4. 7. 2. Receptor Dependency.....270

4. 7. 3. Polyclonal Neutralisations.....274

4. 7. 4. Monoclonal Neutralisations 276

4. 7. 5. Hierarchy of Mutations.....279

4. 7. 6. Final Conclusions.....281

5. Investigating How the HVR-1 Functions.....283

5. 1. Predicting the HVR-1:SR-B1 Interface 285

5. 2. Interrogating HVR-1:SR-B1 Binding 289

5. 3. Investigating the Relationship Between HVR-1 Pocketing and SR-B1 Binding 298

5. 4. Design and Development of an HCV E2 smFRET System 304

5. 5. Concluding Remarks 309

6. Discussion 314

6. 1. Aims and Objectives Revisited 315

6. 2. Current Knowledge Gaps 316

6. 2. 1. The HVR-1 *in situ* with the rest of the E2 Protein 316

6. 2. 2. Systematic Analysis of HVR-1 Sequences.....318

6. 2. 3. How does the HVR-1 Function?.....321

6. 3. A Proposed Model of HVR-1 Function 322

6. 3. 1. Length Polymorphisms Do Alter HVR-1 Function 322

6. 3. 2. The I438V A524T Mutations and Length Polymorphisms in the HVR-1 323

6. 3. 3. HVR-1 Conformation and SR-B1 Dependency.....325

6. 3. 4. HVR-1 Length and SR-B1 Binding.....327

6. 3. 5. A Proposed Model of HVR-1 Function 329

6. 4. Limitations and Future Work 331

6. 5. Final Remarks.....334

List of Tables

Chapter 1 - Introduction

Table 1. 1.	Genotype prevalence percentages used to generate figure 1. 2.
Table 1. 2.	Details of which countries adopted NAT screening for HCV in which year.
Table 1. 3.	Details of currently approved HCV DAAs in the UK and Europe.
Table 1. 4.	Summary of some of the known anti-E2 mAbs.

Chapter 2 - Methods and Materials

Table 2. 1.	Details of the bins used to separate out the initial seven .fasta files from HCV-Glue
Table 5. 2.	Details of substitution models used for each of the trees.
Table 2. 6.	Details of Cell Lines Used for In Vitro Experiments.
Table 2. 4.	Overview of Additional Cut Sites Added to pD607 J6 WT E1E2 Plasmid
Table 2. 5.	Summary of Statistical Analyses
Table 2. 6.	Summary Information for Primary Antibodies Used for Western Blotting.
Table 2. 7.	Summary Information for Secondary Antibodies Used for Western Blotting.
Table 2. 8.	Summary Information for Loading Control Antibodies Used for Western Blotting.

Chapter 3 - Bioinformatic Analysis of the HCV HVR-1

Table 3. 1.	Modal HVR-1 Lengths by Genotype
Table 3. 2.	Breakdown of Length Polymorphisms and Modal Length HVR-1 Sequences by Genotype.

Chapter 4 - Assessing the Phenotypic Consequences of Length Polymorphisms in the HVR-1

Table 4. 7. Summary of the Control Phenotypes.

Table 4. 8. Fold Change of the RLU Figures for the Internal Insertion I438V A524T J6-backbone Length Polymorphism-containing HCVpp versus Empty Vector in the Huh7 OCLN KO Cell Line.

List of Figures

Chapter 1 - Introduction

Figure 1. 1.	HCV - A Global Overview
Figure 1. 2.	Global HCV Genotype Distribution
Figure 1. 3.	Schematic Overview of HCV Entry and Fusion into Hepatocytes
Figure 1. 4.	Overview of HCV Genome Architecture and Viral Protein Membrane Topology
Figure 1. 5.	Overview of the HCV Replication Organelle Structure
Figure 1. 6.	Overview of Currently Licenced DAAs Available in Europe and Their Targets
Figure 1. 7.	Schematic Overview of Viral Fusion Mechanisms
Figure 1. 8.	Pre- and Post-fusion Structures of Class I, II, and III Viral Glycoproteins
Figure 1. 9.	X-Ray Diffraction Crystal Structure of the N-Terminus of H77 E1
Figure 1. 10.	X-Ray Diffraction Crystal Structure of Isolate H HCV E2
Figure 1. 11.	X-Ray Diffraction Crystal Structure of HCV J6 Core E2
Figure 1. 12.	Cryo-EM Structure of Genotype 1a strain AMs0232 E1E2
Figure 1. 13.	Dimer of Dimers Cryo-EM Structure of Genotype 3a HCV S52 E1E2

Chapter 2 - Methods and Materials

Figure 2. 1.	Schematic Representation of the J6-backbone Length Polymorphism Constructs and Controls
Figure 2. 2.	Schematic Representation of the I438V A524T J6-backbone Length Polymorphism Constructs and Controls

Chapter 3 - Bioinformatic Analysis of the HCV HVR-1

Figure 3. 1.	ColabFold Models of J6 WT E2
Figure 3. 2.	Comparison of Cryo-EM HCV S52 Structure with the ColabFold J6 WT E2 Models
Figure 3. 3.	ColabFold Models of E2 Proteins from Each of the Seven HCV Genotypes
Figure 3. 4.	Comparing the Predicted HVR-1 Positioning Across Different Protein Folding Software Packages - Ranked 1 Models
Figure 3. 5.	Relative 3D Positions of the Alpha Carbon of the N-terminus HVR-1 Residues as Predicted by Different Protein Folding Software Arranged Around the ColabFold J6 WT E2 Predicted Structure
Figure 3. 6.	Relative 3D Positions of the Alpha Carbon of the N-terminus HVR-1 Residues as Predicted by Different Protein Folding Software Arranged Around the ColabFold J6 WT E1E2 Predicted Structure.
Figure 3. 7.	Investigating the Prevalence of Length Polymorphisms in the HVR1 Across All HCV Genotypes
Figure 3. 8.	Investigating Universality of Length Polymorphisms Across All HCV Genotypes
Figure 3. 9.	Sequence Logos for All Extracted HVR-1 Sequences

Chapter 4 - Assessing the Phenotypic Consequences of Length Polymorphisms in the HVR-1

Figure 4. 1.	Comparing Controls
Figure 4. 2.	J6 WT Backbone Length Polymorphism-containing HCVpp Infectivity
Figure 4. 3.	I438V A524T J6 Backbone Length Polymorphism-containing HCVpp Average Infectivity

Figure 4. 4. Comparing Infectivity Across J6 WT- and I438V A524T J6-based HCVpp

Figure 4. 5. E2 Western Blots

Figure 4. 6. Infectivity of the HCVpp in Non-permissive Huh7 KO Cell Lines

Figure 4. 7. Infectivity of the HCVpp used in the Receptor Dependency Assays

Figure 4. 8. Infectivity of the HCVpp in Huh7 LDLR KO Cells

Figure 4. 9. Infectivity of the HCVpp in Huh7 SR-B1LDLR DKO Cells

Figure 4. 10. Infectivity of the HCVpp in Huh7 SR-B1 KO

Figure 4. 11. J6-backbone Length Polymorphism-containing HCVpp Polyclonal Neutralisation

Figure 4. 12. I438V A524T J6-backbone Length Polymorphism-containing HCVpp Polyclonal Neutralisation

Figure 4. 13. Monoclonal Neutralisation of the HCVpp using J6.36

Figure 4. 14. Monoclonal Neutralisation of the HCVpp using AR3C

Figure 4. 15. Monoclonal Neutralisation of the HCVpp using AP33

Figure 4. 16. Monoclonal Neutralisation of the HCVpp using AR4A

Chapter 5 - Investigating How the HVR-1 Functions

Figure 5. 1. Predicted E1E2 SR-B1 Interactions

Figure 5. 2. ColabFold Model of the J6 WT sE2 Protein

Figure 5. 3. Representative FACS Dot Plots

Figure 5. 4. Investigating the Binding Efficacy of the sE2 Constructs

Figure 5. 5. Mitigating HVR-1 Pocketing

Figure 5. 6. SR-B1 Binding Efficacy of the J6 and H77 Pocketing-mitigated Constructs

Figure 5. 7. Developing an HCV E2 smFRET System

I dedicate this thesis to my mum, Jacqueline France BSc (Hons). Without your unwavering support and example, none of this would have been possible.

And to the 14-year-old me, who dreamt of being a scientist.

We did it, wee man.

Acknowledgements

I feel I would be being dishonest by omission if I did not acknowledge quite how difficult I have found this PhD, and life to be truthful, the past few years. Therefore, I want to take this moment to say that this thesis is truly the product of the support, encouragement, and help of the people who have been there for me the last few years.

This would not have been possible without each and every single one of you.

Firstly, I want to acknowledge my primary supervisor Dr Joe Grove. From the very first meeting in the Barn, drawing crude E1 “scribbles”, your enthusiasm for viral glycoproteins was palpable and infectious. I knew then that if you let me join your lab, I would have a supervisor who was truly excited by science – curious and passionate, and with a level of knowledge and expertise that I still find astounding. I think we would both agree that I have struggled throughout these past few years, but your understanding and encouragement are what has truly gotten me through this PhD. Thank you, Joe, for your commitment to my development as a scientist, for pushing me to be the best I can be, and for seeing me through to the end of this.

Secondly, I want to thank Professor Ed Hutchinson for both being my secondary supervisor, and for seeing me through my first rotation at the CVR – a wonderfully supportive and encouraging start to my PhD. One thing I genuinely appreciated from my time in the Hutchinson group was the encouragement and faith you had in me Ed, to allow me to take on a project that pushed me to truly develop as a technical scientist. Coming from the world of mycology, I was hesitant to truly dive into the world of virology. However, Ed, with your guidance and belief in my abilities, I felt like I was able to take on the challenge. It truly gave me a confidence in the lab I do not think I would have attained outwith your group and for that I thank you.

Moreover, I want to thank both Dr Adam Fletcher and Dr Olwyn Bryon for their roles as my assessors throughout this PhD project. In my first year, I was nervous going into my first annual review. By my third year, I was excited to engage in a rigorous yet supportive conversation about the project and where it was going. I thank you both for those conversations, and for your guidance and support. Additionally, I do want to thank you both for allowing me to submit much more than the recommended 3,000 word reports – I know it was a lot to read but those annual reviews were incredibly helpful to me in the writing process for the thesis itself so thank you for letting me do that.

From the Grove lab itself, I want to thank each and every member of the lab who has helped me this project throughout the past few years.

I thank Dr Mickey Oliver for both teaching me and for your friendship. I cannot pretend for one moment that this project would be even half as good as it ended up being without your expertise and input. Here is to many more concerts, coffee shops, and road trips (wherein we might occasionally discuss some proteins…)!

Dr Diego Cantoni, my fellow Raptor, I thank you for your help with Western blots, going for fresh air every half an hour, and for the many, many laughs we have shared over the years. It has been my honour to have you on my side in both the lab and on the rugby pitch, and given how competitive you are, I do not think I would want it any other way! (Note, if I knew enough Latin, this paragraph would be written in Latin just for you!)

I also want to thank Dr Kami Toon, I thank you for challenging me on my science, and presenting me with papers and research which helped to expand my understanding of HCV and pseudoparticles – I'm not quite convinced about E2 as the fusogen but you challenged me on that and those conversations have made me a better scientist. I also want to thank you for the laughs in the lab, and for introducing me to Yummy Mummy Time at Rydal Hall.

Charlotte Lewis – I cannot wait to say I knew you when! You are an engaged and capable scientist, and I know you are going to go so, so far! You are so intelligent, so humble, and so hilarious – thank you for everything!

Sarah Little – I know you are going to go so far as well whether that is in the lab, or on Bake Off! Thank you for helping with the HCVpp Western Blots and keeping the lab right for the past few years. More importantly, thank you for all the delicious and spectacular cakes that you have made over the years, and all the amazing birthday/Christmas presents – truly undeserving of your generosity but very much appreciated! I wish you all the best for the future, and I know you are going to thrive in your PhD!

Now, I will thank you Ross Mac Conraoi, but first, I want to commend you for not throwing your laptop out the window when your presentation deleted itself in the wee hours of the morning, just hours before your first conference presentation! I know that you are going to absolutely smash your PhD with Joe and Kevin but thank you for your friendship and the many laughs during your time up in Glasgow.

I thank Dr Spyros Lytras. As Promethius gave humans fire, you gave me the skills to adequately use a computer to run bioinformatics experiments; without you none of the phylogenetic experiments would have gotten to where I wanted them to be – thank you so much!

Dr Anna Sims, truly one of my first friends at the CVR! I miss you, but I know you are thriving in Atlanta! You really welcomed me into the fold of the CVR and made me feel so much less out of place during my first rotation. You helped me understand viruses and statistics – two things I did not really get before starting this PhD - and we did through cry laughing. Thank you!

Speaking of people who welcomed me, and brought me into the fold, my thanks to Dr Lea Meyer. You had the really difficult job of teaching a former mycologist how to do tissue culture and what IAV is, and you did it with such patience and

attentiveness whilst getting your own work done, whilst also looking after the lab. Thank you so much for everything. What you taught me those first few weeks/months were indispensable to the rest of my PhD and potentially my career.

Dr Kieran Lamb – you are getting a special mention not only as a good friend but also for teaching me the nohup command. Life saver, game changer! Thank you!

I also want to give a special mention to both Dr Daniel Weir and Dr Mhairi McCormack – the better thirds(?) of my cohort. With me finally finishing this, that is our MRC cohort officially done! It has been a wild ride, and I wish you all the best for the future. You are both incredibly talented scientists, and both amazing people. I know whatever you both will do in the future, you will both be successful, I wish you both the best!

A massive thank you to both Faye Watson and Lois Mason for allowing me to partake in the Research Goes Viral podcasts/content creation – I loved every interview, video, and moment of silliness we created during my time at the CVR. Another thanks to everyone who let me interview them, and for those who participated in my nonsense!

I would also like to thank Eilidh Rivers and Hagar Sasvari, my fellow PGR reps! Being a PGR rep was rewarding but hard work, and you both made it so much easier!

Now I wish to thank the many friends I have made during my time at the CVR for the many laughs, conversations, support, and some good science: Alex Wilson, Bailey Atkinson, Cal Bentley-Abbot, Chris Davies, Igor Starinskij, Patawee Asamaphan, Elliot Moorhouse, Ewan Smith, Holly Ireland-Jackson, Ilaria Epifano, Kesley Davies, Lauren Orr, Marina Kugler, Rhiannon Heslop, Rosie Geall, Saoirse Healy.

Outwith the CVR, I want to thank my mum for your constant support throughout the last few years and for always letting me explore my passions as a child – it got me here, and this thesis is dedicated to you for that reason.

I want to say thank you to Lauren Mackay, Lauren Robin, Emma Smith, Iina Jaara, Tereza Matlova for your constant friendship for the past years/decades. Particularly over the last few years, your friendship has been a reminder at times when I needed it that the bench is not everything. You have all kept me grounded, encouraged me, and reminded that I am human. Thank you all so much.

Now, I would be remiss if I did not thank my second family: The Glasgow Raptors R.U.F.C. When the science has not been working, and the days were long, and the nights were longer, you were all there to tackle me to the ground, outrun me, and scrum alongside me. Thank you all so, so much. Rugby kept my sanity over the years when nothing else could. I thank you all, but a particular shout out the following folks: Jim and Brian, my favourite props. Doug, my favourite walking injury. Charlie, the younger brother I never had! Mikey, you are a gremlin, so thanks for that, I think! Ken, I thank you for reminding me that everything is actually crabs. Thomas for talking all things biology when no one else cared. Will, you are going to smash your degree – I know it! Thank you again to every single Raptor over the past four years.

I also want to shout out to the people I have worked with over the years. Working part-time and trying to do a PhD was difficult, but you all helped make it worth it so thank you to Conrad Hamilton, Kirsty Hadden, Laura Macaulay, Lucy Napier, and everyone at the Cameron Grill.

Lastly, but no less importantly, I want to thank Dermot Kennedy, Nightwish, and Rocky Flintstone for their contributions to culture which kept me entertained during the many late nights in the lab.

This thesis simply would not have happened without you all there for me, so thank you all.

Declaration

I declare that, except where explicit reference is made to the contribution of others, that this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Note that the following work was performed by others, and appropriate reference is made to their contributions in the text and here:

Dr Michael Oliver: Figure 5. 6. (Bar graph results only), Figure 5. 7.

Miss Sarah Little: Appendix VI, Figure VI. I. and Figure VI. III.

.

.

Signed:

Stephen Devlin

Date: April 2026

Abbreviations

+3 sIVAT	+3 N-Terminus Extension sIVAT
+3 sJ6	+3 N-Terminus Extension sJ6
+4I sIVAT	+4I Internal Insertion sIVAT
+4I sJ6	+4I Internal Insertion sJ6
-3 sJ6	-3 N-Terminus Extension sJ6
ACSL3	Acyl-CoA Synthetase 3
AI	Artificial Intelligence
AS412	Antigenic Site 412
BCR	B-cell Receptor
bNAb	Broadly Neutralising Antibody
BSL2	Biosafety Level Two
BSL3	Biosafety Level Three
CASP	Critical Assessment of Protein Structure Prediction
CDR	Complementarity-Determining Regions
ChAd3	Chimpanzee Adenovirus 3
CHO	Chinese Hamster Ovary
cLD	Cytosolic Lipid Droplets
CLDN-1	Claudin-1
CLEM	Correlative Light and Electron Microscopy
CMV	Cytomegalovirus
CR2	Complement Receptor Type 2
CRE	Cis-Acting RNA Elements
cryo-EM	Cryo-Electron Microscopy
Cα	Alpha Carbon
DAA	Direct-Acting Antivirals
DGAT1	Diacylglycerol Acetyltransferase-1
DMV	Double-Membrane Vesicles
dsRNA	Double-Stranded RNA
EASL	European Association for the Study of the Liver

EGFR	Epidermal Growth Factor Receptor
eIF2	Eukaryotic Initiation Factor 2
eIF3	Eukaryotic Initiation Factor 3
EL1	Extracellular Loop 1
EphA2	Ephrin Receptor A2
EV71	Enterovirus 71
FACS	Fluorescence-Activated Cell Sorting
FRET	Förster Resonance Energy Transfer
GAG	Glycosaminoglycan
gB	Glycoprotein B
GT1	Genotype 1
GT2	Genotype 2
GT3	Genotype 3
GT4	Genotype 4
GT5	Genotype 5
GT6	Genotype 6
GT7	Genotype 7
GTP	Guanine Triphosphate
HCV	Hepatitis C Virus
HCVcc	HCV Cell Culture Viruses
HCVpp	HCV Pseudoparticles
HIV-1	Human Immunodeficiency Virus-1
HNF4α	Hepatocyte Nuclear Factor
HSPG	Heparan Sulphate Proteoglycans
HSV-1	Herpes Simplex Virus-1
HVR-1	Hypervariable Region-1
IAV	Influenza A Virus
IDR	Intrinsically Disordered Regions
IRES	Internal Ribosome Entry Site
iTOL	Interactive Tree of Life
JFH-1	Japanese Fulminant Hepatitis-1
LCS	

LDL	Low-density Lipoprotein
LDLR	Low-Density Lipoprotein Receptor
LEL	Large Extracellular Loop
LIMP-2	Lysosomal Integral Membrane Protein 2
LMICs	Low- And Middle-Income Countries
LVP	Lipoviral Particles
mAbs	Monoclonal Antibodies
MAFFT	Multiple Alignment Using Fast Fourier Transform
MD	Molecular Dynamics
MHC	Major Histocompatibility Complex
miR-122	microRNA-122
MMV	Multimembrane Vesicles
MSA	Multiple Sequence Alignments
MSM	Men Who Have Sex with Men
MTCT	Mother-To-Child Transmission
MTP	Microsomal Triglyceride Transfer Protein
MVA	Modified Vaccinia Ankara
NAb	Neutralising Antibody
NANBH	Non-A, Non-B Hepatitis
NAT	Nucleic Acid Amplification Technology
NCBI	National Centre for Biotechnology Information
NGS	Next-Generation Sequencing
NMR	Nuclear Magnetic Resonance
NPC1L1	Niemann-Pick C1-Like 1
NS	Non-Structural
NTD	N-Terminus Domain
OCLN	Occludin-1
PAMP	Pathogen-associated Molecular Pattern
PCR	Putative Fusion Peptide-containing Region
PD-1	Programmed Cell Death-1
PDB	Research Collaboratory for Structural Bioinformatics Protein Data Bank

PegIFN	Peginterferon
PI4KIIIA	Phosphatidylinositol 4-kinase IIIα
PI4P	Phosphatidylinositol 4-phosphate
PKA	Protein Kinase A
PLA2G4	MAPK-regulated Cytosolic Phospholipase A2
PWID	People Who Inject Drugs
RBV	Ribavirin
RdRp	RNA-dependent RNA polymerase
RLU	Relative Luminescence Unit
RMSD	Root Mean Squared Deviation
RO	Replication Organelle
RTK	Receptor Tyrosine Kinases
sE2	Soluble Ectodomain E2
sIVAT	I438V A524T J6 sE2
sJ6	Wildtype J6 sE2
SL	Stem-Loop Domains
smFRET	Single-Molecule Förster Resonance Energy Transfer
SMV	Single-Membrane Vesicles
SPIEDAC	Strain-Promoted Inverse Electron-Demand Diels-Alder Cycloaddition
SR-B1	Scavenger Receptor – B1
sΔHVR-1	ΔHVR-1 J6 sE2
TAH	Transfusion-associated Hepatitis
tCD81-LEL	Large Extracellular Loop of Tamarin CD81
TCO*A	Trans-Cyclooct-2-En-L-Lysine
TFR1	Transferrin Receptor 1
TLR	Toll-like Receptor
TM	Transmembrane
tRNAi	Ternary Met-Transfer RNA
TTI	Transfusion-Transmitted Infections
UAA	Unnatural Amino Acids
UTR	Untranslated Region

VLDL	Very Low-Density Lipoproteins
VR2	Variable Region 2
VR3	Variable Region 3
vRNA	Viral RNA
VSC	Visual Studio Code
VSV	Vesicular Stomatitis Virus
WHO	World Health Organization
β-GCase	B-Glucocerebrosidase

Chapter 1

Introduction

1. Introduction

1. 1. HCV – A Global Disease

Hepatitis C Virus (HCV) is a +ssRNA hepatotropic virus belonging to the family *Flaviviridae*. Currently, the **World Health Organization** (WHO) estimates that 50 million people globally are living with the virus, with an approximately 1 million new infections each year (World Health Organisation, 2025). HCV was first discovered and classified as the causative agent of **Non-A, Non-B Hepatitis** (NANBH) in 1989 (Choo *et al.*, 1989). In 2011, the first **Direct-Acting Antivirals** (DAAs) – boceprevir and telaprevir, both NS3/4A inhibitors - were introduced to combat HCV infection (Ghany *et al.*, 2011), and then in 2016 the WHO put forward its plan to eliminate the virus by 2030 due to the global burden of disease presented by HCV (Torre *et al.*, 2024).

To eliminate HCV, the WHO focuses on reducing the number of new infections whilst also increasing diagnosis and treatment of those currently living with the virus. However, it was reported by the WHO that in 2019 only 21 % of people with chronic HCV infections globally had been diagnosed and of those, only 13 % had been treated, with the goal to increase this to 90 % diagnosed, and 80 % cured by 2030 (World Health Organisation, 2022). HCV disease burden varies between countries/global regions. According to the CDA Foundation as of 2024, Pakistan, India, Mainland China, Russia, and the United States have the greatest disease burden from HCV in terms of case numbers. However, the countries with the highest disease prevalence are Gabon (4.4 %), Mongolia (3.6 %), Pakistan (3.6 %), Ukraine (3.4 %), and Tajikistan (2.5 %) (Fig. 1.1, adapted from The CDA Foundation, (2025)).

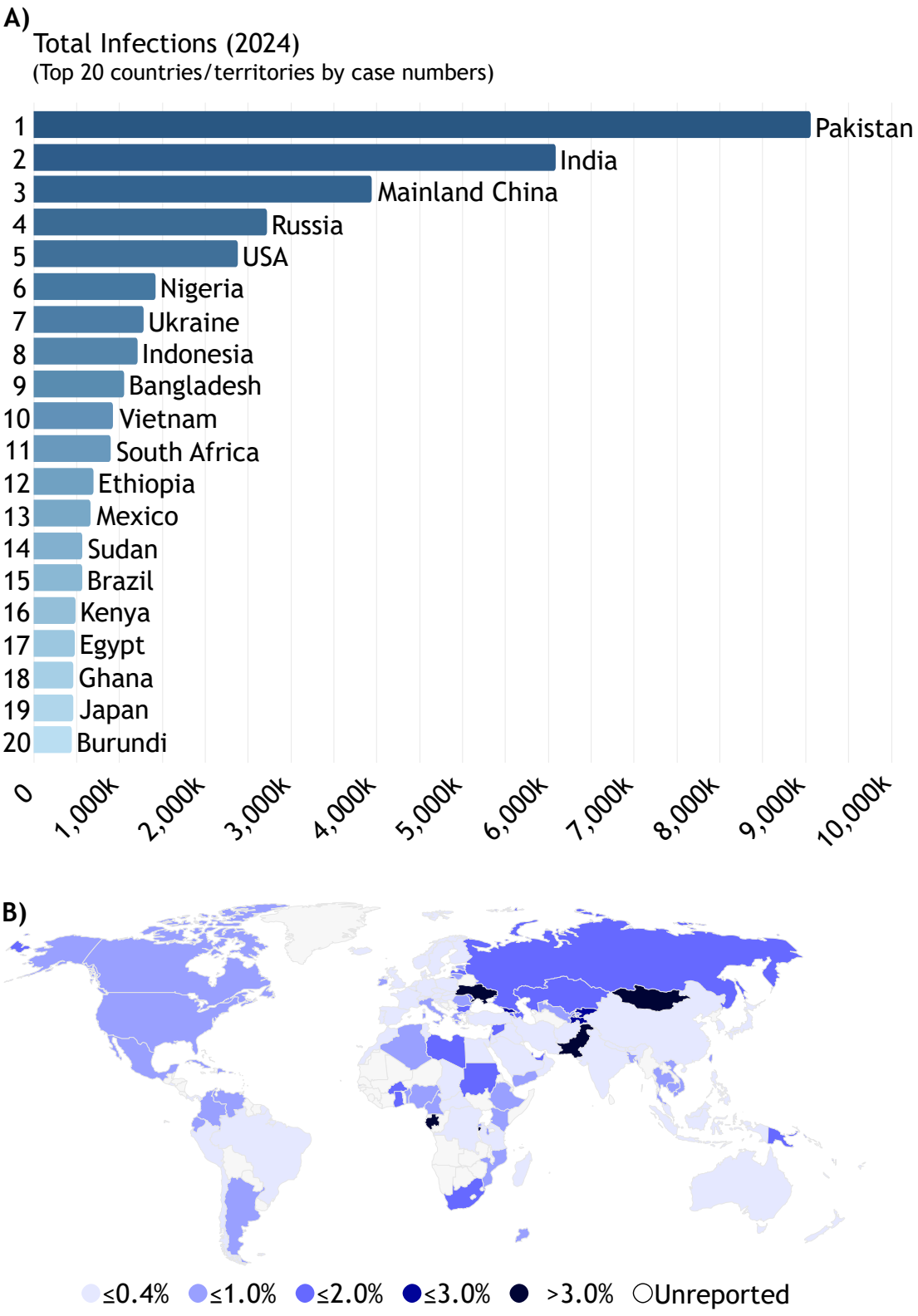


Figure 1. 1. **HCV - A Global Overview.** A) Case numbers in 2024 for the twenty countries with the highest numbers of cases of HCV. B) HCV prevalence percentages by country - the darker blue, the higher the percentage of persons living with HCV within the country. Both figures have been adapted from the Polaris Observatory based on figures from 2024 (The CDA Foundation, 2025)

However, it should be noted that due to the predominately asymptomatic and occult nature of chronic HCV infection, these figures may be an underrepresentation of actual case numbers unless HCV is actively being screened for in individual countries (Stroffolini and Stroffolini, 2024). HCV is a global problem and DAA accessibility remains inequitable on the global scale (Marshall *et al.*, 2024). Thus, efforts are still ongoing to produce an effective HCV vaccine to reduce new HCV infections and, therefore, abetting elimination of HCV by 2030.

1. 1. 1. HCV Genotype Distribution

Genotypes of HCV are distributed differently globally, a summary of which can be found in table 1. 1., and figure 1. 2 which are derived from Messina et al. (2015), and from Mohd Hanafiah et al. (2013); Messina et al. (2015); Roger et al. (2021) respectively. In general, the most prevalent genotype of HCV is genotype 1, followed by genotype 3, and then genotype 2, 4, 6, 5, and then 7 which has only been found in a few individuals from central Africa. However, a more detailed picture can be seen when considering the HCV genotype subtypes. In particular, subtypes 1a, 1b, 2a, and 3a - highly prevalent within high-income countries - are widely distributed. Collectively, these are know as the epidemic subtypes and are generally believed to have been spread within populations through blood product transfusions, iatrogenic transmission, and through **people who inject drugs** (PWID) (Paraskevis et al., 2021). The other HCV subtypes are considered to be endemic and have historically been more geographically restricted. Genotypes 4 and 7 are more prominent in central Africa, with a high number of genotype 4 cases also being found in France and Portugal, which may reflect migration patterns (Guntipalli et al., 2021). Genotype 5 is more prevalent in Southern Africa, and genotype 6 in Southeast Asia. The distribution of the various genotypes of HCV is a reflection of anthropogenic actions which have allowed the virus to spread – the history of blood transfusions, migration, and the increase in numbers of PWID are all reflected in the global distribution of HCV (Mohd Hanafiah et al., 2013; Messina et al., 2015; Roger et al., 2021; The CDA Foundation, 2025)

Table 9. 1. Genotype prevalence percentages used to generate figure 1. 2. Derived from Messina et al. 2015.

WHO GBD Region	GT1 - %	GT2 - %	GT3 - %	GT4 - %	GT6 - %	GT7 - %
Andean Latin America	90.9	1.5	7.6	0	0	0
Australasia	54.2	4.7	39.2	1.3	0	0.5
Caribbean	92.6	3.2	3.5	0.8	0	0
Central Asia	66.6	4.7	28.7	0	0	0
Central Europe	89.2	0.1	9.4	1.3	0	0
Central Latin America	71.7	19.3	8.5	0.4	0	0
Central sub-Saharan Africa	1.7	0.8	0	97.6	0	0
East Asia	58	15.3	10.4	0.1	0	16.2
Eastern Europe	65.1	4.4	30.4	0.1	0	0
Eastern sub-Saharan Africa	37.3	9.2	9.1	30.7	13.7	0
High-income Asia Pacific	74.9	24.5	0.6	0	0	0
High-income North America	75.8	12	10.4	1.2	0.1	0.6
North Africa and Middle East	27.3	0.8	6.3	65.3	0.3	0
South Asia	23.2	2.4	71.6	2.5	0.1	0.1
Southeast Asia	57	18.2	15.4	0.9	0	8.5
Southern Latin America	87	5.7	6.5	0.5	0.4	0
Southern sub-Saharan Africa	26.5	1.2	7.1	6.5	58.8	0
Tropical Latin America	69.3	3.4	26.9	0.3	0.1	0
Western Europe	59	10.8	24.8	4.9	0.5	0
Western sub-Saharan Africa	65.7	23	0	11.3	0.1	0
Totals (excludes Oceania)	46.2	9.1	30.1	8.3	0.8	5.4

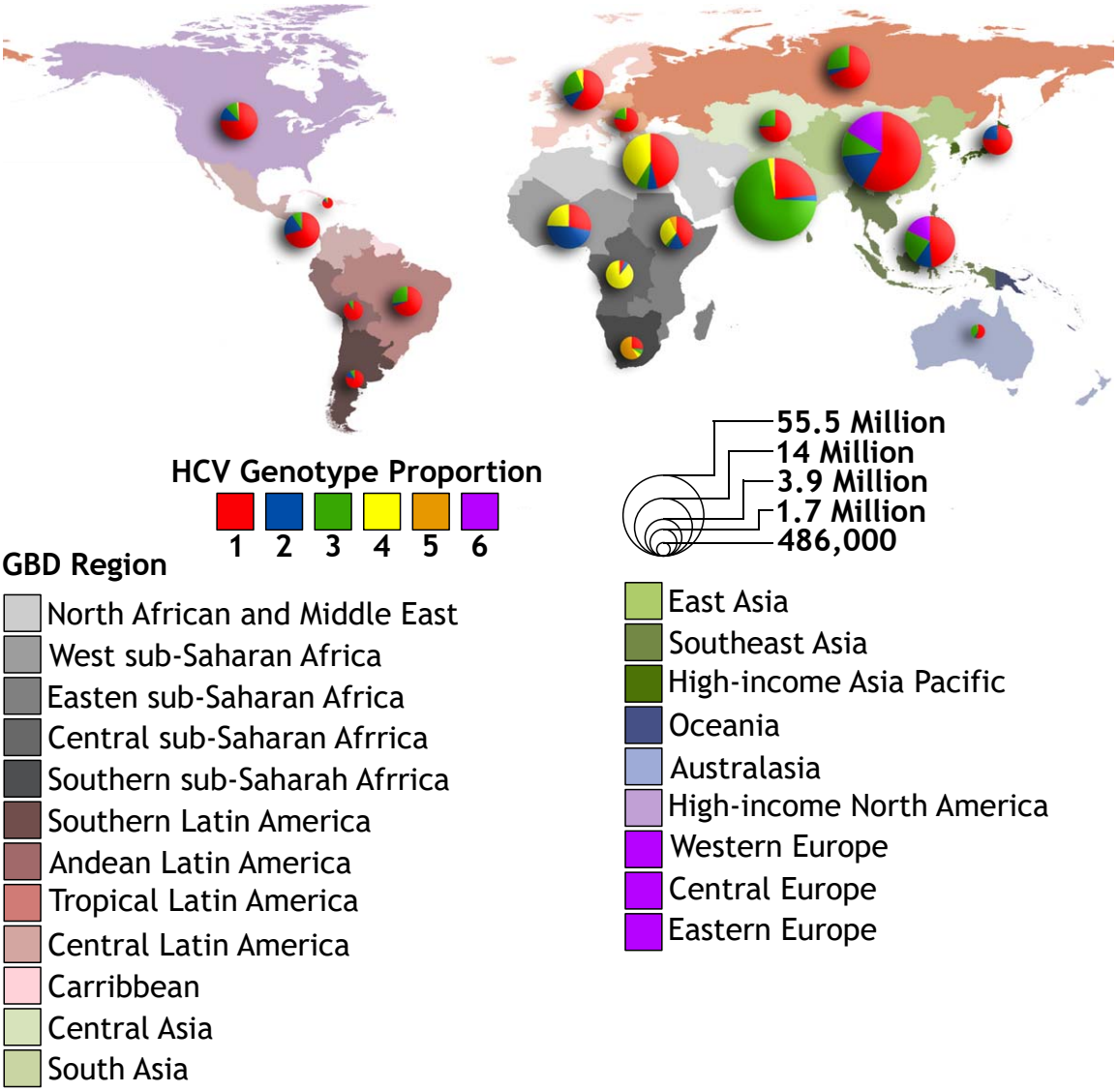


Figure 1. 2. **Global HCV Genotype Distribution.** Relative prevalence of Genotype 1 through 6 by Global Burden of Disease (GBD) regions. Pie chart size is proportional to number of seroprevalent cases in each region. NB: Genotype 7 has only been detected in a few cases in Central Africa. Figure adapted from Messina *et al.* (2015), itself adapted from figures from Hanafiah *et al.* (2013).

1. 1. 2. Transmission of HCV

HCV is a bloodborne virus and is transmitted via blood-to-blood contact. Moreover, HCV has been shown in chimpanzees to have a very low minimum infective dose of approximately 10 viral particles i.e. exposure to ten HCV viral particles is sufficient to infect 50 % of people exposed (Shata *et al.*, 2003). Historically, the major route of transmission of HCV virus was through the transfusion of unscreened, HCV-contaminated blood/blood products. Today, however, the most prominent routes of transmission of HCV include transfusion of HCV-contaminated blood/blood products in **low- and middle-income countries** (LMICs), unsafe injection practices amongst PWID, unsafe use/re-use of medical equipment, sexual behaviours, vertical transmission (mother to child), and body modification practices such as tattooing.

Transfusions of blood/blood products were widely performing during and after World War II – this led to the spread of a lot of bloodborne viruses which were not being screened for at the time. Indeed, this led to the term **transfusion-associated hepatitis** (TAH) wherein a person developed hepatitis as a result of a blood/blood product transfusion. In 1978, it was found that 90 % of TAH cases were as a result of NANBH, (Alter *et al.*, 1978) and in 1989, the etiological agent for these NANBH cases was discovered: HCV (Choo *et al.*, 1989). In 1999, the Committee for Proprietary Medicinal Products in Europe pushed for blood/blood products to be tested for HCV using **nucleic acid amplification technology** (NAT), and in 1999 several countries including the UK and the USA started screening for HCV contamination with several countries commencing screening in the subsequent years (Roth *et al.*, 2012). This intervention reduced the risk of HCV for people receiving blood transfusions in higher income countries. However, the same, unfortunately, cannot be said for those in low- and middle-income countries.

Table 1. 10. Details of which countries adopted NAT screening for HCV in which year. Adapted from Roth et al. (2012).

Year	Countries which implemented NAT testing for screening for HCV
1997	Germany
1998	NA
1999	Austria, Canada, Ireland, Japan, Netherlands, Spain, Switzerland, UK, USA
2000	Australia, Norway, Singapore, Slovenia
2001	France, New Zealand
2002	Belgium, Hong Kong, Italy, Poland
2003	Greece, Portugal
2004	NA
2005	South Africa, South Korea
2006	Estonia, Thailand
2007	Kuwait, Malaysia
2008	Finland, Israel, Latvia, Taiwan
2009	Denmark

For example, in Sub-Saharan Africa, the WHO found in their 2017 report that 1.0 % of blood units screened for HCV tested positive (ranging from 0 % to 4.9 %) and in general 8.7 % of all blood units are discarded over concerns related to **transfusion-transmitted infections** (TTIs). However, they also found that only 41 out of 46 countries in the region were testing all blood/blood products for HCV. Ergo the risk for transfusion-associated HCV infection remains high due to the lack of implementation of screen procedures due to cost and limited availability of screening tools (Thursz and Fontanet, 2014; World Health Organisation, 2017).

Screening of blood/blood products has reduced the number of new HCV infections associated with transfusions in high-income countries which has resulted in injection drug use now being the main route of transmission in these places. In PWID, the main route of transmission is the sharing and re-use of needles. In LMICs, however, the number of PWID is rising but it is still not a major route of transmission. One study estimated that if HCV transmission amongst PWID was removed, 79 % of new cases of HCV in high-income countries would not have occurred, and 38 % of cases in LMICs between 2018 and 2020 – highlighting the disparity of this route of transmission between high-income countries and LMICs (Trickey *et al.*, 2019).

Iatrogenic transmission of HCV is also a risk in both high-income and LMIC countries. In Italy, one study recorded an increased risk of HCV infection following minor surgeries, biopsy, and endoscopy procedures – highlighting the need for adequate sterilisation in healthcare settings (Caminada *et al.*, 2023). However, in LMICs, it is much more common for medical equipment, including injective equipment, to be re-used due to cost and availability with the WHO estimating that nine billion medical procedures are performed with reused equipment (Miller and Pisani, 1999; Hutin, Hauri and Armstrong, 2003).

Sexual transmission of HCV is possible, but the risk varies depending on the people involved. The lowest risk was found to be in monogamous couples wherein one of the partners was living with HCV with the rate of transmission found to be around 0.07 % per year (Terrault *et al.*, 2013). However, the risk of sexual transmission of HCV increased for heterosexual people engaging in unsafe i.e. condomless sexual activity with multiple partners; one study in Italy reported a three-fold higher risk of HCV transmission in individuals with three or more sexual partners in the last six months (Mele *et al.*, 1999). Additionally, co-infection with other sexually transmitted diseases which can damage or disrupt the genital mucosa also increased the risk of sexually transmitting HCV (Terrault, 2002). Lastly, it has been found that the highest risk of sexual transmission of HCV is between **men who have sex with men** (MSM), especially in those co-infected with HIV (Jin *et al.*, 2021).

Vertical transmission of HCV is also possible. **Mother-to-child transmission** (MTCT) is known to correlate with degree of viraemia - higher levels of viraemia in the mother correlating with a higher chance that the virus will spread to the foetus. Overall, risk of vertical transmission is 5.8 % (ranging from 4.2 % to 7.8 %) for mothers only infected with HCV. Whereas, for mothers co-infected with HIV, the percentage risk nearly doubled to 10.8 % (7.6 % to 15.2 %) The type of birth does not increase risk of HCV nor does breastfeeding (Benova *et al.*, 2014; US Preventive Services Task Force, 2020; Terrault *et al.*, 2021).

Lastly, tattooing and inadequate safety procedures around the use of tattooing needles have been found in various studies from Taiwan, Norway, New Mexico, Texas, and Italy to have a correlation with HCV transmission (Ko *et al.*, 1992; Holsen, Harthug and Myrmel, 1993; Balasekaran *et al.*, 1999; Haley and Fischer, 2001; Mariano *et al.*, 2004).

1. 1. 3. Diagnosis and Prognosis

The latest guidelines from the **European Association for the Study of the Liver** (EASL), which the NHS in the UK follows, recommend the use of an enzyme immunoassay to detect anti-HCV antibodies in conjunction with a NAT (such as a PCR test) to confirm the presence of HCV RNA, or testing for the presence of HCV core protein to diagnose a person with a HCV infection (Pawlotsky *et al.*, 2020).

Quick diagnosis following infection can be difficult as HCV begins as an acute infection which is generally asymptomatic, however, approximately 15-30 % of infected persons present with symptomatic acute hepatitis with symptoms appearing within 5-12 weeks of infection, and symptoms lasting between 2-12 weeks from their onset. Primary HCV infection produces mild symptoms which can include nausea, vomiting, malaise, fever, and potentially jaundice (Marcellin, 1999; Orland, Wright and Cooper, 2001).

During the acute phase, approximately 25 % of infected persons will spontaneously clear the infection wherein HCV RNA can no longer be detected in the blood (Micallef, Kaldor and Dore, 2006). Clearance usually happens within 6 months of infection. However, one study on Alaskan natives found spontaneous clearance of HCV could occur after 12 months (Scott *et al.*, 2006). Additionally, the factors which contribute to spontaneous clearance of the virus are multivariate. Factors which are positively associated with clearance include being female, younger age upon infection, lower HCV RNA load, and co-infection with HBV. Spontaneous clearance and current intravenous drug use are negatively associated (Bulteel *et al.*, 2016).

However, there are also biological factors which affect spontaneous clearance. A meta-analysis of 262 studies found that certain HLA and HLA-related genes were involved in promoting or mitigating HCV clearance. For example, the *HLA-DQB1*02* allele was found to be associated with a failure to clear the virus, whilst *HLA-DQB1*03*, *DQB1*03:01*, *DRB1*11* and *DRB1*11:01* alleles promoted clearance (Gauthiez *et al.*, 2017). Additionally, several polymorphisms near the

interferon-λ-3 gene (IFN λ3 or IL28B) have been associated with spontaneous clearance of HCV – these include the rs12979860-C, rs8099917-T and the ss469415590 TT polymorphism (Ge *et al.*, 2009; Tanaka *et al.*, 2009; Prokunina-Olsson *et al.*, 2013).

Moreover, other studies have shown that differences in the virus itself can alter the likelihood of spontaneous clearance. It has been shown that people with a relatively evolutionarily static HCV infection i.e. low complexity of quasispecies are more likely to spontaneously clear the virus (Farci *et al.*, 2000). Moreover, the inverse has also been shown to decrease the likelihood of spontaneous clearance. Higher levels of complexity of quasispecies were found to be associated with persistent viraemia. Additionally, Ray *et al.* also found that persistent viraemia was associated with an E1 sequence with a lower dN/dS ratio, and an HVR-1 with a higher dN/dS ratio (Ray *et al.*, 1999). Another study found that HCV genotype may also play a role in spontaneous clearance with genotype 1 being more likely to spontaneously clear than genotypes 2 or 3 in their study (Harris *et al.*, 2007).

For those that do not clear the infection during the acute phase, the disease progresses to chronicity. During the chronic phase of the infection, patients can develop liver cirrhosis over a long period of time (approximately 20-40 years) which may lead to hepatocellular carcinoma (Alter and Seeff, 2000). Additionally, patients may develop extrahepatic conditions which includes mixed cryoglobulinemia, B-cell non-Hodgkin lymphoma, increased rates of insulin resistance, diabetes, atherosclerosis - which may lead to increased cardiovascular morbidity and mortality – and neurological manifestations of HCV infection which can include fatigue and cognitive impairment (Negro *et al.*, 2015). The development of extrahepatic effects of HCV infection is not well understood, and likely multifactorial in nature.

1. 2. The HCV Life Cycle

1. 2. 1. The HCV Virion

HCV is an +ssRNA enveloped virus of the *Hepacivirus* genus, in the *Flaviviridae* family. HCV virions are heterogeneous in size and range from 40 to 100 nm in diameter. HCV particles consist of the RNA genome interacting with the Core HCV protein, forming the nucleocapsid. The HCV nucleocapsid is surrounded by a host-derived lipid membrane into which the E1E2 glycoproteins are incorporated. Furthermore, the glycoprotein-studded lipid membrane further associates with apolipoproteins, which can include apoE, apoB, apoA-I, apoC-I, apoC-II, apoC-III, and cholesteryl esters producing infectious HCV virions which resemble **very low/low density lipoproteins** (VLDL/LDL) (André *et al.*, 2002; Chang *et al.*, 2007; Meunier, Russell, Engle, *et al.*, 2008; Owen *et al.*, 2009; Merz *et al.*, 2011; Catanese *et al.*, 2013)

All enveloped viruses must commit to the same basic life cycle to commence productive infection and propagation; they must attach and enter target cells, fuse their membrane with that of the target cell, replicate, assemble nascent viral particles, and then egress from the target cell.

1. 2. 2. HCV Entry and Fusion

Infectious HCV **lipoviral particles** (LVPs) are heterogenous, spherical particles with E1E2 glycoproteins on the surface of the particle. The viral particle is further associated with various apolipoproteins, notably apoE. Moreover, studies of the ultrastructure of the HCV LVP suggest that there is a sparsity of E1E2 glycoprotein complexes around the LVP, or indeed the glycoproteins are being masked by host proteins, such as apoE (Catanese *et al.*, 2013). In addition to apolipoprotein apoE, HCV particles also incorporate apoB and apoA-I. Both the E1E2 glycoproteins and apoE are required for efficient viral entry, working in concert with a variety of host factors.

Initial attachment of the HCV LVP is mediated by both the **low-density lipoprotein receptor** (LDLR) and specific **glycosaminoglycan** (GAG) moieties found on **heparan sulphate proteoglycans** (HSPGs) on the surface of hepatocytes. The apoE which is associated with the LVP is able to interact with both the LDLR and the GAG moieties to facilitate adsorption of the LVP on to the host cell (Monazahian *et al.*, 1999; Germi *et al.*, 2002). Following attachment of the LVP to the cell surface, a series of downstream interactions occur between the LVP and various host factors, the most studied of which are **Scavenger Receptor – B1** (SR-B1), **CD81**, **claudin-1** (CLDN-1), and **occludin-1** (OCLN).

SR-B1 is thought to be the first of these four receptors that the HCV LVPs interact with during and following attachment. SR-B1 binds high- and very low-density lipoproteins, and promotes the uptake of cholesterol from high-density lipoproteins to hepatocytes – this highlights the role of SR-B1 in lipid homeostasis (Acton *et al.*, 1996; Rhainds and Brissette, 2004). SR-B1 was discovered to interact with HCV E2 in 2002. HepG2 cells, which lack CD81, were found to be able to bind recombinant HCV E2. Moreover, they found that recombinant E2 proteins which have their HVR-1 deleted were unable to bind these HepG2 cells, and this led to the hypothesis that SR-B1 was able to bind E2 via the HVR-1 and could act as a co-receptor for HCV entry (Scarselli *et al.*, 2002). Further studies elucidated that SR-B1 actually possessed many functions which aided in viral entry outwith interaction with the HVR-1. SR-B1 was found to be able to interact

with the apolipoproteins within the HCV LVP, aiding attachment to the host cell. Additionally, some studies have also suggested that the lipid transfer function of SR-B1 may promote viral entry by transferring the lipids associated with the HCV LVP away, permitting interactions between SR-B1 and HVR-1, which in turn facilitate E2:CD81 interactions (Bankwitz *et al.*, 2010; Thi *et al.*, 2012; Zahid *et al.*, 2013)

Following interaction with the SR-B1 receptor, it is believed that the HCV LVPs interact with CD81, a ubiquitous tetraspannin with numerous roles in host cells. CD81 was the first identified HCV receptor, discovered by Pileri *et al.* in 1998. They cloned human CD81 into WOP cells, a fibroblast cell line from mice, and measured E2 binding (Pileri *et al.*, 1998). Further studies ascertained that it is the **large extracellular loop** (LEL) which binds HCV with a strong affinity, and that this took place following attachment of the HCV LVP to the host cell (Petracca *et al.*, 2000; Bertaux and Dragic, 2006). Moreover, CD81 has been shown to interact with both SR-B1 (Gerold *et al.*, 2015; Palor *et al.*, 2020) and CLDN-1.

CLDN-1 is a tight junction protein in hepatocytes and is a known co-receptor for HCV entry, discovered in 2007 by Evans *et al.* through expressing CLDN-1 in HEK 293T cells. HEK 293T cells are normally resistant to HCVpp infection, but, upon expression of CLDN-1, the HEK 293T cells became HCVpp permissive. However, CLDN-1 does not interact directly with the glycoproteins of HCV, rather, the **extracellular loop 1** (EL1) interacts with CD81 to form a E2:CD81:CLDN-1 complex which then facilitates viral entry through clathrin-mediated endocytosis, with these complexes preferentially forming on the basolateral surface of hepatocyte (Evans *et al.*, 2007; Mee *et al.*, 2009; Harris *et al.*, 2010; Krieger *et al.*, 2010; Farquhar *et al.*, 2012).

The formation of the E2:CD81:CLDN-1 complex is critical for HCV entry into cells. As noted, this interaction is mediated by the LEL of CD81 binding to the E2 of HCV, and then the EL1 of CLDN-1 binding to CD81. However, the formation of these complexes is mediated by the activities of various other co-factors. For example, the formation of the E2:CD81:CLDN-1 complex may be reliant on the

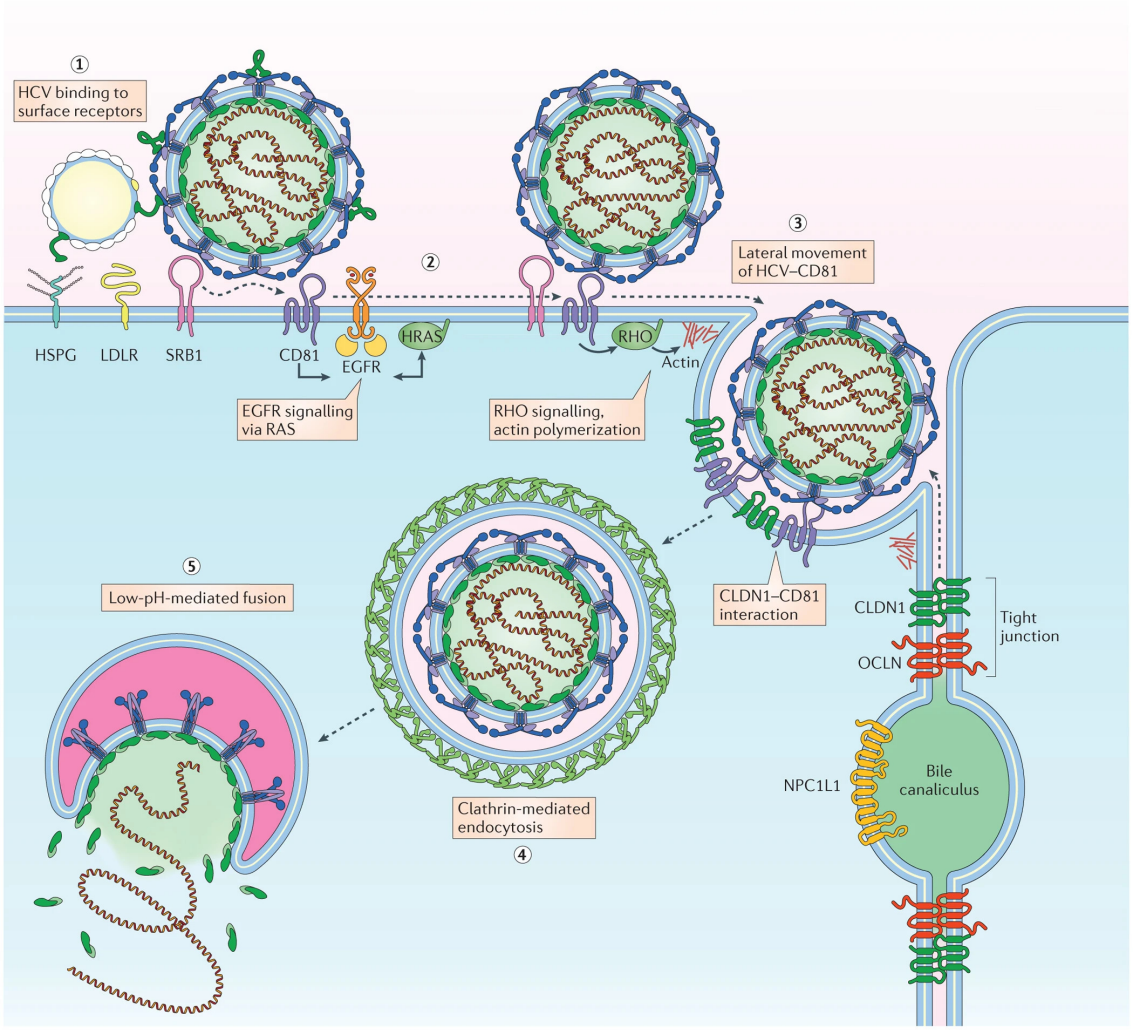
function of **protein kinase A** (PKA) to maintain CLDN-1 in the cell membrane, and prevent its internalisation (Farquhar *et al.*, 2008), maintaining the correct cellular localisation of CLDN-1 for complex formation. When considering cellular localisation, it is the E2:CD81 complex itself which activates RHO GTPases Rac, Rho, and Cdc42 which then re-localise the E2:CD81 complex to the tight junction via actin rearrangements in the host cell (Brazzoli *et al.*, 2008) thus bringing the E2:CD81 complex and CLDN-1 together in local proximity at the tight junction. Further still, two **receptor tyrosine kinases** (RTKs) - **epidermal growth factor receptor** (EGFR) and **ephrin receptor A2** (EphA2) – have also been identified as regulating the formation of the CD81:CLDN-1 complex (Lupberger *et al.*, 2011).

OCLN, another tight junction protein, was found in 2009 to be an essential co-factor in HCV entry. This was discovered by cloning in OCLN into murine cells already expressing human CD81, SR-B1, and CLDN-1, rendering the murine cells permissive to HCV infection (Furuse *et al.*, 1993; Ploss *et al.*, 2009). OCLN features four transmembrane domains, and two extracellular loops and furthermore, the second of these extracellular loops was found to be involved in HCV entry into hepatocytes. Moreover, it is known to function later in the entry events of HCV (Furuse *et al.*, 1993; Liu *et al.*, 2009; Sourisseau *et al.*, 2013). More work is required to understand the exact mechanisms by which OCLN contributes to HCV entry, however, it has been shown that HCV particles recruit OCLN outside of the tight junction, and does so independently of CLDN-1 expression (Deffieu *et al.*, 2022).

Other factors like **transferrin receptor 1** (TFR1) – an iron uptake receptor - and **Niemann-Pick C1-like 1 receptor** (NPC1L1) – involved in cholesterol uptake - have been suggested to be important for HCV entry as well. One study found that TFR1 could act as a receptor for HCV which was important post-CD81 binding, however, its mechanism is yet to be elucidated (Martin and Uprichard, 2013). Lastly, NPC1L1 has also been suggested as an HCV entry factor based on a study which inhibited NPC1L1 both *in vitro* and *in vivo* and found that NPC1L1 was necessary for HCV infection (Sainz *et al.*, 2012) but, again, its mechanism has yet to be described. However, as of today no further work has been

presented on either of these potential HCV receptors and so it remains to be seen the exact role either of these receptors play in HCV entry, if any.

Taken altogether, a picture of the viral entry events for HCV can start to be pieced together: HCV LVPs attach to LDLR/HSPG GAGs/SR-B1. SR-B1 interacts with the apolipoproteins surrounding the HCV LVP and may start to recycle some of the lipids in the LVP, potentially exposing the E2 glycoprotein. The HVR-1 of the E2 can then interact with SR-B1, facilitating E2:CD81 interaction. This activates various RHO GTPases which restructure cortical actin to relocalise the E2:CD81 complex to the tight junction. CLDN-1 is being maintained in the tight junction though PKA, and then when the E2:CD81 complex arrives, the EL2 of CLDN-1 forms a complex with the E2:CD81 complex aided by EGFR and EphA2. This E2:CD81:CLDN-1 complex then induces clathrin-mediated endocytosis. OCLN is being recruited independently of this process, though its exact role in abetting this series of events is still under investigation. The virus is now inside the cell.



Nature Reviews | Microbiology

Figure 1. 3. **Schematic Overview of HCV Entry and Fusion into Hepatocytes.** 1: Initial attachment is facilitated by HSPG, LDLR, and SR-B1 and the HCV LVP. Interaction with SR-B1, potentially mediated by the HVR-1 of E2, allows for E2:CD81 interactions. 2: E2:CD81 interaction induces EGFR and Rho GTPases signalling. 3: Signalling induces lateral movement of the HCV particle attached to CD81 to sites of cell-to-cell contact through cortical actin rearrangements. 4: CLDN-1 interaction with CD81 induces HCV clathrin-mediated endocytosis. 5: Maturation of the endosome lowers the pH, inducing conformational changes in E1E2, facilitating fusion and the deposition of the viral genome into the cytoplasm of the target cell.

To commence productive infection, the HCV LVP must now fuse its membrane to that of the endosomal compartment. It does this through the co-ordinated actions of the viral glycoproteins E1E2. However, this process is not well understood for HCV, and this overarching question formed the backdrop of this PhD. The current understanding of how fusion occurs in HCV, and across viruses is discussed in depth later. In brief, however, viral fusion is mediated by the E1E2 glycoproteins. It is hypothesised that these glycoproteins undergo conformational rearrangements to bring the membrane of the host and viral particle into close apposition, permitting mixing and fusion.

The +ssRNA HCV genome is then released into the cytoplasm of the target cell and viral replication can commence. Given that HCV is a +ssRNA, protein translation can be initiated immediately upon entry to the cell through hi-jacking host translation machinery. However, viral genome replication requires a few specific steps such as the generation of the -ve sense template strand of RNA and assembly of replication organelles, processes which are dependent on viral proteins. Therefore, translation of the genome must necessarily be accomplished before genome replication.

1. 2. 3. HCV Translation – Protein Synthesis

The HCV genome is approximately 9.6 kbps in length and is used for both protein translation and **viral RNA** (vRNA) replication. It contains one open reading frame which is flanked by both a 5' and a 3' **untranslated region** (UTR). The 5'UTR contains an **internal ribosome entry site** (IRES) which allows for translation initiation in a cap-independent manner (Pérard *et al.*, 2013). However, there are also important elements contained within the 3'UTR and ***cis*-acting RNA elements** (CREs) throughout the protein-coding region of the genome which are important for translation (McMullan *et al.*, 2007; You and Rice, 2008)

The IRES is formed by three **stem-loop** (SL) domains within the 5'UTR – SLII, SLIII, and SLIV. Each SL performs a separate function within the IRES. To commence protein translation, host-derived, translationally-competent ribosomes must be recruited to the HCV genome. Recruitment begins with the SLIII recruiting the 40S ribosomal subunit through a conserved, base-pairing interaction between SLIII and a region of the 18S ribosomal RNA (Matsuda and Mauro, 2014).

This facilitates SLII to then bind the 40S ribosomal subunit tightly. SLIV is rather unstable compared to the other SLs within the IRES and contains the start codon for translation which is released from its stem-loop formation upon 40S binding to both SLII and SLIII. The IRES-40S complex associates with **eukaryotic Initiation Factor 3** (eIF3) and another complex composed of **ternary Met-transfer RNA** (tRNAi), **eukaryotic Initiation Factor 2** (eIF2), and **guanine triphosphate** (GTP). Together the IRES-40S-eIF3 complex and the tRNAi-eIF2-GTP complex forms the 48s IRES-preinitiation complex. The hydrolysis of the associated GTP allows for the 60S ribosomal subunit to bind to the IRES-40S complex following the release of the eIFs – displacement of eIF3 is aided by SLIII - and this leads to the generation of the IRES-80S initiation complex which is elongation competent and can begin translation at the rough **endoplasmic reticulum** (ER) (Hashem *et al.*, 2013; Pérard *et al.*, 2013; Quade *et al.*, 2015; Yamamoto *et al.*, 2015)

Moreover, effective, and efficient translation initiation is aided by **microRNA-122** (miR-122) which is found abundantly and specifically within the liver. miR122 has many roles within the cell but, in general, microRNAs modulate human gene expression through mediating or inhibiting degradation of certain mRNAs. miR-122 has been shown to bind to two locations within the 5'UTR and increases viral replication by stabilising the IRES as an RNA chaperone (Jopling *et al.*, 2005; Jopling, Schütz and Sarnow, 2008; Schult *et al.*, 2018). Moreover, association of miR-122 to the 5' end of the HCV genome mitigates degradation by Xrn1 and Xrn2 – both host nucleases which would otherwise degrade the HCV genome, halting translation/replication (Li *et al.*, 2013; Sedano and Sarnow, 2014).

As the HCV genome contains one ORF, translation produces one polyprotein which requires a series of cleavage events to liberate individual proteins from the polyprotein. As noted, genome translation occurs at the rough ER which contains the host proteases required to liberate proteins at the N-terminal of the nascent polyprotein.

1. 2. 4. Polyprotein Processing

After the generation of the full polyprotein within the ER, two host associated enzymes cleave the structural proteins and the vipoporin p7 from the polyprotein. Signal peptide peptidase cleaves the signal sequence from the N-terminus of the nascent E1 proteins, with signal peptidase cleaving E1, E2, and p7 from the polyprotein, and also cleaving the N-terminus of the NS2 protein. Following this, two viral proteases co-ordinate the cleavage events which release the non-structural proteins from the polyprotein – the C-terminus of the NS2 protein is cleaved from NS3 through the NS2-3 proteolytic complex. This liberates NS2 from NS3. NS3 then forms a serine-type protease with its cofactor NS4A which then cleaves the rest of the non-structural proteins from the polyprotein. After all cleavage events have taken place, the viral proteins remain associated with the membrane of the ER (Failla, Tomei and De Francesco, 1994; Bartenschlager *et al.*, 1995; Bartenschlager, Frese and Pietschmann, 2004).

An overview of the genome structure, polyprotein processing events and viral protein membrane topology and functions can be seen in figure 1. 4. which was derived from Tabata, Neufeldt and Bartenschlager, 2020.

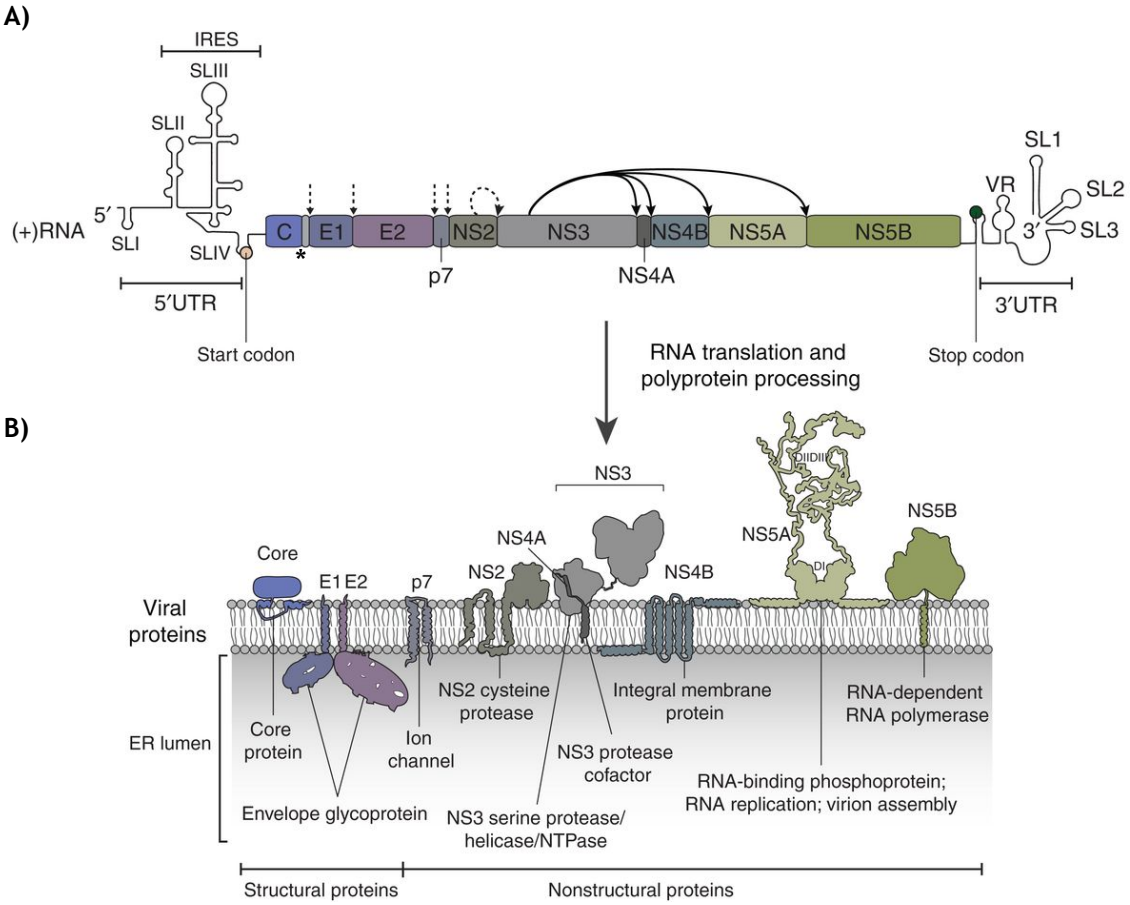


Figure 1. 4. **Overview of HCV Genome Architecture and Viral Protein Membrane Topology.** A) Predicted structures of the 5'UTR and 3'UTR with their predicted functional domains annotated. The two UTRs flank the open reading frame which has viral and host protease cleavage sites denoted. * = cleaved by host signal peptide peptidase. Vertical, dashed arrows = cleaved by host signal peptidase. Curved, dashed arrow = cleaved by NS2/3 protease. Solid arrows = cleaved by NS3/4A protease.

B) Mature viral protein topology with a description of their functions below. The three domains of the NS5A dimer are denoted as DI through DIII. NB: this is the only viral protein represented as a dimer despite other mature viral proteins potentially oligomerising as well (figure adapted from Tabata, Neufeldt and Bartenschlager, 2020.)

1. 2. 5. HCV Genome Replication – Building the Replication Organelle

After liberation from the polyprotein, the HCV **non-structural** (NS) proteins NS3, NS4A, NS4B, NS5A, NS5B begin amplification of the viral genome. Like other +ve sense RNA viruses, HCV induces massive alterations to the host membrane architecture – a feature which was noted even before the aetiological agent for HCV for discovered in 1989, and has been seen in chimpanzee infection models (Jackson, Tabor and Gerety, 1979; Shimizu *et al.*, 1990).

Alteration of the host ER membrane architecture is able to be induced by single HCV protein/protein assemblies such as the NS3/4A complex which can induce membrane perturbations (Egger *et al.*, 2002). Additionally, NS4B is able to induce what is termed the "membranous web" on its own – a condensed membrane structure which was thought to be the site of HCV genome replication (Gosert *et al.*, 2003). However, further work has since shown that the most abundant membrane perturbation within the membranous web is the production of **double-membrane vesicles** (DMVs). **Single-membrane vesicles** (SMVs), and **multimembrane vesicles** (MMVs) have also been detected but SMVs are rarer, and MMVs are thought to arise due to cell stress. DMVs have been found to contain both non-structural proteins and **double-stranded RNA** (dsRNA). Taken together, these two observations point to the DMV being the site of replication for HCV (Ferraris, Blanchard and Roingeard, 2010; Romero-Brey *et al.*, 2012; Paul *et al.*, 2013). Moreover, in 2019, Lee *et al.* developed a system to enable visualisation of the HCV replication organelle. They found that HCV proteins were able to induce ER-derived membranes into wrapping around **cytosolic lipid droplets** (cLDs), around which DMVs were produced. The association between cLDs and virus production had been noted many years early (Boulant, Targett-Adams and McLauchlan, 2007; Miyanari *et al.*, 2007; Jirasko *et al.*, 2010), but it was through the development of a novel **correlative light and electron microscopy** (CLEM) system for HCV that this cLD/HCV protein association could be characterised further. The arrangement of cLD, surrounded by ER-derived membrane with DMVs dotted around allows for HCV to couple both viral RNA replication with assembly (Lee *et al.*, 2019).

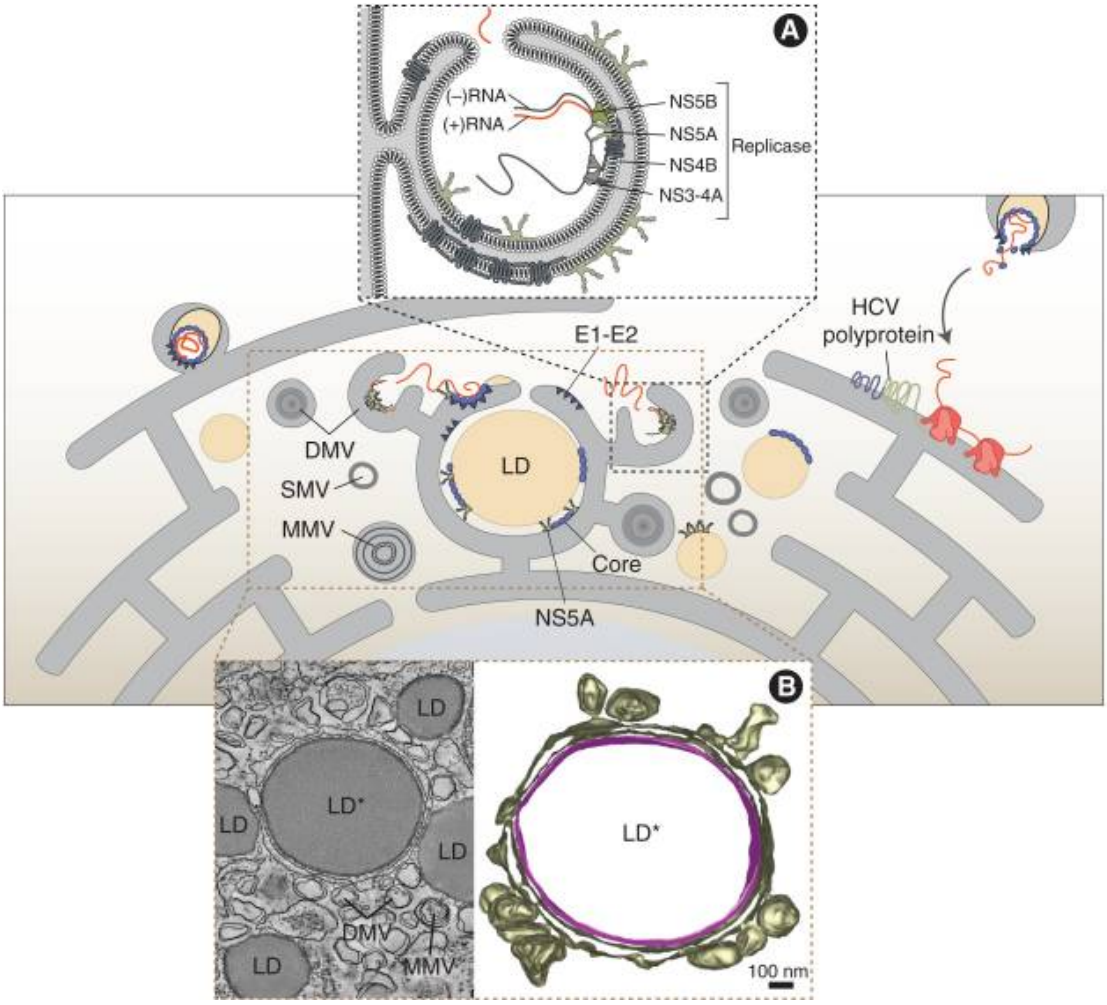


Figure 1. 5. **Overview of the HCV Replication Organelle Structure.** After fusion, the HCV genome travels from the cytoplasm to be translated at the rough endoplasmic reticulum (ER). After translation, viral proteins induce alterations in the membrane architecture of the ER which produces double-membrane vesicles (DMVs), single-membrane vesicles (SMVs), and multimembrane vesicles (MMVs). DMVs are formed from the ER and contain the non-structural (NS) proteins required for genome replication - NS3/4A, NS4B, NS5A, NS5B. Moreover, These DMVs usually associate with lipid droplets and are potential factories of genome replication (A). As the DMVs mature, they go from an open state to a closed state. As long as the DMV is open, nascent strands of replicated +ve sense RNA may be trafficked by NS5A and NS3 to sites of virion assembly wherein there is an enrichment of HCV core proteins and the E1E2 glycoproteins. As noted, HCV proteins induce the formation of DMVs; some of these DMVs derive from ER membranes which themselves surround LDs (tomograph slice showing non-ER membrane associated LDs and LD* in inset B, left. 3D reconstruction of LD*s in inset B, right). These LD*s have been found to contain both E2 and NS5A, whereas lipid droplets not surrounded by the ER membrane do not suggesting that HCV proteins are facilitating the production of this ER membrane-bound lipid droplets (figure adapted from Tabata, Neufeldt and Bartenschlager, 2020, with the image for inset B from Lee et al. 2019).

The **replication organelle** (RO) requires both viral and host factors for its correct assembly. DAAs, which will be discussed in depth in section 1. 3., can also highlight the roles that certain NS proteins play in the formation of the RO. It has been shown that daclatasvir, a NS5A inhibitor, was able to prevent formation of the HCV RO (Berger *et al.*, 2014). Moreover, through expression of individual proteins, or truncated proteins in Huh7 cells via a mutated subgenomic replicon system, it has been shown that both the helicase domain in NS3 and domain I in NS5A are critical for DMV formation (Romero-Brey *et al.*, 2015)

NS4B has also been shown to have a key role in RO biogenesis. NS4B is composed of four transmembrane domains flanked at the N- and C-terminus by two amphipathic α-helices at each terminal (four in total). Additionally, glycine zipper motifs present in transmembrane helices 2 and 3 have been demonstrated to aid in the formation of DMVs, with loss of function glycine zipper mutants producing SMVs with secondary invaginations which may be an intermediate step in DMV production (Paul *et al.*, 2018).

Additionally, several host factors have been identified as having a role in the biogenesis and function of the HCV RO. Firstly, **phosphatidylinositol 4-kinase IIIα** (PI4KIIIA), a lipid kinase, was shown to be recruited and activated by NS5A, producing **phosphatidylinositol 4-phosphate** (PI4P) at HCV protein-containing membranes (Berger *et al.*, 2011). Following this, a slew of other host factors, most notably those involved in transfer of lipids, were identified as having important roles at the HCV RO. These include oxysterol-binding protein, NPC1 (possibly in conjunction with NPC2), and FAPP2 (H. Wang *et al.*, 2014; I. Khan *et al.*, 2014; Stoeck *et al.*, 2017). The localisation of these lipid transfer proteins to the HCV RO may be necessary for the local enrichment of distinct lipids which could facilitate HCV replication and assembly.

1. 2. 6. HCV Genome Replication

Following biogenesis of the RO, the HCV RNA genome can now be replicated. NS5B is the **RNA-dependent RNA polymerase** (RdRp). Like other RdRp proteins, NS5B needs to bind the RNA template, initiate replication, elongate the nascent RNA, and terminate at the end of the template strand of RNA. NS5B contains a palm, a finger, and a thumb domain at the C-terminus along with a trans-membrane domain at the N-terminus which anchors the protein into the membrane (Sesmero and Thorpe, 2015). NS5B, itself, is able to bind to many target RNA sequences, including some host RNAs (Yuhashi *et al.*, 2014). Importantly, the NS5B is recruited to the 3'UTR of the +ve sense RNA strand which contains a poly(U/C) tract, a highly conserved X region, essential *cis*-acting elements, and stem-loop structures required for initiation (Tanaka *et al.*, 1995; Kolykhalov, Feinstone and Rice, 1996; You and Rice, 2008). NS5B then replicates the +ve RNA strand, producing the -ve RNA strand which will function as the template to produce more +ve sense RNA genomes to be packaged. The 3'-UTR of the -ve RNA template is an efficient initiator of RNA synthesis, whilst the inverse is true of the 3' end of +ve RNA strand – taken together, this results in a ten-fold excess of +ve sense RNA strands to -ve sense RNA strands. (Fricke *et al.*, 2015; Niepmann *et al.*, 2018).

However, NS5B does not work alone. The NS3/4A complex also aids in genome replication. It contains a DExD-box RNA helicase domain at its C-terminus which has a role in viral genome replication. There has been a correlation between the ability of the NS3/4A complex to unwind nucleic acids efficiently and overall replicative ability and ergo viral fitness observed (Stross *et al.*, 2016; Zhou *et al.*, 2018).

Lastly, there is the NS5A protein. The role of this protein in viral genome replication and assembly is complex. Notably, it contains three separate domains which are connected by **low-complexity sequences** (LCSs) (Penin *et al.*, 2004) Domain I is essential for HCV RO biogenesis as discussed above. However, Domain I may also play a role in genome replication. Certain residues in both Domain II and III have been identified as aiding in viral genome replication.

However, Domain III has also been implicated in viral particle assembly (Appel *et al.*, 2008; Tellinghuisen *et al.*, 2008; Romero-Brey *et al.*, 2015) The balance of these roles is thought to be mediated by phosphorylation of NS5A which itself may be mediated by the helicase domain of NS3/4A (Yu *et al.*, 2022; Lee *et al.*, 2024).

So, translation has been initiated, the RO has formed, and the HCV genome has been replicated. How these elements are brought together in the release of nascent viral particles shall now be discussed.

1. 2. 7. Particle Assembly and Release

Particle assembly requires the bringing together of viral genomic RNA and viral structural proteins – core, E1, E2. This process for HCV is associated with hepatocyte lipid metabolism. Firstly, the core protein has been found to homodimerise and then relocate to cLDs facilitated by two amphipathic helices contained within the C-terminus of the core protein (Barba *et al.*, 1997; Boulant *et al.*, 2005). Moreover, the host factor **diacylglycerol acetyltransferase-1** (DGAT1) has been implicated in mediating the interaction between the core protein and these cLDs (Herker *et al.*, 2010), and the trafficking of the core protein towards the cLDs has been shown to require both the **MAPK-regulated cytosolic phospholipase A2** (PLA2G4) and its product arachidonic acid (Menzel *et al.*, 2012). At this point, it is thought that newly replicated viral genomes are transferred to the assembly sites by NS5A where they interact with core to form the nucleocapsid (Gawlik and Gallay, 2014). NS5A interacts with the cLD-associated core protein by being localised to the cLD itself by DGAT1, facilitating complex formation between NS5A and core (Camus *et al.*, 2013) and it may be this interaction which forms the nucleocapsid.

However, following association with the cLDs, the core protein needs to be re-trafficked back towards the membrane of the ER to facilitate assembly of HCV virions in the HCVcc system. This has been shown to be mediated by NS2 and p7 – both of which play crucial roles in HCV virion assembly (Boson *et al.*, 2011). Moreover, E1 and E2 are also retained on the ER membrane (Dubuisson *et al.*, 1994) and so the re-trafficking of the core protein to the ER membrane brings these virion components back on to the same membrane. Trafficking of the E1E2 glycoproteins towards the site of assembly has been shown to be mediated by a NS2:p7 complex interacting with a NS3:NS4A complex (Jirasko *et al.*, 2010; Morikawa *et al.*, 2011; Popescu *et al.*, 2011; Stapleford and Lindenbach, 2011).

Now all the players are in place; the nucleocapsid and the glycoproteins have been trafficked to the assembly site on the ER membrane, and the assembly of the immature HCV particle can begin. In general, HCV virion biogenesis is closely related to the VLDL assembly pathway. Due to this, assembly of HCV virions is

aided by several host factors including **microsomal triglyceride transfer protein** (MTP), long chain **acyl-CoA synthetase 3** (ACSL3), and **hepatocyte nuclear factor** (HNF4α) – two proteins involved in VLDL biogenesis, and a transcription factor which regulates the VLDL secretory pathway respectively (Nahmias *et al.*, 2008; Yao and Ye, 2008; Li *et al.*, 2014). However, the biggest class of host-derived factors required are apolipoproteins. Moreover, apoE has been shown to be sufficient on its own to allow for the generation of infectious HCV particles in a non-hepatic cells (Da Costa *et al.*, 2012), highlighting its importance and role in determining hepatotropism for HCV.

Immature virions bud into the lumen of the ER from the ER membrane to be secreted via the *trans*-Golgi secretory pathway. Budding may be aided by the **endosomal-sorting complex required for transport** (ESCRT) pathway (Corless *et al.*, 2010; Ariumi *et al.*, 2011). However, as the nascent, immature virion is supposed to bud into the lumen of the ER, the exact role of the ESCRT pathway is debated as its normal role is to delivery ubiquitinated cargo to the lysosome for degradation (Henne, Buchkovich and Emr, 2011). However, it has also been show that p7 likely aids the egressing virion by neutralising any acidic compartments through forming ion channels in membraned vesicles (Wozniak *et al.*, 2010). Following budding, the immature virions are transported out of the cell via the *trans*-Golgi network as noted. During secretion, the virus particle matures. The E1E2 glycoproteins are modified which includes the maturation of the glycans present on the E1E2 complexes, and a shuffling of the disulphide bonds (Vieyres *et al.*, 2010). Moreover, several studies have shown the link between the VLDL pathway and the assembly and egress of nascent HCV LVPs. In brief, lipid droplets which contain the HCV nucleocapsid fuse with VLDL particles which contain apoB. These nascent nucleocapsid:VLDL particles acquire other apolipoproteins, including apoE and apoC, forming the mature LVP which exits the cell via the Golgi network (Gastaminza, Kapadia and Chisari, 2006; Chang *et al.*, 2007; Huang *et al.*, 2007; Gastaminza *et al.*, 2008; Coller *et al.*, 2012)

1. 3. Direct-acting Antivirals

There are currently three classes of DAAs which are approved for use within the UK and Europe, those which inhibit the NS3/4A complex, the NS5A protein, and the NS5B protein (European Medicines Agency, 2018b; BNF, 2025)

Prior to the advent of DAAs for the treatment of HCV, first line treatment involved the administration of both **peginterferon** (PegIFN) and **ribavirin** (RBV) for 24 weeks or 48 weeks depending on the genotype of the HCV – genotypes 2 and 3 required 24 weeks of treatment, and 48 weeks for genotype 1, 4, 5, and 6. The aim of treatment was to induce a **sustained virologic response** (SVR) which is associated with clearance of the virus and better outcomes for patients (Veldt *et al.*, 2007). However, for genotype 1 – the most prevalent genotype of HCV - only 40-50 % of those treated obtained an SVR. However, after the development of boceprevir – an NS3/4A inhibitor – a higher rate of SVR was found in both treatment naïve patients (ranging from 53 % to 75 % across two studies), and patients who had failed previous PegIFN and RBV treatment (SVR rates ranged from 59 % to 66 %) in patients infected with HCV genotype 1 (Kwo *et al.*, 2010; Bacon *et al.*, 2011; Poordad *et al.*, 2011). At the same time, another NS3/4A inhibitor – telaprevir - was also being brought to market. In phase 3 clinical trials, it was found that telaprevir induced SVR rates of between 69 % and 88 % (Jacobson *et al.*, 2011; Zeuzem *et al.*, 2011).

However, both boceprevir and telaprevir induced additional side effects which lead to tolerability issues with the treatment regimens which could last for 3-6 months. In 2013, second-generation HCV DAAs were approved including simeprevir and sofosbuvir. Simeprevir is another NS3/4A inhibitor. Sofosbuvir, however, is a nucleotide analogue which inhibits the NS5B protein. Then in 2014 combination therapies became approved by the FDA and the European Medicines Agency including ledipasvir/sofosbuvir (Harvoni, approved 2014), a combination of ombitasvir/paritaprevir/ritonavir with dasbuvir (Viekira Pak, approved 2014). Since 2011, many DAAs have come to market, some have been withdrawn, such as boceprevir/telaprevir, due to newer, more tolerable formulations being developed which are more effective. Table 1. 3. details the currently available

anti-HCV DAAs approved for use in the UK and Europe, the types of inhibitor, the genotypes effective against, and the year released (Carter, Connelly and Struble, 2017; European Medicines Agency, 2018b, 2018a; Heo and Deeks, 2018; BNF, 2025).

Table 1. 11. Details of currently approved HCV DAAs in the UK and Europe.

Brand Name	Active Substance	Inhibitor Types	HCV Genotype(s) Susceptible to DAA	Year Authorised
Sovaldi	Sofosbuvir	NS5B	All	2014
Harvoni	Ledipasvir, Sofosbuvir	NS5A, NS5B	1, 3, 4, 5, 6	2014
Zepatier	Grazoprevir, Elbasvir	NS3/4A, NS5A	1 or 4	2016
Epclusa	Velpatasvir, Sofosbuvir	NS5A, NS5B	All	2016
Vosevi	Voxilaprevir, Velpatasvir, Sofosbuvir	NS3/4A, NS5A, NS5B	All	2017
Maviret	Glecaprevir, Pibrentasvir	NS3/4A, NS5A	All	2017

1. 3. 1. Overview of types of DAAs

As noted, there are three classes of DAA available: NS3/4A complex inhibitors, NS5A inhibitors, and NS5B inhibitors.

NS3/4A inhibitors can further be broken down into 3 different classes of drugs: P1-3 macrocyclic, P1-3 acyclic, or P2-4 macrocyclic drugs. All currently approved NS3/4A inhibitors are of the P2-4 macrocyclic class (Schechter and Berger, 1967, 1968; McCauley and Rudd, 2016). They operate similarly by competitively inhibiting the serine protease active site in the N-terminus of the NS3 protein in the NS3/4A complex. The active site is composed of a catalytic triad (S139, H57, D81) in conjunction with an oxyanion hole (G137, S139). P2-4 macrocyclic inhibitors contain a carboxylic acid at the C-terminus of the drug, and mimic transition states. The C-terminus of the drug(s) forms non-covalent interactions with the G137 and the S139 residues, which form the oxyanion hole, and prevent the enzymatic activity of the protease (Martino, Petri and De Rosa, 2024). However, it should also be noted that another macrocyclic NS3/4A drug has been shown to interact with both the protease and helicase domain of NS3, mediated by the P2 and P4 moieties (PDB: 3M5L, Schiering *et al.*, 2011), suggesting that inhibition of NS3/4A by P2-P4 macrocyclic drugs may be facilitated by multiple distinct interactions. Regardless, inhibition of NS3/4A prevents downstream polyprotein processing, and therefore halts downstream replication/assembly/egress of the virus.

NS5A inhibitors operate through a currently unknown mode of action. However, it has been shown both using mathematic modelling, and through *in vitro* work that inhibition of NS5A blocks both RNA synthesis and viral assembly (Guedj *et al.*, 2013; McGivern *et al.*, 2014). Furthermore, analysis of hot spots for mutations, and *in vitro* studies with truncated forms of NS5A showed that domain I of NS5A was the target for inhibition (Qiu *et al.*, 2011; Issur and Götte, 2014). Moreover, the work from Qiu *et al.* suggested that these inhibitors were able to prevent hyperphosphorylation of the NS5A protein – an important feature of NS5A regulation (Masaki *et al.*, 2014). Another study by Fridell *et al.* also pointed to NS5A inhibitors blocking the function of both a basally phosphorylated

NS5A, and prevent hyperphosphorylation i.e. blocking one function directly, and preventing the other by blocking change in phosphorylation state (Fridell *et al.*, 2011). Additionally, it has also been shown that NS5A inhibitors cause a re-trafficking of the NS5A protein to LDs and away from the ER membrane, the site of HCV assembly (Targett-Adams *et al.*, 2011). Together, these data suggest that NS5A inhibitors work on domain I of the protein and prevent RNA replication and viral assembly through modulating or preventing correct phosphorylation of NS5A and physically trafficking the protein away from the site of virion assembly.

The last class of HCV DAAs are NS5B inhibitors. The NS5B is the RdRp, and the only currently approved DAA in the UK and Europe for NS5B is sofosbuvir – a nucleotide analogue. Sofosbuvir (ß-d-2′-Deoxy-2′-α-fluoro-2′-ß-C-methyluridine) is a ProTide-type drug, meaning that it is a prodrug nucleotide inhibitor and is a non-obligate chain terminator due to the presence of the 3'-hydroxyl on the uracil base. Sofosbuvir enters the cell as an analogue of uridine monophosphate where it is phosphorylated, becoming a uridine triphosphate analogue (Sofia *et al.*, 2010; Mehellou, Rattan and Balzarini, 2018). The mechanism of action of sofosbuvir has not been fully elucidated or determined experimentally. However, work on a chemically similar drug, ß-d-2′-Deoxy-2′-fluoro-2′-C-methylcytidine, showed that the mechanism through which NS5B was blocked was through steric clashes between the 2' moieties of the drug and the site of elongation in the NS5B protein; essentially the moieties are such that new NTPs cannot be added on to the elongating RNA strand and thus replication is terminated (Stuyver *et al.*, 2006; Ma *et al.*, 2007; Murakami *et al.*, 2007, 2008). Additionally, further research looked at different moieties on the ribose 2' carbon in sofosbuvir, and found some moieties outwith the 2′-α-fluoro-2′-ß-C-methyl moieties which were capable of inducing chain termination which, again, hints at chain termination being caused by a steric clash and not necessarily a direct function the moieties present on sofosbuvir (Fung *et al.*, 2014).

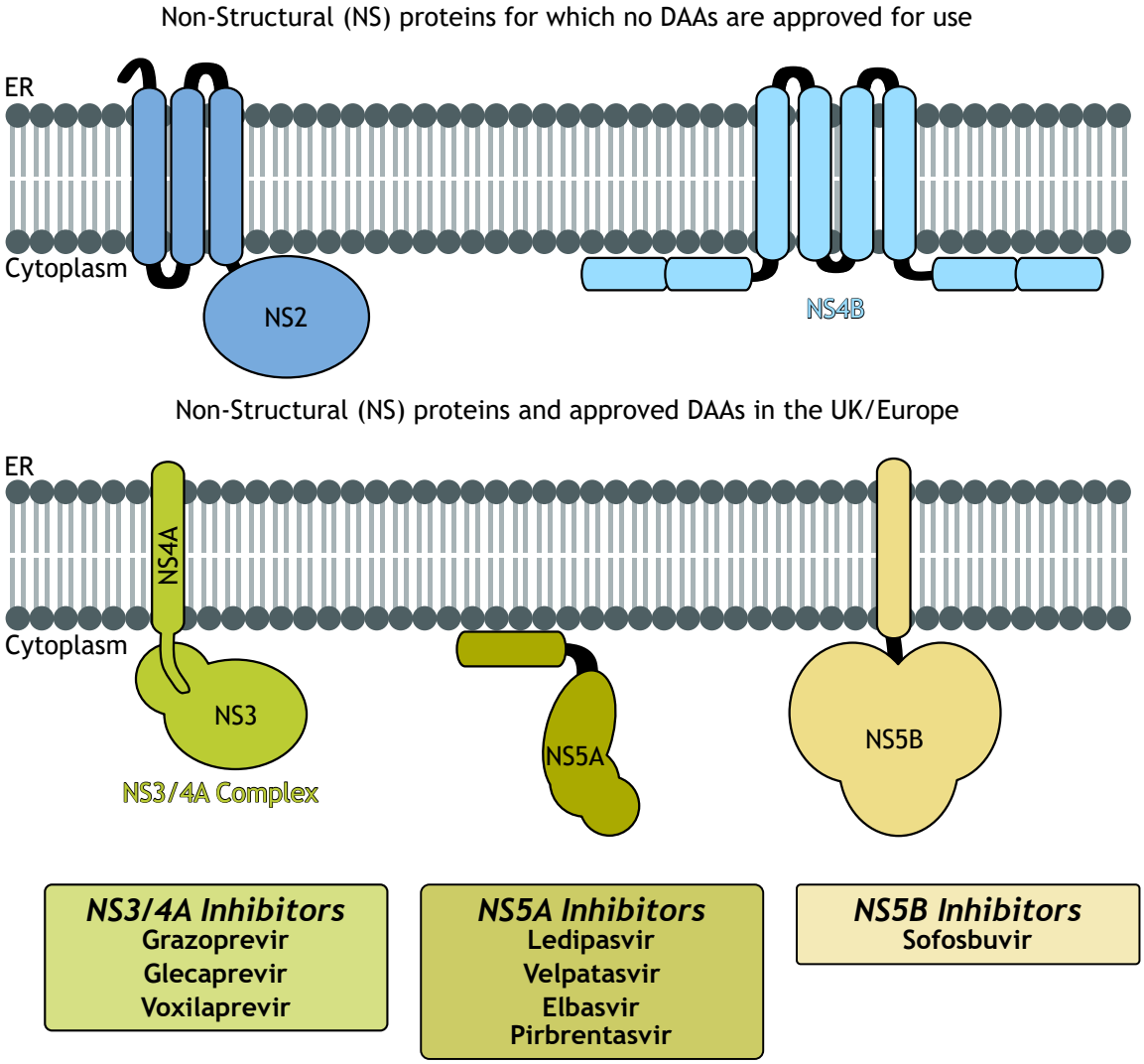


Figure 1. 6. **Overview of Currently Licenced DAAs Available in Europe and Their Targets.** Currently, there are no approved DAAs which target either NS2 or NS4B (top membrane). There are three approved NS3/4A inhibitors, four approved NS5A inhibitors, and one approved NS5B inhibitor (a nucleotide analogue) (bottom membrane). Note the naming conventions: NS3/4A inhibitors are given the -previr suffix, NS5A inhibitors the -svir suffix, and NS5B inhibitors the -buvir suffix.

These currently approved DAAs may be given in combination, and in combination with ribavirin and/or peginterferon-alfa.

Figure adapted from information from the British National Formulary (BNF, 2025).

1. 3. 2. HCV Resistance to DAAs

However, whilst these drugs are effective, it is not to say that they are perfect – escape mutants can arise quite quickly in patients. This is, in part, due to the fact that HCV virions replicate extremely quickly, producing around 10^{12} virions per day (Neumann *et al.*, 1998). A high rate of replication coupled with an RdRp with a high error rate which lacks a 3'-5' proofreading exonuclease activity leads to the generation of quasispecies within people living with HCV. Quasispecies are related but distinct populations of virus contained within one individual which will be unique to them (Weiner *et al.*, 1991; Martell *et al.*, 1992; Pawlotsky, 2006). This means that even treatment naïve patients may have certain populations of HCV quasispecies within their liver which are already resistant to HCV DAAs which, upon administration of DAAs, can be selected and then outcompete the non-resistant quasispecies of HCV, and this can lead to treatment failure.

Work on NS3/4A inhibitors and resistance mechanisms have found that mutations in residues which surround the catalytic triad and the oxyanion hole, with which the current P2-P4 macrocyclic drugs interact, can lead to resistance. Specifically, residues in positions 80 (Q80K), 155 (R155K/Q), 156 (A156V/T), and 168 (D168A/V) in the NS4 proteolytic domain have been associated with conferring resistance when mutated (Lin *et al.*, 2005; He *et al.*, 2008; Sarrazin *et al.*, 2015). The importance of these residues for conferring resistance was further highlighted when a crystal structure showing the interaction between simeprevir and a genotype 1b NS3/4A protein complex was release (PDB: 3KEE), showing that residues Q80, R155, and D168 helped to stabilise the interaction between simeprevir and the protein through a network of salt bridges between those residues (Cummings *et al.*, 2010). Further molecular modelling studies suggested that mutations in those positions can result in reduced interactions between simeprevir and the NS3 protease domain without affecting the proteolytic function of the protein, providing a route for the generation of escape mutants (Xue *et al.*, 2012).

Resistance mechanisms in the NS5A protein are less clear, partly due to the exact molecular mechanism being less clear. However, it has been noted in several studies that positions 31 (L31M) and 93 (Y93H/C) are of particular important for the development of resistance (Lemm *et al.*, 2010). The presence of both of these mutations in Domain I of the NS5A protein led to the belief that Domain I was important for the functioning of the protein. Moreover, it has been shown in patients receiving treatment that Y93H mutations can arise extremely quickly. Strikingly, 100 % of patients contained viral populations which contained this mutations after just 4 days of commencing treatment (Wong *et al.*, 2013). Further studies have shown that M28T, Q30H, and Q30R may also confer resistance (Cheng *et al.*, 2012).

Lastly, there are resistance mutations found within the NS5B protein which confer resistance to sofobuvir. The most documented resistance mutation associated with sofosbuvir resistance is the S282T mutation in NS5B. Threonine is only one methyl group larger than serine; however, it is posited that this is larger enough than serine to cause steric clashes which disfavour the incorporation of sofosbuvir into the growing RNA strand. This mutation does confer a high fitness cost to the protein and replicates less than S282 WT NS5B proteins, suggesting that sofosbuvir has a high barrier to resistance (Migliaccio *et al.*, 2003; Dutartre *et al.*, 2006; Lam *et al.*, 2012; Donaldson *et al.*, 2015) . However, other mutations have been found such as L159F, L320F, S15G, C223H, and V321I which contribute to the resistance of NS5B to nucleotide analogue inhibitors (Lam *et al.*, 2011; Tong *et al.*, 2014).

1. 3. 3. DAAs: The Imperfect Solution

Whilst DAAs are effective, they can drive mutations in HCV which promote escape variants arising in people living with HCV. Moreover, DAAs are expensive. According to information from Public Health Scotland, the NHS Scotland spent £35.9 million on HCV treatments over the 2023/2024 financial year. This broke down to £3.1 million spent on ledipasvir/sofosbuvir (£13,035.69 per unit), £21.6 million on sofosbuvir/velpatasvir (£13,032.44 per unit), £8.5 million on glecaprevir/pibrentasvir (£12,255.92 per unit), and £2.7 million spent on elbasvir/grazoprevir (£11.831.41 per unit) (Public Health Scotland, 2024). These drug regimens are expensive, and this can result in access to these drugs being inequitable. A study looking at access to DAAs found that LMICs had low reimbursement rates when compared to high-income countries which leaves DAA access inequitable (Marshall *et al.*, 2024) Moreover, even in high-income countries, there were found to be social barriers to receiving treatment in PWID: feelings of being unworthy of being cured, difficulty navigating the health care system, and stigma were all found to contribute to a lack of engagement from the PWID community with health care providers (Goodyear *et al.*, 2020; Antoniou *et al.*, 2023, 2025).

All of these issues with DAAs – driving resistance, cost, inequitable access – are compounded by the fact that HCV infections are usually occult with no apparent and obvious symptoms, therefore, unless people are actively being screened, they may not be diagnosed, and cannot start DAA treatment (Stroffolini and Stroffolini, 2024). Lastly, and most crucially, DAAs offer a cure but they do not offer long-term protection from HCV. Some people who were cured of HCV have gone on to become reinfected (Midgard *et al.*, 2016; Martinello *et al.*, 2017). If HCV is to be eliminated, an effective, inexpensive, and long-lasting vaccine must be developed.

1. 4. Vaccine Development

In general terms, vaccines aim to provide the body with immunity to specific pathogens by presenting the immune system with either an inactivated or weakened form of a pathogen unable to cause disease, or constituent parts of a pathogen. This allows the body to mount an immune response which can be remembered and then used to mount a defence in the case the immune system is exposed to the real pathogen in the future.

There is a great deal of evidence to suggest that a vaccine against HCV may be possible. From early chimpanzee studies to studies of human cohorts, there are multiple studies that suggest that reinfection with HCV is less severe than a primary infection. As noted, around 25 % of people who contract HCV will spontaneously clear the virus (Micallef, Kaldor and Dore, 2006). In reinfection studies, chimpanzees which cleared the virus were reinfected, and human cohorts who had spontaneously cleared the virus were monitored for reinfection. These studies proved a few things: firstly, that prior infection was not sufficient for sterilising immunity i.e. there was still detectable viraemia upon reinfection. However, peak viraemia was reduced, and so was the duration of peak viraemia (Farci *et al.*, 1992; Prince *et al.*, 2005; Page *et al.*, 2009; Sacks-Davis *et al.*, 2015). Moreover, many of these reinfection studies pointed out the importance of a strong T-cell response, particularly a strong $CD4^+$ and $CD8^+$ T-cell response, which is associated with viral clearance and protection from chronic infection (Bassett *et al.*, 2001; Major *et al.*, 2002; Nascimbeni *et al.*, 2003). One study looked at the impact of $CD4^+$ T-cells on HCV protection during a chimpanzee reinfection challenge study. They found that by depleting the $CD4^+$ T-cell prior to inoculation that the infection went on to persist despite the presence of memory $CD8^+$ T-cells within the liver of the chimpanzee (Grakoui *et al.*, 2003). Another study depleted CD8+ T-cells prior to rechallenging the chimpanzees with HCV, and found that viraemia was only brought back under control after it was possible to detect CD8+ T-cells in the liver following antibody-mediated depletion (Shoukry *et al.*, 2003). This interplay between $CD4^+$- and $CD8^+$ T-cells and their importance to mediating viral clearance is highlighted when considering what happens when reinfection leads to persistent infection. It has

been shown the failure of $CD4^+$ T-cells to proliferate early in infection and subsequent deletion of any HCV-specific T-cell response is associated with chronic HCV infection (Schulze Zur Wiesch *et al.*, 2012). Furthermore, it has also been demonstrated that exhaustion of $CD8^+$ T-cells early in infection - which can be measured through relative expression of a receptor called **programmed cell death-1** (PD-1) - correlated with persistent infection (Rutebemberwa *et al.*, 2008). Additionally, another group found that exhaustion of CD8+ T-cells can be modulated by $CD4^+$ T-cells (Khan *et al.*, 2022). Taken together, these studies suggest that a strong $CD4^+$- and $CD8^+$ T-cell response are required early on infection to offer protection from chronic HCV.

Regarding B-cells, Osburn *et al.* also found that 60 % of people who become reinfected with HCV produced **neutralising antibodies** (NAbs) during acute infection which were able to neutralise heterologous HCVpp (Osburn *et al.*, 2010). However, other earlier studies did not find a particularly strong B-cell/antibody response to reinfection. However, it has been shown in HCV-naïve people that production of **broadly neutralising antibodies** (bNAbs) early in infection aided in clearance of the virus and that bNAbs production waned as viraemia reduced. Inversely, in chronically infected people, there was a delayed onset of the production of bNAbs which continued as infection persisted (Pestka *et al.*, 2007; Osburn *et al.*, 2014). Moreover, there have been several animal studies which have considered how the administration of bNAbs prior to challenge, or during an established HCV infection would affect disease outcomes. In humanised mice and chimpanzees, prior administration of bNAbs mitigated or prevented the establishment of chronic HCV (Law *et al.*, 2008; Morin *et al.*, 2012; Z.-Y. Keck *et al.*, 2016), and in humanised mice, administration of AR3A, AR3B, and AR4A abrogated an established HCV infection (de Jong *et al.*, 2014)

Altogether, these studies suggest that a protective HCV vaccine will need to induce a strong and proliferative T-cell response and induce the production of bNAbs early upon infection.

1. 4. 1. Development of a T-cell Vaccine

Having shown that it is important to induce a strong T-cell response early after infection to produce protective immunity, the next question to ask how best to achieve this? Ward *et al.* in the early 2000s demonstrated that $CD8^+$ T-cells preferentially targeted the non-structural proteins of HCV (Ward *et al.*, 2002), coupled with the fact that the non-structural proteins are more conserved than the structural proteins, this seemed like a good place to start. In 2006, one chimpanzee study found that HCV challenge after administration of an adenovirus-vectored vaccine containing NS3 through NS5B from HCV genotype 1b elicited heterologous protection from subsequent HCV challenges (Folgori *et al.*, 2006). Based off of this study, a NS HCV T-cell vaccine seemed promising.

A few years later a meta-analysis of 12 different animal vaccine/challenge studies was released, and it found two things which would confound the previous studies: firstly, that level of T-cell response did not seem to correlate with vaccine success; secondly, the vaccines which contained only the structural proteins actually outperformed those that contained either only the non-structural proteins or a mix of structural and non-structural. However, these 12 different studies varied greatly in vaccine type, vaccine regimen, timings for prime-boosts etc. and so the findings may be influenced by the vast difference within the study designs of the included animal trials (Dahari, Feinstone and Major, 2010).

These initial studies eventually led to human studies for T-cell vaccine effectiveness. The first T-cell vaccine brought to clinical trial was that of an HCV core ISCOMATRIX vaccine. ISCOMATRIX is an adjuvant known to induce a robust $CD4^+$- and $CD8^+$ T-cell response (Baz Morelli *et al.*, 2012). The trial demonstrated that the vaccine did induce some response in healthy volunteers, and it was deemed as safe, but it was not considered good enough to move forward on to testing in at risk groups of individuals, with $CD8^+$ T-cell responses only being induced in 2 of the 8 individuals receiving the highest dose of the vaccine (Drane *et al.*, 2009).

Only one T-cell vaccine has been brought to a phase I/II safety and effectiveness trial in at-risk individuals. The vaccine strategy was a **chimpanzee Adenovirus 3** (ChAd3) and a **modified vaccinia Ankara** (MVA) heterologous prime boost with the vectors containing NS3 through NS5B, with the NS5B mutated to prevent replication (NSmut). In initial studies in healthy, not at-risk people, the vaccine was found to be safe and effective at inducing a sustained $CD4^+$- and $CD8^+$ T-cell response which mimicked the T-cell populations found in PWID who had spontaneously cleared the virus i.e. this vaccine seemed to be adequately priming the test cohort for fighting off HCV (Barnes *et al.*, 2012; Swadling *et al.*, 2014). Unfortunately, when the vaccine was moved in phase II trials, testing effectiveness of protection in an at-risk group of 455 PWID, the vaccine was found not to offer protection from chronic HCV. 202 subjects in the trial were given the prime-boost regime of the vaccine, and 199 were given placebos. There was found to be no difference in the development of chronic HCV infections between the two groups, with 14 individuals from each group developing a chronic HCV infection over the course of the trial. (Cox *et al.*, 2019).

One major hurdle to the success of this vaccine may have been the genetic variability between the vaccine strain and the diversity of HCV strains with which the people in the cohort came into contact. The non-structural proteins for this vaccine came from HCV genotype 1b and it was posited that perhaps variation in the epitopes primed against by the vaccine were different from those encountered by the people within the trial. Therefore, efforts are now being made to produce a pan-genotype T-cell vaccine by examining the conserved T-cell epitopes across the whole HCV genome and across all 7 genotypes, it is hoped that this next generation of ChAdOx-vectored vaccines may induced cross-reactive protection from HCV of all genotypes (Kelly *et al.*, 2015; von Delft *et al.*, 2018).

1. 4. 2. Development of a B-cell Vaccine

Given the failure of the ChAd3/MVA NSmut T-cell vaccine, and that a bNAb response early in infection is associated with clearance of the virus, this may suggest that both a T-cell and B-cell response would be required for an effective vaccine.

The primary targets for antibody-mediated neutralisation are the HCV glycoproteins E1 and E2, particularly E2, and bNAbs have been demonstrated to be able to recognise, bind, and neutralise viral quasispecies across the HCV genotypes (Giang *et al.*, 2012; Boisvert *et al.*, 2016; Bailey *et al.*, 2017). The exact mechanism of antibody-mediated neutralisation and how that drives viral clearance for HCV remains unclear, however, one study found two patients which produced bNAbs which targeted the front layer of E2, were able to neutralise autologous HCV strains, and drove evolution of the E2 down a fitness bottleneck, lowering overall fitness of the viral population, promoting clearance (Kinchen *et al.*, 2018). This suggests that the role of bNAbs in HCV infection may be two-fold; direct neutralisation and reducing the fitness of escape mutants through driving their evolution. This is also further proof of concept that the production of bNAbs may aid in viral clearance and should be considered when designing an HCV vaccine.

How T-cell and B-cell vaccines are rationally designed vary greatly due to the how the T-cell and B-cell response operate. In brief, T-cells present short peptides on their **major histocompatibility complex** (MHC) receptors. These peptides can be exogenously or endogenously acquired. Endogenous antigen acquisition involves degradation of the antigen by the proteasome, and then presentation of the antigen on MHC-I receptors which can activate CD8+ T-cells. Exogenously acquired antigens are treated slightly differently. They are degraded in the phagolysosome after engulfment, and peptide antigens are presented on MHC-II receptors, activating CD4+ cells (Abbas, Lichtman and Pillai, 2017).

B-cells, however, undergo several stages of activation and proliferation with the end goal of producing antibodies which are able to bind to specific antigens within the body and mediate downstream processes to facilitate clearing of the antigen/pathogen. In brief, mature, naïve B-cells (IgM+, IgD+), are activated upon binding an antigen in its native 3D conformation via their **B-cell antigen receptor** (BCR) which is composed of membrane bound Ig molecules – either IgM or IgD – and the associated Igα and Igß proteins. Additionally, Helper T-cells and other stimuli through may aid in activating B-cells. The other stimuli can include things such as detection of complement proteins, and **pathogen-associated molecular patterns** (PAMPs) by **complement receptor type 2** (CR2) or **Toll-like receptors** (TLRs) respectively.

Upon activation, the B-cells begin to proliferate and differentiate. During this process, the B-cells may undergo other processes such as heavy chain isotype switching (class switching) where they begin to produce other Ig isotypes rather than just IgM or IgD. Additionally, B-cells may also undergo a process of affinity maturation in the B-cell germinal centres. During this process, follicular dendritic cells in the germinal centres present antigens against which the immune response is being raised. B-cells in the germinal centre dark zone undergo somatic hypermutation of their V genes which alters the conformation of the **complementarity-determining regions** (CDRs). In turn, this alters the affinity of the BCR present on the differentiating B-cell to the antigen presented by the follicular dendritic cell. If the newly differentiated B-cell is able to recognise and bind the antigen presented by the follicular dendritic cell, this interaction prevents the B-cell from undergoing apoptosis. However, if the B-cell cannot bind the antigen, it undergoes apoptosis. Over the course of the immune response, the affinity of the receptors produced by B-cells for the antigen increases.

Following affinity maturation, B-cells may undergo class switching, and then circulate as either memory B-cells or antibody-producing terminally differentiated plasma cells. The production of antibodies can then mediate downstream immune responses through various means such as opsonisation of microbes which can lead to phagocytosis of the pathogen. However, for viruses

it may only be necessary for antibodies to bind to the viral glycoproteins and prevent the binding of the virus to their receptor. This could prevent infection by stopping any virus:host receptor interactions and stopping the glycoproteins from undergoing conformational changes that would typically facilitate viral fusion and the commencement of infection (Abbas, Lichtman and Pillai, 2017; Cyster and Allen, 2019).

All this is to say is that T-cell vaccines do not have to consider the 3D structure of their epitopes, B-cell vaccines do.

As noted, the main targets for antibody-mediated neutralisation of HCV are the glycoproteins E1 and E2. To date, only one B-cell vaccine has been administered to healthy, not at-risk humans in a clinical trial (Frey *et al.*, 2010). The vaccine was based on the E1E2 glycoproteins from genotype 1a purified from **Chinese Hamster Ovary** (CHO) cells and adjuvanted with MF59C.1 (a squalene-in-water emulsion), however, whilst safe and able to induce generation of anti-HCV antibodies and proliferation of CD4+ T-cells, the types of antibodies produced were not as cross-reactive as perhaps necessary for a protective vaccine. Rather, in one follow up study, it was found that only one test participant produced broadly cross-reactive antibodies to all HCV genotypes, whilst others produced antibodies which were capable of neutralising some heterologous HCV strains but not all (Law *et al.*, 2013). Moreover, it was found in another follow-up study that around 50 % of the anti-HCV antibodies produced targeted the HVR-1, which may be a problem as the HVR-1 is thought to function as an epitope decoy and actually mitigate neutralisation, potentially through undergoing rapid mutational escape (Ray *et al.*, 2010).

To improve future B-cell vaccine candidates, there needs to be greater understanding of how the HCV E1E2 glycoproteins work, to better guide rational vaccine design – what are the moving parts that are essential for function which can be targeted for neutralisation. Understanding the moving parts of viral glycoproteins has aided rational vaccine design, particularly for viruses with a class I fusion mechanism such as SARS-CoV-2 (Wrapp *et al.*, 2020; Klasse, Nixon and Moore, 2021). In rationally designing vaccines for viruses, the general

principle is to design a stable, soluble antigen which mimics the structure of the pre-fusion glycoprotein. This can be achieved through a variety of alternations to the native-state glycoprotein such as proline substitutions to disfavour α-helix formation, altering fusion peptide exposure, deleting the cleavage site, and addition of disulphide bonds to stabilise the pre-fusion glycoprotein/glycoprotein trimer (Anasir and Poh, 2019; Sanders and Moore, 2021).

To aid rational vaccine design, it is crucial to know the pre- and post-fusion state of the viral glycoproteins, and moreover, to know how the glycoproteins transition from one state to the other. There are many different facets to this question which each need answered to solve the larger question of how to design an effective B-cell HCV vaccine. For example, work by Kalemera *et al.* presented an optimised HCVpp production method, complimented by work by Chumbe *et al.* who produced a standardised panel of HCVpp which antibodies could be tested against (Kalemera *et al.*, 2021; Chumbe *et al.*, 2024). Taken together, their work allowed for optimised production of specific HCVpp with a diverse range of antigenically distinct E1E2 complexes – a crucial addition to the toolbox for developing an effective and protective HCV vaccine.

Currently, it is known that the E1E2 complex is the HCV fusion machinery. However, it is unknown exactly how these proteins work in concert to bring about viral fusion and this has hampered rational HCV vaccine designs. Current efforts are based on presenting particular epitopes through epitope scaffolding (He *et al.*, 2015; Pierce *et al.*, 2017), optimising the antigenicity of E2 alone (Marín *et al.*, 2020; Pierce *et al.*, 2020), optimising the antigenicity of E1E2 in complex (Cao *et al.*, 2019; Guest *et al.*, 2021), and through novel approaches such as the use of two-component nanoparticles to display E2E1 immunogens (Sliepen *et al.*, 2022). Other efforts to aid in rational vaccine design tackle the problem from the other direction. Instead of asking what the best immunogens are, other groups have investigated the best antibodies for the body to produce to overcome HCV. Using biphasic antibodies Radić *et al.* have started to uncover how to produce the most effective anti-HCV antibodies which could be used as either a therapeutic, or will elucidate exactly what is required of an effective HCV vaccine (Radić *et al.*, 2025). However, none of these approaches have led

to a successful vaccine candidate coming to market yet, and therefore further research into how the glycoproteins of HCV work must be undertaken to aid in further rational vaccine designs.

1. 5. Viral Fusion Mechanisms

To date, three viral fusion mechanisms have been described and characterised. In general, enveloped viruses are studded with fusion proteins which are held in a metastable pre-fusion conformation which, upon interaction with certain cellular triggers, irreversibly transition to a lowest-energy, post-fusion state. The release of energy from the conformational transition is sufficient to bring the two membranes together, overcoming the repulsion between the membranes caused by the hydration force between them. During this conformational change, there is an extension of a fusogenic peptide, a non-polar protein segment, which is inserted into the membrane of the target cell. The extended intermediate state protein then folds back into a “hairpin” confirmation, this pulls the fusogenic peptide, which has been inserted into the target membrane, back towards the transmembrane domain of the viral fusion protein, bringing the two membranes together (Harrison, 2015; Rey and Lok, 2018). However, there are certain variations on this theme.

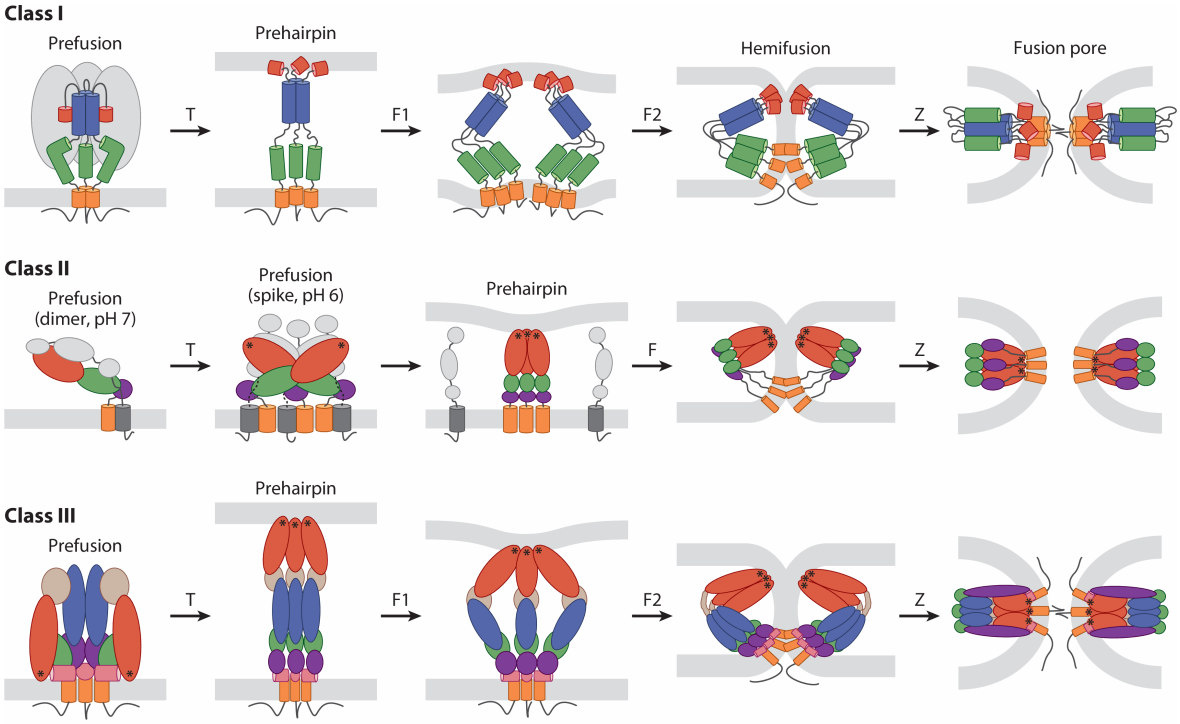


Figure 1. 7. **Schematic Overview of Viral Fusion Mechanisms.** T = specific trigger, F = fold-back, F1, F2, F3 represent progressive stages of fold-back, Z = zippering. Class I is represented by the HIV Env proteins, class II by Semiliki Forest virus (SFV) E1 protein, and class III by Herpes Simplex virus 1 (HSV-1) gB protein. For all three of these classes of viral fusion mechanisms the first stage is triggering in response to a specific molecular cue. This can be a drop in pH, receptor binding, or a combination of them both. Notably, for class II proteins, a drop in pH may facilitate dimeration of the glycoproteins, a necessary step in fusion, which allows the class II glycoproteins to trimerise to form the pre-fusion extended hairpin. Generally, the formation of the extended prefusion hairpin inserts the fusion peptide from the glycoprotein into the target cell membrane. Following this, the proteins will commence a process of folding back which will bring the membranes together. As the membranes are brought together, the membranes go from distinct bilayers, through hemifusion, to the formation of a pore in the membrane through which viral contents - genomic material, proteins etc - can be translocated into the target cell.

Figure derived from White *et al.* (2023).

1. 5. 1. Class I Viral Fusion Mechanisms

Class I fusion proteins exist as homotrimeric, single-chain precursor proteins. Example of viruses which use class I fusion mechanisms include **Influenza A Virus** (IAV) and **Human Immunodeficiency Virus** (HIV). This class is typified by the presence of a single-chain precursor which must undergo maturation to produce N-terminal and C-terminal functional subunits. For IAV, the HA_0 precursor undergoes proteolytic cleavage events to produce HA_1 and HA_2 subunits which are bound together by disulphide bonds. For IAV, HA_1 participates in receptor recognition, and binding. Following recognition, the HA_1 dissociates from the prefusion trimer, and the B-loops found within the HA_2 subunits undergo a loop-to-helix transition, bringing the fusion peptide approximately 100 Å closer to the target membrane, which is followed by a foldback event which involves the C-terminal segment of the HA_2 “leash” domain binding with grooves in the N-terminal helix bundles, this results in the defining feature for class 1 fusion proteins – an α-helical coiled coil in the post-fusion confirmation (Park, Gruenke and White, 2003; Xu and Wilson, 2011; Benton *et al.*, 2020).

1. 5. 2. Class II Viral Fusion Mechanisms

Class II viral proteins are found in the *Orthoflaviviruses* (Dengue, Yellow Fever, and West Nile virus), *Alphaviruses* (Chikungunya, Sindbis, Semliki Forest virus) and *Bunyavirales* (Crimean Congo Haemorrhagic Fever virus, Lassa virus, Rift Valley fever virus) among others. They are structurally distinct from class I fusion proteins, whilst using a similar topological approach to drive membrane fusion. For dengue, fusion is achieved by the envelope, or E protein which, common to all class II fusion proteins, is formed of three domains. These domains are domain I, formed of an 8-stranded ß-barrel into which domain II is inserted. Domain II is comprised of 12 ß-strands and two α-helices and contains the fusion loop. Domain III is made of ten ß-strands and is thought to aid in receptor-binding. On the virion, E proteins form an icosahedral lattice of E protein dimers. Decreased pH within the endosome leads to residue protonation and E dimer disassociation, exposing the fusion loop on domain II, whilst concomitantly allowing the E protein ectodomain to form an extended intermediate. Thusly, low pH for dengue E protein, releases the fusion loop and extends it to where it can interact with the target membrane. Upon interaction of the fusion loop and the target membrane, the E proteins begin to trimerise and conformation changes are induced in the E protein to facilitate collapse of the extended intermediate, bringing the viral and host membranes together (Modis *et al.*, 2003; Modis, 2013; Chao *et al.*, 2014; Villalaín, 2025).

1. 5. 3. Class III Viral Fusion Mechanisms

Lastly, there are class III viral fusion proteins, which combine some elements of both the class I and class II proteins – they contain a central trimeric coiled-coil, and three of the five protein domains are mostly ß-sheets. In 2006, two structures of class III fusion proteins were released – one of the **glycoproteins** (G) of **Vesicular Stomatitis Virus** (VSV), often termed VSV-G, and another of **glycoprotein B** (gB) from **Herpes Simplex Virus 1** (HSV-1). Despite coming from entirely different viruses, their fusion proteins were structurally homologous, and thus thought to represent the same class of viral fusion proteins (Heldwein *et al.*, 2006; Roche *et al.*, 2006). Class III fusion proteins contain five domains. Domain I contains the two predicted fusion loops and depending on the virus can be formed of different ß-sheet arrangements. The second domain contains one or more anti-parallel ß-sheets, the third contains a long, central α-helix which interacts with other monomers within the trimer of fusion proteins to form the trimeric coiled coil, typical of class I fusion proteins. The fourth domain is variable between viruses but is comprised entirely of ß-sheets and is thought to be intrinsically flexible. Lastly, domain V is thought to insert between the other monomers of the trimers, aiding in trimerisation (Backovic and Jardetzky, 2009). Comparison of the pre- and post-fusion states, reveals that domain III undergoes significant structural changes wherein the central α-helix extends, repositioning the C-terminal helices into the grooves formed by the other monomers in the post-fusion state – similar to the leash in the groove conformation of HA. Moreover, whilst the folds of the other domains remain the same, the structural rearrangement of domain III forms an extended intermediate whereby the domain I, containing the fusion loops, is positioned extended towards the target cell membrane (Roche *et al.*, 2007, 2008; Backovic and Jardetzky, 2009; White *et al.*, 2023). Unlike in other glycoproteins, these conformational changes in the VSV-G protein can be reversible – they are induced by low pH and can be reversed when pH returns to neutral. This may be due to the presence of many histidine residues within domain I acting as a sensor, and initiating the conformation changes required for fusion (Backovic and Jardetzky, 2009). Further work has suggested that this VSV-G proteins need to go through a monomeric intermediate state, which allows dimerisation of VSV-G

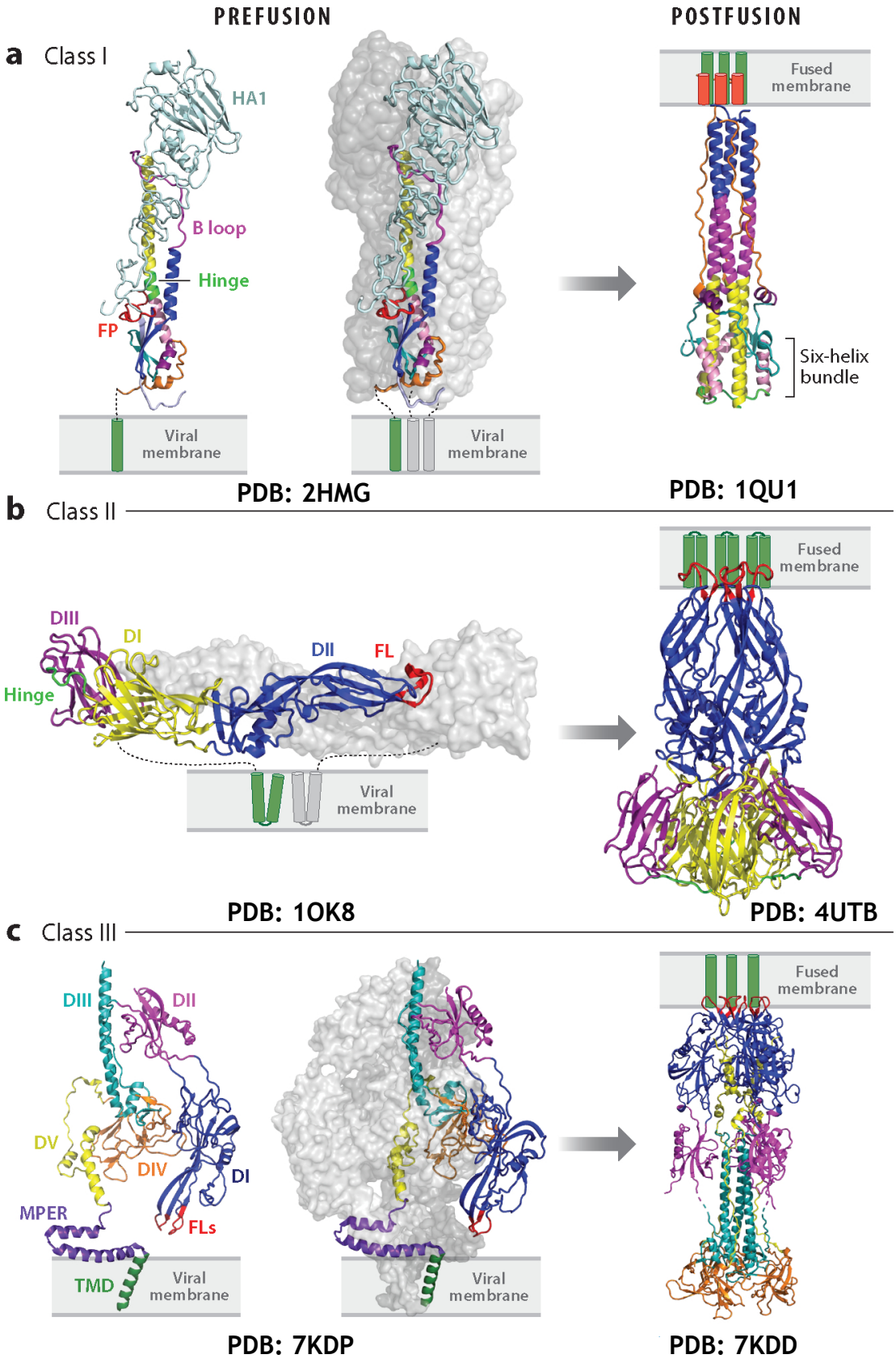


Figure 1. 8. **Pre- and Post-fusion Structures of Class I, II, and III Viral Glycoproteins.** Structures are coloured as per the labels in the left-most structure. Class I is represented by Influenza A virus (IAV) HA_1, class II by Dengue E protein, and class III by Human Cytomegalovirus (HCMV) gB protein.

Figure derived from White *et al.* (2023).

proteins, which in turn results a post-fusion trimeric state but the exact mechanisms of this are as yet unknown (Baquero *et al.*, 2017)

An overview of the differences in pre- and post-fusion structures of class I, class II, and class III structures can be seen in figure 1.8. However, HCV does not appear to display a class I, II, or III fusion mechanism. Instead, it seems to house a novel mechanism for viral fusion. Ergo, to aid in rational design of immunogens which can stimulate a bNAb-response in vaccines, there needs to be a better understanding of how the moving parts in E1E2 operate to facilitate fusion so these epitopes can be targeted.

1. 6. The E1E2 Glycoproteins – Evidence for a Novel Fusion Mechanism and Insights for Vaccine Design

E1E2, the focus on this PhD, were briefly mentioned when discussing viral entry and fusion. Here their roles in viral entry, immune evasion, and escape shall be discussed in more detail.

1. 6. 1. E1

From N- to C-terminus, there are 6 distinct domains within the E1 protein structure: The N-terminus domain, the ß-sheet core, the **putative fusion peptide-containing region** (PCR) contained within a helical hairpin, and a C-terminal loop which runs from the PCR to the transmembrane proximal stem region, and finally the transmembrane domain (Torrents de la Peña *et al.*, 2022; Oliver *et al.*, 2023). E1 is reported to have between 4 and 6 glycosylation sites, (N196, N209, N234, N250[1], N305, and N325[2]). These glycans are thought to form part of the glycan shield and aid in stabilisation of the E1E2 complex but these functions will be discussed more in detail when discussing the E1E2 complex itself. Moreover, E1 contains 4 disulphide bonds which are thought to be important in E1 function and conformation (Meunier *et al.*, 1999; Torrents de la Peña *et al.*, 2022; Carriquí-Madroñal *et al.*, 2023).

[1] Found in genotypes 1b and 6 only
[2] Dependent on whether a proline follows the glycosylation sequon NXS/T in this position

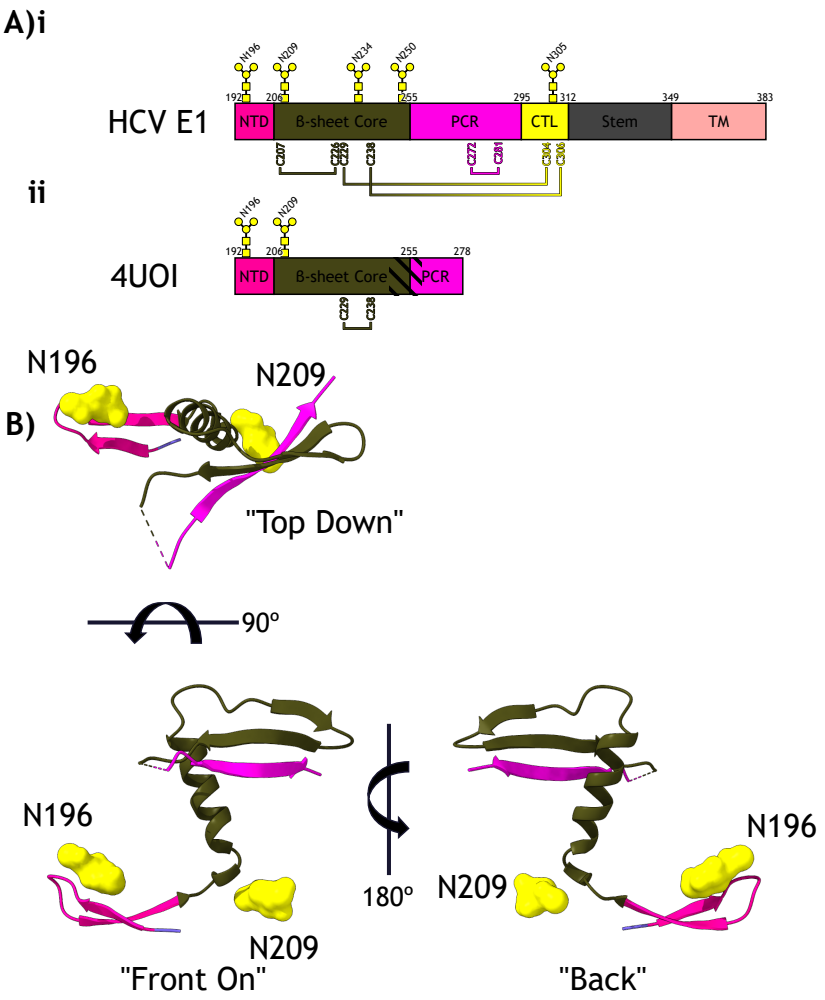


Figure 1. 9. **X-Ray Diffraction Crystral Structure of the N-Terminus of H77 E1 (PDB: 4UOI).** A) i HCV E1 protein architecture; glycans are denoted above, and disulphide bonds below. E1 structure is coloured as per this schematic. NTD = N-Terminal Domain, PCR = Putative fusion peptide-Containing Region, CTL = C-Terminal Loop, TM = Transmembrane domain. A) ii schematic of the E1 protein which was crystallised with the glycans and disulphide bonds which were resolved in the structure denoted. The dashed lines represent parts of the input sequence which were not resolved. B) X-ray crystal structure of 4UOI with glycans. The bottom left structure is "front on", the structure to the right is rotated 180° on the y-axis relative to the first structure, and the top structure is rotated 90° towards the reader relative to the "front on" structure.

To date, there has been one crystal structure of the **N-terminus domain** (NTD) of E1 (Fig. 1.9.). The E1 NTD was crystallised at a pH of 4.8, which is around the pH at which the E1E2 complex may undergo conformational changes to induce viral:host membrane fusion (Bartosch, Vitelli, *et al.*, 2003; Hsu *et al.*, 2003). However, the authors do not state if this low pH was to mimic these physiological conditions which may cause the truncated E1 NTD to fold into its post-fusion conformation or if this was simply the conditions under which the protein successfully crystallised. Moreover, the structure obtained was unexpected, and more akin to a phosphatidylcholine transfer protein structure than other known glycoprotein structures (El Omari *et al.*, 2014). This may hint to E1 operating in a completing unique manner when compared to other fusion proteins. On the other hand, it has also been shown that E1 requires E2 in order to fold correctly (Patel, Patel and McLauchlan, 2001). Therefore, this truncated E1 crystal structure may not be physiologically relevant or correct – it may simply represent an artifact of the crystallisation process. Moreover, subsequent structures which show both E1 and E2 (PDBs: 7T6X and 8RJJ, (Torrents de la Peña *et al.*, 2022; Augestad *et al.*, 2024) do not display any structural similarity with this truncated E1-only structure.

Moreover, machine learning approaches to protein folding (e.g. AlphaFold) have suggested that the unique fold of E1, observed in recent cryo-EM structures, is conserved amongst closely related hepaci-, pesti- and pegiviruses despite a lack of conservation of sequence. This may suggest that the fold is being maintained evolutionarily to conserve the function(s) of E1 across these different genera of viruses. The same cannot be said for the E2 proteins across these genera which are predicted to exhibit mostly different structures; albeit that the C-terminus of the E2 displays structural conservation across these genera (Mifsud *et al.*, 2024). Taken together, this suggests a conserved function for the E1 proteins across these closely related viruses, which could potentially be to act as the fusogen (Oliver *et al.*, 2023).

As noted, E1 contains a PCR (Flint *et al.*, 1999) which has been shown to be necessary for viral fusion (along with two discrete regions in E2) (Lavillette *et al.*, 2007). However, E1 is a small protein, too small to operate as a class II or

class III fusion protein, and it is also lacking structural features of class I fusion proteins such a leucine zipper motif which aids in the foldback mechanism necessary to bring two membranes together (Li and Modis, 2014). Taken together, these data suggest that if E1 is the fusogen, then it is operating in a distinct manner to other viral fusogens. However, E1 does contain some structural motifs which are typical features of other viral fusion proteins such as a GXXXG motif in the transmembrane domain which can facilitate trimerisation of E1 itself (Falson *et al.*, 2015). Previously, this GXXXG domain was thought to aid in heterodimerisation with E2. However, examination of the latest E1E2 complex structure (PDB: 8RJJ) shows that this does not appear to be the case (Ciczora *et al.*, 2007; Augestad *et al.*, 2024).

Additionally, other functions have been described for E1 previously. There is some research to suggest that E1 facilitates the interaction between the E1E2 complex and the receptor CLDN1. Moreover, it has been demonstrated that mutations within the E1 protein can allow for the E1E2 to recognise and bind CLDN6 as opposed to CLDN1. E1, therefore, appears to have a role in determining receptor usage, and consequently, cellular tropism of HCV (Douam *et al.*, 2014; Hopcraft and Evans, 2015; Haddad *et al.*, 2017). Lastly, there is some sparse evidence to suggest that E1 may also directly aid in viral morphogenesis by potentially acting in concert with NS2 (Ma *et al.*, 2011; Tong *et al.*, 2017).

Lastly, several antibodies have been found and characterised which target E1. Whilst many more E2-targeting antibodies have been identified, it is important to note that E1 is a small protein and potentially buried somewhat in the membrane as seen in the recent E1E2 complex structures. This would reduce the possibility of antibodies being able to bind to E1, and may explain the discrepancy in the numbers of antibodies which target each of the HCV glycoproteins (Torrents de la Peña *et al.*, 2022; Augestad *et al.*, 2024). That being said, several anti-E1 antibodies have been discovered.

H-111 was characterised from a patient living with a chronic genotype 1b HCV infection. Research by Keck *et al.* established that this antibody bound to an

epitope in E1 that mapped to R195, N196, S198, G199, and Y201. Other antibodies against E1 include HEPC112 and A6, both of which bind to **antigenic site 112** (AS112, amino acids 230–239) on E1 (Mesalam *et al.*, 2018; Colbert *et al.*, 2019). Two more significant anti-E1 antibodies are IGH526 and IGH505; both of these antibodies were derived from a patient with a chronic genotype 1b HCV infection who cleared the infection following IFN-therapy. These two antibodies were found to be broadly neutralising and targeted overlapping but distinct epitopes in the C-terminus of E1 (amino acids 313 to 327) (Meunier, Russell, Goossens, *et al.*, 2008).

Importantly, however, IGH505 was found to bind to a surface-exposed α-helix in E1, and was used to stabilise the E1E2 crystal structure produced in 2022 (Torrents de la Peña *et al.*, 2022).

Altogether, there is evidence to suggest that E1 may have multiple functions: mediating CLDN usage and interactions, potentially aiding in viral morphogenesis through NS2 interactions, and may be the viral fusogen. However, the evidence to suggest that E1 is the fusogen is mixed. It is small, buried in the membrane, and lacks features consistent with other viral fusogens (e.g. a leucine zipper motif) which would argue against the feasibility of E1 being the fusogen. Conversely, structural predictions of E1 proteins from different related genera of viruses show a conserved fold for the whole of the E1 protein. Whilst this implies a conserved function for these related E1 proteins, it does not directly imply that this function is to mediate fusion. At this point, all that can be said confidently is that more research is required to understand the exact function of E1 and how it contributes to the HCV entry mechanism.

1. 6. 2. E2

The E2 protein is a highly glycosylated protein featuring 11 glycans and several distinct regions. From N- to C-terminus, these regions include: HVR-1, **Antigenic Site 412** (AS412), the front layer, **Variable Region 2** (VR2), the IgG-like ß-sandwich core which contains the CD81 binding loop, **Variable Region 3** (VR3), post-V3, the back layer, the stem region, and finally the **transmembrane** (TM) domain (Kong *et al.*, 2013; A. G. Khan *et al.*, 2014; Castelli *et al.*, 2017; Torrents de la Peña *et al.*, 2022; Augestad *et al.*, 2024). As discussed, structural evidence from studies of E1 suggested that perhaps HCV uses a unique fusion mechanism, this is further substantiated by structural studies of E2.

Further evidence for a novel fusion mechanism came from the first structures of E2 to be released (PDB: 4MWF and 4WEB, Kong *et al.*, 2013; A. G. Khan *et al.*, 2014, Fig. 1. 10, and 1. 11.) which both showed the core of E2 bound to Fabs. Until the release of these structures, it was predicted that perhaps E2 from HCV would resemble other class II fusion proteins as found in other members of the *Flaviviridae* (Yagnik *et al.*, 2000; Krey *et al.*, 2010) and would therefore have a similar fusion mechanism. However, this was not the case.

The first structure was solved through the co-crystallisation of the core of H77 E2 (genotype 1a) with the Fab AR3C (Fig. 1. 10.). The team made several truncations to the full-length E2 protein to enable crystallisation including the removal of the N- and C-terminus, replacement of the VR2 with a short GSSG linker, and removal of N448 and N576, asparagine residues known to be glycosylated. In figure 1. 10. Ai-ii, a schematic of a hypothetical standard HCV E2 protein is shown next to another schematic of the actual E2 sequence used by the researchers to generate the 4MWF crystal structure.

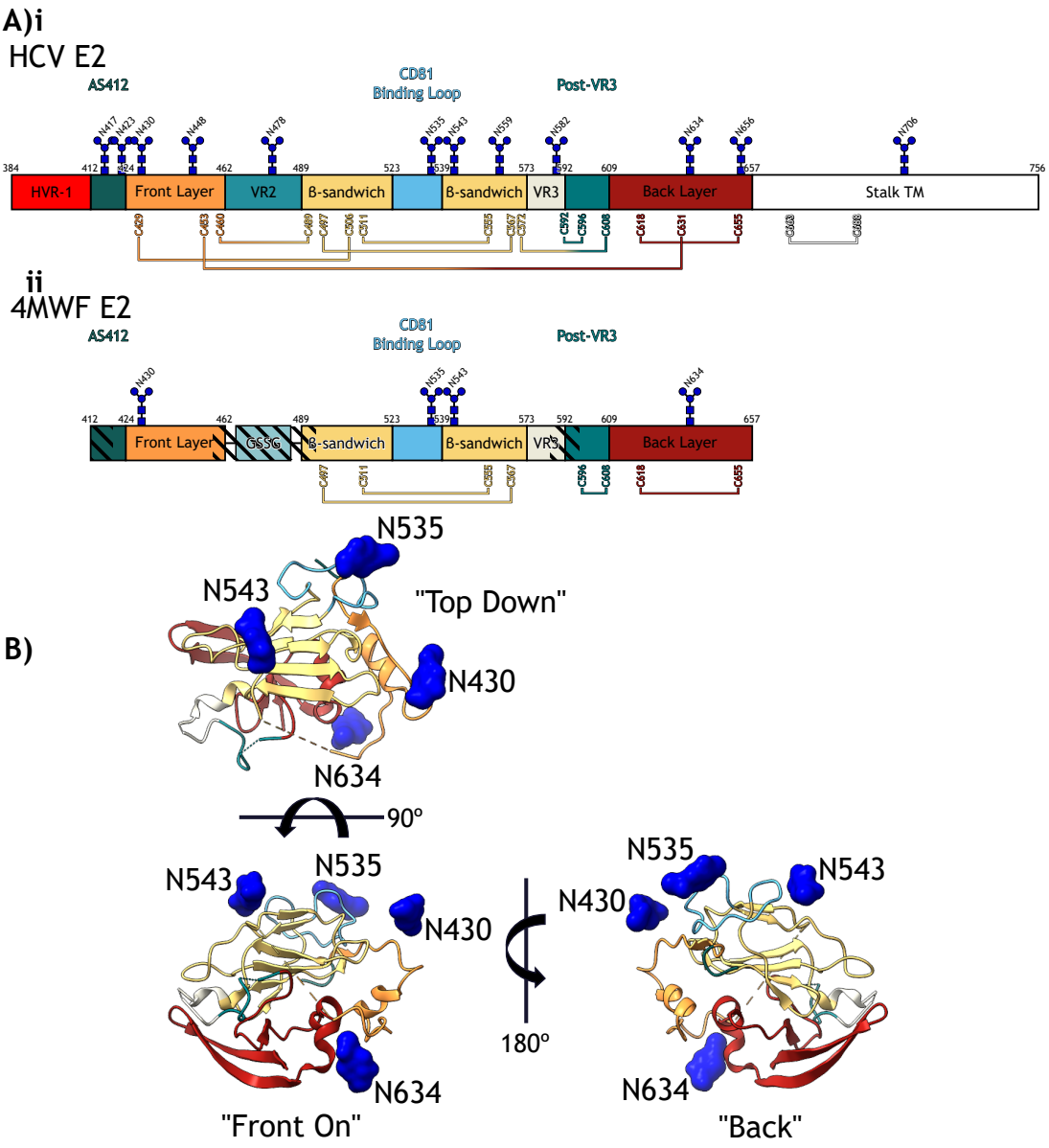


Figure 1. 10. **X-Ray Diffraction Crystral Structure of Isolate H HCV E2 (PDB: 4MWF).** A) i HCV E2 protein architecture; glycans are denoted above, and disulphide bonds below. E2 structure is coloured as per this schematic. HVR-1 = Hypervariable Region 1, AS412 = Antigenic Site 412, VR2 = Variable Region 2, VR3 = Variable Region 3. A) ii schematic of the 4MWF E2 protein which was crystallised with the glycans and disulphide bonds which were fully resolved in the structure denoted. The dashed lines represent parts of the input sequence which were not resolved in the structure. B) X-ray crystal structure of 4MWF with glycans. The bottom left structure - "front on" - was oriented by structural alignment of 4MWF to 8RJJ in ChimeraX v1.9 such that the viral membrane would be below the "front on" model, with the protein shown perpendicular to the plane of the viral membrane. The structure to the right - "back" - is rotated 180° on the y-axis relative to the "front on" structure, and the "top down" structure is rotated 90° towards the reader relative to the "front on" structure.

The second structure came from the core of J6 E2 (genotype 2a, residues 456 - 656) containing an N542Q mutation in complex with the non-neutralising 2A12 Fab (Fig. 1. 11). N542, in J6, is known to be glycosylated which is why the team chose to mutate out this residue to aid in crystallisation. Again, two schematics can be seen in figure 1. 11. – a hypothetical E2 protein, and the actual E2 sequence used by Khan *et al.* Moreover, SAXS experiments from Khan *et al.* showed that E2 did not undergo dramatic conformational re-arrangements at low pH, nor when interacting with the receptor CD81 (A. G. Khan *et al.*, 2014). However, future work by this group does show that interaction of E2 with CD81 does, in fact, induce conformational changes in E2 (Kumar *et al.*, 2021).

Both structures highlighted the compact globular fold of the core domain of the E2 glycoprotein, containing 2 anti-parallel ß-sheets forming a ß-sandwich, an IgG-like fold in the centre of the protein. This ß-sandwich core was flanked by both a front and a back layer. Both structures were in agreement with each other, exhibiting a 0.8 Å RMSD between the Cαs of each structure (Kong *et al.*, 2013; A. G. Khan *et al.*, 2014).

As noted, comparative structural prediction studies found that E2s from closely related genera of viruses differed in their overall structures but shared a conserved C-terminus. It was suggested, therefore, that perhaps E2 played a greater role in mediating the life cycle of the virus, and combatting host defences (Oliver *et al.*, 2023; Mifsud *et al.*, 2024). This potential role in aiding to overcome the host immune system may help in explaining the plethora of anti-E2 antibodies which have been discovered and characterised (Table 1. 4.).

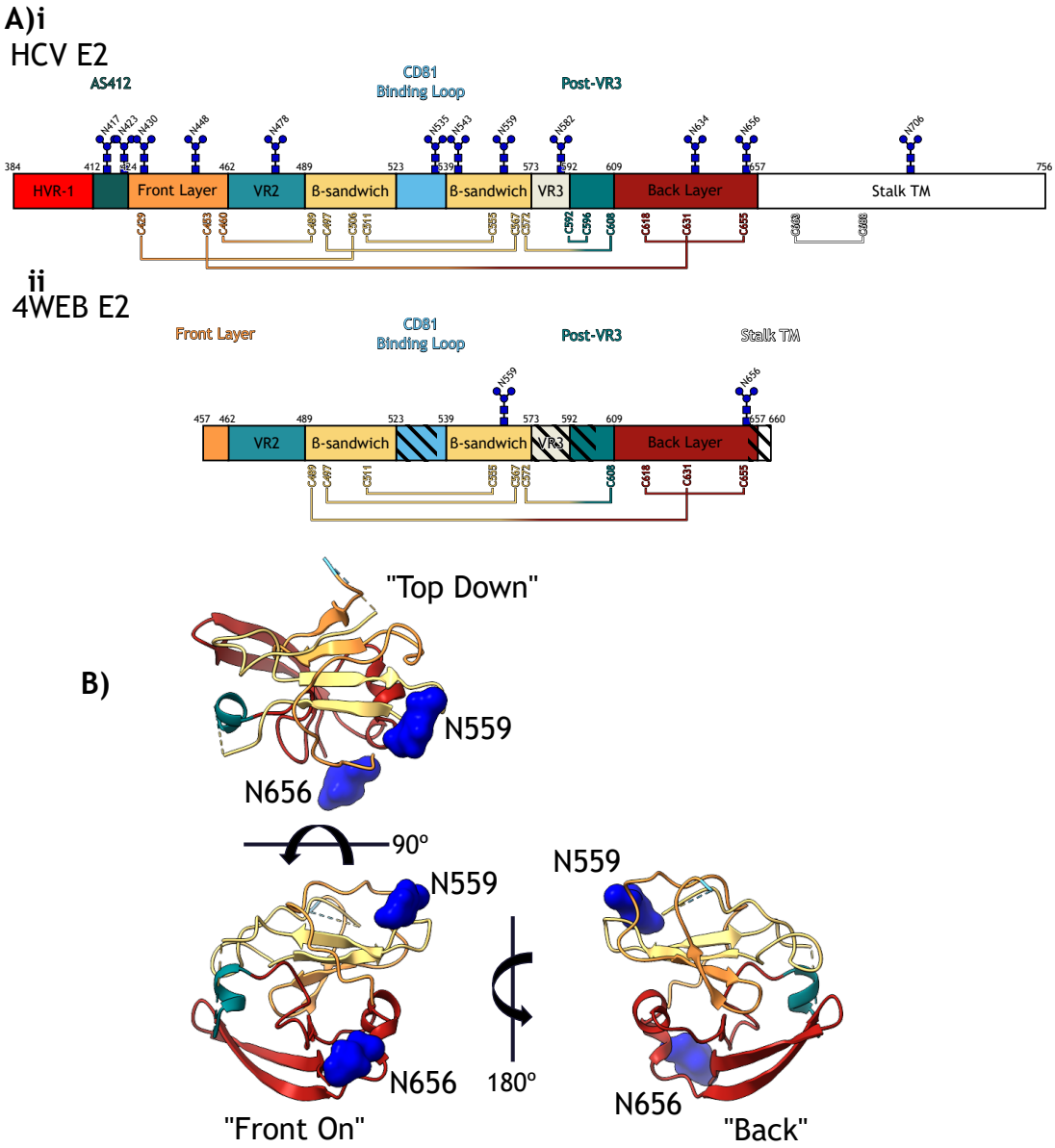


Figure 1. 11. **X-Ray Diffraction Crystral Structure of HCV J6 Core E2 (PDB: 4WEB).** A) i HCV E2 protein architecture; glycans are denoted above, and disulphide bonds below. E2 structure is coloured as per this schematic. HVR-1 = Hypervariable Region 1, AS412 = Antigenic Site 412, VR2 = Variable Region 2, VR3 = Variable Region 3. A) ii schematic of the 4WEB E2 protein which was crystallised with the glycans and disulphide bonds which were fully resolved in the structure denoted. The dashed lines represents parts of the input sequence which were not resolved in the structure. B) X-ray crystal structure of 4WEB with glycans. As with the 4MWF structure in figure 1. 10., the 4WEB structure was structurally aligned with 8RJJ such that the "front on" structure is perpendicular to the plane of the viral membrane. As before, the "back" structure is a 180° rotation of the "front on" structure on the y-axis. The "top down" structure is a 90° rotation toward the reader of the "front on" structure on the x-axis.

Table 1. 12. Summary of some of the known anti-E2 mAbs. Derived from (Sautto et al., 2013; Keck et al., 2018).

Antigenic Domain (A-E)	Epitope (Residues Numbered According to H77)	E2 Domain	Antibodies
HVR-1	1-27	HVR-1	J6.36 H77.16 HEPC98 9/27 J6.103 HCV-AB68
A	581-584, 627-633	Post VR3, and Back Layer	CBH-4B CBH-4D CBH-20 CBH-21 A33
B*	431-439, 529-535	Front layer and CD81 binding loop	HC-1 HC-11 AR3A AR3C CBH-5 HEPC3 HEPC74 HC84.1 HC84.26 HC84.27
C	544-549	B-sandwich	CBH-7 CBH-23 AR1A AR1B HEPC50
D*	420-428, 441-443, 616	AS412, Front layer, 616 in Back layer	See Domain B.
E	412-423	AS412	AP33 HCV1 H77.39 HC33.1 HC33.4

***There is overlap between antigenic domains B, D. For brevity, the list of mAbs has only be written once but applies to both. Moreover, domains B and D are otherwise referred to as Antigenic Region 3 which represents the neutralising face of E2**

Another key role of E2 is that of interacting with CD81. As noted, there are plenty of anti-E2 antibodies which have been characterised and a lot of these target residues which are important for CD81 interaction (Tzarum, Wilson and Law, 2018). It was demonstrated in 1998 that E2 binds directly to the LEL of CD81 (Pileri *et al.*, 1998) and further work demonstrated that E2:CD81 interactions were a priming step for fusion to occur as low pH alone was not enough to render HCV virions non-infectious through causing misfiring of the glycoprotein complex (Tscherne *et al.*, 2006; Sharma *et al.*, 2011). Previously, it has been described how the E2:CD81 interaction recruits CLDN-1 which aids in facilitating clathrin-mediate endocytosis. However, the group that produced the 4WEB structure went on to perform another study which compared various crystal structures and their work has suggested that the E2:CD81-LEL interaction may be required to re-arrange the CD81 binding loop into an extended confirmation which they suggest would allow for E2 to operate as a fusogen, rather than E1 (Kumar *et al.*, 2021, 2023).

However, as with E1, the evidence here is mixed, with E2 lacking any classical structural motifs of other classical viral fusion proteins. Additionally, the proposed E2 fusion loop is unique to HCV i.e. no related E2 proteins contain a homologous structure (Oliver *et al.*, 2023). As previously discussed, the C-terminus of the E2 proteins from HCV and other related genera are conserved but if the E2 proteins are the fusogens, perhaps a higher level of conservation between these related proteins may be expected.

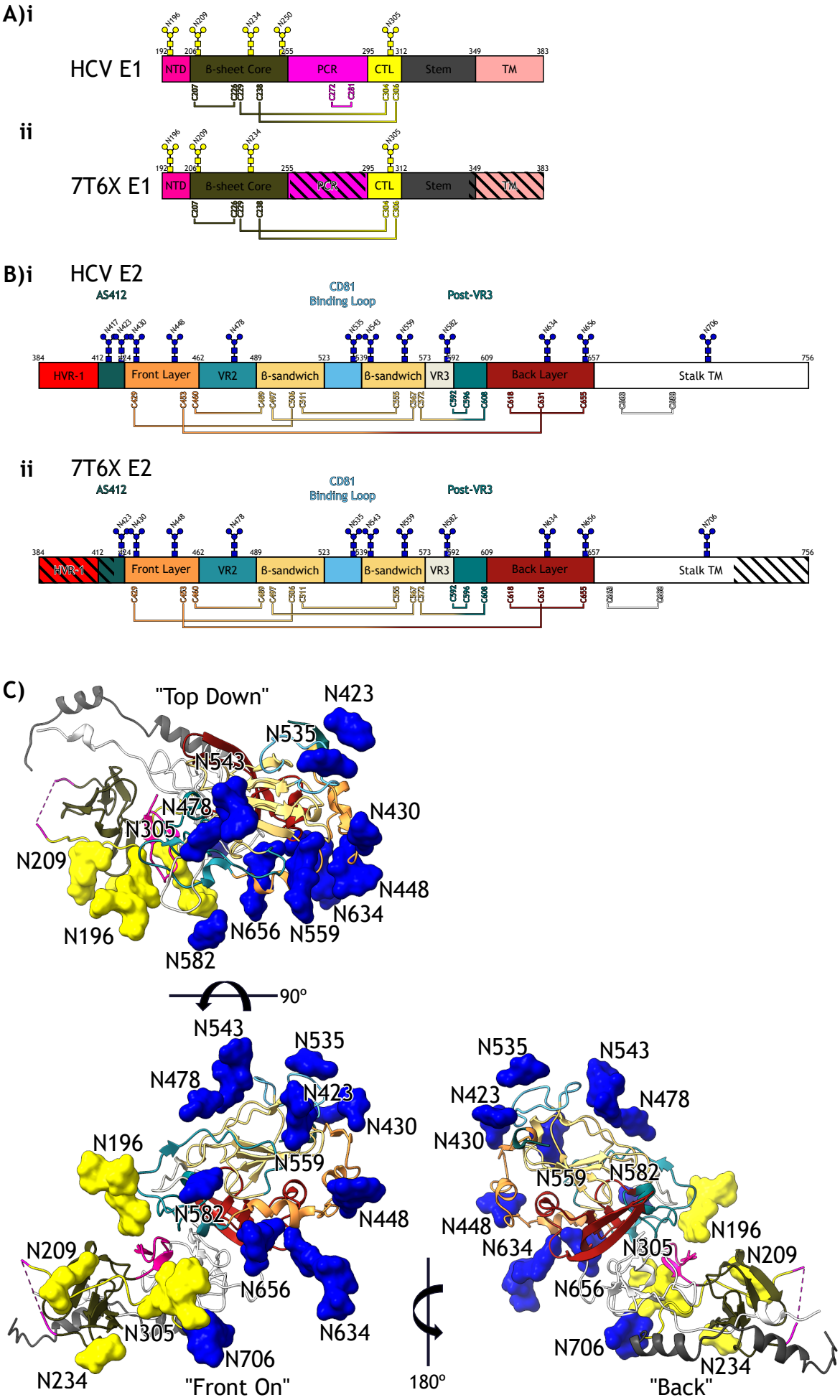
Upon reflection of the E1 and E2 proteins and their potential roles as fusogens, it may be that fusion requires both E1E2 to be working in concert and that HCV does commit viral fusion in an entirely unique way. However, as stated previously, the one statement that can made with certainty is that more research into this area is required.

1. 6. 2. The E1E2 Complex

In 2022, the first structure of E1E2 was released (PDB: 7T6X, Torrents de la Peña *et al.*, 2022, Fig. 1. 11.). Their breakthrough was to express the E1E2 glycoprotein complex from a genotype 1a strain (AMS0232) in conjunction with AR4A to stabilise the E1E2 complex. The structure of E1E2 was resolved in complex with the AR4A antibody and the bNAbs IGH505, which has been discussed, and AT1209, which was found to bind the front layer of E2. This structure highlighted several important discoveries. Firstly, the disulphide bonds present in E1 and E2 were not seen to be involved in heterodimerisation of the E1E2 complex as previously posited (Wahid *et al.*, 2013). Instead, the E1E2 complex is stabilised by both the glycan shield and hydrophobic patches present on both E1 and E2. Specifically, the structure highlighted the role of glycans N196 and N305 on the E1 protein which formed a salt bridge with Q467 in the E2 protein (N196), π-π stacking with E2 W469 (N196), and another salt bridge with E655 (N305) also in E2. Moreover, the glycan-containing face was in opposition to the poorly neutralised hydrophobic back layer of the complex. This hydrophobic back layer has been posited to aid in membrane association or dimerisation (Falson *et al.*, 2015).

This role in dimerisation was then confirmed when the next E1E2 structure was released in 2024 which showed a dimer of heterodimers complex with the hydrophobic patch on the back layer forming part of the E1E2:E1E2 homodimer interface (Augestad *et al.*, 2024). This cryo-EM structure was also able to resolve a region of the E2 protein which had not been resolved in any previous structures: the HVR-1. They found that the HVR-1 was pocketing into a cavity between the base and the stem region of the E2 protein and moreover that this interaction was stabilised by residues at the N-terminus of the HVR-1. Additionally, this structure resolved the conformation of AS412 and they found that it formed a ß-hairpin-like structure which had been recapitulated in other structures of E2 (PDB: 4DGY, 4GAG, Kong, Giang, Robbins, *et al.*, 2012; Potter *et al.*, 2012). Moreover, they found that the pocketing of the HVR-1 and the ß-hairpin AS412 conformations could be abrogated at 37 °C and promoted at 4 °C. Moreover, this structure shows the CD81 binding loop in the bent orientation and

highlights the flexibility of the CD81-interacting region of E2 which may mitigate E2:CD81 interactions. Altogether, this structure demonstrates for the first time to oligomerisation of E1E2 dimers into homodimers, and importantly for this work, the pocketing of the HVR-1 and how this may have downstream impacts on the ability of CD81 to bind to E2.


A)i
HCV E1
NTD
β-sheet Core
PCR
CTL
Stem
TM
ii
7T6X E1
B)i
HCV E2
AS412
CD81 Binding Loop
Post-VR3
HVR-1
Front Layer
VR2
β-sandwich
VR3
Back Layer
Stalk TM
ii
7T6X E2
C)
"Top Down"
N423
N535
N543
N478
N430
N305
N209
N448
N634
N196
N656
N559
N582
90°
"Front On"
180°
"Back"
N234
N706

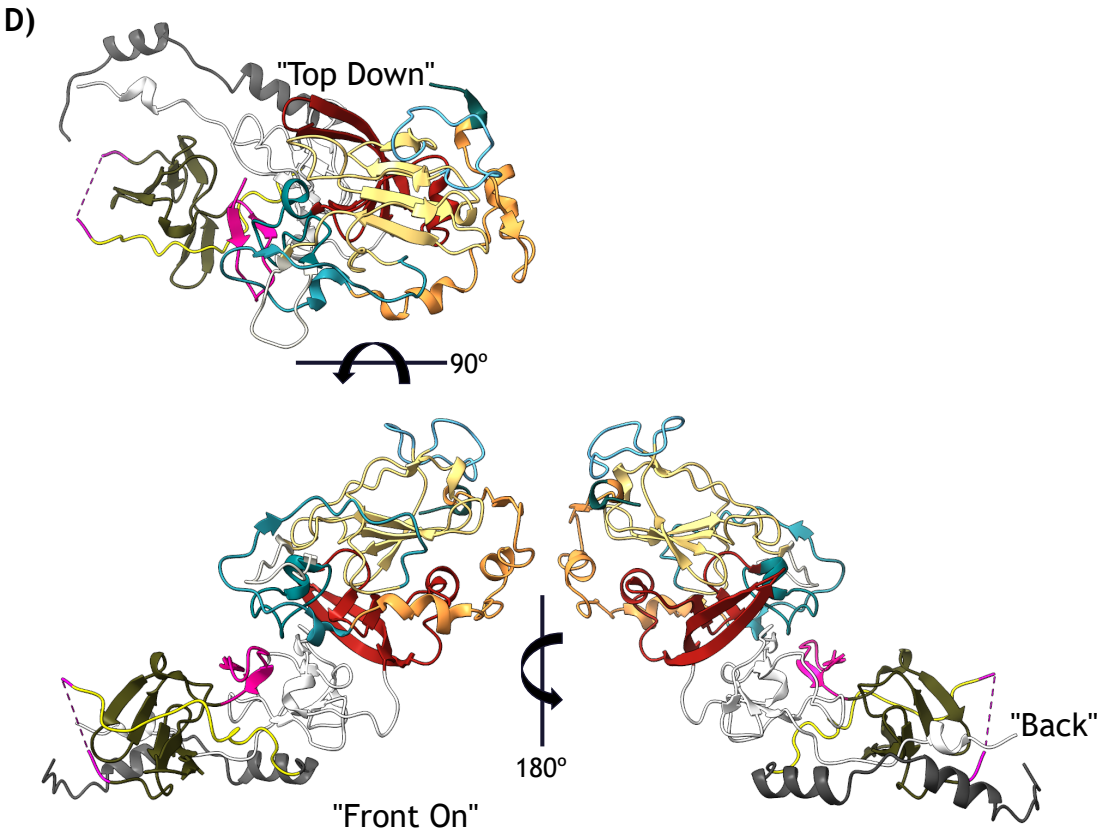
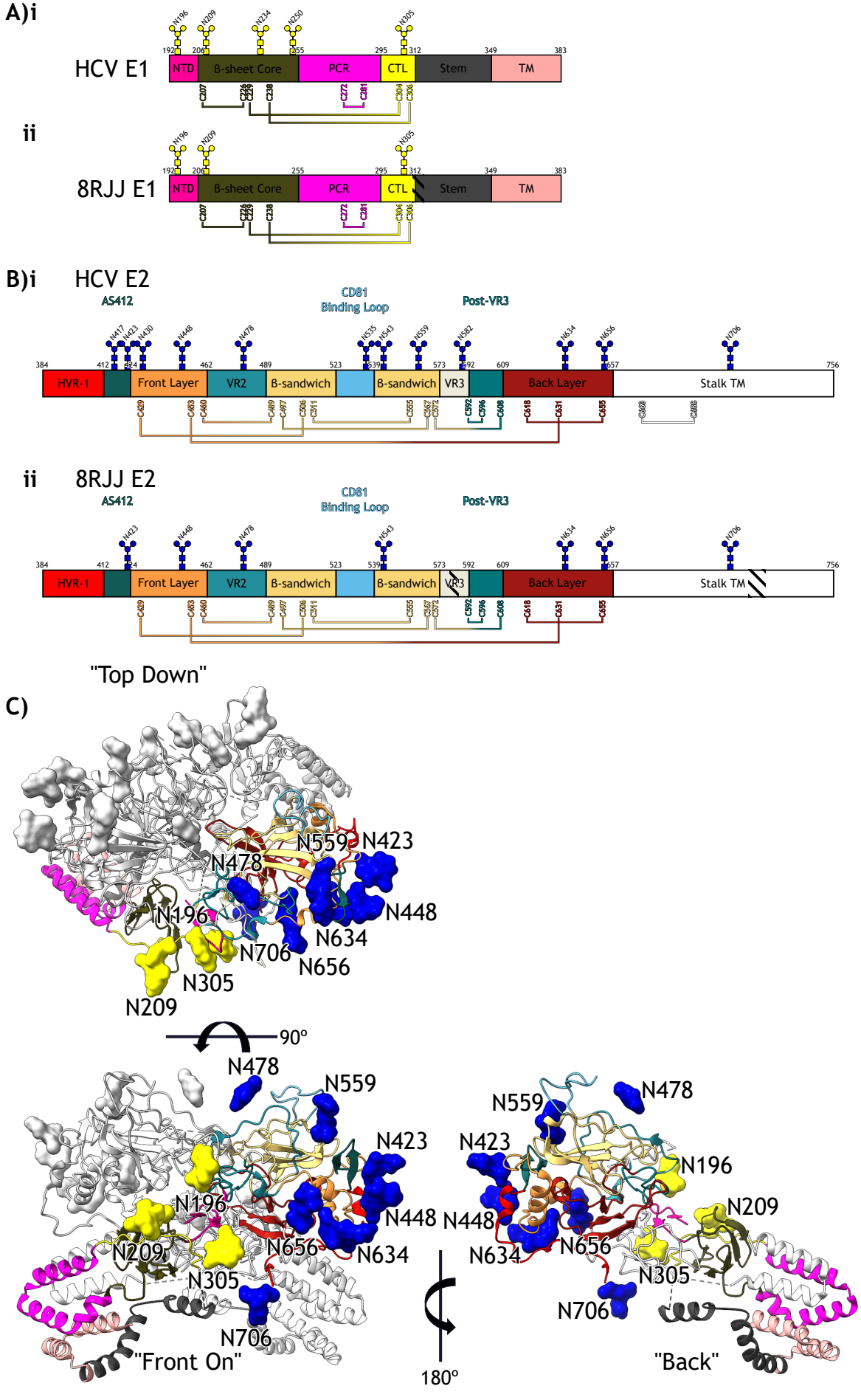


Figure 1. 12. **Cryo-EM Structure of Genotype 1a strain AMS0232 E1E2 (PDB: 7T6X).** A) i HCV E1 protein architecture; glycans are denoted above, and disulphide bonds below. NTD = N-Terminal Domain, PCR = Putative fusion peptide-Containing Region, CTL = C-Terminal Loop, TM = Transmembrane domain. A) ii schematic of the 7T6X E1 protein which was resolved with the glycans and disulphide bonds resolved in the structure denoted. The dashed lines represent parts of the input sequence which were not resolved in the structure. B) i and ii are the same as A) i and ii but for E2. HVR-1 = Hypervariable Region 1, AS412 = Antigenic Site 412, VR2 = Variable Region 2, VR3 = Variable Region 3. Protein structures are coloured as per these schematics. C) Cryo-EM structure of 7T6X with glycans shown. As with both 4MWF and 4WEB structures, the 7T6X structure was structurally aligned with 8RJJ such that the "front on" structure is perpendicular to the plane of the viral membrane, with the resolved glycans facing towards the reader. The "back" structure is a 180° rotation of the "front on" structure on the y-axis. The "top down" structure is a 90° rotation toward the reader of the "front on" structure on the x-axis. D) Same structures as in C) with the glycans not shown.


A)i
HCV E1
NTD
β-sheet Core
PCR
CTL
Stem
TM
ii
8RJJ E1
B)i
HCV E2
AS412
CD81
Binding Loop
Post-VR3
HVR-1
Front Layer
VR2
β-sandwich
VR3
Back Layer
Stalk TM
ii
8RJJ E2
"Top Down"
C)
N559
N423
N478
N448
N196
N634
N706
N656
N305
N209
90°
"Front On"
180°
"Back"

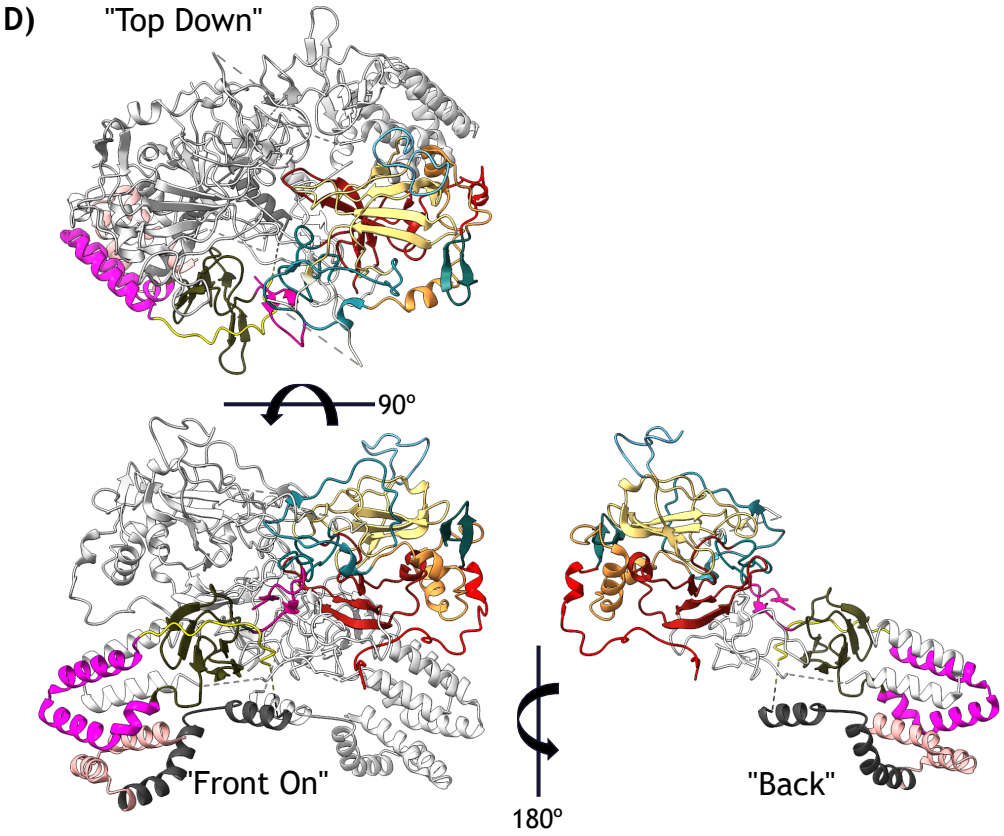


Figure 1. 13. **Dimer of Dimers Cryo-EM Structure of Genotype 3a HCV S52 E1E2 (PDB: 8RJJ).** A) i HCV E1 protein architecture; glycans are denoted above, and disulphide bonds below. NTD = N-Terminal Domain, PCR = Putative fusion peptide-Containing Region, CTL = C-Terminal Loop, TM = Transmembrane domain. A) ii schematic of the 8RJJ E1 protein which was resolved with the glycans and disulphide bonds resolved in the structure denoted. The dashed lines represent parts of the input sequence which were not resolved in the structure. B) i and ii are the same as A) i and ii but for E2. HVR-1 = Hypervariable Region 1, AS412 = Antigenic Site 412, VR2 = Variable Region 2, VR3 = Variable Region 3. Protein structures are coloured as per these schematics. C) Cryo-EM structure of 8RJJ with glycans shown. 8RJJ shows two E1E2 dimers with near-perpendicular ectodomains. The E1E2 dimer facing towards the reader is coloured as per the schematics above, the E1E2 dimer facing away is coloured in light grey. All other structures containing E2 have been structurally aligned to the forward facing E2 in this structure. The "back" structure is a 180° rotation of the "front on" structure on the y-axis with the light grey second E1E2 dimer removed to see the back of the E1E2 dimer. The "top down" structure is a 90° rotation toward the reader of the "front on" structure on the x-axis. D) Same structures as in C) with the glycans not shown.

1. 7. The HVR-1

The HVR-1, located at the very N-terminus of the E2 protein, is a dynamic and disordered tail of 27 amino acids in length. The variability of this region was first noted in the early 1990s (Hijikata *et al.*, 1991; Ogata *et al.*, 1991; Weiner *et al.*, 1991). Further studies in patients demonstrated that the variation of this region was being driven by antibodies, and mediating antigenic drift in the HVR-1 (da Silva Cardoso *et al.*, 1995; Scarselli *et al.*, 1995; Dowd *et al.*, 2009). It has been suggested that early generation of anti-HVR-1 antibodies can prevent HCV infections from progressing to chronic infections (Farci *et al.*, 2000). Moreover, it has been demonstrated that patient-derived antibodies can be cross-reactive for different HVR-1 sequences (Scarselli *et al.*, 1995; Jackson *et al.*, 1997). Together, these data suggest that whilst the HVR-1 will readily mutate, there are some conserved residues which are required for function.

Deletion of the HVR-1 in both HCVcc and HCVpp studies has been associated with an increase in antibody sensitivity (Bartosch *et al.*, 2005; Prentoe *et al.*, 2011). Indeed, work by Forns *et al.* showed that HVR-1 deletion resulted in reduced viral titres whilst still being infectious in chimpanzees, with compensatory mutations in E2 developing over time which increased titres, in particular V514M (Forns, Thimme, *et al.*, 2000). Thus, demonstrating that the HVR-1, whilst not essential for HCV, is important for its normal functioning/maintaining its infectivity. Moreover, in 2002, Scarselli *et al.* demonstrated that deletion of the HVR-1 from E2 resulted in a loss of binding to SR-B1 in HepG2 cells (cells which do not express CD81) which could be restored with the introduction of V514M compensatory mutation, as seen in the chimpanzees. However, when they repeated these binding assays in CHO cells expressing human SR-B1, they found that deletion of the HVR-1 led to a loss of binding which could not be restored by introducing those compensatory mutations. This suggests that the interaction between E2 and SR-B1 is being mediated by the HVR-1, though the exact mechanism of interaction is yet to be elucidated and may not be a direct interaction (Scarselli *et al.*, 2002).

A study of a comparison of 1,382 sequences found that the HVR-1 could be defined as an overall basic stretch of amino acids, with specific residue positions being occupied with basic residues (Penin *et al.*, 2001). They concluded that the conservation of positively charged residues across the HVR-1 may aid with its function of interacting with either lipids, other proteins, or the GAGs on HPSG. However, they also noted the "strong variability" of the HVR-1 and its tolerance for amino acids substitutions (Penin *et al.*, 2001). Other studies have considered the effects of length polymorphisms on the functions of the HVR-1. One study looked at patient-derived sequences from found between a 6.2 % and 9.1 % prevalence of length polymorphisms within the HVR-1 (Bianchettin *et al.*, 2007). Recent studies have found a 3 % rate of incidence for length polymorphisms within the HVR-1 (Olesen *et al.*, 2025).

The extent of polymorphisms – both compositional and length – within the HVR-1 reflect its ability to mutate readily and rapidly which can facilitate immune escape. Previous studies described the HVR-1 as "shielding" certain epitopes on the E1E2 complex (Prentoe *et al.*, 2016), whilst others describe the HVR-1 as an epitope decoy due to the high level of mutations in the HVR-1 correlating with viral persistence (Ray *et al.*, 1999). Moreover, further studies identified that the HVR-1 was under positive selection (Cuevas *et al.*, 2009; Schmidt-Martin *et al.*, 2015). Sequencing of the HVR-1 from immune-complexed and non-immune-complexed HCV virions demonstrated that predominant quasispecies clones contained similar HVR-1 sequences which were not immune-complexed. Conversely, minor clones contained much higher diversity of HVR-1 sequences which were found to be immune-complexed (Korenaga *et al.*, 2001). Altogether, these findings led to the hypothesis that the HVR-1 operates as an immunodominant epitope decoy which either directly sterically protected more conserved, functional epitopes, or "distracted" the humoral response, allowing HCV to persist.

However, further work has shown that the HVR-1 actually influences the conformation of the whole E1E2 complex. To investigate this, Augestad *et al.* generated HVR-1 swap chimeras and performed neutralisation assays, neutralisation assays at different temperatures, investigated SR-B1 dependency,

and performed MD simulations of their E2 HVR-1 chimeras. Through their work, they showed that broad differences in antibody sensitivity between HCV isolates arise from shifts in global E1E2 complex conformational dynamics which can be modulated by the HVR-1. They demonstrated that sequence variation in HVR-1 (residues 400–404) and in the front layer of E2 (residues 414, 431, and 453) regulated the adoption of a neutralisation-resistant "closed" and neutralisation-sensitive "open" E1E2 state. They found that the HVR-1 stabilises "closed" states, while front-layer polymorphisms fine-tuned epitope accessibility. "Closed" states were found to be associated with increased dependence on the entry factor SR-BI and reduced CD81 engagement prior to SR-B1 interaction. Thus, demonstrating the balance between immune evasion and entry efficiency. Augestad *et al.* further established a link between E1E2 conformation and the structure of AS412. They showed that "closed" states preferentially stabilised AS412 in a ß-hairpin conformation. Together, these findings show the importance of overall global E1E2 structure to antibody sensitivity, receptor dependency, and infectivity (Augestad *et al.*, 2020).

Prior work from the Grove lab also provided insight on the ability of HVR-1 to alter global E1E2 dynamics. Firstly, it was demonstrated through MD simulations that the HVR-1 is a highly dynamic, and disordered tail (Stejskal, Lees, *et al.*, 2020). Following this work, the Grove lab then examined the role of the HVR-1 as a safety catch that prevents premature misfiring of the fusion machinery for HCV. In brief, the HVR-1 has an autoinhibitory function on the E1E2 complex which may prevent the conformational changes required for fusion from occurring spontaneously. This, according to the safety catch model, is modulated by the conformational entropy exhibited by the HVR-1. A dynamic HVR-1 that is highly disordered suppresses E1E2 activity. This prevents misfiring events for the E1E2 fusion machinery but also increases the kinetic barrier to be overcome to facilitate downstream conformational changes that permit fusion to occur. On the other hand, a stabilised HVR-1 leads to a hyperreactive phenotype wherein the E1E2 fusion machinery may misfire but also have a reduced kinetic barrier to conformational changes which in turn may reduce receptor dependency. The safety catch is then turned off through interactions with host cell receptors - potentially SR-B1 - to permit fusion. Furthermore, this safety

catch has other implications for the HCV virion. For example, a less disordered/more stable HVR-1 in E2 turns off the safety catch, and leads to an HCV phenotype which is more sensitive to NAbs and less dependent on SR-B1 to facilitate entry (Stejskal, Kalemera, *et al.*, 2020).

Altogether, these last two studies suggest that the dynamics of E1E2 can be modulated by the HVR-1. Modulation of the HVR-1 dynamics through intrinsic stabilisation may, in turn, alter global E1E2 dynamics and fine-tune the "closed"/"open" E1E2 state. Additionally, MD work by Stejskal *et al.* actually looked into the transduction of movement across the E2 protein from the HVR-1 and found that, indeed, movements from the HVR-1 were transduced throughout the rest of the protein. These two studies provide the foundation for considering the dynamics of the HVR-1 as part of a larger picture wherein the dynamics of the HVR-1 can alter infectivity, receptor dependency, and antibody sensitivity through modulating the global E1E2 dynamics.

Lastly, when considering any interaction with CD81, it has previously been shown that interaction between E2 and CD81 is increased with the HVR-1 is deleted (Roccasecca *et al.*, 2003). This intuitively makes sense when considering the structure of the 8RJJ structure as this model showed a pocketed HVR-1 which altered E2 global structure to disfavour CD81 interaction and, therefore, deleting it could prevent the adoption of this CD81-disfavoured conformation (Augestad *et al.*, 2024).

1. 8. Aims and Objectives

Despite decades of intensive study, structural understanding of E2 remains incomplete, particularly with respect to the HVR-1. In all the experimentally determined E2 structures to date, only a single structure includes the HVR-1. Strikingly, this structure shows the HVR-1 in an unexpected conformation – pocketed into the rest of E2 (PDB: 8RJJ). The lack of E2 structures with a resolved HVR-1 is more than likely due to the difficulty in trying to crystallise a, potentially, highly dynamic protein tail. However, taking advantage of recent advances in machine-learning–based protein structure prediction, this project sought to overcome the limitations of traditional protein structural biology by systematically analysing and folding diverse E1E2 and E2 sequences using multiple state-of-the-art structure prediction platforms.

Additionally, this project aimed to investigate the functional role of length polymorphisms within the HVR-1, something which has been noted for a long time but not yet investigated. To date, there have been one or two studies which sought to provide an assessment of the prevalence and distribution of HVR-1 length polymorphisms. However, in examining the literature, these types of investigations were last conducted in the early 2000s. Many, many more HCV sequences have been collected and uploaded since then. This project, therefore, aimed to conduct a large-scale bioinformatic analysis of all publicly available HCV sequences to examine HVR-1 length diversity across all genotypes and to establish a foundation for subsequent functional investigation.

To directly assess the differences in phenotypes brought on by length polymorphisms in the HVR-1, the project further employed HCVpp with E2 constructs containing HVR-1s of defined and different lengths. This approach was designed to facilitate mechanistic analysis of how HVR-1 length affects viral entry, receptor usage, and antibody sensitivity. Altogether, this approach should provide insight into the earliest steps of HCV infection, and immune escape.

Overall, this project was designed to combine state-of-the-art computational structural prediction, large-scale bioinformatic sequence analysis, and functional virology to better understand the HVR-1 and its role in HCV infections.

Chapter 2

Materials and Methods

2. Methods and Materials

2. 1. Bioinformatic Analysis of the HVR-1 of HCV

All Python scripts generated and used during this project can be found in a repository on GitHub.

Available here: https://github.com/S-Devlin/sdevlin_thesis_scripts

All codes were written in Visual Studio Code (VSC).

Version: 1.104.1 (user setup)

Commit: 0f0d87fa9e96c856c5212fc86db137ac0d783365

Date: 2025-09-17T23:36:24.973Z

Electron: 37.3.1

ElectronBuildId: 12404162

Chromium: 138.0.7204.235

Node.js: 22.18.0 V8: 13.8.258.31-electron.0

OS: Windows_NT x64 10.0.26200

Troubleshooting of Python scripts was aided by GitHub Copilot.

2. 1. 1. HCV-Glue: Retrieving Sequence Information

From HCV-Glue (Singer *et al.*, 2018), all currently deposited sequences which covered ≥ 25 % of the E2 region of the HCV genome were downloaded by genotype which gave seven .fasta files containing DNA sequences. The DNA sequences differed in length; some covered a portion of the E2 glycoprotein DNA sequence, others covered the entire sequence of the HCV genome.

2. 1. 2. Extracting HVR-1 Sequence Information

After downloading the seven .fasta files, the DNA sequence for the HVR-1 region was extracted. This required several steps which, in brief, included separating the DNA sequences by sequence length per genotype, pairwise alignment of the DNA sequences, manual extraction of the HVR-1 region, and verifying all extracted HVR-1 sequences were valid i.e. sequence in codons, without internal stop codons, and containing no non-canonical nucleotides.

Firstly, the sequences in each of the initial seven .fasta files were separated by length. A Python script was written (bin_hcv_glue_seqs_by_length.py) which counted the length of the DNA sequences within the initial .fasta files and separated them into different bins according to length. There were 21 bins into which DNA sequences could be sorted which are detailed in table 2. 1. After binning the sequences based on length, the sequences in the same bin were written to a new .fasta file. Prior to alignment, each newly-generated .fasta file was processed by SeqKit (Shen *et al.*, 2016) to remove any duplicate sequences and to generate a secondary file that records the accession numbers of all removed sequences and the sequence with which they share 100 % identity.

Table 2. 1. Details of the bins used to separate out the initial seven .fasta files from HCV-Glue

Bin	Length of DNA Sequence
Bin 1	1 - 500 DNA Bases
Bin 2	501 – 1000 DNA Bases
Bin 3	1001 – 1500 DNA Bases
Bin 4	1501 – 2000 DNA Bases
Bin 5	2001 – 2500 DNA Bases
Bin 6	2501 – 3000 DNA Bases
Bin 7	3001 – 3500 DNA Bases
Bin 8	3501 – 4000 DNA Bases
Bin 9	4001 – 4500 DNA Bases
Bin 10	4501 – 5000 DNA Bases
Bin 11	5001 – 5500 DNA Bases
Bin 12	5501 – 6000 DNA Bases
Bin 13	6001 – 6500 DNA Bases
Bin 14	6501 – 7000 DNA Bases
Bin 15	7001 – 7500 DNA Bases
Bin 16	7501 – 8000 DNA Bases
Bin 17	8001 – 8500 DNA Bases
Bin 18	8501 – 9000 DNA Bases
Bin 19	9001 – 9500 DNA Bases
Bin 20	9501 – 10000 DNA Bases
Bin 21	>10000 DNA Bases/Previously Unbinned

binned .fasta files to produce up to 21 different alignments per genotype which were opened in AliView (Larsson, 2014), an alignment visualisation software, and manually curated.

Manual curation, firstly, ensured that the sequences included the HVR-1/N-terminus of the E2 protein, removing any sequences that did not contain any

Following the removal of duplicate sequences, the remaining sequences were aligned using **Multiple Alignment using Fast Fourier Transform** (MAFFT) (Katoh, Asimenos and Toh, 2009), a high-speed multiple sequence alignment program. A local pairwise alignment was performed on the sequence information relating to the HVR-1 – this was particularly important for the shorter sequences.

Some of the shorter sequences, indeed, covered 25 % of the E2 coding region from the HCV genome but at the C-terminus and therefore did not include any sequence information pertaining to the HVR-1. After removal of any non-HVR-1 sequences, the following regions of the aligned sequences were copied into new .fasta files: the HVR-1, AS412, full-length E2, and E2 minus the HVR-1 – producing 4 different DNA sequence-containing .fasta files for downstream processing.

After successfully extracting all HVR-1 sequences, the newly generated .fasta files per genotype containing the HVR-1 sequences were concatenated to produce one .fasta file per genotype that contained all the extracted HVR-1 sequences – the same was done for the AS412, full-length E2, and E2 minus the HVR-1 sequences.

This resulted in 4 .fasta files per genotype – one containing all of the extracted HVR-1 sequences, another containing the extracted AS412 sequences, a third containing all of the extracted full-length E2 sequences, and finally a .fasta file containing all of the E2 sequences with the HVR-1 region removed. Additionally, the individual .fasta files per genotype were then concatenated to produce four .fasta files that contained all the HVR-1 sequences across all seven genotypes, all the AS412 sequences, full-length E2 sequences, and all E2 sequences without the HVR-1 sequences.

After all the HVR-1 sequences from all seven genotypes were collated, the .fasta file containing the HVR-1 DNA sequences from all seven genotypes was processed by a Python script (codon_checker.py) to check that the DNA sequences correlated to exact codons i.e. any sequence not divisible by three was checked to ensure that the correct DNA sequence information had been extracted. Additionally, the above Python script also detected any premature stop codons

and ambiguous nucleotides. Any sequence that contained a premature stop codon was removed from the HVR-1 sequence .fasta file. Given the prevalence of ambiguous nucleotides throughout the HVR-1 sequences, a further round of filtering was undertaken wherein using another Python script (n_base_filter.py), any sequences containing more than three “N” nucleotides were removed from the HVR-1 sequence .fasta file. Lastly, the HVR-1 DNA sequences were converted to their amino acid sequences and processed using SeqKit to remove any duplicate amino acid sequences.

These HVR-1 DNA/amino acid sequence .fasta files then became the master files to which all other .fasta files were compared. Using another Python script (fasta_comparison.py), these master .fasta files were parsed and the accession numbers contained within the master .fasta files were used to compared to other .fasta files with the same type of sequence information i.e. The .fasta file containing the full E2 DNA sequences were compared to the HVR-1 DNA master .fasta file which had been parsed. Any accession number found in the full E2 DNA sequences .fasta file which were not present in the HVR-1 master .fasta files were removed along with the corresponding sequence information. Using the HVR-1 sequence master .fasta files, all other .fasta files were compared to them, and their sequence information parsed as detailed above.

After extraction of all valid HVR-1 sequences, another Python script (hvr1_length_counter.py) was used to count the lengths of the HVR-1 sequences and produce a .csv file which contained accession numbers, their respective HVR-1 amino acid sequences, and the length of the amino acid sequences. This could be used to investigate and graph out the prevalence of length polymorphisms in the HVR-1.

2. 1. 3. Preparing E1E2/E2 Sequences for Structural Prediction

After extracting all HVR-1 sequences from the HCV-Glue database, a subset of these sequences and their corresponding E1 and E2 sequences were taken forward for structural prediction with ColabFold, AlphaFold3, Boltz-1, and ChAI-1.

After parsing and validation, 9,103 HVR-1 sequences were extracted along with their corresponding E2 sequences. To reduce the number of sequences to fold whilst aiming to capture as much sequence diversity as possible, a series of parsing steps were undertaken. Firstly, the 9,103 E2 sequences were sorted from shortest to longest using SeqKit. From this sorted .fasta, all sequences which included the full-length E2 were extracted and copied to a new .fasta file containing 3,932 sequences. These were then filtered using a Python script that removed any sequences containing non-canonical nucleotides (remove_seqs_with_noncanonical_nucleotides.py), resulting in a finalised subset of 3,453 full-length and valid E2 sequences which could be folded. Following this, MMSeq2 (Kallenborn *et al.*, 2025) was used to cluster these full-length and valid E2 sequences with 80 % sequence identity. The sequences were clustered down into 105 distinct clusters numbered from cluster 0 through to cluster 104. This clustering process also produced a .txt file which detailed the accession numbers and the cluster into which each accession was clustered. This produced a set of 105 E2 sequences which were ready to be folded.

The corresponding E1 sequences were then extracted. To do this, all currently deposited sequences which covered $\geq$ 25 % of the E2 region of the HCV genome were downloaded from HCV-Glue into one .fasta file. Using a Python script (fasta_comparison.py), the accession numbers from the .fasta file containing the 105 E2 sequences were extracted and used to filter this new .fasta file containing all of the sequences from HCV-Glue covering $\geq$ 25 % of E2. This gave a new .fasta file containing 105 sequences which varied in length from E2-only sequences to full genome sequences. These were then manually curated to ensure a corresponding full-length and valid E1 sequence for each of the 105 E2 sequences. Where no full-length or valid E1 sequence was available for a particular accession number, the next accession in the cluster was considered if

possible. This was repeated iteratively, however, for some clusters it was not possible to extract a full-length and valid E1 sequence. This produced a set of 96 E1E2 sequences which were ready to be folded.

Additionally, the sequence of HCV J6 E1E2 and the related constructs that were used in the wet lab experiments were added into the folding datasets to give a final set of 134 E2 sequences and 125 E1E2 sequences.

2. 1. 4. Protein Structure Prediction

The E2 sequences and E1E2 sequences were folded using ColabFold/AlphaFold2 (Mirdita *et al.*, 2022), AlphaFold3 (Abramson *et al.*, 2024), Boltz-1 (Wohlwend *et al.*, 2025), and ChAI-1 (Chai Discovery *et al.*, 2024).

For ColabFold, each sequence/complex of sequences was folded with the same set of parameters: no templating; three recycling steps; five models produced per input; MMSeqs2 used to generate the MSA. For AlphaFold3, the parameters were as follows: no templating; no recycling steps; five models produced per input; JackHMMR used to generate the MSA. For Boltz-1: no templating; three recycling steps; five models produced per input; MMSeqs2 used to generate the MSA. For ChAI-1: no templating; three trunk recycles; five models produced per input; MMSeqs2 used to generate the MSA. Where possible, parameters were maintained across the four different protein folding platforms to better facilitate comparison between the outputs.

2. 1. 5. Comparing the Protein Structure Prediction Outputs

After folding, all models were then viewed using Chimera X v1.9 (Pettersen *et al.*, 2021). The aim was to compare the predicted position of the HVR-1 in relation to the rest of the E2 protein and how this varied by prediction based on platform, sequence, and whether in complex with E1 or not.

To enable comparison, several methods were considered. Firstly, the Euclidean straight-line distance between the very N-terminal residue of the HVR-1 and a conserved valine residue (V708, as per the protein alignment and renumbering in appendix I) was calculated using a Python script (nterm_to_v319_and_plddt.py) which extracted the 3D co-ordinates of the alpha carbon (Cα) of the first residue and the Cα of the conserved valine, and then calculated the straight line distance between those two points using the following calculation:

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2}$$

Which gave the Euclidean straight-line distance between the two Cαs in Ångströms (Å) which was then plotted. However, this merely gave information about the distance of the HVR-1 to a conserved valine residue within the stem region of the E2 protein but did not provide specific information about where the HVR-1 was positioned in relation to the E2 protein.

Therefore, to visualise the predicted spatial relationships between the folded E2 proteins/E1E2 complexes and their respective HVR-1 proteins, all of the predicted models (both E2-only and E1E2 complexes) from all four of the folding platforms were opened in ChimeraX and the N-terminal residues of the many HVR-1s were visualised in comparison to the HCV J6 E2/E1E2 complex as predicted by ColabFold.

Using the HCV J6 WT E2 ColabFold-predicted protein model as a reference, all other models were structurally aligned to this protein model. Following alignment, the individual models were then combined into one .pdb file. Combining the models in ChimeraX generates a log which documents the original .pdb/.cif file name and the new chain ID in the combined .pdb file – this was

referred to as the remapping log. To distinguish between which chain correlated to an E1 protein or an E2 protein after combining the E2-only and E1E2 complex .pdb/.cif files into one combined .pdb file, the length of the chain was used to differentiate. The longest E1 protein in the folding dataset was found to be 194 amino acids in length and therefore, the cut off to distinguish between E1 and E2 chains was set at ≥ 250 amino acids – those less than 250 amino acids were E1 proteins and therefore excluded from coordinate extraction.

Using the remapping log and the combined .pdb file, another Python script (extract_nterm_hvr1_ca_3d_coords.py) was used to parse out the Cα coordinate for the first residue in each chain ≥ 250 amino acids in length in the combined .pdb file, and with the remapping log, re-assign the original file name from before the combination in ChimeraX. This produced an Excel spreadsheet which detailed the original file name, the new chain ID and the first E2 residue Cα coordinates. Moreover, the original file names allowed for the Python script to further parse out which folding software was used to produce the original .pdb/.cif file and further, whether it was E2-only or an E1E2 complex that was folded.

At this point in the analysis, there was an Excel spreadsheet that detailed the original file names, the new chain IDs in the combined .pdb file, the coordinates of the N-terminal residue Cα of the E2s from the combined .pdb file, the folding software that produced each model, and whether it was folded as an E2 model or an E1E2 complex. The last step was to parse out eight separate .pdb files from the combined .pdb file that only contained the ATOM lines relating to the N-terminal residue Cα of the E2s. Using the Excel spreadsheet as reference, another Python script was written (parse_nterm_hvr1_ca_3d_coords_by_folding_software.py) which would go through the Excel spreadsheet and compare the chain IDs and coordinates to ATOM lines in the combined .pdb and would copy them to new .pdb files split by folding software and by type of fold (E2-alone vs E1E2). This resulted in a combined .pdb file which contained the alignment of all the protein folds by all folding software used, and eight separate .pdb files which contained the first N-terminus residue Cα coordinates split by folding software and by E2-only folds and E1E2 complex folds.

These nine .pdb files could then be opened in ChimeraX and used to produce visualisations of the predicted positioning of the N-terminus of the HVR-1 in relation to the HCV J6 WT E2 protein/E1E2 complex as predicted by ColabFold.

2. 1. 6. Phylogenetics

After extracting all HVR-1 sequences, phylogenetic analysis allowed for the investigation of universality of length polymorphisms throughout each genotype and across all genotypes. To permit this, IQ-Tree 2 (Minh *et al.*, 2020) was used to tree the amino acid E2 sequences without the HVR-1 sequence information to prevent the HVR-1 sequence from biasing the tree itself. IQ-Tree 2 has a test mode wherein the data is sampled to select the best substitution model for the phylogenetic analysis, this was performed for each data set to ensure optimal conditions, given the data characteristics and computational resources available. Given that genotype 7 only had 2 unique and valid HVR-1 sequences, an individual tree could not be produced for genotype 7. A full list of substitution models used can be found in table 2. 2.

Table 13. 2. Details of substitution models used for each of the trees.

Genotype	Substitution Model
GT1	JTT+G4
GT2	JTT+I+G4
GT3	JTT+I+G4
GT4	JTT+I+G4
GT5	FLU+I+G4
GT6	JTT+I+G4
GT7	NA
All Genotypes	JTT+G4

Following the production of the 7 trees – GT1 – 6, all GTs – they were imported into the **interactive Tree of Life** (iTOL) online software (Letunic and Bork, 2024) and using the .csv files generated after HVR-1 sequence extraction, the metadata regarding the genotype and/or the length of the HVR-1 could be mapped on to the phylogenetic tree to examine the universality of length polymorphisms throughout the genotypes of HCV.

2. 1. 7. Sequence Logos

Having extracted the HVR-1 sequence information from all 9,103 valid and unique HVR-1 sequences, sequence logos were made in RStudio version 2025.5.1.513 (Posit team, 2025) using the following RStudio packages: ggseqlogo (Wagih, 2017), ggplot2 (Wickham, 2016), and Biostrings (Pagès *et al.*, 2024). All HVR-1 sequences were aligned using MAFFT (Katoh, Asimenos and Toh, 2009) and then two sequence logos were produced – one measuring "bits", and the other "probability". This was then repeated for each of the seven genotypes individually.

2. 2. *In Vitro* Analysis of Length Polymorphisms in HVR-1

2. 2. 1. Cell Lines

Table 2. 14. Details of Cell Lines Used for In Vitro Experiments.

Cell Line	Media and Conditions
HEK 293T CD81 KO	All cells were maintained in Dulbecco's Modified Eagle Medium containing L-glutamine 10 % (v/v) foetal bovine serum, 1 % (v/v) penicillin/streptomycin, and 1 % (v/v) non-essential amino acids.
Huh 7 WT	
Huh 7 CD81 KO*	
Huh 7 CLDN KO*	
Huh 7 LDLR KO*	
Huh 7 OCLN KO*	
Huh 7 SR-B1 KO*	
Huh 7 SR-B1/LDLR DKO*	All cells were kept at 37 °C and 5 % CO_2.
CHO WT	

*** Huh 7 KO cell lines were graciously provided by Dr Yoshiharu Matsuura, Department of Molecular Virology, Research Institute for Microbial Diseases, Osaka University, Osaka, Japan**

2. 2. 2. Pseudoparticle Production

Pseudoparticles containing either HCV J6 E1E2 or mutant glycoproteins were produced via co-transfection. 24 hours prior to co-transfection, $1x10^6$ HEK 293T CD81 KO cells were seeded in 6-well plates (Corning). After 24 hours, the now approximately $2x10^6$ HEK 293T CD81 KO cells were co-transfected using FuGENE6 transfection reagent (Promega) with 1,300 ng of HIV GAGPOL (pCMVΔR8.91), 1,300 ng of the luciferase reporter plasmid (CSLW), and 100 ng of the desired E1E2 expression plasmid. Pseudoparticles were collected from the supernatant at 48 and 72 hours post-transfection and pooled. Transfected cells were lysed for downstream experimental work with 250 µL of RIPA buffer (Thermo Scientific) with 1 % (v/v) protease inhibitor cocktail (Cell Signalling Technology).

2. 2. 3. Generating Length Polymorphism-containing Mutant E2 Constructs

Length polymorphism-containing mutant E2 constructs were designed, informed by previous work undertaken by the Grove lab. These constructs broke down into three groups: N-Terminus Deletions, N-Terminus Extensions, and Internal Insertions. For all experiments, the same controls were used J6 WT E1E2, J6 I438V A524T E1E2, and J6 ΔHVR-1 E1E2. The length polymorphism mutants were cloned into both the J6 WT E1E2 backbone and the J6 I438V A524T E1E2 backbone (Fig. 2. 1. And 2. 2.).

Synthesised sequences (GeneArt, ThermoFisher) which contained these E2 mutations of interest were ligated into the backbone of an HCV J6 E1E2 expression vector and a J6 I438V A524T E1E2 expression vector using the NEB Quick Ligation Kit. This backbone was generated through double digestion of both the HCV J6 E1E2 expression vector and the J6 I438V A524T E1E2 expression vector by BsiWI and BsrDI restriction enzymes (NEB).

Generated plasmids containing our mutations of interest were used to transform NEB DH5α *E. coli* cells which were left to grow at 37 °C overnight on LB agar selection plates supplemented with 100 µg/mL ampicillin. Single colonies were taken from the selection plates and grown overnight in 5 mL Terrific Broth supplemented with 100 µg/mL ampicillin at 37 °C with shaking. Lastly, plasmid DNA was purified using NEB Monarch® plasmid miniprep kit and then sequenced to ensure correct insertion of the mutation of interest.

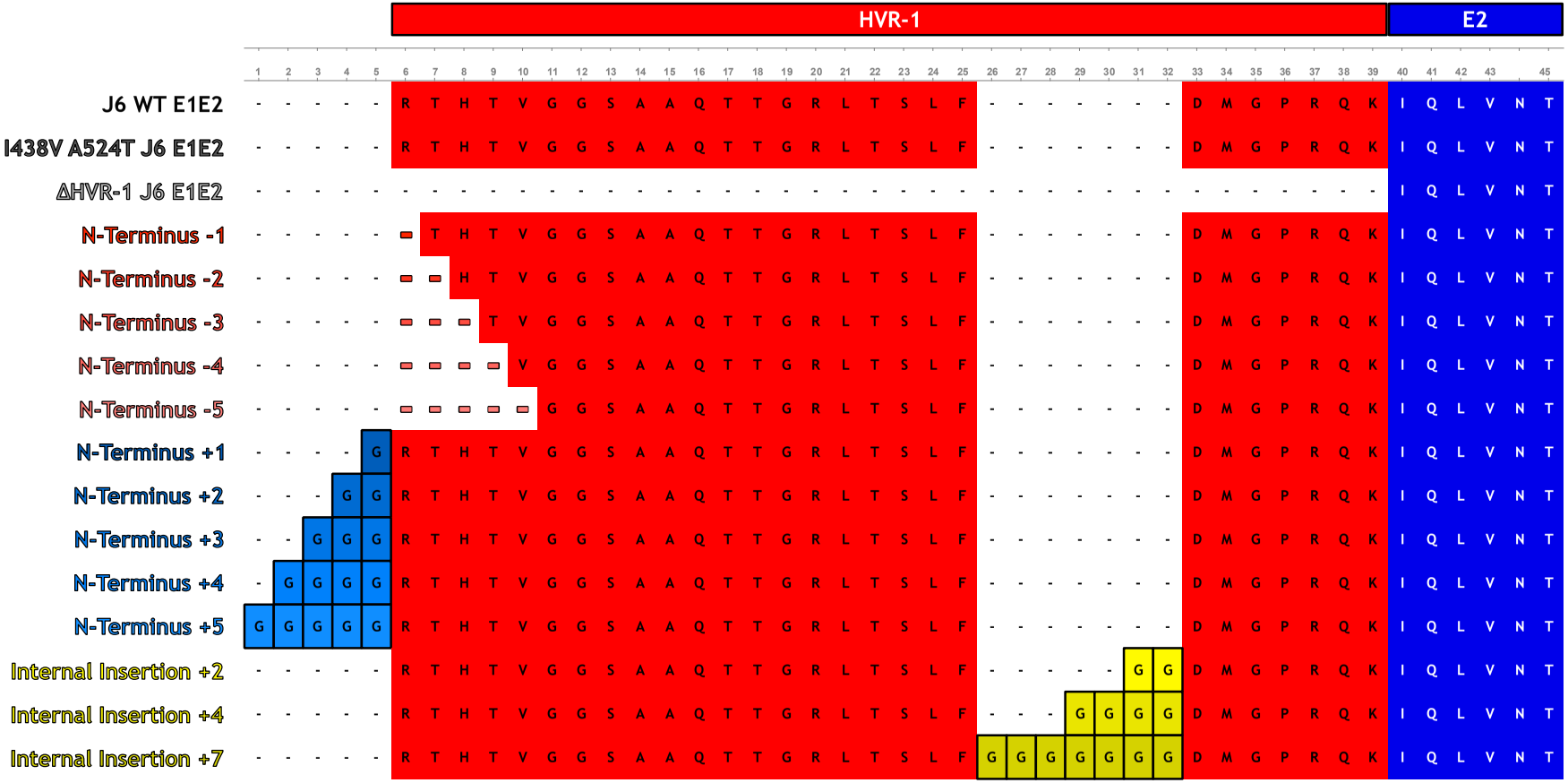


Figure 2. 1. **Schematic Representation of the J6-backbone Length Polymorphism Constructs and Controls.** Sequences were aligned using MAFFT. Sequences were based off of previous work by the Grove Lab, and then gene fragments were ordered from GeneArt (ThermoFisher) and ligated into the J6-backbone using NEB QuickLigase kit. Full details as to how these constructs were generated are found within the text.

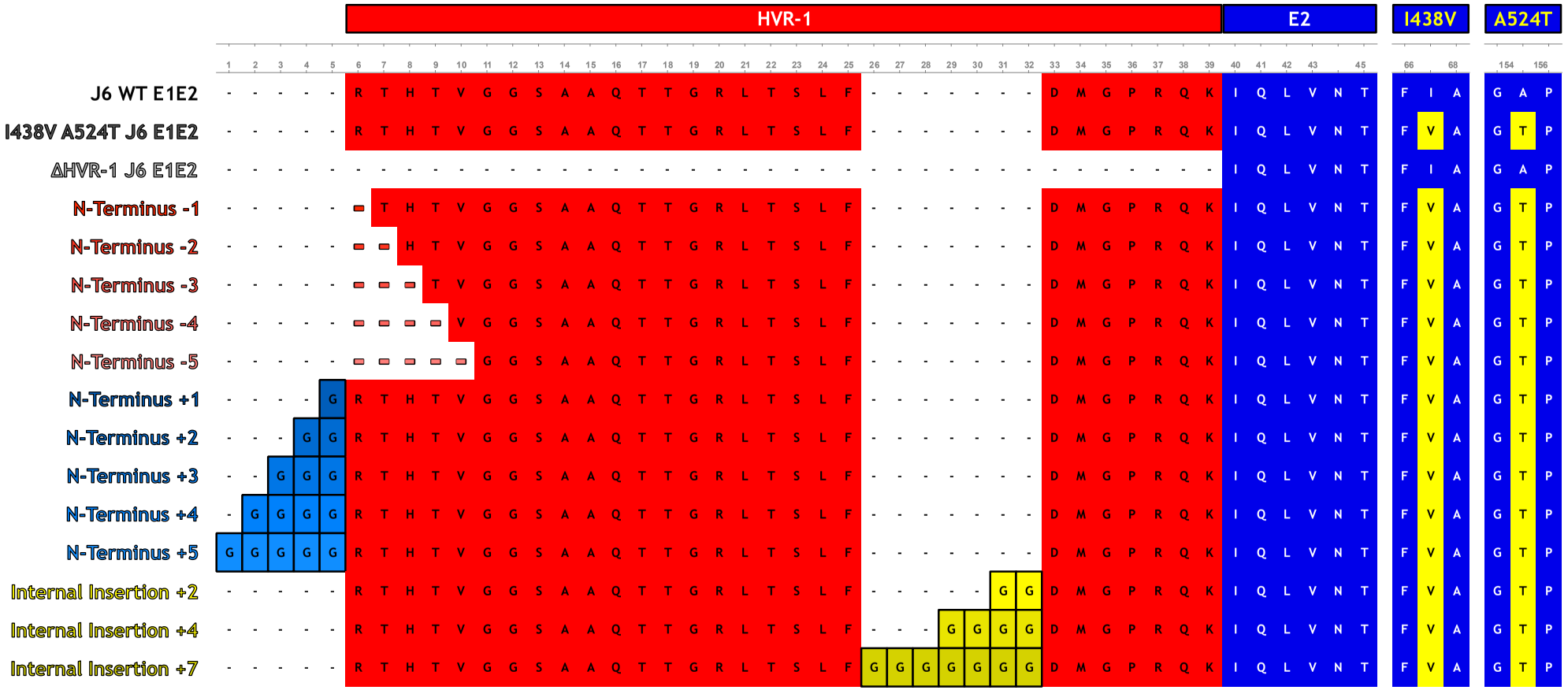


Figure 2. 2. **Schematic Representation of the I438V A524T J6-backbone Length Polymorphism Constructs and Controls.** Sequences were aligned using MAFFT. Sequences were based off of previous work by the Grove Lab, and then gene fragments were ordered from GeneArt (ThermoFisher) and ligated into the I438V A524T J6-backbone using NEB QuickLigase kit. Full details as to how these constructs were generated are found within the text.

2. 2. 4. Pseudoparticle Infection Assays

Huh 7 WT cells were seeded at $3x10^4$ cells per well in a 96-well opaque white plate (Greiner) 24 hours prior to infection with pseudoparticle. Immediately prior to addition of the pseudoparticles, the media was aspirated from the wells and 50 µL of fresh media was overlayed onto the seeded cells. Next, the cells were infected with 50 µL of pseudoparticle. The infected cells were then kept at 37 °C and 5 % CO_2 for 72 hours. Cells were then lysed with the BrightGlo assay kit (Promega) 72 hours post-infection and luciferase activity was measured using a GloMax microplate reader (Promega).

2. 2. 5. Pseudoparticle Receptor Dependency Assays

Huh 7 WT and Huh 7 Receptor KO cell lines – CD81 KO, CLDN KO, LDLR KO, OCLN KO, SR-B1 KO, and SR-B1/LDLR DKO – were seeded at a density of $1.5x10^4$ cells per well in a 96-well opaque white plate (Grenier) 24 hours prior to infection. Immediately prior to adding the pseudoparticles, the media was aspirated from the wells and replaced with 50 µL of fresh media. Next, 50 µL of pseudoparticle was added to the wells. The infected cells were incubated at 37 °C and 5 % CO_2 for 72 hours. Using the BrightGlo assay kit (Promega) the infected cells were lysed, and luciferase activity was measured using a GloMax microplate reader (Promega).

2. 2. 6. Pseudoparticle Neutralisation Assays – Polyclonal Antibodies

Huh 7 WT cells were seeded at $1.5x10^4$ cells per well in a 96-well white, flat-bottomed plate (Greiner) 24 hours prior to infection. Pseudoparticles were incubated with patient-derived IgG for 1 hour before being overlayed on the pre-seeded cells. The patient-derived IgG was a pool from patients infected with HCV GT1 and GT3. The patient-derived IgG was serially diluted three-fold. The highest concentration of IgG was 100 µg/mL and was diluted 5 times which yielded the lowest concentration of 0.412 µg/mL. Before infection, the media on the cells was aspirated off and replaced with 50 µL of fresh media, and then 50 µL of the incubated HCVpp/IgG solution was added to the cells. The infected cells were then incubated at 37 °C and 5 % CO_2 for 72 hours. After 72 hours, the cells were lysed with the BrightGlo assay kit (Promega), and the luciferase was measured with a GloMax microplate reader (Promega) for analysis.

Analysis was performed in Excel (Microsoft). Obtained RLUs for the HCVpp/IgG solution-containing wells were divided by the obtained RLU for the well containing the corresponding HCVpp without IgG. This gave a proportion of retained infectivity which was converted to percentage neutralisation with the following equation:

$$Percentage\ neutralisation\ (\%) = 100 - (proportion\, retained\ infectivity\ \times 100)$$

These were then graphed using GraphPad Prism.

2. 2. 7. Pseudoparticle Neutralisation Assays – Monoclonal Antibodies

Huh 7 WT cells were seeded at $1.5x10^4$ cells per well in a 96-well white flat-bottomed plate (Greiner) 24 hours prior to infection. Pseudoparticles were incubated with four different monoclonals for 1 hour before being overlayed on the pre-seeded cells. The monoclonals were J6.36, AR3C, AP33, and AR4A. The monoclonal IgGs were serially diluted three-fold. The highest concentration of IgG was 10 µg/mL and was diluted 8 times which yielded the lowest concentration of 0.0015 µg/mL. Before infection, the media on the cells was aspirated off and replaced with 50 µL of fresh media, and then 50 µL of the incubated HCVpp/IgG solution was added to the cells. The infected cells were then incubated at 37 °C and 5 % CO_2 for 72 hours. After 72 hours, the cells were lysed with the BrightGlo assay kit (Promega), and the luciferase was measured with a GloMax microplate reader (Promega) for analysis.

Analysis was performed in Excel (Microsoft). Obtained RLUs for the HCVpp/IgG solution-containing wells were divided by the obtained RLU for the well containing the corresponding HCVpp without IgG. This gave a proportion of retained infectivity which was converted to percentage neutralisation with the following equation:

$$Percentage\ neutralisation\ (\%) = 100 - (proportion\, retained\ infectivity\ \times 100)$$

These were then graphed using GraphPad Prism.

2. 2. 8. Lentiviral Vector Production

Lentiviral vectors were produced via co-transfection to transduce CHO cells to express CD81 and SR-B1. 24 hours prior to co-transfection, $1x10^6$ HEK 293T CD81 KO cells were seeded in 6-well plates (Corning). The seeded HEK 293T CD81 KO cells were co-transfected using FuGENE6 transfection reagent (Promega) with 1,300 ng of HIV GAGPOL (pCMV-dR8.91), 400 ng of the VSV-G plasmid (pMD2.g VSV-G), and 1,300 ng of a pDual plasmid containing both the desired receptor and GFP as a reporter. Lentiviral vectors were collected from the supernatant at 48 and 72 hours post-transfection, pooled, and frozen at -20 °C until required.

2. 2. 9. sE2 Production, Purification, and Buffer Exchange

Plasmids for the production of **soluble ectodomain E2** (sE2) proteins were purchased from Twist Bioscience. The plasmids used the pTwist CMV plasmid as a backbone into which the ectodomains of the controls used in the pseudoparticle-based assays and a representative subset of the length polymorphism constructs each with a Twin-Strep-tag® (IBA Lifesciences GmbH) for purification were cloned. The length polymorphisms cloned into the pTwist plasmids included the J6 -3 N-Terminus Deletion, J6 +3 N-Terminus Extension, J6 +4 Internal Insertion, J6 I438V A524T -3 N-Terminus Deletion, J6 I438V A524T +3 N-Terminus Extension, and J6 I438V A524T +4 Internal Insertion. These plasmids were then transfected into HEK 293T CD81 KO cells to produce the various sE2 proteins for purification.

HEK 293T CD81 KO cells were seeded in 6-well plates (Corning) at a density of $1x10^6$ cells/well 24 hours prior to transfection. The pre-seeded HEK 293T CD81 KO cells, now at a density of approximately $2x10^6$ cells/well, were transfected using FuGENE 6 transfection reagent (Promega) with 1,000 ng of the desired sE2 expression plasmid in a ratio of 3:1 FuGENE (µl) to DNA (µg).

sE2 proteins were collected from the supernatant 72 hours post-transfection and pooled for purification. Samples were stored, and all purification steps were done at 4 °C.

Purification was performed using Strep-Tactin® 4Flow® high-capacity resin (50 % suspension) (IBA Lifesciences GmbH) and 10 mL Pierce™ Disposable Columns (Thermo Fisher Scientific Inc). An equilibration buffer (100 mM Tris/HCl pH 8.0, 150 mM NaCl, 1 mM EDTA) and an elution buffer (100 mM Tris/HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin) were made in-house, filtered, and de-gassed prior to use. 2 mL of resin was used per gravity flow column, over which 10 mL of equilibration buffer was passed to equilibrate the resin. Following this, the collected supernatants were passed four times through the columns. After the final pass, the supernatant was collected and retained at 4 °C for Western blot analysis. Following this, 10 mL of equilibration buffer was passed through the column. This eluted equilibration buffer was also retained

for Western blotting. Lastly, the purified protein was eluted in five separate fractions, each produced by running 2 mL of elution buffer over the resin and collecting in separate tubes. Each fraction was retained for Western blotting.

Lastly, the fractions containing sE2 were buffer exchanged into PBS. Using Vivaspin® ultrafiltration spin columns (Sartorius AG). All sE2-containing fractions were loaded into the spin columns and span at 2,000 g until the fraction volume was reduced to ~200 µL. Subsequently, PBS was added to the spin columns in a ratio of 9:1 PBS to sE2-containing fractions ergo 1.8 mL of PBS was added to the spin column and span down again at 2,000 g until the fraction volume was reduced, again, to ~200 µL. This process was repeated two more times. The purified and buffer exchanged sE2 protein samples were then flash frozen in liquid nitrogen, and stored at -80 °C.

2. 2. 10. sE2 Binding Assay

CHO cells were transduced with lentiviral vectors containing pDual plasmids containing CD81 and SR-B1, along with GFP as a reporter. Transduction of the CHO cells was performed by splitting wildtype CHO cells in a T150 flask (Corning®) and then adding 4 mL of lentiviral viral vector-containing supernatant, making the total flask volume up to 20 mL and then allowing the cells to grow to confluency. Successful transduction was confirmed via EVOS. Following transduction, cells were kept in culture for up to two weeks before discarding.

To perform the sE2 binding assay, transduced CHO cells expressing either CD81 or SR-B1 were trypsinised to obtain a single cell suspension and then centrifuged at 1,500 RPM for 5 minutes. The receptor-expressing CHO cells were then resuspended in Traffic Stop buffer (PBS + 1 % BSA + 0.01 % Sodium Azide) to a concentration between 2.5-5x10^6 cells/mL. Traffic Stop buffer was made in the laboratory in a large batch, aliquoted, and frozen at -20 °C. Following resuspension, a 96-well clear round-bottomed microplate (Corning®) had 100 µL of cell suspension added to each well, and the plate was incubated at 37 °C for 20 minutes. During incubation, sE2 stocks were diluted to 10 µg/mL in PBS, and following incubation 100 µL of these diluted sE2 solutions were added to each well, mixed well via pipetting and then incubated for an hour at 37 °C.

The 96-well plate was then centrifuged at 1,500 RPM for 5 minutes. The supernatant was removed, and the cells resuspended in 100 µL PBS, this was then repeated. The cells were then centrifuged once again at 1,500 RPM for 5 minutes and then resuspended in 100 µL of 3 µg/mL StrepMAB-Classic (IBA Lifesciences GmbH) diluted in Traffic Stop buffer and then incubated for 45 minutes at room temperature. After incubation with the primary antibody, the cells were washed twice more; the cells were centrifuged at 1,500 RPM for 5 minutes, supernatant removed, and cells resuspended in 100 µL of PBS. Next the cells were resuspended in 100 µL of Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 594 (Invitrogen) diluted 1 in 1,000 in Traffic Stop buffer. This was then incubated at room temperature for 45 minutes. The cells were then washed two more times following the same steps as before:

centrifugation at 1,500 RPM for 5 minutes; removal of supernatant; resuspension in 100 µL of PBS.

The cells were spun down at 1,500 RPM for 5 minutes after washing and resuspended in 100 µL of PBS. The cells were then inspected on the microscope to ensure a unicellular suspension before fixation. Cells were fixed through the addition of 100 µL of 2 % formaldehyde (Thermo Fisher Scientific Inc.) diluted in PBS – this gave a final concentration of 1 % formaldehyde in each well. The plate was then incubated at room temperature for 10 minutes.

Lastly, the cells were centrifuged again at 1,500 RPM for 5 minutes, the supernatant removed, and the cells resuspended in 200 µL of PBS. The plate was then transferred to the Cytek® Guava® easyCyte™ Flow Cytometer for analysis.

2. 2. 11. Plasmid Cloning for smFRET System Development

SnapGene software (www.snapgene.com) was used to introduce silent cut site mutations into the pD607 J6 WT E1E2 plasmid. For clarity, the plasmid containing the additional cut sites shall be referred to as pD607 J6 E1E2+ throughout this section. Details of the additional cut sites introduced can be found in table 2. 4.

Table 2. 4. Overview of Additional Cut Sites Added to pD607 J6 WT E1E2 Plasmid. A comparison of the pD607 J6 WT E1E2 and the new pD607 J6 WT E1E2 with additional cut site plasmids can be found in appendix II. Note that both the BspDI and ClaI are redundant and therefore added at the same position in the parental plasmid but only BspDI was used for cloning purposes.

Restriction Enzyme	Plasmid Nucleotide Position
KflI	3886
HindIII	4058
XcmI	**4236**
BamHI	**4315**
BstBI	**4440**
BspDI **ClaI**	**4596**

Gene fragments which replaced glycan-linked asparagine residues with amber stop codons (TAG) were ordered from Twist Bioscience (N417, N423, N430, N448, N478, N535, N543, N559, N582, N634, N656). Two additional fragments were purchased which would insert an amber stop at the very N-terminus of the HVR-1, and directly behind F404. These fragments were then attempted to be cloned into the pD607 J6 E1E2+ plasmid.

The gene fragments which replaced N582 and N656 with TAG codons were successfully cloned into the pD607 J6 E1E2+ plasmid.

Both the $N582^{TAG}$ fragment and the pD607 J6 E1E2+ plasmid were digested with the restriction enzymes HindIII (NEB) and BstBI (NEB) and then ligated together

with the NEB Quick Ligation Kit. The same procedure was followed for the $N656^{TAG}$ fragment and the pD607 J6 E1E2+ plasmid but with BstBI (NEB) and PflmI (NEB). Following ligation, NEB DH5α *E. coli* cells were transformed with the TAG-containing J6 E1E2+ plasmids. Following transformation, the NEB DH5α *E. coli* cells were left to grow at 37 °C overnight on LB agar selection plates supplemented with 100 µg/mL ampicillin. Single colonies were taken from the selection plates and grown overnight in 5 mL Terrific Broth supplemented with 100 µg/mL ampicillin at 37 °C with shaking. Lastly, plasmid DNA was purified using NEB Monarch® plasmid miniprep kit and then sequenced to ensure that the N582 and N656 codons were substituted with TAG codons.

2. 2. 12 Pseudoparticle Production for smFRET System Development

Pseudoparticles expressing J6 E1E2^{TAG} glycoproteins were produced via co-transfection of four plasmids. 24 hours prior to co-transfection, 1x10^6 HEK 293T CD81 KO cells were seeded in 6-well plates (Corning). After 24 hours, the now approximately 2x10^6 HEK 293T CD81 KO cells were co-transfected using FuGENE6 transfection reagent (Promega) with 1,300 ng of HIV GAGPOL (pCMVΔR8.91), 1,300 ng of the luciferase reporter plasmid (CSLW), 100 ng of the pD607 J6 E1E2^{TAG} expression plasmids, and 100 ng of the pcDNA3.1(+)_U6 tRNAPyl_CMV NESPylRS(AF) plasmid. The pcDNA3.1(+)_U6 tRNAPyl_CMV NESPylRS(AF) (Addgene plasmid: 182287) encodes both the tRNA synthetase and the tRNA transferase for TCO*A and was graciously provided by Dr James Dunlop.

Trans-Cyclooct-2-en-L-Lysine (TCO*A) was purchased from Sirius Fine Chemicals (SiChem). For long-term storage, a 100 mM stock solution of TCO*A was produced by dissolving the powdered TCO*A in 0.2 M NaOH and 15 % v/v DMSO. This was then stored at −20 °C. This stock solution would be mixed in a ratio of 1:4 with a 1 M HEPES (Sigma-Aldritch) solution prior to supplemented HEK293T CD81 KO producer cell media. Following transfection, 25 µL of the diluted TCO*A in 1 M HEPES was added to the cell media.

Pseudoparticles were collected from the supernatant at 48 and 72 hours post-transfection and pooled. HCVpp^{TAG} were used to infect Huh7 WT cells. Transfected producer cells were lysed for downstream experimental work with 250 µL of RIPA buffer (Thermo Scientific) with 1 % (v/v) protease inhibitor cocktail (Cell Signalling Technology).

2. 2. 13. Statistical Analyses

All statistical analysis was performed in Prism 10.6.0 (GraphPad Software).

Five different analyses were conducted throughout this thesis: ordinary one-way ANOVA tests, ordinary one-way ANOVA tests followed by Dunnett's multiple comparison tests, 2way ANOVA tests followed by Šídák multiple comparison tests, extra sum-of-square F tests, and Welch's T-tests. Full context is provided in the figure legends as to how these tests were performed. However, table 2. 5. provides an overview of which analyses were used for each figure.

Table 2. 5. Summary of Statistical Analyses

Statistical Analysis	Figures in which the analysis was used
Ordinary one-way ANOVA	4. 2. 4. 3. 5. 3. B. 5. 7. B. 5. 7. C.
Ordinary one-way ANOVA followed by Dunnett's multiple comparison tests	4. 1. A. 4. 2. 4. 3. 4. 6. A. 4. 6. B. 4. 7. A. 4. 7. B. 5. 4. B. 5. 7. A.
2way ANOVA followed by Šídák multiple comparisons test	4. 1. B. 4. 4. 4. 6. C. 4. 6. D. 4. 6. E. 4. 7. C. 4. 7. D. 4. 7. E. 4. 8. 5. 4. C. 5. 5. D.
Extra sum-of-squares F tests	4. 1. C. 4. 9. C. 4. 10. C. 4. 11. D. 4. 12. D. 4. 13. D. 4. 14. D.
Welch's T-test	5. 6. C.

2. 2. 14. Western Blots

The transfected HEK 293T CD81 KO cells used to produce the pseudoparticles were lysed and Western blotting was performed on the lysate to measure E1E2 production. Cell lysates were produced through adding 250 µL of RIPA buffer (Thermo Scientific) with 1 % (v/v) protease inhibitor cocktail (Cell Signalling Technology) to the producer HEK cells and placing these at 4 °C for 20 minutes. The resultant lysate was centrifuged at 13,000x RPM in a refrigerated centrifuge at 4 °C for 20 minutes. After centrifugation, 90 µL of the supernatant was transferred into an Eppendorf containing 30 µL of 4x Bolt LDS sample buffer with 5 % (v/v) beta-mercaptoethanol. Samples were thoroughly pipetted to ensure even mixing and then heated to 95 °C for 10 minutes. After reduction, 5 µL of reduced lysate was loaded into a 4–20 % Mini-PROTEAN® TGX™ Precast Protein Gel and the gel was run at 100 V until the dye front reached the edge of the gel cassette. The samples were transferred from the gel onto a nitrocellulose membrane using the Trans-Blot Turbo system (Bio-Rad) according to instructions from the manufacturer.

A solution of 0.1 % (v/v) Tween-20 in PBS (PBS-T) was used to wash the membranes. Additionally, a 5 % (w/v) milk/PBS-T solution was used to block the membranes and as a buffer for both primary and secondary antibodies. Blocking was done at room temperature for 1 hour. Primary antibody incubations were done overnight at 4 °C. Secondary HRP-conjugated antibody incubations were carried over 1 hour at room temperature. The Western blots were developed by overlaying the membranes with SuperSignal™ West Pico PLUS Chemiluminescent Substrate (ThermoFisher). As per the protocol of the manufacturer: the chemiluminescent reagents A and B were mixed in a 1:1 ratio and the membrane was incubated with this mix for five minutes. Imaging was done using the ChemiDoc XRS+ System (BioRad) and ImageLab (BioRad) software.

Membranes were probed with anti-E2 and anti-E1 antibodies. Anti-ß-actin antibodies were used as a loading control. Details of all antibodies used can be found in tables 2. 6. - 8.

Table 2. 6. Summary Information for Primary Antibodies Used for Western Blotting.

Primary Antibody	Antibody Description	Concentration (µg/mL)	Dilution Buffer	Incubation
AP33*	Monoclonal, mouse	0.1	5 % Milk PBS-T	4 °C Overnight
J6.36	Monoclonal, mouse	0.5	5 % Milk PBS-T	4 °C Overnight

***Graciously provided by Professor Arvind Patel, MRC-Centre for Virus Research, University of Glasgow.**

Table 2. 7. Summary Information for Secondary Antibodies Used for Western Blotting.

Secondary Antibody	Antibody Description	Concentration (µg/mL)	Dilution Buffer	Incubation
Anti-mouse Invitrogen, A16078	Polyclonal, goat, HRP-conjugated	0.5	5 % Milk PBS-T	Room temperature, 1 hour

Table 2. 8. Summary Information for Loading Control Antibodies Used for Western Blotting.

Loading Control Antibody	Antibody Description	Concentration (µg/mL)	Dilution Buffer	Incubation
Anti-ß-actin	[unknown] HRP-conjugated	[unknown]	5 % Milk PBS-T	Room temperature, 1 hour

Chapter 3

Interrogating the HVR-1 via Bioinformatics

3. Bioinformatic Analysis of the HCV HVR-1

3. 1. The Protein Folding Problem and Protein Structure Prediction

In the early 1960s, Cyrus Levinthal was considering how proteins folded. He noted that if proteins were to explore every theoretically possible conformation that a single polypeptide chain could adopt in order to reach their native-state conformation, then the length of time required for proteins to fold properly would be akin to the lifetime of the universe. Yet proteins fold correctly in seconds. This became known as Levinthal's Paradox (Levinthal, 1968; Zwanzig, Szabo and Bagchi, 1992).

In 1972, Christian Anfinsen shared the Nobel Prize for Chemistry for his work on ribonuclease proteins and on his proposed Thermodynamic Hypothesis of protein folding. His work led to Anfinsen's Dogma which postulates that the 3D structure of protein is determined solely by its amino acid sequence and, moreover, the amino acids arrange themselves in such a way in 3D that produces the lowest possible Gibbs free energy for the whole system (Anfinsen *et al.*, 1961; Anfinsen, 1973). This can be further extrapolated such that the 3D structure of proteins in physiological conditions should be able to be predicted from the amino acid sequence alone by predicting the sum of all interatomic interactions and picking the structure with the lowest Gibbs free energy. However, with regards to Levinthal's Paradox, simply investigating all possible iterations of potential protein folds would take an exceptionally long time, or unfeasibly large amounts of computing power (Hunter, 2006).

These two interplaying concepts led to the development of the "protein folding problem" which itself is composed of three interrelated yet distinct questions. Firstly, understanding the thermodynamics of protein folding: what are the thermodynamic properties at play which must be balanced for proteins to become folded correctly? Secondly, the mechanisms of folding: what are the mechanisms of protein folding which allow for proteins to overcome Levinthal's

Paradox? Lastly, protein structure prediction: how can the 3D structures of proteins be accurately predicted given the first two questions of the protein folding problem? (Dill *et al.*, 2008). Solving the protein folding problem has been a long-standing research problem, and recent advances in machine learning and **artificial intelligence** (AI) are starting to unravel parts of this problem.

In 1994, the first **Critical Assessment of protein Structure Prediction** (CASP) experiment was organised by the Protein Structure Prediction Centre. These are biennial experiments are comprised of research groups providing unreleased, solved protein structures derived through classical molecular biology experiments, and other bioinformatic-focussed research groups attempting to predict the structures of the same proteins. The researchers then come together and discuss how well certain bioinformatic processes recapitulated the solved protein structures (Moult *et al.*, 1995). Across different CASP experiments, one novel approach which emerged was the analysis of co-evolving amino acid residues. Co-evolution analysis looks across **multiple sequence alignments** (MSA) and compares where and when residues are evolving and which other residues are evolving at the same time. From this, inferences can be made about how these residues interact. If two residues are always co-evolving throughout an MSA, it can be inferred that these two residues are in contact with one another in the 3D protein structure (Lapedes *et al.*, 1999; Marks, Hopf and Sander, 2012). Co-evolutionary analysis and Anfinsen's Dogma allow protein structure predictions to side-step Levinthal's Paradox as contact residues can be inferred without having to consider the folding mechanism through which those residues come together. Rather, inferences on structure can be made based on evolutionary history, amino acid sequence, and lowering Gibbs free energy alone.

The Google DeepMind team entered the CASP13 experiment in 2018, their first entry into the competition. They entered with AlphaFold1 and came first in the *ab initio* or free modelling category of CASP. Moreover, they greatly outperformed the next best protein structure prediction. The big advancement made by AlphaFold1 was to combine co-evolutionary analysis as described above with deep neural networks (AlQuraishi, 2019). Neural networks can be multi-

layered with each layer consisting of various nodes. Each node encodes a function which receives an input, performs some function, and produces an output. These functions can be informed by learned weightings and bias which can modulate the output. Importantly, the output from one layer is the input for the next layer – the more layers, the more complex the inferences can be from a neural network as they can compare more input parameters (Greener *et al.*, 2022) However, the neural network needs to be trained to learn the weightings and bias on each node. AlphaFold1 and subsequent versions of AlphaFold were trained on the **Protein Databank** (PDB). The neural network was shown a limited training dataset from the PDB for which it was given access to the structure, the sequence, and the corresponding MSA of distinct protein domains. Through this training, the neural network effectively learnt the correlation between structure, sequence, and the MSA. This training tunes the weightings and biases in the node functions to allow them to recapitulate 3D protein structure from sequence and MSA alone (Senior *et al.*, 2020). Essentially, AlphaFold predicts protein structures by first producing an MSA based on the input amino acid sequence. It then parses this MSA for co-evolutionary information with its neural network and makes inferences on 3D protein structure from what it has learnt from the PDB. The deep neural networks developed by DeepMind contained hundreds of layers and therefore identified patterns in the protein sequence and co-evolutionary couplings that other algorithms were unable to uncover. These learned inferences were converted into a contact map from which the 3D structure of the protein could be predicted (AlQuraishi, 2019).

Then at the CASP14 experiment, DeepMind returned with AlphaFold2. This updated version of AlphaFold contained a completely different deep learning neural network that was capable of attention-based deep learning, allowing their neural network to achieve self-consistent structure predictions. However, in basic principles, it still worked by examining the input sequence, assembling an MSA, and then mining for co-evolutionary data in the MSA with its neural network. Its neural network was, once again, trained on the PDB. However, DeepMind were able to augment the learning of the neural network through a process known as noisy student self-distillation (Xie *et al.*, 2020). Essentially, this breaks down to showing the neural network the PDB and then asking it to

predict ~350,000 protein structures from UniClust30 protein sequences. Then, highly confident structures from this set of predicted structures were combined with the PDB structures to form the next dataset from which the neural network was to learn. However, data augmentations were applied to this second training dataset such as cropping out parts of the protein structures and using subsets of the MSA. These data augmentations prevented the neural network from simply copying the structures it had predicted before, allowing it to learn more robust patterns in the relationship between amino acid sequences, MSAs, co-evolution, and protein structures (Jumper *et al.*, 2021). AlphaFold2 represented a massive leap forward for protein structure prediction. Where AlphaFold1 focussed on local interactions, considering all co-evolution between residues equally to infer 3D structures, AlphaFold2 was trained to selectively focus on the co-evolution of certain residues whilst also considering the overall global fold of the protein. In real terms, AlphaFold1 considered all co-evolution relationships explored within the MSA and used those to infer structure. AlphaFold2 was able to parse the most salient parts of the MSA, extract more complex patterns from the evolutionary history presented in the MSA, and then consider if the final protein made biological sense (Moult *et al.*, 2020; Jumper *et al.*, 2021).

As noted, the first step in AlphaFold2 predicting the 3D structure of a protein is the assembly of the MSA which can be parsed by the neural network. AlphaFold2 uses jackHMMer (Johnson, Eddy and Portugaly, 2010) and HHblits (Remmert *et al.*, 2012) to build its MSA. Both use hidden Markov models to scour protein sequence databases and collate diverse, evolutionarily related protein structures with great accuracy. However, this is not a computationally trivial task, and when predicting protein structures with AlphaFold2, MSA compilation would often represent the bulk of the runtime for a prediction. Enter ColabFold, a re-packaging of AlphaFold2 that replaced jackHMMer and HHblits with MMSeqs2 (Steinegger and Söding, 2017) which built the MSA 40- to 60-fold faster than previous MSA compilation methods whilst maintaining the same neural network as in AlphaFold2. Additionally, AlphaFold2 was initially not able to be used to effectively predict protein:protein interactions and complex formations until AlphaFold-Multimer was released (preprint: Evans *et al.*, 2022). ColabFold maintained the functionality of AlphaFold-Multimer, meaning ColabFold can

predict the structures of both single proteins and protein complexes. Moreover, ColabFold is an open-source software which allowed for local installation, which made it easier to use. Altogether, ColabFold speeds up MSA assembly, allows for predicting protein complexes, is open-source, and able to be installed locally (Mirdita *et al.*, 2022). Given these practical considerations, ColabFold was used in this project in place of AlphaFold2 for protein structure predictions.

Three years after AlphaFold2/ColabFold, AlphaFold3 was released. AlphaFold3 was entered into the CASP16 experiment in 2024. Across the CASP16 experiment, AlphaFold3 either performed better than AlphaFold2 or on par. However, AlphaFold3 represents an entire re-working of the approach DeepMind took to protein structure prediction (preprint: Elofsson, 2025). As noted, AlphaFold2/ColabFold learnt the relationship between amino acid sequences, MSAs, co-evolution, and protein structures. AlphaFold3 relies less on MSAs to deduce protein structure and instead relies on a diffusion module, predicting protein structures in an analogous way to how AI image generation learnt how to produce different images. To train AlphaFold3, the diffusion module of the neural network was shown known structures from the PDB with noise added to the atomic co-ordinates. Noise was added in different amounts. Low level noise – minor changes and movements to the atoms in the structure – allowed the diffusion module to learn local stereochemistry when denoising these training structures. Denoising high noise structures – structures in which atoms were scattered considerably from their starting position - allowed for the diffusion module to learn about overall protein folding when recapitulating the original structure. Diffusion models can be subject to hallucination, or outputting something which is plausible but incorrect (preprint: Rathkopf, 2025). To prevent hallucinations, AlphaFold3 was trained on AlphaFold2 output structures which would often generate extended loops in unstructured regions of proteins. This trained the AlphaFold3 diffusion model to produce extended loops instead of trying to force these unstructured protein regions into compact domains which may have looked plausible but biologically incorrect (Abramson *et al.*, 2024).

Following the release of AlphaFold3, other generative models of protein structure prediction have been released which work along the same principles: diffusion models trained on the PDB. The two that have been used in this project include Boltz-1 (preprint: Wohlwend *et al.*, 2025) and ChAI-1 (preprint: Chai Discovery *et al.*, 2024). Boltz-1 is very similar to AlphaFold3 but has optimisation of parts of the diffusion model and a more robust self-checking during de-noising with the Kabsch algorithm being used at each inference step (Wohlwend *et al.*, 2025). ChAI-1, again, works very similarly to AlphaFold3 but with a broader scope. ChAI-1 also uses diffusion to predict protein structures. However, it also allows for protein structures to be predicted by evaluation of MSAs and co-evolution analysis, by a large language model with embeddings from ESMFold, or through template modelling. Additionally, ChAI-1 allows for optional restraints to be applied to proteins structures derived from wet lab experiments. For example, if a research team found the interacting residues between an antibody and an epitope, this data can be given to ChAI-1 and considered when folding the antibody and the target together into a complex (Chai Discovery *et al.*, 2024).

Altogether, each folding software used in this project has its own strengths. ColabFold has a stricter focus on co-evolutionary analysis than the diffusion-based software and may therefore highlight interactions between co-evolving residues which may be missed by other software. On the other hand, the diffusion-based software packages can sample multiple conformations and explore conformational space in a way that is less attached to the MSA and therefore may be more exploratory in their outputs. Throughout this project, all four were used to provide a robust overview of the potential structures of the HCV E2 protein and the E1E2 complex. The use of all four allowed comparison between their outputs, and a broader appreciation for the conformation landscape that both the E2 protein and the E1E2 complex explore.

3. 1. 1. ColabFold Structure Prediction and Comparisons

At the start of this project, ColabFold was used to fold J6 WT E2 which produced five models. These were ranked 1 through 5 based on their pLDDT score by ColabFold. These models were then viewed in ChimeraX v1.9. These initial models can be seen in Figure 3. 1. Notably, the HVR-1 is placed in apposition to the rest of E2 in the rank 1 and rank 4 model. More specifically, the HVR-1 is placed in apposition to the stem region of the E2 protein – we termed this conformation “pocketing” of the HVR-1. In contrast, the HVR-1 in the rank 2, 3, and 5 models are predicted to be “free-floating” in 3D space and not taking part in any intramolecular interactions.

At the time of producing these models, the “pocketing” of the HVR-1 was not considered to be particularly important. Rather, it was thought to be merely a product of ColabFold not being able to accurately predict where to position a disordered tail which could be in myriad conformations as seen in the rank 2, 3, and 5 models. That was until a new structure for the S52 HCV E1E2 homodimer complex was released (PDB: 8RJJ, Augestad *et al.*, 2024). Notably, this new structure showed the HCV E1E2 proteins as a dimer of dimers which had not been seen previously but even more strikingly, this structure also showed a pocketed HVR-1 – the first time that the HVR-1 had been resolved in any HCV E2/E1E2 protein structure. Therefore, the J6 WT E2 models that predicted a pocketed HVR-1 were compared to this new S52 structure using the Matchmaker and Contact functions in ChimeraX to compare how similar this novel cryo-EM structure was to the predicted J6 WT E2 models (Fig. 3. 2).

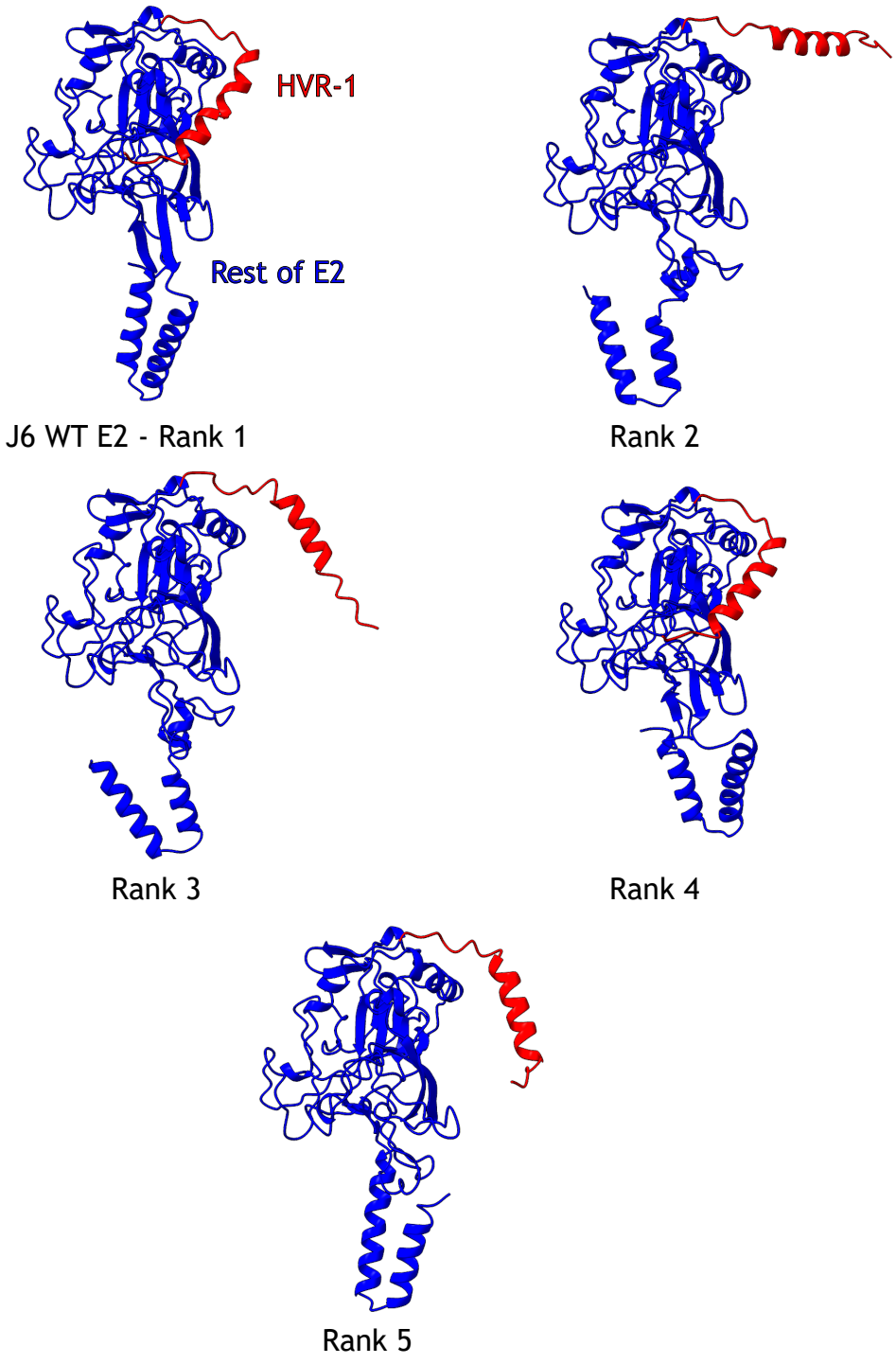


Figure 3. 1. **ColabFold Models of J6 WT E2.** The HVR-1 is highlighted in red, the rest of the E2 protein in blue. These were the initial five models produced using ColabFold. The models are ranked 1 through 5 based on their pLDDT scores. Notably, the HVR-1 is predicted to occupy a distinct position in 3D space in the rank 2, 3 and 5 models. However, the rank 1 and 4 models predict the HVR-1 to be in apposition to the rest of the E2 protein. This "pocketed" HVR-1 was, at the time of running this structure prediction, thought to be a potentially novel HVR-1 conformation.

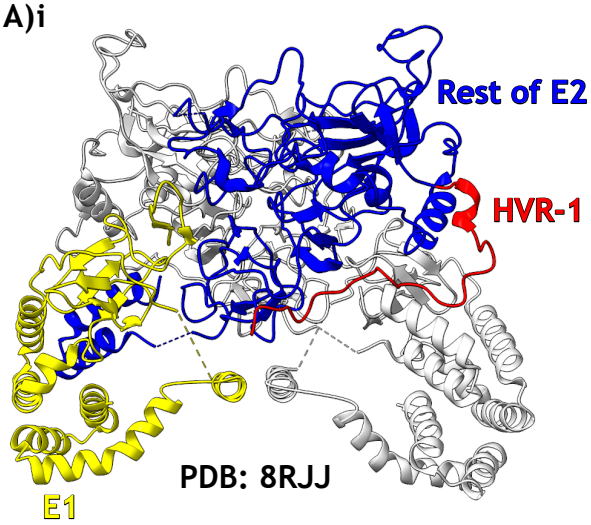
Figure 3. 2. A i is the S52 cryo-EM structure (PDB: 8RJJ) in full. The complex is a dimer of dimers with two E1E2 complexes. The E1E2 complex positioned away from the reader is highlighted in white. The other E1E2 complex is coloured as follows: HVR-1 = red, rest of S52 E2 = blue, S52 E1 = yellow.

Figures 3. 2. A ii is one of the E1E2 dimers from the full model. Next, the rank 1 and rank 4 predicted structures of the J6 WT E2 from ColabFold were aligned to the E2 from the 8RJJ structure. There are shown separately first (Fig. 3. 2. B i, ii, iii) and then aligned together in ChimeraX v1.9 using its Matchmaker function (Fig. 3. 2. B iv). It is striking how similar these models are to this protein structure, particularly in the placement of the HVR-1 into the stem region. Even more striking is that given the upload date of this PDB structure (Deposited: 21-12-2023) and the training data cut-off date for ColabFold (28-8-2019), ColabFold predicted this experimentally validated structure unbiased by having been trained on this structure.

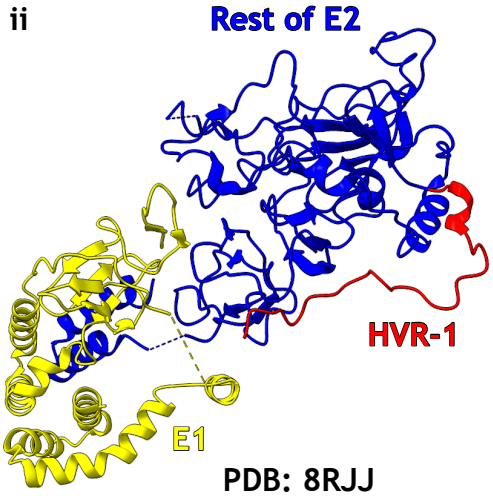
Furthermore, in figure 3. 2. C i-iii, the intramolecular interactions that are stabilising this pocketing of the HVR-1 were investigated. Only backbone:backbone interactions have been considered due to the hypervariability of the HVR-1 which may mean specific amino acids and their side chains may not be present in other E2 proteins.

First, the intramolecular interactions stabilising the pocketed HVR-1 in the S52 E2 have been focussed on (Fig. 3. 2. C i). There are two main regions of interaction that are stabilising the HVR-1. The first region is sandwiching of the very N-terminus of the HVR-1 with interactions above and below the HVR-1 from unstructured loops from the stem region. Above the HVR-1, residues R659 and G660 from the stem region interact with T385, the second residue of the HVR-1. From below, N706 is interacting with E384 – the first residue of the HVR-1 – and I707 is interacting with both T385 and Y386 – the second and third residues of the HVR-1. The second region of interaction involves S391 – the eighth HVR-1 residue – interacting with T636 and L637 from a ß-sheet in the back layer.

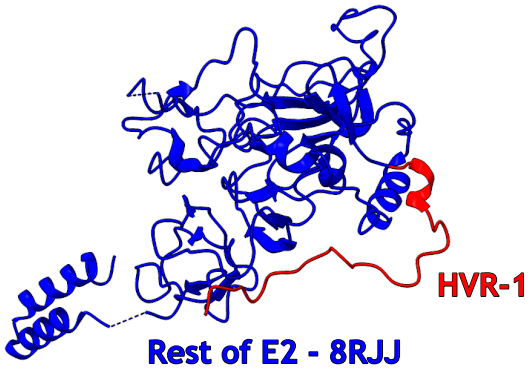
A)i



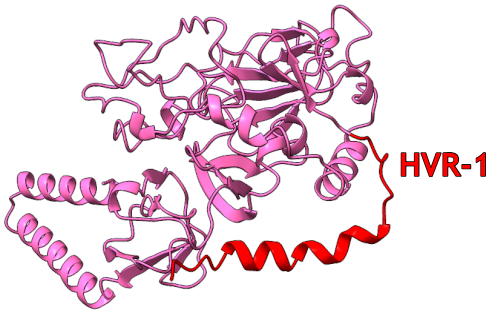
ii



B)i

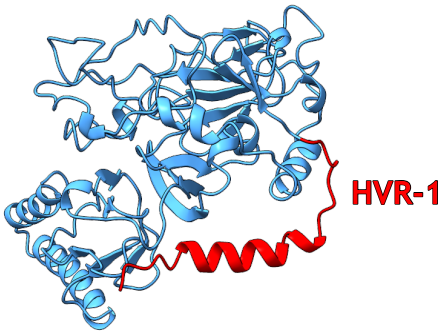


ii



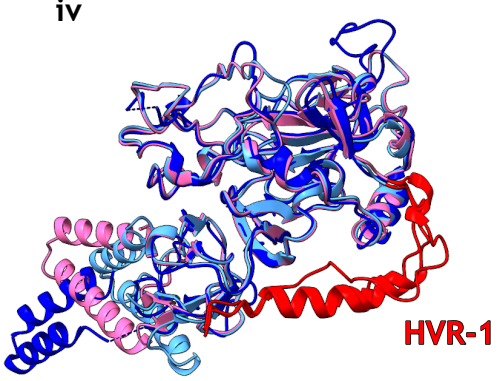
Rest of E2 - ColabFold J6 Rank 1 Model

iii



Rest of E2 - ColabFold J6 Rank 4 Model

iv



All three E2 structure/models

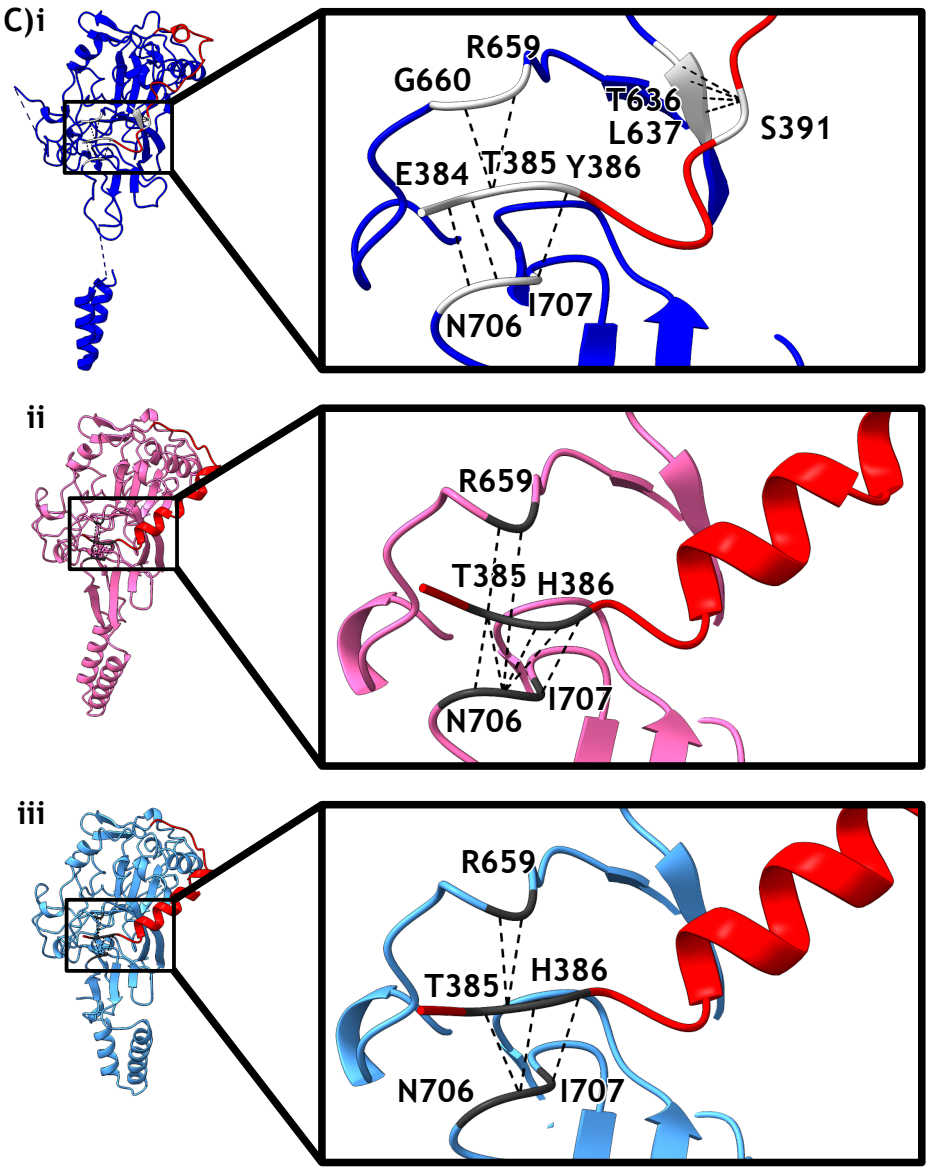


Figure 3. 2. **Comparison of Cryo-EM HCV S52 Structure (PDB: 8RJJ) with the ColabFold J6 WT E2 Models.** A) i Cryo-EM structure of the dimer of dimers complex of E1E2. One dimer is in white; the second dimer is split into yellow for E1, blue for the rest of E2, with the HVR-1 highlighted in red. A) ii E1E2 complex, coloured as before, but the second dimer removed. B) i E2 from the dimer of dimer complex on its own, coloured as before. B) ii Rank 1 J6 WT ColabFold predicted model: HVR-1 in red, rest of E2 in pink. B) iii Rank 4 J6 WT ColabFold predicted model: HVR-1 in red, rest of E2 in light blue. B) iv Structural alignment of the three E2 structures/models in B) i through iii. Structures were aligned in ChimeraX v1.9. C) i-iii The E2 structure/models, highlighting the backbone:backbone interactions between the HVR-1 and the rest of E2. Interactions were defined by residue backbones having an overlap of their Van der Waals radii of less than -0.4 Å. Residues involved in the backbone:backbone interactions are highlighted in a contrasting grey. Residues are renumbered as per the alignments in Appendix I. Contacts were mapped in ChimeraX v1.9.

Note that, for clarity and consistency throughout the thesis, all amino acids have been numbered as per the re-numbering alignment in appendix I.

These interactions are recapitulated in the ColabFold models. Considering the rank 1 model (Fig. 3. 2. C ii), the HVR-1 is, again, being stabilised by stem region unstructured loops above and below it. From above, R659 is predicted to interact with T385 – the second HVR-1 residue in J6. Notably in the ColabFold models, G660 is not predicted to stabilise the N-terminus of the HVR-1. From below, N706 and I707 are predicted to interact with T385, and further I707 is also predicted to interact with H386 – the third HVR-1 residue in J6. Moreover, no residue in the stem region is predicted to interact with the first residue of the HVR-1 as in the S52 structure. The rank 4 model (Fig. 3. 2. C iii) predicts near identical backbone:backbone interactions to stabilise the pocketed HVR-1 as the rank 1 model.

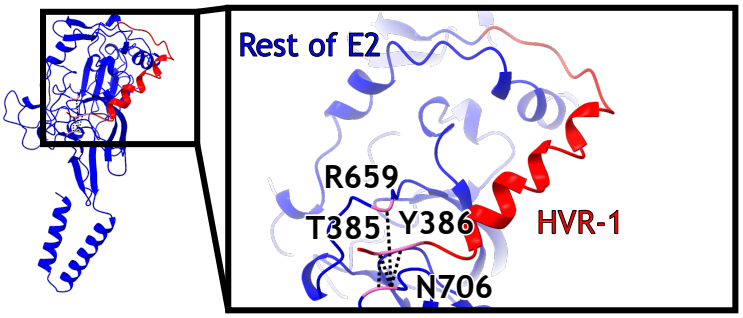
One major difference between the S52 structure and the J6 WT E2 models is the contribution of the back layer in stabilising this pocketed HVR-1 which is seen in the former and not the latter. This, however, may be due to the HVR-1 being predicted to adopt a helical conformation in the J6 models which may ablate these interactions in these predicted models.

Given how ColabFold predicts protein structures, and how it finds residue interactions by examining the co-evolutionary relationships between residues it is extremely surprising that ColabFold has managed to find this interaction. The similarities between these predicted models and the 8RJJ structure are even more surprising considering that 8RJJ could not have been included in the training data for ColabFold. Additionally, the HVR-1 has been known to be highly variable in its amino acid sequence since the early 90s, only a few years after the discovery of HCV as the etiological agent of NANBH in 1989 (Ogata *et al.*, 1991; Weiner *et al.*, 1991; Penin *et al.*, 2001). However, this is only one E2 protein sequence – J6 -and it is only two models of five which have found this interaction, so this is not conclusive evidence that there is an evolutionary relationship between the HVR-1 and the stem region, only identifiable through the application of deep machine learning. Moreover, as noted AlphaFold2 has

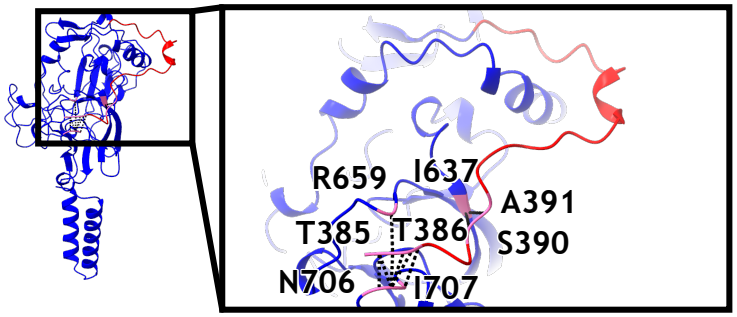
been trained to be self-correcting and ask itself if proteins make biological sense at the level of global protein fold (Jumper *et al.*, 2021; Mirdita *et al.*, 2022). Therefore, this structure may represent a snapshot of where the HVR-1 may be positioned, and indeed, this positioning may represent a minimal energy state of the E2 protein. ELISAs performed by Augestad *et al.* showed reduced binding between the E1E2 dimer of dimers complex and the antibody AR4A at 4 °C, but not at 37 °C (Augestad *et al.*, 2024). In the 8RJJ structure, the pocketed HVR-1 is covering antigenic region 4 – the region that AR4A targets. This suggests that the pocketed HVR-1 may be representative of a low energy state of the E2 protein wherein the HVR-1 disassociates from this pocket at higher temperatures/higher energy states. However, whether this prediction from ColabFold is merely a representation of a low energy state of the E2 or is pointing towards some hitherto unknown evolutionary relationship between the HVR-1 and the stem region is something that may be explored further with folding more E2 proteins with ColabFold.

Next, representative sequences from each of the seven HCV genotypes were folded with ColabFold (Fig. 3. 3). The folded E2s from each of the genotypes, except MH940291 (GT7), predicted a pocketed HVR-1. Moreover, each pocketed HVR-1 recapitulated some, if not all, of the same regions of interaction stabilising the HVR-1:stem region interaction as seen in the S52 structure. MK328019 (GT4), EF420132 (GT6), and MH940291 (GT7) were predicted to have novel HVR-1:stem region intramolecular interactions which were not seen in the S52 structure.

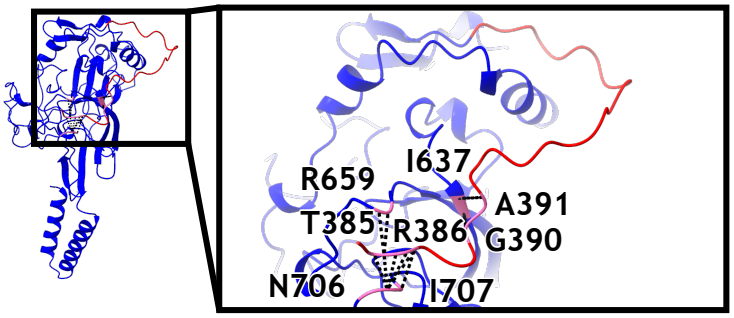
Starting with KJ439778 (GT1), the HVR-1 is stabilised in the pocketed position with the stem region sandwich of unstructured loops as seen in all the previous models. The second residue of the HVR-1 – T385 – is stabilised by the analogous R659 above the HVR-1. The bottom stabilising loop has N706 stabilising both T385 and Y386 – analogous to the same interactions in the 8RJJ structure.



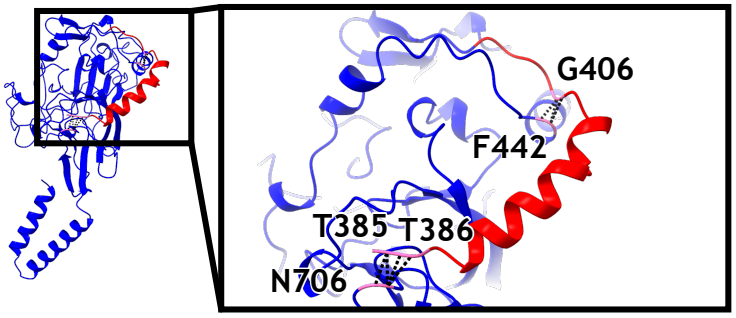
KJ439778 - GT1



JF735119 - GT2



AY545957 - GT3



MK328019 - GT4

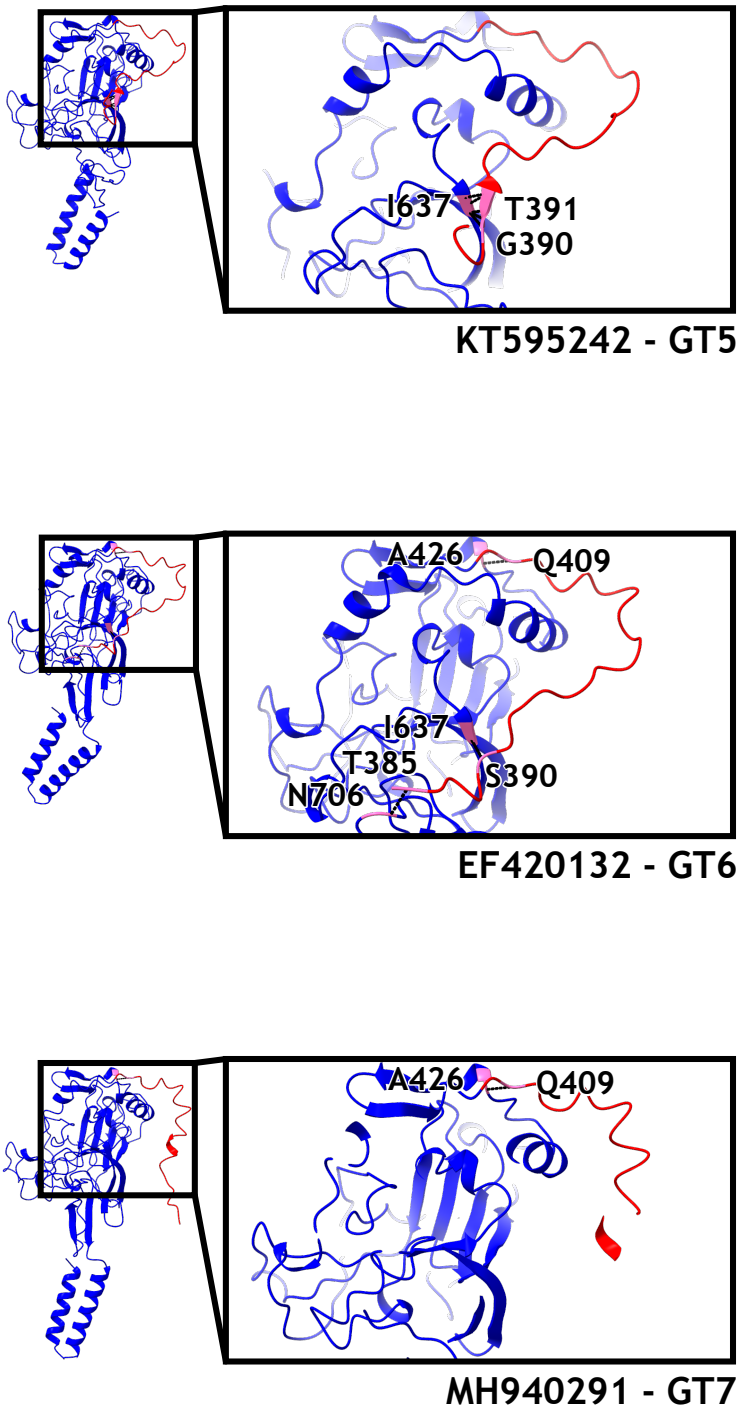


Figure 3. 3. **ColabFold Models of E2 Proteins from Each of the Seven HCV Genotypes.** A representative rank 1 ColabFold model from each of the seven HCV genotypes was investigated to look at HVR-1:rest of E2 interactions across genotypes. The HVR-1s are highlighted in red, the rest of E2 in blue, and the residues predicted to be in contact with one another are highlighted in pink with black dotted lines between contacts. Residue contact predictions were defined by having an overlap of their Van der Waals radii of less than -0.4 Å which was calculated and mapped in ChimeraX v1.9. Residues were renumbered as per the alignment in Appendix I.

JF735199 (GT2), again, is predicted to have a stabilised HVR-1:stem region unstructured loop sandwich conformation. Moreover, it is predicted to have the back layer stabilisation as seen in the S52 model. This model closely resembles the S52 structure with the N-terminus of the HVR-1 being stabilised from above by R659 interacting with T385 (the first residue of the HVR-1) alone. From below the HVR-1, N706 is predicted to interact with both T385 and T386. I316 is predicted to interact with T386. Furthermore, S390 and A391 in the HVR-1 are predicted to interact with I637 in the back layer – like the S391:T636/L637 interactions in the S52 structure.

AY545957 (GT3), like the JF735119 model and S52 structure features the two regions of stabilisation – the unstructured loop stem region:HVR-1 sandwich and the interaction of the HVR-1 with the back layer. Precisely, R659 from the stem region is predicted to interact with T385 in the HVR-1 from above. From below, N706 is predicted to interact with T385 and R386, and I707 is predicted to interact with R386. In the back layer, I637 is predicted to interact with both G390 and A391.

MK328019 (GT4) is predicted to contain a novel interaction between the HVR-1 and the front layer – G406:F442 – this is the only model to contain this interaction. Additionally, it contains an interaction between an unstructured stem region loop and the HVR-1 but in this predicted model, the interaction only comes from below with N706 interacting with both T385 (the first HVR-1 residue in this structure) and T386.

KT595242 (GT5) is only predicted by ColabFold to have the HVR-1:back layer interaction as stabilised by G390 and T391, of HVR-1, interacting with I637. However, this is the only HVR-1 which is predicted to be a ß-strand interacting with the ß-sheet in the back layer. It is also the only prediction in which the HVR-1 appears to double back on itself. This would, in this model, prevent the HVR-1 from interacting with unstructured loops from the stem region as it is simply too far from those loops.

EF420132 (GT6) also features a novel interaction between the HVR-1 and the front layer. This interaction is further back in the HVR-1 than in MK328019 (GT4), taking place between Q409 of the HVR-1 and A426 of the front layer. Other than this, it is still predicted to have the HVR-1 stabilised by an unstructured stem region loop from below - as seen in the T385:N706 interaction – and is also stabilised through HVR-1:back layer interactions – S390:I637.

MH940291 (GT7), as noted, is not predicted to have a pocketed HVR-1. However, there is still one interaction between the HVR-1 and the front layer – Q409:A426. This interaction is analogous to the Q409:A426 interaction predicted in the EF420132 (GT6) model.

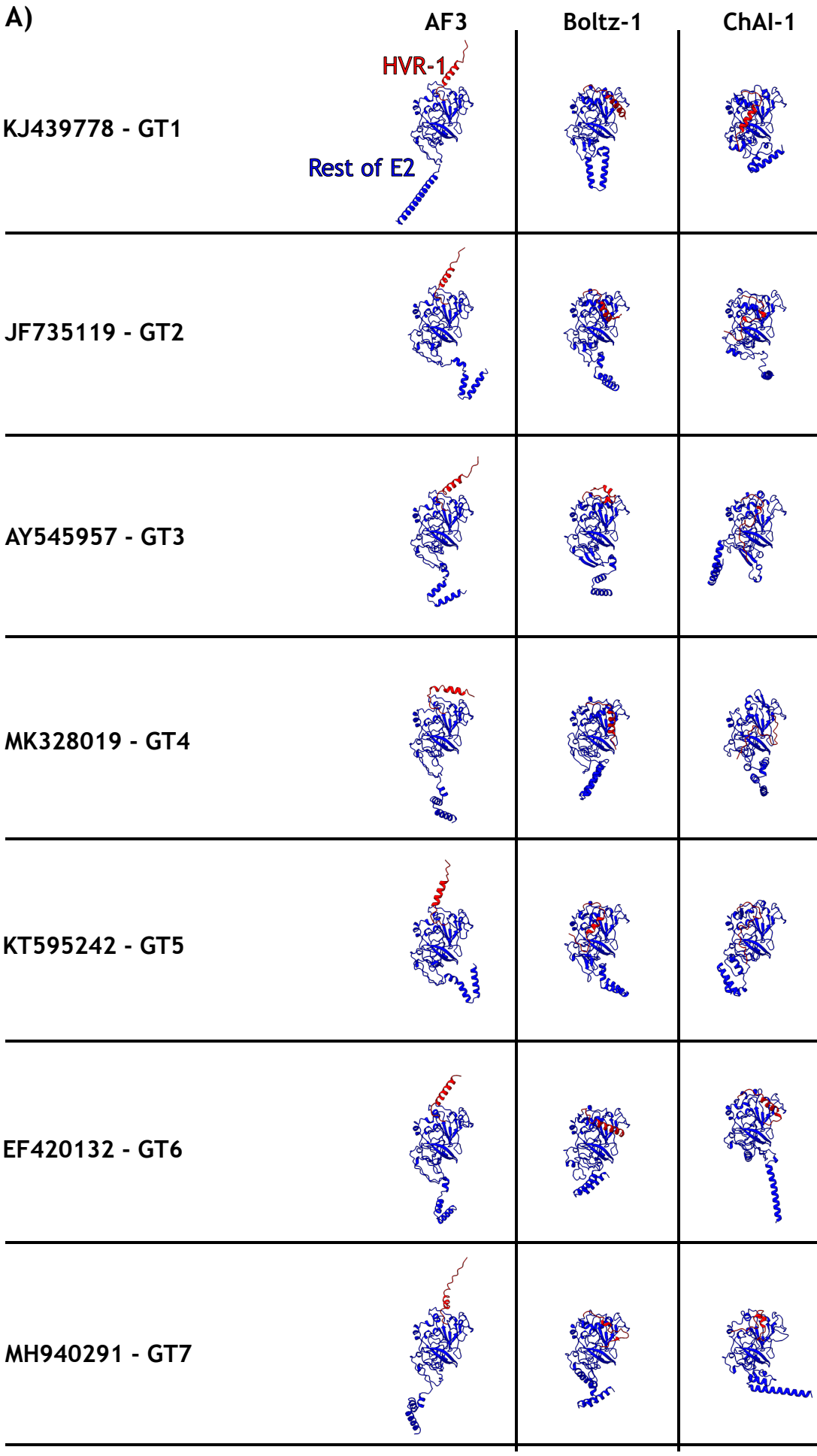
One thing that becomes strikingly obvious when comparing the rank 1 models across the seven genotypes of HCV is how often analogous interactions are seen. Residues outwith the HVR-1 which are consistently predicted to interact with the HVR-1 include R659 (predicted in 3/7 models), N706 (5/7), and I637 (4/7). For R659 and N706, these are mostly seen interacting with threonine residues in either the first or second position of the HVR-1. I637 is seen interacting with either the 7th or 8th residue of the HVR-1. R659 and N706 are 100 % conserved across all seven input sequences. I637 is conserved across five of the seven sequences, being replaced with leucine in KJ439778 (GT1), and a threonine in MK328019 (GT4). Given the conservation of the residues outwith the HVR-1, these predicted interactions could be a reflection of the evolutionary relationships that ColabFold can explore and from which it can extrapolate positional relationships between residues (Jumper *et al.*, 2021; Mirdita *et al.*, 2022). It may also speak to the evolutionary conservation of the pocketing interaction, possibly due to functional importance.

3. 1. 2. Comparisons Between Protein Folding Software Packages

Next, these representative sequences were folded with different folding software packages to see how they would each perform and if they could recapitulate the pocketed HVR-1 conformation as seen in the S52 E2 cryo-EM structure.

First, the representative sequences from each of the seven genotypes were folded by AlphaFold 3, Boltz-1 and ChAI-1 (Fig 3. 4. A). Note, that these models were structurally aligned to one another in ChimeraX before being arranged in the columns such that they should all be in the same orientation. The first column contains the E2 structures as predicted by AlphaFold3. Most strikingly with the AlphaFold3 models is that none of them show a pocketed HVR-1 – all of them position the HVR-1 in different orientations in 3D space. The next column shows the outcome of the Boltz-1 models which do position the HVR-1 in apposition to the rest of the E2 protein but not with the N-terminus of the HVR-1 in the "pocket" between the base and the stem region of the E2 protein for most models – it is only in the KT595242 (GT5) model that the HVR-1 points towards the base/stem region cavity. However, the very N-terminus of the HVR-1 appears to tuck back on itself and therefore does not interact with the same region predicted by ColabFold. Instead, the rest of the Boltz-1 models predict the HVR-1 to be tucked under the CD81 binding loop. The third column shows the models as predicted by ChAI-1. Their HVR-1s, in general, are pointing towards the base/stem region cavity. However, there are kinks in the HVR-1s for GT2, GT3, GT4, and GT5 which may prevent HVR-1:stem region interactions. Overall, the ChAI-1 models are the most variable.

A)


AF3
Boltz-1
ChAI-1
HVR-1
Rest of E2
KJ439778 - GT1
JF735119 - GT2
AY545957 - GT3
MK328019 - GT4
KT595242 - GT5
EF420132 - GT6
MH940291 - GT7

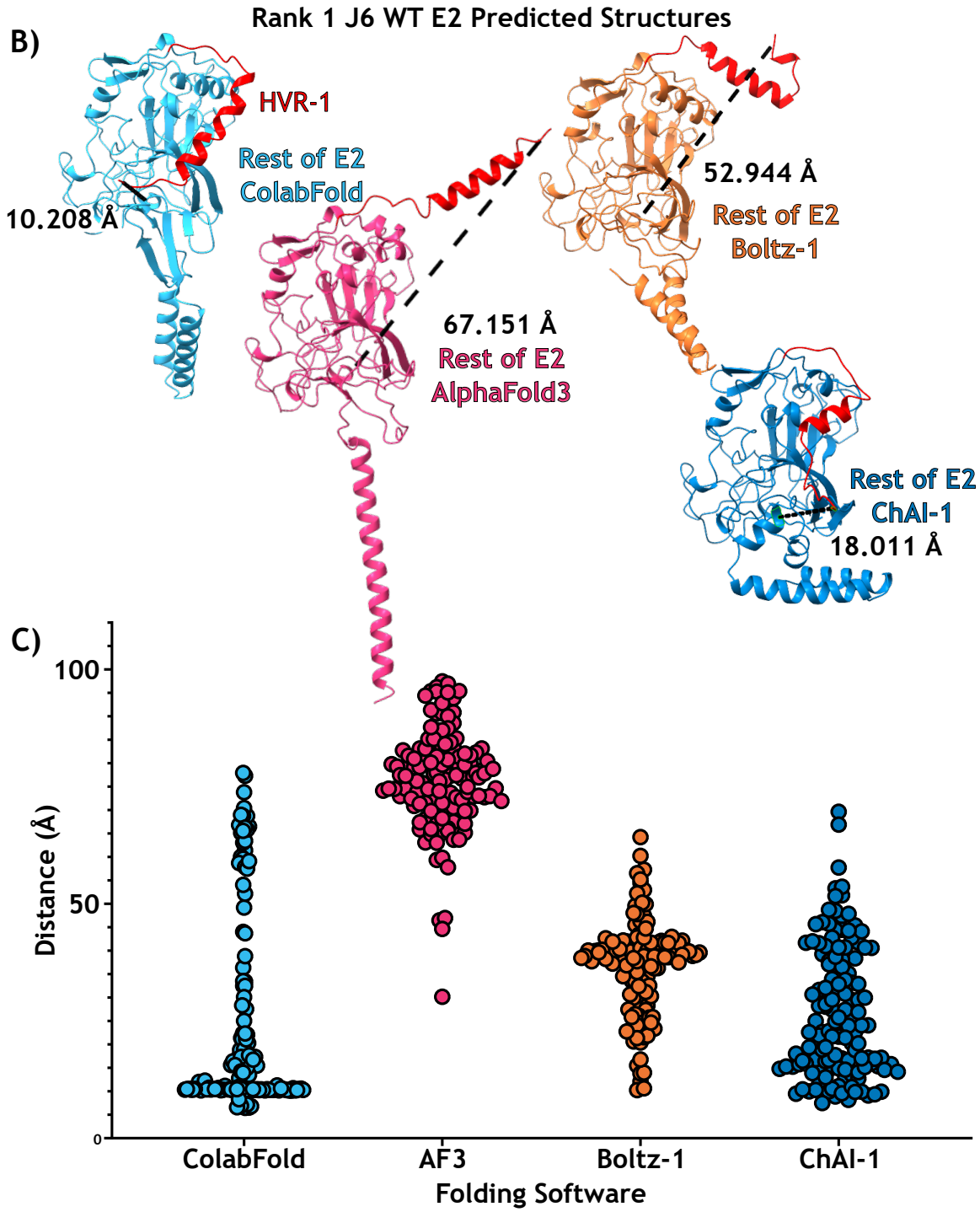


Figure 3. 4. **Comparing the Predicted HVR-1 Positioning Across Different Protein Folding Software Packages - Rank 1 Models Only.** A) Predicted representative E2 structures across each of the genotypes, modelled by AlphaFold 3 in the first column, then Boltz-1 and ChAI-1 in column 2 and 3 respectively. HVR-1: red, Rest of E2: blue. B) Predicted J6 WT E2 structures for each of the protein folding software packages with the straight-line distance from the Cα of R384 to the Cα of V708 measured in ChimeraX. HVR-1 in all models: red, ColabFold: light blue, AlphaFold 3: pink, Boltz-1: orange, ChAI-1: dark blue. C) The distances obtained when measuring from the first HVR-1 residues to V708 across all 134 E2 structures and the four different protein folding software packages. Residues renumbered as per alignment in Appendix I.

None of the diffusion models were able to consistently find the pocketed HVR-1 structure across these seven representative sequences. However, this could reflect these particular sequences. Therefore, to attempt to quantify whether these diffusion models were able to find the pocketed HVR-1 conformation, further representative E2 sequences from across all seven genotypes of HCV were folded with ColabFold, AlphaFold3, Boltz-1, and ChAI-1. Nine-thousand one-hundred and three unique and valid (no non-canonical amino acids) HVR-1 sequences were extracted from 30,289 HCV sequences that covered ≥25 % of the E2 protein. At the same time, the corresponding E2 sequence information was also captured. Following a series of filtering and clustering steps, the number of potential E2 sequences to fold was reduced from 9,103 to 105 unique and valid E2 sequences – a more computationally manageable number of sequences but still representative of all genotypes. To these 105 E2 sequences, the sequences for the mutant J6 E2 constructs designed for the wet lab experiments - see Chapter 4 - were added to give a final dataset of 134 E2 sequences.

To quantify pocketing of the HVR-1, the Euclidean straight-line distance was calculated between the first N-terminal residue Cα of the HVR-1 and the Cα of a conserved residue within the stem region. Across all 105 E2 sequences, V708 was 100 % conserved and within the HVR-1 pocket despite not being predicted to interact directly with the HVR-1 in any of the above models. Therefore, after folding, the distance between the first N-terminal residue Cα of the HVR-1 and the Cα V708 was measured. To provide context for quantification, the J6 WT E2 HVR-1-stem region measurements were first established across the four folding software packages (Fig. 3. 4. B). The ColabFold modelled J6 WT E2, which features a pocketed HVR-1, had a distance of 10.208 Å between the N-terminal HVR-1 residue (R384) and V708. This was considered the baseline - a measurement between the N-terminus residue of the HVR-1 and V708 equal to this distance should represent a pocketed HVR-1. The AlphaFold3 model had 67.151 Å between the HVR-1 and V708 and does not feature a pocketed HVR-1. The Boltz-1 J6 WT E2 model does not predict a pocketed HVR-1 and had an R384 to V708 measurement of 52.944 Å. Lastly, the ChAI-1 model which featured an HVR-1 in apposition to the rest of E2 but without the N-terminus of the HVR-1 directly interacting with the HVR-1 pocket had an R384 to V708 measurement of

18.011 Å. After folding J6 WT E2, each of the 134 E2 sequences were folded with ColabFold, AlphaFold3, Boltz-1 and ChAI-1 and the rank 1 model for each folding software platform for each E2 sequence was measured from the first HVR-1 residue of the E2 protein to the conserved V708 (Fig. 3.4.C).

The ColabFold distances cluster around 10 Å, going up to around 80 Å. Given the previous model of J6 WT E2 and the observation that an approximately 10 Å distance between the first residue and V708 would imply that the HVR-1 is pocketed, it can be surmised that the majority of the E2 folds predicted by ColabFold tend to place the HVR-1 into the stem region pocket. The AlphaFold3 distances cluster around 75-85 Å but are less clustered than ColabFold. Again, taken together with the observations from figure 3. 4. B, this would suggest that AlphaFold 3 never pockets the HVR-1 into the stem region. Interestingly, Boltz-1 distances cluster around 40 Å. It does not appear to correlate with the pocketing conformation of the HVR-1 into the stem region, but the clustering does imply that the HVR-1 is being placed consistently into a similar conformation across the Boltz-1 models. Lastly, the ChAI-1 distances appear more spread out, ranging from 10-70 Å. Those smaller measurements around 10 Å may correlate to a pocketed HVR-1 but the distance measurements are not as clustered as those measurements from the ColabFold models or, indeed, the Boltz-1. In general, the measurements for ChAI-1 are less clustered than the other folding software platforms outputs. Taken together, these data suggest that the diffusion-based protein prediction software packages are not finding the pocketed HVR-1. However, they are exploring other conformations. The lack of clustering for AlphaFold3 and ChAI-1 may suggest they are exploring more conformations than Boltz-1, where the majority of measurements cluster around 40 Å. This suggests that an evolutionary fingerprint that ColabFold identifies, that these other models do not, facilitates the prediction of a pocketed HVR-1. However, this simple straight-line measurement does not speak to the global 3D structure of these E2 models, and it cannot be accurately deduced from this method whether the HVR-1 has been pocketed.

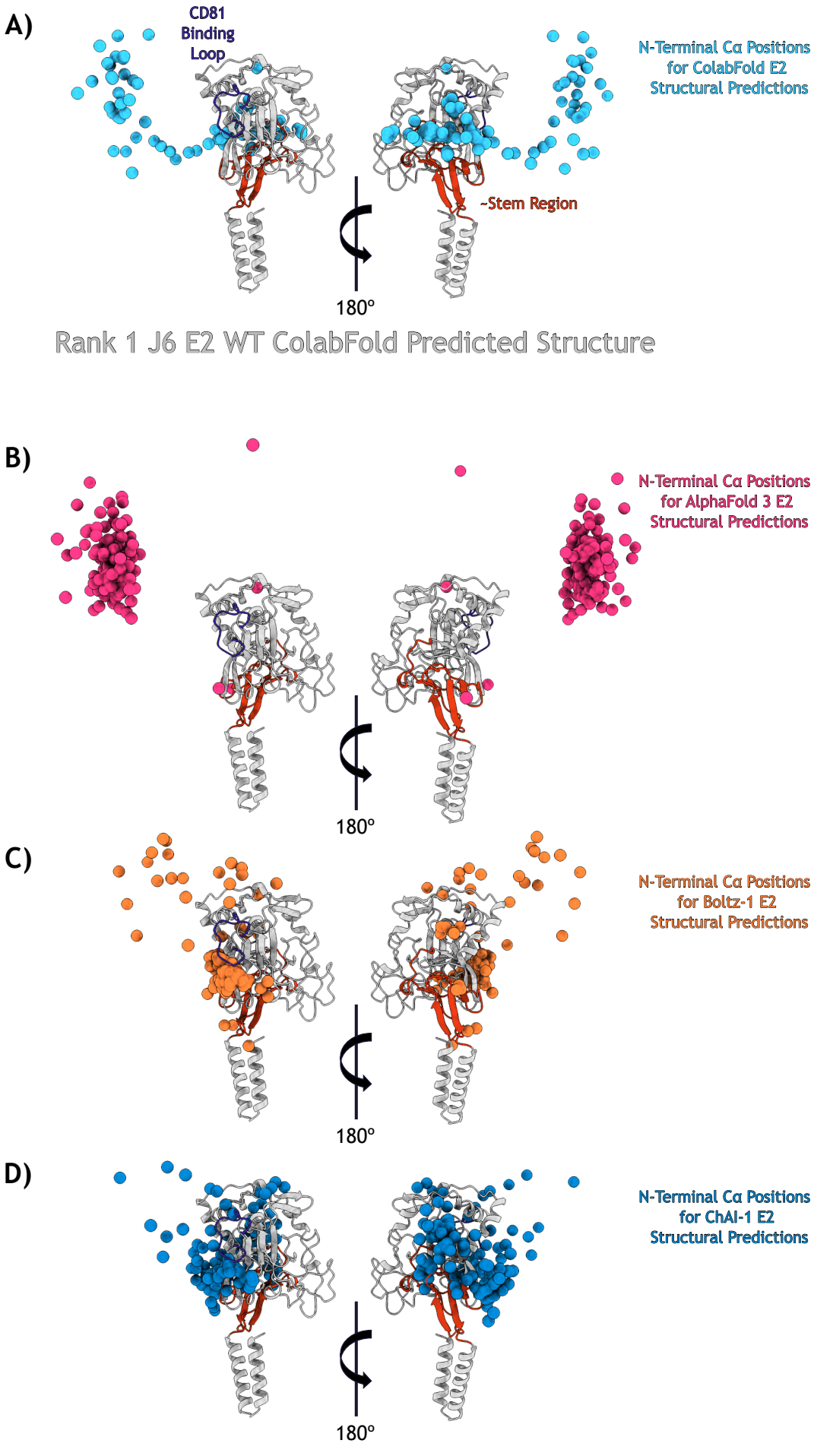
Therefore, to investigate the global 3D structure of the E2 models and how they positioned the HVR-1, the N-terminus residue Cα positional information was

extracted and then modelled around the ColabFold model of J6 WT E2. In figure 3. 5. A, only the N-terminus residue Cαs from the ColabFold predicted models are visible. A minority of N-terminal Cαs are predicted to extend into 3D space. Conversely, the majority of HVR-1s are predicted by ColabFold to lay in apposition to the rest of the E2 protein, "pocketing" the HVR-1 into the stem region; similar to its position in the 8RJJ S52 E1E2 cryo-EM structure. Moreover, this does reflect the distances and clustering of distances around the 10 Å measurement seen in figure 3. 4. C.

AlphaFold 3, however, does not predict the N-terminal residue Cαs to be in apposition to the rest of E2, instead positioning the Cαs in an extended position in 3D-space (Fig. 3. 5. B). The HVR-1 adopts an extended conformation which, again, matches up with the measurements seen in figure 3. 4. C.

Boltz-1, interestingly, predicts the majority of the N-terminal residue Cαs of the HVR-1 to be positioned under the CD81 binding loop as seen in the left-hand model (Fig. 3. 5. C). Whilst this HVR-1/E2 conformation has not been experimentally validated, given the consistency of this fold across 134 E2 sequences, these models may represent a novel HVR-1 conformation. Alternatively, given how the diffusion modules work and how they are suited to exploring conformational possibilities, this may simply represent a biologically possible position.

Lastly, ChAI-1 predicts the most variable positioning of the HVR-1 N-terminal residue Cαs; placing the Cαs either in the stem region pocket or underneath the CD81 binding loop but rarely predicts the HVR-1 in an extended position as AlphaFold3 does (Fig. 3. 5. D).


A)
CD81
Binding
Loop
N-Terminal Cα Positions
for ColabFold E2
Structural Predictions
Stem Region
180°
Rank 1 J6 E2 WT ColabFold Predicted Structure
B)
N-Terminal Cα Positions
for AlphaFold 3 E2
Structural Predictions
180°
C)
N-Terminal Cα Positions
for Boltz-1 E2
Structural Predictions
180°
D)
N-Terminal Cα Positions
for ChAI-1 E2
Structural Predictions
180°

E)

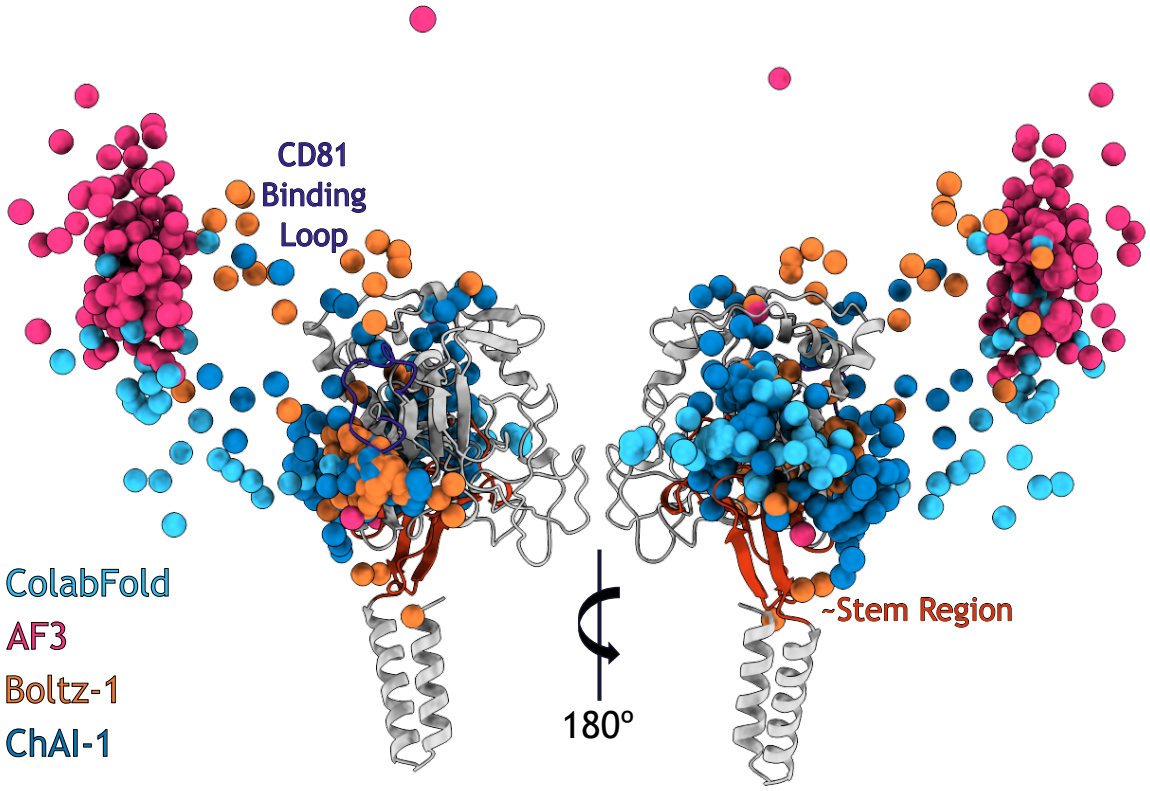
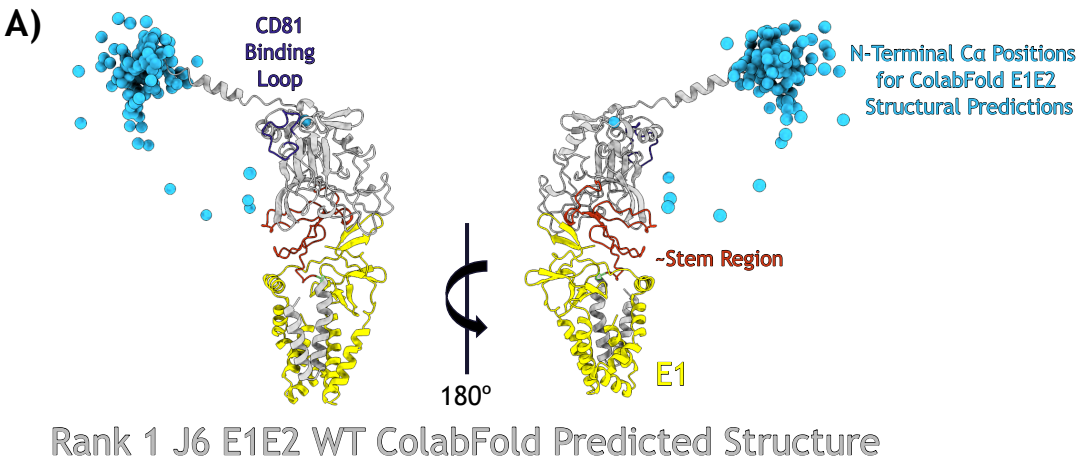


Figure 3. 5. **Relative 3D Positions of the Alpha Carbon of the N-terminus HVR-1 Residues as Predicted by Different Protein Folding Software Arranged Around the ColabFold J6 WT E2 Predicted Structure.** J6 WT E2 ColabFold model: grey, CD81 binding loop: darkest blue, ~Stem region: red, ColabFold N-terminal Cαs: light blue, AlphaFold 3 N-terminal Cαs: pink, Boltz-1 N-terminal Cαs: orange, ChAI-1 N-terminal Cαs: dark blue. A)-D) shows the N-terminal Cα positions as predicted by ColabFold, AlphaFold 3, Boltz-1, and ChAI-1 respectively, framed around the ColabFold predicted structure for J6 WT E2. E) shows the results for all four folding software collated around the ColabFold J6 WT E2 model. Full-length and valid E2 sequences were derived from the 9,103 accessions from which HVR-1 sequences were obtained and then clustered using MMSeqs2 to produce a list of 105 E2 sequences to which the mutant E2 construct sequences were added. This resulted in 136 sequences which were folded using ColabFold, AlphaFold 3, Boltz-1, and ChAI-1. These models were structurally aligned to the ColabFold J6 WT E2 model and the co-ordinate information for the first N-terminal residue Cα was extracted for visualisation - both the structural alignment and visualisation were carried out in ChimeraX v1.9.

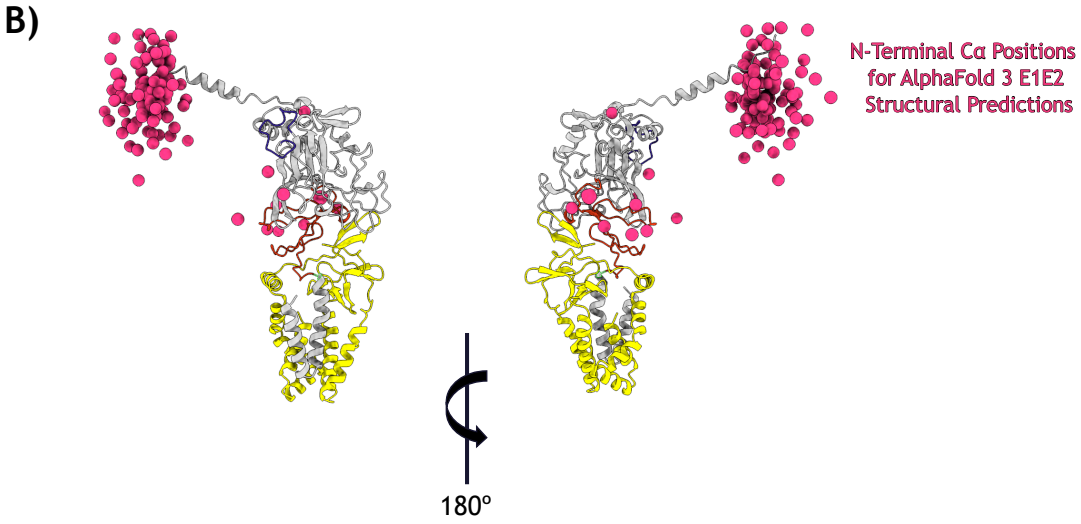
Considering the outputs of all four protein structure prediction software packages together, ColabFold consistently places the HVR-1 into the stem region pocket. AlphaFold3 predicts a “free-floating” HVR-1 which does not lie in apposition to the rest of E2. Boltz-1 predicts a majority of HVR-1s to tuck under the CD81 binding loop. ChAI-1 predicts the HVR-1 to rest in apposition to the rest of E2 – either in the stem region pocket, or underneath the CD81 binding loop.

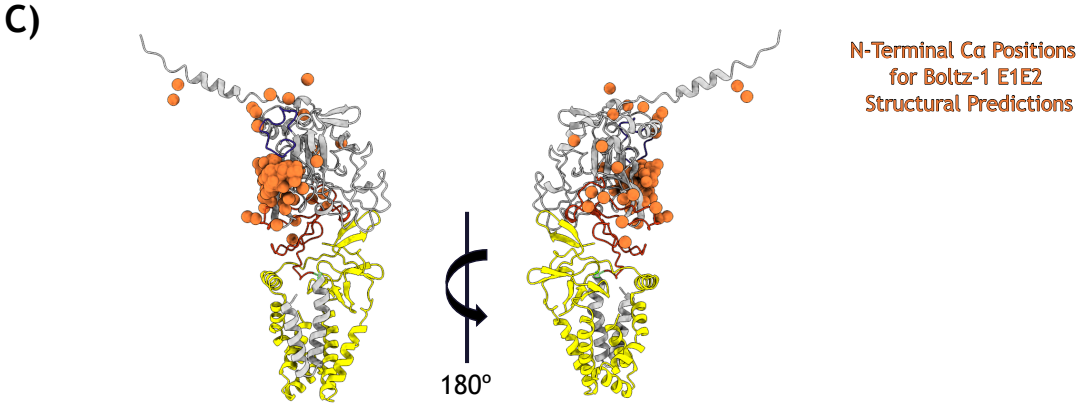
Next, E1E2 complexes were modelled with these four protein folding software platforms and, again, the N-terminal residue Cα positional information was visualised around the J6 WT E1E2 ColabFold model. Before folding any E1E2 complex, the corresponding E1 sequences to the 105 E2 sequences previously folded had to be manually curated to ensure that the sequences were valid i.e. contained no non-canonical amino acids in their sequence which may have obfuscated protein folding. Unfortunately, not each E2 sequence previously folded had a corresponding valid E1 sequence and therefore only 96 E1E2 complexes were folded. Again, the mutant J6 E1E2 constructs that were designed for use in the wet experiments were also folded, however, it should be noted that only the E2 sequences were mutated for the experiments ergo each of these constructs had a J6 WT E1 sequence. These models are included in the E1E2 and E2 alone bubble models in figures 3. 5. and 3. 6.

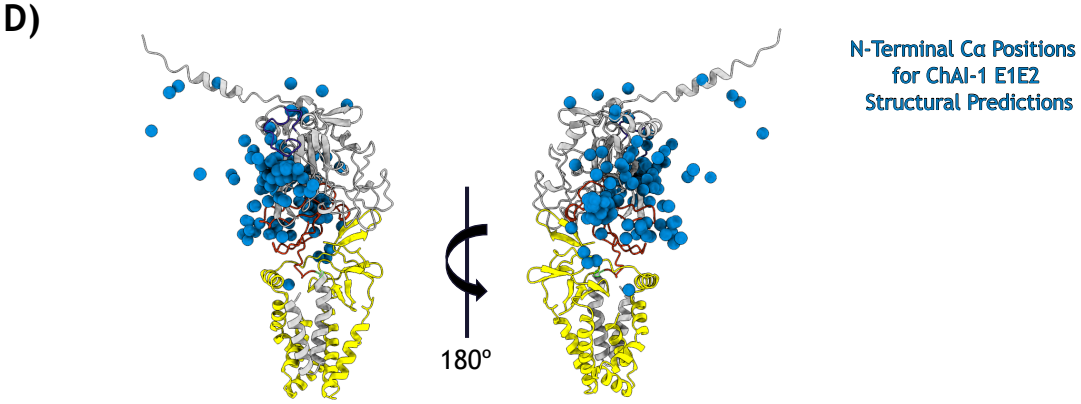
In figure 3. 6. A, ColabFold predicts the N-terminal residue Cα of the HVR-1 to now be extended out into 3D space, similar to AlphaFold3. Interestingly, no N-terminal residue Cαs are predicted to be in the stem region pocket. The ability to find that HVR-1:stem region E2 conformation has been ablated by the addition of E1 to the model. This is the most significant difference between figures 3. 5 and 3. 6.



Rank 1 J6 E1E2 WT ColabFold Predicted Structure







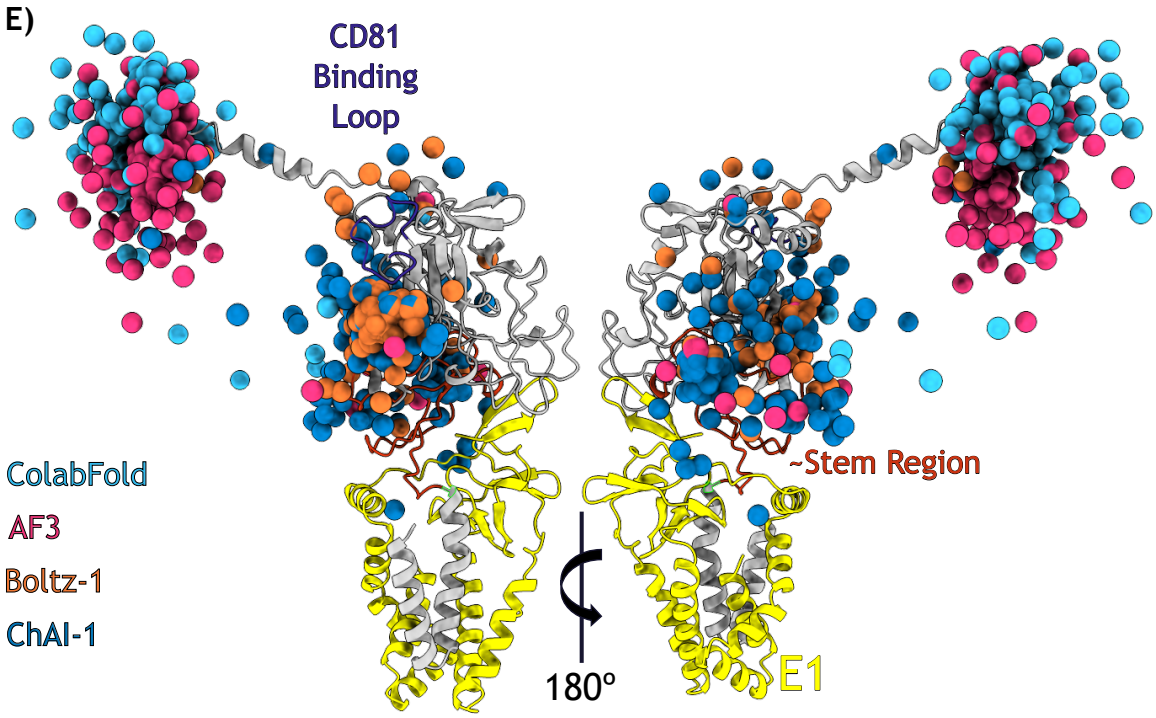


Figure 3. 6. **Relative 3D Positions of the Alpha Carbon of the N-terminus HVR-1 Residues as Predicted by Different Protein Folding Software Arranged Around the ColabFold J6 WT E1E2 Predicted Structure.** J6 WT E2 ColabFold model: grey, J6 WT E1 ColabFold model: yellow, CD81 binding loop: darkest blue, ~Stem region: red, ColabFold N-terminal Cαs: light blue, AlphaFold 3 N-terminal Cαs: pink, Boltz-1 N-terminal Cαs: orange, ChAI-1 N-terminal Cαs: dark blue. A)-D) shows the N-terminal Cα positions as predicted by ColabFold, AlphaFold 3, Boltz-1, and ChAI-1 respectively, framed around the ColabFold predicted structure for the J6 WT E1E2 complex. E) shows the results for all four folding software packages collated around the ColabFold J6 WT E1E2 complex model. From the list of 105 full-length and valid E2s which were folded in Fig. 3.5, the corresponding E1 sequences were interrogated to find full-length and valid E1 sequences of which 96 sequences were found. Again, the mutant construct E1 sequences were added - this was the J6 WT E1 sequence as no mutations were made to the E1 protein in the constructs used in this work. This resulted in 125 E1E2 complex sequences which were folded using ColabFold, AlphaFold3, Boltz-1, and ChAI-1. These models had their E2 proteins structurally aligned to the ColabFold J6 WT E2 model and the co-ordinate information for the first N-terminal residue Cα was extracted for visualisation - both the structural alignment and visualisation were carried out in ChimeraX v1.9.

AlphaFold3, again, predicts the N-terminal residue Cαs in an extended 3D position in the E1E2 complex models (Fig. 3. 6. B). Moreover, the Boltz-1 platform predicts that the N-terminal residue Cαs are tucked under the CD81 binding in the E1E2 complexes, like the E2-alone models (Fig. 3. 6. C). Lastly, ChAI-1 predicts the N-terminal residue Cαs in the E1E2 complexes in the same positions as in the E2-alone models – in apposition to the rest of E2, either in the stem region cavity or tucked under the CD81 binding loop (Fig. 3. 6. D).

When considering all the folding software platforms together (Fig. 3. 6. E), it can be easily seen that for AlphaFold 3, Boltz-1, and ChAI-1 that the HVR-1 is predicted to occupy mostly the same spaces for the E1E2 complexes as when E2 was modelled alone. ColabFold, however, no longer predicts the pocketed HVR-1 conformation when folding E2 in complex with E1. The loss of pocketing may be explained by loss of the evolutionary fingerprint that ColabFold was able to pick up on when looking at E2-alone. ColabFold has an attention-based deep learning architecture. It is able to parse MSAs to find the most salient co-evolutionary signals and convert them into 3D positions (Jumper *et al.*, 2021; Mirdita *et al.*, 2022). However, in the context of heteromeric protein complexes AlphaFold-Multimer, which is used by ColabFold for protein complexes, looks for regions of co-evolution between the two proteins to determine the interface between the two proteins whilst considering the co-evolutionary signals within individual proteins that aid in deciphering 3D protein structure for individual chains (Evans *et al.*, 2022). In considering the most salient co-evolutionary signals from the heterodimeric E1E2 models, it may be that whatever signal was being parsed from the E2-alone MSAs that led to the HVR-1 being positioned in the stem region pocket is ablated or disregarded in the overall E1E2 context. Considering that the diffusion models are less reliant on co-evolutionary signals, and are better able to explore possible protein conformations when compared to ColabFold, it intuitively makes sense why there was no drastic change in HVR-1 position across the E1E2 models and the E2-alone structure (Abramson *et al.*, 2024; Chai Discovery *et al.*, 2024; Wohlwend *et al.*, 2025).

Altogether, considering both figure 3. 5 and figure 3. 6, the HVR-1 is predicted to occupy one of three distinct poses. Firstly, the pocketed HVR-1 which lies in

apposition with the rest of E2, interacting with the stem region of E2. This perhaps unexpected HVR-1 conformation has been experimentally validated (PDB: 8RJJ, Augestad *et al.*, 2024). Secondly, the "free-floating" HVR-1 wherein the HVR-1 is not in apposition to E2 at all. The HVR-1 is a flexible tail (Stejskal, Lees, *et al.*, 2020); this is supported by its lack of resolution in previous E1E2/E2-alone structures (e.g. PDB: 7T6X, Torrents de la Peña *et al.*, 2022). This "free-floating" HVR-1 is likely a snapshot in time of a disordered HVR-1. Lastly, there is the predicted HVR-1:CD81 binding loop interaction which is an entirely novel HVR-1 conformation. It may be that in physiological reality the HVR-1 transitions from one of these conformations or "poses "to another, or that the HVR-1 is required to adopt one of these "poses" to allow for downstream conformation changes to occur to facilitate the correct functioning of the E1E2 complex. Augestad *et al.* noted in their paper describing the 8RJJ structure that when the HVR-1 was pocketed that the AS412 and CD81 binding loop regions of the E2 protein were arranged in such a way as to disfavour CD81 engagement (Augestad *et al.*, 2024). Moreover, it has been demonstrated that deletion of the HVR-1 leads to increased soluble CD81 binding (Bankwitz *et al.*, 2010). HCV LVPs engage with SR-B1 prior to interacting with CD81 (Lindenbach and Rice, 2013). Therefore, it may well be that a pocketed HVR-1 requires interaction with SR-B1 to re-arrange the AS412 and CD81 binding loop conformations to be in more favourable positions to engage CD81.

3. 2. Investigating Length Polymorphisms in the HVR-1

3. 2. 1. Prevalence of Length Polymorphisms in the HVR-1

Machine learning protein structure prediction software packages are predicting these distinct poses, which may be relevant for the biological function of the HVR-1. Naturally, this raises the question of how the occupancy rates of these various poses is regulated. Previous studies have considered how length polymorphisms in the HVR-1 affect its function. One study in 2007 examined the accumulation of insertions in the HVR-1 across 113 patients with and without cryoglobulinemia and found 6.2 % of HVR-1 sequences had insertions in patients with cryoglobulinemia and 9.1 % of HVR-1 sequences contained insertions in patients without cryoglobulinemia (Bianchettin *et al.*, 2007). A more recent study looked at the role of insertions in the HVR-1 on NAb evasion and found that insertions in the HVR-1 altered epitope-specific NAb sensitivity (Olesen *et al.*, 2025). Altering the length of HVR-1, therefore, may have an effect on disease outcomes (Bianchettin *et al.*, 2007) and the antibody sensitivity of the HCV virion. Furthermore, it has been noted in the literature that the HVR-1 can change length throughout the course of infection whether in patients or in cell culture (Abe, Inchauspe and Fujisawa, 1992; Kato *et al.*, 1992; Höhne, Schreier and Roggendorf, 1994; Ramirez *et al.*, 2009). It was hypothesised, therefore, that perhaps length polymorphisms in the HVR-1 may be part of the mechanism of regulating the occupancy rates of these various poses.

Length polymorphisms in the HVR-1 were investigated for prevalence and universality across all HCV genotypes. As previously noted, 9,103 unique and valid HVR-1 sequences were extracted from 30,289 sequences from HCV-Glue. The number of amino acids in each of the extracted HVR-1 sequences was then counted. This was then stratified by genotype and the modal length of HVR-1 was calculated per genotype (Table 3. 1.). The number of HVR-1 sequences containing insertions or deletions relative to the modal length HVR-1 for that genotype was then counted (Table 3. 2.).

Table 3. 1. Modal HVR-1 Lengths by Genotype

	GT1	GT2	GT3	GT4	GT5	GT6	GT7	All GTs
Modal Length HVR-1 (No. of AAs)	27	27	27	27	27	26	27	27

Table 3. 2. Breakdown of Length Polymorphisms and Modal Length HVR-1 Sequences by Genotype

	GT1	GT2	GT3	GT4	GT5	GT6	GT7	All GTs
Deletions	35	14	7	0	1	5	0	62
Modal Length	5,855	622	1,519	258	36	232	2	8,524
Extensions	240	22	25	3	0	227	0	517
Total per GT	6,130	658	1,551	261	37	464	2	9,103

These figures then allowed for the calculation of percentages to investigate the prevalence of length polymorphisms in the HVR-1 across all HCV genotypes (Fig. 3. 7). Unsurprisingly, the highest number of sequences came from genotypes 1 and 3, followed by genotype 2. Collectively, these are known as the epidemic strains of HCV and are known to be the most prevalent strains globally, but particularly within high-income countries which have greater access to testing and sequencing (Paraskevis *et al.*, 2021).

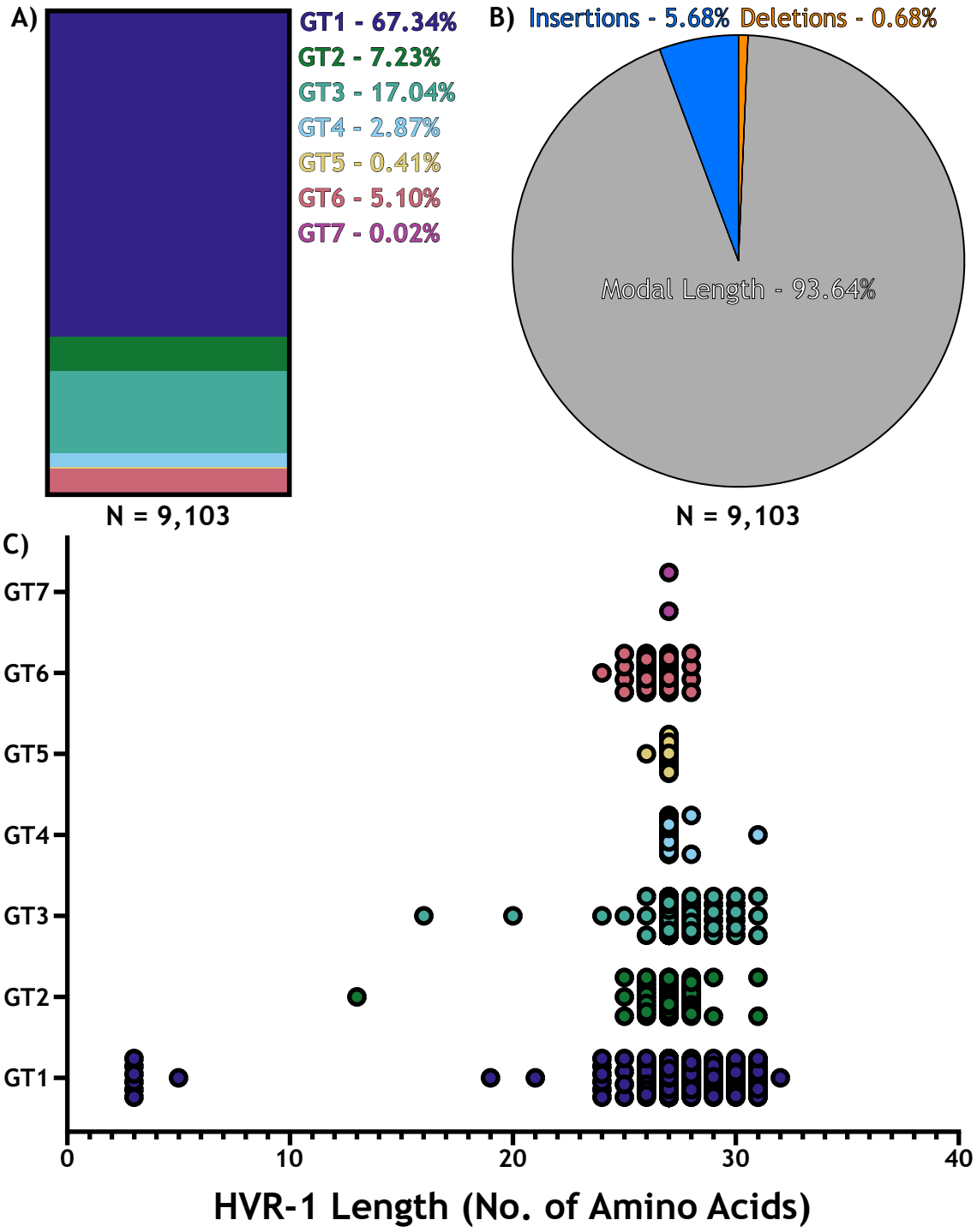


Figure 3. 7. **Investigating the Prevalence of Length Polymorphisms in the HVR-1 Across All HCV Genotypes.** A) Percentages of extracted HVR-1 sequences from each genotype. B) Percentages of length polymorphisms across each of the genotypes grouped into modal length HVR-1s, insertions and deletions. Note that for all genotypes barring genotype 6, the modal length HVR-1 is 27 amino acids in length. For genotype 6, the modal length is 26. C) Distribution of HVR-1 lengths by genotype. Genotype 1 through genotype 7 are down the y-axis; HVR-1 lengths in number of amino acids across the x-axis. Each dot represents an HVR-1 sequence of that length from that genotype.

Inversely, the other four genotypes of HCV are considered endemic strains and are found mostly in LMICs. The disparity in the numbers of sequences between the epidemic and endemic strains that have been sequenced likely reflects the disparity between high-income countries and LMICs in terms of provision and availability of testing and sequencing for HCV (Mohd Hanafiah *et al.*, 2013; Messina *et al.*, 2015; Roger *et al.*, 2021; The CDA Foundation, 2025).

However, the main finding from this analysis was that length polymorphisms are indeed prevalent across all HCV genotypes. In figure 3. 7. B, it is shown that over all HCV genotypes 5.68 % of all HVR-1 sequences were found to contain some form of insertion, and 0.68 % contained some form of deletion. Moreover, there are just over eight times more HVR-1 sequences which contain insertions than there are sequences containing deletions. This ratio of insertions to deletions agrees with the findings in other studies. In considering the studies referenced above, five of the six studies only encountered insertions in the HVR-1 (Abe, Inchauspe and Fujisawa, 1992; Höhne, Schreier and Roggendorf, 1994; Bianchettin *et al.*, 2007; Ramirez *et al.*, 2009; Olesen *et al.*, 2025). Whereas, only one study found both insertions and deletions across the HVR-1 (Kato *et al.*, 1992). These studies and the above data suggest that insertions within the HVR-1 are much more tolerable and stable than deletions. It might therefore be hypothesised that deletions within the HVR-1 are deleterious to viral fitness and therefore selected against throughout the course of HCV infection. These hypotheses, however, would require further experimentation to validate. Lastly, the distribution of HVR-1 lengths by genotypes was graphed (Fig. 3. 8. C). Each dot is equal to one sequence of a particular length from a particular genotype.

To our knowledge, this is the widest analysis of length polymorphisms within the HVR-1 that has been undertaken to date. However, the reported percentages for length polymorphisms across all seven genotypes is likely an underrepresentation of the true prevalence of length polymorphisms in the HVR-1. When extracting the HVR-1 sequence information from HCV-Glue, there was no meta-analysis of the sequencing methods used by the research teams to obtain said sequence information. Short read **next-generation sequencing** (NGS) protocols can struggle with the accurate assessment of genetic variation which includes the

presence of indels (Casals *et al.*, 2012). Moreover, the Olesen *et al.* team noted when investigating length polymorphisms that only 0.7 % of HVR-1 sequences contained some form of insertion when analysing next-generation sequencing contigs. However, this figure jumped to 3 % when the team used Sanger sequencing to determine the E1E2 sequences from patients with HCV (Olesen *et al.*, 2025). Ergo, these reported statistics are likely an underrepresentation of the true prevalence of length polymorphisms in the HVR-1.

Moreover, this issue of underrepresentation may be compounded by the global distribution of HCV and inequitable access to routine testing and sequencing of HCV. For example, genotype 5 is most prevalent in Eastern- and Southern Sub-Saharan Africa, and is considered to be an endemic genotype (Messina *et al.*, 2015). Only 37 genotype 5 sequences have been analysed compared to the 6,130 sequences for genotype 1 which is widely distributed in high-income countries (Paraskevis *et al.*, 2021).

As discussed, it has been shown that length polymorphisms in the HVR-1 may alter disease outcome, and NAb sensitivity (Bianchettin *et al.*, 2007; Olesen *et al.*, 2025). Therefore, having an accurate and global picture of the prevalence of length polymorphisms within the HVR-1 may help in understanding differences in disease progression, and moreover, aid in rational B-cell vaccine design.

3. 2. 2. Universality of Length Polymorphisms in the HVR-1

Having considered the prevalence of length polymorphisms in the HVR-1 sequences across the seven genotypes of HCV, the universality of those length polymorphisms was investigated. Having extracted 9,103 HVR-1 sequences and their corresponding E2 sequences, a phylogenetic analysis of the E2 sequences minus the HVR-1 region was performed. The HVR-1 region was removed to not bias the phylogeny.

The length polymorphism data could then be mapped on to this phylogenetic tree to investigate the universality of length polymorphisms. The inner circle is coloured to represent the genotypes to which the sequences belong. For the most part, the E2 sequences have separated out by genotype. Length polymorphisms are highlighted in blue (insertions) and orange (deletions) around the outer circle of the tree. Length polymorphisms are found throughout this tree with a few clusters.

Most striking is genotype 6 - 50 % of genotype 6 sequences are 26 amino acids in length, and 48.92 % show an insertion of some kind. The genotype 6 insertion E2 sequences are clustered separately from those HVR-1 sequences of 26 amino acids in length. Considering the branch points of the tree, the majority of the 26 amino acid HVR-1 sequences in genotype 6 appear to be a distinct clade of E2 sequences which are an offshoot of the genotype 6 E2 sequences with an HVR-1 of 27+ amino acids. This may suggest that the 26 amino acid HVR-1s found in genotype 6 represent a distinct evolutionary lineage from the other genotype 6 sequences. However, this was not investigated further but would make for an interesting line of inquiry into the history of this genotype of HCV.

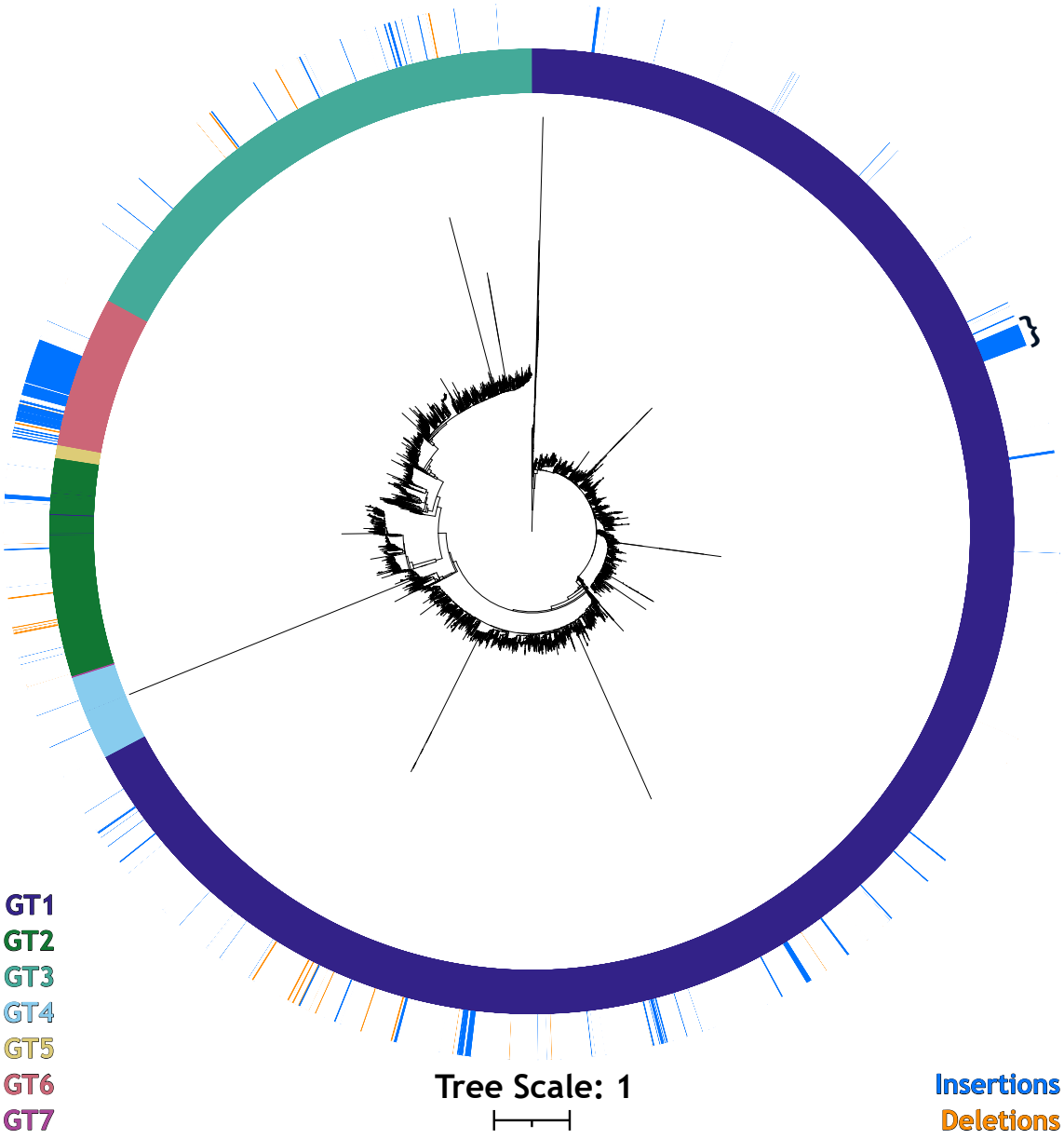


Figure 3. 8. **Investigating Universality of Length Polymorphisms Across All HCV Genotypes.** Scale bar = distance representing 1 substitution per site. Having compared the length polymorphisms across the 9,130 HVR-1 sequences, a phylogenetic tree for all 9,103 corresponding E2 sequences with the HVR-1 removed was produced by IQ-Tree2 using 1,000 ultrafast bootstraps. Annotated in iTOL, the inner circle highlights genotype, and the outer circle shows whether the corresponding HVR-1 sequence featured insertions or deletions compared to the modal HVR-1 length for that genotype. The black bracket next to the cluster of insertions in genotype 1 highlights a cluster of sequences that came from one study in Italy and is discussed further in the text.

One notable finding is the separation of E2 sequences by genotype. This implies that there is, broadly, an evolutionary signature within the E2 protein itself which correlates with genotype. This is, perhaps, surprising as most genotyping efforts for HCV focus on comparing the sequences of more conserved regions of the genome such as the 5'-UTR, core, E1, and NS5B protein sequences (Sandres-Sauné *et al.*, 2003; Murphy *et al.*, 2007; Cai *et al.*, 2013). However, this was not mapped down to any specific region/residues within the E2 protein sequences.

More importantly, length polymorphisms are dispersed throughout the tree, implying that the ability to gain or lose amino acids in the HVR-1 is a universal feature of E2 proteins. However, it should be noted that there are a few clusters of length polymorphism-containing HVR-1 sequences present in genotype 1, and genotype 6. Again, this analysis did not take the source of the sequence or how the sequence was obtained into consideration. These clusters may be a representation of small, localised outbreaks from a single source of infection, or represent repeated sequencing of a single person/small group of people infected with HCV. For example, the largest cluster of insertion sequences in genotype 1 – demarked by a black bracket (Fig. 3. 8.) - all come from one longitudinal study of the evolution of HCV quasispecies in perinatally infected children (Farci *et al.*, 2006). Despite different sources of infection, repeated sequencing of HCV virions from specific individuals can create clusters of length polymorphisms within the phylogenetic tree. Overall, this clustering represents the minority of the representation for length polymorphisms across the phylogenetic tree and length polymorphisms are dispersed throughout the phylogenetic tree. Thus, there is good evidence to suggest that not only are length polymorphisms in the HVR-1 prevalent, with a propensity for insertions in the HVR-1, but they are also universally present across all seven genotypes of HCV.

3. 3. Investigating HVR-1 Composition

Lastly, the 9,103 HVR-1 sequences were aligned using MAFFT to produce sequence logos in RStudio to investigate the conservation of residues across the HVR-1, and to investigate compositional variability. Moreover, the insert sites for the length polymorphisms previously discussed could also be interrogated. Previous studies have considered the conservation of residues across the HVR-1. (Penin *et al.*, 2001). However, this study only considered 1,382 HVR-1 sequences and was performed over two decades ago. Many more HCV sequences have been uploaded to sequence databases since then.

Figure 3. 9. A i is the bits sequence logo for all HVR-1 sequences from across all seven genotypes, figure 3. 9. A ii is the probability sequence logo. Sequence logos based on the bits metric highlight the conservation across the sequence. There are six highly conserved residues across all 9,103 HVR-1: T7 (T385), G12 (G389), G13 (G390), F33 (F403), G44 (G406), and Q47 (Q409). Numbered as per these alignments, and then numbered in brackets as per the alignments in appendix I. Interestingly, previous studies indicated that the conservation of basic residues across the HVR-1 may correlate with increased infectivity (Callens *et al.*, 2005). However, this analysis does not show the same conservation pattern previously reported. This may be a function of the disparity in the number of sequences analysed. Callens *et al.* looked at the sequences of 1,489 non-redundant HVR-1 sequences. This study, however, compared 9,103 HVR-1 sequences.

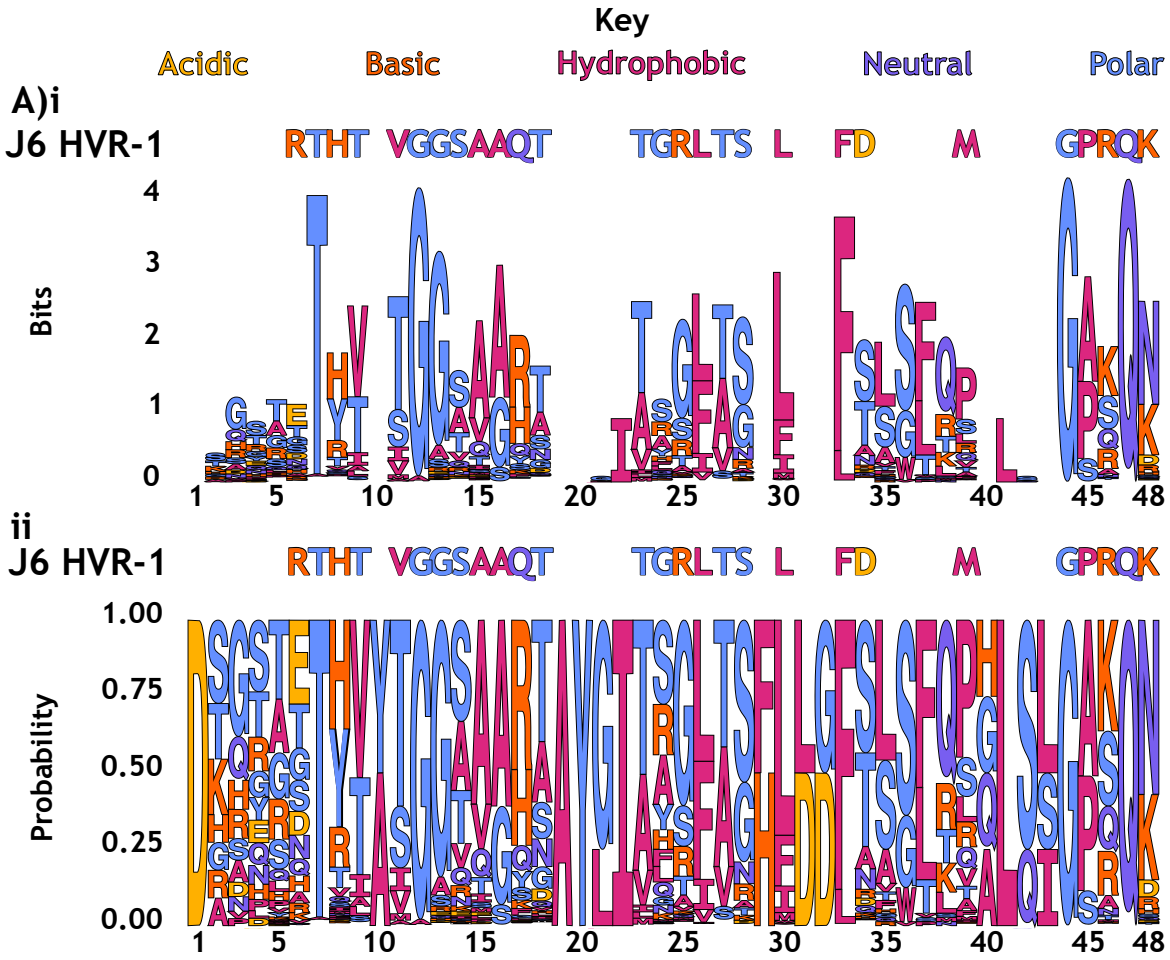


Figure 3. 9. **Sequence Logos for All Extracted HVR-1 Sequences.** A) i and ii sequence logos show the same sequences, the former using bits, and the latter using probability as the y-axis metric measured. For the bits graphs, the taller the letters, the more conserved this amino acid is at this position. Inversely, smaller letters or a lack of letters in a position implies a lack of conservation at this position, or indeed a small proportion of sequences containing an insertion at this position. In the probability graphs, the bigger the letter, the higher the probability of finding this amino acid in this position. However, these graphs are not necessarily informed by level of conservation. For example, the probability graph would suggest that an aspartic acid is present at the N-terminus of the HVR-1 100 % of the time. However, this is only representative of one sequence - the only sequence with residue information at this position. However, the probability graphs do allow for visualisation of which residues are being inserted into the HVR-1 by comparing the positions in the probability graph which are rendered as blank positions in the bits graph. The aligned J6 HVR-1 is rendered above both graphs. All sequence logos were generated from the extracted HVR-1 sequences and then processed in RStudio using the ggseqlogo, ggplot2, and Biostrings packages.

Moreover, as the residues are coloured by their chemical properties, conservation of properties can also be considered. For example, in positions 26, 30, and 33 no particular amino acid is being conserved. However, hydrophobicity is being conserved. This conservation of physicochemical properties of certain positions within the HVR-1 has been noted previously (McAllister *et al.*, 1998; Penin *et al.*, 2001). Additionally, conservative amino acid substitutions are likely indicative of an underlying function in **intrinsically disordered regions** (IDRs) of proteins (Beh, Colwell and Francis, 2012; Schmidt, Barreau and Rohatgi, 2019; Cho *et al.*, 2020). Nonetheless, the exact contributions that each of these specific conserved residues and the conservation of hydrophobic residues makes to HCV infectivity would require further experimentation.

Lastly, in considering where residues are inserted into the HVR-1, two main hotspots for insertions can be seen when comparing the J6 HVR-1 sequence, and the bits sequence logo. As mentioned, the bit sequence logo shows overall conservation of residues at certain positions. Conservation correlates to residue letter height. Empty positions suggest an insertion which is only present in a minority of sequences (Schneider and Stephens, 1990). When comparing the aligned J6 HVR-1 sequence, and the height of residue letters in the insertion sites, the abundance of these insertions across the 9,103 HVR-1 sequences can be deduced. Therefore, two insertion hotspots can be seen – at the very N-terminus (positions 1 – 5), and behind D34 (D404) (positions 35 – 38).

3. 4. Concluding Remarks

It cannot be understated how revolutionary machine learning- and AI-enabled protein structure prediction is for the field of biology as a whole. So much so that the AlphaFold project leads Demis Hassabis and John Jumper have received multiple awards for their contribution to science, culminating in receiving the Nobel Prize for Chemistry in 2024 (The Royal Swedish Academy of Sciences, 2024). Proteins are the functional machines that power life across all its kingdoms, and it has been long understood that the structure of a protein is intrinsically related to its function – known as the structure-function relationship. Resolving the structures for proteins allows for a much greater interrogation of how these proteins function. However, resolving protein structures with classical molecular biology techniques is an incredibly challenging task.

For x-ray crystallography, researchers need to produce a supersaturated solution of a particular protein in a near native-state which is then crystallised. X-rays are fired at this crystal, the diffraction pattern interpreted, and thus the structure can be resolved. There are myriad ways to produce this initial protein crystal – multiple techniques can be used to promote and collect protein crystals, a cocktail of different chemicals can be added to the solution to promote supersaturation/crystallisation, and multiple other properties have to be controlled for such as pH, temperature, and pressure. All this, and even then, some proteins will not crystallise. To top everything off, these variables require empirical testing to find the best possible method for producing a crystal for a specific protein (Kundrot, 2004; McPherson and Gavira, 2014; Timofeev and Samygina, 2023).

Cryo-electron microscopy (cryo-EM) and protein **nuclear magnetic resonance** (NMR) spectroscopy offer alternatives to x-ray crystallography, but these techniques have their own associated problems. For cryo-EM, negative staining can increase background noise and limit resolution. Vitrification, or plunge freezing, can adhere the sample to the air-water interface after spotting the sample onto the grid and this can induce partial sample denaturation (Takizawa

et al., 2017; Lyumkis, 2019). Protein NMR has a relatively low upper limit to the size of molecules it can be used to investigate. It can have poor sensitivity due to issues such as spectral crowding i.e. spectra detected from endogenous NMR probes and not necessarily from the target being interrogated (Kleckner and Foster, 2011).

These issues all require specific optimisation protocols to overcome and may require many iterative rounds of experimentation which can be time consuming and expensive. These difficulties and limitations may aid in explaining the discrepancy in protein sequences available through online databases and the number of protein structures available. As of September 2025, searching for *Hepacivirus hominis* alone on the **National Centre for Biotechnology Information** (NCBI) protein sequence database results in 245,278 results (NCBI, 2025). On the other hand, the PDB has a total 242,573 structures across any organism. In other words, there are more protein sequences that have been uploaded online for a single viral species than there are protein structures uploaded in total. Machine learning- and AI-enabled protein structure prediction will help in bridging this gap as all they require is the amino acid sequence. Considering this project alone, as of September 2025, there are 72 structures uploaded onto the PDB which are labelled as *Hepacivirus hominis* and contain "E2" in the structure description and only one has a resolved HVR-1 in the structure (RCSB Protein Data Bank, 2025). In this project alone, 536 E2 model structures and 500 E1E2 models have been predicted - all of which model the HVR-1. There is now a concerted effort expand the number of structures that are available for researchers to interrogate. Databases of predicted structures have now began to be compiled such as Viro3D (Litvin *et al.*, 2025), the AlphaFold Protein Structure Database (Varadi *et al.*, 2022), and the Big Fantastic Virus Database (Kim *et al.*, 2025) which collate predicted protein structures.

Moreover, the value of protein structure prediction lies not only in the ease of use and its high-throughput nature compared to classical structural biology techniques but in its multiple uses. To date, protein structure prediction has been used in structural biology studies (Fowler and Williamson, 2022; Hu *et al.*, 2022; Hutin *et al.*, 2022), drug discovery (Weng *et al.*, 2022; Zhang *et al.*, 2022;

Ren *et al.*, 2023), protein design studies (Jendrusch, Korbel and Sadiq, 2021; Goverde *et al.*, 2022), protein target prediction (Wang *et al.*, 2022), and protein function prediction (Ma *et al.*, 2022).

The Grove lab itself has leveraged protein structure prediction to investigate the evolution of glycoproteins across the *Flaviviridae*. Firstly, using AlphaFold2, the E1E2 structures of related hepaci-, pegi-, and pestiviruses were interrogated and a cross-genera, conserved E1 protein fold was discovered, suggesting a conserved function for this protein (Oliver *et al.*, 2023). Expanding across the whole of the *Flaviviridae*, the Grove lab and collaborators used ColabFold and phylogenetic analyses to investigate the structure and evolution of glycoproteins across this family of viruses and found two evolutionarily distinct groups within the *Flaviviridae* – those which use class II fusion proteins homologous to the E glycoproteins found in Orthoflaviviruses (such as Dengue virus), and those which use E1E2 glycoproteins. Moreover, through phylogenetic analysis, they found that usage of E1E2 glycoproteins appeared to restrict viruses to infecting vertebrate hosts (Mifsud *et al.*, 2024).

In this project, protein structure prediction is being used to investigate the HVR-1 of the E2 protein of HCV, examining its conformational possibilities, and using these models to generate and validate hypotheses. The HVR-1 is predicted to adopt one of three distinct poses across both E2-alone and E1E2 models, across four different protein folding software packages. Moreover, these HVR-1 conformations may relate to its function. For example, a pocketed HVR-1 may require SR-B1 interaction to permit CD81 engagement of the E2 protein (Lindenbach and Rice, 2013; Augestad *et al.*, 2024). We then postulated that length polymorphisms are modulating the occupancy rates of these distinct conformations through altering or ablating the intramolecular interactions required to stabilise different conformations. For example, deletions within the HVR-1 may render the HVR-1 too short to reach the stem region into which it has been shown to pocket both in the ColabFold predictions, and in the S52 E1E2 structure (PDB: 8RJJ). Length polymorphisms have been a noted feature of the HVR-1 since the early 1990s. Herein is presented possibly the largest, to date, analysis of length polymorphisms across the HVR-1. It was found that length

polymorphisms are, indeed, prevalent, and universal across the seven genotypes of HCV. Moreover, there is a propensity for the HVR-1 to extend which suggests that deletions in the HVR-1 may be particularly deleterious to viral fitness; possibly through preventing certain conformations of the HVR-1. As discussed, shorter HVR-1s may not be able to pocket, and this may lower viral fitness. However, this would need to be assessed experimentally.

Moreover, through the alignment of 9,103 HVR-1 sequences, the two hotspots for insertions within the HVR-1 have been visualised. One hotspot for insertions is seen at the very N-terminus of the HVR-1, and the other is towards the C-terminus of the HVR-1, behind D404. However, the exact impact of deletions in the HVR-1 and of insertions in either of these two hotspots is difficult to determine from the structure predictions alone. Therefore, further experimental work is required and to probe these hypotheses. In the next block of work for this project, HCVpp were generated containing both insertions and deletions at the N-terminus and insertions in the C-terminus. It should be noted here that preliminary work on this alignment of HVR-1 sequences found that the internal insertion site was post-F403 and so this was chosen as the internal insertion site.

Chapter 4

Assessing the Phenotypic Consequences of Length Polymorphisms in the HVR-1

4. Assessing the Phenotypic Consequences of Length Polymorphisms in the HVR-1

In the previous chapter, machine learning- and AI-enabled protein structure prediction allowed possible HVR-1 conformations to be interrogated. Three potential HVR-1 “poses” were discovered – pocketed, “free-floating”, and tucked under the CD81 binding loop. It was posited that the ability, or inability, to adopt these poses may modulate the function of the HVR-1. Then, through counting the lengths of 9,103 HVR-1 sequences and phylogenetic analysis of those sequences, length polymorphisms within the HVR-1 were found to be prevalent and universal. Additionally, there was found to be a propensity for inserting additional amino acids into the HVR-1 rather than deleting residues. Furthermore, insertion hot spots were determined through the alignment of the HVR-1 sequences. Altogether, these preliminary bioinformatic experiments led to two central hypotheses that were to be evaluated through molecular virology experiments: length polymorphisms in the HVR-1 alter the phenotype of HCV; the change in phenotype is caused by modulating the occupancy rate of the conformational poses the HVR-1 adopts.

To assess the first of these two hypotheses, a panel of HCVpp constructs were generated which contained length polymorphisms at the N-terminus of the HVR-1 and at the C-terminal insertion site.

At the N-terminus, the first five residues of the HVR-1 were deleted one residue at a time to investigate the effects of shortening the HVR-1. This produced five N-Terminus Deletion constructs. The first construct had the first amino acid removed, the second the first two amino acids removed etc. Based on the lack of HVR-1 sequences which exhibited deletions, it was posited that this would be deleterious to the HCVpp.

Moreover, up to an additional five glycine residues were added one at a time to the N-terminus of the HVR-1. This produced five N-Terminus Extension constructs. The first with one additional glycine residue, the second with two

additional glycine residues etc. Furthermore, analysis of the HVR-1 sequences found an internal insertion site at the C-terminus of the HVR-1. As noted in the previous chapter, preliminary data suggested that the internal insertion site was behind F404 which is where an additional two, four, or seven glycine residues were added. This produced three Internal Insertion constructs. Glycine was chosen as the amino acid for insertions for a few reasons: firstly, it is small and flexible; second, and more importantly, preliminary analyses of HVR-1 sequences revealed an enrichment of glycine residues in the HVR-1 sequences with insertions.

Altogether, five N-Terminus Deletions, five N-Terminus Extensions, and three Internal Insertion constructs were cloned into both a wildtype J6 backbone, and an I438V A524T J6 backbone. This resulted in a total of twenty-six constructs which were assessed for infectivity, receptor dependency, and antibody sensitivity.

4. 1. HCV Pseudoparticles: Their History and Applications

4. 1. 1. The History and Development of HCV Pseudoparticles

Viral pseudoparticles are virus-like particles which are replication defective. These particles are derived from the core "skeleton" of one virus, expressing the glycoproteins of another virus, containing some reporter gene. These pseudoparticles can be used to investigate viral glycoprotein functions in various contexts such as during viral entry and neutralisation (Rizatdinova, Ershova and Astrakhantseva, 2025). In this project, HIV-1-derived HCVpp were used to investigate length polymorphisms in the HVR-1. HCVpp were generated through the co-transfection of HEK293T CD81 KO cells with three plasmids: one which encoded the HIV-1 *gag-pol* genes (pCMVΔR8.91); a second a luciferase reporter gene (CSLW); and a third encoding for the HCV glycoproteins E1E2 controls and constructs.

The first HIV-1-based system for pseudotyping was developed in 1996 (Naldini *et al.*, 1996). Further improvements to the safety and efficacy of the HIV-1-based pseudotyping came a year later in 1997. This 2nd generation HIV-1 pseudotyping system uses the pCMVΔR8.91 plasmid which encodes for the *gag*, *pol*, *tat*, and *rev* genes from HIV-1. These genes are under a **cytomegalovirus** (CMV) promoter. Additionally, the packaging signal (Ψ) element was removed from this plasmid and was transposed into the plasmid that encodes the reporter gene. It was found that these HIV-1-derived retroviral vector particles could stably transduce non-dividing cells *in vitro* (Zufferey *et al.*, 1997). The *gag* gene encodes for several proteins including the HIV-1 capsid protein, matrix protein, and the nucleoprotein – this forms the core skeleton of the pseudoparticle upon which glycoproteins from different viruses can be expressed. The *pol* gene encodes both the reverse transcriptase and integrase proteins among others which facilitates the reverse transcription of the RNA of the reporter gene back into DNA and its integration into the genome of the target cell. Lastly, the *tat* and *rev* genes encode a transcription activator and RNA export regulator respectively (Seitz, 2016). The luciferase reporter plasmid (CSLW) contains both the HIV-1 packaging signal (Ψ) element and the firefly luciferase gene, along with other regulatory elements. This permits the packaging of the luciferase

gene into the HIV-1 core as coded for by the pCMVΔR8.91 plasmid. This leaves the HCV glycoproteins.

Three research groups published papers in 2003 which described processes by which HCVpp could be generated. Whilst there were variations in the exact process, each method from the three research groups were analogous in their general principles. Overexpression of the HCV E1E2 proteins resulted in a minority of E1E2 complexes being expressed on the cell surface. After assembly of the retroviral core, including packaging of the reporter gene, the retroviral core buds from the cell surface. This incorporates the HCV E1E2 glycoproteins onto the retroviral particle (Bartosch, Dubuisson and Cosset, 2003; Drummer, Maerz and Poumbourios, 2003; Hsu *et al.*, 2003). However, further studies have suggested that assembly of the HCVpp takes place in intracellular compartments which can traffic and fuse with the cell surface, releasing HCVpp into the supernatant (Sandrin *et al.*, 2005). Either way, HCVpp accumulate in the supernatant of producer cells which can be used to investigate the functions of the viral glycoproteins.

HCVpp have been used consistently since the development of the system in 2003. However, following the development of the HCVpp system, a cell culture system for propagation of full virus was developed in 2005. There are some key differences between HCVpp and **HCV cell culture** (HCVcc) viruses.

4. 1. 2. Comparison of HCV Pseudoparticles with HCV Cell Culture Particles and Applications of HCV Pseudoparticles

To compare the HCVpp system to that of the HCVcc system, first the history of the development of the HCVcc system will be discussed. In brief, one of the first steps in the development of the HCVcc system was the development of a subgenomic HCV replicon system in 1999 (Lohmann *et al.*, 1999). Further developments in the subgenomic replicon system came in 2003 with the development of a subgenomic replicon system for HCV genotype 2a. The sequence for this new replicon system (NS3 through NS5B) was cloned from a Japanese person with HCV, suffering from fulminant hepatitis. The full-length cloned sequence (core to NS5B) from this person was termed "**Japanese Fulminant Hepatitis – 1**" or JFH-1 (Kato *et al.*, 2001). In 2003, Kato *et al.* took this sequence and developed a subgenomic replicon system. Critically, they found that this new JFH-1 replicon efficiently replicated without requiring any adaptive mutations (Kato *et al.*, 2003). Further work by three different research groups demonstrated that chimeras consisting of the JFH-1 replication genes (NS3 through NS5B) and the core through to NS2 genes from differing HCV genomes or, indeed, the complete wildtype JFH-1 genome replicated efficiently in Huh7 cells, producing infectious viral particles (Lindenbach *et al.*, 2005; Wakita *et al.*, 2005; Zhong *et al.*, 2005; Gottwein *et al.*, 2007; Scheel *et al.*, 2008).

HCVpp and HCVcc particles do, however, differ in their composition. One critical difference between the two is the association of host-derived components, specifically the association of apolipoproteins with HCVcc particles. It has been shown in several studies that HCVcc particles form associations with apoE, apoB, apoA1, and apoC1 (Meunier, Russell, Engle, *et al.*, 2008; Merz *et al.*, 2011; Catanese *et al.*, 2013). HCVpp, however, do not associate with apolipoproteins (Wasilewski, Ray and Bailey, 2016; Riva and Dubuisson, 2019). However, in general, HCVpp are sufficient for to study E1E2 glycoprotein function, and have been used in myriad studies to investigate HCV entry events and receptor interactions (Bartosch, Dubuisson and Cosset, 2003; McKeating *et al.*, 2004; Lavillette *et al.*, 2005). Additionally, the lack of apolipoproteins found in

association with the HCVpp allows for a more direct assessment of the intrinsic effects of altering the length of the HVR-1, despite being a less physiologically analogous model to native HCV.

Moreover, HCVpp can be used to characterise and investigate neutralising antibodies. In 2016, Wasilewski *et al.* compared the antibody sensitivity of a panel of HCVpp and HCVcc particles expressing analogous E1E2 glycoproteins and found a strong positive correlation between the IC50 values obtained from the HCVpp and HCVcc particles. Moreover, HCVpp have been used in many studies to investigate the neutralisation of HCV (Bartosch, Bukh, *et al.*, 2003; Meunier *et al.*, 2005).

Lastly, HCVpp are replication deficient and can therefore be produced and used in experiments in a **biosafety level two** (BSL2) laboratory. Whereas HCVcc must be contained to **biosafety level three** (BSL3) laboratories. Therefore, on a purely pragmatic basis, this can make HCVpp "easier" to work with for experiments due to the lack of additional infection control measures that must be undertaken in a BSL3 setting.

Altogether, HCVpp are a good model to study viral entry events, receptor dependency, antibody sensitivity, and can be used in a BSL2 laboratory. Therefore, HCVpp were used in this study to investigate changes in the length of the HVR-1 on phenotype.

4. 2. Comparing the Controls

Four different controls have been used throughout this project where and when appropriate: J6 WT E1E2, I438V A524T J6 E1E2, ΔHVR-1 J6 E1E2, and Empty Vector (i.e. pseudoparticles produced without glycoproteins on the HIV-1 core). To provide context for the experimental results, these controls shall first be discussed and compared.

The J6 WT E1E2 sequence was derived from the sequencing of HCV from the blood plasma of a Japanese person living with HCV in 1991 (Okamoto *et al.*, 1991). It has since become a prototypical laboratory strain of HCV from genotype 2a. The core-NS2 gene regions from J6 were combined with the NS3 to NS5B gene regions from JFH-1 during the development of the HCVcc system (Lindenbach *et al.*, 2005). It was found by Lindenbach *et al.* that the J6/JFH-1 chimera was replication competent and that the centrifuged and filtered supernatant from cells transfected with the J6/JFH-1 chimera was able to induce productive infection in naïve cells (Lindenbach *et al.*, 2005). Following this breakthrough in the HCVcc system, J6 started to be used as a prototypical strain in the laboratory.

Further work with the J6/JFH-1 HCVcc molecular clone by in the Grove lab looked at the evolution of adaptations in this clone through putting the J6/JFH-1 into continuous culture for twenty weeks in Huh 7.5 cells and evaluating genome evolution through NGS throughout these twenty weeks. By day 42, two mutations began to be seen in a minority of sequences – I438V and A524T. However, by day 100, these mutations were found in 100 % of sequences (Stejskal *et al.*, 2022). Note that throughout this project, this mutant has been called I438V A524T J6 as per the original publication. However, with reference to the alignment used for renumbering residues throughout this thesis in appendix 1, these mutations would be I438V and A527T. These mutations are found in the front layer and the CD81 binding loop of the E2 protein, respectively.

Interestingly, position 438 has been found in other studies as an important site for mutations which can alter antibody sensitivity. El-Diwany *et al.* found that

mutations in position 438 modulated both HC33.4 and AR4A antibody sensitivity. Specifically, they found that an L438V mutation in a genotype 1a HCVpp expressing patient-derived E1E2 glycoproteins increased sensitivity to both antibodies. Moreover, they found that this mutation reduced SR-B1 binding (El-Diwany *et al.*, 2017). Interestingly, the I438V A524T J6 is also more antibody sensitive and has reduced SR-B1 binding. Moreover, a L438I mutation in a genotype 1b HCVpp also expressing patient-derived glycoproteins has been shown to confer moderate resistance to HC84.26 (Bailey *et al.*, 2015). Lastly, Olesen *et al.* generated HVR-1 swap chimeras in the HCVcc system. They found a F438V mutation in a J4/JFH-1 recombinant HCVcc virus which had its HVR-1 swapped for that of S52. The F438V mutation was found to increase AR3A sensitivity which they posit is due to this mutation increasing the binding of the AR3A to its epitope.

Position 524 (527 as per the alignment in appendix 1) in the E2 protein is located within the CD81 binding loop. Early studies demonstrated that antibodies binding to the CD81 binding loop (residues 523 – 538 as per the alignment in appendix 1) prevents CD81 engagement (Owsianka *et al.*, 2001; Roccasecca *et al.*, 2003). At the time, it was thought that the CD81 binding loop engaged directly with CD81. However, later studies by Kumar *et al.* resolved the structure of a complex consisting of the ectodomain of J6 E2, with the HVR-1 removed, bound to both the **large extracellular loop of tamarin CD81** (tCD81-LEL) and a crystallisation chaperone antibody 2A12. They found that the CD81 binding site was discontinuous and comprised of residues from AS412, the front layer (residues 412 to 445 as per alignment in appendix 1), central residues from the CD81 binding loop (524, 528 to 530), and residues in the back layer (623, 624) (PDB: 7MWX, Kumar *et al.*, 2021). Whilst the binding of CD81 does not span the entirety of the CD81 binding loop, it was found to induce re-arrangement of the unstructured loop into an extended state. Kumar *et al.* posited that this CD81 binding loop re-arrangement may be a necessary step in viral fusion, and they suggest that E2 may be the HCV fusogen. Therefore, any mutations within the CD81 binding loop may modulate phenotype by altering the propensity or ability to extend the CD81 binding loop upon CD81 engagement. Additionally, other studies have examined the interaction between CD81, residues in the CD81

binding loop, and antibody sensitivity. Further work by both Zhao *et al.* and Deng *et al.* demonstrated that residue A524 (527) in H77 (genotype 1a) was important for 1H8 and 1H8-like antibodies to bind to E2 but did not facilitate CD81 binding (Zhao *et al.*, 2014; Deng *et al.*, 2021). Altogether, these previous studies contextualise the importance of both residue 438 and 524 (527) for antibody sensitivity and receptor dependency.

Further work by the Grove lab then characterised this novel I438V A524T J6 mutant and found that it had a distinct phenotype compared to wildtype J6. They found that in the HCVcc system, I438V A524T J6 was more infective than the wildtype J6. Moreover, in both the HCVcc and HCVpp systems, it was discovered that I438V A524T J6 was less dependent on SR-B1 than wildtype J6, yet more antibody sensitive across all the antibodies tested in their work – these changes to the phenotype may have been expected given the previous work on mutations in position 438 and 524 (527) in the E2 protein. Additionally, further work was conducted to interrogate the mechanism through which I438V A524T J6 altered HCVpp/HCVcc phenotype. Stejskal *et al.* posited that this distinct I438V A524T J6 phenotype was caused by the constitutive stabilisation of the HVR-1. **Molecular Dynamics** (MD) simulations were performed and found that the HVR-1 of I438V A524T J6 had a reduced **root mean squared deviation** (RMSD) across the MD than wildtype J6. RMSD is a measure in Ångströms of movement of atoms from their start position over the course of a simulation. Dynamic cross-correlation analysis of MD simulations of both I428V A524T J6 and wildtype J6 found that movement in the HVR-1 was transmitted across the rest of the protein in wildtype J6 but not in I438V A524T J6. Altogether, the work performed by Stejskal *et al.* discovered I438V A524T J6 and characterised this mutant as being more infective, less SR-B1 dependent, and more antibody sensitive than wildtype J6 which was posited to be caused by the stabilisation of the HVR-1.

The third control used in these assays is ΔHVR-1 J6. Removal of the HVR-1 from E2 was first performed in 1997 to perform immunisation studies in BALB/c mice from cDNA (Saito *et al.*, 1997). Further studies found that the removal of the HVR-1 did not prevent viral propagation in chimpanzees transfected intra-

hepatically with ΔHVR-1 H77 RNA transcripts (Forns, Thimme, *et al.*, 2000). Further work by Forns *et al.* started to characterise the ΔHVR-1 H77 phenotype *in vitro*. They found that ΔHVR-1 H77 folded correctly and was able to bind CD81 (Forns, Allander, *et al.*, 2000). Additional studies continued to characterise ΔHVR-1 E2 proteins. Removal of the HVR-1 was found to lower SR-B1 dependency (Prentoe *et al.*, 2014), prevent SR-B1 binding (Scarselli *et al.*, 2002), and increase antibody sensitivity (Bartosch *et al.*, 2005; Z. Keck *et al.*, 2016).

Lastly, the Empty Vector control replaces the typical glycoprotein-encoding plasmid with a plasmid containing no glycoprotein sequence. This produces a HIV-1 core pseudoparticle containing the luciferase gene but with no glycoproteins. This allows for the measurement of stochastic membrane fusion, and it is the negative control which has been used across all the assays described below.

4. 2. 1. Comparing the Controls – The Results

To provide further context on the phenotypes described above, the results obtained by each of the four controls from the initial infectivity, receptor dependency, and neutralisation assays have been extracted and presented alone. Firstly, the infectivity of the four controls has been presented (Fig. 4. 1. A). In comparing the infectivity, it is shown that the J6 WT E1E2 HCVpp and the I438V A524T J6 E1E2 HCVpp are as infective as one another. ΔHVR-1 J6 E1E2 HCVpp are significantly less infective than either of the other two controls. However, all three of these controls are producing log_{10}(RLU) figures higher than the background, as represented by the solid black bar.

Next, the receptor dependency of the four controls is shown across three Huh7 receptor knockout cell lines (Fig. 4. 1. B). Note that the results for the Huh7 CD81 KO, Huh7 CLDN-1 KO, and CD81 OCLN KO are not shown here as these receptor KO cell lines do not support HCVpp infection. The full data demonstrating that these cells lines are not permissive to HCVpp infection are shown in figure 4. 6.

Across figure 4. 1. B i – iii, the bottom graph shows the log_{10}(RLU)s obtained in the Huh7 WT cells in the clear bars and the log_{10}(RLU)s obtained in the various receptor knockout Huh7 cells in the diagonally striped bars. A 2way ANOVA was performed between these log_{10}(RLU) figures to examine statistical differences between the two log_{10}(RLU)s in the different cell lines. In the top graphs, the differences in raw RLU figures are presented as percentages. These percentages were calculated by dividing the RLU obtained in the Huh7 KO cell line by the RLU obtained in the Huh7 WT cells.

In considering the Huh7 LDLR KO cell line, there is no statistically significant difference between the infectivity seen in the Huh7 WT cells and the LDLR KO cells. Previous studies have proposed that LDLR:HCV interaction is mediated by the associated apolipoproteins (Owen *et al.*, 2009; Albecka *et al.*, 2012). Therefore, this lack of difference in infectivity intuitively makes sense

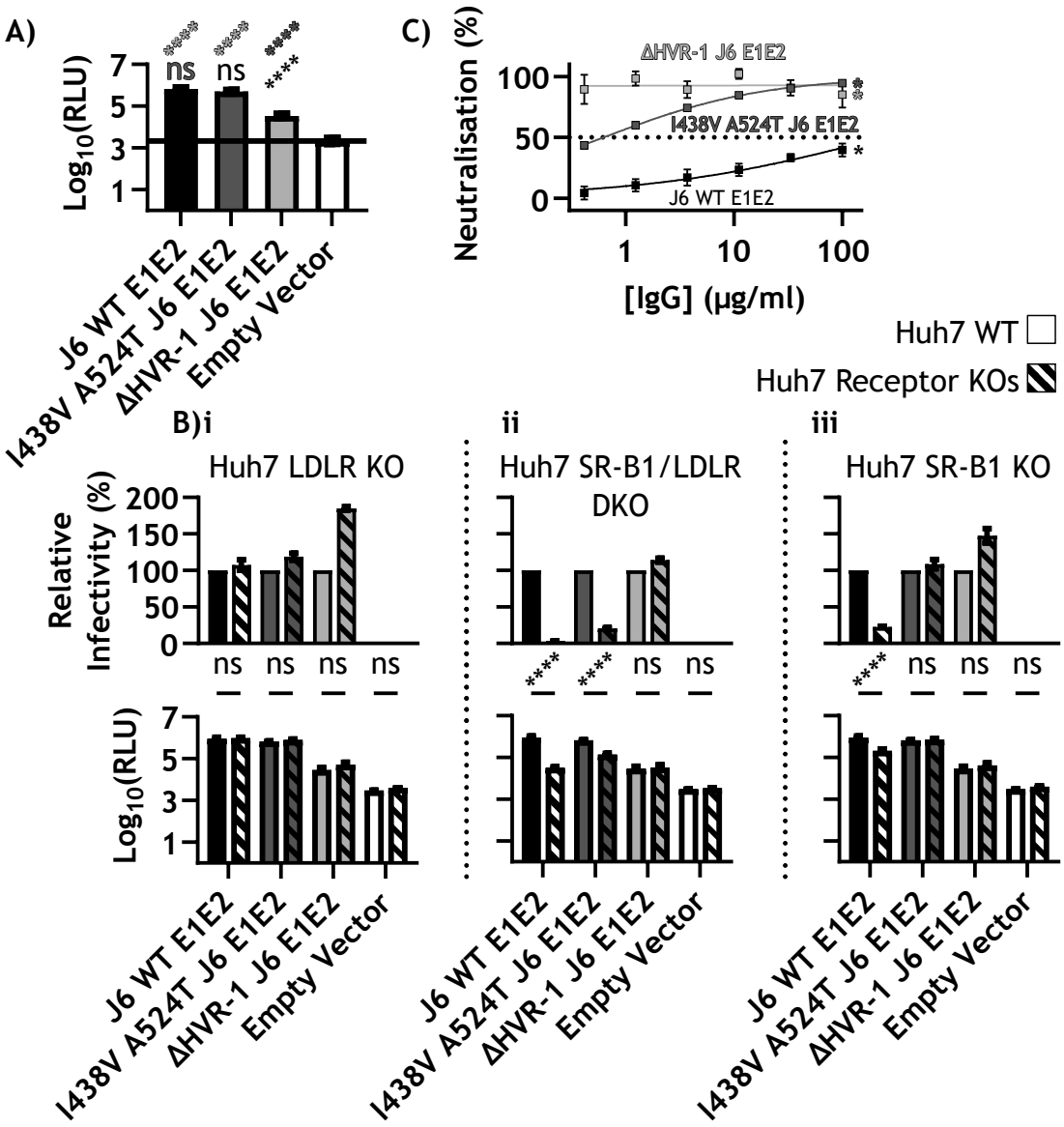


Figure 4. 1. **Comparing Controls.** RLU = Relative Luminescence Units. A) Infectivity across the four controls J6 WT E1E2, I438V A524T J6 E1E2, ΔHVR-1 J6, and Empty Vector. Iterative ordinary one-way ANOVA tests were performed to compare each of the controls to the others i.e. J6 WT was compared to all other controls, then I438V A524T J6 was compared to all other controls. B) Receptor dependency across the controls. Only LDLR KO, SR-B1/LDLR DKO, and SR-B1 KO are shown as the other Huh7 KO cells lines are not permissive to infection. B)i is the percentage relative infectivity; ii the RLUs obtained. The plain bars are the % Relative Infectivity/RLUs obtained by infecting Huh7 WT cells, and the striped bars are the % Relative Infectivity/RLUs obtained from infecting the various Huh7 KO cell lines. a 2way ANOVA test was performed to compare the $\log_{10}$(RLU) obtained in the Huh7 WT cells versus the various receptor KOs. C) Percentage neutralisation with patient-derived polyclonal IgG. Dotted line represents 50% neutralisation. Iterative extra sum-of-squares F tests were performed to ensure the curves were significantly different from one another.

For P-values; ns = $P > 0.05$, * = $P \leq 0.05$, *** = $P \leq 0.001$, **** = $P \leq 0.0001$. All graphs represent three biological replicates with error bars denoting the SEM.

considering the lack of apolipoprotein association with the HCVpp which means that HCVpp are unable to use LDLR as a co-receptor.

When comparing the infectivity of these controls in the Huh7 SR-B1/LDLR DKO cells, the ΔHVR-1 J6 E1E2 HCVpp shows no significant drop in $\log_{10}$(RLU). However, the other two controls do. Specifically, I438V A524T J6 HCVpp do not appear to be dependent on either LDLR or SR-B1 when they are knocked out one at a time. However, I438V A524T J6 HCVpp are significantly less infective (~20 % retained infectivity) in the Huh7 SR-B1/LDLR DKO cells. This is perhaps a surprising result but is discussed in the full context of the length polymorphism-containing HCVpp data. In brief, previous work has demonstrated that there is a synergistic effect in reducing HCVpp infectivity when both SR-B1 and LDLR are knocked out of Huh7.5 cells (Zapatero-Belinchón *et al.*, 2021).

However, there is a difference in the Huh7 SR-B1 KO cells. Where the J6 WT E1E2 HCVpp produce significantly lower $\log_{10}$(RLU) figures in the Huh7 SR-B1 KO cells, there is no difference for both the I438V A524T J6 HCVpp and the ΔHVR-1 J6 E1E2 HCVpp. This recapitulates the phenotypes described for SR-B1 dependency for I438V A524T J6 HCVcc and ΔHVR-1 J6 E1E2 HCVcc in the work by Stejskal *et al.* (2022), which they posit is due to a change in the dynamics between the HVR-1 and the rest of the E2 protein as discussed above.

Notably, across all three assays there is no significant difference in RLUs obtained by the Empty Vector control across any Huh7 KO cell line compared to the Huh7 WT cells.

Lastly, in figure 4. 1. C the results from a polyclonal neutralisation assay performed on the three HCVpp controls displaying glycoproteins are shown. HCVpp were incubated with varying concentrations of patient-derived IgG for one hour prior to overlaying the HCVpp onto Huh7 WT cells. Linear regression was performed in Prism v1.10, and then iterative extra sum-of-squares F-tests were performed between each curve to determine that each curve described a statistically different series of data points. J6 WT HCVpp are highly antibody

resistant, only reaching ~40 % neutralisation even at the highest concentration of IgG. Conversely, I438V A524T J6 HCVpp are antibody sensitive, around ~45 % neutralised even at the lowest concentration of IgG, reaching ~95 % neutralised at the highest IgG concentration. Lastly, the ΔHVR-1 J6 E1E2 HCVpp are the most antibody sensitive across these three controls. At the lowest IgG concentration, the ΔHVR-1 J6 E1E2 HCVpp are around ~95 % neutralised. These phenotypes are, once again, recapitulations of the phenotypes described by Stejskal *et al.* (2022).

Altogether, the phenotypes for each of the controls with regards to infectivity, receptor dependency, and antibody sensitivity can now be described (Table 4. 1.)

Table 4. 15. Summary of the Control Phenotypes. Plus signs (+) have been assigned based on the relative phenotypes of each of the controls in relation to one another as shown in figure 4. 1. Minus signs (-) represent a lack of dependency on specific receptors.

Control	Infectivity	LDLR Dependency	SR-B1 Dependency	SR-B1/LDLR Dependency	Antibody Sensitivity
J6 WT E1E2 HCVpp	+++	-	+++	+++	+
I438V A524T J6 E1E2 HCVpp	+++	-	-	++	+++
ΔHVR-1 J6 E1E2 HCVpp	+	-	-	-	++++

These are the phenotypes to which the length polymorphism-containing HCVpp will be compared.

4. 2. HCVpp Infectivity

4. 2. 1. J6 WT Backbone HCVpp Infectivity

As described, it was hypothesised that altering the length of the HVR-1 may alter the phenotype of HCVpp by modulating the occupancy rate of certain poses adopted by the HVR-1. In this chapter, the first half of this hypothesis was to be examined: do alterations in the length of the HVR-1 alter HCVpp phenotype?

The first assay undertaken to investigate this hypothesis was to evaluate the infectivity of the J6-backbone length polymorphism-containing HCVpp when infecting Huh7 WT cells. J6-backbone length polymorphism-containing HCVpp were produced in HEK 293T CD81 KO cells with luciferase as a reporter gene, overlaid onto Huh7 WT cells, and luciferase activity was measured after 72 hours to give a **relative luminescence unit** (RLU) measurement as a proxy for viral entry. This experiment was key for two reasons. Firstly, it would measure whether these J6-backbone length polymorphism-containing HCVpp were viable/infectious and therefore able to be characterised. Secondly, it would measure any differences in entry efficacy induced by changing the length of the HVR-1.

Crucially, all the HCVpp constructs were viable/infective. Additionally, it was found that altering the length of the HVR-1 in the J6 backbone could modulate HCV entry in Huh7 cells expressing the full complement of known required HCV entry factors (Fig. 4. 2.). Various ANOVA statistical analyses were used to examine the relationships between the infectivity of the length polymorphism construct HCVpp, their groupings and how these relate to the four controls mentioned prior. Strikingly, the HCVpp grouped by length polymorphism type.

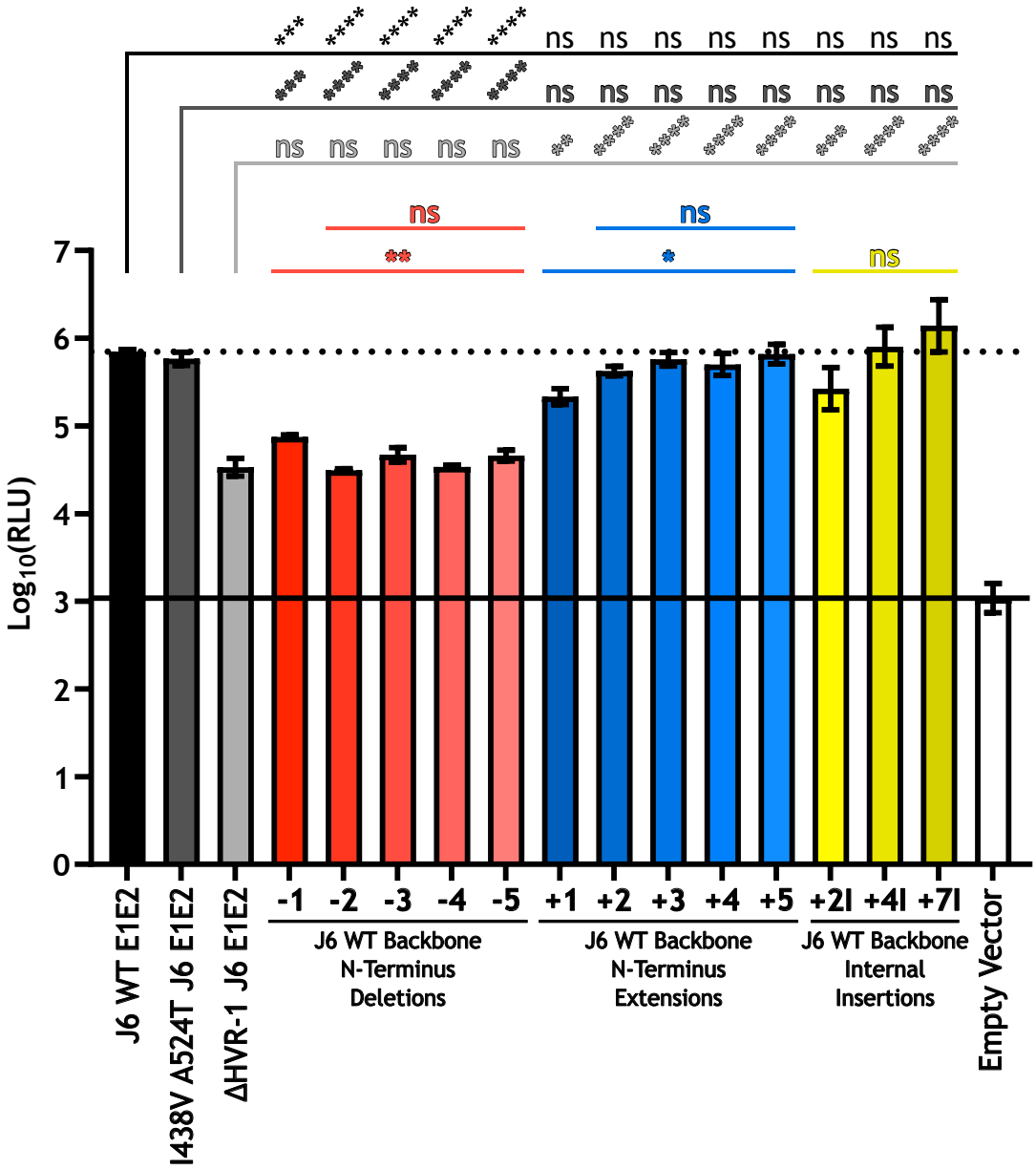


Figure 4. 2. **J6 WT Backbone Length Polymorphism-containing HCVpp Infectivity.** RLU = Relative Luminescence Units. Average $\log_{10}$(RLU) measurements for the HVR-1 length polymorphism constructs in the J6 WT backbone over 3 biological replicates. The dotted line represents the average $\log_{10}$(RLU) given by the J6 WT E1E2 HCVpp. The 'Empty Vector' bar is a measure of $\log_{10}$(RLU) produced by pseudoparticles with no glycoproteins and are, therefore, a measure of the background signal in this assay, demarked with a solid line. Ordinary one-way ANOVA tests followed by Dunnett's multiple comparisons tests were performed with the length polymorphism constructs and one control at a time to avoid bias from the other controls. Moreover, to test if the HCVpp were clustering by length polymorphism type, additional ordinary one-way ANOVA tests were performed for each length polymorphism group. For the N-Terminus Deletions and the N-Terminus Extensions, it was noted that -1 and +1 were outliers in their clusters. Therefore, additional ordinary one-way ANOVAs were performed on subsets of the HCVpp groups excluding these two HCVpp to confirm if the rest of the HCVpp clustered. For P-values; ns = $P > 0.05$, * = $P \leq 0.05$, ** = $P \leq 0.01$, *** = $P \leq 0.001$, **** = $P \leq 0.0001$. Error bars represent the SEM across 3 biological replicates.

In considering the wildtype J6-backbone N-Terminus Deletion HCVpp, statistically, they are significantly less infectious than both the wildtype J6 HCVpp and the I438V A524T J6 HCVpp. Moreover, these N-Terminus Deletion constructs are as infective as ΔHVR-1 J6 HCVpp. These data suggest that removing even one residue from the N-terminus of the HVR-1 is akin to removing the whole HVR-1.

However, when considering the grouping within the N-Terminus Deletion HCVpp group, it has been shown that the -1 N-Terminus Deletion HCVpp is statistically different from the rest of the N-Terminus Deletion HCVpp. Interestingly, it has been demonstrated that the second position threonine residue – present in wildtype J6 – is highly conserved (Section 3. 3., Figure 3. 9.), and it appears that once this threonine residue is removed, the infectivity is statistically the same for the -2 through -5 N-Terminus Deletion HCVpp. This may imply that this highly conserved residue is functionally important.

Overall, the reduction in infectivity caused by removal of residues from the N-terminus of the HVR-1 may explain the lack of deletions found from the analysis of 9,103 HVR-1 sequences. It was posited that deletions in the HVR-1 may be fewer in number than insertions because deletions may be deleterious to viral fitness, and these data support that hypothesis.

In examining the N-Terminus Deletions data in the context of the results of the E2/E1E2 protein structure prediction work, this may be due to the inability of a shortened HVR-1 to be stabilised through intramolecular interactions in one of the HVR-1 conformations. For example, a shortened HVR-1 may not be able to reach the HVR-1 pocket and form stable interactions.

Conversely, addition of glycine residues to either the N-terminus or insertion of glycine residues after F404 (Internal Insertions) does not significantly modulate HCVpp infectivity when compared to the infectivity of both the J6 WT E1E2 HCVpp and I438V A524T J6 E1E2 HCVpp controls.

This intuitively makes sense to explain the prevalence of insertions in the HVR-1 found in Chapter 3 (Section 3. 2.). Further ordinary one-way ANOVA analyses revealed that the +2 through +5 N-Terminus Extension HCVpp formed one

statistically similar grouping. However, the +1 N-Terminus Extension construct did not fit into that group. This, whilst curious, would require further investigation as to why the +1 HCVpp were less infectious than the rest of the N-Terminus Extension group.

Analysis of the length polymorphisms found across the 9,103 HVR-1 sequences did not suggest that the addition of one single residue to the HVR-1 would be deleterious i.e. +1 residue was the most common length polymorphism found across all genotypes barring genotypes 5 and 7 which contained sequences with no insertions and no length polymorphisms, respectively (Appendix III). However, this may reflect the low number of HVR-1 sequences for both genotypes.

Moreover, the Internal Insertion HCVpp group were all, statistically, as infectious as both the wildtype J6 HCVpp and the I438V A524T J6 HCVpp and each other. Once again, in the context of the predicted structures in Chapter 3 (Section 3. 1.), this lack of significant change to infectivity may also be explained by the fact that the HVR-1 segments in the wildtype J6-backbone Internal Insertion HCVpp could all, theoretically, still participate in the intramolecular interactions predicted between the HVR-1 and the rest of the E2 protein.

4. 2. 2. I438V A524T J6 Backbone HCVpp Infectivity

Next, the same length polymorphisms were cloned into the I438V A524T J6 backbone. As in the previous experiment, HCVpp were produced, overlaid onto Huh7 WT cells, and 72 hours later an RLU measurement was taken. Again, this experiment was performed first to ascertain whether the I438V A524T J6-backbone length polymorphism-containing HCVpp were viable/infective and if altering the length of the HVR-1 would result in changes in infectivity in a different HCVpp backbone. Additionally, this experiment would also investigate the interplay of the length polymorphisms and the I438V A524T mutations.

Once more, the length polymorphisms in the I438V A524T J6-backbone length polymorphism-containing HCVpp were found to modulate HCV entry in Huh7 WT cells (Fig. 4. 3.). Similar to the J6-backbone length polymorphism-containing HCVpp, the I438V A524T J6-backbone N-Terminus Deletions exhibited reduced infectivity when compared to the J6 WT E1E2 HCVpp. However, the -1 N-Terminus Deletion I438V A524T HCVpp was as infectious, statistically, as the I438V A524T J6 E1E2 HCVpp, with -2 through -5 showing comparatively reduced infectivity. Additionally, the N-Terminus Deletions were again found in the I438V A524T backbone to be, statistically, as infective as the ΔHVR-1 J6 E1E2 HCVpp.

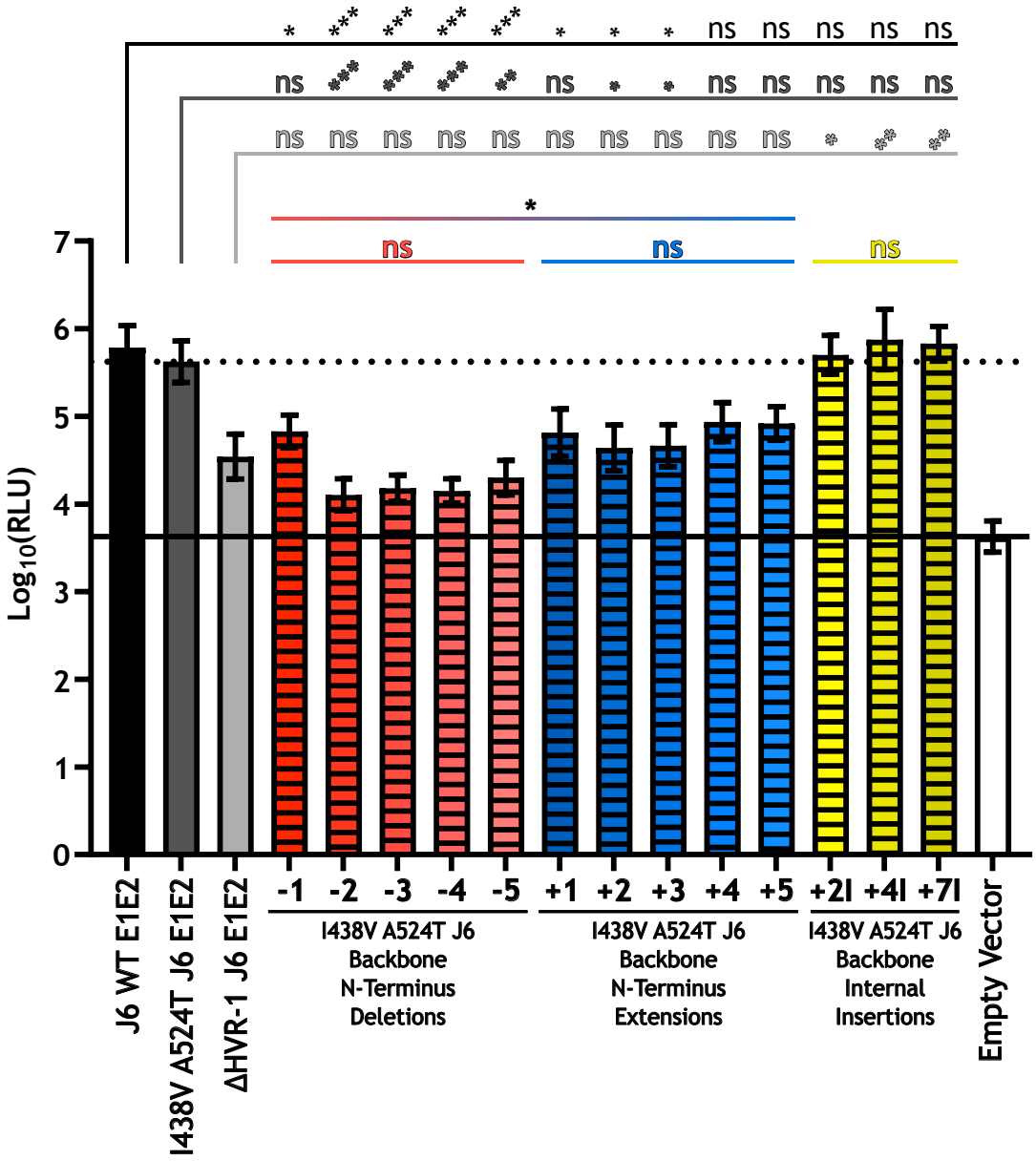


Figure 4. 3. **I438V A524T J6 Backbone Length Polymorphism-containing HCVpp Infectivity.** RLU = Relative Luminescence Units. Average log_{10}(RLU) measurements for the HVR-1 length polymorphism constructs in the J6 WT backbone over 3 biological replicates. The dotted line represents the average log_{10}(RLU) given by the I438V A524T J6 WT E1E2 HCVpp. The 'Empty Vector' bar is a measure of log_{10}(RLU) produced by pseudoparticles with no glycoproteins and are, therefore, a measure of the background signal in this assay, demarked with a solid line. Ordinary one-way ANOVA tests followed by Dunnett's multiple comparisons tests were performed with the length polymorphism constructs and one control at a time to avoid bias from the other controls. Moreover, to test if the HCVpp were clustering by length polymorphism type, additional ordinary one-way ANOVA tests were performed for each length polymorphism group. For the N-Terminus Deletions and the N-Terminus Extensions, it was noted they were much closer in log10(RLU) figures. Therefore, additional ordinary one-way ANOVA was performed on both of these HCVpp groups to confirm if the all of the N-terminus length polymorphism HCVpp clustered. For P-values; ns = $P > 0.05$, * = $P \leq 0.05$, ** = $P \leq 0.01$, *** = $P \leq 0.001$, **** = $P \leq 0.0001$. Error bars represent the SEM across 3 biological replicates.

Altogether, the I438V A524T J6-backbone N-Terminus Deletion HCVpp formed one statistical grouping. However, as noted, the I438V A524T J6-backbone -1 N-Terminus Deletion HCVpp did produce a higher log_{10}(RLU), and was statistically the same as the I438V A524T J6 HCVpp control in this assay. This, again, may be due to the presence of the conserved threonine residue but this would require further investigation.

Notably, the log_{10}(RLU) figures for the other I438V A524T J6-backbone N-Terminus Deletion HCVpp were close to the noise i.e. the log_{10}(RLU) figure for the Empty Vector, demarked by the solid line in the figure. This may suggest that the I438V A524T J6-backbone N-Terminus Deletion HCVpp were not as infective as their wildtype J6 backbone counterparts, or less stable. Therefore, in general, the characterisation for this group of HCVpp should be interpreted cautiously.

Additionally, the N-Terminus Extensions also exhibited a marked reduction in infectivity in the I438V A524T J6-backbone. This differed from the J6-backbone HCVpp wherein the N-Terminus Extensions were as infective as the J6 WT E1E2 HCVpp and the I438V A524T J6 E1E2 HCVpp. More specifically, considering the statistical analysis, none of the log_{10}(RLU) figures obtained by the N-Terminus Extensions are significantly different from the ΔHVR-1 J6 E1E2 RLU figures. Moreover, the I438V A524T J6-backbone +1, +4, and +5 N-Terminus Extension HCVpp gave RLUs which were not significantly different from those obtained by I438V A524T J6 E1E2, and +4 and +5 were not significantly different from J6 WT E1E2.

Altogether, the I438V A524T J6-backbone N-Terminus Extension HCVpp showed an overall reduction in infectivity. They appeared much closer to the I438V A524T J6-backbone N-Terminus Deletion HCVpp with regards to infectivity. To quantify this, an ordinary one-way ANOVA was performed across all the N-Terminus length polymorphism-containing HCVpp. There was a significant difference across this N-Terminus grouping, with a p-value of ≤0.05 (*). However, in the J6-backbone, the p-value was ≤0.0001 (****, Appendix V).

Altogether, this suggests that the N-Terminus length polymorphism constructs in the I438V A524T backbone are more similar to one another than in the wildtype J6 backbone.

Conversely, the Internal Insertions remain as infective as both the J6 WT E1E2 HCVpp and the I438V A524T J6 E1E2 HCVpp. However, there is a slight reduction in the infectivity of the +7 Internal Insertion HCVpp relative to +4 Internal Insertion in this assay and compared to the +7 Internal Insertion in the J6-backbone.

In the J6 WT backbone infectivity assay, it was found that, for the most part, the HCVpp with the same type of length polymorphisms grouped together i.e. the N-Terminus Deletions were as infective as one another, and the N-Terminus Extensions and Internal Insertions were as infective as one another when compared to the controls. Here, the HCVpp phenotypes group by position of the length polymorphisms within the HVR-1. HCVpp with length polymorphisms at the N-terminus in the I438V A524T J6-backbone are more alike than the Internal Insertions which broadly retain their infectivity in either backbone.

4. 2. 3. Comparing Infectivity Across Backbones

To explore whether the I438V A524T mutations were affecting those constructs with length polymorphisms at the N-terminus differently to the Internal Insertions, the log_{10}(RLU) figures obtained by the HCVpp in both backbones were compared. To facilitate comparison, the log_{10}(RLU) figures obtained by the Empty Vector HCVpp in each assay was averaged and subtracted from the other log_{10}(RLU) figures obtained by the other HCVpp in that assay. Next, a 2way ANOVA was performed to compare between the two assays (Fig. 4. 4.).

When considering the N-Terminus Deletions, there is a statistically significant reduction in infectivity in the I438V A524T J6-backbone compared to the wildtype J6-backbone. Across all the N-Terminus Extensions, there is a significant reduction in infectivity in the I438V A524T J6-backbone. However, the Internal Insertion constructs are statistically as infective in either backbone. Taken together, these data suggest that the effects of the I438V A524T mutations are more pronounced in the N-terminus length polymorphism constructs than on the Internal Insertion constructs.

These data would suggest that there is a hierarchy of mutations which impact the HCVpp phenotype wherein the Internal Insertions are not affected by the I438V A524T mutations, but the N-terminus length polymorphisms are. Therefore, it may be concluded that the Internal Insertion mutations have a greater impact on phenotype than the I438V A524T mutations which, in turn, have a greater impact than alterations to the length of the HVR-1 at the N-terminus.

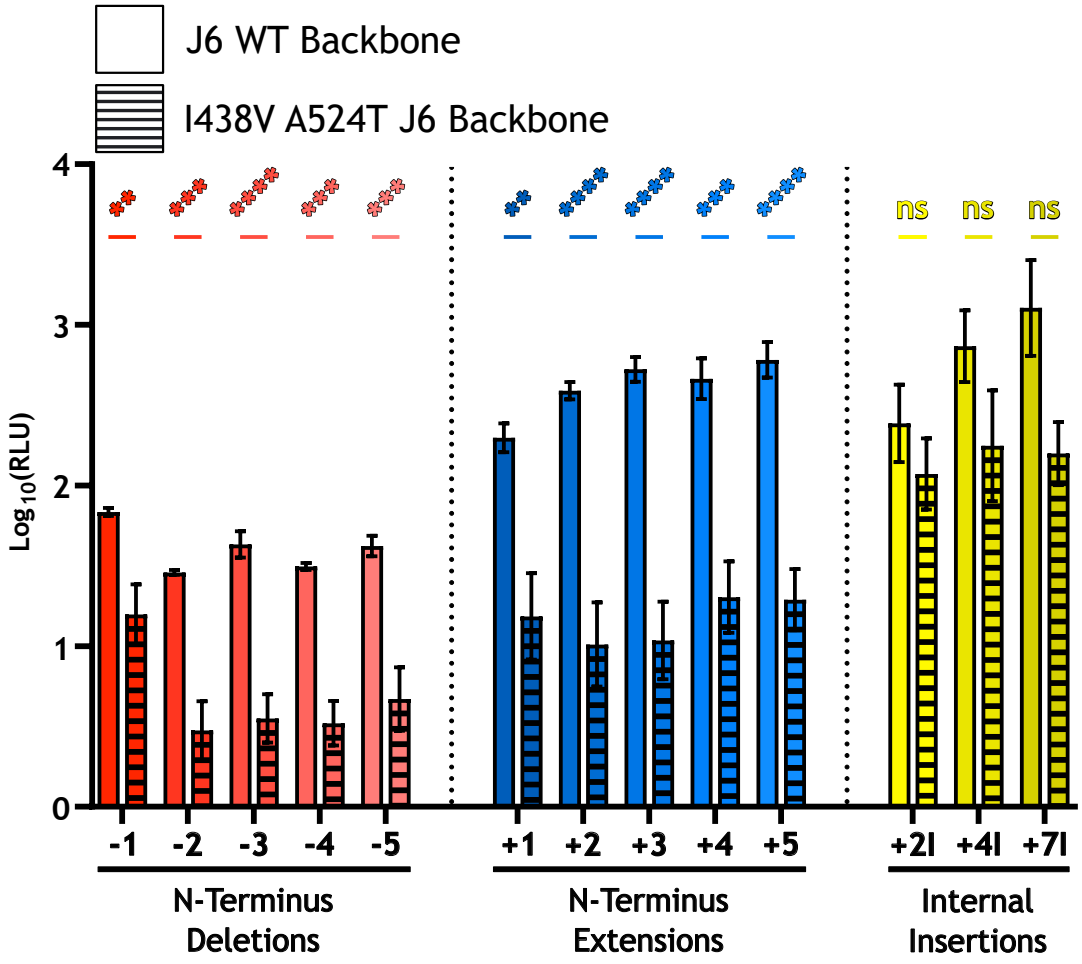


Figure 4. 4. **Comparing Infectivity Across J6 WT- and I438V A524T J6-based HCVpp.** RLU = Relative Luminescence Units. For the length polymorphism-containing HCVpp, the un-striped bars are those length polymorphisms in the J6 WT E1E2 backbone, the striped bars are in the I438V A524T J6 E1E2 backbone. To facilitate comparison between the two mutant groups, the average Empty Vector $\log_{10}$(RLU) was subtracted from the $\log_{10}$(RLU) figures obtained for the length polymorphism-containing HCVpp. Next, given the clustering of the HCVpp found in the prior statistical analyses, 2way ANOVA tests were performed followed by Šídák's multiple comparisons tests on each length polymorphism group separately i.e. the N-Terminus Deletions were tested together, then the N-Terminus Extensions etc. to measure differences in infectivity in mutants with the same length polymorphism mutation in the two different backbones. For P-values; ns = $P > 0.05$, ** = $P \leq 0.01$, *** = $P \leq 0.001$, **** = $P \leq 0.0001$. Error bars represent the SEM across 3 biological replicates.

4. 3. E2 Western Blots

Before continuing to investigate receptor dependency and neutralisation, the HEK 293T CD81 KO used to produce the J6-backbone length polymorphism-containing HCVpp were Western blotted to quantify E2 production (Fig 4. 5.). This was to ensure that differences in infectivity were not a result of differences in production. Currently, there are no commercially available antibodies that detect wildtype J6 E1 via Western blotting and therefore its presence could not be validated.

E2 was detected for each of the J6-backbone length polymorphism-containing HCVpp producer cell lysates through Western blotting. The uncropped blots are included in appendix VI. These are representative blots of one round of J6-backbone length polymorphism-containing HCVpp production.

Across each blot, E2 expression levels in the producer cell lysates appear to be internally consistent. Moreover, the actin loading control is also consistent across each of the three blots. This aids in ensuring that differences in measured infectivity were not a result of differences in HCVpp production efficiency. Additionally, for all Western blots shown, the Empty Vector negative control lane remains empty for E2 and positive for ß-actin.

Western blots were attempted for the I438V A524T-backbone length polymorphism-containing HCVpp producer cell lysates. However, perhaps due to sample degradation, E2 was not detected from the producer cell lysates used to produce the I438V A524T J6-backbone HCVpp used in the above and following assays. Due to time constraints, these Western blots were attempted again by Sarah Little. However, E2 was not consistently detected for each of the length polymorphism constructs. However, the Western blots that Sarah Little produced can be found in appendix VI.

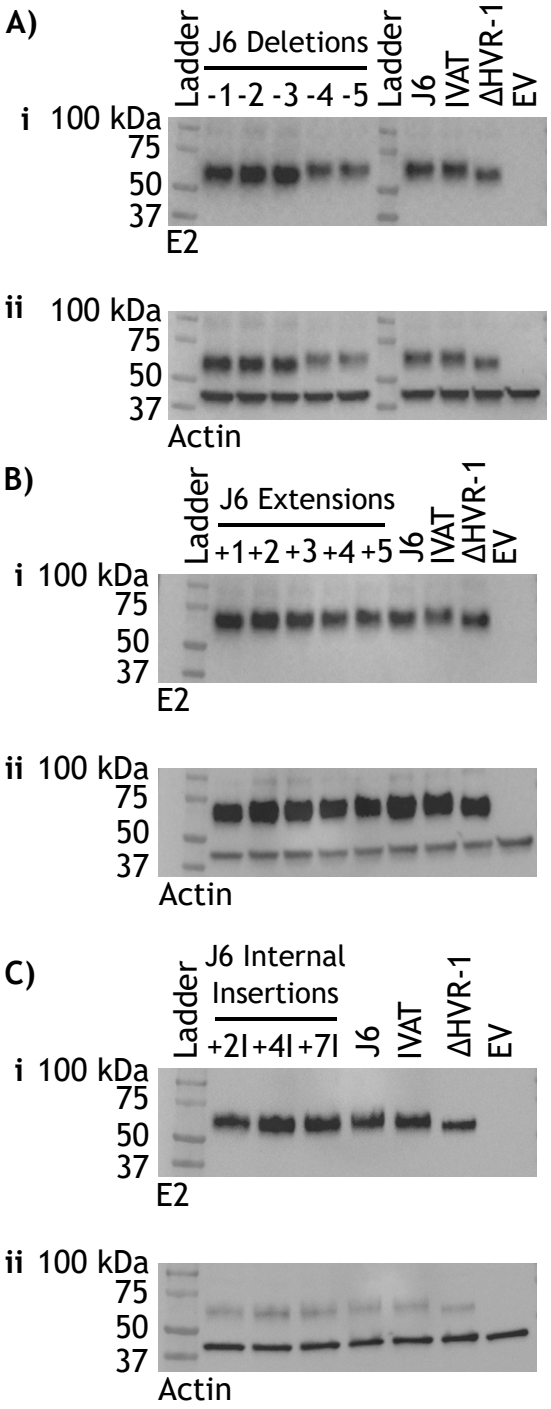


Figure 4. 5. **E2 Western Blots.** E2 molecular weight: ~60 kDa, actin molecular weight: 42 kDa. Lysates of the HEK 293T CD81 KO cells used to produce the HCVpp used in assays were Western blotted to quanitfy E2 production to ensure differences in phenotype were not caused by differences in expression. Actin was used as a loading control, and blotted for on the same nitrocellulose membrane as the E2 blots.
A) J6 N-Terminus deletions. B) J6 N-Terminus extensions. C) J6 Internal Insertions.
J6 = J6 WT E1E2
IVAT = I438V A524T J6 E1E2
ΔHVR-1 = ΔHVR-1 J6 E1E2
EV = Empty Vector

4. 4. HCVpp Receptor Dependency

After discovering that altering HVR-1 length altered infectivity, it was hypothesised that differences in infectivity may be caused by a difference in receptor usage. Therefore, it was then investigated whether changes in the length of the HVR-1 would alter receptor dependency.

In brief, HCVpp in both the wildtype J6 and the I438V A524T backbones were produced as before and overlaid on to Huh7 cells with various key HCV receptors knocked out. The KO cell lines used were Huh7 CD81 KOs, Huh7 CLDN-1 KOs, Huh7 LDLR KOs, Huh7 OCLN KOs, Huh7 SR-B1 KOs, and Huh7 SR-B1/LDLR DKOs. Along with infecting these cell lines, Huh7 WT cells were also infected and used to compare to the infectivity in the various KOs. These receptor KO cell lines were chosen because each of those receptors has been implicated as being involved in the entry of HCV previously. In brief, there are myriad studies on the interactions between CD81, SR-B1, and the HCV E1E2 proteins. However, first the CLDN-1, OCLN, and LDLR receptors shall be discussed as the current literature either does not support a direct interaction between the receptor and the E1E2 proteins (CLDN-1), research is still ongoing to determine how the receptor functions (OCLN), or the receptor is unable to bind to the HCVpp used in these experiments due to the lack of associated apolipoproteins (LDLR).

CLDN-1 was first discovered as co-receptor for HCV entry in 2007 by Evans *et al.* However, further mechanistic studies found that the EL1 of CLDN-1 interacts with CD81 rather than E1 or E2 and forms the E2:CD81:CLDN-1 complex. This permits clathrin-mediated endocytosis in turn leading to viral entry (Evans *et al.*, 2007; Mee *et al.*, 2009; Harris *et al.*, 2010; Krieger *et al.*, 2010; Farquhar *et al.*, 2012). Therefore, because CLDN-1 does not interact directly with either E1 or E2, it was hypothesised that differences in the length of the HVR-1 would not alter the usage of CLDN-1. Moreover, since its discovery as a co-receptor, it has been shown that CLDN-1 is necessary for HCV entry. Therefore, it was also hypothesised that knocking out CLDN-1 would render the Huh7 non-permissive to HCVpp infection (Evans *et al.*, 2007).

OCLN, first discovered in 2009, is another essential co-receptor in HCV entry (Ploss *et al.*, 2009). OCLN is known to function later in the entry events of HCV (Liu *et al.*, 2009; Sourisseau *et al.*, 2013). Additionally, it has been shown that HCV particles recruit OCLN outside of the tight junction, and does so independently of CLDN-1 expression (Deffieu *et al.*, 2022). However, more work is required to understand the exact mechanisms by which OCLN contributes to HCV entry. It was hypothesised that because OCLN is an essential HCV receptor that knocking this receptor out would also render the Huh7 cells non-permissive to HCV infection. Additionally, the lack of mechanistic understanding as to how OCLN functions in HCV entry made it difficult to entirely hypothesise how alterations in the length of the HVR-1 might alter usage of this receptor. However, this will be returned to as there is one surprising result found in these experiments that may, through future work, tease apart how OCLN contributes to HCV entry.

Lastly, for these three receptors, LDLR was discovered in 1999 as a potential co-receptor for HCV entry, LDLR has not been shown to interact directly with the E1E2 proteins – much like CLDN-1. Rather, LDLR is thought to bind to HCV LVPs by binding on to the apolipoproteins which surround HCV LVPs, in particular apoE. (Monazahian *et al.*, 1999; Germi *et al.*, 2002). Additionally, further work in 2012 actually examined the differences in LDLR usage by HCVcc and HCVpp and found that HCVpp are unaffected by the presence of LDLR, and they concluded that this was due to the fact that LDLR and HCVpp cannot interact with one another (Albecka *et al.*, 2012). Therefore, it was hypothesised that there would be no difference in infectivity for any of the HCVpp in either Huh7 WT cells or Huh7 LDLR KO cells as the HCVpp would not be able to use LDLR as a co-receptor.

In summary, it was not expected that the receptor dependency experiments done in the cell lines wherein the above three receptors were knocked out would be particularly informative for the outlined reasons, however, they were undertaken alongside the experiments examining CD81 and SR-B1 usage for completeness.

In brief, CD81 is an essential HCV receptor. It was discovered in 1998 by Pileri *et al.* Further studies discovered that HCV binds to the LEL of CD81 (Pileri *et al.*, 1998; Petracca *et al.*, 2000; Bertaux and Dragic, 2006). Again, this receptor is essential for HCV entry and, therefore, it was hypothesised that knocking it out would render Huh7 cell non-permissive for HCVpp infection.

However, there are several studies which, taken together, point to the HVR-1 being important for engagement of CD81. The first hint that the HVR-1 can alter CD81 binding comes from Bankwitz *et al.* who in 2010 found that the deletion of HVR-1 did not significantly reduce the overall dependence of HCV on CD81 for entry. However, deletion of the HVR-1 increased the sensitivity of the virus to neutralisation by soluble CD81 LEL. Moreover, soluble CD81 also precipitated the ΔHVR-1 particles more efficiently than wildtype particles. Their interpretation of these data was that the HVR-1 was physically masking or occluding the CD81 binding site on E2 and when HVR-1 is removed, that binding site becomes constitutively exposed (Bankwitz *et al.*, 2010).

However, more recent studies put these results in a slightly different context. Kumar *et al.* released two structures of the E2 protein in 2021 and 2023, respectively. Their first structure was that of the ectodomain of the E2 protein with the HVR-1 deleted binding to tamarin CD81 (ΔHVR1 eE2:tCD81-LEL) (PDB: 7MWX, Kumar *et al.*, 2021). This structure revealed two key conformational changes. Upon CD81 binding, residues 418–422 of E2 become ordered and change their conformation. This displacement of residues 418-422 allows the CD81-binding loop to extend outward by 13–15 Å towards the host membrane. They suggested that E2 may even be the fusogen for HCV based on this finding. Their next paper examined another E2 structure which resolved the core and stem regions of E2 in complex with the 2A12 Fab (PDB: 8DK6, Kumar *et al.*, 2023). They compared this structure to many other published E2 structures and found that the CD81 binding loop is highly flexible and can adopt multiple conformations. They demonstrated that the CD81 binding loop exists in at least three distinct conformations which they termed retracted, bent, and extended (PDB: 4WEB, Khan *et al.*, 2014; PDB: 5TCX Zimmerman *et al.*, 2016; PDB: 7MWW & 7MWX, Kumar *et al.*, 2021).

All this came together with the release of the dimer of dimers E1E2 complex structure which shows a pocketed HVR-1 which induces a CD81 binding loop conformation that disfavours CD81 engagement (PDB: 8RJJ, Augestad *et al.*, 2024). In this structure, the CD81 binding loop is in the bent conformation described previously, oriented in the opposite direction relative to where CD81 binds. Augestad *et al.* suggest this may be a strategy to prevent non-productive membrane interactions before the virus reaches the appropriate entry compartment for productive viral entry. Moreover, they conclude that the HVR-1 is the key regulator locking the E1E2 complex in this CD81-resistant state. The pocketed HVR-1 means that the AS412 region is found in a hairpin conformation which in turn keeps the CD81 binding loop in this bent conformation, disfavouring CD81 engagement.

Lastly, SR-B1, which, whilst not an essential receptor for HCVpp entry (Kalemera *et al.*, 2019a; Zapatero-Belinchón *et al.*, 2021) plays an important role in HCV viral entry. Knocking out SR-B1 should leave Huh7 cells permissive for HCVpp entry, facilitating an interrogation into the effects of length polymorphisms in the HVR-1 on SR-B1 dependency. SR-B1 participates in lipid homeostasis and binds high- and very low-density lipoproteins. Additionally, it promotes the uptake of cholesterol by hepatocytes (Acton *et al.*, 1996; Rhainds and Brissette, 2004).

SR-B1 was first discovered as an HCV receptor 2002, more specifically, it was found that recombinant ΔHVR-1 E2 proteins were unable to bind HepG2 cells expressing SR-B1. This led to the hypothesis that the HVR-1 was interacting with SR-B1 (Scarselli *et al.*, 2002). Further studies have also suggested that SR-B1 may promote viral entry by transferring the lipids associated with the HCV LVP away, permitting interactions between SR-B1 and HVR-1 which facilitate E2:CD81 interactions in turn (Bankwitz *et al.*, 2010; Thi *et al.*, 2012; Zahid *et al.*, 2013). However, the role of lipid transfer away from HCV LVP and how that aids in establishment of the SR-B1:HVR-1 interaction is poorly understood.

Given that SR-B1 is non-essential for HCVpp entry, and that there is a putative interaction between the HVR-1 and SR-B1, we hypothesised that length polymorphisms in the HVR-1 may alter the dependency of the HCVpp on SR-B1.

4. 4. 1. Non-Permissive KO Cell Lines

As mentioned previously, some receptors are essential for HCVpp entry into Huh7 cells. Knocking these receptors out renders Huh7 non-permissive to infection by HCVpp. Therefore, these non-permissive cell lines become uninformative in characterising the effects of altering the length of the HVR-1 on infectivity. However, for completeness, the results of these assays are shown in figure 4. 6.

In brief, both figure 4. 6. A) and B) are identical but for the wildtype J6-backbone length polymorphism-containing HCVpp, and the I438V A524T J6-backbone HCVpp, respectively. Additionally, the results for each of the cell lines is presented in the same order: i - Huh7 CD81 KO, ii - Huh7 CLDN-1 KO, and iii - Huh7 OCLN KO cells, respectively. A one-way ANOVA test followed by a Dunnett's Multiple comparison test was performed to compare the infectivity of each HCVpp in each cell line to the infectivity of the Empty Vector HCVpp - the negative control; its average $\log_{10}$(RLU) figure is also demarked by a solid line on the graphs in figure 4. 6.

In examining the $\log_{10}$(RLU) values for the HCVpp evaluated, it can be seen that none of the HCVpp in the Huh7 CD81 KO cells, nor the Huh7 CLDN-1 KO cells are infective. Ergo, for all of the HCVpp constructs and controls tested both CD81 and CLDN-1 are essential receptors required for the entry of HCVpp into Huh7 cells.

However, this might not be the case for OCLN. As noted, there is a lack of understanding as to how OCLN contributes to HCV entry. For all of the controls and the length polymorphism-containing HCVpp constructs in the wildtype J6 backbone, OCLN was found to be an essential receptor i.e. knocking OCLN out of the Huh7 rendered these cells non-permissive to those HCVpp.

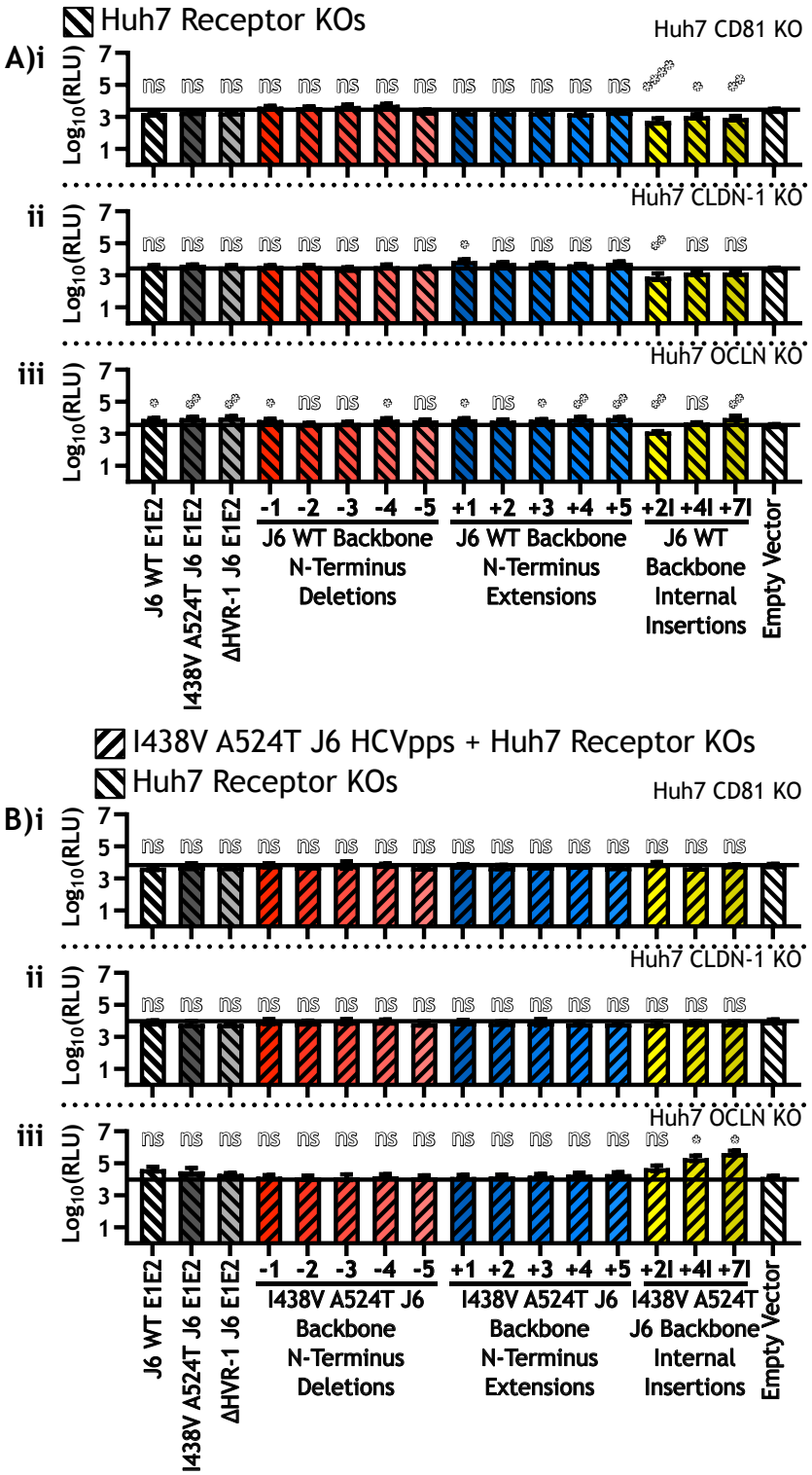


Figure 4. 6. **Infectivity of the HCVpp in Non-permissive Huh7 KO Cell Lines.** A) i - iii, log_{10}(RLU) values for the controls and length polymorphism-containing HCVpp in the wildtype J6 backbone in the Huh7 CD81 KO, Huh7 CLDN-1 KO, and Huh7 OCLN KO cell lines, respectively. B) i - iii, as in A) i - iii but with the I438V A524T J6 backbone HCVpp and controls. Ordinary one-way ANOVA statistical tests were performed to compare the log_{10}(RLU) values for each HCVpp to the Empty Vector control infecting the same cell line.
P-values; ns = $P > 0.05$, * = $P \leq 0.05$, ** = $P \leq 0.01$, **** = $P \leq 0.0001$. Error bars represent the SEM across 3 biological replicates.

However, curiously, it would appear that the I438V A524T J6-backbone Internal Insertion HCVpp are not OCLN dependent and were still able to infect the Huh7 OCLN KO cells – consider the fold changes in infectivity for each of these constructs compared to Empty Vector (Table 2.).

At 3.89-fold more infective than Empty Vector, the I438V A524T J6-backbone +2I Internal Insertion HCVpp might not be considered robustly infective in this cell line. However, the I438V A524T J6-backbone +7I Internal Insertion HCVpp produced an RLU signal ~35-fold greater than Empty Vector (Table 4. 2.). Moreover, increasing the number of glycine residues in the internal insertion site increased the infectivity of these HCVpp in a stepwise manner. This is a surprising result but was consistent across all three repeats of this experiment. Unfortunately, investigating this further was outwith the scope of this PhD project. However, it does offer an interesting opening into investigating how OCLN contributes to HCV entry.

Table 4. 16. Fold Change of the RLU Figures for the Internal Insertion I438V A524T J6-backbone Length Polymorphism-containing HCVpp versus Empty Vector in the Huh7 OCLN KO Cell Line.

HCVpp	Fold Change vs. Empty Vector
Empty Vector	1.00
+2I Internal Insertion I438V A524T J6 HCVpp	3.90
+4I Internal Insertion I438V A524T J6 HCVpp	16.89
+7I Internal Insertion I438V A524T J6 HCVpp	34.49

To date, it is understood that OCLN functions late HCV entry, being recruited to the site of viral entry, independent of CLDN-1 (Furuse *et al.*, 1993; Liu *et al.*, 2009; Ploss *et al.*, 2009; Sourisseau *et al.*, 2013; Deffieu *et al.*, 2022). These I438V A524T J6-backbone Internal Insertion HCVpp constructs may offer a route for further investigation of this process.

4. 4. 2. Huh7 WT Cells

Figure 4. 7. shows the infectivity of the HCVpp controls and length polymorphism-containing constructs in both backbones, with the wildtype J6-backbone constructs in figure 4. 7. A., and the I438V A524T J6-backbone constructs in figure 4. 7. B.

In general, the infectivity trends shown in figures 4. 2. And 4. 3. are recapitulated here. The log_{10}(RLU) figure from each HCVpp has been compared to each of the three controls separately through an ordinary one-way ANOVA followed by a Dunnett's multiple comparison test. Whilst the general trends are the same as reported previously, there are some differences between the exact statistical relationships presented here and in figures 4. 2. and 4. 3. Whilst these differences are subtle and likely reflect an improvement in technique over time, a change in cells, or different passage states for the cells, they are statistically different from the results obtained for these HCVpp in the infectivity assay. For this reason, it is important that these results are the ones to which the receptor dependency assay results are going to be compared as these infections were performed at the same time as infecting the various receptor KO Huh7 cell lines.

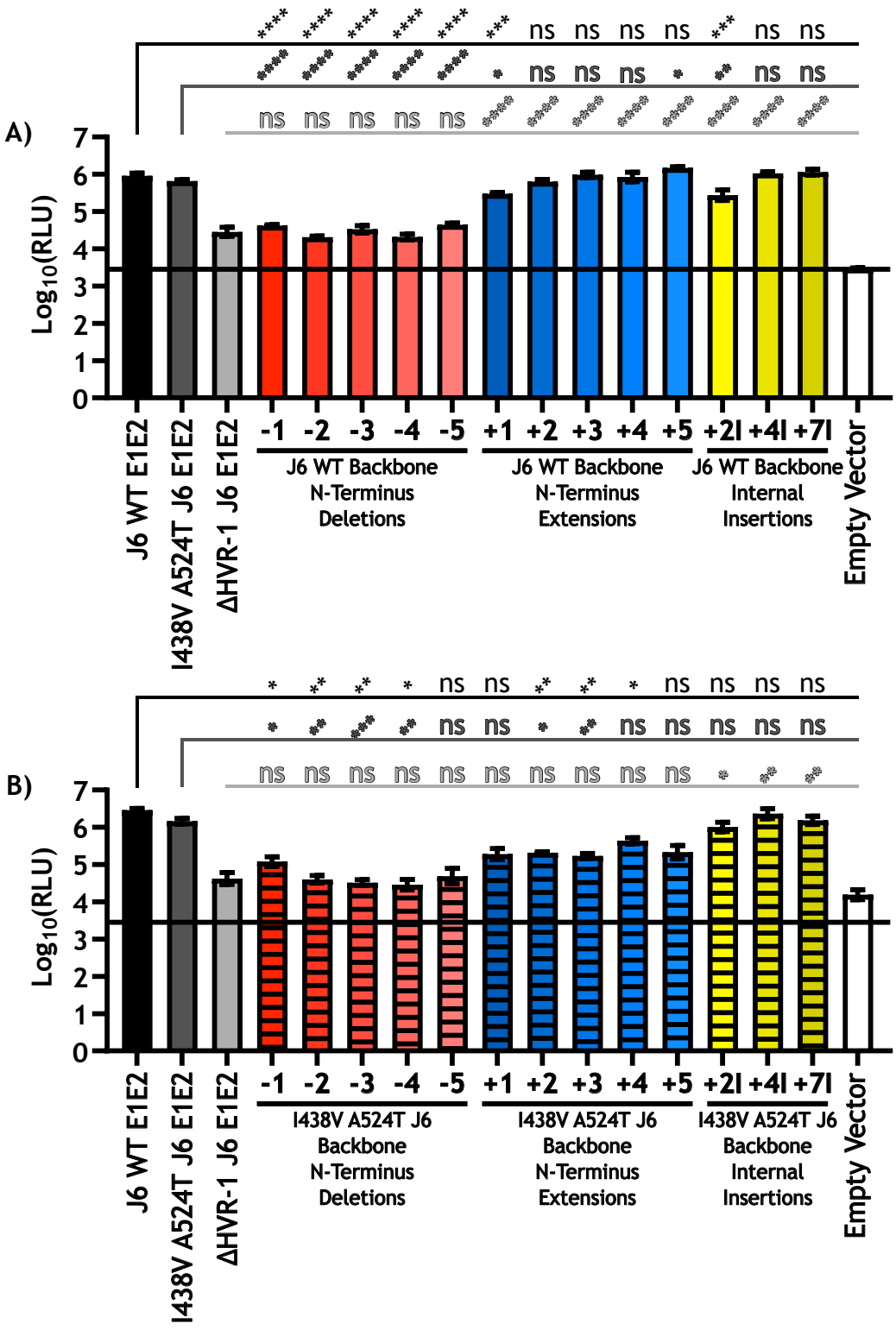


Figure 4. 7. **Infectivity of the HCVpp used in the Receptor Dependency Assays.** A) Log_{10}(RLU) measurements of the wildtype J6-backbone length polymorphism-containing HCVpp when infecting Huh7 WT cells. Ordinary one-way ANOVA tests followed by Dunnett's multiple comparisons tests were performed between each control and the HCVpp constructs. The raw RLU measurements were used to calculate percentage relative infectivity to give an accurate figure of difference in infectivity in each Huh7 KO cell line compared to the Huh7 WT cells. B) as in A) but for the I438V A524T J6-backbone length polymorphism-containing HCVpp. P-values; ns = P > 0.05, * = P ≤ 0.05, ** = P ≤ 0.01, *** = P ≤ 0.001, **** = P ≤ 0.0001. Error bars represent the SEM across 3 biological replicates.

4. 4. 3. Huh7 LDLR KO Cells

Figures 4. 8. through 4. 10. each consist of two paired upper and lower graphs and show the results of the receptor dependency assay performed in Huh7 LDLR KO cells, Huh7 SR-B1/LDLR DKO cells, and Huh7 SR-B1 KO cells, respectively. The first pair of graphs (A) shows the infectivity of the wildtype J6-backbone length polymorphism-containing HCVpp in each of the Huh7 receptor KO cell lines, the second pair (B) show the same but for the I438V A524T J6-backbone HCVpp.

In each pair, the lower graph shows the $\log_{10}$(RLU) figures obtained by each HCVpp and the controls, while the upper graph shows the percentage relative infectivity derived from those raw values. For example, the bottom graph (Fig. 4. 8. A. ii) shows the $\log_{10}$(RLU) figures obtained from each of the wildtype J6-backbone length polymorphism-containing HCVpp and the controls infecting Huh7 LDLR KO cells. The top graph (Fig. 4. 8. A. i) was produced by taking the raw RLU figure obtained in the Huh7 LDLR KO cell line and dividing by the RLU figure obtained by that same HCVpp when infecting the Huh7 WT cells. This proportion of retained infectivity was then converted into a percentage. The Huh7 WT % Relative Infectivity was always 100 %. A lower percentage relative infectivity in the KO cells implied dependence on that receptor for entry, whereas a similar or higher percentage implied independence from, or inability to use, that receptor. Lastly, for statistical validation, a 2way ANOVA test was performed to compare the $\log_{10}$(RLU) figures for each HCVpp in both the Huh7 WT cells and the Huh7 receptor KO cells.

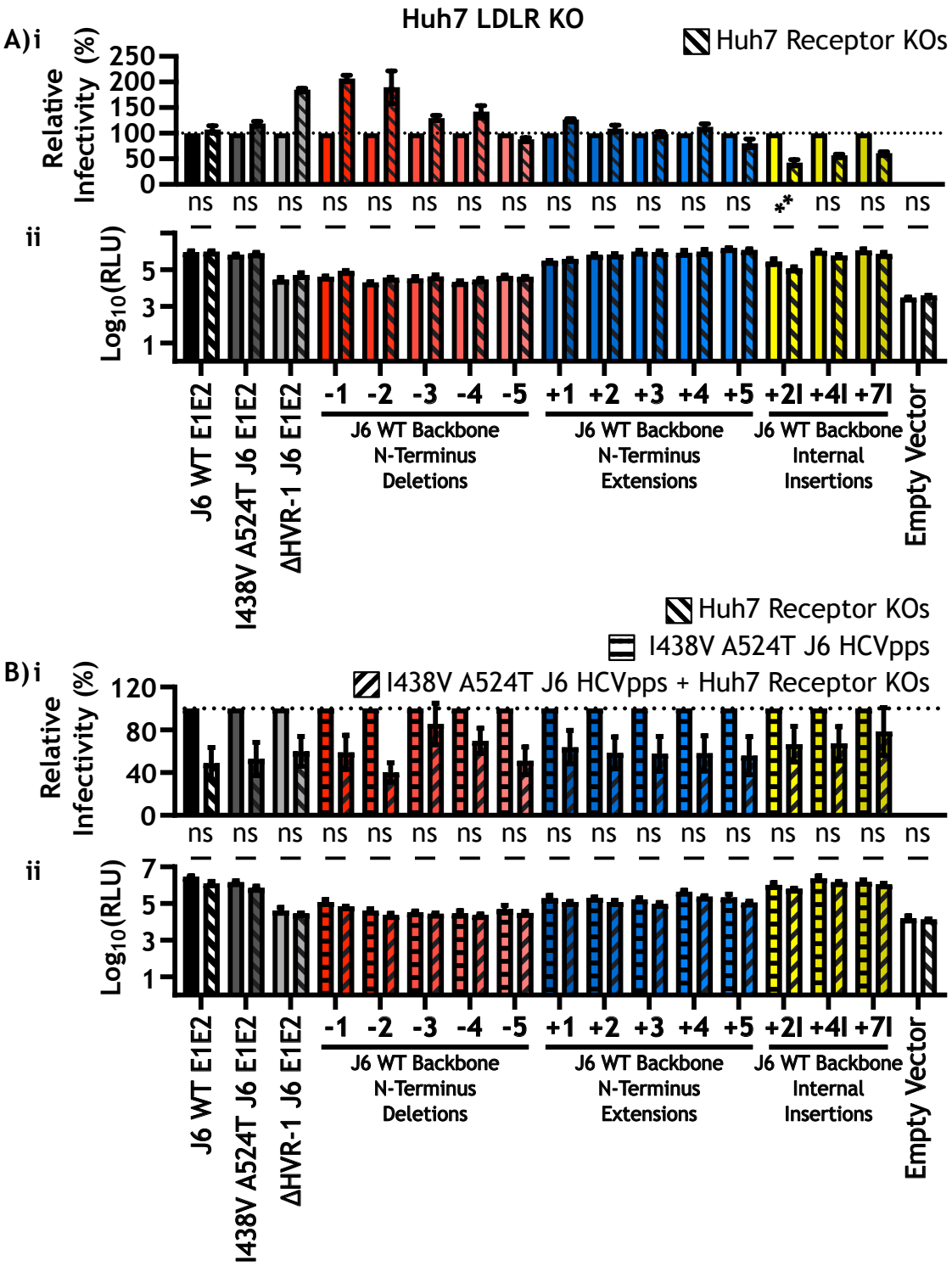


Figure 4. 8. **Infectivity of the HCVpp in Huh7 LDLR KO Cells.** RLU = Relative Luminescence Units. No pattern on the bar = Huh7 WT cells. Downward diagonal stripes = Huh7 KO cells. Horizontal stripes = I438V A524T-backbone HCVpp in Huh7 WT cells. Upward diagonal stripes = I438V A524T-backbone HCVpp in Huh7 KO cells. A) i Percentage relative infectivity when comparing RLU obtained in Huh7 WT cells and Huh7 LDLR KO cells. A) ii Comparison of Log_{10}(RLU) figures in Huh7 WT cells and Huh7 LDLR KO cells. 2way ANOVA tests followed by Šídák multiple comparison tests were performed to compare the log_{10}(RLU) figure obtained in the Huh7 WT cells and the Huh7 receptor KO cells for each HCVpp. P-values; ns = $P > 0.05$, ** = $P \leq 0.01$. Error bars represent the SEM across 3 biological replicates.

There is no statistical difference between the $\log_{10}$(RLU) figures obtained from infecting the Huh7 WT cells and the Huh7 LDLR KO cells for every HCVpp bar the +2I Internal Insertion J6 HCVpp, which has a reduced infectivity (Fig. 4. 8.). This result, intuitively, makes sense. As discussed previously, it has been proposed that the LDLR:HCV interaction relies on the apolipoproteins which associate with HCV virions (Owen *et al.*, 2009; Albecka *et al.*, 2012). These apolipoproteins, crucially, do not associate with HCVpp and this may render HCVpp unable to use LDLR as a receptor for viral entry. The reduced infectivity seen for the +2I Internal Insertion J6 HCVpp may reflect experimental variation over time and is unlikely to be a product of an underlying biological discovery.

Therefore, the HCVpp – no matter the length of the HVR-1 or the presence of the I438V A524T mutations – cannot use this receptor and therefore infectivity is not affected by the presence of LDLR or lack thereof.

4. 4. 4. Huh7 SR-B1/LDLR DKO Cells

Figure 4. 9. shows the infectivity of the length polymorphism-containing HCVpp in both backbones infecting Huh7 SR-B1/LDLR DKO cells. Across all the HCVpp used to infect these cells bar the ΔHVR-1 and Empty Vector controls, there is a statistically significant reduction in infectivity.

On first glance, these results appear unintuitive. It has been demonstrated in the previous figure (Fig. 4. 8.) that removing LDLR as a co-receptor makes no difference in infectivity for HCVpp due to the inability of HCVpp to use LDLR as a receptor. Therefore, it could be hypothesised that the Huh7 SR-B1/LDLR DKO cells would be as permissive to infection as Huh7 SR-B1 KO cells, which will be discussed in the next section. However, as will be demonstrated, the phenotypes shown in figure 4. 9. do not mirror those in figure 4. 10.

However, work by Zapatero-Belinchón *et al.* who examined the effects of lipoprotein receptors – namely SR-B1 and LDLR – and lipid-lowering drugs on HCV infectivity demonstrated something interesting. In their work, they noted that removal of both SR-B1 and LDLR drastically reduced both J6 HCVpp and H77 HCVpp infectivity (Zapatero-Belinchón *et al.*, 2021).

Moreover, they used CRISPR-Cas9 gene editing to produce multiple Huh7 SR-B1/LDLR DKO cell lines and they found that the infectivity in each of these cell lines was statistically significantly different from one another. Therefore, it can be posited that this reduction in infectivity is, perhaps, more than the sum of its parts and nods to this reduction in infectivity being a synergistic rather than simply additive effect.

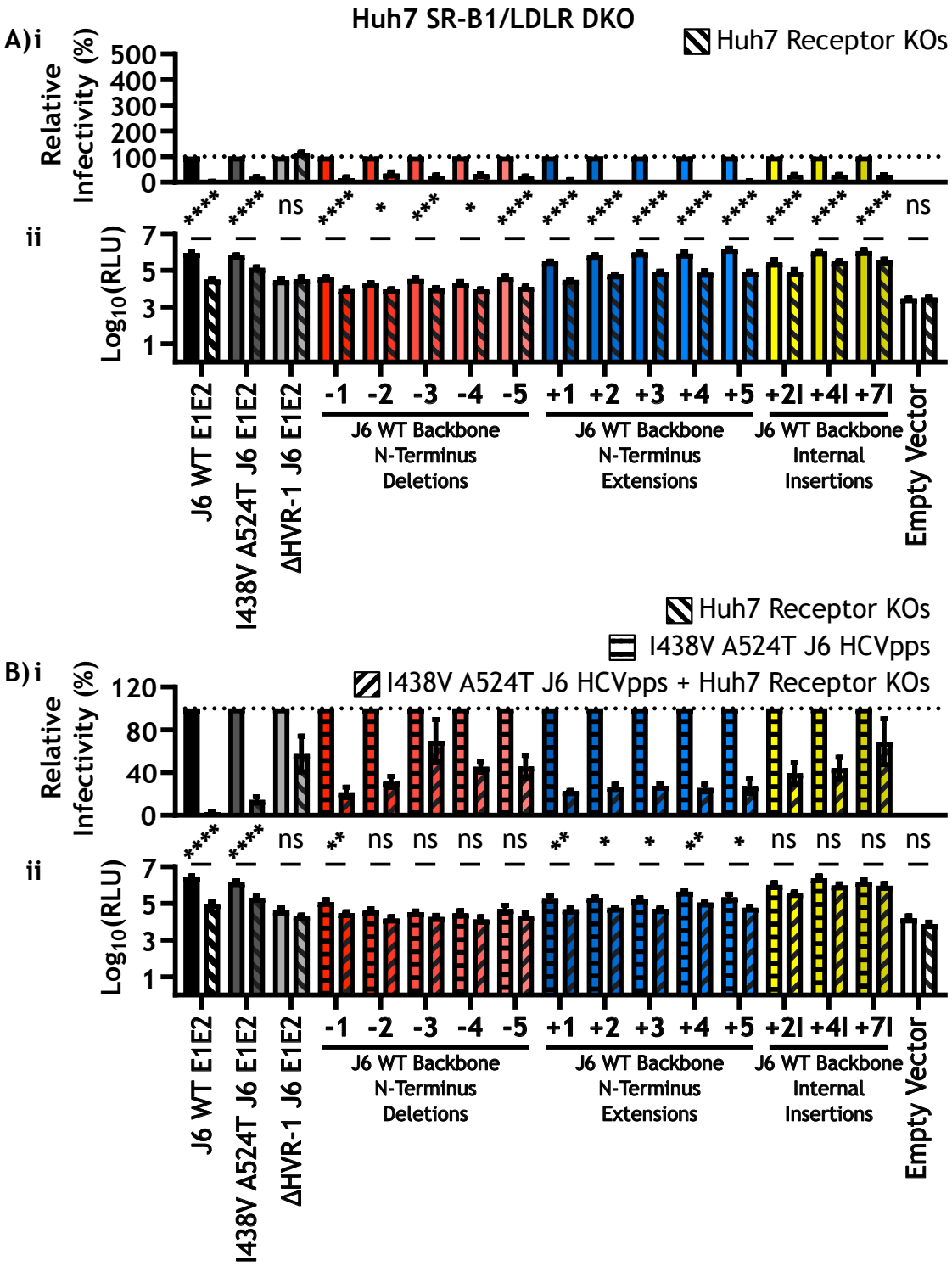


Figure 4. 9. **Infectivity of the HCVpp in Huh7 SR-B1/LDLR DKO Cells.** RLU = Relative Luminescence Units. No pattern on the bar = Huh7 WT cells. Downward diagonal stripes = Huh7 KO cells. Horizontal stripes = I438V A524T-backbone HCVpp in Huh7 WT cells. Upward diagonal stripes = I438V A524T-backbone HCVpp in Huh7 KO cells. A) i Percentage relative infectivity when comparing RLU obtained in Huh7 WT cells and Huh7 SR-B1/LDLR DKO cells. A) ii Comparison of Log_{10}(RLU) figures in Huh7 WT cells and Huh7 SR-B1/LDLR DKO cells. 2way ANOVA tests followed by Šídák multiple comparison tests were performed to compare the log_{10}(RLU) figure obtained in the Huh7 WT cells and the Huh7 receptor KO cells for each HCVpp.
P-values; ns = $P > 0.05$, * = $P \leq 0.05$, ** = $P \leq 0.01$, **** = $P \leq 0.0001$. Error bars represent the SEM across 3 biological replicates.

The Huh7 receptor KO cell lines were generated by Dr Yoshiharu Matsuura, Osaka University, using CRISPR-Cas9 technology (Yamamoto *et al.*, 2016). It has long been known that CRISPR-Cas9, for all its effectiveness, can induce off-target effects (Wu, Kriz and Sharp, 2014; Tasan and Zhao, 2017; Fu *et al.*, 2022). Unless these off-target effects are screened for effectively, the single Huh7 receptor KO cells which were grown to produce these KO cell lines may also have carried other mutations in other genes/proteins which could have downstream effects on cell permissiveness for HCVpp infection.

As demonstrated in the work of Zapatero-Belinchón *et al.,* two different Huh7 SR-B1/LDLR DKO cell lines produced statistically significantly different infectivity results for the same H77 HCVpp (2021). Therefore, without knowing whether off-target effects are present, or what their downstream consequences might be in the Huh7 SR-B1/LDLR DKO cells, all that can truly be said is that removal of these two receptors in these specific Huh7 SR-B1/LDLR DKO cells leads to an overall reduction in infectivity for any HCVpp which have a HVR-1.

Additionally, it should be noted that none of the Huh7 KO cell lines used in these experiments were screened for potential off-target effects. Indeed, off-target effects in any of these Huh7 KO cell lines could render particular cell lines more or less permissive to HCVpp infection. However, that being said, each cell line used in these experiments were derived from a single cell and should therefore be internally consistent. Ergo, differences in infectivity can be attributed to differences in the length of the HVR-1 and the different E2 backbones into which the various length polymorphisms were cloned. However, the off-target effects were discussed for the Huh7 SR-B1/LDLR DKO cells due to the incongruence of the expected results and the actual results i.e. it may be expected that this DKO cell line would recapitulate the results of the Huh7 SR-B1 KO cells due to the inability of HCVpp to interact with LDLR itself.

4. 4. 5. Huh7 SR-B1 KO Cells

The Huh7 SR-B1 KO context is the most informative (Fig. 4. 10.). As before, figure 4. 10. A. shows the results for the wildtype J6-backbone length polymorphism-containing HCVpp, and figure 4. 10. B. the I438V A524T J6-backbone HCVpp. As in the Huh7 WT cells, the infectivity figures for the HCVpp in the Huh7 SR-B1 KO cells formed groups based on type of length polymorphism in the HVR-1. Therefore, they shall be discussed as groups.

For the wildtype J6-backbone N-Terminus Deletion HCVpp, they showed an increased infectivity in the Huh7 SR-B1 KO cells compared to the Huh7 WT cells bar the -1 N-Terminus Deletion J6 HCVpp. This suggests two things: these HCVpp are SR-B1 independent; and furthermore, the increase in infectivity beyond what was seen in the Huh7 WT cells suggests they can engage other receptors more readily, leading to an increase in infectivity in the Huh7 SR-B1 KO cells.

To return to the results for the wildtype J6-backbone -1 N-Terminus Deletion HCVpp, briefly, this did not show increased infectivity in the Huh7 SR-B1 KO cells unlike the rest of the N-Terminus Deletion HCVpp constructs. This, as posited when discussing the infectivity data, may be due to the retention of the conserved second position threonine that the other wildtype J6-backbone N-Terminus Deletion HCVpp lack.

For the wildtype J6-backbone N-Terminus Extension HCVpp, again, there is a phenotypic grouping of these HCVpp constructs when comparing the $\log_{10}$(RLU) figures. However, it has been demonstrated that statistically, the +3 and +5 N-Terminus Extension J6 constructs show reduced entry efficacy/infectivity in the Huh7 SR-B1 KO cells. Moreover, whilst the differences are not statistically significant, all the other N-Terminus Extension HCVpp constructs do show a reduced infectivity in the Huh7 SR-B1 KO cells compared to their infectivity in the Huh7 WT cells.

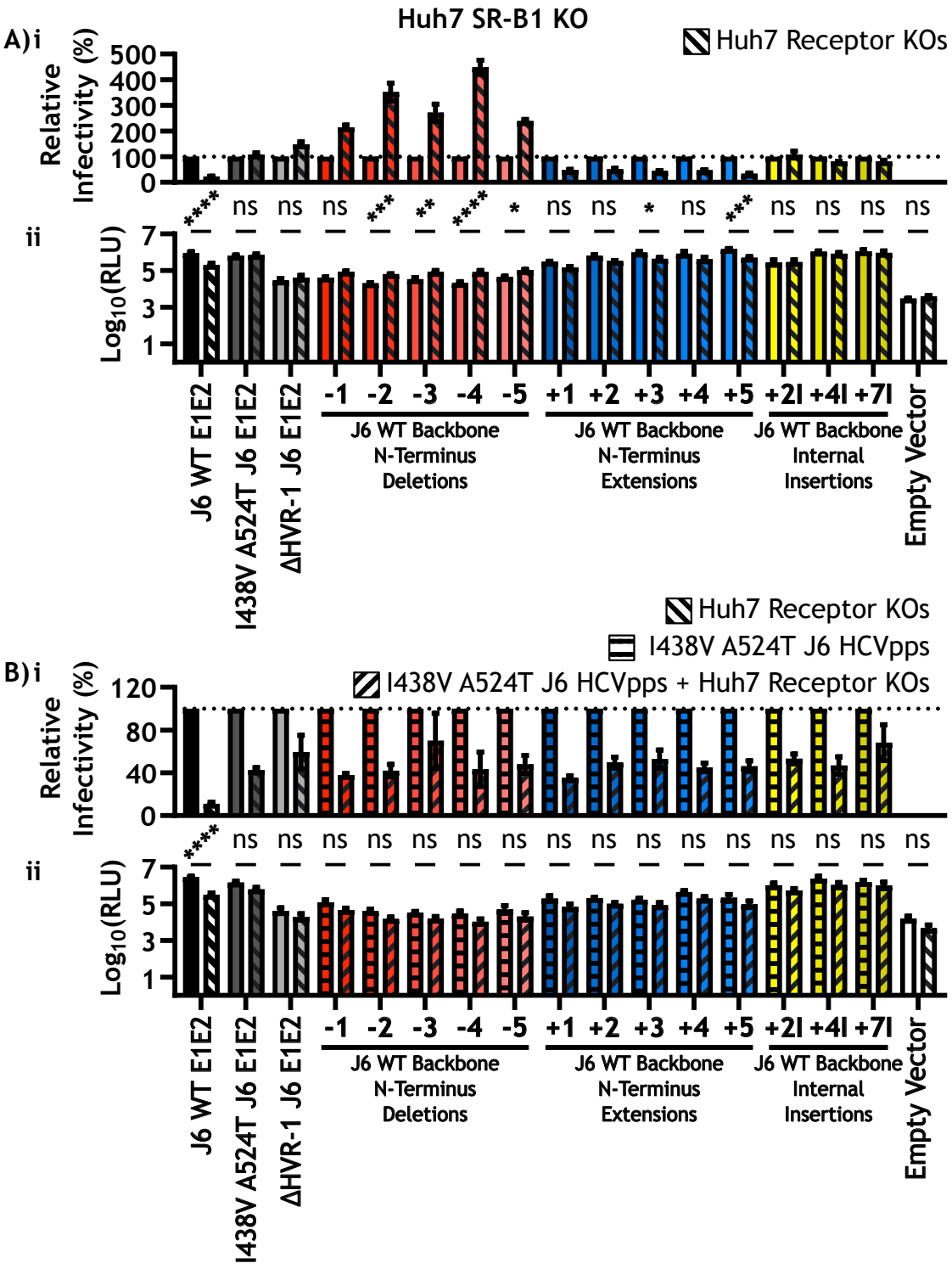


Figure 4. 10. **Infectivity of the HCVpp in Huh7 SR-B1 KO Cells.** RLU = Relative Luminescence Units. No pattern on the bar = Huh7 WT cells. Downward diagonal stripes = Huh7 KO cells. Horizontal stripes = I438V A524T-backbone HCVpp in Huh7 WT cells. Upward diagonal stripes = I438V A524T-backbone HCVpp in Huh7 KO cells. A) i Percentage relative infectivity when comparing RLU obtained in Huh7 WT cells and Huh7 SR-B1 KO cells. A) ii Comparison of Log_{10}(RLU) figures in Huh7 WT cells and Huh7 SR-B1 KO cells. 2way ANOVA tests followed by Šídák multiple comparison tests were performed to compare the log_{10}(RLU) figure obtained in the Huh7 WT cells and the Huh7 receptor KO cells for each HCVpp. P-values; ns = $P > 0.05$, * = $P \leq 0.05$, ** = $P \leq 0.01$, *** = $P \leq 0.001$, **** = $P \leq 0.0001$. Error bars represent the SEM across 3 biological replicates.

Altogether, for this group of HCVpp, there is either a reduction in infectivity, or no difference. This implies that there is some dependence on SR-B1 as a receptor for these HCVpp constructs, though it is not as pronounced as for the wildtype J6 HCVpp.

Lastly, the Internal Insertion J6 HCVpp constructs showed no statistical difference in infectivity across the Huh7 WT cells and the Huh7 SR-B1 KO cells. Therefore, these Internal Insertion HCVpp are SR-B1 independent. However, unlike the N-Terminus Deletion J6 HCVpp constructs, there is no increase in infectivity in the Huh7 SR-B1 KO cell line. Therefore, it can be posited that there is no increase in the engagement of other receptors.

The results for the I438V A524T J6-backbone length polymorphism-containing HCVpp differ greatly from those in the wildtype J6 backbone. Where before there were differences in SR-B1 receptor usage which could be related to the length of the HVR-1, here there is no statistically significant difference in infectivity in either the Huh7 WT or the Huh7 SR-B1 KO cell lines. This makes sense in the broader context of the I438V A524T mutations themselves which induce SR-B1 independence (Stejskal *et al.*, 2022) and this appears to be maintained in all length polymorphism HCVpp.

4. 5. HCVpp Neutralisation – Polyclonal Antibodies

Much research has been undertaken to investigate the relationship between HCV, the humoral immune response, and the HVR-1. Previous studies concluded that the HVR-1 was either "shielding" certain epitopes on the E1E2 complex, or was operating as an epitope decoy (Ray *et al.*, 1999; Korenaga *et al.*, 2001; Prentoe *et al.*, 2016). Altogether, the suggestion was that the HVR-1 was either "distracting" the humoral immune response or was sterically protecting particular E1E2 epitopes. However, this understanding of how the HVR-1 contributed to persistence was challenged in 2020 by Augestad *et al.* who demonstrated that, actually, the HVR-1 alters the conformation of the whole E1E2 complex. They demonstrated that the HVR-1 was able to stabilise neutralisation-resistant "closed" E1E2 complex states. Moreover, and more importantly given the prior section, they found that these "closed" states were highly dependent on SR-B1 as a co-receptor for viral entry (Augestad *et al.*, 2020).

Compare this to the paper by Stejskal *et al.* wherein the I438V A524T mutations were first characterised, and a clearer idea of the potential downstream effects of HVR-1 conformational switching on the global E1E2 conformation and, therefore, phenotype becomes more apparent. Stejskal *et al.* proposed that the I438V A524T mutations were constitutively stabilising the HVR-1 – this would shift the equilibrium of global E1E2 conformational states. Furthermore, in that paper, Stejskal *et al.* examined the transduction of movement throughout the E2 protein from the HVR-1 and found that there was a transduction of movement across the whole protein from the HVR-1 (Stejskal *et al.*, 2022).

Lastly, more work undertaken by Augestad *et al.* provides another key piece of this puzzle. Returning to their 8RJJ structure of the E1E2 complexes in a dimer of dimers, they assert that this structure shows the E1E2 complex in the "closed" conformation. The HVR-1 is pocketed and is inducing conformational changes along the rest of the E2 protein which render the complex antibody-resistant. Taken together, these three critical papers provide a solid foundation

for what might be expected of the length polymorphism-containing HCVpp when it comes to neutralisation.

It has already been discussed how a shorter HVR-1 might not be able to pocket due to physically not being able to reach between the base and the stem regions of the E2. Furthermore, it was hypothesised that insertions in the C-terminus of the HVR-1 would release the HVR-1 from conformational constraints more than insertions at the N-terminus, and lead to more rapid sampling of conformational space for the Internal Insertion HCVpp than the N-Terminus Extensions.

Given this information, and the finding by Augestad *et al.* that neutralisation-resistant, "closed" conformations correlated with SR-B1 dependency (Augestad *et al.*, 2020), it could be hypothesised that those HCVpp which were SR-B1 independent may also be antibody sensitive, and this may be explained by the lack of a stably pocketed HVR-1 which aided in stabilising the "closed" E1E2 complex.

4. 5. 1. J6 WT Backbone HCVpp Neutralisation – Polyclonal Antibodies

To examine the effects of length polymorphisms within the HVR-1 on its ability to mitigate neutralisation, the length polymorphism-containing HCVpp in both backbones were produced, incubated with polyclonal IgG antibodies - derived from patient sera - for 1 hour and then overlaid onto Huh7 WT cells. After 72 hours, luciferase activity was measured. Across both figure 4. 11. and figure 4. 12. the figure panels are laid out in the same manner.

Figure 4. 11. A. shows the neutralisation curves obtained for each of the individual HCVpp, coloured as per the key next to this figure. Figure 4. 11. B. i through iii show the curves for the different groups of length polymorphisms – the N-Terminus Deletion HCVpp, N-Terminus Extension HCVpp, and the Internal Insertion HCVpp, respectively. Figure 4. 11. C. shows the results of iterative extra sum-of-squares F tests which were performed between all the individual curves shown in figure 4. 11. A. to find the minimal number of curves required to describe the neutralisation data presented herein. Lastly, from the curves, the IC50 values were calculated and plotted by grouping in figure 4. 11. D.

Only the wildtype J6 HCVpp and I438V A524T J6 HCVpp results are shown in each of these figures. As discussed, when looking at figure 4. 1., these two controls represent antibody resistant, and antibody sensitive HCVpp, respectively. Considering the neutralisation data for the ΔHVR-1 HCVpp as shown in figure 4. 1., the ΔHVR-1 HCVpp were almost 100 % neutralised even at the lowest concentration of antibody and did not add into the interpretation of the results for the length polymorphism-containing HCVpp. Therefore, the ΔHVR-1 data was removed from these figures for visual clarity.

The most striking quality about the data presented in figures 4. 11. A. and B. i through iii is the distinct grouping of the wildtype J6-backbone length polymorphism-containing HCVpp by length polymorphism type. This was noted before when considering the infectivity assay and the receptor dependency assay results and once again, that trend of grouping by length polymorphism type continues for neutralisation.

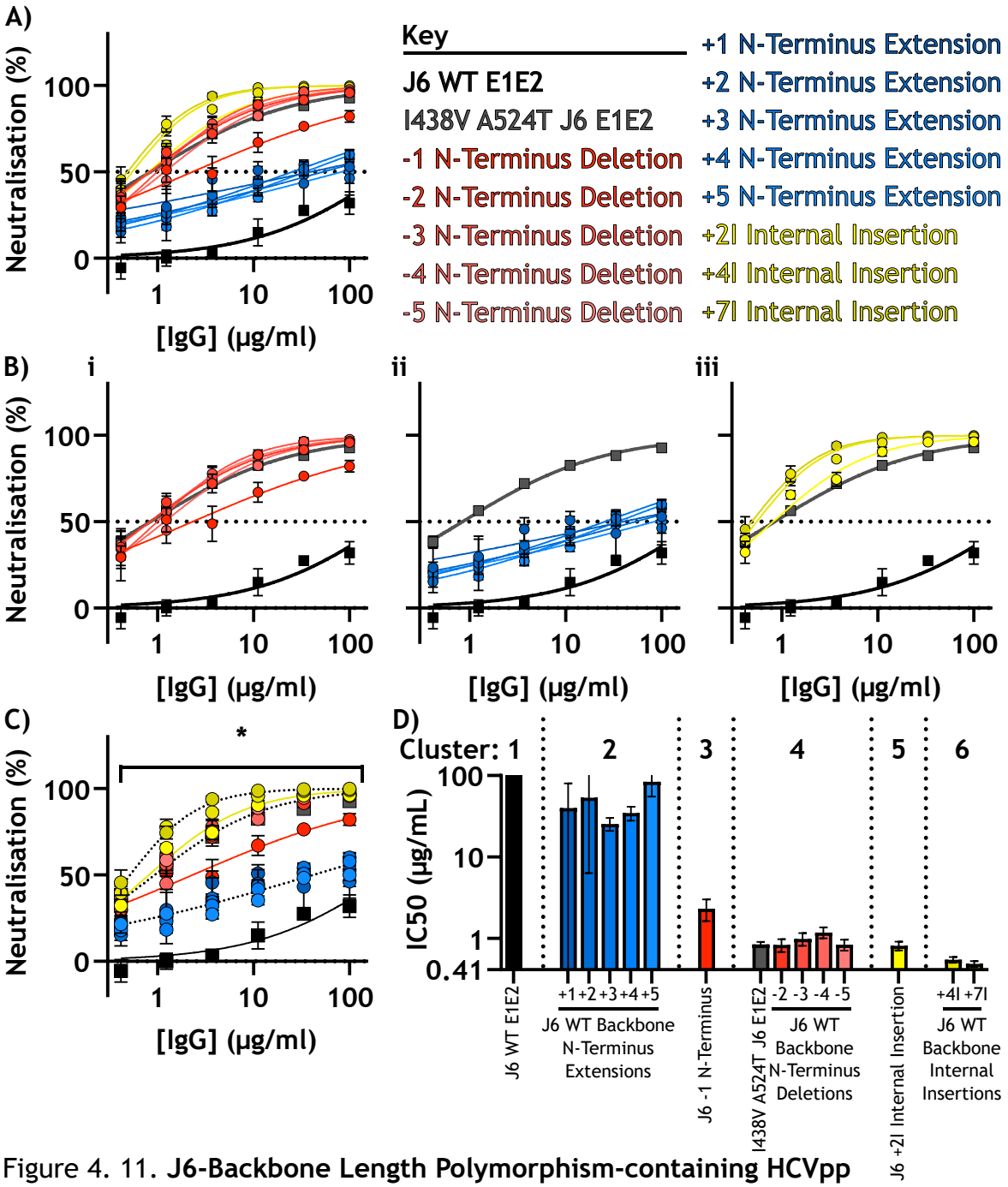


Figure 4. 11. **J6-Backbone Length Polymorphism-containing HCVpp Polyclonal Neutralisation.** A) Neutralisation curves obtained from neutralising the J6-backbone length polymorphism-containing HCVpp with patient-derived anti-HCV polyclonal antibodies. Curves were fitted with a 4-parameter non-linear regression with the hill slope constrained to be >0. B) Same as in A) but separated by length polymorphism type: i N-Terminus Deletions, ii N-Terminus Extensions, iii Internal Insertions. Due to the apparent clustering of the curves by length polymorphism type, iterative extra sum-of-squares F tests were performed between all curves to calculate the minimal number of curves required to describe the above data and to cluster the HCVpp. C) Six statistically different curves are required to describe these data. A black dotted line represents a cluster, whereas a coloured curve represents a single HCVpp which did not cluster. D) IC50s calculated from the non-linear regression, clustered by the extra sum-of-squares F test. Y-axis range = range of [IgG] tested.
* = statistically different curves i.e. p-value > 0.05 when comparing the individual curves within the cluster.

Here it is shown that the wildtype J6-backbone N-Terminus Deletion HCVpp are all antibody sensitive. However, in considering the grouping of the curves and the IC50 grouping shown in figures 4. 11. C. and 4. 11. D., it can be seen again that the -1 N-Terminus Deletion has a statistically different phenotype to the other four N-Terminus Deletions. The wildtype J6-backbone -1 N-Terminus Deletion HCVpp forms its own group (group 3 as per figure 4. 11. D.). Again, this recapitulates the statistical differences that have been demonstrated for the wildtype J6-backbone -1 N-Terminus Deletion HCVpp in both figures 4. 2. and 4. 10.

As discussed for both figures previously, this difference in phenotype between the -1 N-Terminus Deletion and the other N-Terminus Deletions may be a result of the retention of the highly conserved threonine as shown in figure 3. 9.

Moreover, in considering the framework of the three papers mentioned earlier (Augestad *et al.*, 2020, 2024; Stejskal *et al.*, 2022), and the fact that these HCVpp, likely, cannot stably pocket their HVR-1, these HCVpp being antibody sensitive intuitively makes sense. If they cannot pocket their HVR-1, they likely do not enter the “closed” antibody resistant state, and therefore are easily neutralised, with approximately 40 % neutralisation even at the lowest concentration of IgG.

The wildtype J6-backbone N-Terminus Extension HCVpp, on the other hand, are antibody resistant. In fact, they are the most resistant of the wildtype J6-backbone length polymorphism-containing HCVpp groups, reaching only ~50-60 % neutralisation at the highest concentration of IgG. Moreover, in considering figure 4. 11. C. and 4. 11. D., the wildtype J6-backbone constructs form their own group, second only to wildtype J6 HCVpp with respect to calculated IC50 values.

As discussed previously, the wildtype J6-backbone N-Terminus Extension HCVpp should, hypothetically, have the ability to pocket their HVR-1. It has previously been hypothesised that these HCVpp would have a more conformationally labile

HVR-1 than the wildtype J6 due to its increased length, but less labile than either the N-Terminus Deletions or Internal Insertions. This hypothesis appears to bear out in these neutralisation data. The trio of papers described above present a model wherein a pocketed HVR-1 locks the E1E2 complex in an antibody resistant conformation. A more conformationally labile HVR-1 may reduce the amount of time the HVR-1 spends pocketed and therefore increase antibody sensitivity by increasing the time that the E1E2 spends in the "open" conformation.

This hypothesis is compounded by the results for the wildtype J6-backbone Internal Insertions, which are the most antibody sensitive of the wildtype J6-backbone HCVpp groups. The wildtype J6-backbone Internal Insertion HCVpp are approximately 50 % neutralised, even at the lowest concentration of IgG. In the model proposed so far, these HCVpp have the most flexible HVR-1s, with the additional glycine residues at the C-terminus of the HVR-1 liberating the whole HVR-1 from any conformation constraints it may otherwise be under. It is hypothesised that whilst the Internal Insertion HCVpp can pocket their HVR-1, they may be less likely to do so due to the location of the insertions. These neutralisation data, along with the previous data about infectivity and receptor dependency support this hypothesis.

One thing to note, from figures 4. 11. C. and 4. 11. D. is that the wildtype J6-backbone +2I Internal Insertion does, statistically, exhibit a different phenotype with regards to neutralisation than the other two wildtype J6-backbone Internal Insertion HCVpp.

The distinct behaviour of +2I Internal Insertion raises the possibility that internal insertions exert an additive effect on HVR-1 flexibility, with phenotypic consequences scaling with insertion length. The more glycine residues added, the more flexibility at the base of the HVR-1, the more the HVR-1 is liberated from conformational constraints. However, to map down this potentially additive effect, additional constructs that include +1, +3, +5, and +6 glycine residues at the internal insertion site would need to be characterised.

Altogether, the neutralisation data presented here is consistent with the model presented by the three papers that have been discussed. Moreover, in linking this data back to the results of the previous sections, a clearer understanding of the relationship between the conformational state of the HVR-1, receptor dependency, and neutralisation can be established.

A pocketed HVR-1 is dependent on SR-B1 and renders the E1E2 complex antibody resistant. The occupancy rate of the pocket can be modulated by changing the length of the HVR-1 to either ablate pocketing or to disfavour it. This, in turn, affects both receptor dependency and neutralisation in strikingly predictable ways.

4. 5. 2. I438V A524 T J6 Backbone HCVpp Neutralisation – Polyclonal Antibodies

Figure 4. 12. follows the same panel layout as figure 4. 11., showing neutralisation curves for the I438V A524T J6-backbone length polymorphism-containing HCVpp.

These data present a much different story about the effects of length polymorphisms in the HVR-1 than before. However, it should be noted here that a direct comparison between the neutralisation results for the length polymorphism-containing HCVpp in the wildtype J6 backbone and the I438V A524T J6 backbone is not possible.

This was due to antibody availability constraints. As discussed before, and in the Methods and Materials chapter (Section 2. 2. 6.), the HCVpp were incubated with patient-derived IgG. These patients were chronically infected with either GT1 or GT3. However, due to the limited volume of IgG available for use across these experiments, two different polyclonal antibody stocks were used to neutralise the wildtype J6-backbone HCVpp versus the I438V A524T J6-backbone HCVpp. That being said, the results obtained for both the wildtype J6 control, and the I438V A524T J6 control do correlate well between this assay and the previous assay. Therefore, whilst trends can be discussed, a direct comparison would not be appropriate.

Once more, the I438V A524T J6-backbone N-Terminus Deletion HCVpp are well neutralised, exhibiting the lowest IC50 figures, and all of them being above ~40 % neutralised at even the lowest concentration of IgG. Additionally, it should be noted that the raw RLU figures for these HCVpp would occasionally dip below background in some technical repeats of this experiment – this experiment was performed five times in the end to achieve a suitable data set. This raises the question of whether these results reflect genuine antibody-mediated neutralisation or simply reflect the instability of these HCVpp which were not able to infect the Huh7 cells following the one-hour incubation with the polyclonal antibodies.

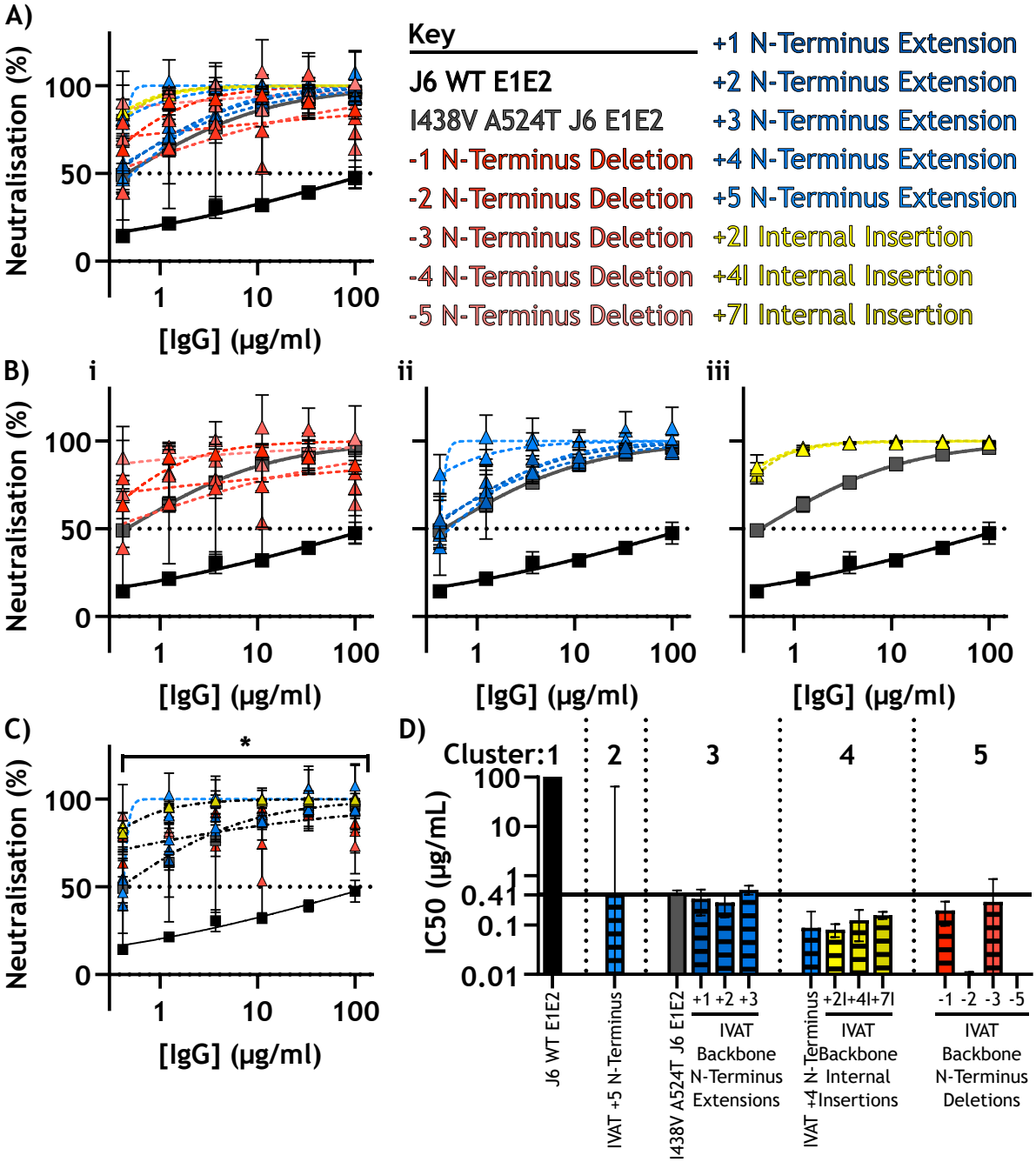


Figure 4. 12. **I438V A524T J6-Backbone Length Polymorphism-containing HCVpp Polyclonal Neutralisation.** A) Neutralisation curves obtained from neutralising the I438V A524T J6-backbone length polymorphism-containing HCVpp with patient-derived anti-HCV polyclonal antibodies. Curves were fitted with a 4-parameter non-linear regression with the hill slope constrained to be >0. B) Same as in A) but separated by length polymorphism type: i N-Terminus Deletions, ii N-Terminus Extensions, iii Internal Insertions. Iterative extra sum-of-squares F tests were performed between all curves to calculate the minimal number of curves required to describe the above data and to cluster the HCVpp. C) Five statistically different curves are required to describe these data. A black dot dash line represents a cluster, whereas a coloured curve represents a single HCVpp which did not cluster. D) IC50s calculated from the non-linear regression, clustered by the extra sum-of-squares F test. Bars are cut off at the maximum [IgG] tested. Solid line represents minimum [IgG] tested.
* = statistically different curves i.e. p-value > 0.05 when comparing the individual curves within the cluster.

This may explain why the curves for these HCVpp do not group together in the same manner as before – unlike the wildtype J6-backbone N- Terminus Deletion HCVpp, which grouped tightly, the curves here are more sporadic. Therefore, whilst this data would support the hypothesis that the inability to pocket the HVR-1 leads to antibody sensitivity, this conclusion should be drawn cautiously as the neutralisation curves for the I438V A524T J6 N-Terminus Deletion HCVpp cannot be attributed definitively to only the effects of the antibodies with which the HCVpp were incubated.

However, there is a stark difference between the neutralisation profiles of the I438V A524T J6 N-Terminus Extension HCVpp and the wildtype J6 N-Terminus Extension HCVpp. The latter were antibody resistant, and this was attributed to being able to pocket their HVR-1 which would result in the "closed" E1E2 complex conformation.

However, in the I438V A524T J6 backbone, the N-Terminus Extension HCVpp are highly antibody sensitive, with all these HCVpp being ~50 % neutralised at even the lowest concentration of IgG. Moreover, when examining the IC50 values in figure 4. 12. D., the solid horizontal line represents the lowest concentration of antibody tested in this assay – none of the bars associated with the I438V A524T J6 N-Terminus Extension HCVpp reach past this line, indicating that the IC50 falls below the lowest antibody concentration tested, meaning these HCVpp are hypothetically effectively neutralised before the titration even begins.

The I438V A524T J6-backbone +5 N-Terminus Extension HCVpp does group separately from the other four constructs (group 2), but its IC50 value is still comparable to the four other constructs. Moreover, the other four I438V A524T J6 N-Terminus Extension HCVpp grouped with the I438V A524T J6 HCVpp control. In considering all the data thus far, there is a good amount of evidence to support the notion that the I438V A524T mutations are more phenotypically dominant than insertion of additional amino acid residues at the N-terminus of the HVR-1.

Lastly, the I438V A524T J6 Internal Insertion HCVpp are well neutralised, even at the lowest concentration of IgG these HCVpp were ~80 % neutralised. This, again, is similar to the same phenotype shown in figure 4. 11. for the wildtype J6-backbone Internal Insertion HCVpp. The I438V A524T J6-backbone Internal Insertion HCVpp are, numerically, more neutralised at lower concentrations of IgG than their wildtype J6-backbone counterparts.

However, a direct comparison cannot be made as these two neutralisation assays were performed with two different batches of patient-derived IgG. The trends across both figures 4. 11. and 4. 12. for the Internal Insertion HCVpp in both backbones are the same, this does lend credence to the fact that the I438V A524T mutations do not appear to be able to alter the relationship between the HVR-1s of the Internal Insertions and the rest of their E2 proteins, as was shown for the N-Terminus Deletion HCVpp and N-Terminus Extension HCVpp.

However, when the neutralisation of these HCVpp was performed by select monoclonal antibodies, a slightly more nuanced picture of the relationship between the Internal Insertion constructs in their separate backbones may be elucidated.

4. 6. HCVpp Neutralisation – Monoclonal Antibodies

Following on from the polyclonal neutralisation assays, monoclonal neutralisation assays were performed with select monoclonal antibodies which would target different epitopes on the E2 protein of the HCVpp.

Due to the constraint of available volumes of antibody for this work, and the phenotypic grouping that has been demonstrated throughout the infectivity, receptor dependency, and polyclonal neutralisation assays only six length polymorphism-containing HCVpp constructs were used in these assays – one from each of the length polymorphism type groups across the two backbones. Altogether, the -3 N-Terminus Deletion, the +3 N-Terminus Extension, and the +4I Internal Insertion constructs in both backbones were used in these experiments. The selected monoclonal antibodies used in these assays were J6.36, AR3C, AP33, and AR4A.

In brief, the select length polymorphism-containing HCVpp constructs were produced in HEK 293T CD81 KO cells, incubated with the select monoclonals for one hour prior to being overlaid onto Huh7 WT cells and left for 72 hours. Following this, the cells were lysed and luciferase activity measured.

Of note, whilst the I438V A524T J6 -3 N-Terminus Deletion HCVpp were produced, the data derived from them was found to be uninterpretable due to consistently producing luciferase signals below background. It was believed that this was due to the inherent instability of these I438V A524T J6-backbone N-Terminus Deletion HCVpp. These data have been omitted from each of the figure panels.

Additionally, it was found that the +4 Internal Insertion constructs in both backbones were unaffected by J6.36 due to disruption of its epitope by the insertion of the additional glycine residues. As these data are not informative beyond saying that the epitope was disrupted and therefore neutralisation was

ablated, these HCVpp constructs have been removed from the data in figure 4. 13.

Lastly, each panel is laid out in the same manner. For example, figure 4. 13. A. shows the epitope of the J6.36 antibody mapped on to an AlphaFold 3 structure of the J6 E1E2 complex. Additionally, it should be noted that the epitope mapping which was used to colour these E1E2 complex models came from a review by Sautto *et al.* (2013) for J6.36, AR3C, and AP33. The epitope mapping for the AR4A antibody came from the E1E2 complex structure produced by Torrents de la Peña *et al.* (PDB: 7T6X, 2022).

Figure 4. 13. B. shows the neutralisation curves for both the wildtype J6 HCVpp and the I438V A524T J6 HCVpp controls, and each of the length polymorphism-containing HCVpp assessed.

Figure 4. 13. C. i and ii. further break down these data into the curves obtained by each of the control HCVpp and then the wildtype J6-backbone HCVpp and the I438V A524T J6-backbone HCVpp, respectively.

Figure 4. 13. D. shows the statistically distinct curves produced after iterative extra sum-of-squares F tests were performed to determine the minimal number of curves required to describe the full data set.

Lastly, figure 4. 13. E. shows the IC50 values calculated from the neutralisation curves set out by the groupings of the curves as found in figure 4. 11. D.

This layout is repeated for figures 4. 14., 4. 15., and 4. 16. but for the AR3C, AP33, and AR4A neutralisation data, respectively.

4. 6. 1. HCVpp Neutralisation – J6.36

J6.36 was first generated and characterised in 2011 by Sabo *et al.* They inoculated mice with either H77 or J6 sE2 proteins interperitoneally between 2 and 5 times until such point the antibody titres in the mice were sufficiently high. They then created hybridomas through splenocyte:myeloma fusion to produce immortalised cell lines which would produce single antibodies. They then characterised those antibodies and found that J6.36 was a potent inhibitor of HCV at a post-attachment step (Sabo *et al.*, 2011).

Further work on J6.36 found that its epitope actually consisted of two residues: F403 and G406 (Cowton *et al.*, 2018). As discussed previously, the glycine residues are inserted right after the F403 residue in the HVR-1. Therefore, it was believed, and demonstrated, that this antibody would not neutralise the Internal Insertion HCVpp.

Also in 2011, Prentoe *et al.* examined the effects of removing the glycans from the E2 protein. Notably, the glycans on the E1E2 proteins are thought to have several roles. Two of the most important are to stabilise the E1E2 complex (Torrents de la Peña *et al.*, 2022), and to allow the E1E2 complex to evade the immune system. Early studies proposed that the glycan shield on E2 protects HCV from neutralising antibodies through direct steric shielding of neutralisation epitopes (Falkowska *et al.*, 2007; Helle *et al.*, 2007, 2010; Prentoe *et al.*, 2011). However, more recent work by Prentoe *et al.* has challenged this model, suggesting instead that glycans contribute to immune evasion by stabilising "closed", neutralisation-resistant E1E2 conformations, co-ordinating with the HVR-1 to regulate the equilibrium between "open" and "closed" states (Prentoe *et al.*, 2019). They do note, however, that J6.36 binds directly to the HVR-1 itself and, therefore, its neutralisation is less dependent on the conformational "open"/"closed" equilibrium that the glycans are modulating, compared to antibodies targeting other epitopes.

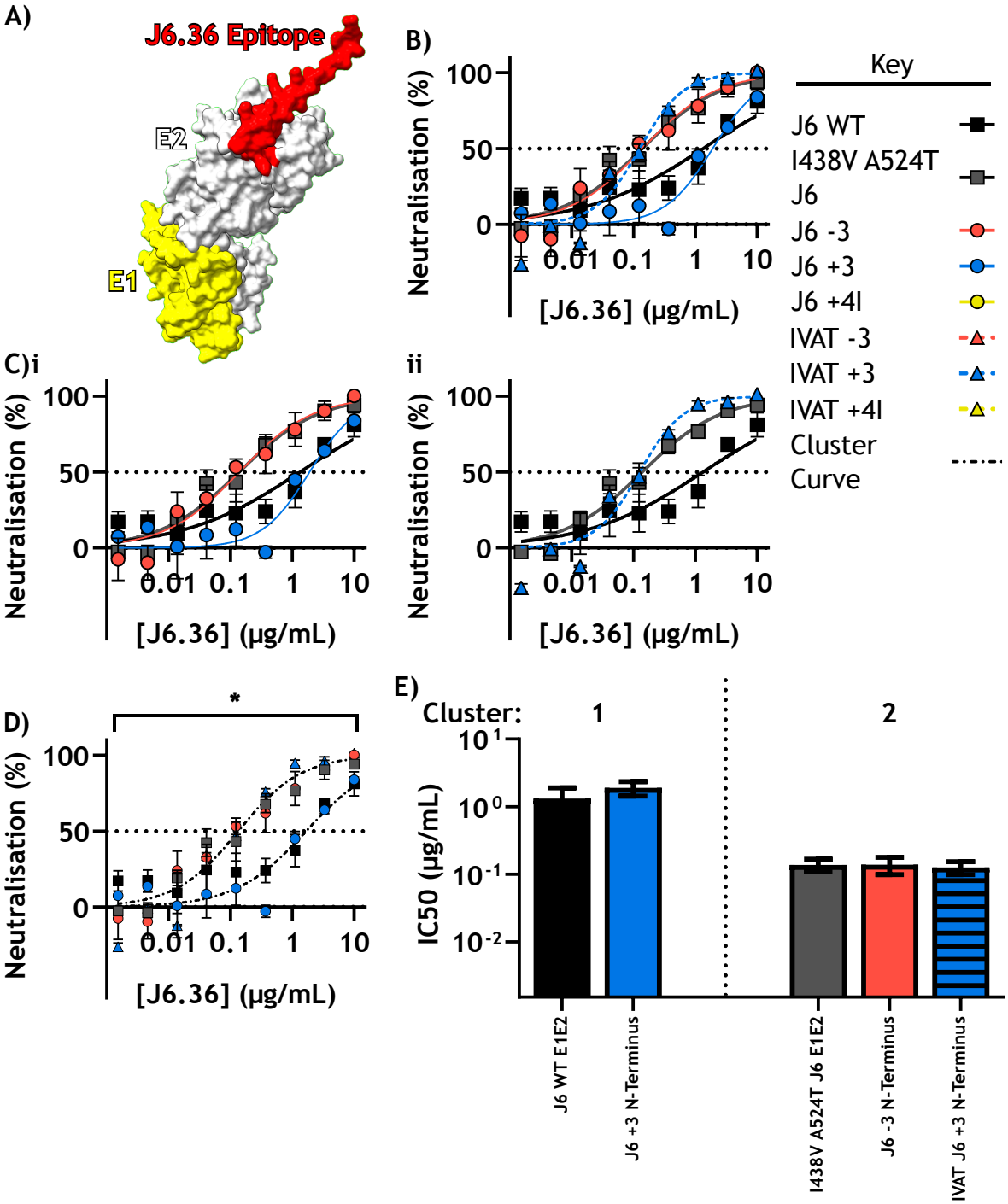


Figure 4. 13. **Monoclonal Neutralisation of the HCVpp using J6.36.** A) Mapping of the J6.36 epitope on an AlphaFold 3 structure of E1E2. E1 = yellow, E2 = white, J6.36 epitope = red. B) Curves obtained from neutralising HCVpp in both backbones with J6.36. Curves were fitted with a 4-parameter non-linear regression (hill slope > 0). C) Same as in B) but separated by backbone: i J6 WT, ii I438V A524T J6. Iterative extra sum-of-squares F tests were performed between curves to calculate the minimal number of curves required to describe the data and cluster the HCVpp. D) Two statistically different curves describe these data. A black dash-dotted line represents a cluster. E) IC50s calculated from the non-linear regression. Y-axis range = range of [J6.36] tested.
The Internal Insertion HCVpp were not neutralised due to disruption of the J6.36 epitope and therefore not shown. The I438V A524T J6 -3 N-Terminus HCVpp was too close to the background to be interpretable and therefore not shown.
* = statistically different curves

All this is to say that removal of the glycans from E2 was found to modulate the effectiveness of J6.36 neutralisation. Given that the HVR-1 has been shown to be central to the regulation of the conformational equilibrium between “open” and “closed” E1E2 states, and that J6.36 binds directly within the HVR-1 at residues F403 and G406, it was of interest to investigate whether altering the length of the HVR-1 would modulate the ability of J6.36 to neutralise the length polymorphism-containing HCVpp. Specifically, it was hypothesised that changes in HVR-1 length may alter the conformational accessibility of the J6.36 epitope, either by modulating the occupancy of the pocketed state wherein the HVR-1 epitope may be less accessible, or by directly disrupting the epitope itself, as in the case of the Internal Insertion constructs.

From the data, it has been shown that the -3 N-Terminus Deletion is well neutralised, with an IC50 value of approximately 0.1 µg/mL. This correlates well with the neutralisation profile for the other wildtype J6-backbone N-Terminus Deletion HCVpp as described previously. It has been described above that movement in the HVR-1 can induce “open” or “closed” states for the E1E2 complex which in turn may modulate antibody sensitivity. Here, however, the epitope is the HVR-1 itself.

Therefore, it can be considered that pocketing the HVR-1 may be the only defence that the HCVpp have against J6.36 – to sequester the HVR-1 away from the antibodies. The data presented here for the -3 N-Terminus Deletion fits into this model wherein shorter HVR-1s either cannot pocket or cannot stably pocket and are therefore more likely to be exposed thus increasing antibody sensitivity to J6.36.

The results for the N-Terminus Extensions in both backbones, considering the above, do make intuitive sense. To compare the results for the two backbones, it is shown that the wildtype J6-backbone N-Terminus Extension HCVpp are more resistant to J6.36 than their I438V A524T J6-backbone counterpart.

Considering the pocketing model, it can be hypothesised that the resistance to J6.36 neutralisation shown here by the wildtype J6-backbone +3 N-Terminus Extension HCVpp is mediated by the fact that these HCVpp are able to pocket their HVR-1 and sequester it away from the J6.36 antibodies.

However, where in previous figures it has been shown that the wildtype J6-backbone N-Terminus Extensions have a phenotype similar to that of wildtype J6 HCVpp, here they have an identical phenotype statistically and are grouped together (Fig. 4. 11. D. and fig. 4. 11. E.). The implication here is that the rate of exposure of the J6.36 epitope is the same for both of these HCVpp. Therefore, this suggests that both are equally able to sequester their HVR-1s and do so at a similar rate and with similar stability which may or may not be a result of pocketing the HVR-1. In summary: whatever the wildtype J6 HCVpp are doing to mitigate neutralisation by J6.36, the wildtype J6-backbone +3 N-Terminus Extension HCVpp are doing this just as well.

Lastly, in considering the data for the I438V A524T J6-backbone N-Terminus Extension support the hypothesis that the I438V A524T mutations are dominant over the N-Terminus Extension mutations as here the I438V A524T J6-backbone +3 N-Terminus Extension HCVpp have a statistically identical phenotype for J6.36 neutralisation as the I438V A524T J6 HCVpp themselves. However, what is interesting to note about this is that the implication for a high level of J6.36 sensitivity is that if the I438V A524T mutations do stabilise the HVR-1 as hypothesised previously, these mutations are doing so in a way that leaves the HVR-1 exposed for J6.36 neutralisation i.e. there is no evidence herein to suggest that the HVR-1 is being sequestered whilst being stabilised. Therefore, it is unlikely to be stabilised in the pocket.

4. 6. 2. HCVpp Neutralisation – AR3C

The AR3 family of antibodies were first identified in 2008 by Law *et al.* using phage display library of Fab fragments. They found that AR3 antibodies reacted with genotype 1a and genotype 2a HCV. They took this to suggest that the epitope for these AR3 antibodies was highly conserved. They further noted that AR3A-D where able to prevent CD81 binding to HCV (Law *et al.*, 2008). One of the first E2 structures that was released actually shows E2 in complex with the Fab fragment of AR3C (PDB: 4MWF, Kong *et al.*, 2013). From this structure, it was determined that the residues to which AR3C binds are in the front layer and one residue in the back layer. Specifically, they found that AR3C bound to T425, N428, C429, L438, L441, F442, and Y443 in the front layer, and W616 (627) in the back layer. Residue numbering is presented as in the paper, with the numbers as per the protein renumbering alignment found in appendix I in brackets (Kong *et al.*, 2013; Cowton *et al.*, 2018).

Based on the model established throughout this chapter, several predictions can be made about how the length polymorphism-containing HCVpp will perform in AR3C neutralisation assays. The wildtype J6-backbone N-Terminus Deletions were predicted to be well neutralised. The wildtype J6-backbone N-Terminus Extensions were predicted to be more antibody resistant, though more sensitive than wildtype J6 HCVpp themselves. The I438V A524T J6-backbone N-Terminus Extensions were predicted to be well neutralised given the dominance of the I438V A524T mutations demonstrated previously. Lastly, the Internal Insertion HCVpp in either backbone were predicted to be well neutralised. As will be shown, these predictions were borne out by the data.

Indeed, the wildtype J6-backbone N-Terminus Deletions were well neutralised. However, at the lowest concentration of AR3C, these HCVpp were ~25 % neutralised, reaching near 100 % neutralisation at the highest concentration. Based on the model presented above, this may be because these HCVpp may not be able to pocket their HVR-1 and, if true, may not enter a stably “closed” conformation.

The wildtype J6-backbone +3 N-Terminus Extension HCVpp were found to be highly AR3C resistant, similar to wildtype J6 HCVpp with IC50 values calculated to be greater than the range of AR3C concentrations assessed. Given that the wildtype J6-backbone +3 N-Terminus Extension HCVpp should be able to pocket their HVR-1, their resistance to AR3C is consistent with the pocketing model.

Moreover, as seen in the J6.36 neutralisation data, the wildtype J6-backbone +3 N-Terminus Extension HCVpp have, statistically, the same phenotype as the wildtype J6 HCVpp. Nevertheless, it cannot be confidently stated that these HCVpp have the same phenotype with regards to AR3C as this assay did not cover the range of antibody concentrations that would be required to really investigate the dose-response relationship between these HCVpp and AR3C. Another round of experimentation with a higher starting concentration of AR3C would be required to determine definitively whether these phenotypes are truly identical or whether the current data simply lacks the resolution to distinguish between them.

Next, it was predicted that the I438V A524T J6-backbone N-Terminus Extensions would be well neutralised due to the dominance of the I438V A524T mutations which appear to supersede the N-Terminus Extension mutations and render the HCVpp antibody sensitive, potentially through stabilisation of the HVR-1. The data here do support that prediction as the I438V A524T J6-backbone +3 N-Terminus Extension HCVpp was well neutralised. However, these data do allow for a further investigation as to how this antibody sensitivity is brought about. As noted previously, MD simulations by Stejskal *et al.* predicted that movement in

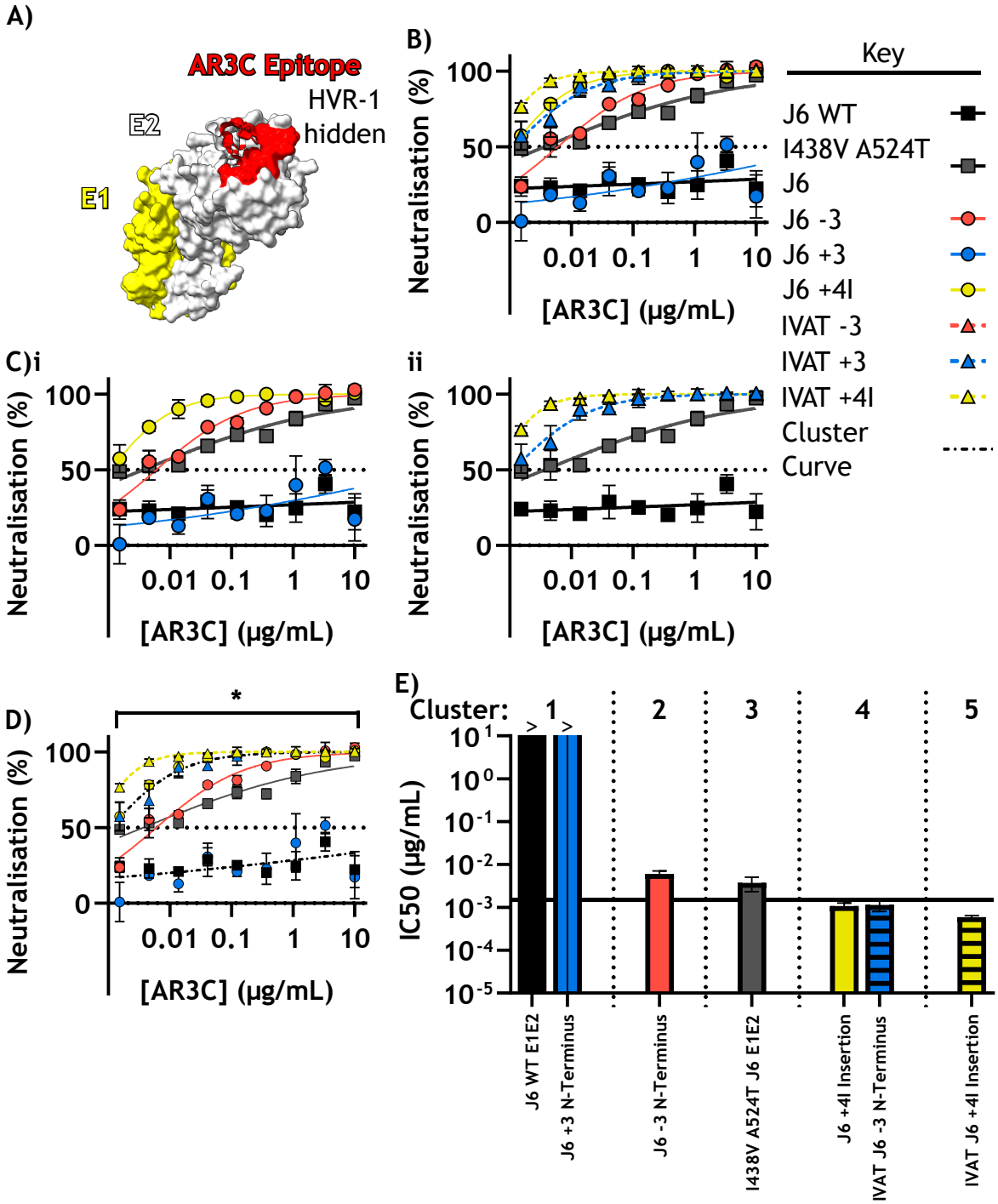


Figure 4. 14. **Monoclonal Neutralisation of the HCVpp using AR3C.** A) Mapping of the AR3C epitope on an AlphaFold 3 structure of E1E2. E1 = yellow, E2 = white, AR3C epitope = red. B) Curves obtained from neutralising HCVpp in both backbones with AR3C. Curves were fitted with a 4-parameter non-linear regression (hill slope > 0). C) Same as in B) but separated by backbone: i J6 WT, ii I438V A524T J6. Iterative extra sum-of-squares F tests were performed to calculate the minimal number of curves required to describe and cluster the HCVpp. D) Five statistically different curves describe these data. A black dash-dotted line = a cluster. E) IC50s calculated from the non-linear regression. The solid bar represents the minimum [AR3C] tested, the top of the y-axis is the highest [AR3C] tested. A ">" at the top of the bar = IC50 greater than the range of antibody concentrations tested.
I438V A524T J6 -3 N-Terminus HCVpp not shown as uninterpretable.
* = statistically different curves

the HVR-1 could be transduced throughout the rest of the E2 protein itself, in particular the front layer – where the epitopes for AR3C are found. However, crucially, Stejskal *et al.* also noted that this transduction of movement from the HVR-1 throughout the rest of the E2 protein was not found in their I438V A524T J6 MD simulations (Stejskal *et al.*, 2022). At the time, it was suggested that this was because of a constitutively stabilised HVR-1.

However, in this context, it may be suggested that whilst, hypothetically, the I438V A524T J6 +3 N-Terminus Extension HCVpp should be able to pocket their HVR-1, doing so may not transduce movement across the front layer of the E2 protein in such a way as to render the E1E2 in a "closed" antibody resistant conformation. However, determining if the HVR-1 is stabilised by the I438V A524T mutations, or if these mutations render the E1E2 complex to be in a constitutively "open" state would require further experimentation.

Lastly, it was predicted that the Internal Insertion HCVpp would be well neutralised in either backbone, given that their conformationally labile HVR-1s are unlikely to stably pocket and therefore unlikely to stabilise the "closed", antibody resistant E1E2 conformation. This was reflected in the data presented in figure 4. 14. Both Internal Insertion HCVpp constructs were above 50 % neutralisation at the lowest AR3C concentration assessed and approaching 100 % neutralisation at 0.1 µg/mL.

However, the statistical analysis indicates that the two Internal Insertion HCVpp are significantly different from one another depending on backbone i.e. they are found in different groups in figures 4. 14. D. and 4. 14. E. This would imply that the I438V A524T mutations are having some effect on the Internal Insertions and would provide nuance to the hierarchy of mutations hypothesis previously discussed. That said, both constructs are so sensitive to AR3C that the assay cannot meaningfully distinguish between them. They are both effectively at the ceiling of detection.

Therefore, this statistical difference is more likely a reflection of the limited dynamic range of this assay than a genuine biological difference. A lower starting concentration of AR3C would be required to determine whether the I438V A524T mutations truly alter the AR3C sensitivity of the Internal Insertion HCVpp.

4. 6. 3. HCVpp Neutralisation – AP33

The stocks of AP33 used throughout this PhD were graciously provided by Professor Arvind Patel, MRC-University of Glasgow, Centre for Virus Research. Additionally, Professor Arvind Patel and his team discovered and characterised AP33 in the early 2000s (Owsianka *et al.*, 2001). It was found at the time to be a potent inhibitor of different forms of E2 and HCV virus-like particles. Moreover, it was known to bind to AS412, which is immediately downstream of the HVR-1 and stretches from residue 412 to 423. Structural studies by Kong *et al.* successfully resolved the structure of AP33 binding a peptide based on AS412. They found that the AS412 peptide bound to AP33 in a ß-hairpin conformation. Based off this structure, the relevant residues for AP33 binding were determined to be L413, N415, G418, and W420 (PDB: 4G6A, Kong *et al.*, 2012; Cowton *et al.*, 2018).

Briefly, the 8RJJ dimer of dimers model for the E1E2 complex shall be discussed in relation to the conformation of AS412. In their structure, Augestad *et al.* found that the pocketed HVR-1 led to an AS412 that was locked into the ß-hairpin conformation (PDB: 8RJJ, Augestad *et al.*, 2024).

On face value, it may then be predicted that in this AP33 neutralisation assay that any HVR-1 that is able to pocket, and therefore induce the ß-hairpin conformation in the AS412, may be more sensitive to AP33 than those HCVpp that cannot pocket their HVR-1 such as the wildtype J6-backbone N-Terminus Deletions. However, this is not supported by the data presented in figure 4. 15.

Rather, what is demonstrated is that the wildtype J6-backbone -3 N-Terminus Deletion is AP33 sensitive, reaching approximately 100 % neutralisation at the highest concentration of AP33. Moreover, the IC50 value for these HCVpp was lower than that of both control HCVpp and the wildtype J6-backbone +3 N-Terminus Extension HCVpp (Fig. 4. 15. E.).

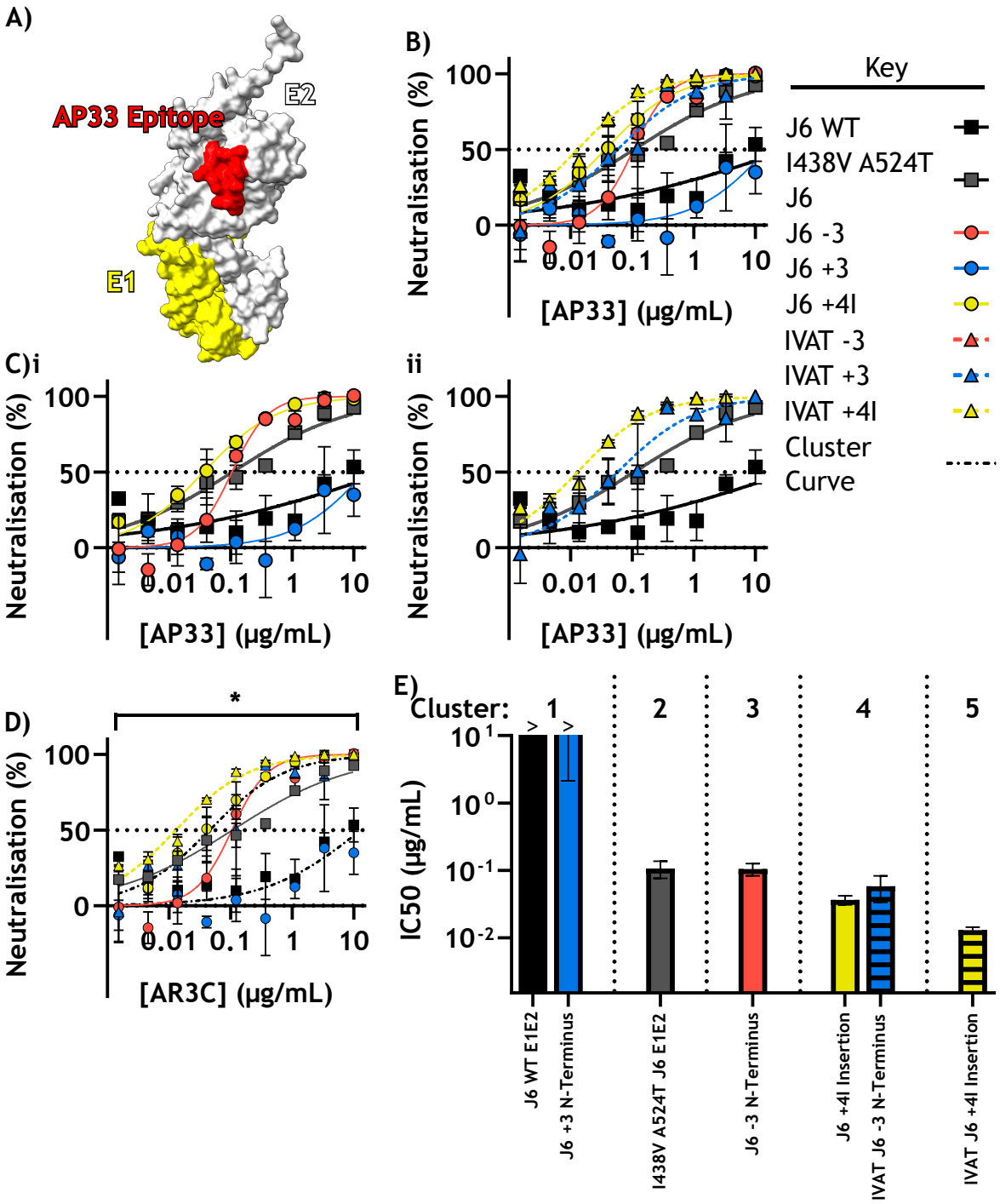


Figure 4. 15. **Monoclonal Neutralisation of the HCVpp using AP33.** A) Mapping of the AP33 epitope on an AlphaFold 3 structure of E1E2. E1 = yellow, E2 = white, AP33 epitope = red. B) Curves obtained from neutralising HCVpp in both backbones with AP33. Curves were fitted with a 4-parameter non-linear regression (hill slope > 0). C) Same as in B) but separated by backbone: i J6 WT, ii I438V A524T J6. Iterative extra sum-of-squares F tests were performed to calculate the minimal number of curves required to describe and cluster the HCVpp. D) Five statistically different curves describe these data. A black dash-dotted line = a cluster. E) IC50s calculated from the non-linear regression. The y-axis range is the range of [AP33] tested. A ">" at the top of the bar = IC50 greater than the range of antibody concentrations tested.
I438V A524T J6 -3 N-Terminus HCVpp not shown as uninterpretable.
* = statistically different curves

Therefore, the idea that not being able to pocket the HVR-1 may offer some sort of protection from neutralisation by AP33 is not supported by these data. However, it has been known since the early 2000s that some antibodies are able to induce conformational changes in their epitopes (Roguin and Retegui, 2003) and as recently as 2024, it was demonstrated that two SARS-CoV-2-specific antibodies were able to induce conformational changes in the RBD (Sun *et al.*, 2024).

Taken together, these may suggest that even if AS412 is not in the ß-hairpin conformation, AP33 may be able to induce this conformation when neutralising the length polymorphism-containing HCVpp. Therefore, the inability to pocket the HVR-1 and therefore lock the AS412 into the ß-hairpin conformation may not offer protection from AP33. Moreover, in considering the data from section 4. 3. of this chapter, AP33 was used in reducing Western blots to detect the presence of E2 from the wildtype J6-backbone length polymorphism-containing HCVpp. This implies that AP33 can bind to both its linear and conformational epitopes.

In considering what the data does show, it recapitulates the patterns of neutralisation that have so far been demonstrated in both the polyclonal neutralisations and the above monoclonal neutralisations. As discussed, the wildtype J6-backbone N-Terminus Deletions are well neutralised. However, these data do not necessarily speak directly to the lack of pocketing leading to an “open” E1E2 complex which is antibody sensitive. As noted above, what may be happening here is that AP33 is able to induce the ß-hairpin conformation in AS412 and prevent downstream conformational changes which facilitate viral entry.

Next, the +3 N-Terminus Extension HCVpp in both backbones shall be considered. Once more, there is this dichotomy between the two wherein the wildtype J6-backbone HCVpp is antibody resistant. However, upon introduction of the I438V A524T mutations, these HCVpp become antibody sensitive. This recapitulates what has been demonstrated before.

In brief, it has already been discussed that both of these HCVpp should be able to pocket their HVR-1s, however, it has been shown that AP33 can neutralise HCVpp which are predicted to be pocketing incompetent and therefore this should not affect neutralisation by AP33. Therefore, the difference here in antibody sensitivity is likely attributable to the I438V A524T mutations themselves. Indeed, the I438V A524T J6 HCVpp is well neutralised in this assay.

Lastly, the Internal Insertions shall be considered. Previously, it was discussed that the dynamic range for the AR3C experiment was not sufficient to truly disentangle whether the statistical difference found for the Internal Insertion HCVpp in both backbones reflected the underlying biology or the statistical analyses themselves. Here, however, it is shown that the Internal Insertion HCVpp do separate out into statistically different groupings.

Moreover, from figure 4. 15. B., this assay captures more of the dynamic range for this neutralisation assay, and therefore this statistical result may be more dependable. Therefore, in this case, it does appear that there is some effect that the I438V A524T mutations are imposing on the Internal Insertion HCVpp constructs. Notably, this represents the first instance in all the assays discussed so far where a statistically resolvable difference in phenotype between the wildtype J6-backbone and I438V A524T J6-backbone Internal Insertion HCVpp has been detected. The calculated IC50 values are 0.037 ± 0.0056 µg/mL for the wildtype J6-backbone +4I Internal Insertion HCVpp versus 0.013 ± 0.0015 µg/mL, i.e. a 2.85-fold change.

However, this would require further investigation and a clearer idea as to exactly how the I438V A524T mutations impose their phenotype. Moreover, further studies would need to elucidate how the Internal Insertions alter the relationship between the HVR-1 and the rest of the E2 protein. Having definitive answers to both of these research questions would allow the question as to how these mutations intersect with one another to be answered fully.

4. 6. 4. HCVpp Neutralisation – AR4A

Lastly, the select HCVpp were neutralised using AR4A,; this monoclonal antibody was discovered via phage display technology using an exhaustive-panning strategy from a patient with a chronic genotype 1a HCV infection (Giang *et al.*, 2012). This antibody was found to only be able to find correctly folded E1E2 complexes. Moreover, they demonstrated that this antibody was able to neutralise HCVcc and HCVpp from genotypes other than genotype 1a. They concluded that this meant the epitope bound by AR4A must be highly conserved to enable this level of cross-genotype neutralisation (Giang *et al.*, 2012).

In 2022, the first structure of the HCV E1E2 complex was released which used AR4A as a chaperone antibody. From this structure, the epitope of AR4A could be mapped precisely to residues in the back layer of E2. More precisely, AR4A binds the following residues in E2: W646 (657), G649 (660), L667 (678), Q671 (682), P676 (682), I696 (707, D698 (709) (Torrents de la Peña *et al.*, 2022). Residue numbers in brackets are references to the re-numbering scheme denoted in appendix I.

Further characterisation work helped to generate the steric shielding hypothesis to describe how the HVR-1 was able to mitigate neutralisation (Prentoe *et al.*, 2016; Velázquez-Moctezuma *et al.*, 2019). These papers found that AR4A can access its epitope regardless of whether the HVR-1 was present on the HCVpp or not. Therefore, it may be hypothesised that changing the length of the HVR-1 may not modulate AR4A sensitivity.

However, as noted previously, the model for HVR-1-mediated antibody protection has evolved. Rather than a simple epitope decoy or steric shielding of the epitopes found on E2, the HVR-1 is able to modulate global envelope dynamics and this is what results in antibody resistance/sensitivity through the induction of “open” or “closed” E1E2 complex conformations (Augestad *et al.*, 2020). For AR4A, however, these two mechanisms of HVR-1-mediated antibody resistance are not mutually exclusive.

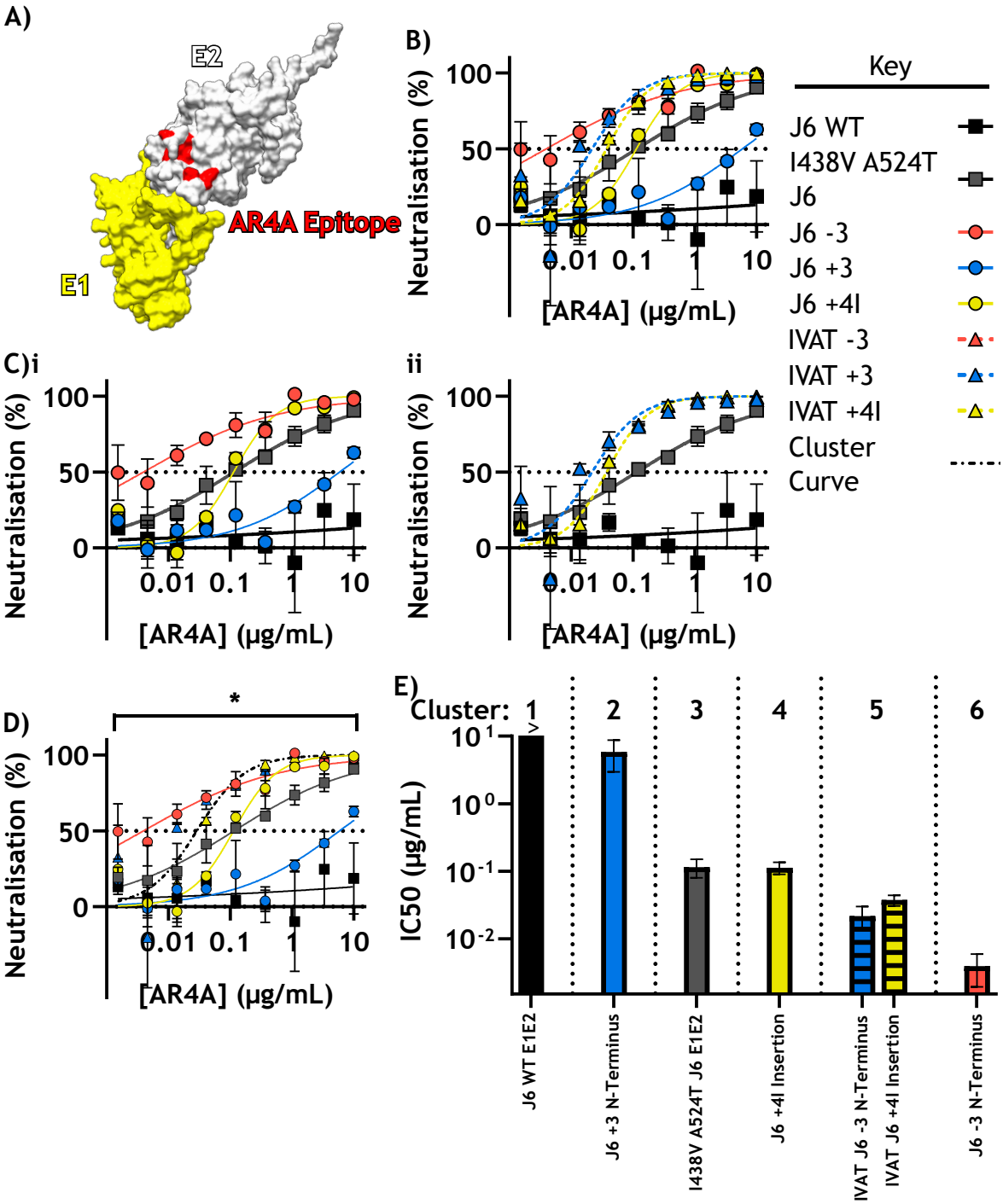


Figure 4. 16. **Monoclonal Neutralisation of the HCVpp using AR4A.** A) Mapping of the AR4A epitope on an AlphaFold 3 structure of E1E2. E1 = yellow, E2 = white, AR4A epitope = red. B) Curves obtained from neutralising HCVpp in both backbones with AR4A. Curves were fitted with a 4-parameter non-linear regression (hill slope > 0). C) Same as in B) but separated by backbone: i J6 WT, ii I438V A524T J6. Iterative extra sum-of-squares F tests were performed to calculate the minimal number of curves required to describe and cluster the HCVpp. D) Six statistically different curves describe these data. A black dash-dotted line = a cluster. E) IC50s calculated from the non-linear regression. The solid bar represents the minimum [AR4A] tested, the top of the y-axis is the highest [AR4A] tested. A ">" at the top of the bar = IC50 greater than the range of antibody concentrations tested.
I438V A524T J6 -3 N-Terminus HCVpp not shown as uninterpretable.
* = statistically different curves

In their 8RJJ structure, which shows a pocketed HVR-1, Augestad *et al.* note that the pocketed HVR-1 occludes the AR4A epitope i.e. it sterically shields the E1E2 from AR4A neutralisation (Augestad *et al.*, 2024). Therefore, whilst in general the ability to modulate antibody sensitivity may be due to modulating the global E1E2 conformation, for AR4A having an HVR-1 which cannot pocket may lead to increased AR4A sensitivity at least. This is confirmed by the data in figure 4. 16.

The wildtype J6-backbone -3 N-Terminus Deletion HCVpp are the most AR4A sensitive HCVpp evaluated. Their calculated IC50 value was 0.0040 ± 0.0020 µg/mL, the lowest of all the HCVpp assessed. Given the 8RJJ dimer of dimers structure of the E1E2 complex (Augestad *et al.*, 2024), this is perhaps the most compelling evidence that the wildtype J6-backbone -3 N-Terminus Deletion HCVpp are unable to pocket their HVR-1. Therefore, they are unable to sterically shield AR4A from its epitope in the stem region and are, therefore, AR4A sensitive. This result makes intuitive sense, and moreover, offers the most compelling evidence that these shortened HVR-1s cannot pocket their HVR-1s. However, whilst this finding lends weight to previous conclusions based on the possibility that the shortened HVR-1s are not able to stably bind into the pocket, it is still not direct evidence.

With this framing, that AR4A sensitivity may be a proxy for ability to pocket, or frequency of pocketing the HVR-1, the other data provides further supporting evidence. Consider the N-Terminus Extensions in both backbones. It has been discussed that whilst both of these HCVpp should be able to pocket their HVR-1, they are differently sensitive to AR4A neutralisation and statistically appear in different neutralisation groups (Fig. 4. 16. D. and fig. 4. 16. E.).

It was previously hypothesised that this may be due to constitutive stabilisation of the HVR-1 induced by the I438V A524T mutations, but this picture becomes clearer with these data – whatever the I438V A524T mutations are doing is preventing the HVR-1 from pocketing stably and therefore leaving the AR4A epitope exposed.

Moreover, as discussed previously for the wildtype J6-backbone +3 N-Terminus Extensions, it was proposed that the HVR-1 on these HCVpp would be more conformationally labile than that of wildtype J6 HCVpp and therefore more likely to explore conformational space than wildtype J6 HCVpp. Again, this is borne out in this data. The wildtype J6-backbone +3 N-Terminus Extension HCVpp are more sensitive to neutralisation by AR4A than wildtype J6 HCVpp. This implies that comparatively, the HVR-1 of the wildtype J6-backbone +3 N-Terminus Extension HCVpp is spending less time in the pocket sterically shielding the AR4A epitope.

As before, the Internal Insertion HCVpp data support the idea that the HVR-1 of these Internal Insertion HCVpp constructs is the most conformationally labile and therefore spending less time in the pocket. Moreover, the data here demonstrate that there is a statistically significant difference between the neutralisation of the wildtype J6-backbone +4I Internal Insertion HCVpp and those in the I438V A524T J6 backbone.

Therefore, in discussing the hierarchy of mutations, it does appear now that the I438V A524T mutations are altering the phenotype of the Internal Insertion HCVpp in some manner. However, this was only observable through these monoclonal neutralisation experiments. Therefore, whilst the I438V A524T mutations and the Internal Insertions mutations do appear to have a synergistic effect on phenotype, it is less pronounced than in other contexts. Again, this may be because the HVR-1 in these Internal Insertion HCVpp constructs is the most conformationally labile and therefore less subject to the phenotypic effects of the I438V A524T mutations.

4. 7. Concluding Remarks

Based on the data presented in chapter 3, the following hypothesis was made: length polymorphisms in the HVR-1 may alter the phenotype of HCV; the change in phenotype may be caused by modulating the occupancy rate of the conformational poses the HVR-1 adopts. The aim of this chapter was to examine the first half of that hypothesis through molecular virology. Having examined the data presented herein, it can be stated that yes, length polymorphisms in the HVR-1 do alter the phenotype of HCVpp. However, unexpectedly, the data presented in this chapter also begins to provide clarity on the second half of the central hypothesis – the ability to pocket the HVR-1 provides a rational explanation for the HCVpp phenotypes.

Length polymorphisms do alter phenotype, and they do so in a position-dependent manner that was not expected at the outset. Insertions at the C-terminus of the HVR-1 lead to a distinct phenotype compared to length polymorphisms at the N-terminus. Additionally, insertions and deletions at the N-terminus of the HVR-1 led to two distinct phenotypes.

4. 7. 1. Infectivity

In discussing the infectivity of the length polymorphism-containing HCVpp; in Chapter 3, it was demonstrated that shorter HVR-1 sequences are less abundant than lengthened HVR-1 sequences. It was posited that this was due to a shorter HVR-1 being deleterious to the infectivity of HCV. The infectivity data for the wildtype J6-backbone HCVpp (figure 4. 2.) would support that assertion in the HCVpp system. The wildtype J6-backbone N-Terminus Deletion HCVpp were found to be less infective than either the N-Terminus Extensions or the Internal Insertion HCVpp in the same backbone. Conversely, the N-Terminus Extensions and the Internal Insertion HCVpp were found to be as infective as both the wildtype J6 and I438V A524T J6 HCVpp.

To date, only one other paper has looked at how insertions in the HVR-1 would alter infectivity of HCVcc (Olesen *et al.*, 2025). Notably, they did not find any HVR-1 sequences in their dataset which were shortened. They first amplified and subcloned the E1E2 sequence from 131 HCV patient samples and performed Sanger sequencing on select patient samples. They observed that 3 % of the HVR-1 sequences contained insertions. To evaluate whether these insertions had any effect on infectivity, they swapped the full patient-derived HVR-1 sequence into the closest by homology JFH1-based Core-NS2 HCV recombinant. For example, the genotype 1a HVR-1 sequences were swapped into H77 which is also genotype 1a. Furthermore, they then isolated the individual inserted residues and cloned those specific residues into the appropriate site in the HVR-1. For example, one genotype 1b sequence had an insertion of SRK at the N-terminus of the HVR-1. Therefore, they inserted SRK into the HVR-1 from H77 at the N-terminus. They found that, broadly, in the full HVR-1 swap chimeras that infectivity was reduced or not affected, and that in the HCVcc with specific residues added – H77 with SRK residues for example - that infectivity was also reduced. However, they do note that this decrease in infectivity is likely caused by the chimeric nature of the HCVcc and not the insertions themselves (Olesen *et al.*, 2025).

Whilst the work presented here and the work performed by Olesen *et al.* are not a perfect mirror of one another i.e. the same experiments performed with

HCVpp and HCVcc, it does pose an interesting question: would performing these exact same experiments with the same sequences yield comparable results in the HCVcc system? In general, their results showed either no change or reduced infectivity for their insertions which differs ever so subtly from the data presented here wherein any extension to the HVR-1 in these experiments, statistically, results in no change compared to wildtype when in the wildtype J6 backbone i.e. no reduction in infectivity is seen when just adding residues into the HVR-1.

However, when comparing the infectivity of the HCVpp in the wildtype J6 backbone and the I438V A524T J6 backbone, the N-Terminus Deletions, and the Internal Insertion HCVpp exhibit similar phenotypes in both backbones. On the other hand, there was a reduction in infectivity for the N-Terminus Extensions in the I438V A524T J6 backbone. These data also hint at something Olesen *et al.* concluded in their paper: the function of the HVR-1 is determined not only by its own sequence alone but also its relationship to the rest of the E2 sequence.

4. 7. 2. Receptor Dependency

We recapitulated that CD81 and CLDN-1 are absolutely essential for HCV entry irrespective of HVR-1 length. Furthermore, the majority of the HCVpp assessed were found to be OCLN dependent. Notably, however, the I438V A524T J6-backbone Internal Insertion HCVpp appear to be OCLN independent; this may present a tool for further investigations on the role of OCLN in HCV entry.

Whilst the Huh7 LDLR KO and Huh7 SR-B1/LDLR DKO were permissive to infection, the results obtained were inscrutable in characterising the effects of length polymorphisms in the HVR-1 on receptor usage. The infections of Huh7 LDLR KO cells, matched our expectations; HCVpp do not have associated apolipoproteins and cannot engage this receptor, therefore LDLR knock out should not effect HCVpp entry (Monazahian *et al.*, 1999; Germi *et al.*, 2002; Owen *et al.*, 2009; Albecka *et al.*, 2012).

The results for the Huh7 SR-B1/LDLR DKO cell line were, on first consideration, unexpected. It was hypothesised that the knocking out of LDLR would make no difference, as seen in the Huh7 LDLR KO cells. Therefore, it was hypothesised that perhaps infecting the Huh7 SR-B1/LDLR DKO cells with the HCVpp would recapitulate the infectivity phenotypes for the HCVpp when infecting Huh7 SR-B1 KO cells. This was not the case. In returning to the literature, one other paper to date has compared the infectivity of HCVpp in Huh7 SR-B1/LDLR DKO cells (Zapatero-Belinchón *et al.*, 2021). This study used multiple Huh7 SR-B1/LDLR DKO cell lines edited with CRISPR-Cas9 technologies. Critically, they found that there was a difference in infectivity for the same HCVpp across their different Huh7 SR-B1/LDLR DKO cell lines. This may point to unknown off-target effects, or artifacts of single cell clonal expansion, generating variability in how permissive the cells are to HCVpp. Therefore, we concluded that Huh7 SR-B1/LDLR DKO cell lines cannot reliably be used to characterise the effects of length polymorphisms in the HVR-1. Again, as noted, off-target effects may be present in each of the Huh7 KO cell lines but each cell line was internally consistent, and an investigation into off-target effects in these cell lines would only help in, perhaps, elucidating why there is a synergistic effect in knocking

out both SR-B1 and LDLR in reducing the entry of HCVpp when HCVpp can only interact with SR-B1 and not LDLR.

In the Huh7 SR-B1 single KO context, however, it was possible to examine the effects of length polymorphisms in the HVR-1 on SR-B1 dependency. The wildtype J6-backbone N-Terminus Deletions were more infective in the Huh7 SR-B1 KO cells than in Huh7 WT cells. The wildtype J6-backbone N-Terminus Extensions were approximately 50 % as infective, and the wildtype J6-backbone Internal Insertions were as infective in either cell line. Moreover, none of the I438V A524T J6-backbone HCVpp exhibited altered infectivity in the Huh7 SR-B1 KO cells.

One of the more striking observations from the receptor dependency assay was that the wildtype J6-backbone N-Terminus Deletion HCVpp were more infective upon the removal of SR-B1. At first glance, this appears counterintuitive. However, it was shown in the 8RJJ E1E2 complex structure that a pocketed HVR-1 resulted in an E2 conformation which disfavoured engagement of the CD81 (PDB: 8RJJ, Kumar *et al.*, 2021, 2023; Augestad *et al.*, 2024). We hypothesise that a shortened HVR-1 may be unable to pocket, as the HVR-1 would simply not be long enough to reach the HVR-1 binding cleft and form intramolecular interactions. Therefore, if the wildtype J6-backbone N-Terminus Deletion HCVpp are unable to pocket their HVR-1, they may be better able to engage CD81 directly.

This structural interpretation also finds support in a quantitative modelling framework. Previous work by Kalemera *et al.* (2019) sought to build a mathematical model for the acquisition of CD81 either intrinsically or through the initial engagement of SR-B1. They found that there was a low intrinsic rate of CD81 engagement in their mathematical model. They suggested that this would reflect the rare spontaneous sampling of a CD81-accessible conformation, rather than a truly independent binding pathway (Kalemera *et al.*, 2019b). Integrating these structural and functional observations with the present data, it may be that this low level of intrinsic CD81 engagement may reflect the conformation state of the HVR-1.

Altering the length of the HVR-1 may alter the conformational states that the HVR-1 is able to sample. As noted, movement in the HVR-1 is predicted to be transduced across the E2 protein (Stejskal *et al.*, 2022). Altering the rate of conformational switching in the HVR-1 may therefore have downstream consequences for the ability of the E2 protein itself to adopt certain conformations. In turn, this may alter the rate of intrinsic CD81 engagement. This may also explain why the N-Terminus Deletion HCVpp have an increased infectivity in the Huh7 SR-B1 KO cells.

For the wildtype J6-backbone N-Terminus Extensions, it was found that they were SR-B1 dependent, but less so than the wildtype J6 HCVpp. For the N-Terminus Deletion constructs, SR-B1 independence was hypothesised to be caused by an inability to pocket. Here, the N-Terminus Extension HCVpp constructs may represent an intermediate state in which HVR-1 pocketing is reduced rather than ablated. The increased length of the HVR-1 may shift the conformational equilibrium towards un-pocketed states, increasing the rate of intrinsic CD81 engagement and thereby reducing SR-B1 dependency without eliminating it entirely.

It was hypothesised that the wildtype J6-backbone Internal Insertion HCVpp contained an HVR-1 which should be able to pocket. However, there was no difference in infectivity for the wildtype J6-backbone Internal Insertion HCVpp in either the Huh7 WT cells or the Huh7 SR-B1 KO cells – thus, these HCVpp are SR-B1 independent.

The distinct phenotypes found for the N-Terminus Extension and the Internal Insertion HCVpp may be understood through a simple mechanical analogy. Consider a rigid structure attached to a fixed point like a lever attached to a wall. Adding length to the distal end of the lever will alter its dynamics and the conformational states it can explore, but the fundamental constraint remains at the point of attachment to the rigid structure. By contrast, introducing flexibility directly at the point of attachment i.e. inserting flexible glycine residues at the base of the HVR-1 may remove the primary constraint on motion entirely.

In turn, this may explain why the wildtype J6-backbone Internal Insertion HCVpp constructs achieve full SR-B1 independence whilst the wildtype J6-backbone N-Terminus Extension constructs retain partial SR-B1 dependency. The Internal Insertions may be increasing the rate of HVR-1 conformational sampling. Whilst the wildtype J6-backbone N-Terminus Extension HCVpp constructs may also be modulating the rate of conformational sampling, this is likely to be less pronounced.

Considering the results of the I438V A524T J6-backbone length polymorphism-containing HCVpp, it was found that there was no change in infectivity for these HCVpp in the Huh7 SR-B1 KO context. The I438V A524T mutations have previously been shown to induce SR-B1 independence (Stejskal *et al.*, 2022), therefore the simple explanation is that the I438V A524T mutations, regardless of the length of the HVR-1, continue to induce SR-B1 independence.

However, as discussed for the wildtype J6-backbone N-Terminus Deletion HCVpp there was an increase in infectivity for these HCVpp in the Huh7 SR-B1 KO context. The proposed explanation for this was that the inability to pocket may disfavour SR-B1 engagement and may favour intrinsic CD81 engagement. That increase in infectivity is not seen in these experiments. It can be suggested, therefore, that the combination of the I438V A524T mutations and a shorter HVR-1 might be disfavouring the intrinsic engagement of CD81. Although, this would appear to contradict the data presented by Stejskal *et al.* (2022) who deduced that the estimated CD81 acquisition via both the SR-B1-mediated and intrinsic routes was enhanced by the I438V A524T mutations.

4. 7. 3. Polyclonal Neutralisations

When looking at the prior structural and experimental work, a relationship between the conformational space that the HVR-1 was able to explore, SR-B1 dependency, and how this related to the "open" or "closed" state of the E1E2 complex i.e. the degree of antibody sensitivity of the E1E2 complex (Augestad *et al.*, 2020, 2024; Stejskal *et al.*, 2022) can be formed.

Distilling the key messages from those papers, and the E1E2/E2 alone structure predictions presented in chapter 3, the following hypothesis can be derived: an HVR-1 which can stably pocket, can stably hold the E1E2 complex in the "closed" conformation. In turn, this leads to antibody resistance and SR-B1 dependence. It was hypothesised that changes in the length of the HVR-1 would alter the ability to pocket and therefore alter antibody sensitivity by shifting the equilibrium between the "open" and "closed" states of E2 protein.

Consequently, it was hypothesised that the shorter HVR-1s found within the N-Terminus Deletion HCVpp in both backbones would not be able to stabilise the "closed" E1E2 conformation, through not being able to pocket their HVR-1 and would therefore be antibody sensitive – this was demonstrated to be the case (Fig. 4. 11. and fig. 4. 12.). However, as noted when discussing the neutralisation of the N-Terminus Deletion HCVpp in the I438V A524T backbone, the curves obtained for these HCVpp did not group as they did in the wildtype J6 backbone. As noted, this may reflect the inherent instability of the I438V A524T J6-backbone N-Terminus Deletion HCVpp and therefore difficult to interpret meaningfully.

Furthermore, it was hypothesised that extending the HVR-1 at the N-terminus would produce HCVpp which could pocket their HVR-1 but perhaps less stably than wildtype J6 due to additional glycine residues and an increase in local entropy. It was therefore hypothesised that these would transition more often from the "open" to "closed" state and may even struggle to place the HVR-1 back into the pocket to induce the "closed" state once it was removed. In terms of neutralisation, this was supported by the data. Whilst the N-Terminus

Extensions were the most antibody resistant of the HCVpp constructs in the wildtype J6 backbone, they were more antibody sensitive than wildtype J6 HCVpp.

However, there is a stark difference between the neutralisation profiles of the I438V A524T J6-backbone N-Terminus Extensions and the wildtype J6-backbone N-Terminus Extensions. The latter were comparatively antibody resistant, and this could be attributed to these HCVpp being able to pocket their HVR-1 which could result in the "closed" E1E2 complex conformation.

However, in the I438V A524T J6 backbone, the N-Terminus Extensions are comparatively more antibody sensitive, with all these constructs being ~50 % neutralised at even the lowest concentration of antibody. Moreover, when examining the IC50 values in figure 4. 12. D., the solid horizontal line represents the lowest concentration of antibody tested in this assay – none of the bars associated with the I438V A524T J6 N-Terminus Extensions reach past this line, indicating that the IC50 falls below the lowest antibody concentration tested, meaning these HCVpp are hypothetically effectively neutralised before the titration of the IgG even begins. It was previously demonstrated by Stejskal *et al.* that the I438V A524T J6 construct was also highly antibody sensitive. Therefore, the simplest explanation of that the introduction of the I438V A524T mutations into the N-Terminus Extension HCVpp constructs are sufficient to render the HCVpp antibody sensitive. This does imply that the phenotypic effects of the I438V A524T mutations dominate the effects of extending the HVR-1 at the N-terminus.

For the Internal Insertion HCVpp in both backbones, it was thought that the increase in flexibility at the base of the HVR-1 through the insertion of up to 7 glycine residues would liberate the HVR-1 from its conformational constraints. Therefore, these HVR-1s may never pocket despite the hypothetical ability to do so. Again, this was borne out in the data presented here wherein the Internal Insertions HCVpp in both backbones were well neutralised. Therefore, the HVR-1 may not to be pocketing often, if at all, for these HCVpp and that leads to SR-B1 independence and antibody sensitivity.

4. 7. 4. Monoclonal Neutralisations

To summarise, the J6.36 antibody neutralisation results gave a clearer view on HVR-1 exposure directly. The AR3C antibody elucidated a deeper understanding of global front layer accessibility that could be modulated by movement in the HVR-1 and conformational transduction. The AP33 antibody facilitated examination of AS412 conformation independently of HVR-1 pocketing. Lastly, the AR4A reports most directly on which HCVpp were HVR-1 pocketing-competent and provided hints as to the frequency of HVR-1 pocketing through steric occlusion. Altogether, these data represent a multi-pronged approach towards a better understanding of how the HVR-1 aids HCVpp in resisting neutralisation.

In considering each of the antibodies in turn, J6.36 was the first antibody evaluated against the length polymorphism-containing HCVpp. The J6.36 antibody binds to the HVR-1 directly (Sabo *et al.*, 2011; Cowton *et al.*, 2018). Therefore, the results from this assay are a proxy for HVR-1 exposure. The strongest conclusion that can be drawn from this work that has not already been stated is that if the I438V A524T mutations are stabilising the HVR-1, they are stabilising the HVR-1 in a conformation that is exposed and may sit outwith the pocket at the stem region.

However, in considering once more the 8RJJ E1E2 structure, and the results obtained in figure 4. 13., this may be an overinterpretation of the data presented (Augestad *et al.*, 2024). The 8RJJ E1E2 structure does not explicitly negate the hypothesis that even whilst pocketed the HVR-1 could be sufficiently exposed for J6.36 to bind. Additionally, it should be noted that in chapter 3 the HVR-1 was predicted to either be “free-floating”, pocketed, or bound under the CD81 binding loop. Therefore, it may be that the I438V A524T mutations stabilise the HVR-1 under the CD81 binding loop. However, to answer these questions, more research would be required involving a larger panel of length polymorphism-containing HCVpp, or HCVpp with wildtype length HVR-1s which have their ability to pocket ablated in other ways.

As noted, the AR3C antibody allowed for a deeper interrogation into the conformational dynamics of the front layer and how length polymorphisms in the HVR-1 may affect epitope exposure in the front layer by altering the equilibrium of the “open” and “closed” E2 conformational states. Firstly, these data presented here agree with the model of HVR-1-mediated antibody resistance as posited by Augestad *et al.* (2020). The wildtype J6-backbone -3 N-Terminus Deletion HCVpp were well neutralised by AR3C. This possibly suggests that these HCVpp rarely go into the “closed” conformational state, possibly due to the inability of the HVR-1 to pocket. The wildtype J6-backbone +3 N-Terminus Extension HCVpp were AR3C resistant which may correlate with the ability of these HCVpp to pocket their HVR-1 and enter the “closed” E2 state. Lastly, the wildtype J6-backbone +4I Internal Insertion HCVpp are well neutralised which may reflect the increased conformational lability of the HVR-1 at its base. In turn, this may disfavour the adoption of the “closed” conformation of the E2 protein.

However, one caveat to these results is that for the dynamic range for neutralisation for the wildtype J6-backbone -3 N-Terminus Deletion HCVpp, the Internal Insertion HCVpp in both backbones, and the I438V A524T J6-backbone HCVpp was lower than the antibody concentrations evaluated here. These HCVpp have calculated IC50 values lower than the lowest concentration of antibody assessed. Therefore, whilst the trends still hold up to the model, further experimental work would need to be performed in an appropriate range of dilutions to truly understand the dose-response relationship of these HCVpp to the AR3C antibody.

It was predicted that the wildtype J6-backbone -3 N-Terminus Deletion HCVpp may be AP33 resistant. These HCVpp may not be able to pocket their HVR-1, and therefore might not be able to lock their AS412 into the ß-hairpin conformation that AP33 has been shown to bind (Kong, Giang, Nieusma, *et al.*, 2012; Augestad *et al.*, 2024). However, these HCVpp were well neutralised, this suggests that pocketing competency does not modulate AP33 sensitivity in a binary fashion where inability to pocket the HVR-1 renders AP33 inert. Rather, what might be happening is that the structure resolved by Kong *et al.* shows one possible way in

which AP33 can engage with its epitope, that AP33 is able to induce conformational changes in its epitope upon binding, or that even with an HVR-1 which does not pocket, AS412 still switches between different conformations. Given that AP33 can recognise the linear version of its epitope (Section 4. 3.), the first explanation is perhaps more reasonable. Additionally, there is no direct evidence to support directly the other two hypotheses.

Lastly, the AR4A data in figure 4. 14. provides the most direct evidence that the ability to pocket the HVR-1 alters phenotype. Moreover, this is the most convincing evidence that the N-Terminus Deletion HCVpp are pocketing incompetent, because the AR4A epitope occlusion by the pocketed HVR-1 means AR4A sensitivity can be used as a direct structural proxy for pocketing frequency. Importantly, this retroactively validates the mechanistic interpretations made throughout all the preceding assays which were all built on the assumption that these HCVpp could not pocket. However, it should be stated clearly that this is still indirect evidence for the fact that the shorter HVR-1s are pocketing incompetent.

Altogether, the overarching conclusion that can be drawn from these data is that HVR-1 length does modulate antibody sensitivity. Moreover, it is likely that HVR-1 length determines pocketing competence, and in turn, pocketing competence determines the conformational equilibrium of the E1E2 complex. As demonstrated by Augestad *et al.* (2020), this in turn governs antibody sensitivity in a manner that is consistent, predictable, and mechanistically interpretable across multiple independent assays.

4. 7. 5. Hierarchy of Mutations

Given all of the data presented above, another conclusion which can be drawn is that there appears to be a hierarchy of mutations across the length polymorphisms and the I438V A524T mutations.

It was previously discussed when examining figure 4. 3., that there appeared to be a phenotypic hierarchy wherein the Internal Insertion constructs were unaffected by the I438V A524T mutations but that both the N-Terminus Deletions and Extensions were. The N-Terminus Deletions and Extensions were statistically less infective in the I438V A524T J6 backbone, and than in the wildtype J6 backbone. However, the Internal Insertion constructs were found to be as infective in either backbone. The interpretation of these data was that the I438V A524T mutations were able to dominate the phenotypes associated with the lengthening or shortening of the HVR-1 at the N-terminus, but not for the Internal Insertions.

In considering the SR-B1 dependency assay, it was also demonstrated that the I438V A524T mutations did affect those HCVpp with length polymorphisms at the N-terminus. Wherein the I438V A524T mutations ablated the increase in infectivity found for the wildtype J6-backbone N-Terminus Deletion HCVpp and decreased the dependency on SR-B1 in the N-Terminus Extension HCVpp in the I438V A524T J6-backbone. However, there was no change in SR-B1 dependency for the Internal Insertion HCVpp in either backbone wherein both sets of Internal Insertion HCVpp were SR-B1 independent. Therefore, the phenotypes described for the Internal Insertion HCVpp in the I438V A524T J6 backbone cannot be ascribed to either the I438V A524T mutations or the internally inserted glycine residues alone.

For the polyclonal neutralisation assay, the same pattern emerges. In considering the N-Terminus Deletion HCVpp, insertion of the I438V A524T mutations rendered the HCVpp, perhaps, more unstable than in the wildtype J6 backbone. Moreover, for the N-Terminus Extension HCVpp, there was a switch from antibody resistant in the wildtype J6 backbone to antibody sensitive in the

I438V A524T backbone. However, there was no change in phenotype for the Internal Insertion HCVpp in either backbone, both being well neutralised. Again, there was a change in phenotype for those HCVpp with length polymorphisms at the N-terminus of the HVR-1 upon the introduction of the I438V A524T mutations which was not present across the Internal Insertion HCVpp in the two backbones.

Altogether, at this point it was suggested that there was a hierarchy of mutations wherein length changes at the N-terminus of the HVR-1 are altered by the presence of the I438V A524T mutations. Whereas insertion of residues at the C-terminus of the HVR-1 overcomes the phenotypic effects of the I438V A524T mutations.

That being said, the AP33 neutralisation data provide some evidence that the I438V A524T mutations do affect Internal Insertion HCVpp phenotype, albeit less prominently than for N-Terminus Extensions. This was also corroborated by the data for the AR4A neutralisations presented in figure 4. 16. Rather than overturning the hierarchy of mutations hypothesis which has been evidenced very well up until this assay, the AP33 and AR4A neutralisation data offer a refinement of that hierarchy: conformational lability of the Internal Insertion HVR-1s appears to be the dominant phenotypic driver, with the I438V A524T mutations exerting a detectable but secondary effect. Therefore, the hierarchy may still stand wherein the phenotypic consequences of the I438V A524T mutations are less pronounced than those of the Internal Insertions, but they do intersect and cause a synergistic effect on HCVpp phenotype.

4. 7. 6. Final Conclusions

The central hypothesis generated from the results of chapter 3 was as follows: length polymorphisms in the HVR-1 may alter the phenotype of HCV; the change in phenotype may be caused by modulating the occupancy rate of the conformational poses the HVR-1 adopts. The work in this chapter was undertaken to assess the first half of that hypothesis.

Yes, length polymorphisms in the HVR-1 do alter the phenotype of HCVpp with respects to infectivity, receptor dependency, and antibody sensitivity. Moreover, the modulation of phenotype by length polymorphisms in the HVR-1 is position-dependent.

Given the data presented here, a unified picture of how the HVR-1 contributes to HCVpp phenotype can now be considered. The HVR-1 is a disordered and flexible tail (Stejskal, Lees, *et al.*, 2020). However, the HVR-1 is able to pocket into the HVR-1 binding cleft at the stem region of the E2 protein (Augestad *et al.*, 2024). This alters global and local E1E2 conformational dynamics and leads to a "closed" state which is antibody resistant and SR-B1 dependent (Augestad *et al.*, 2020). This may be driven by allosteric effects which are transduced from the HVR-1 across the rest of the E2 protein (Stejskal *et al.*, 2022). However, the HVR-1 is likely not constitutively in the pocket, rather there is an equilibrium of pocketing versus "free-floating" which can be fine-tuned by mutations such as I438V A524T (Stejskal *et al.*, 2022), or indels in the HVR-1 (Chapter 4).

However, whilst this model does explain how modulation of the occupancy rate of pocketing could explain differences in phenotype for the HCVpp, no experimental work in this chapter directly measured this – this can only be inferred from phenotypic readouts. Notably, the AR4A monoclonal neutralisation assay as a proxy for pocketing frequency is the closest any assay described herein comes to a direct examination of pocketing frequency, but it remains indirect and falls short of disentangling pocket occupancy in the HCVpp with insertions in their HVR-1s. However, this question formed the basis for the work to be undertaken in chapter 5.

Chapter 5

Investigating How the HVR-1 Functions

5. Investigating How the HVR-1 Functions

As noted previously, there is a long standing association between the HVR-1 and HCV infectivity, receptor dependency, and antibody sensitivity (Bartosch *et al.*, 2005; Bankwitz *et al.*, 2010; Prentoe *et al.*, 2011, 2019; Stejskal, Kalemera, *et al.*, 2020; Olesen *et al.*, 2021, 2025). However, previous studies lack a real mechanistic understanding as to how the HVR-1 regulates these characteristics. Having already investigated the relationship between length polymorphisms in the HVR-1 and the phenotype of HCVpp, the mechanism through which these different phenotypes are brought about was now investigated.

It was discovered that shorter HVR-1s were less dependent on SR-B1 as an entry factor. However, the exact interaction between the HVR-1 and the SR-B1 is, as yet, poorly understood. To begin to answer this question, we returned to machine learning-enabled protein folding software and co-folded the E1E2 complex from two different prototypical lab strains of HCV (J6 and H77) with SR-B1 to examine how ColabFold would predict how the E1E2:SR-B1 complex might form.

As noted in Chapter 4, the phenotypes of the HCVpp were similar within length polymorphism types i.e. all the N-Terminus Deletions had a similar phenotype regardless of the exact length of the HVR-1 etc. Notably, it was also discovered in Chapter 4, that the N-Terminus Deletion HCVpp were not dependent on SR-B1 for viral entry. Therefore, to further investigate the relationship between the HVR-1 length and SR-B1, soluble E2 proteins containing select HVR-1 length polymorphisms in both the wildtype J6- and I438V A524T J6-backbones were produced, purified, and used in an SR-B1 binding assay. This allowed a more nuanced view of SR-B1 usage versus binding.

Additionally, the ColabFold models seen in Chapter 3 (Section 3. 1.) predicted that the HVR-1 would pocket into the HVR-1 binding cleft which was later validated in the E1E2 dimer of dimers resolved structure (PDB: 8RJJ, Augestad *et al.*, 2024). We hypothesised that the inability to pocket HVR-1 could be

determining the distinct SR-B1 dependency phenotypes observed in the N-Terminus Deletion HCVpp. To further investigate the relationship between the ability to pocket the HVR-1, SR-B1 dependency, and the ability to bind SR-B1, Dr Michael Oliver used the HVR-1 sequence information extracted from the 9,103 HVR-1 sequences (Section 3. 3.) to generate mutant constructs wherein the HVR-1 length had not been altered but which should disfavour pocketing of the HVR-1. He then used these constructs to investigate the relationship between pocketing, SR-B1 usage, and SR-B1 binding.

At this point, there are multiple strands of evidence which pointed to the fact that the HVR-1 was not merely a disordered tail which would randomly sample space around the E1E2 complex. Rather, the HVR-1 may be adopting several poses and moving between them. These different poses may result in E1E2 complexes with transiently different phenotypes. Consider the “open/closed” E2 phenotype as discussed by Prentoe *et al.* wherein they posit that movement within the HVR-1 induces global conformational changes across the E2 protein and can modulate its antibody sensitivity (Prentoe *et al.*, 2019). Additionally, the I438V A524T mutations have been discussed at length already throughout this thesis, but in short, the predicted reduction in HVR-1 flexibility induced by the I438V A524T mutations altered HCVpp phenotype. In that paper, they look at how motion in the HVR-1 is transduced across the E2 as a whole and found that the I438V A524T reduced the transduction of motion across the protein (Stejskal, Kalemera, *et al.*, 2020). Altogether, these hypotheses and papers suggest that the movement of the HVR-1 is important for its function. However, to date, its movement has not yet been directly measured, although we have indirect evidence of HVR-1 disorder due to the lack of resolution in E2 crystal structures. To this end, in the final component of experimental work, I aimed to produce an smFRET system for measurement of movement within the HVR-1. Initial steps were made but due to time constraints, this part of the project did not achieve its aims.

Altogether, this last section of the project aimed to provide some mechanistic understanding of how the HVR-1 contributed to the phenotypes reported in Chapter 4 bioinformatically, indirectly, and directly.

5. 1. Predicting the HVR-1:SR-B1 Interface

SR-B1 is a member of the CD36 family of proteins which, in general, have roles in lipid metabolism and ligand binding, and contain a channel through the centre of the protein which is thought to play a role in lipid transport. In 2013, the structure of **Lysosomal Integral Membrane Protein 2** (LIMP-2) was resolved using x-ray crystallography (PDB: 4F7B, Neculai *et al.*, 2013). LIMP-2 is a related protein to SR-B1, both belonging to the CD36 superfamily. LIMP-2 binds **ß-glucocerebrosidase** (ß-GCase) and mediates its delivery to the lysosome. Their work mapped down the interaction between LIMP-2 and ß-GCase to the apical tri-helical bundle. This apical tri-helical bundle is found across LIMP-2, CD36, and importantly SR-B1 so Neculai *et al.* posited that this conserved apical tri-helical bundle may be important for each of these proteins and binding to their respective ligands.

Further work supported this hypothesis, and moreover, showed that this helical bundle can be co-opted by pathogens to propagate infection. In 2016, Hsieh *et al.* solved an x-ray crystal structure of the PfEMP1 protein from *Plasmodium falciparum*, a causative agent of malaria, binding to the tri-helical bundle on CD36 (PDB: 5LGD, Hsieh *et al.*, 2016). In 2019, a cryo-EM structure of LIMP-2 was released in complex with **enterovirus 71** (EV71) – a causative agent of hand, foot, and mouth disease – and once again, the virus was bound to LIMP-2 via interaction with the tri-helical bundle (PDB: 6I2K, Zhou *et al.*, 2019).

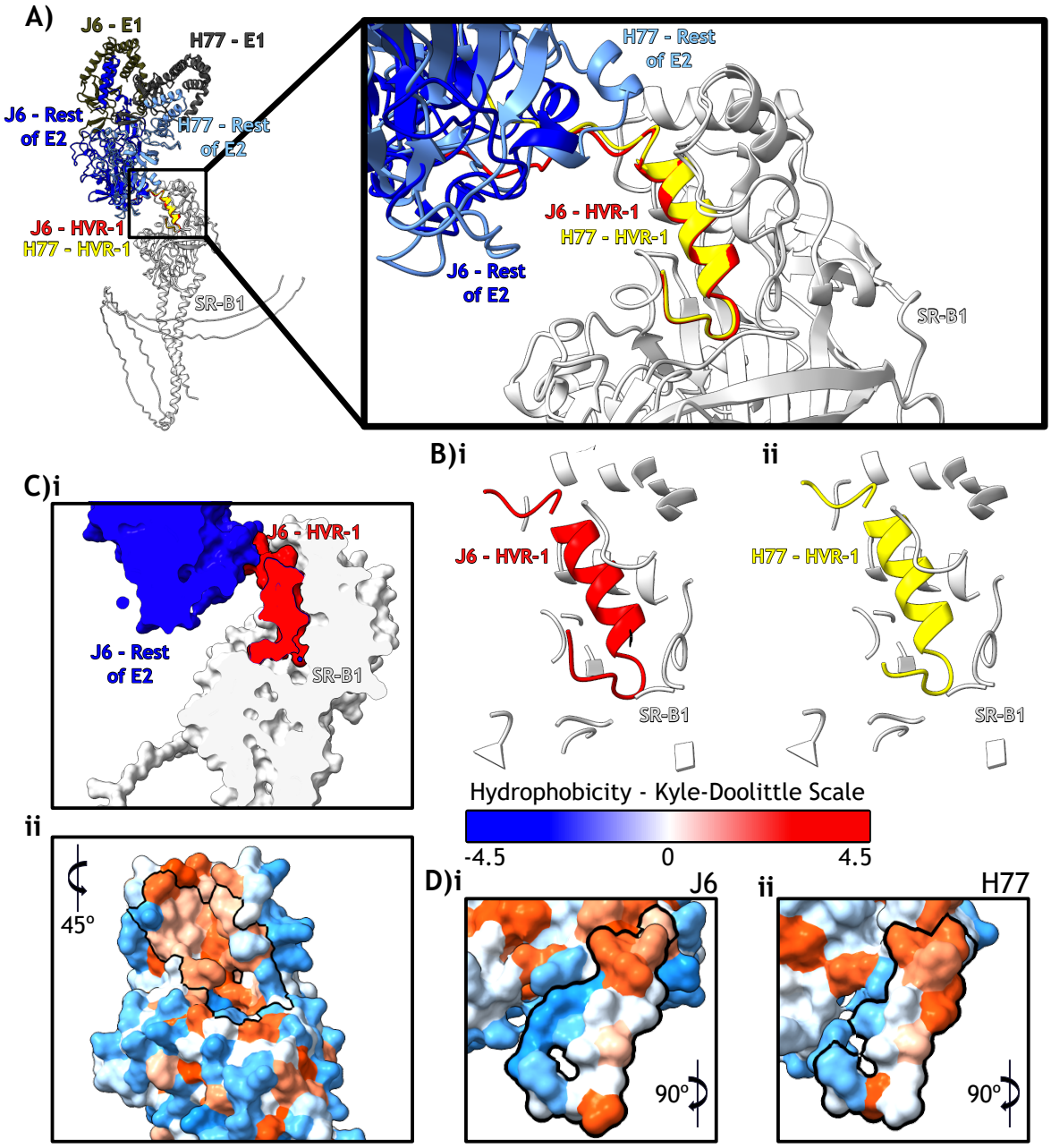


Figure 5. 1. **Predicted E1E2:SR-B1 Interactions.** Using ColabFold, the interactions between the E1E2 dimers of both J6 and H77 with SR-B1 were predicted. A) Both complex models were structurally aligned based on SR-B1, and despite differences in the overall E1E2 3D positions, the HVR-1s are positioned homologously, as shown by the punch in. B) i Slice through of the surface punch in, showing the HVR-1 positioned in the SR-B1 lipid transfer channel. B) ii Surface representation of the transfer channel entrance, coloured by residue hydrophobicity. The entrance is highlighted in black. Model rotated right 45° relative to A). C) i and ii HVR-1:SR-B1 interacting residues. Residues involved in the interaction had Van der Waals overlaps ≥ -0.4 Å. C) i and ii Surface hydrophobicity representation of the HVR-1 - outlined in black - from both J6 and H77 turned left 90° relative to B). Together B), C), and D) predict an interaction between the HVR-1 with the lipid transfer channel found in SR-B1 which may be caused by hydrophobic forces bringing the hydrophobic face of the HVR-1 into the hydrophobic SR-B1 channel.
NB: The H77 E1E2:SR-B1 prediction was performed by Dr Michael Oliver.

Concerning SR-B1, in 2020, work by Langlois *et al.* mapped the key domain in SR-B1 which facilitates productive *Plasmodium* infection through human/mouse SR-B1 chimera studies and mapped down to the tri-helical bundle as being a key determinant for successful infection (Langlois *et al.*, 2020).

The consistency of the findings of these studies – that ligands are binding to the apical tri-helical bundles of CD36 family member proteins - may suggest that the tri-helical bundle of SR-B1 is where the HVR-1 may be interacting, if indeed they interact directly. To investigate this, SR-B1 was modelled in complex with E1E2 from both J6 and H77 strains of HCV by ColabFold and the Rank 1 models were investigated further in ChimeraX v1.9. (Fig. 5. 1). Whilst there are differences between the exact 3D positioning of the E1E2 complexes in relation to one another, the predicted position of the HVR-1 for both J6 and H77 is identical, and moreover, the residues with which the HVR-1s are predicted to interact are identical despite their sequence divergence (Fig. 5. 1. Bi, ii). Compared to other complex structures - such as PDB: 5LGD and 6I2K which place the plasmodium interaction at the very apical tip of SR-B1 – ColabFold predicts that HVR-1 is actually positioned in the channel which spans the length of SR-B1 (Fig. 5. 1. Ci). Considering the entrance to this channel, shown in figure 5. 1. Cii, represents a local area of increased hydrophobicity, the hydrophobicity of HVR-1 was then considered. Figure 5. 1. Di, ii is a view of J6 and H77 HVR-1 from the entrance of the channel and it shows the hydrophobic face of the HVR-1 pushing in towards the channel. This would suggest that this interaction is being mediated by the coming together of the hydrophobic face of the HVR-1 into the SR-B1 hydrophobic lipid transfer channel.

This is, of course, a prediction and therefore should be considered to represent a biological possibility rather than definitive biological reality. Moreover, these are two predicted structures that show an entirely unique interface with SR-B1 when compared to other CD36 family proteins and their ligands. However, if this model is true, hypotheses can be made regarding how length polymorphisms in the HVR-1 may disrupt or promote this interaction.

Shortening the HVR-1 may reduce SR-B1 binding as there would be fewer residues to enter the SR-B1 channel and form interactions with the inner channel residues and stabilise the E2:SR-B1 interaction. Conversely, increasing the length of the HVR-1 may have the opposite effect and increase SR-B1 binding. However, this may not be the full picture; the HVR-1 may also need to maintain local hydrophobicity to drive interaction with the SR-B1 channel. Breaking this down further, the N-Terminus Extensions may increase the binding of E2 to SR-B1 by increasing the number of residues in the SR-B1 channel, anchoring the two proteins together through increased residue:residue interactions.

Considering the Internal Insertion constructs, glycine residues have been added to the C-Terminus end of the HVR-1 which is outside of the SR-B1 channel in the ColabFold model, so may not improve E2:SR-B1 binding directly. However, the glycine residues would increase the flexibility of C-Terminal region of the HVR-1. Some intrinsically disordered proteins are proposed to work by "fly-casting" for their binding partner which can increase their binding efficacy for their ligands (Mollica *et al.*, 2016). Increasing the flexibility at the base of the HVR-1 tail may, therefore, increase E2:SR-B1 binding by increasing the odds that a more flexible HVR-1 is better able to "cast" and capture its SR-B1 partner.

5. 2. Interrogating HVR-1:SR-B1 Binding

To investigate this model of E2:SR-B1 binding, sE2 constructs containing length-polymorphisms in the HVR-1 were cloned, expressed, and purified. The purified sE2 proteins were then incubated with CHO cells transduced to express SR-B1 (and GFP to discriminate transduced cells). Using **Fluorescence-activated Cell Sorting** (FACS), the relative binding of sE2 constructs to SR-B1 could be measured.

Select length-polymorphisms were cloned into both a J6 WT sE2 and an I438V A524T J6 sE2 backbone. An overview of the J6 WT sE2 protein architecture and predicted structure can be seen in figure 5. 2. Both the J6 WT and the I438V A524T J6 sE2 constructs are truncated at the C-terminal transmembrane domain to drive secretion. Given the phenotypic grouping exhibited by the different length polymorphism types, a representative from each length polymorphism type was cloned into the sE2 constructs. This led to six different sE2 constructs which could be evaluated for SR-B1 binding: -3 N-Terminus Deletion J6, +3 N-Terminus Extension J6, +4I Internal Insertion J6, -3 N-Terminus Deletion I438V A524T J6, +3 N-Terminus Extension I438V A524T J6, and +4I Internal Insertion I438V A524T J6.

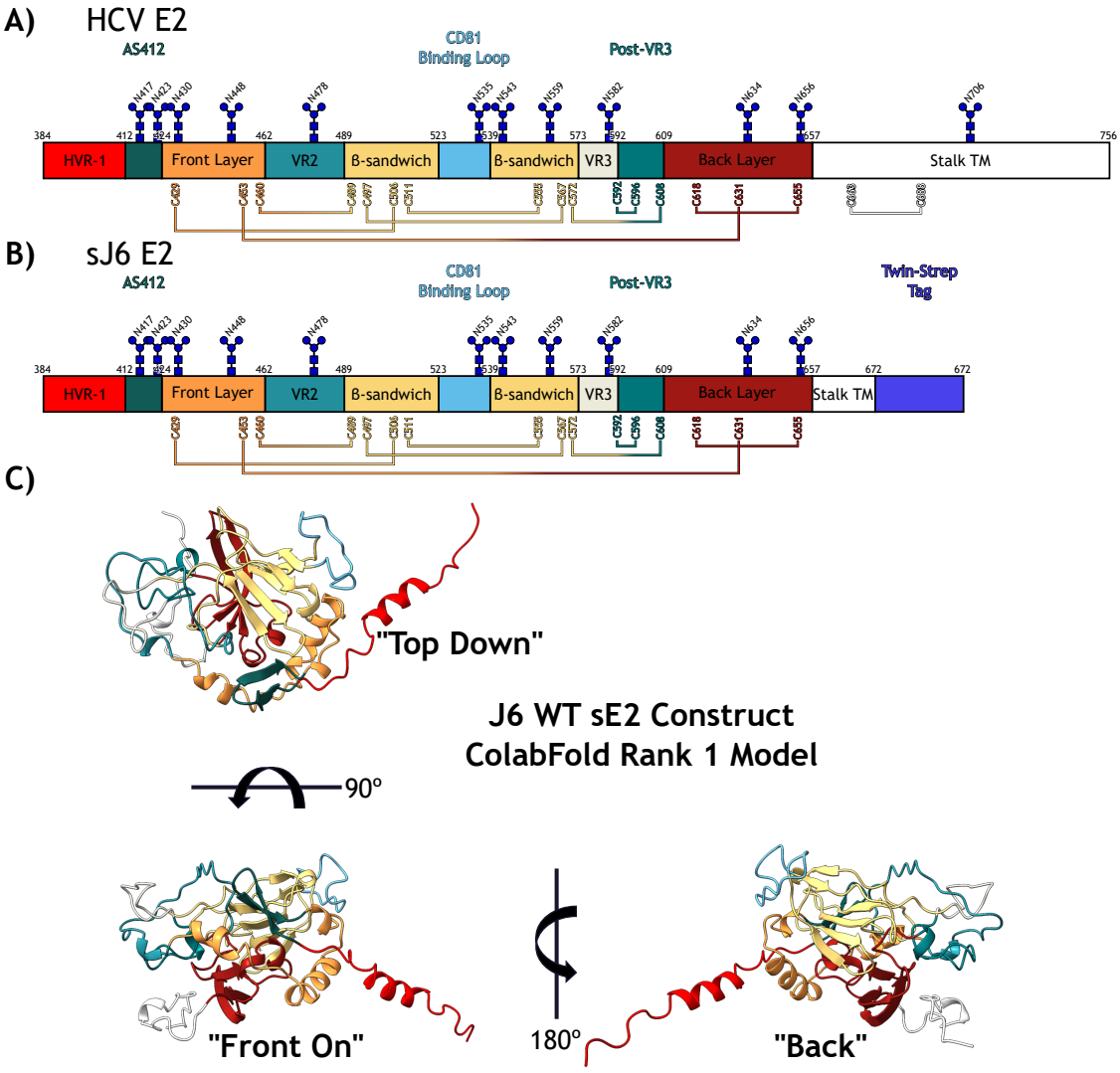


Figure 5. 2. **ColabFold Model of the J6 WT sE2 Protein.** A) Hypothetical full length E2 schematic. B) Schematic of the J6 WT sE2 protein construct. Structures in C) are coloured as per these schematics. C) Rank 1 ColabFold models of the J6 WT sE2 construct. The Twin-Strep tag was not modelled. Notably, the sE2 protein lacks the HVR-1 binding cleft into which the HVR-1 is pocketed, ergo these sE2 constructs cannot pocket the HVR-1.

Five of the six constructs were successfully expressed in HEK293T CD81 KO cells and purified using gravity flow columns and Strep-Tactin®XT 4Flow® resin. However, the -3 N-Terminus Deletion I438V A524T J6 construct could not be purified to sufficiently high concentrations to permit downstream experimental work. Various alanine scanning and high-throughput mutagenesis studies have shown that, in general, is it fairly easy the interfere with the stability and conformation of E2 through single point mutations (W. Wang *et al.*, 2014; Pierce *et al.*, 2016; Gopal *et al.*, 2017). It may be reasoned, therefore, that the combination of the I438V A524T mutations and the HVR-1 truncation were deleterious to the stability of the sE2 protein. Previous work by the Grove lab did note that the I438V A524T mutations within the E2 protein do decrease the stability of the E2 protein, and so this result is not particularly surprising (Stejskal, Kalemera, *et al.*, 2020).

Additionally, the **wildtype J6 sE2** (sJ6**), I438V A524T J6 sE2** (sIVAT), and the **ΔHVR-1 J6 sE2** (sΔHVR-1) proteins were expressed and purified to function as controls. These proteins have known phenotypes with regards to SR-B1 binding and can therefore be used to contextualise the results obtained from the length polymorphism-containing sE2 proteins. In brief, sJ6 binds SR-B1 more than sIVAT, and sΔHVR-1 does not bind SR-B1 (Stejskal, Kalemera, *et al.*, 2020).

Figure 5. 3. contains representative dot blots from the first repeat of the sE2:SR-B1 binding assay. Briefly, CHO cells were transduced to express both GFP and SR-B1. GFP expression functioned as a proxy for SR-B1 expression. The sE2 constructs had a Twin-Strep tag present which could be targeted by the

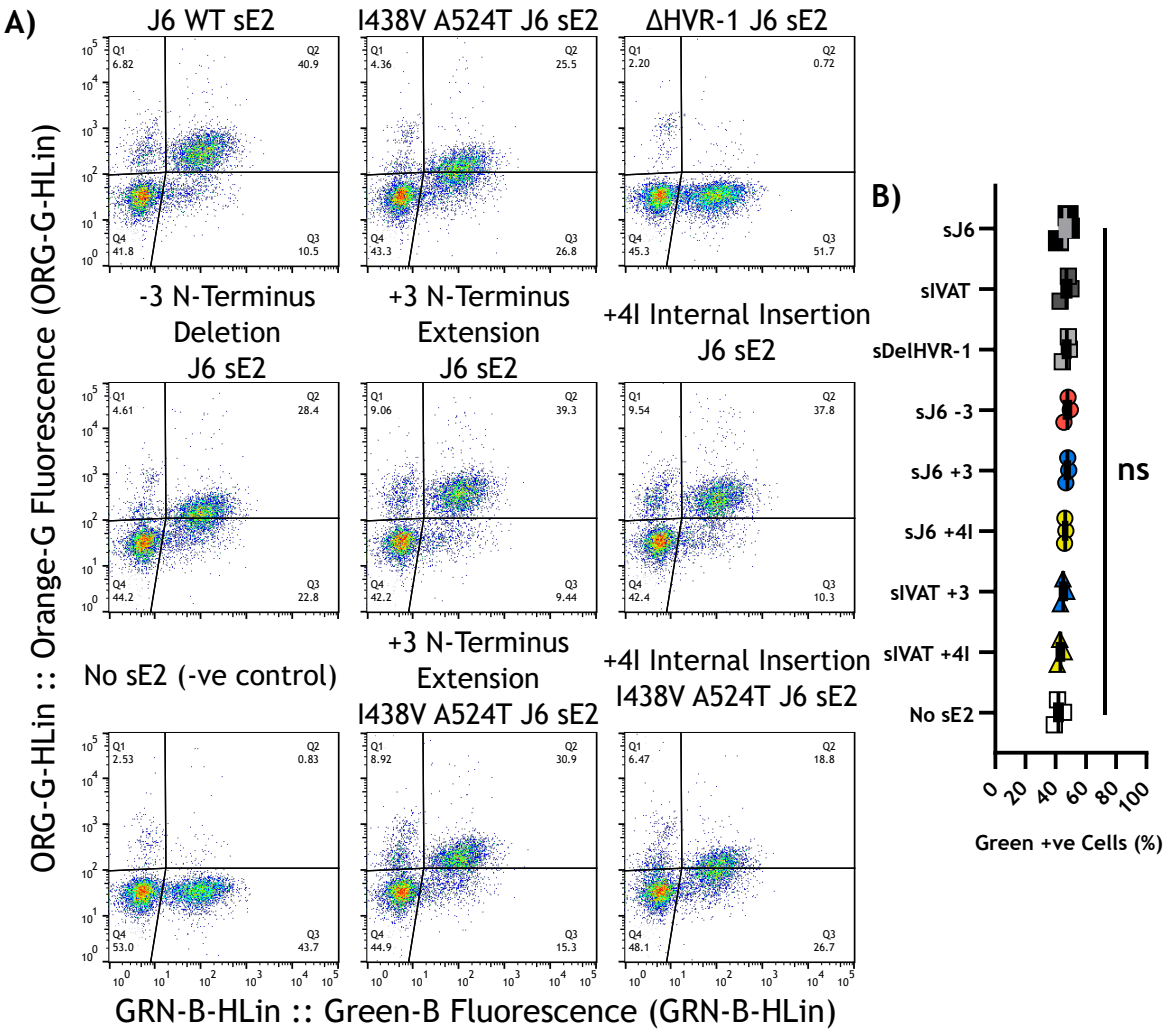


Figure 5. 3. **Representative FACS Dot Plots.** A) Representative dot blots from one repeat of the sE2 binding assay. CHO cells were transduced with a pDual vector containing both GFP and SR-B1, measurement of GFP was a proxy for SR-B1 expression. Across the x-axis is the green fluorescence signal given by the transduced GFP within the CHO cells, allowing SR-B1 expressing CHO cells to be gated appropriately. Across the y-axis is the red fluorescence signal, given by the Alexa Fluor™ 594 secondary antibody which was bound to StrepMAB-Classic, itself bound to the TwinStrep tag on the sE2 constructs. Ergo, green +ve cells represented CHO cells which should be expressing SR-B1, and double +ve cells were those with a sE2 construct bound to SR-B1. The quadrants are as follows: Q1 - red +ve, green -ve, Q2 - red +ve, green +ve, Q3 - red -ve, green +ve, Q4 - red -ve, green -ve. Importantly, the no sE2 control showed no shift in the green +ve population towards Q2. B) Average percentage of green +ve cells for each of the sE2 constructs across three biological repeats of the sE2 binding assay. An ordinary one-way ANOVA was performed to ensure there was no significant difference in the percentage of cells successfully transduced. Therefore, any difference in red +ve cells was due to the sE2 constructs alone and not a result of transduction efficacy.

StrepMAB-Classic primary antibody, which was detected with an Alexa Fluor™ 594 secondary antibody. Therefore, cells counted in Q2 (green +ve, red +ve) are CHO cells expressing SR-B1 with bound sE2 proteins. Notably, each of the controls behaved as expected with 40.9 % of cells in Q2 for sJ6, and 25.5 % and 0.72 % in Q2 for sIVAT and sΔHVR-1, respectively. Moreover, the no sE2 control had 0.83 % of cells in Q2.

In considering the length polymorphism-containing sE2 proteins, each of them was able to bind SR-B1, ranging from 18.8 % to 39.3 % of cells in Q2. However, the percentages of cells counted in a quadrant is a not a true reflection of the brightness of the fluorescence i.e. the number of sE2 proteins which are bound to SR-B1 which is a more valid measurement for binding efficacy comparison.

Before continuing to that analysis, the average total number of GFP +ve cells were counted and compared to ensure that any differences in SR-B1 binding could not be accounted for by differences in transduction efficacy (Fig. 5. 3. B) which was found to be consistent across all three repeats of this experiment.

Next, the geometric mean red fluorescence intensity of green +ve cells was then compared. In figure 5. 4. A, representative cytometry histograms from one repeat of the experiment show the differences in geometric mean fluorescence intensity. As the histogram peaks move to the right, the more intense the red fluorescence given by those green +ve cells is i.e. the more copies of sE2 are bound to the SR-B1. This can be further quantified to look at relationships between the sE2 controls and the length polymorphism-containing sE2 proteins.

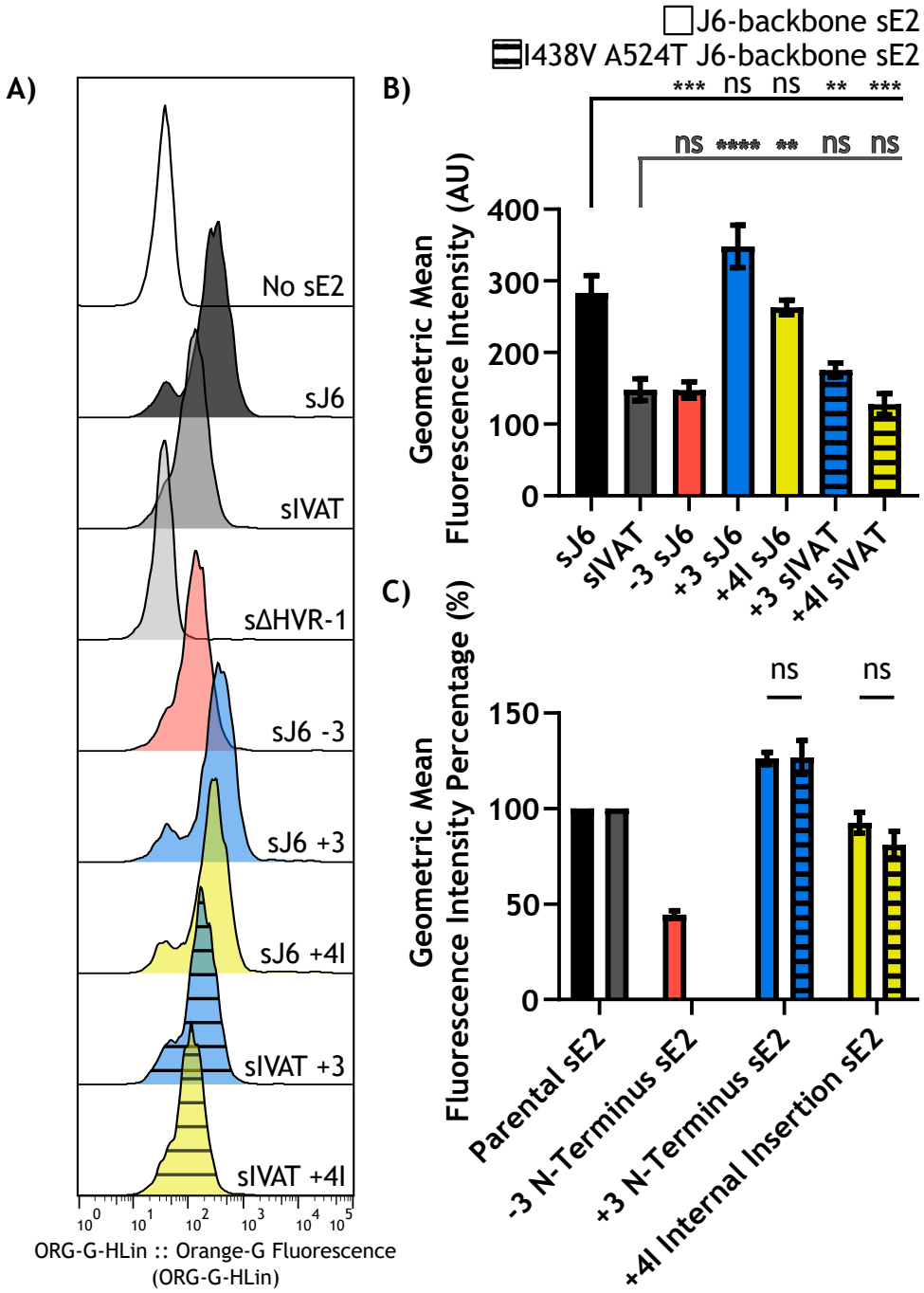


Figure 5. 4. **Investigating the Binding Efficacy of the sE2 Constructs.** A) Representative cytometry histograms of sE2 binding to transduced CHO cells expressing SR-B1 with the no sE2 -ve control. B) Geometric mean red fluorescence intensity of green +ve cells. Ordinary one-way ANOVA tests followed by Dunnett's multiple comparison tests were performed between both of the controls - J6 WT sE2 (sJ6), I438V A524T J6 sE2 (sIVAT) and the mutant constructs (i.e. sJ6 vs constructs and then sIVAT vs constructs). * = p-value <0.05. C) Geometric mean red fluorescence intensity of green +ve cells as a percentage of the parental construct. A 2way ANOVA followed by Šídák multiple comparison tests was performed to evaluate the differences between the percentage fluorescence intensity of the sJ6-based constructs and the sIVAT-based constructs, and no significant difference was found.

Figure 5. 4. B compares the geometric mean fluorescence found across three biological repeats of this experiment. Note that sΔHVR-1 is not shown as it does not bind SR-B1. **-3 N-Terminus Deletion sJ6** (-3 sJ6) has a reduced capacity for binding SR-B1, which is statistically different from its parental sE2, but is not significantly different from sIVAT. In considering the HVR-1:SR-B1 binding model, this would support the hypothesis that if the HVR-1 is being inserted into the SR-B1 lipid transport channel that a reduction in HVR-1 length reduces the stability of this interaction. However, this reduction in SR-B1 binding may also be due to truncations of the HVR-1 leading to a distinct set of resides at the N-terminus of the HVR-1 being inserted into the SR-B1 channel which are disfavoured compared to WT sJ6.

Conversely, **+3 N-Terminus Extension sJ6** (+3 sJ6) has increased SR-B1 binding compared to sJ6 but this increase is not statistically significant. Again, if the HVR-1:SR-B1 interaction prediction is correct, the increase in the length of the HVR-1 may aid in explaining this subtle increase in SR-B1 binding. The additional glycine residues may also favour insertion into the SR-B1 binding channel over WT sJ6 as they are small, and relatively chemically neutral as compared to the first three residues of WT sJ6 which are arginine, threonine, and histidine.

However, the **+4 Internal Insertion sJ6** (+4I sJ6) binds SR-B1 with the same efficacy of sJ6. It was hypothesised that if the predicted HVR-1:SR-B1 interaction was correct, then the internal insertion mutants would have increased binding efficacy to SR-B1 by being able to “fly-cast” their binding partner more efficiently. This does not appear to be the case. However, in considering the HVR-1:SR-B1 binding model, the HVR-1 is predicted to be inserted into a channel. Having an extremely flexible HVR-1 which can move quite freely from its base may disfavour remaining in the channel as much as the increased flexibility enables better “fly-casting”. However, this “fly-casting” hypothesis was not evaluated directly, nor does the current data speak to this, and therefore this is conjecture which requires further experimentation. However, should this model be correct, this may explain why the +4I sJ6 protein is only as effective at binding SR-B1 as sJ6. Alternatively, there have been no changes to residues at the N-terminus of the HVR-1 which inserts into the SR-B1

and without changing these residues, insertion is neither promoted nor disfavoured, hence being similar to WT sJ6 in binding efficacy.

These trends were recapitulated in the I438V A524T J6-backbone sE2 proteins. However, these proteins were, overall, less able to bind SR-B1 than their sJ6-backbone counterparts. The **+3 N-Terminus sIVAT** (+3 sIVAT) protein exhibited increased SR-B1 binding than sIVAT but did not reach statistical significance. The **+4 Internal Insertion sIVAT** (+4I sIVAT) showed a similar SR-B1 binding efficacy when compared to sIVAT. The explanations for these phenotypes may be similar to their counterparts in the sJ6-backbone. This becomes clearer when examining the binding efficacy of the length polymorphism-containing sE2 relative to their respective parental sE2 proteins (Fig. 5. 4. C).

Comparing the relative binding between the different length polymorphisms in either background shows no significant differences between the N-Terminus Extensions, and the Internal Insertions. Both the N-Terminus Extensions are slightly more efficacious at binding SR-B1 than their parental sE2 proteins, though not significantly so. The Internal Insertions are as efficacious as their parental sE2 proteins. Taken together, this suggests that the length polymorphisms in the HVR-1 alter SR-B1 binding in a manner independent of the I438V A524T mutations. Moreover, the I438V A524T mutations are resulting in a reduced capacity for binding SR-B1, implying they are somehow limiting HVR-1 availability to bind SR-B1. Previous work by the Grove lab posited that the HVR-1 was being constitutively stabilised by the I428V A524T mutations and held in apposition to the rest of E2 which may result in this restriction of HVR-1 availability for SR-B1 binding (Stejskal, Kalemera, *et al.*, 2020).

However, none of this can be stated too firmly as there is no direct experimental evidence that E2 and SR-B1 bind through insertion into the SR-B1 channel, and nor does the evidence here support that interaction directly i.e. there is no evidence presented here which demonstrates exactly how the sE2 constructs and the SR-B1 proteins are interacting. That being said, that should not detract from the potential validity of the ColabFold prediction of the HVR-1:SR-B1 interaction which will require further experimentation to determine whether this predicted

interaction reflects true physiological reality. However, it has been shown that length polymorphisms within the HVR-1 do alter SR-B1 binding efficacy. Moreover, the I428V A524T mutations reduce SR-B1 binding in an independent manner to how the length changes in the HVR-1 affect sE2 phenotype.

5. 3. Investigating the Relationship Between HVR-1 Pocketing and SR-B1 Binding

Note: the work conducted in this section was a collaboration that combined my previous HVR-1 sequence analysis (Section 3. 3. Investigating HVR-1 Composition) with experimental work conducted by Dr Michael Oliver.

At this point, there are two linked hypotheses concerning the HVR-1 which have been proposed and indirectly assessed. These hypotheses look to explain the relationship between the HVR-1 and SR-B1.

Firstly, it has been demonstrated that the HVR-1 can bind in apposition into the E2 stem region, termed HVR-1 pocketing herein. This coincides with AS412 and CD81 binding loop conformations which disfavour CD81 binding (PDB: 8RJJ, Augestad *et al.*, 2024). The J6 N-Terminus Deletions showed a decreased dependency on SR-B1 to enter cells and are not predicted to pocket due to having too short an HVR-1 to reach the HVR-1 binding cleft. Therefore, it was hypothesised that the ability to pocket the HVR-1 led to SR-B1 dependency. Moreover, the HVR-1:SR-B1 interaction is required to remove the HVR-1 from the pocket which leads to downstream conformational changes in both AS412 and the CD81 binding loop which would favour CD81 binding and, eventually, viral entry. Secondly, SR-B1 dependency is uncoupled from the ability to bind SR-B1. In considering the wildtype J6-backbone N-Terminus Deletion constructs and the Internal Insertion constructs in both backbones, it has been demonstrated in chapter 4. 4. that these HCVpp have a reduced dependency for SR-B1 as a receptor. However, in the above sE2:SR-B1 interaction experiments, it has been demonstrated that the -3 sJ6 protein and the +4I sE2 proteins in both backbones are able to bind SR-B1. Ergo, the dependency on SR-B1 for HCVpp entry is uncoupled from the ability to bind SR-B1.

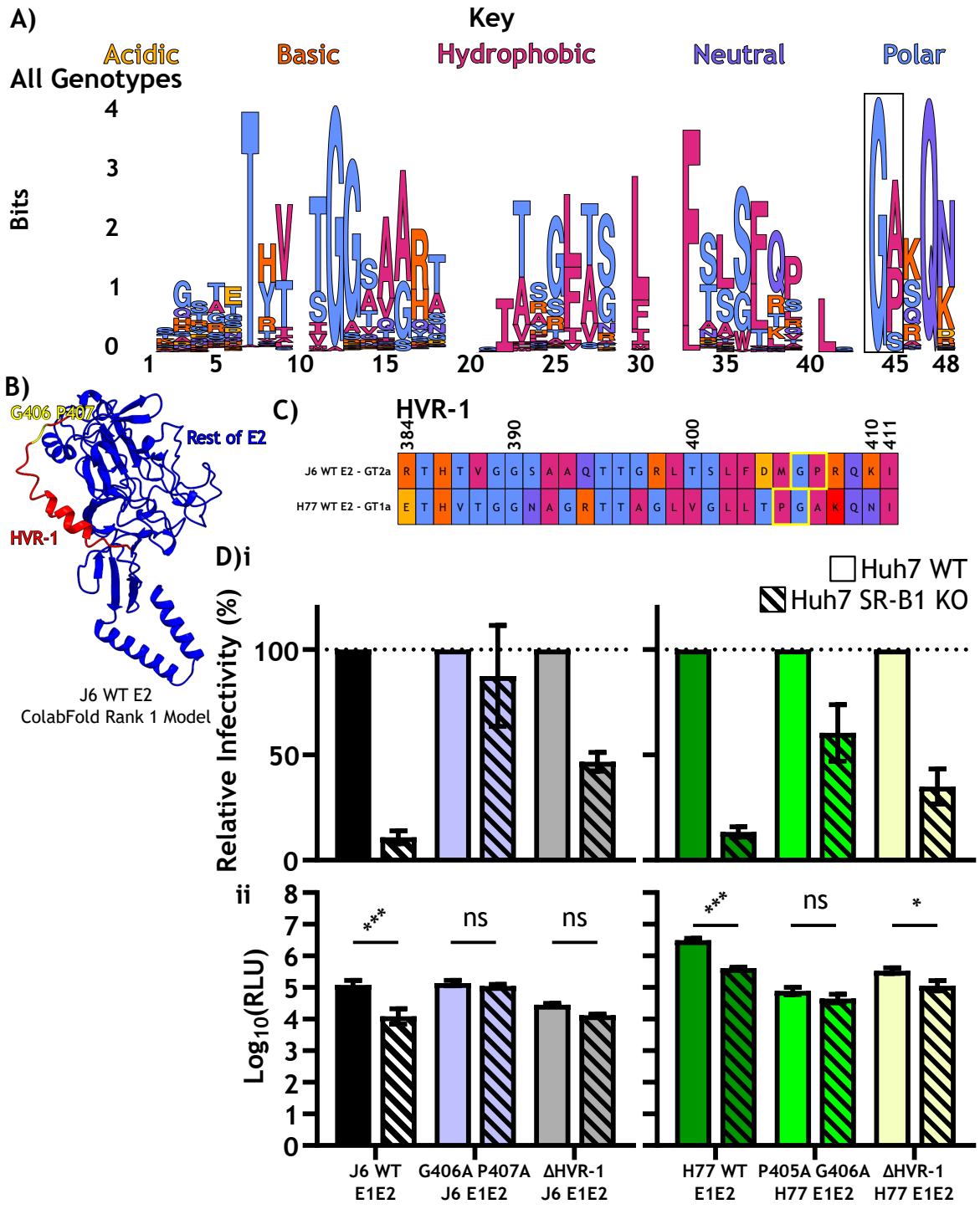


Figure 5. 5. **Mitigating HVR-1 Pocketing.** The experimental work presented in D) was performed by Dr Michael Oliver. A) Bits sequence logo of 9,103 HVR-1 sequences. Residue 44 is a highly conserved glycine, followed by a conserved proline. These residues are near the end of the HVR-1, as seen in B) and C) where these residues are highlighted in yellow in both the ColabFold J6 model structure and the J6 and H77 HVR-1 alignment. D) HCVpp were generated which mutated these helix-breaking residues, which should mitigate HVR-1 pocketing, and then used to infect Huh7 WT and Huh7 SR-B1 KO cells. D) i percentage relative infectivity when comparing Huh7 SR-B1 KO RLU to Huh7 WT RLU. D) ii Log_{10}(RLU) when infecting Huh7 WT and SR-B1 KO cells. A 2way ANOVA followed by Šídák multiple comparison tests was performed between log_{10}(RLU)s obtained in Huh7 WT and SR-B1 KO. P-values; ns = $P > 0.05$, * = $P \leq 0.05$, *** = $P \leq 0.001$. Error bars represent SEM across four technical replicates.

To further investigate the relationship between these two linked hypotheses, that pocketing of the HVR-1 and SR-B1 usage/binding were interlinked, Dr Michael Oliver conducted a series of experiments based on previous analysis of HVR-1 sequences.

Based on the comparison of 9,103 HVR-1 sequences, it was noted that there was a highly conserved glycine residue towards the C-terminus end of the HVR-1, followed by conserved proline residue (Fig. 5. 6. A). Both glycine and proline residues can prevent α-helix formation, and form hinges (Stejskal, Kalemera, *et al.*, 2020). It was hypothesised that these conserved residues may be providing the flexibility/hinging to allow the HVR-1 to pocket into the HVR-1 binding cleft. If these residues were mutated, it may be expected that the resulting construct would be unable to pocket and therefore become SR-B1 independent.

To assess this, we mutated out the two helix-breaking, flexible residues at the C-terminus of the HVR-1 in both J6 and H77. G406 and P407 from J6 and P405 and G406 from H77 were mutated to alanine residues (Fig. 5. 6. C). HCVpp containing these double mutations were then produced and used to infect both Huh7 WT cells and Huh7 SR-B1 KO cells along with their parental HCVpp and a ΔHVR-1 construct of each parental HCVpp (Fig. 5. 6. Di and ii). Figure 5. 6. Dii compares the $\log_{10}$RLUs obtained by each HCVpp in both the Huh7 WT cells and the Huh7 SR-B1 KO cells. A 2way ANOVA was performed to compare the mean $\log_{10}$RLU obtained in both cell types. It was found that the WT parental strains had a significant drop in infectivity in the Huh7 SR-B1 KO cells compared to Huh7 WT cells. However, this was not the same for the HCVpp double mutants – both of which showed no significant drop in infectivity in Huh7 SR-B1 KO cells. ΔHVR-1 J6 E1E2 had no drop in infection across the two cell lines. ΔHVR-1 H77 did, however, have a small yet statistically significant drop in infection in the Huh7 SR-B1 KO cell as compared to in the Huh7 WT cells.

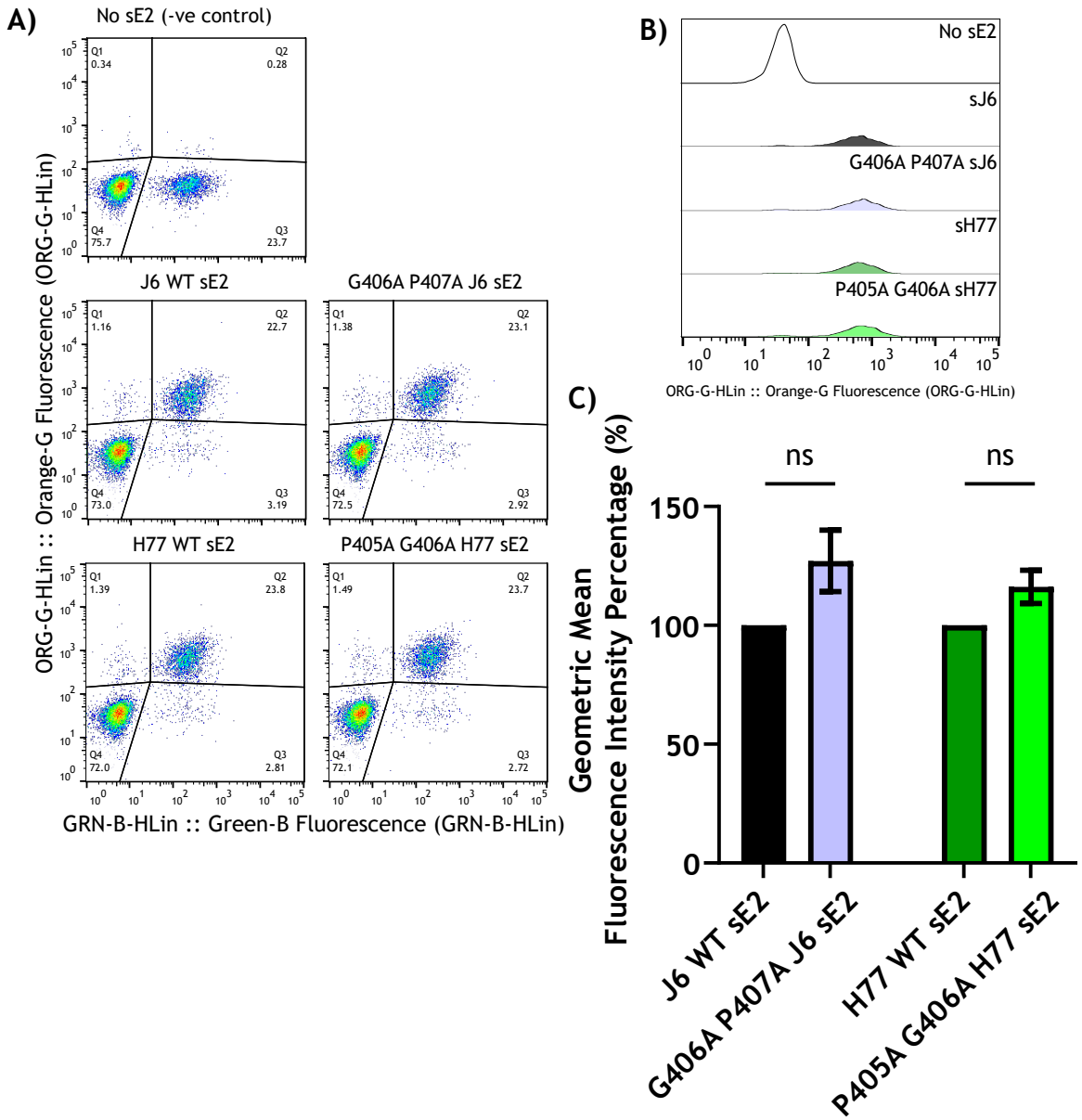


Figure 5. 6. **SR-B1 Binding Efficacy of the J6 and H77 Pocketing-mitigated Constructs.** All work in this figure was performed by Dr Michael Oliver. A) Representative FACS dot blots from the J6 WT sE2, H77 WT sE2, and their derived pocketing-mitigated double mutants. As before, CHO cells were transduced with both GFP and SR-B1, GFP expression was a proxy for SR-B1 expression. Across the x-axis is the green fluorescence signal within the CHO cells. Across the y-axis is the red fluorescence signal, given by the Alexa FluorTM 594 secondary antibody, bound to StrepMAB-Classic, itself bound to the TwinStrep tag on the sE2 constructs. Green +ve cells represented CHO cells which should be expressing SR-B1, and double +ve cells were those with a sE2 construct bound to SR-B1. The quadrants are as follows: Q1 - red +ve, green -ve, Q2 - red +ve, green +ve, Q3 - red -ve, green +ve, Q4 - red -ve, green -ve. Importantly, the no sE2 control showed no shift in the green +ve population towards Q2. B) Representative cytometry histograms of sE2 binding to transduced CHO cells expressing SR-B1 with the no sE2 -ve control. C) Percentage geometric mean fluorescence intensity of the pocketing-mitigated double mutants relative to parental construct. Welch's T-tests were performed between the double mutant constructs and their parental construct. P-values < 0.05 = ns.

These data would support the hypothesis that pocketing of the HVR-1 is what leads to SR-B1 dependency if, indeed, the double mutants in the J6 and H77 backbones are mitigating pocketing of the HVR-1. Next, the ability for these constructs to bind SR-B1 was investigated to better understand the relationship between the ability to pocket the HVR-1, SR-B1 dependency, and the ability to bind SR-B1.

We expressed and purified sE2 constructs for both J6 and H77 containing the helix-breaking double mutants and used these sE2 constructs to investigate SR-B1 binding in CHO cells expressing both GFP and SR-B1, as done before with the length polymorphism-containing sE2 constructs.

It has previously been shown in the length polymorphism sE2 assay that mutations at the N-terminus of the HVR-1 can alter SR-B1 binding. However, increasing the flexibility of the HVR-1 at the C-terminus of the HVR-1 has no effect on SR-B1 binding. As these helix-breaking constructs do not change any residues at the N-terminus but hypothetically reduce flexibility at the C-terminus of the HVR-1, it can be hypothesised that these mutations will have little to no effect on SR-B1 binding.

This is borne out in the SR-B1 binding assay. Figure 5. 7. A. displays representative dot plots from this assay and, importantly, the -ve control shows no shift in cell population towards the double +ve quadrant. However, each of the constructs can bind SR-B1 and shift the GFP +ve cell population into the double +ve quadrant. To quantify this, the geometric mean red fluorescence intensity of green +ve cells was compared and Welch t-tests were performed between the helix-breaking double mutant constructs and their parental sE2. For both double mutants, there was a non-significant increase in SR-B1 binding.

These data show that mutating out flexible residues at the C-terminus of the HVR-1 - which should disfavour HVR-1 pocketing – can mitigate SR-B1 dependency whilst maintaining the ability to bind SR-B1. These data, therefore, support the two hypotheses stated previously. Firstly, the ability to pocket the

HVR-1 leads to SR-B1 dependency, and that the HVR-1:SR-B1 interaction is required to remove the HVR-1 from the pocket. Secondly, the ability to bind SR-B1 is uncoupled from the need to use SR-B1 as an entry factor. Whilst these hypotheses best describe the data herein presented, the ability to pocket or not has only been assumed and not measured directly. To attempt to measure this directly, a **single-molecule Förster resonance energy transfer** (smFRET) system for HCV E2 was designed.

5. 4. Design and Development of an HCV E2 smFRET System

smFRET is an application of **Förster resonance energy transfer** (FRET) wherein donor and acceptor pairs within in a single molecule are excited and detected, allowing the measurement of conformational change. In brief, the donor fluorophore is excited upon exposure to an exogenous photon. The donor transfers its excitation energy to the acceptor via non-radiative dipole–dipole coupling. As the excited acceptor electron returns to ground state, it can release an acceptor-derived photon which can then be detected. Notably, this ratio of detected donor- and acceptor-derived photons allows estimation of the distance between the donor and acceptor. High donor emission and low acceptor emission indicate low FRET efficiency, consistent with a larger donor–acceptor separation. The inverse ratio implies that the two fluorophores are close together. As this ratio changes over time, the change in distance can also be measured. This makes smFRET an effective tool for examining conformational changes in proteins (Roy, Hohng and Ha, 2008; Schuler, 2013; Sasmal *et al.*, 2016). Hence, to investigate the pocketing hypothesis and the interaction between the HVR-1 and SR-B1 more directly, an smFRET system was designed and the initial steps to get the system up and running were undertaken, however we were unable to go beyond initial pilot experiments due to time constraints.

Several smFRET systems have been developed for other viral glycoproteins including HIV-1 (Munro *et al.*, 2014), IAV (Das *et al.*, 2018), and SARS-CoV-2 (Lu *et al.*, 2020). Each of the above studies took a similar approach to developing their smFRET system. Each used pseudotyped viruses with site-specific labelling to enable binding of the FRET donor and acceptors. For the development of the IAV smFRET system (Das *et al.*, 2018), they used a genetic code expansion technique involving amber stop codons and **unnatural amino acids** (UAAs). This technique involves mutating certain codons to be amber stop codons (UAG). As there are no mammalian tRNAs able to recognise the UAG amber stop codon, this would typically result in halting translation. However, co-transfection of an orthogonal tRNA/tRNA synthetase pair into producer cells supplemented with exogenous UAA can be used to permit site-specific incorporation of UAA in the nascent protein, which in turn is incorporated into viral pseudoparticles.

Moieties present on the UAA can then be used for click chemistry to bind specific donor and acceptor pairs onto the viral pseudoparticles.

The first step in developing an smFRET system of HCV was to decide which UAA to use for labelling. For the HCV E2 system, **trans-Cyclooct-2-en-L-Lysine** (TCO*A) was chosen as the UAA. TCO*A is derivative of pyrrolysine and allows for copper-free click chemistry through **Strain-Promoted Inverse Electron-Demand Diels-Alder Cycloaddition** (SPIEDAC) reactions with tetrazines. This was for purely pragmatic reasons - Dr James Dunlop at the Centre for Virus Research had, successfully, used site-specific TCO*A labelling for his own work and graciously provided $tRNA^{TCO*A}$/$tRNA^{TCO*A}$ synthase plasmids and advice.

Next, the residues to be mutated had to be considered. TCO*A is relatively large for an amino acid (255.32 Daltons). To avoid issues with steric bulk ultimately affecting the ability of E2 to fold or function correctly, the glycan-linked asparagine residues were considered. Prentoe *et al.*, in 2019, studied the effects of mutating the glycan-linked asparagine residues. They found that mutation of the asparagine residues was permissible and still resulted in infectious virus, but this was in a strain, glycan position, and mutant residue specific manner. For example, they found that N423D in S52 and J6 ablated viral infectivity. However, in H77, N423D reduced infectivity but not did not prevent it entirely. When the team put N423Q into J6, however, this restored infectivity relative to N423D. Given this intersectional complexity, it was decided that each glycan-linked asparagine would be mutated to TCO*A, evaluated for infectivity, and then assessed for expression of E2. Any single TCO*A-containing E2s found to permit HCVpp infectivity would then be shortlisted. Informed by the 3D positioning of the shortlisted TCO*As in relation to one another, pairs of amber stop codons would then be cloned into single plasmids and evaluated for infectivity, and E2 expression. To facilitate easier cloning between plasmids, a J6 WT E1E2 plasmid with additional cut sites introduced through silent nucleotide mutations was designed and synthesised (J6 Plus).

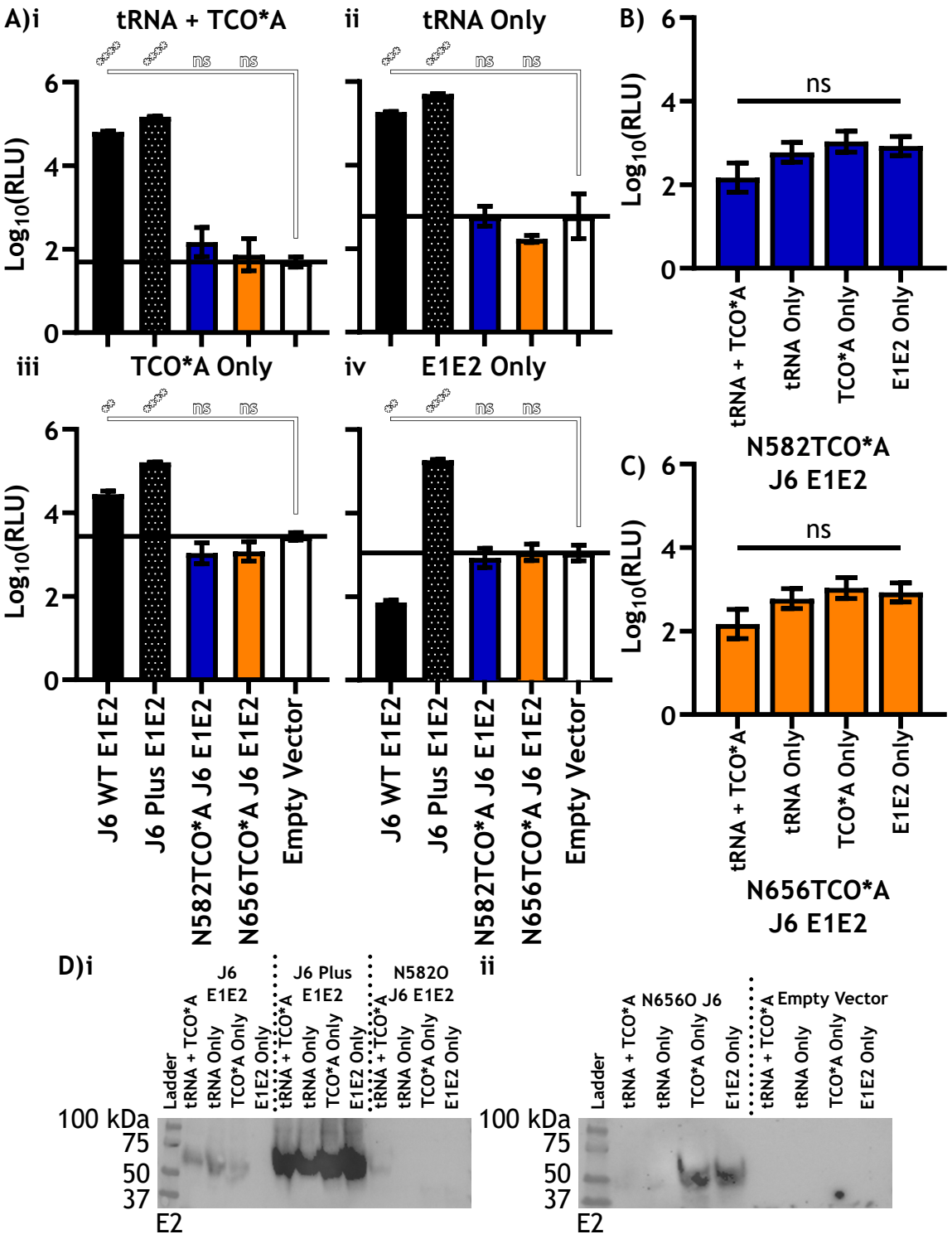


Figure 5. 7. **Developing an HCV E2 smFRET System.** A) Log_{10}(RLU) of smFRET constructs infecting Huh7 WT cells under four conditions: i) $tRNA^{TCO*A}$ + exogenous TCO*A ii) $tRNA^{TCO*A}$ alone iii) exogenous TCO*A alone, iv) E1E2 plasmids only. Ordinary one-way ANOVA followed by Dunnett's multiple comparison tests compared the infectivity of each construct to Empty Vector. P-values: ns > 0.05, * ≤ 0.05, ** ≤ 0.01. Error bars = SEM across three technical replicates. B) and C) Log_{10}(RLU) for N582O J6 and N656O J6 constructs across conditions. One-way ANOVAs showed no significant differences (ns > 0.05). D) Cell lysate Western blot (AP33) shows J6 E1E2 vs J6 Plus E1E2 differences correlating with infectivity. Weak signal in tRNA+TCO*A N582O lane. Bands in N656O TCO*A-only and E1E2-only lanes suggest truncated E2 production.

We began with a pilot experiment in which two glycan-linked asparagine residues were successfully mutated to TCO*As – N582TCO*A and N656TCO*A.

The first test conducted was to look at infectivity of the amber stop codon-containing plasmids in the presence of $tRNA^{TCO*A}$/$tRNA^{TCO*A}$ synthase and exogenous TCO*A, $tRNA^{TCO*A}$/$tRNA^{TCO*A}$ synthase alone, exogenous TCO*A alone, or neither. Both J6 WT E1E2 HCVpp and J6 Plus E1E2 HCVpp were used as positive controls, and Empty Vector as a negative control and measure of background (Fig. 5. 8. Ai – iv). Across the four conditions, neither N582TCO*A J6 nor N656TCO*A J6 attain log_{10}RLUs significantly greater than the Empty Vector HCVpp, implying that these HCVpp – if produced – are non-infective.

There is consistency between the J6 WT E1E2 and J6 Plus E1E2 across three of the four conditions. In the E1E2 only infection, however, J6 WT E1E2 HCVpp does not reach above background. This is, perhaps, explained by the poor transfection which resulted in no E2 being detected by Western blot (Fig. 5. 8. Di). However, this does suggest that the co-transfection of $tRNA^{TCO*A}$/$tRNA^{TCO*A}$ synthase and the addition of exogenous TCO*A are not deleterious to the production of functional HCVpp.

To confirm that there was no difference between the log_{10}RLUs obtained for N582TCO*A J6 E1E2 and N656TCO*A J6 E1E2 across the four conditions evaluated, a one-way ANOVA was performed and there was no significant difference found. Without $tRNA^{TCO*A}$/$tRNA^{TCO*A}$ synthase or exogenous TCO*A, there should be no productive translation of the N582TCO*A E2 and the N656TCO*A E2 proteins and therefore no infectivity. To examine protein expression levels, a Western blot was performed on lysates from the producer cells in each condition to confirm whether N582TCO*A or N656TCO*A (Fig. 5. 8. Di – ii). In the "tRNA + TCO*A" lane for N582TCO*A J6 E1E2, there is a small, faint band which is not seen in other lanes. This may be evidence for a $tRNA^{TCO*A}$/$tRNA^{TCO*A}$ synthase and exogenous TCO*A dependent expression of E2. However, the band is faint, and further repeats would be required to validate this result.

For N656TCO*A J6 E1E2, there are two large bands present in the “TCO*A only” and “E1E2 only” lanes. They are slightly smaller than the canonical E2 size of ~ 60 kDa, implying that these bands represent some level of production of a truncated E2. This suggests that the plasmid containing the amber stop codon is being translated in these cells but potentially the ribosome is stalling due to the lack of charged $tRNA^{TCO*A}$ which can lead to the release of truncated proteins from the ribosome (Subramaniam, Zid and O’Shea, 2014; Darnell, Subramaniam and O’Shea, 2018). However, this truncated E2 band is not present in the tRNA only lane. Without being charged with TCO*A, the $tRNA^{TCO*A}$ would be disfavoured by the EF-Tu complex for entry into the elongation ribosome (Nissen *et al.*, 1996; Raina and Ibba, 2014). Therefore, the uncharged $tRNA^{TCO*A}$ should not interfere with normal translation of the N656CO*A J6 E1E2 plasmid itself. Hypothetically, this condition could produce the truncated E2 as in the “TCO*A only” and the “E1E2 only” lanes. It should be noted that this was a single repeat of the experiment and may be the result of a poor transfection as seen in the J6 WT E1E2 in the E1E2 only conditions.

5. 5. Concluding Remarks

The aim of this section of work was to investigate how, mechanistically, the HVR-1 contributed to phenotypes described in Chapter 4. As noted in the introduction to this chapter, myriad studies have examined the role of the HVR-1 in infectivity, receptor dependency, and antibody sensitivity (Bartosch *et al.*, 2005; Bankwitz *et al.*, 2010; Prentoe *et al.*, 2011; Stejskal, Kalemera, *et al.*, 2020; Olesen *et al.*, 2021, 2025) but little has been proven about how the HVR-1 modulates these aspects of the phenotype of HCV.

To start, ColabFold was used to investigate the interaction between E2 and SR-B1. As noted, other studies examined the interaction between SR-B1 and related proteins in the CD36 superfamily of proteins and how they bound their ligands (Neculai *et al.*, 2013; Hsieh *et al.*, 2016; Zhou *et al.*, 2019; Langlois *et al.*, 2020). Consistently, they found that the ligands bound to these CD36 superfamily proteins through the conserved apical tri-helical bundle. However, the ColabFold models that are shown here show a different story wherein the HVR-1 is inserted into the channel of the SR-B1 protein. However, it should be noted that this prediction is not consistently found across every E2/SR-B1 co-fold experiment performed in ColabFold. As discussed in the introduction to Chapter 3, ColabFold relies on analysis of the co-evolution of certain residues to accurately make predictions (Lapedes *et al.*, 1999; Marks, Hopf and Sander, 2012; AlQuraishi, 2019) but it may well be that there are too few co-evolutionary clues in the MSAs produced for these two proteins for ColabFold to predict the interaction consistently and with high confidence. Nonetheless, if this prediction is accurate, it may explain why the shorter HVR-1s did not make use of SR-B1 as a co-receptor; as they may not be long enough to enter the channel and/or form sufficient intramolecular interactions to stabilise the interface for long enough to permit downstream conformational changes in E2, priming CD81 engagement.

To further investigate this interaction, sE2s with HVR-1s of differing lengths were produced, purified, and used in an SR-B1 binding assay. It was found that HVR-1s of any length were able to bind SR-B1. However, the different HVR-1 lengths did modulate SR-B1 binding although not statistically significantly so for

the longer HVR-1s (i.e. comparing J6-backbone proteins to sJ6, and I438V A524T-backbone proteins to sIVAT). However, the -3 sJ6 protein did exhibit reduced binding to SR-B1. This result lends credence to the ColabFold model and the hypothesis that shortening the HVR-1 reduces SR-B1 dependency through reducing the number of residues available in the SR-B1 channel which can stabilise the interaction. However, -3 sJ6, +4I sJ6, and +4I sIVAT did, indeed, bind SR-B1 whereas in Chapter 4. 4., it was demonstrated that the wildtype J6-backbone N-Terminus Deletion HCVpp, and the Internal Insertion HCVpp in both backbones were not dependent on SR-B1 for viral entry. Altogether, this suggests that the need to use SR-B1 as a receptor for entry is uncoupled from the ability to bind SR-B1.

Importantly, this assay also provided more clarity on the hierarchy of mutations in comparing the wildtype J6-backbone length polymorphism sE2s with those in the I438V A524T-backbone. Overall, the sIVAT-backbone sE2 proteins bound SR-B1 less than their sJ6-backbone counterparts, reflecting the difference between sJ6 and sIVAT. However, when compared to each respective parental sE2 protein, the length polymorphisms introduced into each background bound SR-B1 in a statistically similar manner. This suggests that however the length polymorphisms are bringing about their phenotype, it is distinct from how the I438V A524T mutations cause their phenotype. Moreover, the overall reduction in the SR-B1 binding from the sIVAT-backbone mutations implies a hierarchy in that the phenotype from the I438V A524T mutations is exerted over and above the phenotypes associated with altering the length of the HVR-1.

Although the experimental data shown so far supports the idea that the HVR-1 may insert into the SR-B1 channel, we next wanted to investigate whether the pocketing of the HVR-1 which was prevalent in the ColabFold models of E2 (Chapter 3 Section 3. 1.) and in the E1E2 dimer of dimers structure (PDB: 8RJJ, Augestad *et al.*, 2024) was important for this E2:SR-B1 interaction. Dr Michael Oliver produced two constructs based on the hypothesis that the pocketed HVR-1 conformation was important for SR-B1 engagement. Dr Michael Oliver produced two constructs wherein conserved, helix-breaking residues found at the C-terminus of the HVR-1 in both J6 and H77 were replaced with alanine residues. It

was hypothesised that mutating these helix-breaking residues to alanine residues would disfavour the HVR-1 from hinging and folding down into the HVR-1 binding cleft by favouring alpha-helix formation across the HVR-1. Interestingly, his work found a strikingly similar relationship between his pocketing-deficient constructs and the N-Terminus Deletion constructs described in Chapter 3. Both groups of constructs were SR-B1 independent for viral entry and yet were able to bind SR-B1. Altogether, this would suggest that for HVR-1 to interact with SR-B1 to facilitate viral entry, it has to be able to pocket itself into the HVR-1 binding cleft. Yet the inability to pocket does not intrinsically prevent HVR-1:SR-B1 interactions, which may be driven by the N-terminus of HVR-1.

Altogether, the investigation of conserved helix-breaking residues at the C-terminus of HVR-1 has provided compelling evidence for the HVR-1 pocketing hypothesis. The demonstration that mutating glycine and proline residues to alanine abolishes SR-B1 dependency while maintaining SR-B1 binding capacity strongly supports the model wherein HVR-1 pocketing creates SR-B1 dependency, and HVR-1:SR-B1 interaction is required to sequester the HVR-1, preventing conformational sampling which enables subsequent conformational changes necessary for CD81 binding and viral entry. These findings suggest a two-step entry mechanism where SR-B1 serves not only as an attachment factor, but as a conformational switch.

At this point, there are several indirect streams of evidence which support the idea that movement in the HVR-1 is important for its function and modulating its ability to move between, or adopt, certain conformations such as pocketing may further modulate the phenotype of HCV. For example, a few theories have been posited about how the HVR-1 functions to modulate antibody sensitivity. In brief, early theories posited that the HVR-1 operated as a immunodominant epitope decoy for conserved functional epitopes either through direct steric shielding, or through “distracting” the humoral response (Ray *et al.*, 1999; Korenaga *et al.*, 2001; Cuevas *et al.*, 2009; Schmidt-Martin *et al.*, 2015; Prentoe *et al.*, 2016). However, as noted in Section 1. 7. of the introduction, these hypotheses have been supplanted by additional studies by Augestad *et al.* who demonstrated that the HVR-1 can modulate the overall E1E2 complex

conformation and render the complexes antibody resistant or sensitive (Augestad *et al.*, 2020). Additionally, work by Stejskal *et al.* demonstrated that movements in the HVR-1 are sterically transduced throughout the entirety of E2 (Stejskal, Kalemera, *et al.*, 2020), However, the former study relied on specific antibodies for neutralisation assays to make inferences, the latter used MD simulations to predict the effects of movement of the HVR-1 on the E2 protein as a whole. Ergo, to date there has been no direct study of movement of the HVR-1 and how that influences the phenotype of HCV.

To that end, the last project attempted was to develop an smFRET system to allow detection of movement within the HVR-1. This would then allow researchers to examine how movement in the HVR-1 is altered by length and compositional polymorphisms, how mutations outwith the HVR-1 affect its movement, and crucially how the HVR-1 interacts with various receptors including SR-B1. Ultimately, whilst not fruitful, an smFRET system for the E2 protein of HCV would be an invaluable tool in rationally designing a vaccine for HCV. If there was a clear and distinct idea of the conformations and conformational changes that the E1E2 complex had to undergo to facilitate viral entry and ultimately fusion, further vaccine studies would know exactly where to put the wrench into the works of HCV, and hopefully move humanity one step closer to eliminating HCV globally.

Chapter 6

Discussion

6. Discussion

The nature of the HVR-1 has rendered it fundamentally resistant to study through classical structural biology – it is a dynamic and disordered tail at the N-terminus of the E2 protein of HCV. To date, there are 72 deposited E2 structures on the PDB; only one of these structures actually resolved the structure of the HVR-1 (Augestad *et al.*, 2024; RCSB Protein Data Bank, 2025). It has long been understood that the HVR-1 influences the infectivity, receptor dependency and immune escape of HCV (da Silva Cardoso *et al.*, 1995; Scarselli *et al.*, 1995; Ray *et al.*, 1999; Korenaga *et al.*, 2001; Penin *et al.*, 2001; Roccasecca *et al.*, 2003; Dowd *et al.*, 2009; Prentoe *et al.*, 2016; Augestad *et al.*, 2020; Stejskal, Lees, *et al.*, 2020; Olesen *et al.*, 2025). However, due to the difficulties in visualising the structure of the HVR-1 in relation to the rest of the E2 protein, mapping the structure-function relationship between this N-terminus tail and the overall function of the E1E2 complex has remained an intractable problem.

Understanding HVR-1 and exactly how it contributes to the function of HCV virions is of critical importance. To date, there exists no vaccine against HCV. Currently, the WHO estimates that 50 million people are currently living with HCV with approximately 1 million new infections per year (World Health Organisation, 2025). Whilst DAAs are available and offer cure rates of approximately 95 % of people living with HCV, they are expensive (Public Health Scotland, 2024) and access remains inequitable (Marshall *et al.*, 2024). Therefore, an effective prophylactic vaccine is required to meet the goal of the WHO to eliminate HCV by 2030 (Torre *et al.*, 2024). One of the key issues with the development of an effective vaccine is the gaps in our knowledge on how the HVR-1 contributes to function, how it aids immune escape, and how to overcome these difficulties. Through the use of state-of-the-art machine learning-enabled protein structure prediction software, high throughput bioinformatics, and classical virology, this PhD aimed to expand our knowledge of how the HVR-1 functions.

6. 1. Aims and Objectives Revisited

To deepen understanding of the structure-function relationship between the HVR-1 and the rest of the E1E2 complex and the E2 protein in isolation, four different protein structure prediction software platforms were employed to produce 1,036 models of the E1E2 complex (500 models) and E2 alone (536 models).

Concurrently, 9,103 unique HVR-1 sequences were extracted from a total of 30,289 sequences on HCV-Glue. This enabled bioinformatic examination of the sequence variation in the HVR-1. Moreover, this permitted analysis of the prevalence and universality of length polymorphisms within the HVR-1 through more traditional bioinformatic approaches.

Lastly, the results from these two branches of research led to some classical virology experiments. Using both HCVpp and sE2 proteins containing length polymorphisms in the HVR-1, the effects of those length polymorphisms in the HVR-1 were examined in relation to infectivity, receptor dependency, and antibody sensitivity.

Altogether, due to the difficulties of studying the HVR-1 using classical structural biology, this three-pronged approach was undertaken with the aim of building a more complete picture of HVR-1 function than what classical structural approaches alone could provide.

6. 2. Current Knowledge Gaps

6. 2. 1. The HVR-1 *in situ* with the rest of the E2 Protein

The first gap in our understanding of the HVR-1 has already been discussed: there is a dearth of structures of E1E2/E2 alone which actually resolve the structure of the HVR-1, likely a result of the disorder and dynamics of the HVR-1 itself. Our understanding of how the HVR-1 relates, conformationally, to the rest of the E2 protein is lacking. Given the advent of machine learning-enabled protein structure prediction around the start of this PhD, the question was asked: what can these protein structure platforms reveal about the structural relationship between the HVR-1 and the rest of the E2 protein?

Using ColabFold (Jumper *et al.*, 2021; Mirdita *et al.*, 2022), AlphaFold 3 (Abramson *et al.*, 2024), Boltz-1 (Wohlwend *et al.*, 2025), and ChAI-1 (Chai Discovery *et al.*, 2024), we generated 536 models of the E2 proteins in isolation, and 500 models of E1E2 dimeric complexes.

Through systematic examination of these models, there were several key findings. The HVR-1 was predicted to be found in one of three distinct poses. These were termed the “free-floating” pose, the pocketed HVR-1 conformation, and the pose wherein the HVR-1 is tucked under the CD81 binding loop. Interestingly, these poses appeared to be characteristic of particular protein folding software platforms. For example, only ColabFold predicted a pocketed HVR-1, and only when folding E2 alone. However, when predicting the structure of the E1E2 complex, ColabFold positioned the HVR-1 in the “free-floating” pose. AlphaFold 3 consistently, across E1E2 complex models and E2 alone models, positions the HVR-1 in the “free-floating” pose. Boltz-1 consistently positioned the HVR-1 tucked under the CD81 binding loop for both E1E2 models and E2 alone models. Lastly, ChAI-1 positioned the HVR-1 in apposition to the rest of the E2 protein, in between the HVR-1 pocket and under the CD81 binding loop.

We found the pocketed HVR-1 to be the most interesting of these poses. Initially, the importance of the pocketed HVR-1 pose was not entirely appreciated until Augestad *et al.* (2024) released their structure of E1E2 (PDB: 8RJJ). This structure was surprising for two important reasons: it showed the E1E2 complex as a dimer of dimers; and they resolved the structure of the HVR-1. This demonstrated that the HVR-1 is being sequestered into what they term the HVR-1 binding cleft, what we term the HVR-1 pocket. This lent confidence to the ColabFold models we had already generated. Importantly, it also made us not dismiss – as potential artifacts/hallucinations - the other poses for the HVR-1 as predicted by AlphaFold 3, Boltz-1, and ChAI-1.

However, it should be noted that whilst this structure (PDB: 8RJJ) gave us more confidence in the models we had produced, none of these platforms were able to accurately predict the E1E2 dimer of dimers structure which was shown by Augestad *et al.* This does suggest that there is, still, improvements to be made for machine learning-enabled protein structure prediction, or indeed, that biological reality is sometimes stranger than what can be predicted or assumed based on previous knowledge.

6. 2. 2. Systematic Analysis of HVR-1 Sequences

The second knowledge gap that this PhD project sought to fill was that of performing a systematic analysis of HVR-1 sequences. It has been noted in many papers that the HVR-1 is prone to compositional and length polymorphisms (Hijikata *et al.*, 1991; Ogata *et al.*, 1991; Weiner *et al.*, 1991; Kato *et al.*, 1992; Ray *et al.*, 1999; Kumagai *et al.*, 2007; Prentoe *et al.*, 2016; Law *et al.*, 2018; Olesen *et al.*, 2025). Note, this list of papers is non-exhaustive, almost every paper that discusses the HVR-1 notes its compositional lability and may note its propensity for indel mutations. However, to date, only two papers have performed a systematic analysis of HVR-1 sequence. Penin *et al.*, in 2001, reported an analysis of 1,382 non-redundant HVR-1 sequences; in 2005, Callens *et al.* performed a similar analysis with 1,489 non-redundant HVR-1 sequences (Penin *et al.*, 2001; Callens *et al.*, 2005).

Using HCV-Glue (Singer *et al.*, 2018), an online database which scrapes other sequence repositories for HCV sequence information, 30,289 HCV sequences which covered at least 25 % of the E2 protein were downloaded, and 9,103 non-redundant HVR-1 sequences were extracted from these sequences. To date, this is the largest analysis of HVR-1 sequences.

Our key findings from this analysis were many-fold. However, most importantly, length polymorphisms are prevalent and universal throughout the various genotypes of HCV. We found length polymorphisms in 6 of the 7 genotypes. More precisely, we found both insertions and deletions across 4 out of the 7 genotypes (GT1, GT2, GT3, and GT6). However, for GT4 only insertions were found, and for GT5 only one HVR-1 sequence was found to contain a deletion. For GT7, neither insertions nor deletions were found in the HVR-1 sequences. However, the lack of both insertions and/or deletions found in these three genotypes – GT4, GT5, and GT7 – may reflect the fact that these are the genotypes for which we have the fewest sequences. This is likely a sampling artefact and not a reflection of the underlying biology of the HVR-1/E2 proteins for these genotypes wherein indels are not biologically permissible.

Moreover, in considering the phylogenetic analysis which permitted analysis of the universality of length polymorphisms across the seven genotypes of HCV, it should be noted that E2 sequences without the HVR-1 were used to produce the phylogenetic tree. This methodology was used to avoid bias from the HVR-1 sequences in producing the phylogenetic tree itself. In the analysis presented herein, it was found that the E2 sequences separated out by genotype for the most part. However, Sliepen *et al.* did produce a phylogenetic tree of 512 randomly selected E1E2 sequences, followed by a phylogenetic tree of the 512 HVR-1 sequences alone. They found that the E1E2 sequences, as in this work, separated out by genotype. However, they found that the HVR-1 sequences themselves did not cluster by genotype. This suggests that compositional and length polymorphisms within the HVR-1 are not restricted by genotype i.e. there is no detectable evolutionary signal within the 512 HVR-1 sequences that Sliepen *et al.* analysed that caused the HVR-1 sequences to cluster by genotype (Sliepen *et al.*, 2022). Indirectly, this further supports the fact that both compositional and length polymorphisms are universally present across all seven genotypes of HCV, and that, for example, there is no particular polymorphism which is sequestered to one genotype which may be detected by phylogenetic analysis.

In taking a step back and examining the genotypes, we found that for GT6 the modal HVR-1 was 26 amino acids in length. For every other genotype, the modal HVR-1 length was 27. Additionally, we found that the HVR-1 had a propensity for insertions with 5.68 % of sequences featuring an insertion, and only 0.68 % of sequences containing a deletion. Including those sequences in GT6, which as per appendix III, only contained five sequences which were shorter than 26 amino acids in length. This may suggest that 26 amino acids is a possible lower length threshold, below which continuing to shorten the HVR-1 is deleterious. This may aid in explaining two observations that were made when looking at the sequence logos in chapter 3 (Section 3. 3.), and the characterisation of the HCVpp in chapter 4. It was noted that there was a highly conserved threonine residue which in wildtype J6 is found in the second position of the HVR-1. When considering the results from chapter 4, we found a consistent statistically significant difference between the phenotype of the wildtype J6-backbone -1 N-Terminus Deletion construct HCVpp and the other four wildtype J6-backbone N-

Terminus Deletion constructs. Altogether, shortening the HVR-1 below 26 amino acids may be intrinsically deleterious, or it may be that the removal of this threonine residue is the critical determinant in phenotype. Dr Michael Oliver did begin to examine the effects of removing this threonine residue by mutating it to an alanine residue. However, this project was unable to be completed due to time constraints but remains an interesting question.

Moreover, in examining 9,103 HVR-1 sequences we were able to identify two separate hotspots for insertions – at the very N-terminus of the HVR-1, and at the C-terminus, following D404 in the wildtype J6 HVR-1. Moreover, the results from chapter 4 and 5 support the hypothesis that HCV phenotype is altered not only by the presence of insertions, but also by where the insertions are located within the HVR-1 as evidenced by the difference in phenotypes between the N-Terminus Extension HCVpp and the Internal Insertion HCVpp.

In further examining the C-terminus of the HVR-1, Dr Michael Oliver noted the high level of conservation of small, flexible amino acid residues (Chapter 5, Section 5. 3.) at almost the very base of the HVR-1 (G406 and P407 in wildtype J6). Through mutating these flexible amino acids in both J6 and H77 at the C-terminus end of their respective HVR-1s, Dr Michael Oliver was able to recapitulate the SR-B1 dependency and binding phenotypes found for the wildtype J6-backbone N-Terminus Deletion HCVpp/sE2 proteins. Therefore, these highly conserved flexible amino acids at the C-terminus of the HVR-1, we assert, are critical for the ability of the HVR-1 to pocket into the binding cleft. This insight was enabled only through the extraction, alignment, and assessment of 9,103 HVR-1 sequences.

6. 2. 3. How does the HVR-1 Function?

The work performed in chapter 3 was done to evaluate our initial hypothesis: length polymorphisms in the HVR-1 are prevalent and universal. State-of-the-art protein structure prediction software can be used to elucidate more structural information about the relationship between the HVR-1 and the rest of the E2 protein.

Length polymorphisms were found to be prevalent and universal. Moreover, the HVR-1 was predicted to be positioned in one of three distinct poses – the "free-floating" HVR-1, the pocketed HVR-1, the HVR-1 tucked under the CD81 binding loop. After validating the initial hypothesis, our hypotheses evolved. Therefore, the work presented in chapters 4 and 5 aimed to assess the following central hypotheses: the HVR-1 vacillates between (at least) three distinct conformational poses. Length polymorphisms modulate the rate at which the HVR-1 switches between these poses. This has downstream, measurable effects on HCV phenotype with regards to infectivity, SR-B1 dependency, and neutralisation sensitivity. In particular, we assert that the ability to pocket the HVR-1 is important for modulating HCV phenotype.

Together, the results from chapters 4 and 5 allowed for the proposal of a model for HVR-1 function.

6. 3. A Proposed Model of HVR-1 Function

6. 3. 1. Length Polymorphisms Do Alter HVR-1 Function

Part of our central hypothesis states that length polymorphisms in the HVR-1 would modulate phenotype. The work presented in chapter 4 supports this hypothesis: length polymorphisms in the HVR-1 do alter the phenotypes of HCVpp. However, something that was not hypothesised but emerged from the data in chapter 4 is that the position of insertions into the HVR-1 also alters HCVpp phenotype. Consider the differences between the phenotypes characterised for the N-Terminus Extension HCVpp and the Internal Insertion HCVpp. Specifically, if we compare the wildtype J6-backbone +4 N-Terminus Extension HCVpp and the wildtype J6-backbone +4I Internal Insertion HCVpp, both contained four additional glycine residues but at the N-terminus and at the C-terminus, respectively. Both HCVpp have different phenotypes when considering infectivity, SR-B1 dependency, and antibody sensitivity.

6. 3. 2. The I438V A524T Mutations and Length Polymorphisms in the HVR-1

Previous work by the Grove lab characterised these I438V A524T mutations (Stejskal *et al.*, 2022). They found that the I438V A524T J6 HCVpp had a distinct phenotype from wildtype J6 HCVpp. The I438V A524T J6 HCVpp were found to be hyperreactive, SR-B1 independent, and antibody sensitive. Previous work by the Grove lab concluded that the I438V A524T mutations altered phenotype through constitutive stabilisation of the HVR-1. Given this, we investigated the effects of combining both the I438V A524T mutations and length polymorphisms.

Interestingly, what we found is that the I438V A524T mutations altered the phenotypes of the length polymorphism-containing HCVpp. Initially, based on the infectivity, receptor dependency, and polyclonal neutralisation assays, it was believed that there was a hierarchy of mutations. For these assays, there was little difference for the results obtained for the Internal Insertion HCVpp in either backbone. However, the phenotypes of both the N-Terminus Deletion and N-Terminus Extension HCVpp were altered upon the introduction of the I438V A524T mutations. Therefore, after these assays, we hypothesised that the I438V A524T mutations were dominant over the length polymorphism mutations at the N-terminus of the HVR-1, but not dominant over the C-terminus insertions.

However, when considering the monoclonal neutralisation assays, the wildtype J6-backbone Internal Insertion HCVpp and the I438V A524T J6-backbone Internal Insertion HCVpp were statistically different. Therefore, whilst the effects of the I438V A524T mutations were perhaps less prominent on the Internal Insertion HCVpp, the I438V A524T mutations were subtly altering their phenotype. Altogether, this suggests that the effects of the length polymorphisms and the I438V A524T mutations are distinct yet additive. Currently, we do not have a concrete idea as to how the I438V A524T mutations produce their phenotype from a mechanistic perspective. The work by Stejskal *et al.* relied on MD simulations and indirect characterisation of phenotypes to ascertain how the I438V A524T mutations produce their phenotype (Stejskal *et al.*, 2022), much as we have within this thesis. Therefore, to truly disentangle how the length

polymorphisms and the I438V A524T mutations contribute to the phenotypes of the HCVpp characterised in the I438V A524T J6 backbone, more experimental work needs to be completed. Later, we shall discuss our attempts to generate an smFRET system for the E2 protein of HCV. Crucially, an smFRET system would allow us to better understand how both length polymorphisms and the I438V A524T mutations contribute to phenotype separately, and together.

This was further evidenced by the sE2:SR-B1 binding assay. This will be discussed in more detail in the next section. However, in relation to the I438V A524T mutations, it was found that the sJ6 protein was able to bind SR-B1 more than sIVAT. Moreover, when considering the length polymorphisms, depending on the normalisation approach used to assess the ability to bind SR-B1, an interesting pattern emerged (Chapter 5, Section 5. 2., Figure 5. 4.).

When normalising the ability of length variant sE2s to bind SR-B1 relative to their parental sE2 protein, it was found that there was no difference between the +3 and +4I sE2 proteins in either backbone to bind SR-B1. However, it was found that the sIVAT proteins, and the derived length polymorphism-containing sE2 proteins were less able to bind SR-B1 than their wildtype J6 counterparts. Altogether, this does suggest that whatever the I438V A524T mutations are doing, it is altering the relationship between the HVR-1 and the rest of the E2 protein. This may be through constitutive HVR-1 stabilisation, or through sequestration of the HVR-1 in a particular conformation – for example tucked next to the CD81 binding loop as suggested by Boltz modelling - or indeed through modulating the transduction of movement along the HVR-1 throughout the rest of the E2 protein. Ultimately, this would require much more work to investigate and map exactly how these mutations alter HVR-1 or E2 function.

6. 3. 3. HVR-1 Conformation and SR-B1 Dependency

In further examining the relationship between the HVR-1 and SR-B1, we have found evidence – although indirect – that SR-B1 dependency may be defined by the ability to pocket the HVR-1. Three important pieces of evidence come to the fore when discussing this concept: the wildtype J6-backbone N-Terminus Deletion HCVpp infecting the Huh7 SR-B1 KO cells (Chapter 4, Section 4. 4., Figure 4. 6.), their AR4A neutralisation assay results (Chapter 4, Section 4. 6. 4., Figure 4. 14.), and the work of Dr Michael Oliver in chapter 5 (Chapter 5, Section 5. 3., Figure 5. 6.).

We suggested that the wildtype J6-backbone -3 N-Terminus Deletion HCVpp either do not pocket their HVR-1 or do so less than the other length polymorphism-containing HCVpp. This was evidenced by the fact that these HCVpp were the most neutralised by AR4A. According to Augestad *et al.* (2024) and their 8RJJ E1E2 complex structure, the pocketed HVR-1 occludes the AR4A epitope. Given how sensitive the wildtype J6-backbone -3 N-Terminus Deletion HCVpp were to AR4A, we concluded that these HCVpp must be unable to physically block the AR4A epitope. Therefore, we believe that this is indirect evidence that these HCVpp cannot pocket their HVR-1.

Prior to this experiment, we hypothesised that these shorter HVR-1s might not be able to physically reach on engage with the HVR-1 pocket. In examining the results for the wildtype J6-backbone N-Terminus Deletion HCVpp infecting the Huh7 SR-B1 KO cells, we saw that there was actually a statistically significant increase in infectivity when SR-B1 was removed from the Huh7 cells. According to the mathematical model of SR-B1 and CD81 acquisition, it is suggested that acquisition of CD81 is enhanced in the presence of SR-B1 (Kalemera *et al.*, 2019b). However, here we have demonstrated that when the HVR-1 is shorter, there is increased infectivity when SR-B1 is absent. Therefore, we posit that the ability to pocket the HVR-1 is part of what drives SR-B1 dependency. Moreover, having a shorter HVR-1 which cannot pocket may be shifting the equilibrium of intrinsic CD81 acquisition.

This was further supported by work by Dr Michael Oliver who, off the back of the sequence logos in chapter 3 (Chapter 3, Section 3. 3., Figure 3. 9.), noted that there were conserved small, flexible, and potentially helix-breaking amino acid residues at the C-terminus of the HVR-1. It was hypothesised that mutating these flexible, hinge residues may result in HCVpp which cannot pocket their HVR-1.

Therefore, we examined whether the G406A P407A J6 HCVpp and P405A G406A H77 mutants would be SR-B1 independent – the results in Chapter 5, Section 5. 3., Figure 5. 6. support this hypothesis. However, the mutants that Dr Michael Oliver generated did not show the same increase in infectivity in the Huh7 SR-B1 KO cells which was found in the wildtype J6-backbone N-Terminus Deletion HCVpp. We posited that the lack of increase in infectivity may mean that the shift in equilibrium towards intrinsic CD81 acquisition was somewhat different for theses mutants vs the N-Terminus Deletion HCVpp. However, further experimental work would be required to examine the intrinsic acquisition of CD81 by the wildtype J6-backbone N-Terminus Deletion HCVpp.

6. 3. 4. HVR-1 Length and SR-B1 Binding

Notably, dependence on SR-B1 as a co-receptor for HCVpp entry was found to be uncoupled from the intrinsic ability of our length polymorphism-containing sE2 proteins to bind SR-B1. All of our sE2 proteins bound to SR-B1 in our binding assay. Moreover, it should be noted that due to the specific truncations used to generate sE2 proteins, the HVR-1 binding cleft is no longer present (Chapter 5, Section 5. 2., Figure 5. 2.). Ergo, none of the sE2 proteins had the capacity to pocket their HVR-1.

On one hand, we have shown that, for example, the wildtype J6-backbone N-Terminus Deletion HCVpp are not dependent on SR-B1 as a co-receptor for entry. Now, we are demonstrating that shortening the HVR-1 does not remove the intrinsic ability of the HVR-1 to bind SR-B1. Consider the results for the AR4A monoclonal neutralisation assay, wherein it was indirectly shown that the wildtype J6-backbone -3 N-Terminus Deletion HCVpp were pocketing incompetent. Taken together, this may suggest that SR-B1 dependency is correlated with the ability to productively pocket the HVR-1. In turn, this may suggest that SR-B1 does not require the HVR-1 to be in the pocket to bind, but does require the HVR-1 to be in the pocket, or be able to pocket, to productively lead to downstream conformational changes which permit CD81 acquisition and, ultimately, viral fusion to take place. However, this is an extrapolation of indirect evidence, therefore, we must be cautious as to not overstate what these findings could mean.

In considering the predicted structures for HVR-1:SR-B1 binding produced by ColabFold (Chapter 5., Section 5. 1., Figure 5. 1.), it can be imagined how a pocketed HVR-1 and SR-B1 might interact. Note, once more, this will require further experimentation to determine. However, as a hypothesis, it is not outwith the realm of possibility that a pocketed HVR-1 presents the hydrophobic face of the HVR-1 outward. This is positioned towards the hydrophobic residues surrounding the entrance to the SR-B1 lipid transfer channel. This allows the HVR-1 to be inserted into the SR-B1 lipid transfer channel, binding SR-B1. In turn, this may liberate the HVR-1 from the binding cleft. Allosteric movement is

transduced along the HVR-1 into the front layer and the rest of the E2 protein, allowing for CD81 acquisition. Again, this hypothesis will require further experimentation to determine its validity, but as it stands, this interaction model remains plausible.

6. 3. 5. A Proposed Model of HVR-1 Function

Given all of the above, we can put together a model of how the HVR-1 functions to aid in viral entry, how it determines receptor dependency, and how it aids in immune escape. Now, as with the potential HVR-1:SR-B1 binding interaction, this model is a hypothesis based on the data presented in this thesis. However, our data is indirect, and whilst all of our data supports this model, there is every possibility that this proposed model is partially, or entirely, incorrect.

That being said, we do believe that this work continues the work on the function of the HVR-1 that was started by Augestad *et al.* (2020, 2024) and Stejskal *et al.* (2022). Together with their work, we present the following proposed model of HVR-1 function:

The HVR-1 vacillates between three (or more) distinct "poses", or conformational states (Chapter 3).

One of these "poses", what we have termed the pocketed HVR-1 pose, results in an E2 conformation which Augestad *et al.* call the "closed" conformation which is antibody resistant (Augestad *et al.*, 2020, 2024).

Following attachment of the HCV virion to the target cell, a pocketed HVR-1 is positioned such that it requires SR-B1 acquisition and use as a co-receptor. This drives SR-B1 dependency (Chapter 4 and 5).

The HVR-1 potentially inserts into the lipid transfer channel of the SR-B1 (Chapter 5) or is otherwise removed from the HVR-1 binding cleft by SR-B1. Liberation from the HVR-1 pocket further drives conformational changes across the rest of the E2 protein (Stejskal *et al.*, 2022).

In turn, this allows CD81 acquisition and downstream conformational changes to occur across the E1E2 complex which ultimately result in viral fusion. Moreover, we believe that mutations within the HVR-1 and outwith the HVR-1 (such as the

I438V A524T mutations) are able to fine-tune the conformational lability of the HVR-1, which in turn would fine-tune each step of this proposed model.

6. 4. Limitations and Future Work

Whilst we have proposed a model for how the HVR-1 contributes to the overall function of the HCV glycoproteins, there are still many unanswered questions and a few limitations to this study which should be considered.

Firstly, as noted, our model is based on indirect observations and the characterisation of the phenotypes of HCVpp with length polymorphisms in the HVR-1. There was an attempt to produce an smFRET system which would allow for direct observation of the movement of the HVR-1. This, we believe, would have directly validated the model we have now proposed. Unfortunately, due to technical difficulties and time constraints, only the initial cloning steps for inserting UAAs into two positions across the J6 E2 protein were completed.

We believed that a functional smFRET system for the E2 protein would have provided valuable insight into how the HVR-1 functions, would have validated (or invalidated) the model we have proposed, and would have allowed for direct observations of how the HVR-1 functions. Had this system been completed, we would have been able to measure HVR-1 dynamics in solution, incorporate this smFRET system into the sE2:SR-B1 binding assay, and potentially disentangle the relationship between the HVR-1 and SR-B1. We may also have been able to uncover the relationship between the HVR-1 and the other co-receptors; relationships which were unable to be adequately parsed out in our assays.

Understanding the HVR-1, we believe, is one key step in understanding how to stop HCV virions from entering target cells. A functional smFRET system for the E2 protein of HCV would have deepened our understanding of the movements of the HVR-1 and how these contribute to function. To use a malaphor: this might have given us a better idea of which wrenches are required to be thrown in to which parts of the HCV E1E2 glycoproteins to produce an effective prophylactic vaccine. In other words, we would have had a clearer understanding of which parts of the HVR-1 and the rest of E2 are most amenable to disruption, preventing HCV viral fusion. Therefore, we believe that the development of an smFRET system for HCV E2 would have provided valuable insight into not only

the study of the molecular virology of HCV, but perhaps also into the development of an HCV vaccine itself. However, this is speculation on a system which was not fully developed, and therefore, we cannot truly know what information or insight would have been gleaned from an smFRET system for the E2 protein.

Other limitations to this work which should be considered in the future include: the effects of length polymorphisms in the HVR-1 were only studied in the wildtype J6 backbone, and the I438V A524T J6 backbone; most of our characterisation assays were performed in HCVpp; only glycine residues were inserted into the HVR-1.

To tackle the first limitation, these assays should be conducted in different backbones to examine whether the effects seen herein are strain specific. For example, repeating these assays in the H77 or S52 backbone would allow for assessing the generalisability of our central hypothesis. It may be found that, indeed, length polymorphisms in the HVR-1 lead to similar phenotypes across various different backbones. However, as we have noted, there is a conformational cross talk between the HVR-1 and the rest of the E2 protein. We posit that mutations within the HVR-1 and across the rest of the E2 protein work additively to fine-tune the function of the HVR-1. Therefore, it is just as plausible that length polymorphisms in the HVR-1 in H77 or S52 might render HCVpp with distinct phenotypes to those already described in the wildtype J6 backbone and the I438V A524T J6 backbone.

In considering HCVpp, it would be invaluable to evaluate these length polymorphisms in the HCVcc system. As noted previously, HCVpp lack the apolipoproteins which surround HCV LVPs, HCVcc do not. Repeating these experiments in the HCVcc system would allow for interrogation of the effects of the length polymorphisms in a more physiologically relevant system. Additionally, putting the length polymorphisms in the HCVcc system would allow for other experiments to be considered. One such experiment would be continual passaging of the HCVcc with length polymorphisms in culture with sequencing being performed throughout the continual passaging. This would

answer whether inserted residues are selected against; whether shorter HVR-1s regain their lost amino acid residues over time; if the HVR-1s remain the same length that they started at, do compensatory mutations arise elsewhere throughout the E1E2 complex? This would allow an examination of the evolutionary constraints on length polymorphisms in the HVR-1, and moreover, an investigation into the cross talk between the HVR-1 and the rest of the E2 protein through examination of compensatory mutations outwith the HVR-1.

Lastly, all of the inserted residues in the HVR-1 were glycine residues. Glycine was chosen due to its overrepresentation in the insertions from an initial analysis of the HVR-1 sequences that were extracted. Moreover, glycine residues are small, flexible, and comparatively chemically simple amino acid residues. Therefore, the logical next question to ask is whether replacing the inserted glycine residues with other residues would result in the same phenotypes. If we instead inserted alanine residues, or proline residues for example, would the phenotypes described be the same? This would allow us to delineate any effects intrinsic to the glycine residues themselves versus the change in length in the HVR-1.

6. 5. Final Remarks

Notably, and this is an important omission from this study, we still do not have a clear understanding of how HCV viral fusion occurs. We have proposed a model which aims to explain the molecular biology of one small, but important, section of one of two glycoproteins on HCV. This is not to diminish the work presented herein, but rather to highlight just how much we do not currently know about how HCV virions enter into cells. Thirty-seven years after the first characterisation of HCV (Choo *et al.*, 1989), we may have a clearer understanding as to how the HVR-1 functions. However, if there is one thing that I have learnt from my time studying HCV, it is that HCV is a virus which eludes simplicity or straightforward explanation, and therefore this proposed model may be slightly off-base or, even, entirely wrong.

This project has contributed 536 new models of the HCV E2 protein, and 500 new models of the E1E2 protein complex. Although these are predictions, what they allowed us to infer and validate experimentally demonstrates that these protein structure prediction software platforms produce models which can generate tractable and testable hypotheses, particularly about protein regions which have been classically difficult to visualise such as the HVR-1. The significance of these tools is perhaps best illustrated by the award of the 2024 Nobel Prize in Chemistry to Demis Hassabis and John M. Jumper for their work on protein structure prediction (The Royal Swedish Academy of Sciences, 2024). These tools for protein structure prediction are truly revolutionary for biology. Classical structural biology, whilst still an invaluable tool for validation, can take many, many years to arrive at the same structure that these tools can predict in minutes. There is still some way to go, however. As mentioned, none of the protein structure prediction software platforms tested were able to recapitulate the E1E2 dimer of dimers structure as resolved by Augestad *et al.* (2024). This, however, does not diminish the impact of protein structure predictions. Rather it highlights how these two tools should be used complementarily.

Additionally, if our hypothesis about the pocketing of the HVR-1 turns out to be true, this may have direct implications on HCV vaccine design. We believe that

the conformational state of the HVR-1 influences the global conformational state of the E2 protein, and perhaps the E1E2 complex as a whole. Therefore, future approaches to rational HCV vaccine design may wish to consider designing HCV antigens which stabilise (or mimic) a specific conformational state of E2 against which broadly neutralising antibodies against the “closed” E2 conformation, which appears to be induced partly by a pocketed HVR-1, can be raised.

Understanding the conformational behaviour of the HVR-1 and how it affects the overall conformation of the E1E2 complex is, we believe, essential for designing antigens that correctly present HCV E1E2 epitopes. However, there are many ways this can be approached, and further work would be required to delineate how this is achieved. For example, one thought worth exploring is whether the ability to pocket the HVR-1 is intrinsic to the sequence and length of the HVR-1, or if it is also dependent on the conformation and sequence of the pocket itself. Perhaps one approach to designing future immunogens would be to fill the pocket with bulky amino acid residues such as tyrosine or tryptophan to establish whether this stabilises the “closed” E2 conformation. This would allow for testing antibodies against a “closed” E2 without the stochastic and unpredictable nature of the HVR-1.

To conclude, this thesis set out to characterise the structural and functional consequences of HVR-1 length polymorphisms, to examine their relationship with SR-B1 dependency and antibody sensitivity, and to begin developing direct biophysical tools to observe HVR-1 dynamics. The first two aims were met: through the largest bioinformatic analysis of HVR-1 length polymorphisms to date, the generation of the first systematic structural hypotheses about HVR-1 conformational behaviour using modern computational tools, and rigorous functional virology, we have placed these hypotheses on solid experimental footing. The central finding - that HVR-1 pocketing competence, rather than length per se, is the mechanistic lever governing SR-B1 dependency and antibody sensitivity - advances our understanding of the earliest steps of HCV entry and immune evasion and has direct implications for structure-based vaccine antigen design. The third aim, the development of a functional smFRET system for direct observation of HVR-1 dynamics, remains partially complete - the groundwork has

been laid, but its full realisation and the validation of our proposed model in a more physiologically relevant HCVcc system represent the most important next steps for this field.

This work, we believe, furthers our understanding of the HVR-1 and its functions. The elimination of HCV remains one of the great challenges of modern virology, and whilst this thesis represents one contribution among many that will be required to meet that challenge, we do hope that it will aid in that goal in whatever small way it may.

Appendix I

Amino Acid Numbering Schema

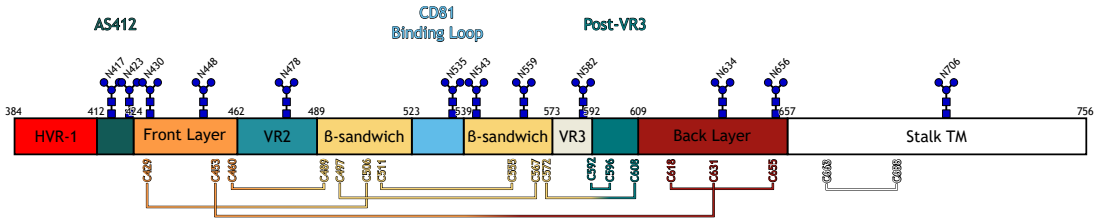
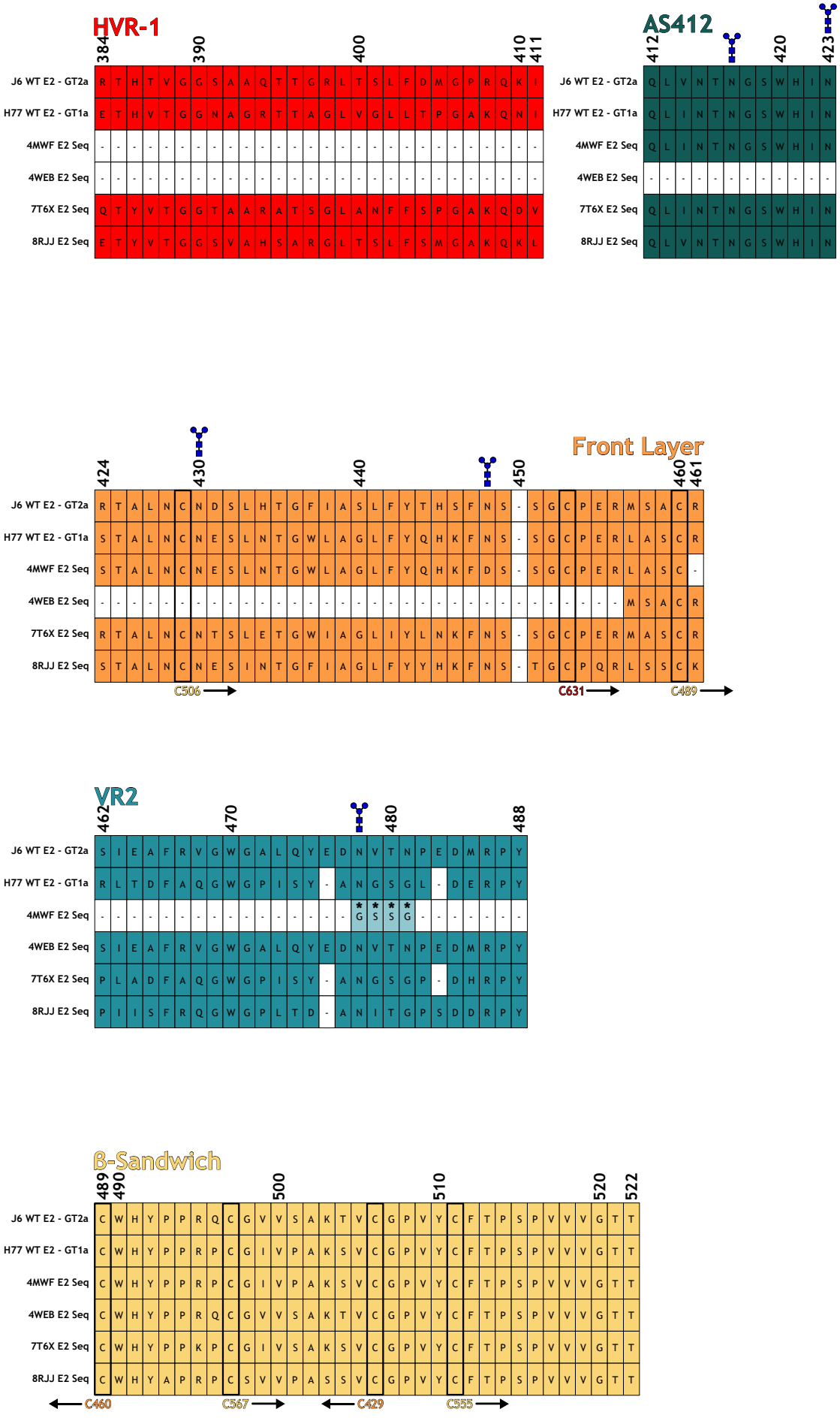
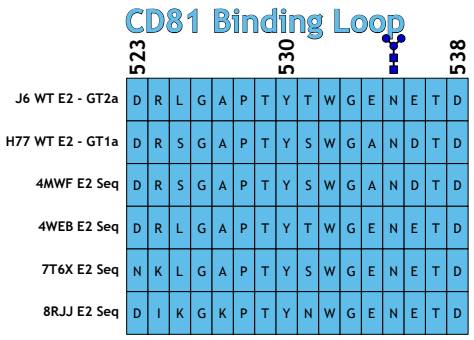
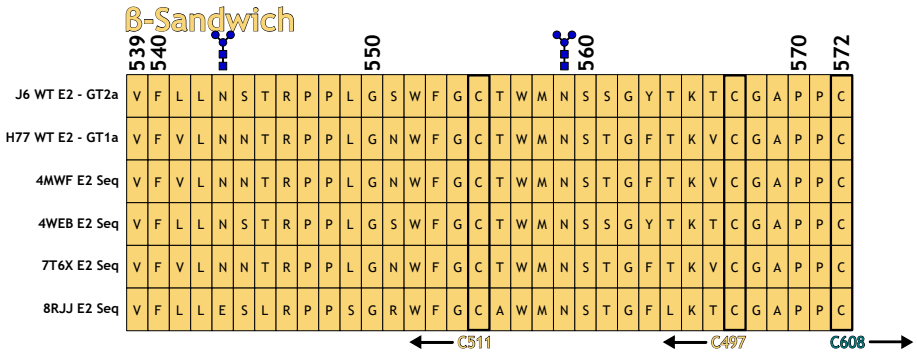
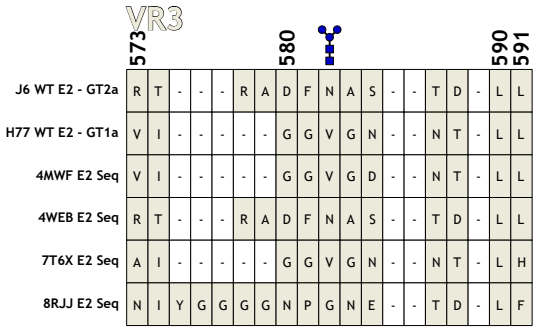


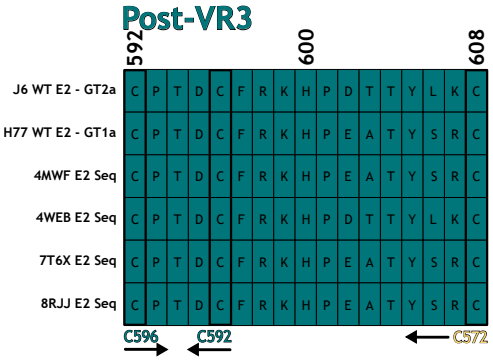
Figure I. I. **Schematic Representation of the E2 Protein Architecture.** Aligned E2 sequences are coloured as per the domains in this schematic. Glycan positions are marked above, disulphide bonds below. Residues re-numbered as per the alignments in figure I. II. and I. III.

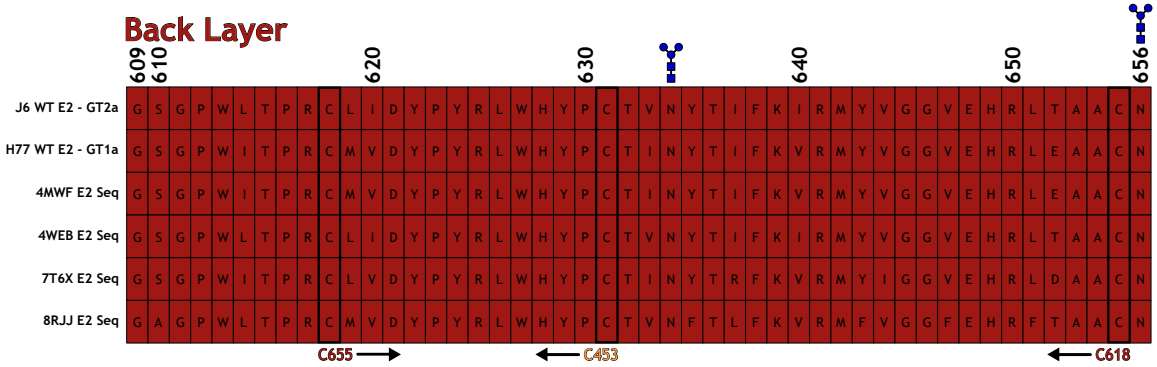

HVR-1
AS412
Front Layer
VR2
β-Sandwich
J6 WT E2 - GT2a
H77 WT E2 - GT1a
4MWF E2 Seq
4WEB E2 Seq
7T6X E2 Seq
8RJJ E2 Seq
C506
C631
C489
C460
C567
C429
C555


CD81 Binding Loop
523
530
538
J6 WT E2 - GT2a D R L G A P T Y T W G E N E T D
H77 WT E2 - GT1a D R S G A P T Y S W G A N D T D
4MWF E2 Seq D R S G A P T Y S W G A N D T D
4WEB E2 Seq D R L G A P T Y T W G E N E T D
7T6X E2 Seq N K L G A P T Y S W G E N E T D
8RJJ E2 Seq D I K G K P T Y N W G E N E T D


β-Sandwich
539
540
550
560
570
572
J6 WT E2 - GT2a V F L L N S T R P P L G S W F G C T W M N S S G Y T K T C G A P P C
H77 WT E2 - GT1a V F V L N N T R P P L G N W F G C T W M N S T G F T K V C G A P P C
4MWF E2 Seq V F V L N N T R P P L G N W F G C T W M N S T G F T K V C G A P P C
4WEB E2 Seq V F L L N S T R P P L G S W F G C T W M N S S G Y T K T C G A P P C
7T6X E2 Seq V F V L N N T R P P L G N W F G C T W M N S T G F T K V C G A P P C
8RJJ E2 Seq V F L L E S L R P P S G R W F G C A W M N S T G F L K T C G A P P C
C511
C497
C608


VR3
573
580
590
591
J6 WT E2 - GT2a R T - - - R A D F N A S - - T D - L L
H77 WT E2 - GT1a V I - - - - - G G V G N - - N T - L L
4MWF E2 Seq V I - - - - - G G V G D - - N T - L L
4WEB E2 Seq R T - - - R A D F N A S - - T D - L L
7T6X E2 Seq A I - - - - - G G V G N - - N T - L H
8RJJ E2 Seq N I Y G G G G N P G N E - - T D - L F


Post-VR3
592
600
608
J6 WT E2 - GT2a C P T D C F R K H P D T T Y L K C
H77 WT E2 - GT1a C P T D C F R K H P E A T Y S R C
4MWF E2 Seq C P T D C F R K H P E A T Y S R C
4WEB E2 Seq C P T D C F R K H P D T T Y L K C
7T6X E2 Seq C P T D C F R K H P E A T Y S R C
8RJJ E2 Seq C P T D C F R K H P E A T Y S R C
C596
C592
C572


Back Layer
609
610
620
630
640
650
656
J6 WT E2 - GT2a G S G P W L T P R C L I D Y P Y R L W H Y P C T V N Y T I F K I R M Y V G G V E H R L T A A C N
H77 WT E2 - GT1a G S G P W I T P R C M V D Y P Y R L W H Y P C T I N Y T I F K V R M Y V G G V E H R L E A A C N
4MWF E2 Seq G S G P W I T P R C M V D Y P Y R L W H Y P C T I N Y T I F K V R M Y V G G V E H R L E A A C N
4WEB E2 Seq G S G P W L T P R C L I D Y P Y R L W H Y P C T V N Y T I F K I R M Y V G G V E H R L T A A C N
7T6X E2 Seq G S G P W I T P R C L V D Y P Y R L W H Y P C T I N Y T R F K V R M Y I G G V E H R L D A A C N
8RJJ E2 Seq G A G P W L T P R C M V D Y P Y R L W H Y P C T V N F T L F K V R M F V G G F E H R F T A A C N
C655
C453
C618

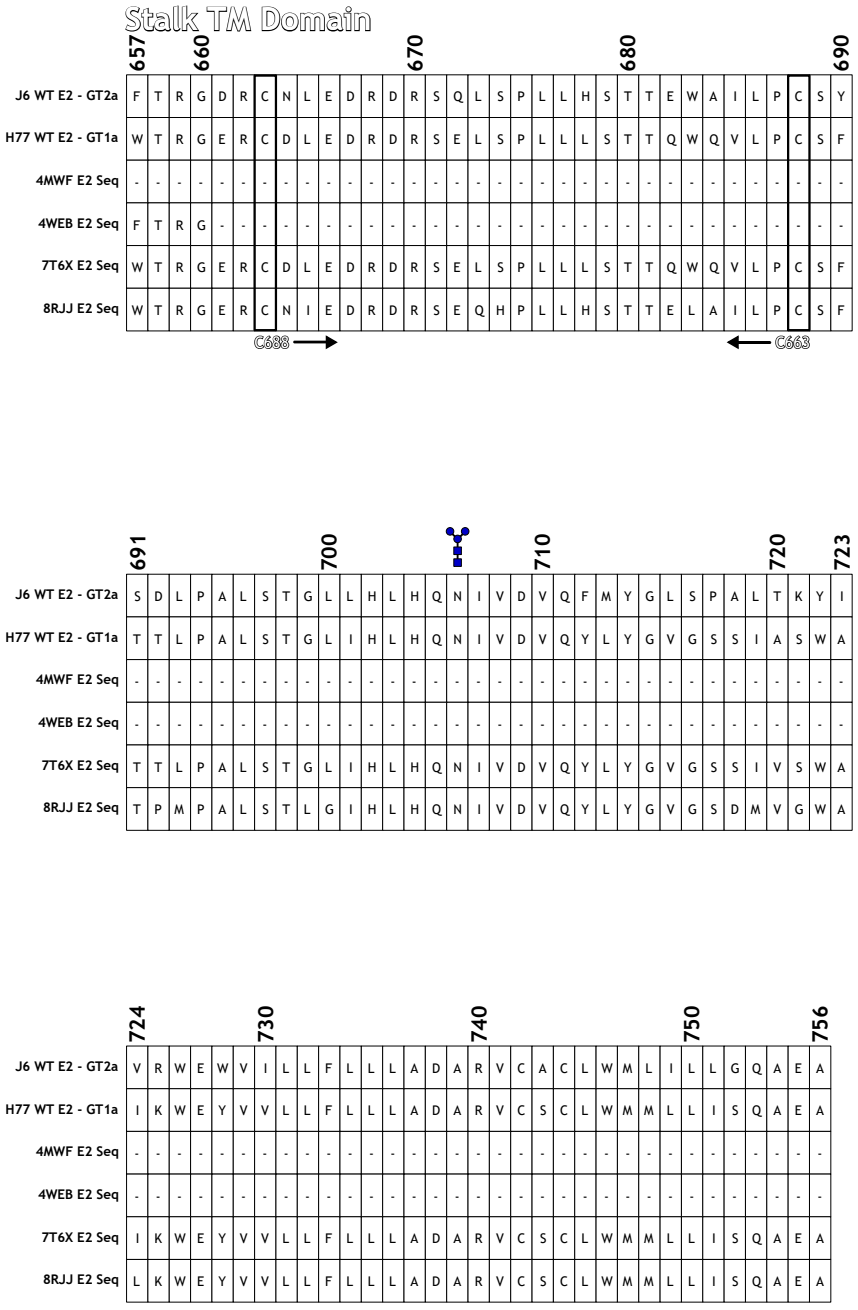
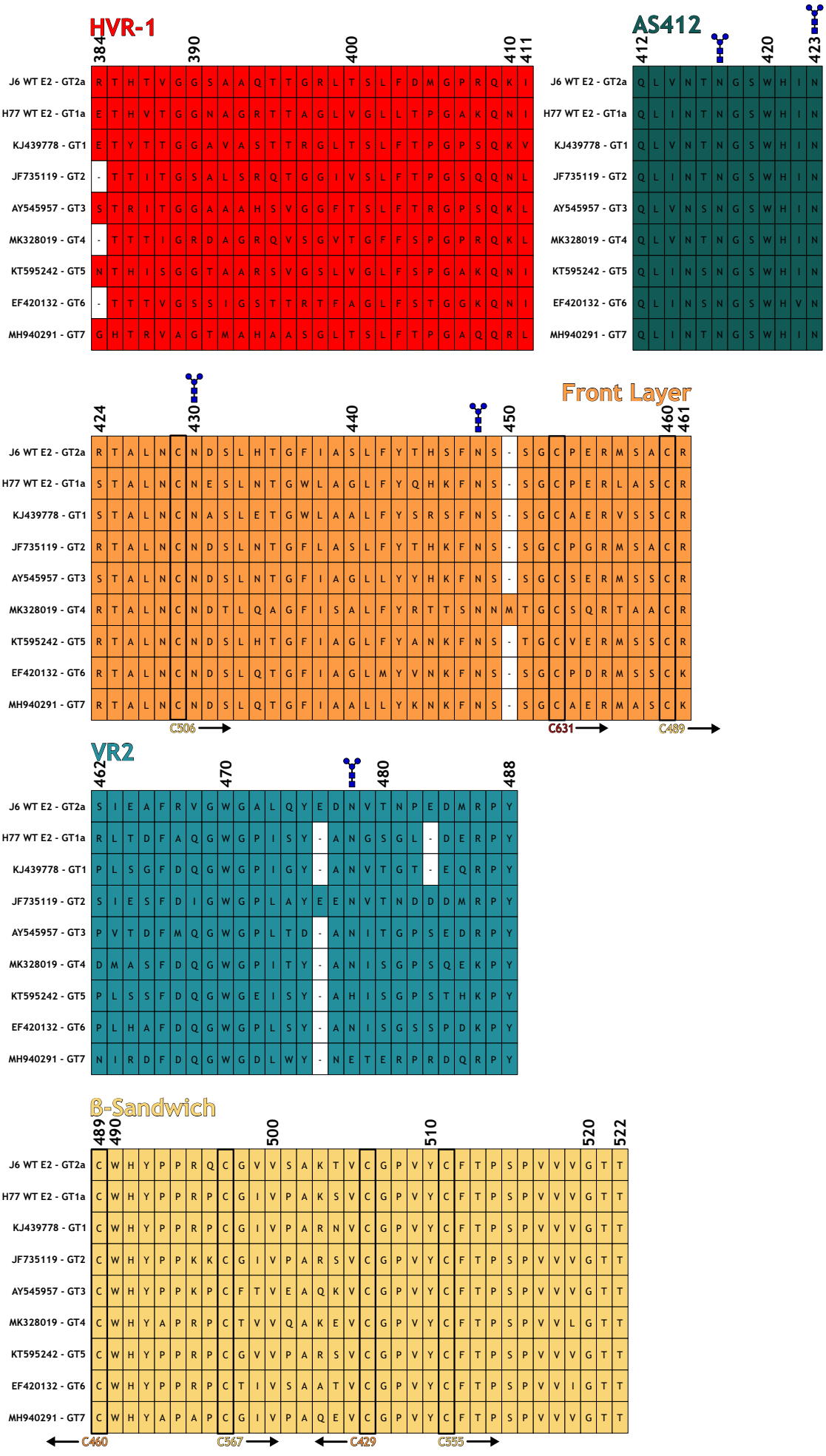
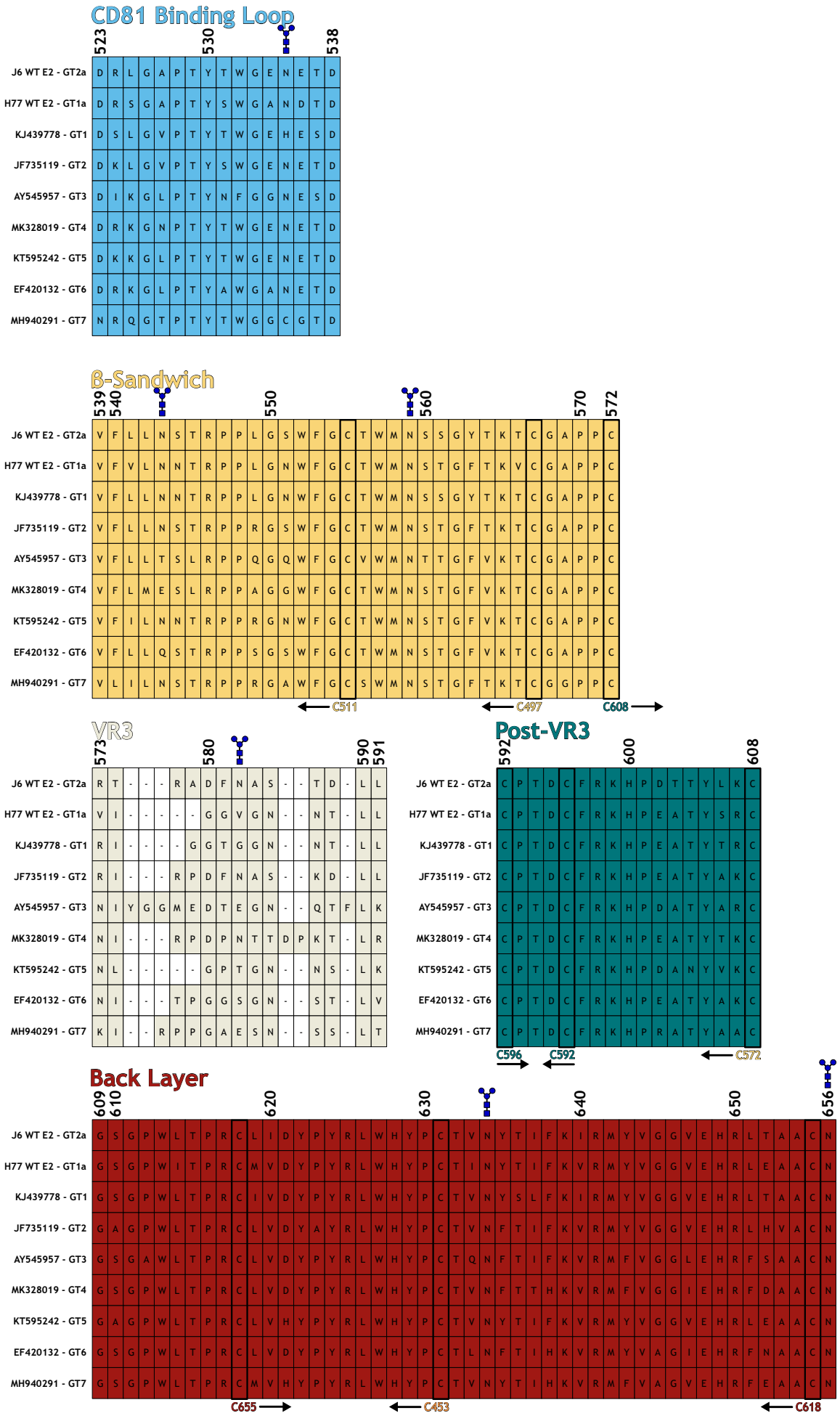


Figure I. II. **Alignment of the E2 Sequences**. For consistent residue numbering across this thesis, all of the E2 sequences directly mentioned in the text have been aligned using MAFFT. Residues in the E2/E1E2 structures referenced in the Introduction have been re-numbered as per this alignment. Glycans are demarked above the alignment with a schematic representation. Cysteine residues highlighted by a black box form disulphide bridges with their canonical partner cysteine which is noted below the alignment. Note that the residues demarked with a * for 4MWF in VR2 represent non-canonical residues which were cloned into 4MWF to aid in crystallisation by replacing the flexible loop of the VR2 which these four residues.


HVR-1
AS412
Front Layer
VR2
ß-Sandwich
J6 WT E2 - GT2a
H77 WT E2 - GT1a
KJ439778 - GT1
JF735119 - GT2
AY545957 - GT3
MK328019 - GT4
KT595242 - GT5
EF420132 - GT6
MH940291 - GT7
C506
C631
C489
C460
C567
C429
C555


CD81 Binding Loop
523
530
538
J6 WT E2 - GT2a D R L G A P T Y T W G E N E T D
H77 WT E2 - GT1a D R S G A P T Y S W G A N D T D
KJ439778 - GT1 D S L G V P T Y T W G E H E S D
JF735119 - GT2 D K L G V P T Y S W G E N E T D
AY545957 - GT3 D I K G L P T Y N F G G N E S D
MK328019 - GT4 D R K G N P T Y T W G E N E T D
KT595242 - GT5 D K K G L P T Y T W G E N E T D
EF420132 - GT6 D R K G L P T Y A W G A N E T D
MH940291 - GT7 N R Q G T P T Y T W G G C G T D
ß-Sandwich
539
540
550
560
570
572
J6 WT E2 - GT2a V F L L N S T R P P L G S W F G C T W M N S S G Y T K T C G A P P C
H77 WT E2 - GT1a V F V L N N T R P P L G N W F G C T W M N S T G F T K V C G A P P C
KJ439778 - GT1 V F L L N N T R P P L G N W F G C T W M N S S G Y T K T C G A P P C
JF735119 - GT2 V F L L N S T R P P R G S W F G C T W M N S T G F T K T C G A P P C
AY545957 - GT3 V F L L T S L R P P Q G Q W F G C V W M N T T G F V K T C G A P P C
MK328019 - GT4 V F L M E S L R P P A G G W F G C T W M N S T G F V K T C G A P P C
KT595242 - GT5 V F I L N N T R P P R G N W F G C T W M N S T G F V K T C G A P P C
EF420132 - GT6 V F L L Q S T R P P S G S W F G C T W M N S T G F V K T C G A P P C
MH940291 - GT7 V L I L N S T R P P R G A W F G C S W M N S T G F T K T C G G P P C
C511
C497
C608
VR3
573
580
590
591
J6 WT E2 - GT2a R T - - - R A D F N A S - - T D - L L
H77 WT E2 - GT1a V I - - - - - G G V G N - - N T - L L
KJ439778 - GT1 R I - - - - G G T G G N - - N T - L L
JF735119 - GT2 R I - - - R P D F N A S - - K D - L L
AY545957 - GT3 N I Y G G M E D T E G N - - Q T F L K
MK328019 - GT4 N I - - - R P D P N T T D P K T - L R
KT595242 - GT5 N L - - - - - G P T G N - - N S - L K
EF420132 - GT6 N I - - - T P G G S G N - - S T - L V
MH940291 - GT7 K I - - R P P G A E S N - - S S - L T
Post-VR3
592
600
608
J6 WT E2 - GT2a C P T D C F R K H P D T T Y L K C
H77 WT E2 - GT1a C P T D C F R K H P E A T Y S R C
KJ439778 - GT1 C P T D C F R K H P E A T Y T R C
JF735119 - GT2 C P T D C F R K H P E A T Y A K C
AY545957 - GT3 C P T D C F R K H P D A T Y A R C
MK328019 - GT4 C P T D C F R K H P E A T Y T K C
KT595242 - GT5 C P T D C F R K H P D A N Y V K C
EF420132 - GT6 C P T D C F R K H P E A T Y A K C
MH940291 - GT7 C P T D C F R K H P R A T Y A A C
C596
C592
C572
Back Layer
609
610
620
630
640
650
656
J6 WT E2 - GT2a G S G P W L T P R C L I D Y P Y R L W H Y P C T V N Y T I F K I R M Y V G G V E H R L T A A C N
H77 WT E2 - GT1a G S G P W I T P R C M V D Y P Y R L W H Y P C T I N Y T I F K V R M Y V G G V E H R L E A A C N
KJ439778 - GT1 G S G P W L T P R C I V D Y P Y R L W H Y P C T V N Y S L F K I R M Y V G G V E H R L T A A C N
JF735119 - GT2 G A G P W L T P R C L V D Y A Y R L W H Y P C T V N F T I F K V R M Y V G G V E H R L H V A C N
AY545957 - GT3 G S G A W L T P R C L V D Y P Y R L W H Y P C T Q N F T I F K V R M F V G G L E H R F S A A C N
MK328019 - GT4 G S G P W L T P R C L V D Y P Y R L W H Y P C T V N F T T H K V R M F V G G I E H R F D A A C N
KT595242 - GT5 G A G P W L T P R C L V H Y P Y R L W H Y P C T V N Y T I F K V R M Y V G G V E H R L E A A C N
EF420132 - GT6 G S G P W L T P R C L V D Y P Y R L W H Y P C T L N F T I H K V R M Y V A G I E H R F N A A C N
MH940291 - GT7 G S G P W L T P R C M V H Y P Y R L W H Y P C T V N Y T I H K V R M F V A G V E H R F E A A C N
C655
C453
C618

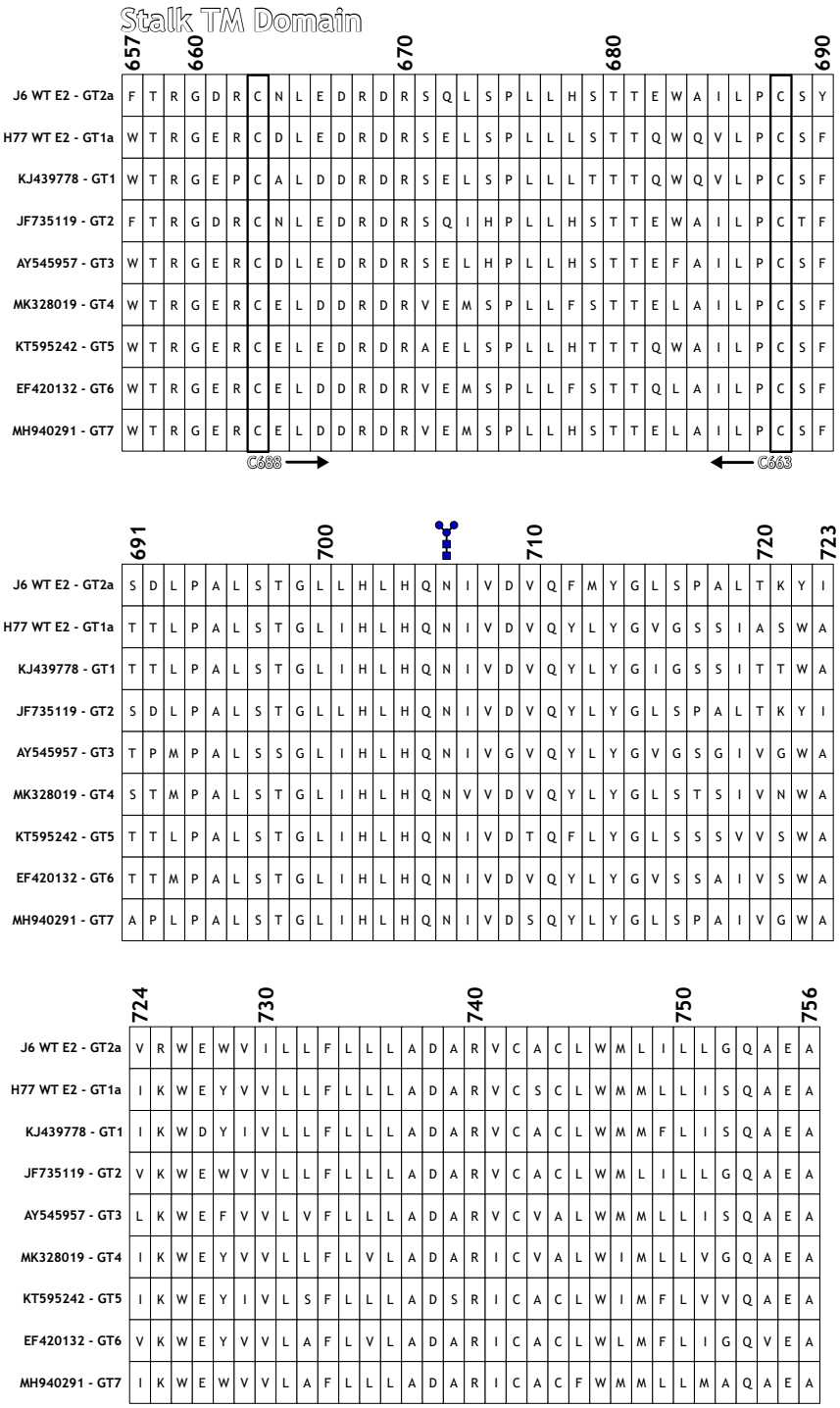


Figure I. III. **Alignment of the E2 Sequences.** For consistent residue numbering across this thesis, all of the E2 sequences directly mentioned in the text have been aligned using MAFFT. Residues in the E2/E1E2 structures referenced in the Chapter 3 have been re-numbered as per this alignment. Glycans are demarked above the alignment with a schematic representation. Cysteine residues highlighted by a black box form disulphide bridges with their canonical partner cysteine which is noted below the alignment.

Appendix II

Plasmid Maps

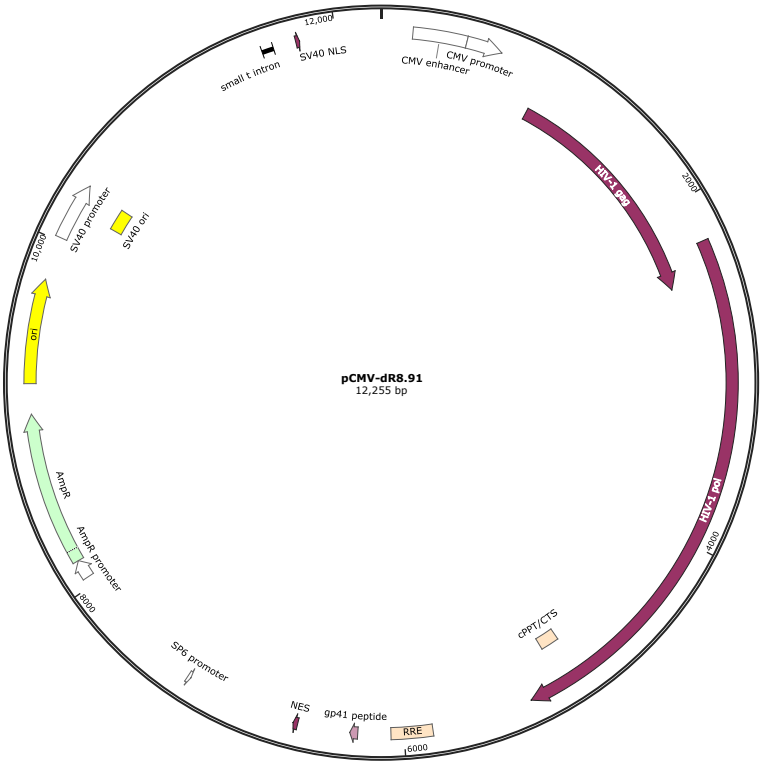


Figure II. I. **pCMVΔR8.91 Plasmid Map.** Plasmid maps were generated and annotated in SnapGene software (www.snapgene.com). The pCMVΔR8.91 contains the HIV-1 *gag-pol* genes. Critically, the packaging signal (Ψ) was removed from this plasmid, and added to the luciferase reporter plasmid for efficient packaging of the luciferase gene. This plasmid was used in conjunction with the CSLW (luciferase-containing plasmid), and the pD607 HCV E1E2 plasmids to produce HCVpp.

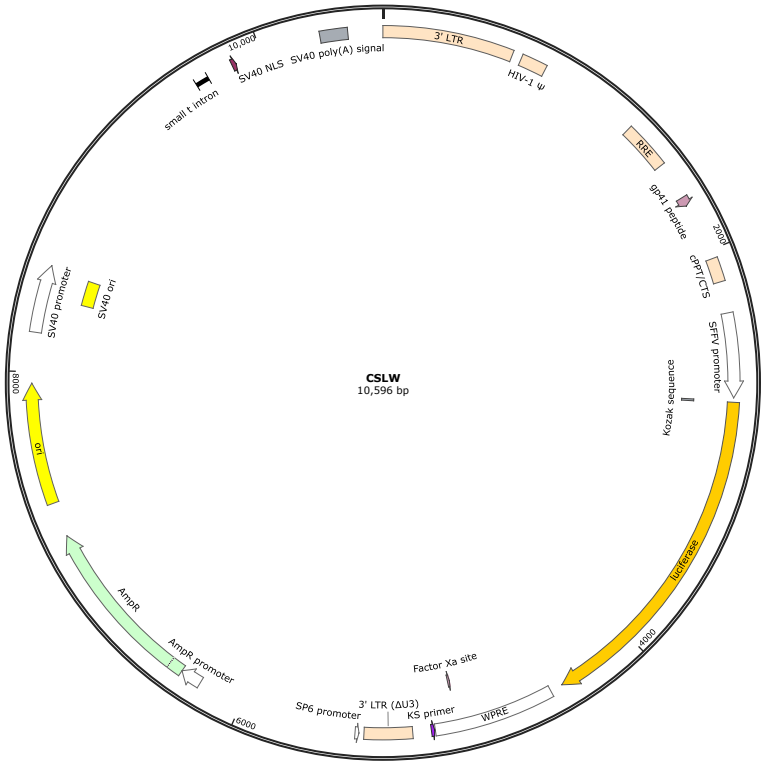


Figure II. II. **CSLW Plasmid Map.** Plasmid maps were generated and annotated in SnapGene software (www.snapgene.com). The CSLW contains the luciferase gene, and a cis-acting HIV-1 packaging signal (Ψ). This plasmid was used in conjunction with the pCMVΔR8.91, and the pD607 HCV E1E2 plasmids to produce HCVpp.

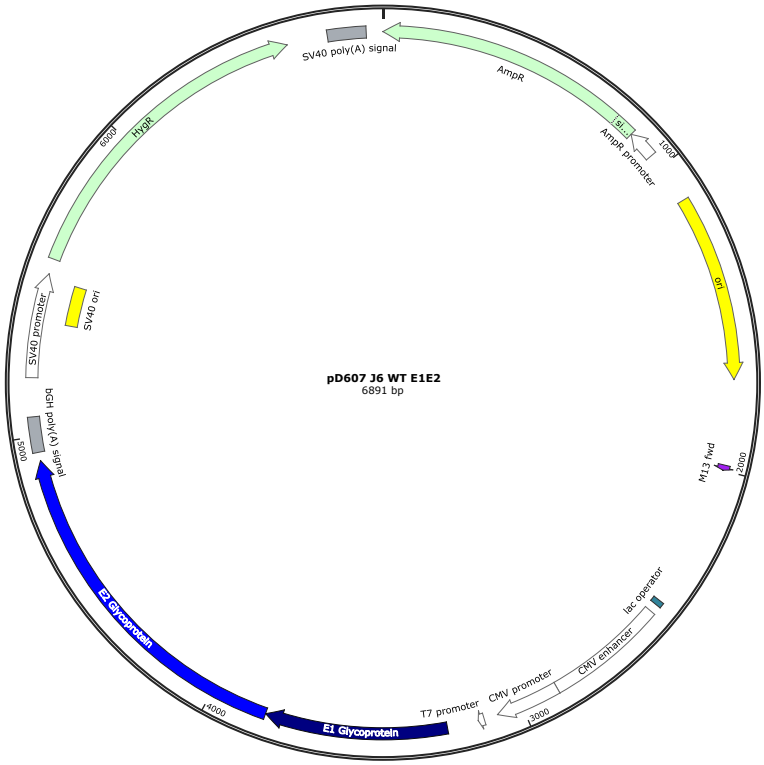


Figure II. III. **pD607 J6 WT E1E2 Plasmid Map.** Plasmid maps were generated and annotated in SnapGene software (www.snapgene.com). The pD607 J6 WT E1E2 contains the E1E2 protein coding sequences. This was the parental plasmid into which the length polymorphisms were cloned to produce the J6-backbone HVR-1 length polymorphism constructs. This plasmid was used in conjunction with the pCMVΔR8.91, and the CSLW plasmids to produce HCVpp.

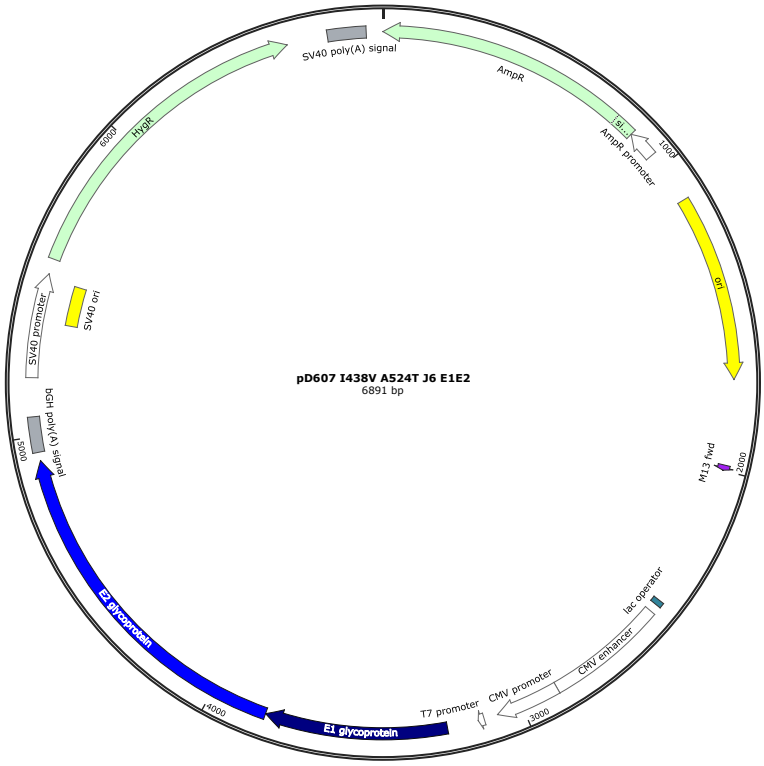


Figure II. IV. **pD607 I438V A524T J6 WT E1E2 Plasmid Map.** Plasmid maps were generated and annotated in SnapGene software (www.snapgene.com). The pD607 I428V A524T J6 WT E1E2 contains the E1E2 protein coding sequences. This was the parental plasmid into which the length polymorphisms were cloned to produce the I438V A524T J6-backbone HVR-1 length polymorphism constructs. This plasmid was used in conjunction with the pCMVΔR8.91, and the CSLW plasmids to produce HCVpp.

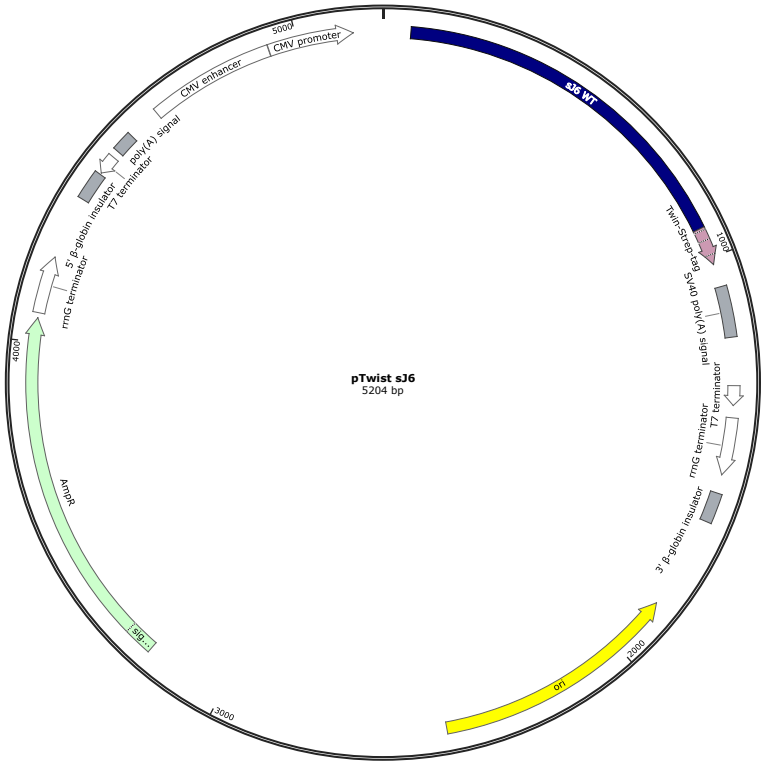


Figure II. V. **pTwist sJ6 Plasmid Map.** Plasmid maps were generated and annotated in SnapGene software (www.snapgene.com). The pTwist sJ6 plasmid contains the coding sequence for the soluble J6 construct (residue 384 to 672, as per residue re-numbering in appendix I) with a Twin-Strep tag at the C-terminus to allow for protein purification. Individual plasmids with the coding sequences for the sI438V A524T J6, sΔHVR-1 J6, sJ6 -3 N-Terminus Deletion, sJ6 +3 N-Terminus Extension, sJ6 +4 Internal Insertion, sI438V A524T J6 -3 N-Terminus Deletion, sI438V A524T J6 +3 N-Terminus Extension, and sI438V A524T J6 +4 Internal Insertion constructs were purchased directly from Twist Bioscience and cloned into the pTwist vector.

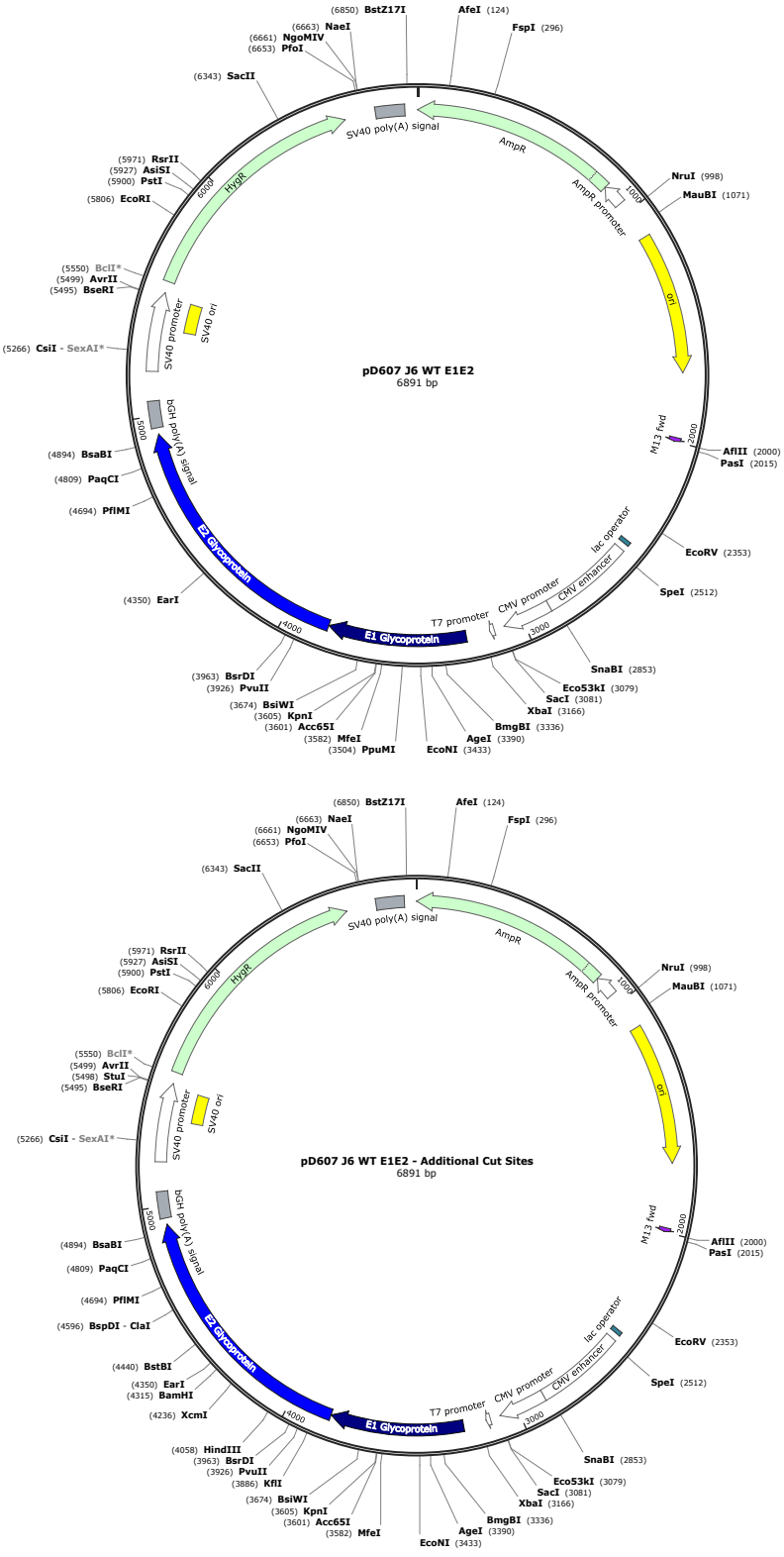


Figure II. VI. **Addition of Silent pD607 J6 WT E1E2.** Plasmid maps were generated and annotated in SnapGene software (www.snapgene.com). During the development of the E2 smFRET system, a version of the pD607 J6 WT E1E2 plasmid was cloned into the parental pD607 J6 WT E1E2 plasmid wherein additional silent cut sites were added across the E2 protein. This would have better facilitated downstream cloning efforts in the process of developing the smFRET system. The silent cut site mutations were designed in SnapGene, the gene fragment ordered from Twist Bioscience and cloned into the parental plasmid.

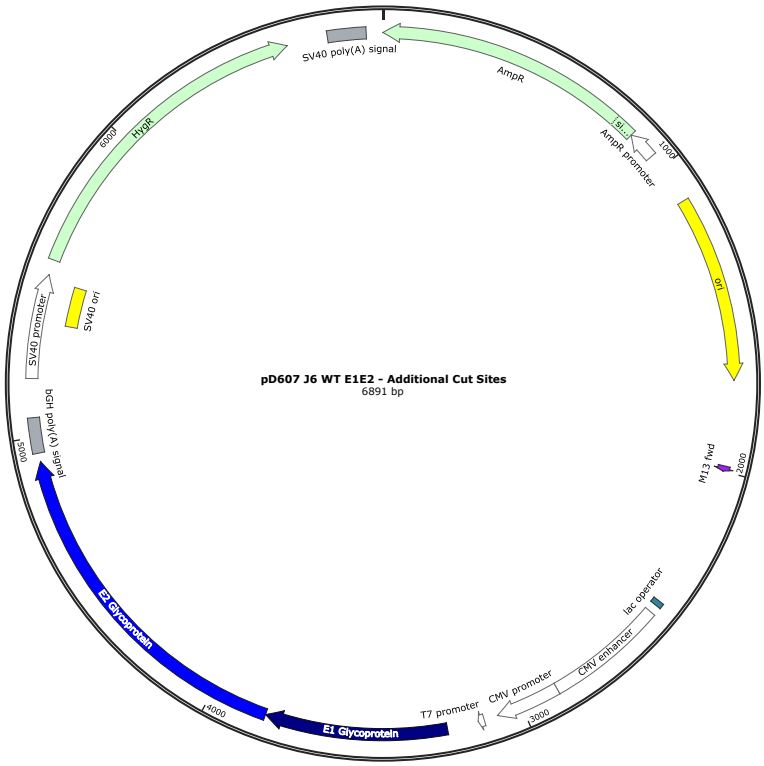


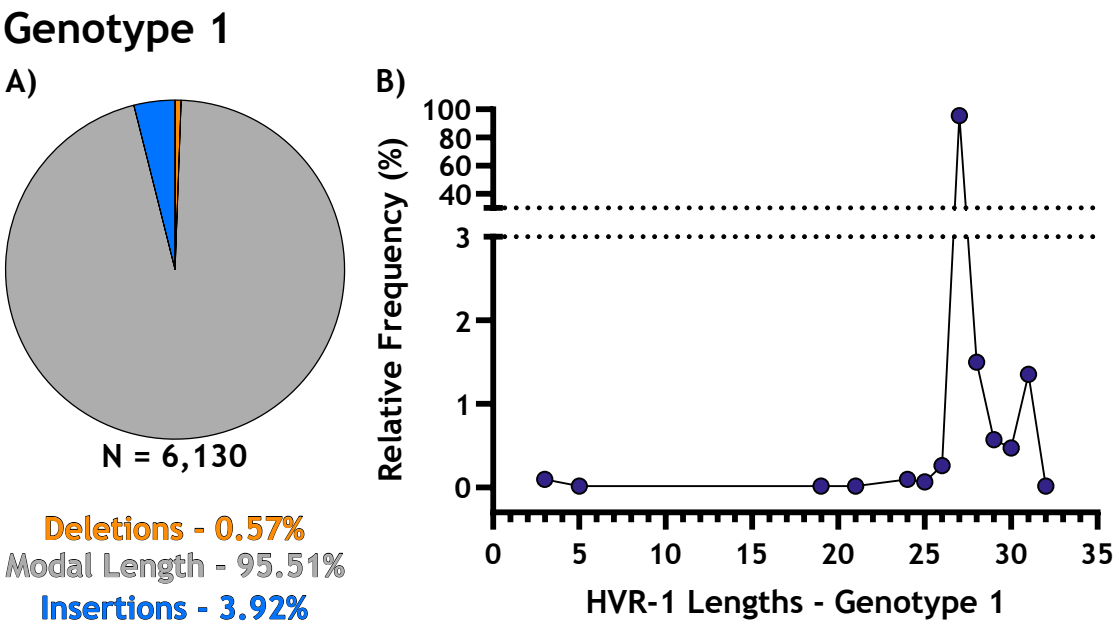
Figure II. VII. **pD607 J6 WT E1E2 - Additional Cut Sites (pD607 J6 Plus E1E2) Plasmid Map.** Plasmid maps were generated and annotated in SnapGene software (www.snapgene.com). This plasmid encodes J6 WT E1E2. However, additional cut sites have been added across the E2 protein via silent mutations in the nucleotide sequence. Silent mutations were planned and designed in SnapGene, the $E2^{silent\ mutations}$ gene was ordered from Twist Biosciences, and cloned into the pD607 J6 WT E1E2 plasmid backbone. This plasmid, termed pD607 J6 Plus E1E2, was used for cloning out specific glycan-linked asparagine residue codons and replacing them with Amber stop codons (TAG). This allowed for site-directed UAA (TCO*A) incorporation into the E2 protein. During this project, only two glycan-linked asparagine residues were able to be successfully substituted with Amber stop (TAG) codons: N582 and N656 (residue numbers as per the alignment in appendix I). The additional cut sites were introduced to ease the swapping of E2 gene segments between plasmids. For example, the additional cut sites would have provided more options for subcloning both the $N582^{TAG}$ and the $N656^{TAG}$ gene segments into one plasmid which would potentially be suitable for smFRET studies. HCVpp were produced with the $N582^{TAG}$ and $N656^{TAG}$ pD607 J6 Plus E1E2 plasmids in conjunction with the pCMVΔR8.91 plasmid, the CSLW plasmid, and the pcDNA3.1(+)_U6 tRNAPyl_CMV NESPylRS(AF) plasmid which would allow for the production of HCVpp with TCO*A incorporated into the E2 protein at the TAG codon sites. These incorporated TCO*A residues could undergo copper-free SPIEDAC reactions with tetrazide-containing FRET-capable fluorophores.



Figure II. VIII. **pcDNA3.1(+)_U6 tRNAPyl_CMV NESPylRS(AF) Plasmid Map.** Plasmid maps were generated and annotated in SnapGene software (www.snapgene.com). This plasmid encodes a codon-optimised Y306A Y384F mutant of *Methanosarcina mazei*-derived pyrrolysine (Pyl) tRNA synthetase fused to a nuclear export signal (NESPylRSAF) & tRNACUAPyl. This plasmid was co-transfected along with the pCMVΔR8.91 plasmid, the CSLW plasmid, and the pD607 J6 Plus E1E2-derived plasmids containing Amber stop (TAG) codons in place of glycan-linked asparagine residues.

Appendix III

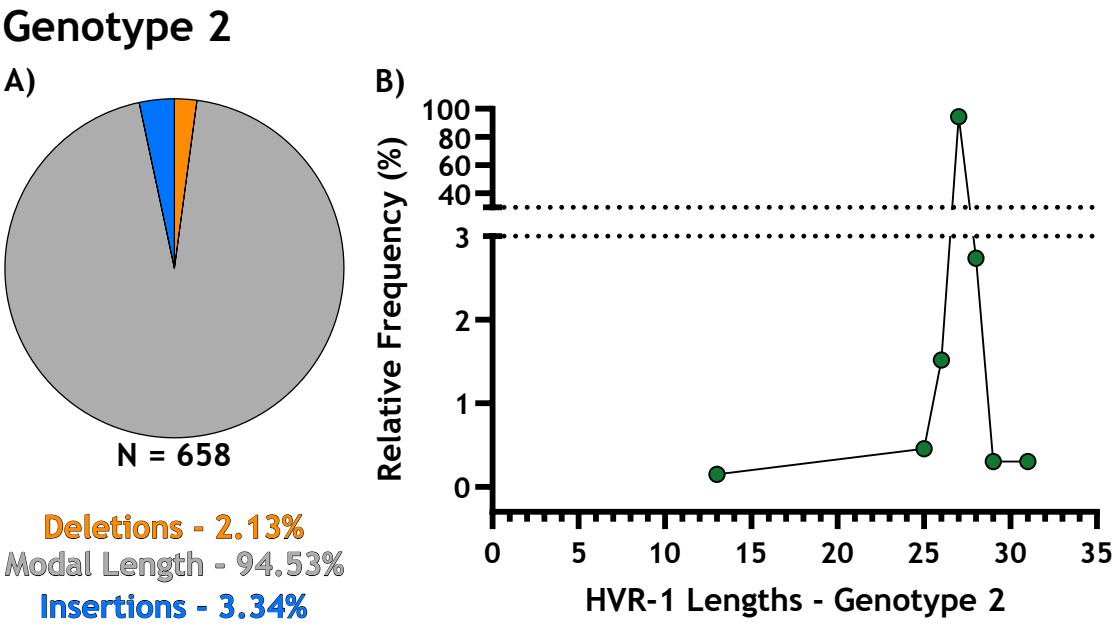
HVR-1 Lengths by Genotype



C)

HVR-1 Lengths	Frequency of HVR-1 Lengths	Relative Frequency (%)
3	6	0.10
5	1	0.02
19	1	0.02
21	1	0.02
24	6	0.10
25	4	0.07
26	16	0.26
27	5,855	95.51
28	92	1.50
29	35	0.57
30	29	0.47
31	83	1.35
32	1	0.02

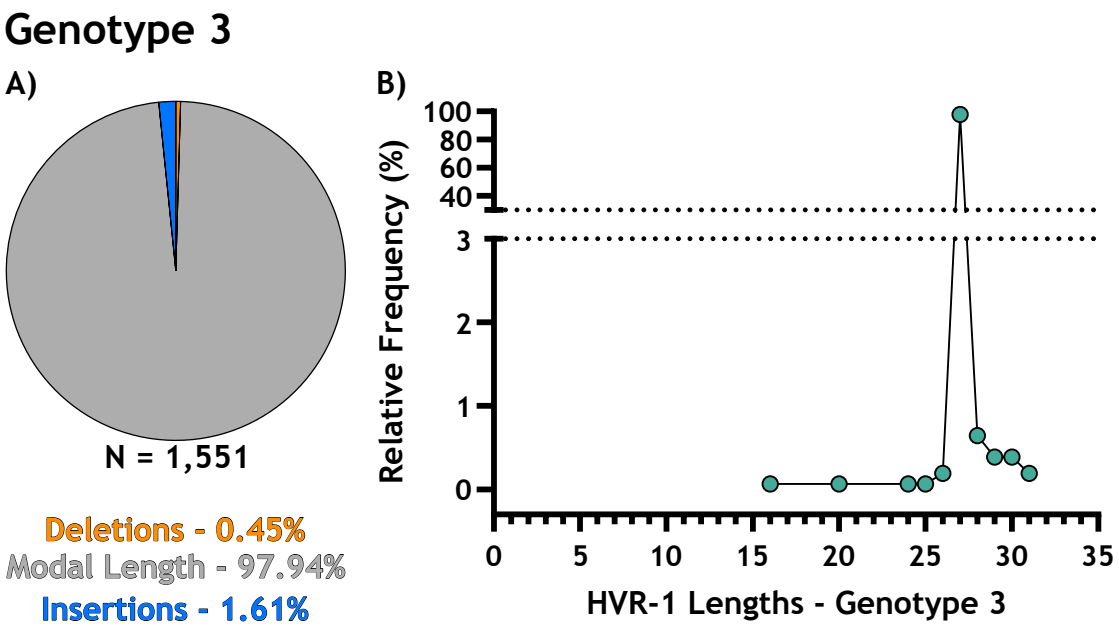
Figure III. I. **HVR-1 Length Polymorphisms in Genotype 1.** A) Proportion of HVR-1 lengths which are modal length, contain insertions, or contain deletions. B) HVR-1 lengths graphed against percentage frequency. Note that the y-axis, showing percentage frequency, are split with the bottom axes running from 0 to 3%, the top axes from 30 to 100%. The x-axis is the HVR-1 lengths. C) Table of the figures used to generate B), showing the number of each length HVR-1.



C)

HVR-1 Lengths	Frequency of HVR-1 Lengths	Relative Frequency (%)
13	1	0.15
25	3	0.46
26	10	1.52
27	622	94.53
28	18	2.74
29	2	0.30
31	2	0.30

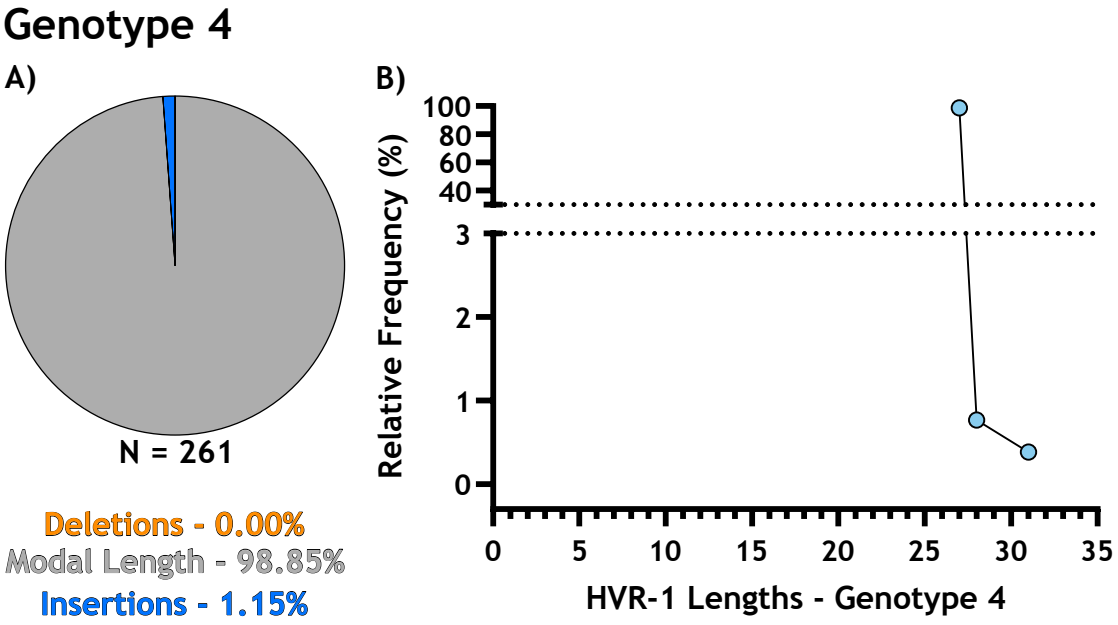
Figure III. II. **HVR-1 Length Polymorphisms in Genotype 2.** A) Proportion of HVR-1 lengths which are modal length, contain insertions, or contain deletions. B) HVR-1 lengths graphed against percentage frequency. Note that the y-axis, showing percentage frequency, are split with the bottom axes running from 0 to 3%, the top axes from 30 to 100%. The x-axis is the HVR-1 lengths. C) Table of the figures used to generate B), showing the number of each length HVR-1.



C)

HVR-1 Lengths	Frequency of HVR-1 Lengths	Relative Frequency (%)
16	1	0.06
20	1	0.06
24	1	0.06
25	1	0.06
26	3	0.19
27	1519	97.94
28	10	0.64
29	6	0.39
30	6	0.39
31	3	0.19

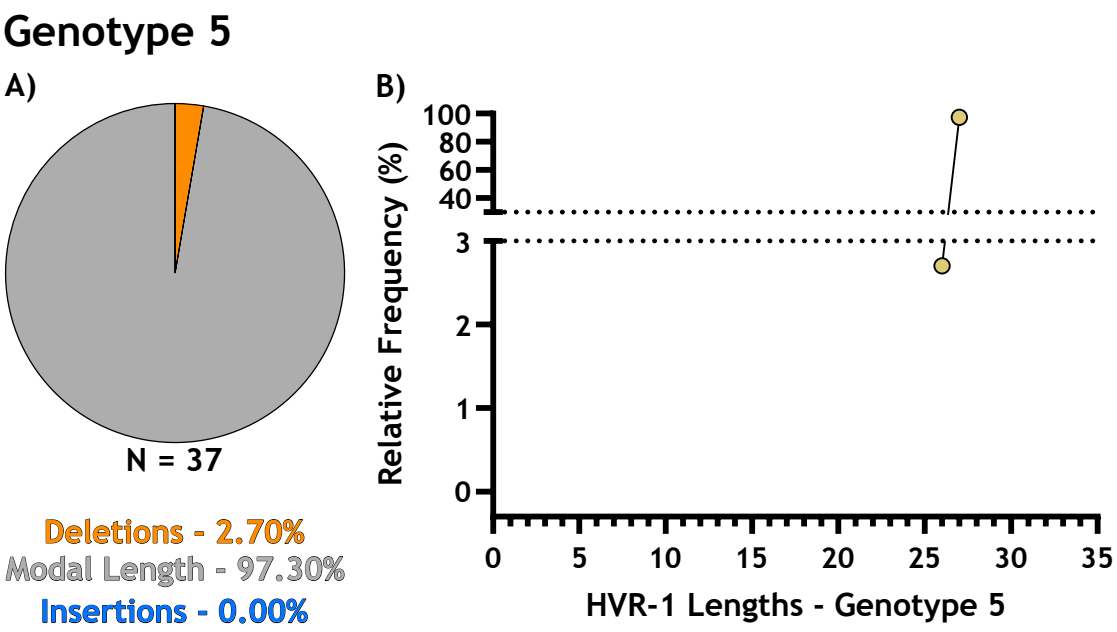
Figure III. III. **HVR-1 Length Polymorphisms in Genotype 3.** A) Proportion of HVR-1 lengths which are modal length, contain insertions, or contain deletions. B) HVR-1 lengths graphed against percentage frequency. Note that the y-axis, showing percentage frequency, are split with the bottom axes running from 0 to 3%, the top axes from 30 to 100%. The x-axis is the HVR-1 lengths. C) Table of the figures used to generate B), showing the number of each length HVR-1.



C)

HVR-1 Lengths	Frequency of HVR-1 Lengths	Relative Frequency (%)
27	258	98.85
28	2	0.77
31	1	0.38

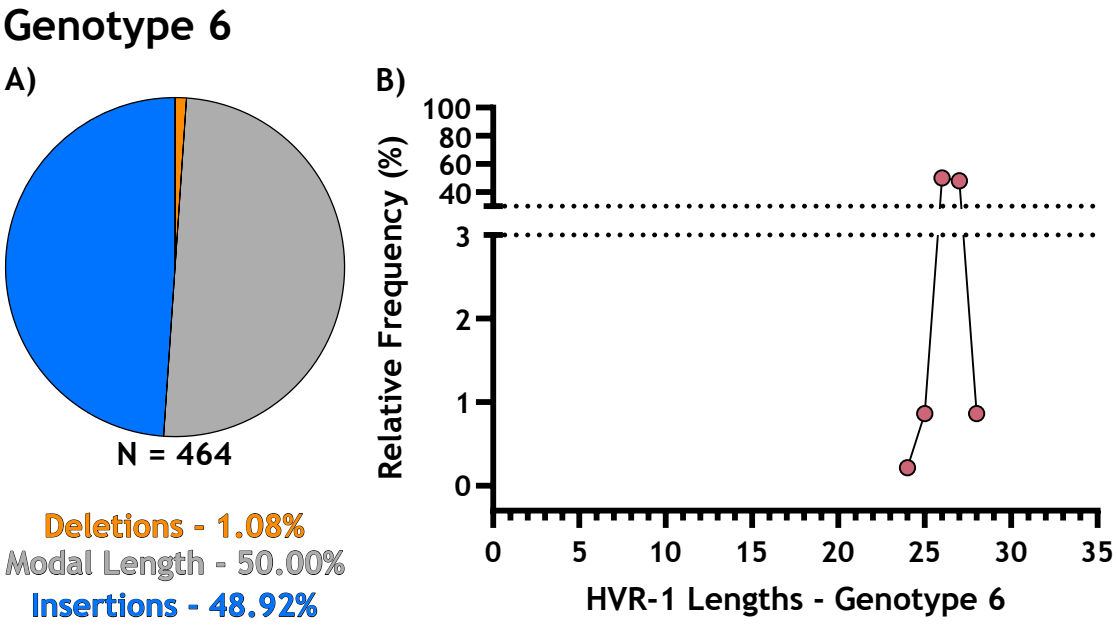
Figure III. IV. **HVR-1 Length Polymorphisms in Genotype 4.** A) Proportion of HVR-1 lengths which are modal length, contain insertions, or contain deletions. B) HVR-1 lengths graphed against percentage frequency. Note that the y-axis, showing percentage frequency, are split with the bottom axes running from 0 to 3%, the top axes from 30 to 100%. The x-axis is the HVR-1 lengths. C) Table of the figures used to generate B), showing the number of each length HVR-1.



C)

HVR-1 Lengths	Frequency of HVR-1 Lengths	Relative Frequency (%)
26	1	2.70
27	36	97.30

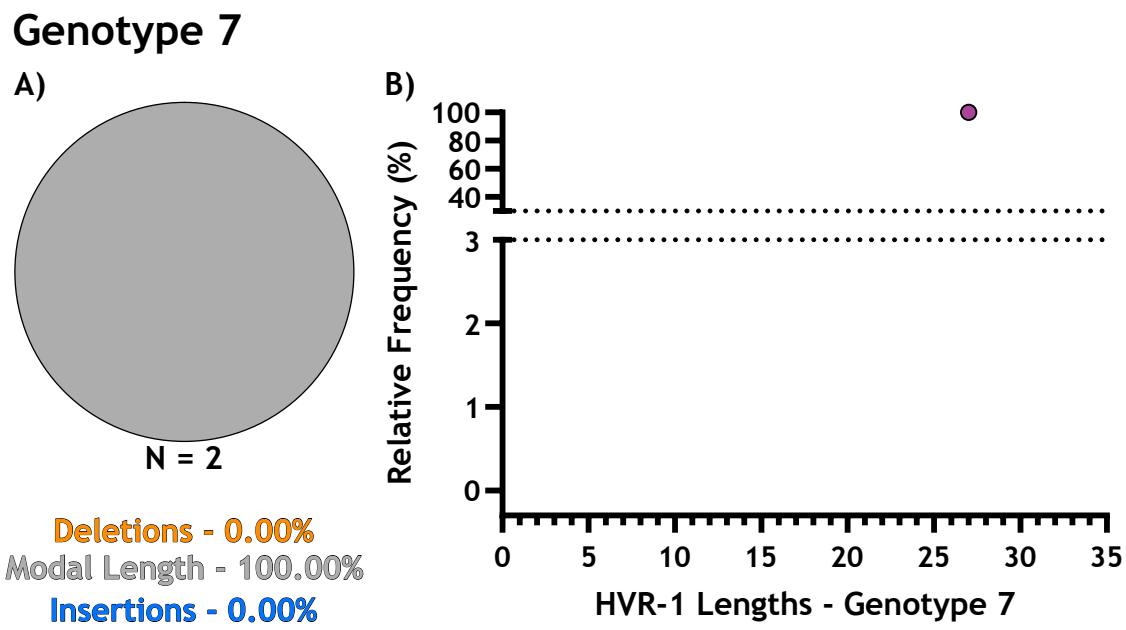
Figure III. V. **HVR-1 Length Polymorphisms in Genotype 5.** A) Proportion of HVR-1 lengths which are modal length, contain insertions, or contain deletions. B) HVR-1 lengths graphed against percentage frequency. Note that the y-axis, showing percentage frequency, are split with the bottom axes running from 0 to 3%, the top axes from 30 to 100%. The x-axis is the HVR-1 lengths. C) Table of the figures used to generate B), showing the number of each length HVR-1.



C)

HVR-1 Lengths	Frequency of HVR-1 Lengths	Relative Frequency (%)
24	1	0.22
25	4	0.86
26	232	50.00
27	223	48.06
28	4	0.86

Figure III. VI. **HVR-1 Length Polymorphisms in Genotype 6.** A) Proportion of HVR-1 lengths which are modal length, contain insertions, or contain deletions. B) HVR-1 lengths graphed against percentage frequency. Note that the y-axis, showing percentage frequency, are split with the bottom axes running from 0 to 3%, the top axes from 30 to 100%. The x-axis is the HVR-1 lengths. C) Table of the figures used to generate B), showing the number of each length HVR-1.



C)

HVR-1 Lengths	Frequency of HVR-1 Lengths	Relative Frequency (%)
27	2	100.00

Figure III. VII. **HVR-1 Length Polymorphisms in Genotype 7.** A) Proportion of HVR-1 lengths which are modal length, contain insertions, or contain deletions. B) HVR-1 lengths graphed against percentage frequency. Note that the y-axis, showing percentage frequency, are split with the bottom axes running from 0 to 3%, the top axes from 30 to 100%. The x-axis is the HVR-1 lengths. C) Table of the figures used to generate B), showing the number of each length HVR-1.

Appendix IV

Phylogenetic Trees by Genotype

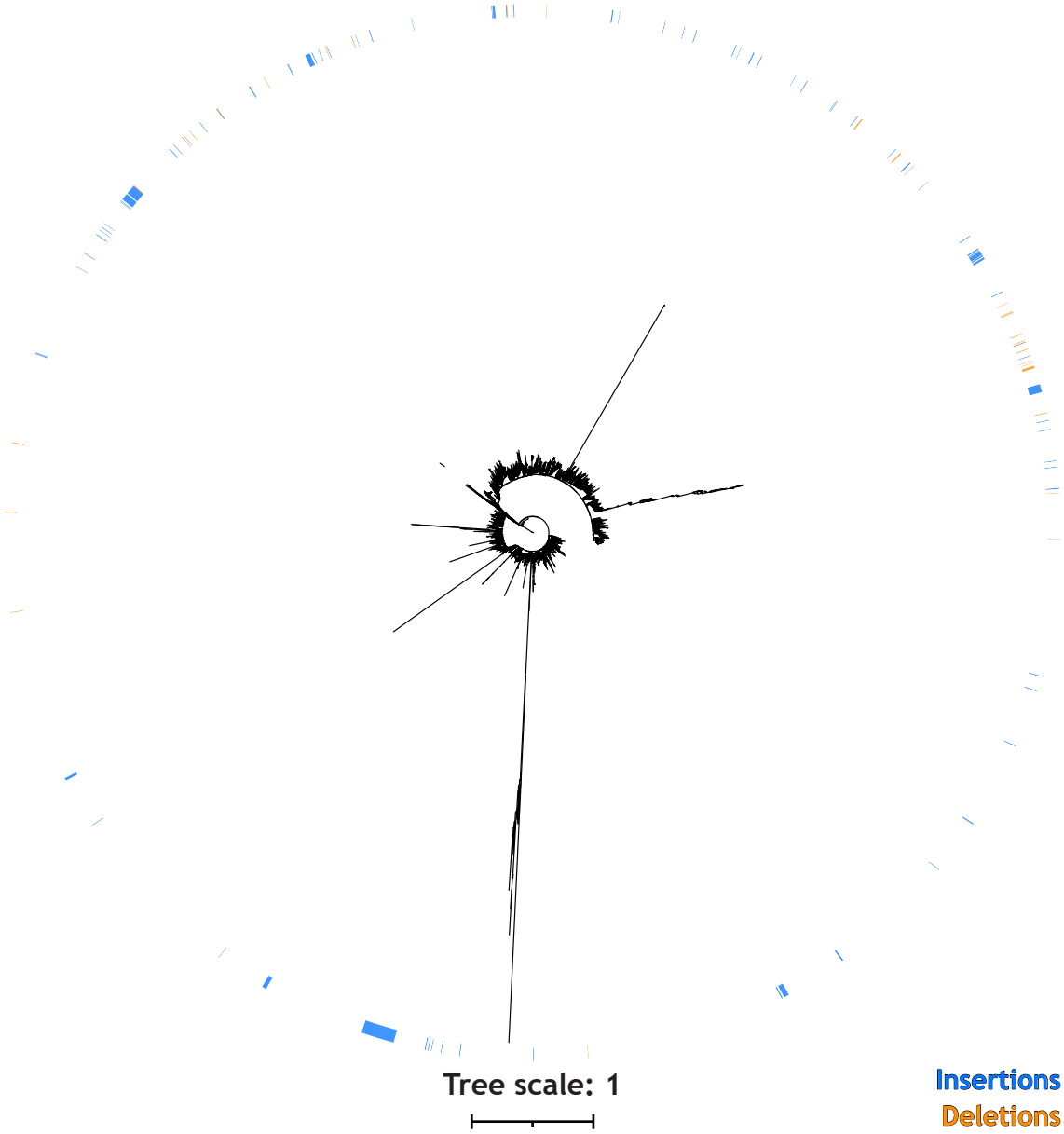


Figure IV. I. **Genotype 1 Phylogenetic Tree.** Phylogenetic tree for the 6,130 genotype 1 E2 sequences with the HVR-1 removed was produced by IQ-Tree2 using 1,000 ultrafast bootstraps. Annotated in iTOL, the coloured bars surrounding the tree show whether the corresponding HVR-1 sequences featured insertions or deletions compared to the model HVR-1 length (27 amino acids) for genotype 1.

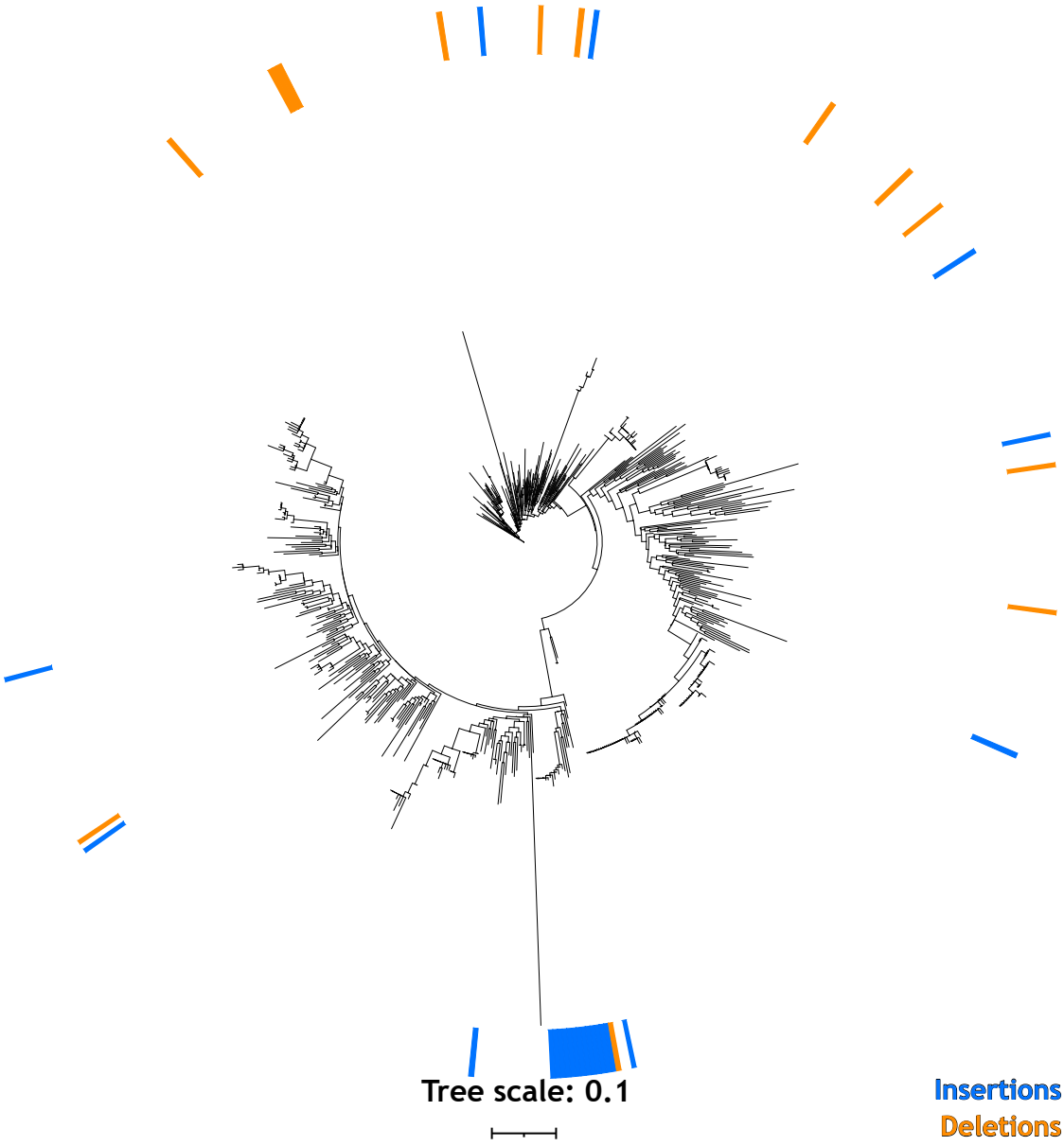


Figure IV. II. **Genotype 2 Phylogenetic Tree.** Phylogenetic tree for the 658 genotype 2 E2 sequences with the HVR-1 removed was produced by IQ-Tree2 using 1,000 ultrafast bootstraps. Annotated in iTOL, the coloured bars surrounding the tree show whether the corresponding HVR-1 sequence featured insertions or deletions compared to the modal HVR-1 length (27 amino acids) for genotype 2.

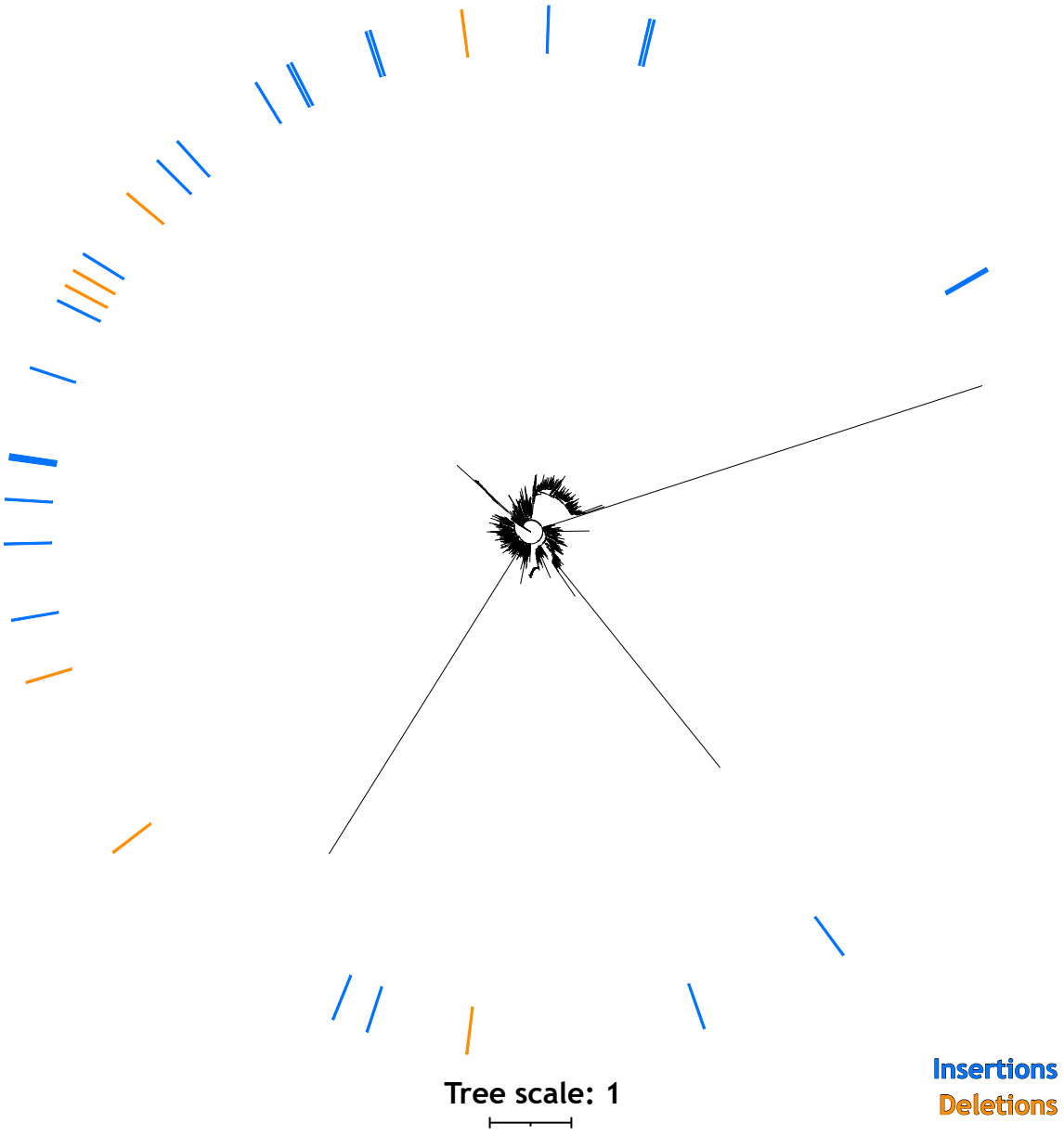


Figure IV. III. **Genotype 3 Phylogenetic Tree.** Phylogenetic tree for the 1,551 genotype 3 E2 sequences with the HVR-1 removed was produced by IQ-Tree2 using 1,000 ultrafast bootstraps. Annotated in iTOL, the coloured bars surrounding the tree show whether the corresponding HVR-1 sequence featured insertions or deletions compared to the modal HVR-1 length (27 amino acids) for genotype 3.

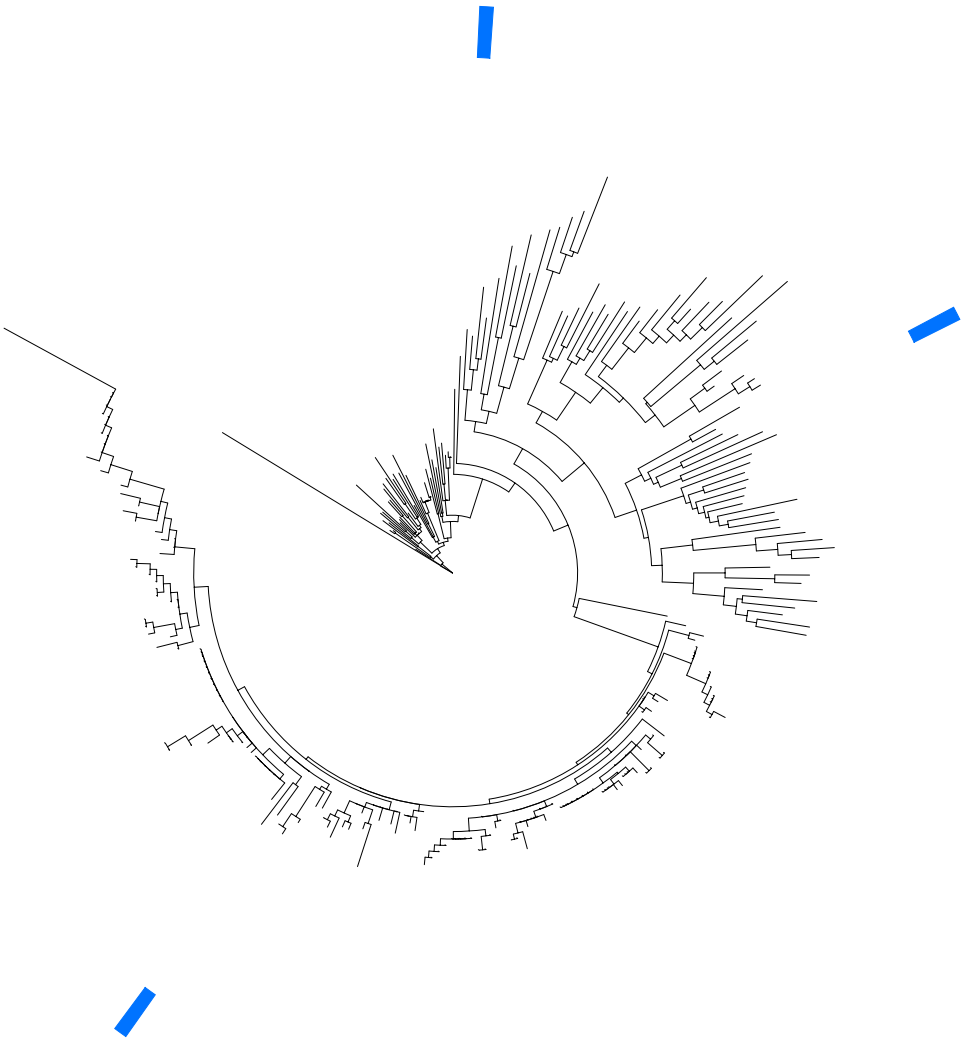


Figure IV. IV. **Genotype 4 Phylogenetic Tree.** Phylogenetic tree for the 261 genotype 4 E2 sequences with the HVR-1 removed was produced by IQ-Tree2 using 1,000 ultrafast bootstraps. Annotated in iTOL, the coloured bars surrounding the tree show whether the corresponding HVR-1 sequence featured insertions or deletions compared to the modal HVR-1 length (27 amino acids) for genotype 4.

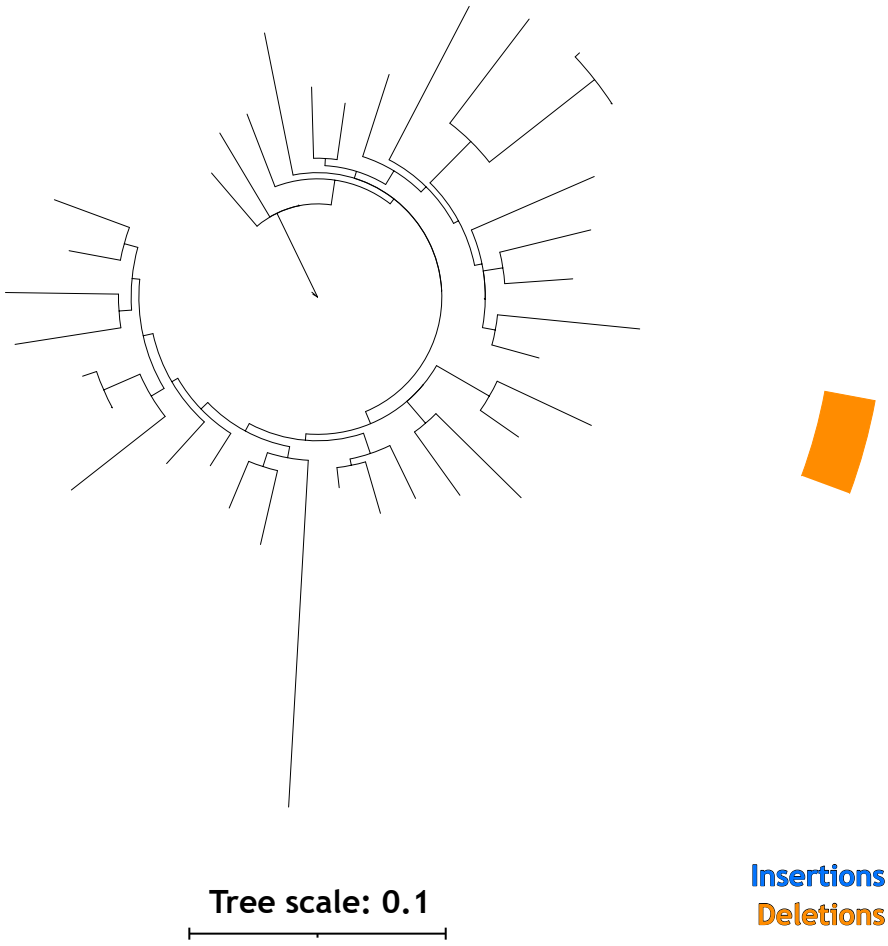


Figure IV. V. **Genotype 5 Phylogenetic Tree.** Phylogenetic tree for the 37 genotype 5 E2 sequences with the HVR-1 removed was produced by IQ-Tree2 using 1,000 ultrafast bootstraps. Annotated in iTOL, the coloured bars surrounding the tree show whether the corresponding HVR-1 sequence featured insertions or deletions compared to the modal HVR-1 length (27 amino acids) for genotype 5.

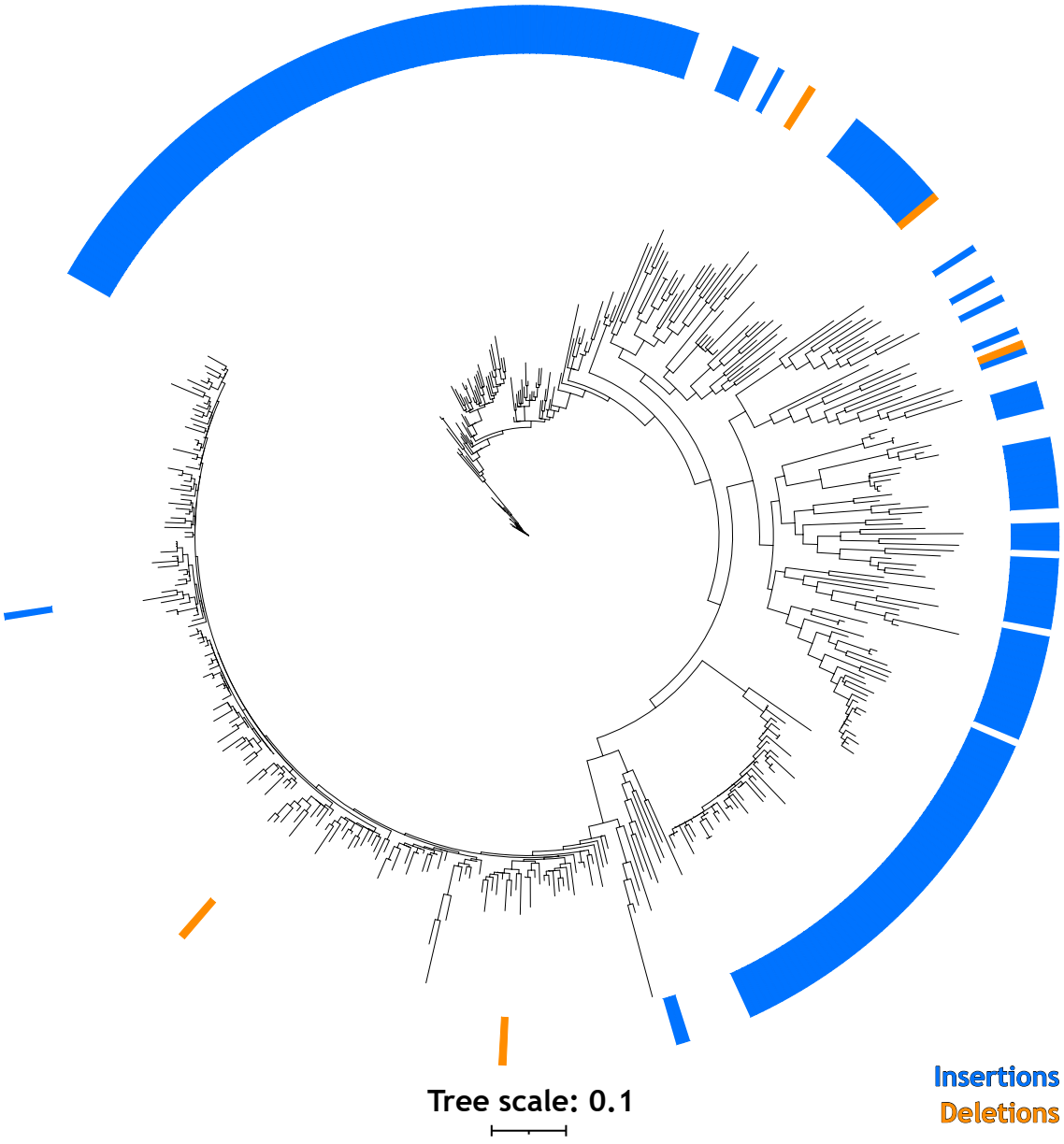


Figure IV. VI. **Genotype 6 Phylogenetic Tree.** Phylogenetic tree for the 464 genotype 6 E2 sequences with the HVR-1 removed was produced by IQ-Tree2 using 1,000 ultrafast bootstraps. Annotated in iTOL, the coloured bars surrounding the tree show whether the corresponding HVR-1 sequence featured insertions or deletions compared to the modal HVR-1 length (26 amino acids) for genotype 6.

Appendix V

Wildtype J6-backbone HCVpp Infectivity – Additional Statistical Analysis

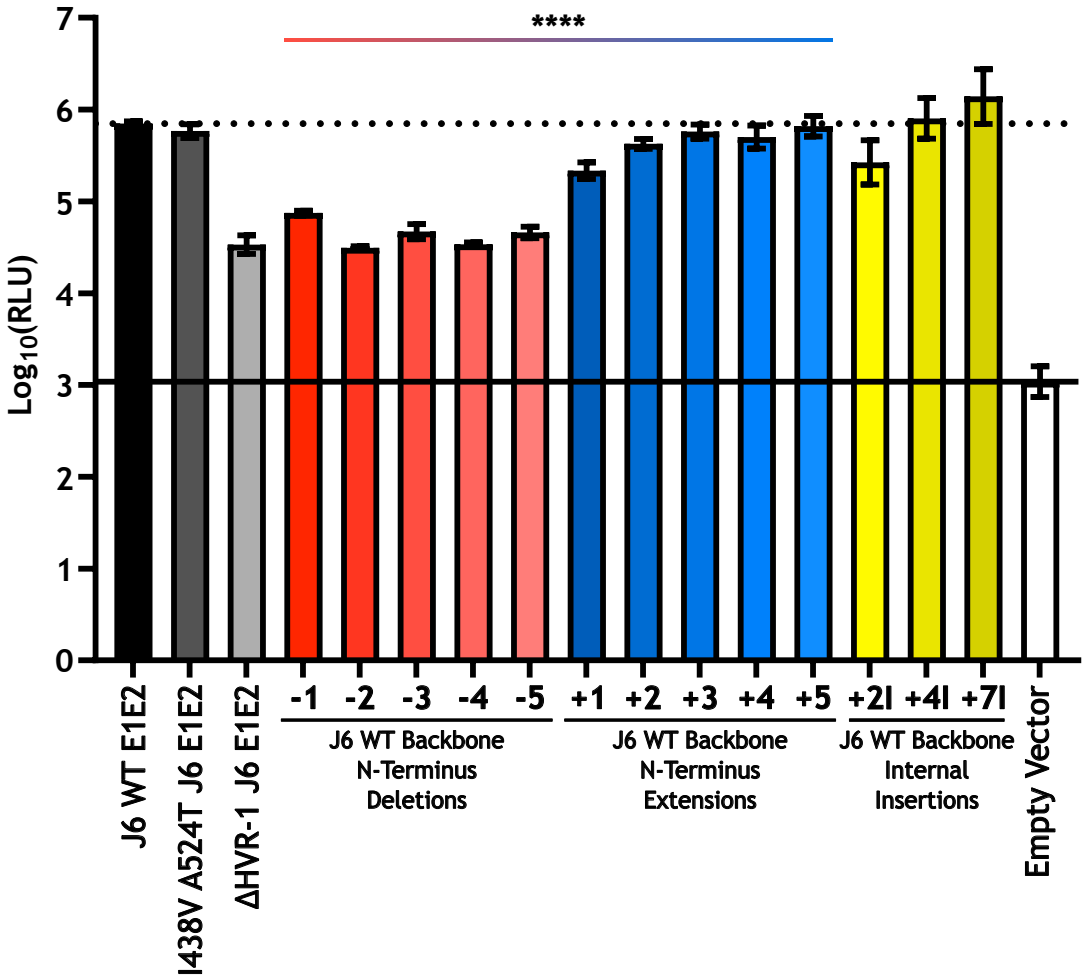


Figure V. I. **Statistical Analysis Comparing the Infectivity of the N-Terminus Deletions and the N-Terminus Extensions in the Wildtype J6-backbone.** An ordinary one-way ANOVA test was performed to compare the log_{10}(RLU) figures obtained by the wildtype J6-backbone N-Terminus Deletion and Extension HCVpp in the wildtype J6-backbone. In the I438V A524T J6-backbone, it was found that there was a reduction in infectivity in the N-Terminus Extension HCVpp as measured by the log_{10}(RLU) figures. At first glance, it appeared that the N-Terminus Extensions in the I438V A524T backbone were approximately as infective as the N-Terminus Deletions. An ordinary one-way ANOVA (as seen in Figure 4. 3.) revealed that the two groups of HCVpp produced statistically different log_{10}(RLU) values with a P-value of ≤ 0.05. For comparison, the same analysis was performed between the two groups of HCVpp in the wildtype J6 backbone. Here the statistical difference between these two groups produces a P-value ≤ 0.0001.
P-values; **** = P ≤ 0.0001. Error bars represent the SEM across 3 biological replicates.

Appendix VI

Additional Western Blots

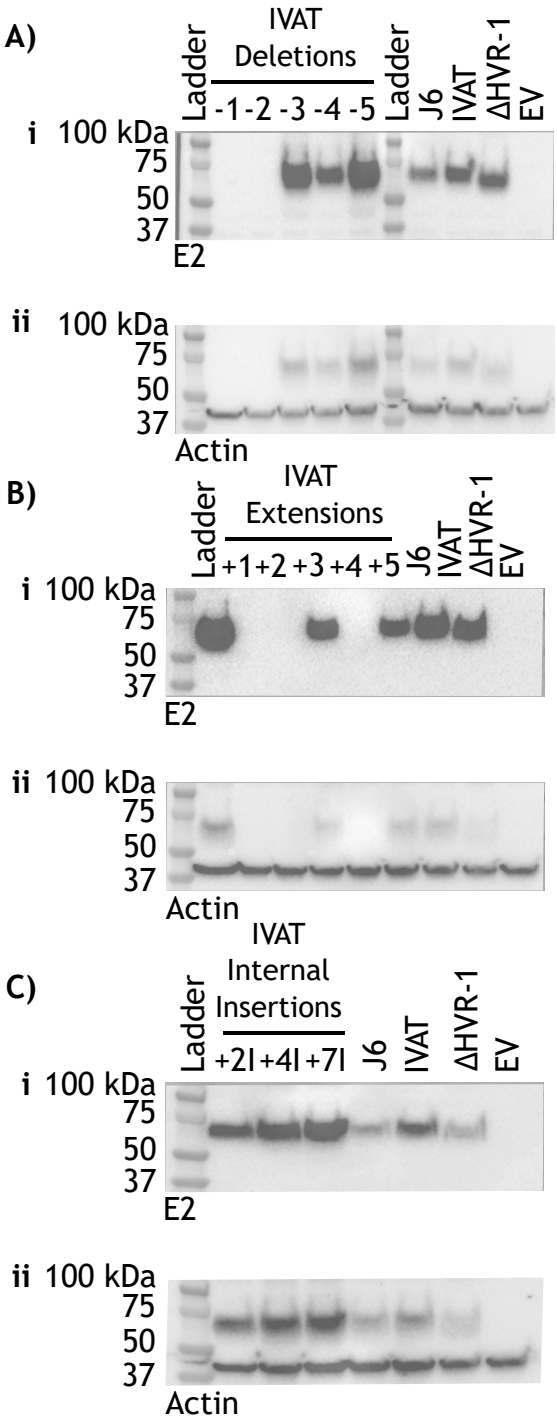
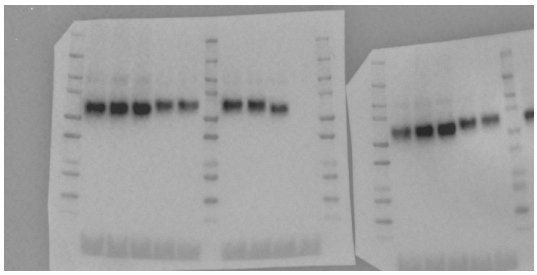
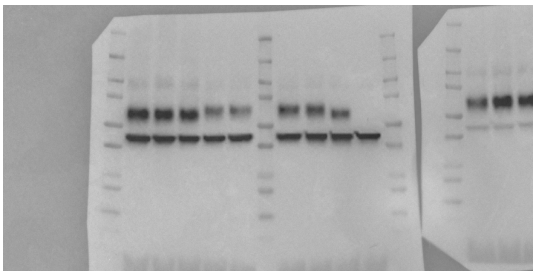


Figure VI. I. **I438V A524T J6 E2 Western Blots.** E2 molecular weight: ~60 kDa, actin molecular weight: 42 kDa. Lysates of the HEK 293T CD81 KO cells used to produce the I438V A524T-backbone HCVpp used in assays were Western blotted to quantify E2 production to ensure differences in phenotype were not caused by differences in expression. Actin was used as a loading control and blotted for on the same nitrocellulose membrane as the E2 blots.
A) I438V A524T J6 N-Terminus deletions. B) I438V A524T J6 N-Terminus extensions. C) I438V A524T J6 Internal Insertions.
J6 = J6 WT E1E2
IVAT = I438V A524T J6 E1E2
ΔHVR-1 = ΔHVR-1 J6 E1E2
EV = Empty Vector
Western Blots performed by Miss Sarah Little.

J6 Deletions

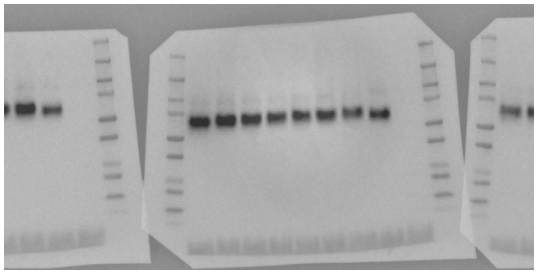


E2

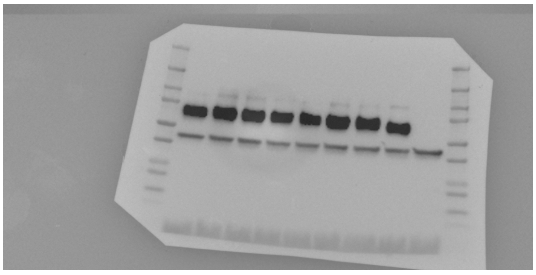


Actin

J6 Extensions

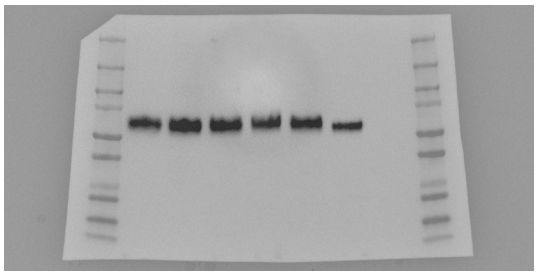


E2

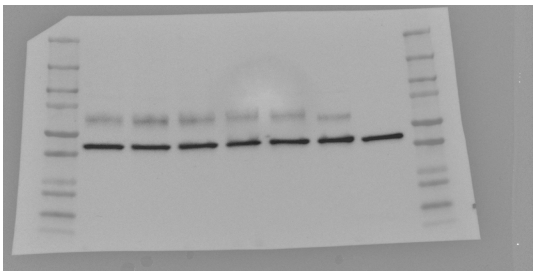


Actin

J6 Internal Insertions



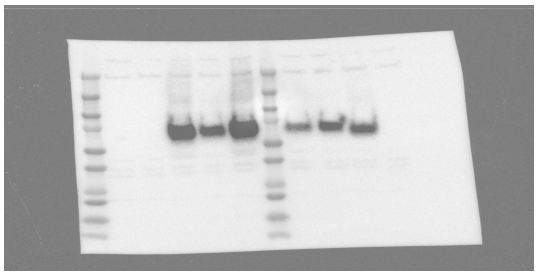
E2



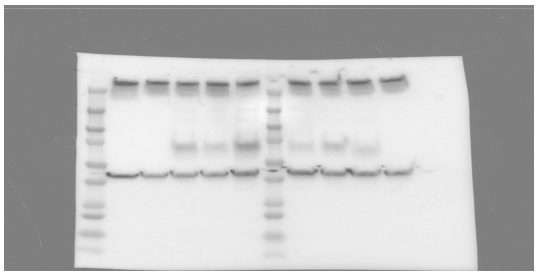
Actin

Figure VI. II. **Uncropped Western Blots for the Wildtype J6-backbone HCVpp.** Samples arranged as per the labelling in Chapter 4, Figure 4. 5. From top to bottom, N-Terminus Deletion HCVpp, N-Terminus Extension HCVpp, Internal Insertion HCVpp. From left to right, E2 blots, followed by actin blots.

I438V A524T J6 Deletions

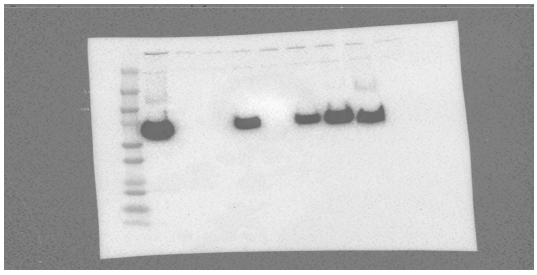


E2

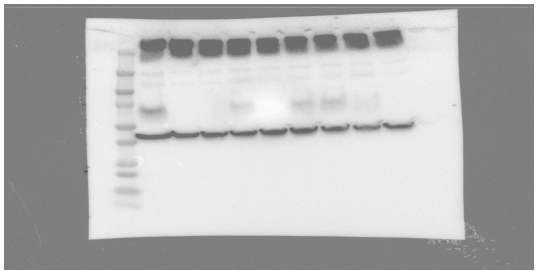


Actin

I438V A524T J6 Extensions

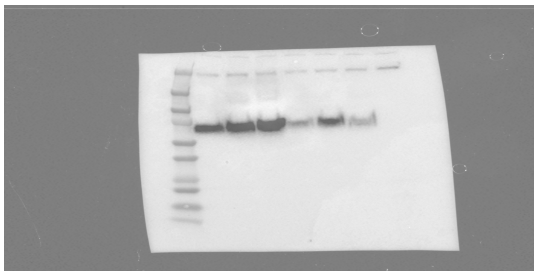


E2

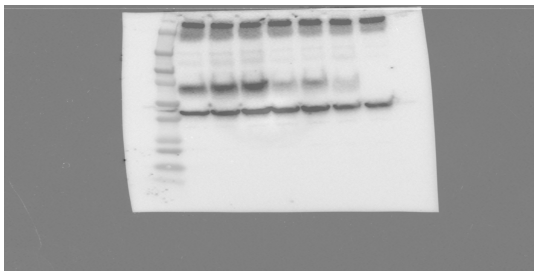


Actin

I438V A524T J6 Internal Insertions



E2



Actin

Figure VI. III. **Uncropped Western Blots for the I438V A524T J6-backbone HCVpp.** Samples arranged as per the labelling in Appendix VI, Figure VI. I. From top to bottom, N-Terminus Deletion HCVpp, N-Terminus Extension HCVpp, Internal Insertion HCVpp. From left to right, E2 blots, followed by actin blots.

References

Abbas, A.K., Lichtman, A.H. and Pillai, S. (2017) 'Cellular and Molecular Immunology', *Cellular and Molecular Immunology*. Ninth Edition. Elsevier.

Abe, K., Inchauspe, G. and Fujisawa, K. (1992) 'Genomic Characterization and Mutation Rate of Hepatitis C Virus Isolated From a Patient Who Contracted Hepatitis During an Epidemic of non-A, non-B Hepatitis in Japan', *Journal of General Virology*, 73(10), pp. 2725–2729. Available at: https://doi.org/10.1099/0022-1317-73-10-2725.

Abramson, J. *et al.* (2024) 'Accurate structure prediction of biomolecular interactions with AlphaFold 3', *Nature*, 630(8016), pp. 493–500. Available at: https://doi.org/10.1038/s41586-024-07487-w.

Acton, S. *et al.* (1996) 'Identification of Scavenger Receptor SR-BI as a High Density Lipoprotein Receptor', *Science*, 271(5248), pp. 518–520. Available at: https://doi.org/10.1126/science.271.5248.518.

Albecka, A. *et al.* (2012) 'Role of low-density lipoprotein receptor in the hepatitis C virus life cycle', *Hepatology*, 55(4), pp. 998–1007. Available at: https://doi.org/10.1002/hep.25501.

AlQuraishi, M. (2019) 'AlphaFold at CASP13', *Bioinformatics*, 35(22), pp. 4862–4865. Available at: https://doi.org/10.1093/bioinformatics/btz422.

Alter, H.J. *et al.* (1978) 'Transmissible agent in non-A, non-B hepatitis', *Lancet (London, England)*, 1(8062), pp. 459–463. Available at: https://doi.org/10.1016/s0140-6736(78)90131-9.

Alter, H.J. and Seeff, L.B. (2000) 'Recovery, Persistence, and Sequelae in Hepatitis C Virus Infection: A Perspective on Long-Term Outcome', *Seminars in Liver Disease*, 20(1), pp. 17–36. Available at: https://doi.org/10.1055/s-2000-9505.

Anasir, M.I. and Poh, C.L. (2019) 'Structural Vaccinology for Viral Vaccine Design', *Frontiers in Microbiology*, 10. Available at: https://doi.org/10.3389/fmicb.2019.00738.

André, P. *et al.* (2002) 'Characterization of Low- and Very-Low-Density Hepatitis C Virus RNA-Containing Particles', *Journal of Virology*, 76(14), pp. 6919–6928. Available at: https://doi.org/10.1128/jvi.76.14.6919-6928.2002.

Anfinsen, C.B. *et al.* (1961) 'THE KINETICS OF FORMATION OF NATIVE RIBONUCLEASE DURING OXIDATION OF THE REDUCED POLYPEPTIDE CHAIN', *Proceedings of the National Academy of Sciences of the United States of America*, 47(9), pp. 1309–1314. Available at: https://doi.org/10.1073/pnas.47.9.1309.

Anfinsen, C.B. (1973) ‘Principles that Govern the Folding of Protein Chains’, *Science*, 181(4096), pp. 223–230. Available at: https://doi.org/10.1126/science.181.4096.223.

Antoniou, T. *et al.* (2023) ‘Accessing hepatitis C direct acting antivirals among people living with hepatitis C: a qualitative study’, *International Journal for Equity in Health*, 22(1), p. 112. Available at: https://doi.org/10.1186/s12939-023-01924-4.

Antoniou, T. *et al.* (2025) ‘Impact of Policy Changes Expanding Access to Direct-Acting Antivirals on Hepatitis C Virus–Related Hospitalizations in People With HIV: A Population-Based Study’, *Open Forum Infectious Diseases*, 12(1), p. ofaf003. Available at: https://doi.org/10.1093/ofid/ofaf003.

Appel, N. *et al.* (2008) ‘Essential role of domain III of nonstructural protein 5A for hepatitis C virus infectious particle assembly’, *PLoS pathogens*, 4(3), p. e1000035. Available at: https://doi.org/10.1371/journal.ppat.1000035.

Ariumi, Y. *et al.* (2011) ‘The ESCRT System Is Required for Hepatitis C Virus Production’, *PLOS ONE*, 6(1), p. e14517. Available at: https://doi.org/10.1371/journal.pone.0014517.

Augestad, E.H. *et al.* (2020) ‘Global and local envelope protein dynamics of hepatitis C virus determine broad antibody sensitivity’, *Science Advances*, 6(35), p. eabb5938. Available at: https://doi.org/10.1126/sciadv.abb5938.

Augestad, E.H. *et al.* (2024) ‘The hepatitis C virus envelope protein complex is a dimer of heterodimers’, *Nature*, 633(8030), pp. 704–709. Available at: https://doi.org/10.1038/s41586-024-07783-5.

Backovic, M. and Jardetzky, T.S. (2009) ‘Class III viral membrane fusion proteins’, *Current opinion in structural biology*, 19(2), pp. 189–196. Available at: https://doi.org/10.1016/j.sbi.2009.02.012.

Bacon, B.R. *et al.* (2011) ‘Boceprevir for Previously Treated Chronic HCV Genotype 1 Infection’, *The New England journal of medicine*, 364(13), pp. 1207–1217. Available at: https://doi.org/10.1056/NEJMoa1009482.

Bailey, J.R. *et al.* (2015) ‘Naturally selected hepatitis C virus polymorphisms confer broad neutralizing antibody resistance’, *The Journal of Clinical Investigation*, 125(1), pp. 437–447. Available at: https://doi.org/10.1172/JCI78794.

Bailey, J.R. *et al.* (2017) ‘Broadly neutralizing antibodies with few somatic mutations and hepatitis C virus clearance’, *JCI Insight*, 2(9), p. e92872. Available at: https://doi.org/10.1172/jci.insight.92872.

Balasekaran, R. *et al.* (1999) ‘A Case-Control Study of Risk Factors for Sporadic Hepatitis C Virus Infection in The Southwestern United States’, *Official journal*

of the American College of Gastroenterology | ACG, 94(5), p. 1341. Available at: https://doi.org/10.1111/j.1572-0241.1999.01084.x.

Bankwitz, D. *et al.* (2010) 'Hepatitis C Virus Hypervariable Region 1 Modulates Receptor Interactions, Conceals the CD81 Binding Site, and Protects Conserved Neutralizing Epitopes', *Journal of Virology*, 84(11), pp. 5751–5763. Available at: https://doi.org/10.1128/JVI.02200-09.

Baquero, E. *et al.* (2017) 'Structural intermediates in the fusion-associated transition of vesiculovirus glycoprotein', *The EMBO Journal*, 36(5), pp. 679–692. Available at: https://doi.org/10.15252/embj.201694565.

Barba, G. *et al.* (1997) 'Hepatitis C virus core protein shows a cytoplasmic localization and associates to cellular lipid storage droplets', *Proceedings of the National Academy of Sciences*, 94(4), pp. 1200–1205. Available at: https://doi.org/10.1073/pnas.94.4.1200.

Barnes, E. *et al.* (2012) 'Novel Adenovirus-Based Vaccines Induce Broad and Sustained T Cell Responses to HCV in Man', *Science Translational Medicine*, 4(115), pp. 115ra1-115ra1. Available at: https://doi.org/10.1126/scitranslmed.3003155.

Bartenschlager, R. *et al.* (1995) 'Complex formation between the NS3 serine-type proteinase of the hepatitis C virus and NS4A and its importance for polyprotein maturation', *Journal of Virology*, 69(12), pp. 7519–7528. Available at: https://doi.org/10.1128/JVI.69.12.7519-7528.1995.

Bartenschlager, R., Frese, M. and Pietschmann, T. (2004) 'Novel Insights into Hepatitis C Virus Replication and Persistence', *Advances in Virus Research*. Academic Press, pp. 71–180. Available at: https://doi.org/10.1016/S0065-3527(04)63002-8.

Bartosch, B., Vitelli, A., *et al.* (2003) 'Cell Entry of Hepatitis C Virus Requires a Set of Co-receptors That Include the CD81 Tetraspanin and the SR-B1 Scavenger Receptor*', *Journal of Biological Chemistry*, 278(43), pp. 41624–41630. Available at: https://doi.org/10.1074/jbc.M305289200.

Bartosch, B., Bukh, J., *et al.* (2003) 'In vitro assay for neutralizing antibody to hepatitis C virus: Evidence for broadly conserved neutralization epitopes', *Proceedings of the National Academy of Sciences*, 100(24), pp. 14199–14204. Available at: https://doi.org/10.1073/pnas.2335981100.

Bartosch, B. *et al.* (2005) 'An Interplay between Hypervariable Region 1 of the Hepatitis C Virus E2 Glycoprotein, the Scavenger Receptor BI, and High-Density Lipoprotein Promotes both Enhancement of Infection and Protection against Neutralizing Antibodies', *Journal of Virology*, 79(13), pp. 8217–8229. Available at: https://doi.org/10.1128/jvi.79.13.8217-8229.2005.

Bartosch, B., Dubuisson, J. and Cosset, F.-L. (2003) 'Infectious Hepatitis C Virus Pseudo-particles Containing Functional E1–E2 Envelope Protein Complexes', *The Journal of Experimental Medicine*, 197(5), pp. 633–642. Available at: https://doi.org/10.1084/jem.20021756.

Bassett, S.E. *et al.* (2001) 'Protective immune response to hepatitis C virus in chimpanzees rechallenged following clearance of primary infection', *Hepatology (Baltimore, Md.)*, 33(6), pp. 1479–1487. Available at: https://doi.org/10.1053/jhep.2001.24371.

Baz Morelli, A. *et al.* (2012) 'ISCOMATRIX: a novel adjuvant for use in prophylactic and therapeutic vaccines against infectious diseases', *Journal of Medical Microbiology*, 61(7), pp. 935–943. Available at: https://doi.org/10.1099/jmm.0.040857-0.

Beh, L.Y., Colwell, L.J. and Francis, N.J. (2012) 'A core subunit of Polycomb repressive complex 1 is broadly conserved in function but not primary sequence', *Proceedings of the National Academy of Sciences of the United States of America*, 109(18), pp. E1063–E1071. Available at: https://doi.org/10.1073/pnas.1118678109.

Benova, L. *et al.* (2014) 'Vertical Transmission of Hepatitis C Virus: Systematic Review and Meta-analysis', *Clinical Infectious Diseases*, 59(6), pp. 765–773. Available at: https://doi.org/10.1093/cid/ciu447.

Benton, D.J. *et al.* (2020) 'Structural transitions in influenza haemagglutinin at membrane fusion pH', *Nature*, 583(7814), pp. 150–153. Available at: https://doi.org/10.1038/s41586-020-2333-6.

Berger, C. *et al.* (2014) 'Daclatasvir-like inhibitors of NS5A block early biogenesis of hepatitis C virus-induced membranous replication factories, independent of RNA replication', *Gastroenterology*, 147(5), pp. 1094-1105.e25. Available at: https://doi.org/10.1053/j.gastro.2014.07.019.

Berger, K.L. *et al.* (2011) 'Hepatitis C Virus Stimulates the Phosphatidylinositol 4-Kinase III Alpha-Dependent Phosphatidylinositol 4-Phosphate Production That Is Essential for Its Replication ▿ ', *Journal of Virology*, 85(17), pp. 8870–8883. Available at: https://doi.org/10.1128/JVI.00059-11.

Bertaux, C. and Dragic, T. (2006) 'Different Domains of CD81 Mediate Distinct Stages of Hepatitis C Virus Pseudoparticle Entry', *Journal of Virology*, 80(10), pp. 4940–4948. Available at: https://doi.org/10.1128/jvi.80.10.4940-4948.2006.

Bianchettin, G. *et al.* (2007) 'Analysis of Hepatitis C Virus Hypervariable Region 1 Sequence from Cryoglobulinemic Patients and Associated Controls', *Journal of Virology*, 81(9), pp. 4564–4571. Available at: https://doi.org/10.1128/JVI.02104-06.

BNF (2025) *Hepatitis | Treatment summaries | BNF content published by NICE*. Available at: https://bnf.nice.org.uk/treatment-summaries/hepatitis/ (Accessed: 21 August 2025).

Boisvert, M. *et al.* (2016) 'Novel E2 Glycoprotein Tetramer Detects Hepatitis C Virus-Specific Memory B Cells', *Journal of Immunology (Baltimore, Md.: 1950)*, 197(12), pp. 4848–4858. Available at: https://doi.org/10.4049/jimmunol.1600763.

Boson, B. *et al.* (2011) 'A Concerted Action of Hepatitis C Virus P7 and Nonstructural Protein 2 Regulates Core Localization at the Endoplasmic Reticulum and Virus Assembly', *PLOS Pathogens*, 7(7), p. e1002144. Available at: https://doi.org/10.1371/journal.ppat.1002144.

Boulant, S. *et al.* (2005) 'Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features', *Journal of Virology*, 79(17), pp. 11353–11365. Available at: https://doi.org/10.1128/JVI.79.17.11353-11365.2005.

Boulant, S., Targett-Adams, P. and McLauchlan, J. (2007) 'Disrupting the association of hepatitis C virus core protein with lipid droplets correlates with a loss in production of infectious virus', *The Journal of General Virology*, 88(Pt 8), pp. 2204–2213. Available at: https://doi.org/10.1099/vir.0.82898-0.

Brazzoli, M. *et al.* (2008) 'CD81 Is a Central Regulator of Cellular Events Required for Hepatitis C Virus Infection of Human Hepatocytes', *Journal of Virology*, 82(17), pp. 8316–8329. Available at: https://doi.org/10.1128/jvi.00665-08.

Bulteel, N. *et al.* (2016) 'Factors associated with spontaneous clearance of chronic hepatitis C virus infection', *Journal of Hepatology*, 65(2), pp. 266–272. Available at: https://doi.org/10.1016/j.jhep.2016.04.030.

Cai, Q. *et al.* (2013) 'Comparison of three different HCV genotyping methods: Core, NS5B sequence analysis and line probe assay', *International Journal of Molecular Medicine*, 31(2), pp. 347–352. Available at: https://doi.org/10.3892/ijmm.2012.1209.

Callens, N. *et al.* (2005) 'Basic residues in hypervariable region 1 of hepatitis C virus envelope glycoprotein e2 contribute to virus entry', *Journal of Virology*, 79(24), pp. 15331–15341. Available at: https://doi.org/10.1128/JVI.79.24.15331-15341.2005.

Caminada, S. *et al.* (2023) 'Risk of parenterally transmitted hepatitis following exposure to invasive procedures in Italy: SEIEVA surveillance 2000-2021', *Journal of Hepatology*, 79(1), pp. 61–68. Available at: https://doi.org/10.1016/j.jhep.2023.03.002.

Camus, G. *et al.* (2013) 'Diacylglycerol Acyltransferase-1 Localizes Hepatitis C Virus NS5A Protein to Lipid Droplets and Enhances NS5A Interaction with the Viral Capsid Core *', *Journal of Biological Chemistry*, 288(14), pp. 9915–9923. Available at: https://doi.org/10.1074/jbc.M112.434910.

Cao, L. *et al.* (2019) 'Functional expression and characterization of the envelope glycoprotein E1E2 heterodimer of hepatitis C virus', *PLoS Pathogens*, 15(5), p. e1007759. Available at: https://doi.org/10.1371/journal.ppat.1007759.

Carriquí-Madroñal, B. *et al.* (2023) 'Genetic and pharmacological perturbation of hepatitis-C virus entry', *Current Opinion in Virology*, 62, p. 101362. Available at: https://doi.org/10.1016/j.coviro.2023.101362.

Carter, W., Connelly, S. and Struble, K. (2017) 'Reinventing HCV Treatment: Past and Future Perspectives', *The Journal of Clinical Pharmacology*, 57(3), pp. 287–296. Available at: https://doi.org/10.1002/jcph.830.

Casals, F. *et al.* (2012) 'Next-generation sequencing approaches for genetic mapping of complex diseases', *Journal of Neuroimmunology*, 248(1), pp. 10–22. Available at: https://doi.org/10.1016/j.jneuroim.2011.12.017.

Castelli, M. *et al.* (2017) 'A Biologically-validated HCV E1E2 Heterodimer Structural Model', *Scientific Reports*, 7(1), p. 214. Available at: https://doi.org/10.1038/s41598-017-00320-7.

Catanese, M.T. *et al.* (2013) 'Ultrastructural analysis of hepatitis C virus particles', *Proceedings of the National Academy of Sciences*, 110(23), pp. 9505–9510. Available at: https://doi.org/10.1073/pnas.1307527110.

Chai Discovery *et al.* (2024) 'Chai-1: Decoding the molecular interactions of life'. Cold Spring Harbor Laboratory. Available at: https://doi.org/10.1101/2024.10.10.615955.

Chang, K.-S. *et al.* (2007) 'Human Apolipoprotein E Is Required for Infectivity and Production of Hepatitis C Virus in Cell Culture', *Journal of Virology*, 81(24), pp. 13783–13793. Available at: https://doi.org/10.1128/JVI.01091-07.

Chao, L.H. *et al.* (2014) 'Sequential conformational rearrangements in flavivirus membrane fusion', *eLife*. Edited by R. Jahn, 3, p. e04389. Available at: https://doi.org/10.7554/eLife.04389.

Cheng, G. *et al.* (2012) '1172 ANTIVIRAL ACTIVITY AND RESISTANCE PROFILE OF THE NOVEL HCV NS5A INHIBITOR GS-5885', *Journal of Hepatology*, 56, p. S464. Available at: https://doi.org/10.1016/S0168-8278(12)61184-2.

Cho, M.-H. *et al.* (2020) 'Hidden dynamic signatures drive substrate selectivity in the disordered phosphoproteome', *Proceedings of the National Academy of Sciences of the United States of America*, 117(38), pp. 23606–23616. Available at: https://doi.org/10.1073/pnas.1921473117.

Choo, Q.-L. *et al.* (1989) 'Isolation of a cDNA cLone Derived from a Blood-Borne Non-A, Non-B Viral Hepatitis Genome', *Science*, 244(4902), pp. 359–362. Available at: https://doi.org/10.1126/science.2523562.

Chumbe, A. *et al.* (2024) 'A panel of hepatitis C virus glycoproteins for the characterization of antibody responses using antibodies with diverse recognition and neutralization patterns', *Virus Research*, 341, p. 199308. Available at: https://doi.org/10.1016/j.virusres.2024.199308.

Ciczora, Y. *et al.* (2007) 'Transmembrane domains of hepatitis C virus envelope glycoproteins: residues involved in E1E2 heterodimerization and involvement of these domains in virus entry', *Journal of Virology*, 81(5), pp. 2372–2381. Available at: https://doi.org/10.1128/JVI.02198-06.

Colbert, M.D. *et al.* (2019) 'Broadly Neutralizing Antibodies Targeting New Sites of Vulnerability in Hepatitis C Virus E1E2', *Journal of Virology*, 93(14), pp. e02070-18. Available at: https://doi.org/10.1128/JVI.02070-18.

Coller, K.E. *et al.* (2012) 'Molecular Determinants and Dynamics of Hepatitis C Virus Secretion', *PLoS Pathogens*. Edited by J.J. Ou, 8(1), p. e1002466. Available at: https://doi.org/10.1371/journal.ppat.1002466.

Corless, L. *et al.* (2010) 'Vps4 and the ESCRT-III complex are required for the release of infectious hepatitis C virus particles', *Journal of General Virology*, 91(2), pp. 362–372. Available at: https://doi.org/10.1099/vir.0.017285-0.

Cowton, V.M. *et al.* (2018) 'Predicting the Effectiveness of Hepatitis C Virus Neutralizing Antibodies by Bioinformatic Analysis of Conserved Epitope Residues Using Public Sequence Data', *Frontiers in Immunology*, 9. Available at: https://doi.org/10.3389/fimmu.2018.01470.

Cox, A.L. *et al.* (2019) 'LB10. A Randomized, Double-Blind, Placebo-Controlled Efficacy Trial of a Vaccine to Prevent Chronic Hepatitis C Virus Infection in an at-Risk Population', *Open Forum Infectious Diseases*, 6(Supplement_2), p. S997. Available at: https://doi.org/10.1093/ofid/ofz415.2493.

Cuevas, J.M. *et al.* (2009) 'The role of positive selection in hepatitis C virus', *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 9(5), pp. 860–866. Available at: https://doi.org/10.1016/j.meegid.2009.05.007.

Cummings, M.D. *et al.* (2010) 'Induced-Fit Binding of the Macrocyclic Noncovalent Inhibitor TMC435 to its HCV NS3/NS4A Protease Target', *Angewandte Chemie International Edition*, 49(9), pp. 1652–1655. Available at: https://doi.org/10.1002/anie.200906696.

Cyster, J.G. and Allen, C.D.C. (2019) 'B Cell Responses: Cell Interaction Dynamics and Decisions', *Cell*, 177(3), pp. 524–540. Available at: https://doi.org/10.1016/j.cell.2019.03.016.

Da Costa, D. *et al.* (2012) 'Reconstitution of the entire hepatitis C virus life cycle in nonhepatic cells', *Journal of Virology*, 86(21), pp. 11919–11925. Available at: https://doi.org/10.1128/JVI.01066-12.

Dahari, H., Feinstone, S.M. and Major, M.E. (2010) 'Meta-Analysis of Hepatitis C Virus Vaccine Efficacy in Chimpanzees Indicates an Importance for Structural Proteins', *Gastroenterology*, 139(3), pp. 965–974. Available at: https://doi.org/10.1053/j.gastro.2010.05.077.

Darnell, A.M., Subramaniam, A.R. and O'Shea, E.K. (2018) 'Translational control through differential ribosome pausing during amino acid limitation in mammalian cells', *Molecular cell*, 71(2), pp. 229-243.e11. Available at: https://doi.org/10.1016/j.molcel.2018.06.041.

Das, D.K. *et al.* (2018) 'Direct visualization of the conformational dynamics of single influenza hemagglutinin trimers', *Cell*, 174(4), pp. 926-937.e12. Available at: https://doi.org/10.1016/j.cell.2018.05.050.

Deffieu, M.S. *et al.* (2022) 'Occludin stalls HCV particle dynamics apart from hepatocyte tight junctions, promoting virion internalization', *Hepatology (Baltimore, Md.)*, 76(4), pp. 1164–1179. Available at: https://doi.org/10.1002/hep.32514.

von Delft, A. *et al.* (2018) 'The generation of a simian adenoviral vectored HCV vaccine encoding genetically conserved gene segments to target multiple HCV genotypes', *Vaccine*, 36(2), pp. 313–321. Available at: https://doi.org/10.1016/j.vaccine.2017.10.079.

Deng, L. *et al.* (2021) 'A conserved epitope III on hepatitis C virus E2 protein has alternate conformations facilitating cell binding or virus neutralization', *Proceedings of the National Academy of Sciences of the United States of America*, 118(28), p. e2104242118. Available at: https://doi.org/10.1073/pnas.2104242118.

Dill, K.A. *et al.* (2008) 'The protein folding problem', *Annual Review of Biophysics*, 37, pp. 289–316. Available at: https://doi.org/10.1146/annurev.biophys.37.092707.153558.

Donaldson, E.F. *et al.* (2015) 'Clinical evidence and bioinformatics characterization of potential hepatitis C virus resistance pathways for sofosbuvir', *Hepatology*, 61(1), p. 56. Available at: https://doi.org/10.1002/hep.27375.

Douam, F. *et al.* (2014) 'Critical interaction between E1 and E2 glycoproteins determines binding and fusion properties of hepatitis C virus during cell entry', *Hepatology*, 59(3), p. 776. Available at: https://doi.org/10.1002/hep.26733.

Dowd, K.A. *et al.* (2009) 'Selection pressure from neutralizing antibodies drives sequence evolution during acute infection with hepatitis C virus',

Gastroenterology, 136(7), pp. 2377–2386. Available at: https://doi.org/10.1053/j.gastro.2009.02.080.

Drane, D. *et al.* (2009) 'Priming of CD4+ and CD8+ T cell responses using a HCV core ISCOMATRIX™ vaccine: A phase I study in healthy volunteers', *Human Vaccines*, 5(3), pp. 151–157. Available at: https://doi.org/10.4161/hv.5.3.6614.

Drummer, H.E., Maerz, A. and Poumbourios, P. (2003) 'Cell surface expression of functional hepatitis C virus E1 and E2 glycoproteins', *FEBS Letters*, 546(2–3), pp. 385–390. Available at: https://doi.org/10.1016/S0014-5793(03)00635-5.

Dubuisson, J. *et al.* (1994) 'Formation and intracellular localization of hepatitis C virus envelope glycoprotein complexes expressed by recombinant vaccinia and Sindbis viruses', *Journal of Virology*, 68(10), pp. 6147–6160. Available at: https://doi.org/10.1128/JVI.68.10.6147-6160.1994.

Dutartre, H. *et al.* (2006) 'General catalytic deficiency of hepatitis C virus RNA polymerase with an S282T mutation and mutually exclusive resistance towards 2'-modified nucleotide analogues', *Antimicrobial Agents and Chemotherapy*, 50(12), pp. 4161–4169. Available at: https://doi.org/10.1128/AAC.00433-06.

Egger, D. *et al.* (2002) 'Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex', *Journal of Virology*, 76(12), pp. 5974–5984. Available at: https://doi.org/10.1128/jvi.76.12.5974-5984.2002.

El Omari, K. *et al.* (2014) 'Unexpected structure for the N-terminal domain of hepatitis C virus envelope glycoprotein E1', *Nature Communications*, 5, p. 4874. Available at: https://doi.org/10.1038/ncomms5874.

El-Diwany, R. *et al.* (2017) 'Extra-epitopic hepatitis C virus polymorphisms confer resistance to broadly neutralizing antibodies by modulating binding to scavenger receptor B1', *PLoS Pathogens*, 13(2), p. e1006235. Available at: https://doi.org/10.1371/journal.ppat.1006235.

Elofsson, A. (2025) 'AlphaFold3 at CASP16'. bioRxiv, p. 2025.04.10.648174. Available at: https://doi.org/10.1101/2025.04.10.648174.

European Medicines Agency (2018a) *Maviret | European Medicines Agency (EMA)*. Available at: https://www.ema.europa.eu/en/medicines/human/EPAR/maviret (Accessed: 23 August 2025).

European Medicines Agency (2018b) *Medicine Data | European Medicines Agency (EMA)*. Available at: https://www.ema.europa.eu/en/medicines/download-medicine-data (Accessed: 22 August 2025).

Evans, M.J. *et al.* (2007) 'Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry', *Nature*, 446(7137), pp. 801–805. Available at: https://doi.org/10.1038/nature05654.

Evans, R. *et al.* (2022) 'Protein complex prediction with AlphaFold-Multimer'. bioRxiv, p. 2021.10.04.463034. Available at: https://doi.org/10.1101/2021.10.04.463034.

Failla, C., Tomei, L. and De Francesco, R. (1994) 'Both NS3 and NS4A are required for proteolytic processing of hepatitis C virus nonstructural proteins', *Journal of Virology*, 68(6), pp. 3753–3760. Available at: https://doi.org/10.1128/JVI.68.6.3753-3760.1994.

Falkowska, E. *et al.* (2007) 'Hepatitis C Virus Envelope Glycoprotein E2 Glycans Modulate Entry, CD81 Binding, and Neutralization', *Journal of Virology*, 81(15), pp. 8072–8079. Available at: https://doi.org/10.1128/jvi.00459-07.

Falson, P. *et al.* (2015) 'Hepatitis C Virus Envelope Glycoprotein E1 Forms Trimers at the Surface of the Virion', *Journal of Virology*, 89(20), pp. 10333–10346. Available at: https://doi.org/10.1128/JVI.00991-15.

Farci, P. *et al.* (1992) 'Lack of protective immunity against reinfection with hepatitis C virus', *Science (New York, N.Y.)*, 258(5079), pp. 135–140. Available at: https://doi.org/10.1126/science.1279801.

Farci, P. *et al.* (2000) 'The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies', *Science (New York, N.Y.)*, 288(5464), pp. 339–344. Available at: https://doi.org/10.1126/science.288.5464.339.

Farci, P. *et al.* (2006) 'Evolution of hepatitis C viral quasispecies and hepatic injury in perinatally infected children followed prospectively', *Proceedings of the National Academy of Sciences of the United States of America*, 103(22), pp. 8475–8480. Available at: https://doi.org/10.1073/pnas.0602546103.

Farquhar, M.J. *et al.* (2008) 'Protein Kinase A-Dependent Step(s) in Hepatitis C Virus Entry and Infectivity', *Journal of Virology*, 82(17), pp. 8797–8811. Available at: https://doi.org/10.1128/jvi.00592-08.

Farquhar, M.J. *et al.* (2012) 'Hepatitis C Virus Induces CD81 and Claudin-1 Endocytosis', *Journal of Virology*, 86(8), pp. 4305–4316. Available at: https://doi.org/10.1128/jvi.06996-11.

Ferraris, P., Blanchard, E. and Roingeard, P. (2010) 'Ultrastructural and biochemical analyses of hepatitis C virus-associated host cell membranes', *The Journal of General Virology*, 91(Pt 9), pp. 2230–2237. Available at: https://doi.org/10.1099/vir.0.022186-0.

Flint, M. *et al.* (1999) 'Functional analysis of cell surface-expressed hepatitis C virus E2 glycoprotein', *Journal of Virology*, 73(8), pp. 6782–6790. Available at: https://doi.org/10.1128/JVI.73.8.6782-6790.1999.

Folgori, A. *et al.* (2006) 'A T-cell HCV vaccine eliciting effective immunity against heterologous virus challenge in chimpanzees', *Nature Medicine*, 12(2), pp. 190–197. Available at: https://doi.org/10.1038/nm1353.

Forns, X., Allander, T., *et al.* (2000) 'Characterization of Modified Hepatitis C Virus E2 Proteins Expressed on the Cell Surface', *Virology*, 274(1), pp. 75–85. Available at: https://doi.org/10.1006/viro.2000.0419.

Forns, X., Thimme, R., *et al.* (2000) 'Hepatitis C virus lacking the hypervariable region 1 of the second envelope protein is infectious and causes acute resolving or persistent infection in chimpanzees', *Proceedings of the National Academy of Sciences*, 97(24), pp. 13318–13323. Available at: https://doi.org/10.1073/pnas.230453597.

Fowler, N.J. and Williamson, M.P. (2022) 'The accuracy of protein structures in solution determined by AlphaFold and NMR', *Structure (London, England: 1993)*, 30(7), pp. 925-933.e2. Available at: https://doi.org/10.1016/j.str.2022.04.005.

Frey, S.E. *et al.* (2010) 'Safety and immunogenicity of HCV E1E2 vaccine adjuvanted with MF59 administered to healthy adults', *Vaccine*, 28(38), pp. 6367–6373. Available at: https://doi.org/10.1016/j.vaccine.2010.06.084.

Fricke, M. *et al.* (2015) 'Conserved RNA secondary structures and long-range interactions in hepatitis C viruses', *RNA (New York, N.Y.)*, 21(7), pp. 1219–1232. Available at: https://doi.org/10.1261/rna.049338.114.

Fridell, R.A. *et al.* (2011) 'Distinct Functions of NS5A in Hepatitis C Virus RNA Replication Uncovered by Studies with the NS5A Inhibitor BMS-790052 ▿', *Journal of Virology*, 85(14), pp. 7312–7320. Available at: https://doi.org/10.1128/JVI.00253-11.

Fu, R. *et al.* (2022) 'Systematic decomposition of sequence determinants governing CRISPR/Cas9 specificity', *Nature Communications*, 13(1), p. 474. Available at: https://doi.org/10.1038/s41467-022-28028-x.

Fung, A. *et al.* (2014) 'Efficiency of Incorporation and Chain Termination Determines the Inhibition Potency of 2′-Modified Nucleotide Analogs against Hepatitis C Virus Polymerase', *Antimicrobial Agents and Chemotherapy*, 58(7), pp. 3636–3645. Available at: https://doi.org/10.1128/AAC.02666-14.

Furuse, M. *et al.* (1993) 'Occludin: a novel integral membrane protein localizing at tight junctions', *The Journal of Cell Biology*, 123(6 Pt 2), pp. 1777–1788. Available at: https://doi.org/10.1083/jcb.123.6.1777.

Gastaminza, P. *et al.* (2008) 'Cellular Determinants of Hepatitis C Virus Assembly, Maturation, Degradation, and Secretion', *Journal of Virology*, 82(5), pp. 2120–2129. Available at: https://doi.org/10.1128/JVI.02053-07.

Gastaminza, P., Kapadia, S.B. and Chisari, F.V. (2006) 'Differential biophysical properties of infectious intracellular and secreted hepatitis C virus particles', *Journal of Virology*, 80(22), pp. 11074–11081. Available at: https://doi.org/10.1128/JVI.01150-06.

Gauthiez, E. *et al.* (2017) 'A systematic review and meta-analysis of HCV clearance', *Liver International*, 37(10), pp. 1431–1445. Available at: https://doi.org/10.1111/liv.13401.

Gawlik, K. and Gallay, P.A. (2014) 'HCV core protein and virus assembly: what we know without structures', *Immunologic Research*, 60(1), pp. 1–10. Available at: https://doi.org/10.1007/s12026-014-8494-3.

Ge, D. *et al.* (2009) 'Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance', *Nature*, 461(7262), pp. 399–401. Available at: https://doi.org/10.1038/nature08309.

Germi, R. *et al.* (2002) 'Cellular glycosaminoglycans and low density lipoprotein receptor are involved in hepatitis C virus adsorption', *Journal of Medical Virology*, 68(2), pp. 206–215. Available at: https://doi.org/10.1002/jmv.10196.

Gerold, G. *et al.* (2015) 'Quantitative Proteomics Identifies Serum Response Factor Binding Protein 1 as a Host Factor for Hepatitis C Virus Entry', *Cell Reports*, 12(5), pp. 864–878. Available at: https://doi.org/10.1016/j.celrep.2015.06.063.

Ghany, M.G. *et al.* (2011) 'An Update on Treatment of Genotype 1 Chronic Hepatitis C Virus Infection: 2011 Practice Guideline by the American Association for the Study of Liver Diseases', *Hepatology (Baltimore, Md.)*, 54(4), pp. 1433–1444. Available at: https://doi.org/10.1002/hep.24641.

Giang, E. *et al.* (2012) 'Human broadly neutralizing antibodies to the envelope glycoprotein complex of hepatitis C virus', *Proceedings of the National Academy of Sciences*, 109(16), pp. 6205–6210. Available at: https://doi.org/10.1073/pnas.1114927109.

Goodyear, T. *et al.* (2020) '"Everybody living with a chronic disease is entitled to be cured": Challenges and opportunities in scaling up access to direct-acting antiviral hepatitis C virus treatment among people who inject drugs', *International Journal of Drug Policy*, 81, p. 102766. Available at: https://doi.org/10.1016/j.drugpo.2020.102766.

Gopal, R. *et al.* (2017) 'Probing the antigenicity of hepatitis C virus envelope glycoprotein complex by high-throughput mutagenesis', *PLOS Pathogens*, 13(12), p. e1006735. Available at: https://doi.org/10.1371/journal.ppat.1006735.

Gosert, R. *et al.* (2003) 'Identification of the hepatitis C virus RNA replication complex in Huh-7 cells harboring subgenomic replicons', *Journal of Virology*,

77(9), pp. 5487–5492. Available at: https://doi.org/10.1128/jvi.77.9.5487-5492.2003.

Gottwein, J.M. *et al.* (2007) 'Robust Hepatitis C Genotype 3a Cell Culture Releasing Adapted Intergenotypic 3a/2a (S52/JFH1) Viruses', *Gastroenterology*, 133(5), pp. 1614–1626. Available at: https://doi.org/10.1053/j.gastro.2007.08.005.

Goverde, C. *et al.* (2022) 'De novo protein design by inversion of the AlphaFold structure prediction network'. bioRxiv, p. 2022.12.13.520346. Available at: https://doi.org/10.1101/2022.12.13.520346.

Grakoui, A. *et al.* (2003) 'HCV persistence and immune evasion in the absence of memory T cell help', *Science (New York, N.Y.)*, 302(5645), pp. 659–662. Available at: https://doi.org/10.1126/science.1088774.

GraphPad Software (no date) 'GraphPad Prism'. Boston, MA: GraphPad Software.

Greener, J.G. *et al.* (2022) 'A guide to machine learning for biologists', *Nature Reviews Molecular Cell Biology*, 23(1), pp. 40–55. Available at: https://doi.org/10.1038/s41580-021-00407-0.

Guedj, J. *et al.* (2013) 'Modeling shows that the NS5A inhibitor daclatasvir has two modes of action and yields a shorter estimate of the hepatitis C virus half-life', *Proceedings of the National Academy of Sciences*, 110(10), pp. 3991–3996. Available at: https://doi.org/10.1073/pnas.1203110110.

Guest, J.D. *et al.* (2021) 'Design of a native-like secreted form of the hepatitis C virus E1E2 heterodimer', *Proceedings of the National Academy of Sciences of the United States of America*, 118(3), p. e2015149118. Available at: https://doi.org/10.1073/pnas.2015149118.

Guntipalli, P. *et al.* (2021) 'Worldwide prevalence, genotype distribution and management of hepatitis C', *Acta Gastro-Enterologica Belgica*, 84(4), pp. 637–656. Available at: https://doi.org/10.51821/84.4.015.

Haddad, J.G. *et al.* (2017) 'Identification of Novel Functions for Hepatitis C Virus Envelope Glycoprotein E1 in Virus Entry and Assembly', *Journal of Virology*, 91(8), pp. e00048-17. Available at: https://doi.org/10.1128/JVI.00048-17.

Haley, R.W. and Fischer, R.P. (2001) 'Commercial Tattooing as a Potentially Important Source of Hepatitis C Infection: Clinical Epidemiology of 626 Consecutive Patients Unaware of Their Hepatitis C Serologic Status', *Medicine*, 80(2), p. 134.

Harris, H.E. *et al.* (2007) 'Does the clinical outcome of hepatitis C infection vary with the infecting hepatitis C virus type?', *Journal of Viral Hepatitis*, 14(3), pp. 213–220. Available at: https://doi.org/10.1111/j.1365-2893.2006.00795.x.

Harris, H.J. *et al.* (2010) 'Claudin association with CD81 defines hepatitis C virus entry', *The Journal of Biological Chemistry*, 285(27), pp. 21092–21102. Available at: https://doi.org/10.1074/jbc.M110.104836.

Harrison, S.C. (2015) 'Viral membrane fusion', *Virology*, 0, pp. 498–507. Available at: https://doi.org/10.1016/j.virol.2015.03.043.

Hashem, Y. *et al.* (2013) 'Hepatitis-C-virus-like internal ribosome entry sites displace eIF3 to gain access to the 40S subunit', *Nature*, 503(7477), pp. 539–543. Available at: https://doi.org/10.1038/nature12658.

He, L. *et al.* (2015) 'Approaching rational epitope vaccine design for hepatitis C virus with meta-server and multivalent scaffolding', *Scientific Reports*, 5, p. 12501. Available at: https://doi.org/10.1038/srep12501.

He, Y. *et al.* (2008) 'Relative replication capacity and selective advantage profiles of protease inhibitor-resistant hepatitis C virus (HCV) NS3 protease mutants in the HCV genotype 1b replicon system', *Antimicrobial Agents and Chemotherapy*, 52(3), pp. 1101–1110. Available at: https://doi.org/10.1128/AAC.01149-07.

Heldwein, E.E. *et al.* (2006) 'Crystal structure of glycoprotein B from herpes simplex virus 1', *Science (New York, N.Y.)*, 313(5784), pp. 217–220. Available at: https://doi.org/10.1126/science.1126548.

Helle, F. *et al.* (2007) 'The neutralizing activity of anti-hepatitis C virus antibodies is modulated by specific glycans on the E2 envelope protein', *Journal of Virology*, 81(15), pp. 8101–8111. Available at: https://doi.org/10.1128/JVI.00127-07.

Helle, F. *et al.* (2010) 'Role of N-Linked Glycans in the Functions of Hepatitis C Virus Envelope Proteins Incorporated into Infectious Virions', *Journal of Virology*, 84(22), pp. 11905–11915. Available at: https://doi.org/10.1128/jvi.01548-10.

Henne, W.M., Buchkovich, N.J. and Emr, S.D. (2011) 'The ESCRT Pathway', *Developmental Cell*, 21(1), pp. 77–91. Available at: https://doi.org/10.1016/j.devcel.2011.05.015.

Heo, Y.-A. and Deeks, E.D. (2018) 'Sofosbuvir/Velpatasvir/Voxilaprevir: A Review in Chronic Hepatitis C', *Drugs*, 78(5), pp. 577–587. Available at: https://doi.org/10.1007/s40265-018-0895-5.

Herker, E. *et al.* (2010) 'Efficient hepatitis C virus particle formation requires diacylglycerol acyltransferase-1', *Nature Medicine*, 16(11), pp. 1295–1298. Available at: https://doi.org/10.1038/nm.2238.

Hijikata, M. *et al.* (1991) 'Hypervariable regions in the putative glycoprotein of hepatitis C virus', *Biochemical and Biophysical Research Communications*,

175(1), pp. 220–228. Available at: https://doi.org/10.1016/s0006-291x(05)81223-9.

Höhne, M., Schreier, E. and Roggendorf, M. (1994) 'Sequence variability in the env-coding region of hepatitis C virus isolated from patients infected during a single source outbreak', *Archives of Virology*, 137(1), pp. 25–34. Available at: https://doi.org/10.1007/BF01311170.

Holsen, D.S., Harthug, S. and Myrmel, H. (1993) 'Prevalence of antibodies to hepatitis C virus and association with intravenous drug abuse and tattooing in a national prison in Norway', *European Journal of Clinical Microbiology and Infectious Diseases*, 12(9), pp. 673–676. Available at: https://doi.org/10.1007/BF02009378.

Hopcraft, S.E. and Evans, M.J. (2015) 'Selection of a Hepatitis C Virus With Altered Entry Factor Requirements Reveals a Genetic Interaction Between the E1 Glycoprotein and Claudins', *Hepatology (Baltimore, Md.)*, 62(4), pp. 1059–1069. Available at: https://doi.org/10.1002/hep.27815.

Hsieh, F.-L. *et al.* (2016) 'The structural basis for CD36 binding by the malaria parasite', *Nature Communications*, 7, p. 12837. Available at: https://doi.org/10.1038/ncomms12837.

Hsu, M. *et al.* (2003) 'Hepatitis C virus glycoproteins mediate pH-dependent cell entry of pseudotyped retroviral particles', *Proceedings of the National Academy of Sciences of the United States of America*, 100(12), pp. 7271–7276. Available at: https://doi.org/10.1073/pnas.0832180100.

Hu, L. *et al.* (2022) 'Novel fold of rotavirus glycan-binding domain predicted by AlphaFold2 and determined by X-ray crystallography', *Communications Biology*, 5(1), p. 419. Available at: https://doi.org/10.1038/s42003-022-03357-1.

Huang, H. *et al.* (2007) 'Hepatitis C virus production by human hepatocytes dependent on assembly and secretion of very low-density lipoproteins', *Proceedings of the National Academy of Sciences of the United States of America*, 104(14), pp. 5848–5853. Available at: https://doi.org/10.1073/pnas.0700760104.

Hunter, P. (2006) 'Into the fold', *EMBO Reports*, 7(3), pp. 249–252. Available at: https://doi.org/10.1038/sj.embor.7400655.

Hutin, S. *et al.* (2022) 'The Vaccinia Virus DNA Helicase Structure from Combined Single-Particle Cryo-Electron Microscopy and AlphaFold2 Prediction', *Viruses*, 14(10), p. 2206. Available at: https://doi.org/10.3390/v14102206.

Hutin, Y.J.F., Hauri, A.M. and Armstrong, G.L. (2003) 'Use of injections in healthcare settings worldwide, 2000: literature review and regional estimates', *BMJ*, 327(7423), p. 1075. Available at: https://doi.org/10.1136/bmj.327.7423.1075.

Issur, M. and Götte, M. (2014) 'Resistance Patterns Associated with HCV NS5A Inhibitors Provide Limited Insight into Drug Binding', *Viruses*, 6(11), pp. 4227–4241. Available at: https://doi.org/10.3390/v6114227.

Jackson, D., Tabor, E. and Gerety, R.J. (1979) 'Acute non-A, non-B hepatitis: specific ultrastructural alterations in endoplasmic reticulum of infected hepatocytes', *Lancet (London, England)*, 1(8128), pp. 1249–1250. Available at: https://doi.org/10.1016/s0140-6736(79)91938-x.

Jackson, P. *et al.* (1997) 'Reactivity of synthetic peptides representing selected sections of hepatitis C virus core and envelope proteins with a panel of hepatitis C virus-seropositive human plasma', *Journal of Medical Virology*, 51(1), pp. 67–79. Available at: https://doi.org/10.1002/(sici)1096-9071(199701)51:1%3C67::aid-jmv11%3E3.0.co;2-1.

Jacobson, I.M. *et al.* (2011) 'Telaprevir for previously untreated chronic hepatitis C virus infection', *The New England Journal of Medicine*, 364(25), pp. 2405–2416. Available at: https://doi.org/10.1056/NEJMoa1012912.

Jendrusch, M., Korbel, J.O. and Sadiq, S.K. (2021) 'AlphaDesign: A de novo protein design framework based on AlphaFold'. bioRxiv, p. 2021.10.11.463937. Available at: https://doi.org/10.1101/2021.10.11.463937.

Jin, F. *et al.* (2021) 'Prevalence and incidence of hepatitis C virus infection in men who have sex with men: a systematic review and meta-analysis', *The Lancet. Gastroenterology & Hepatology*, 6(1), pp. 39–56. Available at: https://doi.org/10.1016/S2468-1253(20)30303-4.

Jirasko, V. *et al.* (2010) 'Structural and Functional Studies of Nonstructural Protein 2 of the Hepatitis C Virus Reveal Its Key Role as Organizer of Virion Assembly', *PLOS Pathogens*, 6(12), p. e1001233. Available at: https://doi.org/10.1371/journal.ppat.1001233.

Johnson, L.S., Eddy, S.R. and Portugaly, E. (2010) 'Hidden Markov model speed heuristic and iterative HMM search procedure', *BMC Bioinformatics*, 11, p. 431. Available at: https://doi.org/10.1186/1471-2105-11-431.

de Jong, Y.P. *et al.* (2014) 'Broadly neutralizing antibodies abrogate established hepatitis C virus infection', *Science Translational Medicine*, 6(254), p. 254ra129. Available at: https://doi.org/10.1126/scitranslmed.3009512.

Jopling, C.L. *et al.* (2005) 'Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA', *Science (New York, N.Y.)*, 309(5740), pp. 1577–1581. Available at: https://doi.org/10.1126/science.1113329.

Jopling, C.L., Schütz, S. and Sarnow, P. (2008) 'Position-dependent function for a tandem microRNA miR-122-binding site located in the hepatitis C virus RNA genome', *Cell Host & Microbe*, 4(1), pp. 77–85. Available at: https://doi.org/10.1016/j.chom.2008.05.013.

Jumper, J. *et al.* (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596(7873), pp. 583–589. Available at: https://doi.org/10.1038/s41586-021-03819-2.

Kalemera, M. *et al.* (2019a) 'Building a mechanistic mathematical model of hepatitis C virus entry', *PLOS Computational Biology*, 15(3), p. e1006905. Available at: https://doi.org/10.1371/journal.pcbi.1006905.

Kalemera, M. *et al.* (2019b) 'Building a mechanistic mathematical model of hepatitis C virus entry', *PLOS Computational Biology*, 15(3), p. e1006905. Available at: https://doi.org/10.1371/journal.pcbi.1006905.

Kalemera, M.D. *et al.* (2021) 'Optimized cell systems for the investigation of hepatitis C virus E1E2 glycoproteins', *Journal of General Virology*, 102(1), p. 001512. Available at: https://doi.org/10.1099/jgv.0.001512.

Kallenborn, F. *et al.* (2025) 'GPU-accelerated homology search with MMseqs2'. bioRxiv, p. 2024.11.13.623350. Available at: https://doi.org/10.1101/2024.11.13.623350.

Kato, N. *et al.* (1992) 'Characterization of hypervariable regions in the putative envelope protein of hepatitis C virus', *Biochemical and Biophysical Research Communications*, 189(1), pp. 119–127. Available at: https://doi.org/10.1016/0006-291x(92)91533-v.

Kato, T. *et al.* (2001) 'Sequence analysis of hepatitis C virus isolated from a fulminant hepatitis patient', *Journal of Medical Virology*, 64(3), pp. 334–339. Available at: https://doi.org/10.1002/jmv.1055.

Kato, T. *et al.* (2003) 'Efficient replication of the genotype 2a hepatitis C virus subgenomic replicon', *Gastroenterology*, 125(6), pp. 1808–1817. Available at: https://doi.org/10.1053/j.gastro.2003.09.023.

Katoh, K., Asimenos, G. and Toh, H. (2009) 'Multiple Alignment of DNA Sequences with MAFFT', in D. Posada (ed.) *Bioinformatics for DNA Sequence Analysis*. Totowa, NJ: Humana Press (Methods in Molecular Biology), pp. 39–64. Available at: https://doi.org/10.1007/978-1-59745-251-9_3.

Keck, M.-L. *et al.* (2018) 'Mapping Determinants of Virus Neutralization and Viral Escape for Rational Design of a Hepatitis C Virus Vaccine', *Frontiers in Immunology*, 9, p. 1194. Available at: https://doi.org/10.3389/fimmu.2018.01194.

Keck, Z. *et al.* (2016) 'Antibody Response to Hypervariable Region 1 Interferes with Broadly Neutralizing Antibodies to Hepatitis C Virus', *Journal of Virology*, 90(6), pp. 3112–3122. Available at: https://doi.org/10.1128/JVI.02458-15.

Keck, Z.-Y. *et al.* (2016) 'Affinity maturation of a broadly neutralizing human monoclonal antibody that prevents acute hepatitis C virus infection in mice',

Hepatology (Baltimore, Md.), 64(6), pp. 1922–1933. Available at: https://doi.org/10.1002/hep.28850.

Kelly, C. *et al.* (2015) 'Cross-reactivity of hepatitis C virus specific vaccine-induced T cells at immunodominant epitopes', *European Journal of Immunology*, 45(1), pp. 309–316. Available at: https://doi.org/10.1002/eji.201444686.

Khan, A.G. *et al.* (2014) 'Structure of the core ectodomain of the hepatitis C virus envelope glycoprotein 2', *Nature*, 509(7500), pp. 381–384. Available at: https://doi.org/10.1038/nature13117.

Khan, I. *et al.* (2014) 'Modulation of hepatitis C virus genome replication by glycosphingolipids and four-phosphate adaptor protein 2', *Journal of Virology*, 88(21), pp. 12276–12295. Available at: https://doi.org/10.1128/JVI.00970-14.

Khan, S.M. *et al.* (2022) 'Impact of CD4 T cells on intratumoral CD8 T-cell exhaustion and responsiveness to PD-1 blockade therapy in mouse brain tumors', *Journal for ImmunoTherapy of Cancer*, 10(12). Available at: https://doi.org/10.1136/jitc-2022-005293.

Kim, R.S. *et al.* (2025) 'BFVD—a large repository of predicted viral protein structures', *Nucleic Acids Research*, 53(D1), pp. D340–D347. Available at: https://doi.org/10.1093/nar/gkae1119.

Kinchen, V.J. *et al.* (2018) 'Broadly Neutralizing Antibody Mediated Clearance of Human Hepatitis C Virus Infection', *Cell Host & Microbe*, 24(5), pp. 717-730.e5. Available at: https://doi.org/10.1016/j.chom.2018.10.012.

Klasse, P.J., Nixon, D.F. and Moore, J.P. (2021) 'Immunogenicity of clinically relevant SARS-CoV-2 vaccines in nonhuman primates and humans', *Science Advances*, 7(12), p. eabe8065. Available at: https://doi.org/10.1126/sciadv.abe8065.

Kleckner, I.R. and Foster, M.P. (2011) 'An introduction to NMR-based approaches for measuring protein dynamics', *Biochimica et biophysica acta*, 1814(8), pp. 942–968. Available at: https://doi.org/10.1016/j.bbapap.2010.10.012.

Ko, Y.-C. *et al.* (1992) 'Tattooing as a risk of hepatitis C virus infection', *Journal of Medical Virology*, 38(4), pp. 288–291. Available at: https://doi.org/10.1002/jmv.1890380411.

Kolykhalov, A.A., Feinstone, S.M. and Rice, C.M. (1996) 'Identification of a highly conserved sequence element at the 3' terminus of hepatitis C virus genome RNA', *Journal of Virology*, 70(6), pp. 3363–3371. Available at: https://doi.org/10.1128/JVI.70.6.3363-3371.1996.

Kong, L., Giang, E., Robbins, J.B., *et al.* (2012) 'Structural basis of hepatitis C virus neutralization by broadly neutralizing antibody HCV1', *Proceedings of the*

National Academy of Sciences, 109(24), pp. 9499–9504. Available at: https://doi.org/10.1073/pnas.1202924109.

Kong, L., Giang, E., Nieusma, T., *et al.* (2012) 'Structure of Hepatitis C Virus Envelope Glycoprotein E2 Antigenic Site 412 to 423 in Complex with Antibody AP33', *Journal of Virology*, 86(23), pp. 13085–13088. Available at: https://doi.org/10.1128/jvi.01939-12.

Kong, L. *et al.* (2013) 'Hepatitis C Virus E2 Envelope Glycoprotein Core Structure', *Science*, 342(6162), pp. 1090–1094. Available at: https://doi.org/10.1126/science.1243876.

Korenaga, M. *et al.* (2001) 'A possible role of hypervariable region 1 quasispecies in escape of hepatitis C virus particles from neutralization', *Journal of Viral Hepatitis*, 8(5), pp. 331–340. Available at: https://doi.org/10.1046/j.1365-2893.2001.00305.x.

Krey, T. *et al.* (2010) 'The disulfide bonds in glycoprotein E2 of hepatitis C virus reveal the tertiary organization of the molecule', *PLoS pathogens*, 6(2), p. e1000762. Available at: https://doi.org/10.1371/journal.ppat.1000762.

Krieger, S.E. *et al.* (2010) 'Inhibition of hepatitis C virus infection by anti-claudin-1 antibodies is mediated by neutralization of E2–CD81–Claudin-1 associations†', *Hepatology*, 51(4), p. 1144. Available at: https://doi.org/10.1002/hep.23445.

Kumagai, N. *et al.* (2007) 'Complexity of the HVR-1 quasispecies and disease activity in patients with hepatitis C', *European Journal of Clinical Investigation*, 37(7), pp. 566–572. Available at: https://doi.org/10.1111/j.1365-2362.2007.01825.x.

Kumar, A. *et al.* (2021) 'Structural insights into hepatitis C virus receptor binding and entry', *Nature*, 598(7881), pp. 521–525. Available at: https://doi.org/10.1038/s41586-021-03913-5.

Kumar, A. *et al.* (2023) 'Regions of hepatitis C virus E2 required for membrane association', *Nature Communications*, 14, p. 433. Available at: https://doi.org/10.1038/s41467-023-36183-y.

Kundrot, C.E. (2004) 'Which strategy for a protein crystallization project?', *Cellular and Molecular Life Sciences CMLS*, 61(5), pp. 525–536. Available at: https://doi.org/10.1007/s00018-003-3260-z.

Kwo, P.Y. *et al.* (2010) 'Efficacy of boceprevir, an NS3 protease inhibitor, in combination with peginterferon alfa-2b and ribavirin in treatment-naive patients with genotype 1 hepatitis C infection (SPRINT-1): an open-label, randomised, multicentre phase 2 trial', *Lancet (London, England)*, 376(9742), pp. 705–716. Available at: https://doi.org/10.1016/S0140-6736(10)60934-8.

Lam, A.M. *et al.* (2011) 'Hepatitis C Virus Nucleotide Inhibitors PSI-352938 and PSI-353661 Exhibit a Novel Mechanism of Resistance Requiring Multiple Mutations within Replicon RNA', *Journal of Virology*, 85(23), pp. 12334–12342. Available at: https://doi.org/10.1128/jvi.05639-11.

Lam, A.M. *et al.* (2012) 'Genotype and subtype profiling of PSI-7977 as a nucleotide inhibitor of hepatitis C virus', *Antimicrobial Agents and Chemotherapy*, 56(6), pp. 3359–3368. Available at: https://doi.org/10.1128/AAC.00054-12.

Langlois, A.-C. *et al.* (2020) 'Molecular determinants of SR-B1-dependent Plasmodium sporozoite entry into hepatocytes', *Scientific Reports*, 10(1), p. 13509. Available at: https://doi.org/10.1038/s41598-020-70468-2.

Lapedes, A.S. *et al.* (1999) 'Correlated Mutations in Models of Protein Sequences: Phylogenetic and Structural Effects', *Lecture Notes-Monograph Series*, 33, pp. 236–256.

Larsson, A. (2014) 'AliView: a fast and lightweight alignment viewer and editor for large datasets', *Bioinformatics*, 30(22), pp. 3276–3278. Available at: https://doi.org/10.1093/bioinformatics/btu531.

Lavillette, D. *et al.* (2005) 'Characterization of host-range and cell entry properties of the major genotypes and subtypes of hepatitis C virus†', *Hepatology*, 41(2), p. 265. Available at: https://doi.org/10.1002/hep.20542.

Lavillette, D. *et al.* (2007) 'Characterization of fusion determinants points to the involvement of three discrete regions of both E1 and E2 glycoproteins in the membrane fusion process of hepatitis C virus', *Journal of Virology*, 81(16), pp. 8752–8765. Available at: https://doi.org/10.1128/JVI.02642-06.

Law, J.L.M. *et al.* (2013) 'A hepatitis C virus (HCV) vaccine comprising envelope glycoproteins gpE1/gpE2 derived from a single isolate elicits broad cross-genotype neutralizing antibodies in humans', *PloS One*, 8(3), p. e59776. Available at: https://doi.org/10.1371/journal.pone.0059776.

Law, J.L.M. *et al.* (2018) 'Role of the E2 Hypervariable Region (HVR1) in the Immunogenicity of a Recombinant Hepatitis C Virus Vaccine', *Journal of Virology*, 92(11), pp. e02141-17. Available at: https://doi.org/10.1128/JVI.02141-17.

Law, M. *et al.* (2008) 'Broadly neutralizing antibodies protect against hepatitis C virus quasispecies challenge', *Nature Medicine*, 14(1), pp. 25–27. Available at: https://doi.org/10.1038/nm1698.

Lee, J.-Y. *et al.* (2019) 'Spatiotemporal Coupling of the Hepatitis C Virus Replication Cycle by Creating a Lipid Droplet- Proximal Membranous Replication Compartment', *Cell Reports*, 27(12), pp. 3602-3617.e5. Available at: https://doi.org/10.1016/j.celrep.2019.05.063.

Lee, W.-P. *et al.* (2024) 'Ser235 phosphorylation of hepatitis C virus NS5A is required for NS5A dimerization and drug resistance', *Life Sciences*, 337, p. 122338. Available at: https://doi.org/10.1016/j.lfs.2023.122338.

Lemm, J.A. *et al.* (2010) 'Identification of hepatitis C virus NS5A inhibitors', *Journal of Virology*, 84(1), pp. 482–491. Available at: https://doi.org/10.1128/JVI.01360-09.

Letunic, I. and Bork, P. (2024) 'Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool', *Nucleic Acids Research*, 52(W1), pp. W78–W82. Available at: https://doi.org/10.1093/nar/gkae268.

Levinthal, C. (1968) 'Are there pathways for protein folding?', *Journal de Chimie Physique*, 65, pp. 44–45. Available at: https://doi.org/10.1051/jcp/1968650044.

Li, X. *et al.* (2014) 'Hepatocyte nuclear factor 4α and downstream secreted phospholipase A2 GXIIB regulate production of infectious hepatitis C virus', *Journal of Virology*, 88(1), pp. 612–627. Available at: https://doi.org/10.1128/JVI.02068-13.

Li, Y. *et al.* (2013) 'Competing and noncompeting activities of miR-122 and the 5' exonuclease Xrn1 in regulation of hepatitis C virus replication', *Proceedings of the National Academy of Sciences of the United States of America*, 110(5), pp. 1881–1886. Available at: https://doi.org/10.1073/pnas.1213515110.

Li, Y. and Modis, Y. (2014) 'A novel membrane fusion protein family in Flaviviridae?', *Trends in Microbiology*, 22(4), pp. 176–182. Available at: https://doi.org/10.1016/j.tim.2014.01.008.

Lin, C. *et al.* (2005) 'In vitro studies of cross-resistance mutations against two hepatitis C virus serine protease inhibitors, VX-950 and BILN 2061', *The Journal of Biological Chemistry*, 280(44), pp. 36784–36791. Available at: https://doi.org/10.1074/jbc.M506462200.

Lindenbach, B.D. *et al.* (2005) 'Complete Replication of Hepatitis C Virus in Cell Culture', *Science*, 309(5734), pp. 623–626. Available at: https://doi.org/10.1126/science.1114016.

Lindenbach, B.D. and Rice, C.M. (2013) 'The ins and outs of hepatitis C virus entry and assembly', *Nature Reviews Microbiology*, 11(10), pp. 688–700. Available at: https://doi.org/10.1038/nrmicro3098.

Litvin, U. *et al.* (2025) 'Viro3D: a comprehensive database of virus protein structure predictions', *Molecular Systems Biology* [Preprint]. Available at: https://doi.org/10.1038/s44320-025-00147-9.

Liu, S. *et al.* (2009) 'Tight Junction Proteins Claudin-1 and Occludin Control Hepatitis C Virus Entry and Are Downregulated during Infection To Prevent

Superinfection', *Journal of Virology*, 83(4), pp. 2011–2014. Available at: https://doi.org/10.1128/jvi.01888-08.

Lohmann, V. *et al.* (1999) 'Replication of Subgenomic Hepatitis C Virus RNAs in a Hepatoma Cell Line', *Science*, 285(5424), pp. 110–113. Available at: https://doi.org/10.1126/science.285.5424.110.

Lu, M. *et al.* (2020) 'Real-Time Conformational Dynamics of SARS-CoV-2 Spikes on Virus Particles', *Cell Host & Microbe*, 28(6), pp. 880-891.e8. Available at: https://doi.org/10.1016/j.chom.2020.11.001.

Lupberger, J. *et al.* (2011) 'EGFR and EphA2 are host factors for hepatitis C virus entry and possible targets for antiviral therapy', *Nature Medicine*, 17(5), pp. 589–595. Available at: https://doi.org/10.1038/nm.2341.

Lyumkis, D. (2019) 'Challenges and opportunities in cryo-EM single-particle analysis', *The Journal of Biological Chemistry*, 294(13), pp. 5181–5197. Available at: https://doi.org/10.1074/jbc.REV118.005602.

Ma, H. *et al.* (2007) 'Characterization of the metabolic activation of hepatitis C virus nucleoside inhibitor beta-D-2'-Deoxy-2'-fluoro-2'-C-methylcytidine (PSI-6130) and identification of a novel active 5'-triphosphate species', *The Journal of Biological Chemistry*, 282(41), pp. 29812–29820. Available at: https://doi.org/10.1074/jbc.M705274200.

Ma, W. *et al.* (2022) 'Enhancing Protein Function Prediction Performance by Utilizing AlphaFold-Predicted Protein Structures', *Journal of Chemical Information and Modeling*, 62(17), pp. 4008–4017. Available at: https://doi.org/10.1021/acs.jcim.2c00885.

Ma, Y. *et al.* (2011) 'Hepatitis C Virus NS2 Protein Serves as a Scaffold for Virus Assembly by Interacting with both Structural and Nonstructural Proteins', *Journal of Virology*, 85(1), pp. 86–97. Available at: https://doi.org/10.1128/JVI.01070-10.

Major, M.E. *et al.* (2002) 'Previously infected and recovered chimpanzees exhibit rapid responses that control hepatitis C virus replication upon rechallenge', *Journal of Virology*, 76(13), pp. 6586–6595. Available at: https://doi.org/10.1128/jvi.76.13.6586-6595.2002.

Marcellin, P. (1999) 'Hepatitis C: the clinical spectrum of the disease', *Journal of Hepatology*, 31, pp. 9–16. Available at: https://doi.org/10.1016/S0168-8278(99)80368-7.

Mariano, A. *et al.* (2004) 'Role of beauty treatment in the spread of parenterally transmitted hepatitis viruses in Italy', *Journal of Medical Virology*, 74(2), pp. 216–220. Available at: https://doi.org/10.1002/jmv.20182.

Marín, M.Q. *et al.* (2020) 'Optimized Hepatitis C Virus (HCV) E2 Glycoproteins and their Immunogenicity in Combination with MVA-HCV', *Vaccines*, 8(3), p. 440. Available at: https://doi.org/10.3390/vaccines8030440.

Marks, D.S., Hopf, T.A. and Sander, C. (2012) 'Protein structure prediction from sequence variation', *Nature biotechnology*, 30(11), pp. 1072–1080. Available at: https://doi.org/10.1038/nbt.2419.

Marshall, A.D. *et al.* (2024) 'Direct-acting antiviral therapies for hepatitis C infection: global registration, reimbursement, and restrictions', *The Lancet Gastroenterology & Hepatology*, 9(4), pp. 366–382. Available at: https://doi.org/10.1016/S2468-1253(23)00335-7.

Martell, M. *et al.* (1992) 'Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution', *Journal of Virology*, 66(5), pp. 3225–3229. Available at: https://doi.org/10.1128/JVI.66.5.3225-3229.1992.

Martin, D.N. and Uprichard, S.L. (2013) 'Identification of transferrin receptor 1 as a hepatitis C virus entry factor', *Proceedings of the National Academy of Sciences of the United States of America*, 110(26), pp. 10777–10782. Available at: https://doi.org/10.1073/pnas.1301764110.

Martinello, M. *et al.* (2017) 'HCV reinfection incidence among individuals treated for recent infection', *Journal of Viral Hepatitis*, 24(5), pp. 359–370. Available at: https://doi.org/10.1111/jvh.12666.

Martino, S.D., Petri, G.L. and De Rosa, M. (2024) 'Hepatitis C: The Story of a Long Journey through First, Second, and Third Generation NS3/4A Peptidomimetic Inhibitors. What Did We Learn?', *Journal of Medicinal Chemistry*, 67(2), pp. 885–921. Available at: https://doi.org/10.1021/acs.jmedchem.3c01971.

Masaki, T. *et al.* (2014) 'Involvement of hepatitis C virus NS5A hyperphosphorylation mediated by casein kinase I-α in infectious virus production', *Journal of Virology*, 88(13), pp. 7541–7555. Available at: https://doi.org/10.1128/JVI.03170-13.

Matsuda, D. and Mauro, V.P. (2014) 'Base pairing between hepatitis C virus RNA and 18S rRNA is required for IRES-dependent translation initiation in vivo', *Proceedings of the National Academy of Sciences of the United States of America*, 111(43), pp. 15385–15389. Available at: https://doi.org/10.1073/pnas.1413472111.

McAllister, J. *et al.* (1998) 'Long-Term Evolution of the Hypervariable Region of Hepatitis C Virus in a Common-Source-Infected Cohort', *Journal of Virology*, 72(6), pp. 4893–4905. Available at: https://doi.org/10.1128/jvi.72.6.4893-4905.1998.

McCauley, J.A. and Rudd, M.T. (2016) 'Hepatitis C virus NS3/4a protease inhibitors', *Current Opinion in Pharmacology*, 30, pp. 84–92. Available at: https://doi.org/10.1016/j.coph.2016.07.015.

McGivern, D.R. *et al.* (2014) 'Kinetic Analyses Reveal Potent and Early Blockade of Hepatitis C Virus Assembly by NS5A Inhibitors', *Gastroenterology*, 147(2), pp. 453-462.e7. Available at: https://doi.org/10.1053/j.gastro.2014.04.021.

McKeating, J.A. *et al.* (2004) 'Diverse Hepatitis C Virus Glycoproteins Mediate Viral Infection in a CD81-Dependent Manner', *Journal of Virology*, 78(16), pp. 8496–8505. Available at: https://doi.org/10.1128/jvi.78.16.8496-8505.2004.

McMullan, L.K. *et al.* (2007) 'Evidence for a functional RNA element in the hepatitis C virus core gene', *Proceedings of the National Academy of Sciences of the United States of America*, 104(8), pp. 2879–2884. Available at: https://doi.org/10.1073/pnas.0611267104.

McPherson, A. and Gavira, J.A. (2014) 'Introduction to protein crystallization', *Acta Crystallographica Section F: Structural Biology Communications*, 70(1), pp. 2–20. Available at: https://doi.org/10.1107/S2053230X13033141.

Mee, C.J. *et al.* (2009) 'Polarization Restricts Hepatitis C Virus Entry into HepG2 Hepatoma Cells', *Journal of Virology*, 83(12), pp. 6211–6221. Available at: https://doi.org/10.1128/jvi.00246-09.

Mehellou, Y., Rattan, H.S. and Balzarini, J. (2018) 'The ProTide Prodrug Technology: From the Concept to the Clinic', *Journal of Medicinal Chemistry*, 61(6), pp. 2211–2226. Available at: https://doi.org/10.1021/acs.jmedchem.7b00734.

Mele, A. *et al.* (1999) 'Heterosexual transmission of hepatitis C in Italy', *Journal of Medical Virology*, 57(2), pp. 111–113. Available at: https://doi.org/10.1002/(SICI)1096-9071(199902)57:2%3C111::AID-JMV4%3E3.0.CO;2-C.

Menzel, N. *et al.* (2012) 'MAP-Kinase Regulated Cytosolic Phospholipase A2 Activity Is Essential for Production of Infectious Hepatitis C Virus Particles', *PLOS Pathogens*, 8(7), p. e1002829. Available at: https://doi.org/10.1371/journal.ppat.1002829.

Merz, A. *et al.* (2011) 'Biochemical and Morphological Properties of Hepatitis C Virus Particles and Determination of Their Lipidome*', *Journal of Biological Chemistry*, 286(4), pp. 3018–3032. Available at: https://doi.org/10.1074/jbc.M110.175018.

Mesalam, A.A. *et al.* (2018) 'Development and characterization of a human monoclonal antibody targeting the N-terminal region of hepatitis C virus envelope glycoprotein E1', *Virology*, 514, pp. 30–41. Available at: https://doi.org/10.1016/j.virol.2017.10.019.

Messina, J.P. *et al.* (2015) 'Global distribution and prevalence of hepatitis C virus genotypes', *Hepatology (Baltimore, Md.)*, 61(1), pp. 77–87. Available at: https://doi.org/10.1002/hep.27259.

Meunier, J.-C. *et al.* (1999) 'Analysis of the glycosylation sites of hepatitis C virus (HCV) glycoprotein E1 and the influence of E1 glycans on the formation of the HCV glycoprotein complex.', *Journal of General Virology*, 80(4), pp. 887–896. Available at: https://doi.org/10.1099/0022-1317-80-4-887.

Meunier, J.-C. *et al.* (2005) 'Evidence for cross-genotype neutralization of hepatitis C virus pseudo-particles and enhancement of infectivity by apolipoprotein C1', *Proceedings of the National Academy of Sciences*, 102(12), pp. 4560–4565. Available at: https://doi.org/10.1073/pnas.0501275102.

Meunier, J.-C., Russell, R.S., Engle, R.E., *et al.* (2008) 'Apolipoprotein C1 Association with Hepatitis C Virus', *Journal of Virology*, 82(19), pp. 9647–9656. Available at: https://doi.org/10.1128/jvi.00914-08.

Meunier, J.-C., Russell, R.S., Goossens, V., *et al.* (2008) 'Isolation and characterization of broadly neutralizing human monoclonal antibodies to the e1 glycoprotein of hepatitis C virus', *Journal of Virology*, 82(2), pp. 966–973. Available at: https://doi.org/10.1128/JVI.01872-07.

Micallef, J.M., Kaldor, J.M. and Dore, G.J. (2006) 'Spontaneous viral clearance following acute hepatitis C infection: a systematic review of longitudinal studies', *Journal of Viral Hepatitis*, 13(1), pp. 34–41. Available at: https://doi.org/10.1111/j.1365-2893.2005.00651.x.

Midgard, H. *et al.* (2016) 'Hepatitis C reinfection after sustained virological response', *Journal of Hepatology*, 64(5), pp. 1020–1026. Available at: https://doi.org/10.1016/j.jhep.2016.01.001.

Mifsud, J.C.O. *et al.* (2024) 'Mapping glycoprotein structure reveals Flaviviridae evolutionary history', *Nature*, 633(8030), pp. 695–703. Available at: https://doi.org/10.1038/s41586-024-07899-8.

Migliaccio, G. *et al.* (2003) 'Characterization of resistance to non-obligate chain-terminating ribonucleoside analogs that inhibit hepatitis C virus replication in vitro', *The Journal of Biological Chemistry*, 278(49), pp. 49164–49170. Available at: https://doi.org/10.1074/jbc.M305041200.

Miller, M.A. and Pisani, E. (1999) 'The cost of unsafe injections.', *Bulletin of the World Health Organization*, 77(10), pp. 808–811.

Minh, B.Q. *et al.* (2020) 'IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era', *Molecular Biology and Evolution*, 37(5), pp. 1530–1534. Available at: https://doi.org/10.1093/molbev/msaa015.

Mirdita, M. *et al.* (2022) 'ColabFold: making protein folding accessible to all', *Nature Methods*, 19(6), pp. 679–682. Available at: https://doi.org/10.1038/s41592-022-01488-1.

Miyanari, Y. *et al.* (2007) 'The lipid droplet is an important organelle for hepatitis C virus production', *Nature Cell Biology*, 9(9), pp. 1089–1097. Available at: https://doi.org/10.1038/ncb1631.

Modis, Y. *et al.* (2003) 'A ligand-binding pocket in the dengue virus envelope glycoprotein', *Proceedings of the National Academy of Sciences of the United States of America*, 100(12), pp. 6986–6991. Available at: https://doi.org/10.1073/pnas.0832193100.

Modis, Y. (2013) 'Class II Fusion Proteins', *Madame Curie Bioscience Database [Internet]*. Landes Bioscience. Available at: https://www.ncbi.nlm.nih.gov/books/NBK154547/ (Accessed: 25 August 2025).

Mohd Hanafiah, K. *et al.* (2013) 'Global epidemiology of hepatitis C virus infection: New estimates of age-specific antibody to HCV seroprevalence', *Hepatology*, 57(4), p. 1333. Available at: https://doi.org/10.1002/hep.26141.

Mollica, L. *et al.* (2016) 'Binding Mechanisms of Intrinsically Disordered Proteins: Theory, Simulation, and Experiment', *Frontiers in Molecular Biosciences*, 3, p. 52. Available at: https://doi.org/10.3389/fmolb.2016.00052.

Monazahian, M. *et al.* (1999) 'Low density lipoprotein receptor as a candidate receptor for hepatitis C virus', *Journal of Medical Virology*, 57(3), pp. 223–229. Available at: https://doi.org/10.1002/(SICI)1096-9071(199903)57:3%3C223::AID-JMV2%3E3.0.CO;2-4.

Morikawa, K. *et al.* (2011) 'Nonstructural protein 3-4A: the Swiss army knife of hepatitis C virus', *Journal of Viral Hepatitis*, 18(5), pp. 305–315. Available at: https://doi.org/10.1111/j.1365-2893.2011.01451.x.

Morin, T.J. *et al.* (2012) 'Human monoclonal antibody HCV1 effectively prevents and treats HCV infection in chimpanzees', *PLoS pathogens*, 8(8), p. e1002895. Available at: https://doi.org/10.1371/journal.ppat.1002895.

Moult, J. *et al.* (1995) 'A large-scale experiment to assess protein structure prediction methods', *Proteins: Structure, Function, and Bioinformatics*, 23(3), pp. ii–iv. Available at: https://doi.org/10.1002/prot.340230303.

Moult, J. *et al.* (2020) 'Critical assessment of techniques for protein structure prediction, fourteenth round.', pp. 22–24. Available at: https://www.predictioncenter.org/casp14/doc/CASP14_Abstracts.pdf.

Munro, J.B. *et al.* (2014) 'Conformational dynamics of single HIV-1 envelope trimers on the surface of native virions', *Science*, 346(6210), pp. 759–763. Available at: https://doi.org/10.1126/science.1254426.

Murakami, E. *et al.* (2007) 'Mechanism of activation of beta-D-2'-deoxy-2'-fluoro-2'-c-methylcytidine and inhibition of hepatitis C virus NS5B RNA polymerase', *Antimicrobial Agents and Chemotherapy*, 51(2), pp. 503–509. Available at: https://doi.org/10.1128/AAC.00400-06.

Murakami, E. *et al.* (2008) 'The Mechanism of Action of β-d-2′-Deoxy-2′-Fluoro-2′-C-Methylcytidine Involves a Second Metabolic Pathway Leading to β-d-2′-Deoxy-2′-Fluoro-2′-C-Methyluridine 5′-Triphosphate, a Potent Inhibitor of the Hepatitis C Virus RNA-Dependent RNA Polymerase', *Antimicrobial Agents and Chemotherapy*, 52(2), pp. 458–464. Available at: https://doi.org/10.1128/AAC.01184-07.

Murphy, D.G. *et al.* (2007) 'Use of Sequence Analysis of the NS5B Region for Routine Genotyping of Hepatitis C Virus with Reference to C/E1 and 5′ Untranslated Region Sequences', *Journal of Clinical Microbiology*, 45(4), pp. 1102–1112. Available at: https://doi.org/10.1128/jcm.02366-06.

Nahmias, Y. *et al.* (2008) 'Apolipoprotein B–dependent hepatitis C virus secretion is inhibited by the grapefruit flavonoid naringenin', *Hepatology*, 47(5), p. 1437. Available at: https://doi.org/10.1002/hep.22197.

Naldini, L. *et al.* (1996) 'In Vivo Gene Delivery and Stable Transduction of Nondividing Cells by a Lentiviral Vector', *Science*, 272(5259), pp. 263–267. Available at: https://doi.org/10.1126/science.272.5259.263.

Nascimbeni, M. *et al.* (2003) 'Kinetics of CD4+ and CD8+ memory T-cell responses during hepatitis C virus rechallenge of previously recovered chimpanzees', *Journal of Virology*, 77(8), pp. 4781–4793. Available at: https://doi.org/10.1128/jvi.77.8.4781-4793.2003.

NCBI (2025) *Hepacivirus hominis - Protein - NCBI*. Available at: https://www.ncbi.nlm.nih.gov/protein/?term=Hepacivirus+hominis (Accessed: 25 September 2025).

Neculai, D. *et al.* (2013) 'Structure of LIMP-2 provides functional insights with implications for SR-BI and CD36', *Nature*, 504(7478), pp. 172–176. Available at: https://doi.org/10.1038/nature12684.

Negro, F. *et al.* (2015) 'Extrahepatic Morbidity and Mortality of Chronic Hepatitis C', *Gastroenterology*, 149(6), pp. 1345–1360. Available at: https://doi.org/10.1053/j.gastro.2015.08.035.

Neumann, A.U. *et al.* (1998) 'Hepatitis C Viral Dynamics in Vivo and the Antiviral Efficacy of Interferon-α Therapy', *Science*, 282(5386), pp. 103–107. Available at: https://doi.org/10.1126/science.282.5386.103.

Niepmann, M. *et al.* (2018) 'Signals Involved in Regulation of Hepatitis C Virus RNA Genome Translation and Replication', *Frontiers in Microbiology*, 9, p. 395. Available at: https://doi.org/10.3389/fmicb.2018.00395.

Nissen, P. *et al.* (1996) 'The ternary complex of aminoacylated tRNA and EF-Tu-GTP. Recognition of a bond and a fold', *Biochimie*, 78(11), pp. 921–933. Available at: https://doi.org/10.1016/S0300-9084(97)86714-4.

Ogata, N. *et al.* (1991) 'Nucleotide sequence and mutation rate of the H strain of hepatitis C virus', *Proceedings of the National Academy of Sciences of the United States of America*, 88(8), pp. 3392–3396. Available at: https://doi.org/10.1073/pnas.88.8.3392.

Okamoto, H. *et al.* (1991) 'Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions', *The Journal of General Virology*, 72 (Pt 11), pp. 2697–2704. Available at: https://doi.org/10.1099/0022-1317-72-11-2697.

Olesen, C.H. *et al.* (2021) 'In vitro adaptation and characterization of attenuated hypervariable region 1 swap chimeras of hepatitis C virus', *PLoS Pathogens*, 17(7), p. e1009720. Available at: https://doi.org/10.1371/journal.ppat.1009720.

Olesen, C.H. *et al.* (2025) 'Prevalence of HCV HVR1 insertions and their role in antibody evasion', *Hepatology (Baltimore, Md.)*, 82(1), pp. 199–211. Available at: https://doi.org/10.1097/HEP.0000000000001114.

Oliver, M.R. *et al.* (2023) 'Structures of the Hepaci-, Pegi-, and Pestiviruses envelope proteins suggest a novel membrane fusion mechanism', *PLOS Biology*, 21(7), p. e3002174. Available at: https://doi.org/10.1371/journal.pbio.3002174.

Orland, J.R., Wright, T.L. and Cooper, S. (2001) 'Acute Hepatitis C', *Hepatology*, 33(2), p. 321. Available at: https://doi.org/10.1053/jhep.2001.22112.

Osburn, W.O. *et al.* (2010) 'Spontaneous control of primary hepatitis C virus infection and immunity against persistent reinfection', *Gastroenterology*, 138(1), pp. 315–324. Available at: https://doi.org/10.1053/j.gastro.2009.09.017.

Osburn, W.O. *et al.* (2014) 'Clearance of Hepatitis C infection is associated with early appearance of broad neutralizing antibody responses', *Hepatology (Baltimore, Md.)*, 59(6), pp. 2140–2151. Available at: https://doi.org/10.1002/hep.27013.

Owen, D.M. *et al.* (2009) 'Apolipoprotein E on hepatitis C virion facilitates infection through interaction with low-density lipoprotein receptor', *Virology*, 394(1), pp. 99–108. Available at: https://doi.org/10.1016/j.virol.2009.08.037.

Owsianka, A. *et al.* (2001) 'Functional analysis of hepatitis C virus E2 glycoproteins and virus-like particles reveals structural dissimilarities between different forms of E2', *The Journal of General Virology*, 82(Pt 8), pp. 1877–1883. Available at: https://doi.org/10.1099/0022-1317-82-8-1877.

Page, K. *et al.* (2009) 'Acute hepatitis C virus infection in young adult injection drug users: a prospective study of incident infection, resolution, and reinfection', *The Journal of Infectious Diseases*, 200(8), pp. 1216–1226. Available at: https://doi.org/10.1086/605947.

Pagès, H. *et al.* (2024) 'Biostrings: Efficient manipulation of biological strings'. Available at: https://bioconductor.org/packages/Biostrings.

Palor, M. *et al.* (2020) 'Cholesterol sensing by CD81 is important for hepatitis C virus entry', *The Journal of Biological Chemistry*, 295(50), pp. 16931–16948. Available at: https://doi.org/10.1074/jbc.RA120.014761.

Paraskevis, D. *et al.* (2021) 'Classification, Genetic Diversity and Global Distribution of Hepatitis C Virus (HCV) Genotypes and Subtypes', in A. Hatzakis (ed.) *Hepatitis C: Epidemiology, Prevention and Elimination : Volume 1*. Cham: Springer International Publishing, pp. 55–69. Available at: https://doi.org/10.1007/978-3-030-64649-3_3.

Park, H.E., Gruenke, J.A. and White, J.M. (2003) 'Leash in the groove mechanism of membrane fusion', *Nature Structural & Molecular Biology*, 10(12), pp. 1048–1053. Available at: https://doi.org/10.1038/nsb1012.

Patel, J., Patel, A.H. and McLauchlan, J. (2001) 'The Transmembrane Domain of the Hepatitis C Virus E2 Glycoprotein Is Required for Correct Folding of the E1 Glycoprotein and Native Complex Formation', *Virology*, 279(1), pp. 58–68. Available at: https://doi.org/10.1006/viro.2000.0693.

Paul, D. *et al.* (2013) 'Morphological and biochemical characterization of the membranous hepatitis C virus replication compartment', *Journal of Virology*, 87(19), pp. 10612–10627. Available at: https://doi.org/10.1128/JVI.01370-13.

Paul, D. *et al.* (2018) 'Glycine Zipper Motifs in Hepatitis C Virus Nonstructural Protein 4B Are Required for the Establishment of Viral Replication Organelles', *Journal of Virology*, 92(4), pp. e01890-17. Available at: https://doi.org/10.1128/JVI.01890-17.

Pawlotsky, J.-M. (2006) 'Hepatitis C Virus Population Dynamics During Infection', in E. Domingo (ed.) *Quasispecies: Concept and Implications for Virology*. Berlin, Heidelberg: Springer, pp. 261–284. Available at: https://doi.org/10.1007/3-540-26397-7_9.

Pawlotsky, J.-M. *et al.* (2020) 'EASL recommendations on treatment of hepatitis C: Final update of the series☆', *Journal of Hepatology*, 73(5), pp. 1170–1218. Available at: https://doi.org/10.1016/j.jhep.2020.08.018.

Penin, F. *et al.* (2001) 'Conservation of the conformation and positive charges of hepatitis C virus E2 envelope glycoprotein hypervariable region 1 points to a role in cell attachment', *Journal of Virology*, 75(12), pp. 5703–5710. Available at: https://doi.org/10.1128/JVI.75.12.5703-5710.2001.

Penin, F. *et al.* (2004) 'Structure and function of the membrane anchor domain of hepatitis C virus nonstructural protein 5A', *The Journal of Biological Chemistry*, 279(39), pp. 40835–40843. Available at: https://doi.org/10.1074/jbc.M404761200.

Pérard, J. *et al.* (2013) 'Structure of the full-length HCV IRES in solution', *Nature Communications*, 4, p. 1612. Available at: https://doi.org/10.1038/ncomms2611.

Pestka, J.M. *et al.* (2007) 'Rapid induction of virus-neutralizing antibodies and viral clearance in a single-source outbreak of hepatitis C', *Proceedings of the National Academy of Sciences of the United States of America*, 104(14), pp. 6025–6030. Available at: https://doi.org/10.1073/pnas.0607026104.

Petracca, R. *et al.* (2000) 'Structure-Function Analysis of Hepatitis C Virus Envelope-CD81 Binding', *Journal of Virology*, 74(10), pp. 4824–4830. Available at: https://doi.org/10.1128/jvi.74.10.4824-4830.2000.

Pettersen, E.F. *et al.* (2021) 'UCSF ChimeraX: Structure visualization for researchers, educators, and developers', *Protein Science: A Publication of the Protein Society*, 30(1), pp. 70–82. Available at: https://doi.org/10.1002/pro.3943.

Pierce, B.G. *et al.* (2016) 'Global mapping of antibody recognition of the hepatitis C virus E2 glycoprotein: Implications for vaccine design', *Proceedings of the National Academy of Sciences*, 113(45), pp. E6946–E6954. Available at: https://doi.org/10.1073/pnas.1614942113.

Pierce, B.G. *et al.* (2017) 'Structure-Based Design of Hepatitis C Virus Vaccines That Elicit Neutralizing Antibody Responses to a Conserved Epitope', *Journal of Virology*, 91(20), p. 10.1128/jvi.01032-17. Available at: https://doi.org/10.1128/jvi.01032-17.

Pierce, B.G. *et al.* (2020) 'Structure-Based Design of Hepatitis C Virus E2 Glycoprotein Improves Serum Binding and Cross-Neutralization', *Journal of Virology*, 94(22), pp. e00704-20. Available at: https://doi.org/10.1128/JVI.00704-20.

Pileri, P. *et al.* (1998) 'Binding of hepatitis C virus to CD81', *Science (New York, N.Y.)*, 282(5390), pp. 938–941. Available at: https://doi.org/10.1126/science.282.5390.938.

Ploss, A. *et al.* (2009) 'Human occludin is a hepatitis C virus entry factor required for infection of mouse cells', *Nature*, 457(7231), pp. 882–886. Available at: https://doi.org/10.1038/nature07684.

Poordad, F. *et al.* (2011) 'Boceprevir for untreated chronic HCV genotype 1 infection', *The New England Journal of Medicine*, 364(13), pp. 1195–1206. Available at: https://doi.org/10.1056/NEJMoa1010494.

Popescu, C.-I. *et al.* (2011) 'NS2 Protein of Hepatitis C Virus Interacts with Structural and Non-Structural Proteins towards Virus Assembly', *PLOS Pathogens*, 7(2), p. e1001278. Available at: https://doi.org/10.1371/journal.ppat.1001278.

Posit team (2025) 'RStudio: Integrated Development Environment for R'. Boston, MA: Posit Software, PBC. Available at: http://www.posit.co/.

Potter, J.A. *et al.* (2012) 'Toward a Hepatitis C Virus Vaccine: the Structural Basis of Hepatitis C Virus Neutralization by AP33, a Broadly Neutralizing Antibody', *Journal of Virology*, 86(23), pp. 12923–12932. Available at: https://doi.org/10.1128/jvi.02052-12.

Prentoe, J. *et al.* (2011) 'Hypervariable Region 1 Differentially Impacts Viability of Hepatitis C Virus Strains of Genotypes 1 to 6 and Impairs Virus Neutralization', *Journal of Virology*, 85(5), pp. 2224–2234. Available at: https://doi.org/10.1128/JVI.01594-10.

Prentoe, J. *et al.* (2014) 'Hypervariable region 1 deletion and required adaptive envelope mutations confer decreased dependency on scavenger receptor class B type I and low-density lipoprotein receptor for hepatitis C virus', *Journal of Virology*, 88(3), pp. 1725–1739. Available at: https://doi.org/10.1128/JVI.02017-13.

Prentoe, J. *et al.* (2016) 'Hypervariable region 1 shielding of hepatitis C virus is a main contributor to genotypic differences in neutralization sensitivity', *Hepatology (Baltimore, Md.)*, 64(6), pp. 1881–1892. Available at: https://doi.org/10.1002/hep.28705.

Prentoe, J. *et al.* (2019) 'Hypervariable region 1 and N-linked glycans of hepatitis C regulate virion neutralization by modulating envelope conformations', *Proceedings of the National Academy of Sciences of the United States of America*, 116(20), pp. 10039–10047. Available at: https://doi.org/10.1073/pnas.1822002116.

Prince, A.M. *et al.* (2005) 'Protection against chronic hepatitis C virus infection after rechallenge with homologous, but not heterologous, genotypes in a chimpanzee model', *The Journal of Infectious Diseases*, 192(10), pp. 1701–1709. Available at: https://doi.org/10.1086/496889.

Prokunina-Olsson, L. *et al.* (2013) 'A variant upstream of IFNL3 (IL28B) creating a new interferon gene IFNL4 is associated with impaired clearance of hepatitis C virus', *Nature Genetics*, 45(2), pp. 164–171. Available at: https://doi.org/10.1038/ng.2521.

Public Health Scotland (2024) *Dispenser payments and prescription cost analysis - Financial year 2023 to 2024 - Dispenser payments and prescription cost analysis - Publications - Public Health Scotland*. Available at: https://publichealthscotland.scot/publications/dispenser-payments-and-

prescription-cost-analysis/dispenser-payments-and-prescription-cost-analysis-financial-year-2023-to-2024/ (Accessed: 24 August 2025).

Qiu, D. *et al.* (2011) 'The effects of NS5A inhibitors on NS5A phosphorylation, polyprotein processing and localization', *Journal of General Virology*, 92(11), pp. 2502–2511. Available at: https://doi.org/10.1099/vir.0.034801-0.

Quade, N. *et al.* (2015) 'Cryo-EM structure of Hepatitis C virus IRES bound to the human ribosome at 3.9-Å resolution', *Nature Communications*, 6, p. 7646. Available at: https://doi.org/10.1038/ncomms8646.

Radić, L. *et al.* (2025) 'Bispecific antibodies against the hepatitis C virus E1E2 envelope glycoprotein', *Proceedings of the National Academy of Sciences*, 122(15), p. e2420402122. Available at: https://doi.org/10.1073/pnas.2420402122.

Raina, M. and Ibba, M. (2014) 'tRNAs as regulators of biological processes', *Frontiers in Genetics*, 5, p. 171. Available at: https://doi.org/10.3389/fgene.2014.00171.

Ramirez, S. *et al.* (2009) 'Hepatitis C Virus Compartmentalization and Infection Recurrence after Liver Transplantation', *American Journal of Transplantation*, 9(7), pp. 1591–1601. Available at: https://doi.org/10.1111/j.1600-6143.2009.02666.x.

Rathkopf, C. (2025) 'Hallucination, reliability, and the role of generative AI in science'. arXiv. Available at: https://doi.org/10.48550/arXiv.2504.08526.

Ray, R. *et al.* (2010) 'Characterization of antibodies induced by vaccination with Hepatitis C virus envelope glycoproteins', *The Journal of Infectious Diseases*, 202(6), pp. 862–866. Available at: https://doi.org/10.1086/655902.

Ray, S.C. *et al.* (1999) 'Acute hepatitis C virus structural gene sequences as predictors of persistent viremia: hypervariable region 1 as a decoy', *Journal of Virology*, 73(4), pp. 2938–2946. Available at: https://doi.org/10.1128/JVI.73.4.2938-2946.1999.

RCSB Protein Data Bank (2025) *RCSB PDB: Homepage*. Available at: https://www.rcsb.org/ (Accessed: 25 September 2025).

Remmert, M. *et al.* (2012) 'HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment', *Nature Methods*, 9(2), pp. 173–175. Available at: https://doi.org/10.1038/nmeth.1818.

Ren, F. *et al.* (2023) 'AlphaFold accelerates artificial intelligence powered drug discovery: efficient discovery of a novel CDK20 small molecule inhibitor', *Chemical Science*, 14(6), pp. 1443–1452. Available at: https://doi.org/10.1039/d2sc05709c.

Rey, F.A. and Lok, S.-M. (2018) 'Common Features of Enveloped Viruses and Implications for Immunogen Design for Next-Generation Vaccines', *Cell*, 172(6), pp. 1319–1334. Available at: https://doi.org/10.1016/j.cell.2018.02.054.

Rhainds, D. and Brissette, L. (2004) 'The role of scavenger receptor class B type I (SR-BI) in lipid trafficking: Defining the rules for lipid traders', *The International Journal of Biochemistry & Cell Biology*, 36(1), pp. 39–77. Available at: https://doi.org/10.1016/S1357-2725(03)00173-0.

Riva, L. and Dubuisson, J. (2019) 'Similarities and Differences Between HCV Pseudoparticle (HCVpp) and Cell Culture HCV (HCVcc) in the Study of HCV', in M. Law (ed.) *Hepatitis C Virus Protocols*. New York, NY: Springer, pp. 33–45. Available at: https://doi.org/10.1007/978-1-4939-8976-8_2.

Rizatdinova, S.N., Ershova, A.E. and Astrakhantseva, I.V. (2025) 'Pseudotyped Viruses: A Useful Platform for Pre-Clinical Studies Conducted in a BSL-2 Laboratory Setting', *Biomolecules*, 15(1), p. 135. Available at: https://doi.org/10.3390/biom15010135.

Roccasecca, R. *et al.* (2003) 'Binding of the hepatitis C virus E2 glycoprotein to CD81 is strain specific and is modulated by a complex interplay between hypervariable regions 1 and 2', *Journal of Virology*, 77(3), pp. 1856–1867. Available at: https://doi.org/10.1128/jvi.77.3.1856-1867.2003.

Roche, S. *et al.* (2006) 'Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G', *Science (New York, N.Y.)*, 313(5784), pp. 187–191. Available at: https://doi.org/10.1126/science.1127683.

Roche, S. *et al.* (2007) 'Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G', *Science (New York, N.Y.)*, 315(5813), pp. 843–848. Available at: https://doi.org/10.1126/science.1135710.

Roche, S. *et al.* (2008) 'Structures of vesicular stomatitis virus glycoprotein: membrane fusion revisited', *Cellular and molecular life sciences: CMLS*, 65(11), pp. 1716–1728. Available at: https://doi.org/10.1007/s00018-008-7534-3.

Roger, S. *et al.* (2021) 'HCV virology and diagnosis', *Clinics and Research in Hepatology and Gastroenterology*, 45(3), p. 101626. Available at: https://doi.org/10.1016/j.clinre.2021.101626.

Roguin, L.P. and Retegui, L.A. (2003) 'Monoclonal Antibodies Inducing Conformational Changes on the Antigen Molecule', *Scandinavian Journal of Immunology*, 58(4), pp. 387–394. Available at: https://doi.org/10.1046/j.1365-3083.2003.01320.x.

Romero-Brey, I. *et al.* (2012) 'Three-Dimensional Architecture and Biogenesis of Membrane Structures Associated with Hepatitis C Virus Replication', *PLOS Pathogens*, 8(12), p. e1003056. Available at: https://doi.org/10.1371/journal.ppat.1003056.

Romero-Brey, I. *et al.* (2015) 'NS5A Domain 1 and Polyprotein Cleavage Kinetics Are Critical for Induction of Double-Membrane Vesicles Associated with Hepatitis C Virus Replication', *mBio*, 6(4), p. e00759. Available at: https://doi.org/10.1128/mBio.00759-15.

Roth, W.K. *et al.* (2012) 'International survey on NAT testing of blood donations: expanding implementation and yield from 1999 to 2009', *Vox Sanguinis*, 102(1), pp. 82–90. Available at: https://doi.org/10.1111/j.1423-0410.2011.01506.x.

Roy, R., Hohng, S. and Ha, T. (2008) 'A practical guide to single-molecule FRET', *Nature Methods*, 5(6), pp. 507–516. Available at: https://doi.org/10.1038/nmeth.1208.

Rutebemberwa, A. *et al.* (2008) 'High-programmed death-1 levels on hepatitis C virus-specific T cells during acute infection are associated with viral persistence and require preservation of cognate antigen during chronic infection', *Journal of Immunology (Baltimore, Md.: 1950)*, 181(12), pp. 8215–8225. Available at: https://doi.org/10.4049/jimmunol.181.12.8215.

Sabo, M.C. *et al.* (2011) 'Neutralizing Monoclonal Antibodies against Hepatitis C Virus E2 Protein Bind Discontinuous Epitopes and Inhibit Infection at a Postattachment Step ▿', *Journal of Virology*, 85(14), pp. 7005–7019. Available at: https://doi.org/10.1128/JVI.00586-11.

Sacks-Davis, R. *et al.* (2015) 'Hepatitis C Virus Reinfection and Spontaneous Clearance of Reinfection--the InC3 Study', *The Journal of Infectious Diseases*, 212(9), pp. 1407–1419. Available at: https://doi.org/10.1093/infdis/jiv220.

Sainz, B. *et al.* (2012) 'Identification of the Niemann-Pick C1–like 1 cholesterol absorption receptor as a new hepatitis C virus entry factor', *Nature Medicine*, 18(2), pp. 281–285. Available at: https://doi.org/10.1038/nm.2581.

Saito, T. *et al.* (1997) 'Plasmid DNA-based immunization for hepatitis C virus structural proteins: immune responses in mice', *Gastroenterology*, 112(4), pp. 1321–1330. Available at: https://doi.org/10.1016/s0016-5085(97)70146-x.

Sanders, R.W. and Moore, J.P. (2021) 'Virus vaccines: proteins prefer prolines', *Cell Host & Microbe*, 29(3), pp. 327–333. Available at: https://doi.org/10.1016/j.chom.2021.02.002.

Sandres-Sauné, K. *et al.* (2003) 'Determining hepatitis C genotype by analyzing the sequence of the NS5b region', *Journal of Virological Methods*, 109(2), pp. 187–193. Available at: https://doi.org/10.1016/S0166-0934(03)00070-3.

Sandrin, V. *et al.* (2005) 'Assembly of functional hepatitis C virus glycoproteins on infectious pseudoparticles occurs intracellularly and requires concomitant incorporation of E1 and E2 glycoproteins', *Journal of General Virology*, 86(12), pp. 3189–3199. Available at: https://doi.org/10.1099/vir.0.81428-0.

Sarrazin, C. *et al.* (2015) 'Prevalence of the hepatitis C virus NS3 polymorphism Q80K in genotype 1 patients in the European region', *Antiviral Research*, 116, pp. 10–16. Available at: https://doi.org/10.1016/j.antiviral.2015.01.003.

Sasmal, D.K. *et al.* (2016) 'Single-Molecule Fluorescence Resonance Energy Transfer in Molecular Biology', *Nanoscale*, 8(48), pp. 19928–19944. Available at: https://doi.org/10.1039/c6nr06794h.

Sautto, G. *et al.* (2013) 'Structural and Antigenic Definition of Hepatitis C Virus E2 Glycoprotein Epitopes Targeted by Monoclonal Antibodies', *Clinical and Developmental Immunology*, 2013, p. 450963. Available at: https://doi.org/10.1155/2013/450963.

Scarselli, E. *et al.* (1995) 'Occurrence of antibodies reactive with more than one variant of the putative envelope glycoprotein (gp70) hypervariable region 1 in viremic hepatitis C virus-infected patients', *Journal of Virology*, 69(7), pp. 4407–4412. Available at: https://doi.org/10.1128/JVI.69.7.4407-4412.1995.

Scarselli, E. *et al.* (2002) 'The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus', *The EMBO Journal*, 21(19), pp. 5017–5025. Available at: https://doi.org/10.1093/emboj/cdf529.

Schechter, I. and Berger, A. (1967) 'On the size of the active site in proteases. I. Papain', *Biochemical and Biophysical Research Communications*, 27(2), pp. 157–162. Available at: https://doi.org/10.1016/S0006-291X(67)80055-X.

Schechter, I. and Berger, A. (1968) 'On the active site of proteases. III. Mapping the active site of papain; specific peptide inhibitors of papain', *Biochemical and Biophysical Research Communications*, 32(5), pp. 898–902. Available at: https://doi.org/10.1016/0006-291X(68)90326-4.

Scheel, T.K.H. *et al.* (2008) 'Development of JFH1-based cell culture systems for hepatitis C virus genotype 4a and evidence for cross-genotype neutralization', *Proceedings of the National Academy of Sciences*, 105(3), pp. 997–1002. Available at: https://doi.org/10.1073/pnas.0711044105.

Schiering, N. *et al.* (2011) 'A macrocyclic HCV NS3/4A protease inhibitor interacts with protease and helicase residues in the complex with its full-length target', *Proceedings of the National Academy of Sciences of the United States of America*, 108(52), pp. 21052–21056. Available at: https://doi.org/10.1073/pnas.1110534108.

Schmidt, H.B., Barreau, A. and Rohatgi, R. (2019) 'Phase separation-deficient TDP43 remains functional in splicing', *Nature Communications*, 10(1), p. 4890. Available at: https://doi.org/10.1038/s41467-019-12740-2.

Schmidt-Martin, D. *et al.* (2015) 'Intensive temporal mapping of hepatitis C hypervariable region 1 quasispecies provides novel insights into hepatitis C virus

evolution in chronic infection', *The Journal of General Virology*, 96(8), pp. 2145–2156. Available at: https://doi.org/10.1099/vir.0.000149.

Schneider, T.D. and Stephens, R.M. (1990) 'Sequence logos: a new way to display consensus sequences.', *Nucleic Acids Research*, 18(20), pp. 6097–6100. Available at: https://doi.org/10.1093/nar/18.20.6097.

Schuler, B. (2013) 'Single-molecule FRET of protein structure and dynamics - a primer', *Journal of Nanobiotechnology*, 11(Suppl 1), p. S2. Available at: https://doi.org/10.1186/1477-3155-11-S1-S2.

Schult, P. *et al.* (2018) 'microRNA-122 amplifies hepatitis C virus translation by shaping the structure of the internal ribosomal entry site', *Nature Communications*, 9(1), p. 2613. Available at: https://doi.org/10.1038/s41467-018-05053-3.

Schulze Zur Wiesch, J. *et al.* (2012) 'Broadly directed virus-specific CD4+ T cell responses are primed during acute hepatitis C infection, but rapidly disappear from human blood with viral persistence', *The Journal of Experimental Medicine*, 209(1), pp. 61–75. Available at: https://doi.org/10.1084/jem.20100388.

Scott, J.D. *et al.* (2006) 'High rate of spontaneous negativity for hepatitis C virus RNA after establishment of chronic infection in Alaska Natives', *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 42(7), pp. 945–952. Available at: https://doi.org/10.1086/500938.

Sedano, C.D. and Sarnow, P. (2014) 'Hepatitis C virus subverts liver-specific miR-122 to protect the viral genome from exoribonuclease Xrn2', *Cell Host & Microbe*, 16(2), pp. 257–264. Available at: https://doi.org/10.1016/j.chom.2014.07.006.

Seitz, R. (2016) 'Human Immunodeficiency Virus (HIV)', *Transfusion Medicine and Hemotherapy*, 43(3), pp. 203–222. Available at: https://doi.org/10.1159/000445852.

Senior, A.W. *et al.* (2020) 'Improved protein structure prediction using potentials from deep learning', *Nature*, 577(7792), pp. 706–710. Available at: https://doi.org/10.1038/s41586-019-1923-7.

Sesmero, E. and Thorpe, I.F. (2015) 'Using the Hepatitis C Virus RNA-Dependent RNA Polymerase as a Model to Understand Viral Polymerase Structure, Function and Dynamics', *Viruses*, 7(7), pp. 3974–3994. Available at: https://doi.org/10.3390/v7072808.

Sharma, N.R. *et al.* (2011) 'Hepatitis C virus is primed by CD81 protein for low pH-dependent fusion', *The Journal of Biological Chemistry*, 286(35), pp. 30361–30376. Available at: https://doi.org/10.1074/jbc.M111.263350.

Shata, M.T. *et al.* (2003) 'Exposure to low infective doses of HCV induces cellular immune responses without consistently detectable viremia or seroconversion in chimpanzees', *Virology*, 314(2), pp. 601–616. Available at: https://doi.org/10.1016/s0042-6822(03)00461-6.

Shen, W. *et al.* (2016) 'SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation', *PLOS ONE*, 11(10), p. e0163962. Available at: https://doi.org/10.1371/journal.pone.0163962.

Shimizu, Y.K. *et al.* (1990) 'Early events in hepatitis C virus infection of chimpanzees', *Proceedings of the National Academy of Sciences of the United States of America*, 87(16), pp. 6441–6444. Available at: https://doi.org/10.1073/pnas.87.16.6441.

Shoukry, N.H. *et al.* (2003) 'Memory CD8+ T cells are required for protection from persistent hepatitis C virus infection', *The Journal of Experimental Medicine*, 197(12), pp. 1645–1655. Available at: https://doi.org/10.1084/jem.20030239.

da Silva Cardoso, M. *et al.* (1995) 'The serology of hepatitis C virus (HCV) infection: antibody crossreaction in the hypervariable region 1', *Archives of Virology*, 140(10), pp. 1705–1713. Available at: https://doi.org/10.1007/BF01384335.

Singer, J.B. *et al.* (2018) 'GLUE: a flexible software system for virus sequence data', *BMC Bioinformatics*, 19(1), p. 532. Available at: https://doi.org/10.1186/s12859-018-2459-9.

Sliepen, K. *et al.* (2022) 'Induction of cross-neutralizing antibodies by a permuted hepatitis C virus glycoprotein nanoparticle vaccine candidate', *Nature Communications*, 13(1), p. 7271. Available at: https://doi.org/10.1038/s41467-022-34961-8.

Sofia, M.J. *et al.* (2010) 'Discovery of a ß-d-2′-Deoxy-2′-α-fluoro-2′-ß-C-methyluridine Nucleotide Prodrug (PSI-7977) for the Treatment of Hepatitis C Virus', *Journal of Medicinal Chemistry*, 53(19), pp. 7202–7218. Available at: https://doi.org/10.1021/jm100863x.

Sourisseau, M. *et al.* (2013) 'Temporal Analysis of Hepatitis C Virus Cell Entry with Occludin Directed Blocking Antibodies', *PLOS Pathogens*, 9(3), p. e1003244. Available at: https://doi.org/10.1371/journal.ppat.1003244.

Stapleford, K.A. and Lindenbach, B.D. (2011) 'Hepatitis C Virus NS2 Coordinates Virus Particle Assembly through Physical Interactions with the E1-E2 Glycoprotein and NS3-NS4A Enzyme Complexes', *Journal of Virology*, 85(4), pp. 1706–1717. Available at: https://doi.org/10.1128/JVI.02268-10.

Steinegger, M. and Söding, J. (2017) 'MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets', *Nature Biotechnology*, 35(11), pp. 1026–1028. Available at: https://doi.org/10.1038/nbt.3988.

Stejskal, L., Kalemera, M.D., *et al.* (2020) 'An Entropic Safety Catch Controls Hepatitis C Virus Entry and Antibody Resistance'. bioRxiv, p. 2020.11.11.377218. Available at: https://doi.org/10.1101/2020.11.11.377218.

Stejskal, L., Lees, W.D., *et al.* (2020) 'Flexibility and intrinsic disorder are conserved features of hepatitis C virus E2 glycoprotein', *PLoS computational biology*, 16(2), p. e1007710. Available at: https://doi.org/10.1371/journal.pcbi.1007710.

Stejskal, L. *et al.* (2022) 'An entropic safety catch controls hepatitis C virus entry and antibody resistance', *eLife*, 11, p. e71854. Available at: https://doi.org/10.7554/eLife.71854.

Stoeck, I.K. *et al.* (2017) 'Hepatitis C Virus Replication Depends on Endosomal Cholesterol Homeostasis', *Journal of Virology*, 92(1), pp. e01196-17. Available at: https://doi.org/10.1128/JVI.01196-17.

Stroffolini, T. and Stroffolini, G. (2024) 'Prevalence and Modes of Transmission of Hepatitis C Virus Infection: A Historical Worldwide Review', *Viruses*, 16(7), p. 1115. Available at: https://doi.org/10.3390/v16071115.

Stross, C. *et al.* (2016) 'Natural HCV variants with increased replicative fitness due to NS3 helicase mutations in the C-terminal helix α18', *Scientific Reports*, 6, p. 19526. Available at: https://doi.org/10.1038/srep19526.

Stuyver, L.J. *et al.* (2006) 'Inhibition of hepatitis C replicon RNA synthesis by beta-D-2'-deoxy-2'-fluoro-2'-C-methylcytidine: a specific inhibitor of hepatitis C virus replication', *Antiviral Chemistry & Chemotherapy*, 17(2), pp. 79–87. Available at: https://doi.org/10.1177/095632020601700203.

Subramaniam, A.R., Zid, B.M. and O'Shea, E.K. (2014) 'An Integrated Approach Reveals Regulatory Controls on Bacterial Translation Elongation', *Cell*, 159(5), pp. 1200–1211. Available at: https://doi.org/10.1016/j.cell.2014.10.043.

Sun, H. *et al.* (2024) 'Two antibodies show broad, synergistic neutralization against SARS-CoV-2 variants by inducing conformational change within the RBD', *Protein & Cell*, 15(2), pp. 121–134. Available at: https://doi.org/10.1093/procel/pwad040.

Swadling, L. *et al.* (2014) 'A human vaccine strategy based on chimpanzee adenoviral and MVA vectors that primes, boosts, and sustains functional HCV-specific T cell memory', *Science Translational Medicine*, 6(261), p. 261ra153. Available at: https://doi.org/10.1126/scitranslmed.3009185.

Tabata, K., Neufeldt, C.J. and Bartenschlager, R. (2020) 'Hepatitis C Virus Replication', *Cold Spring Harbor Perspectives in Medicine*, 10(3), p. a037093. Available at: https://doi.org/10.1101/cshperspect.a037093.

Takizawa, Y. *et al.* (2017) 'While the revolution will not be crystallized, biochemistry reigns supreme', *Protein Science*, 26(1), pp. 69–81. Available at: https://doi.org/10.1002/pro.3054.

Tanaka, T. *et al.* (1995) 'A novel sequence found at the 3' terminus of hepatitis C virus genome', *Biochemical and Biophysical Research Communications*, 215(2), pp. 744–749. Available at: https://doi.org/10.1006/bbrc.1995.2526.

Tanaka, Y. *et al.* (2009) 'Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C', *Nature Genetics*, 41(10), pp. 1105–1109. Available at: https://doi.org/10.1038/ng.449.

Targett-Adams, P. *et al.* (2011) 'Small molecules targeting hepatitis C virus-encoded NS5A cause subcellular redistribution of their target: insights into compound modes of action', *Journal of Virology*, 85(13), pp. 6353–6368. Available at: https://doi.org/10.1128/JVI.00215-11.

Tasan, I. and Zhao, H. (2017) 'Targeting Specificity of the CRISPR/Cas9 System', *ACS Synthetic Biology*, 6(9), pp. 1609–1613. Available at: https://doi.org/10.1021/acssynbio.7b00270.

Tellinghuisen, T.L. *et al.* (2008) 'Identification of Residues Required for RNA Replication in Domains II and III of the Hepatitis C Virus NS5A Protein', *Journal of Virology*, 82(3), pp. 1073–1083. Available at: https://doi.org/10.1128/JVI.00328-07.

Terrault, N.A. (2002) 'Sexual activity as a risk factor for hepatitis C', *Hepatology (Baltimore, Md.)*, 36(5 Suppl 1), pp. S99-105. Available at: https://doi.org/10.1053/jhep.2002.36797.

Terrault, N.A. *et al.* (2013) 'Sexual transmission of hepatitis C virus among monogamous heterosexual couples: The HCV partners study', *Hepatology*, 57(3), pp. 881–889. Available at: https://doi.org/10.1002/hep.26164.

Terrault, N.A. *et al.* (2021) 'Viral hepatitis and pregnancy', *Nature Reviews Gastroenterology & Hepatology*, 18(2), pp. 117–130. Available at: https://doi.org/10.1038/s41575-020-00361-w.

The CDA Foundation (2025) *HCV Global Distribution*. Available at: https://cdafound.org/polaris/global-distribution/ (Accessed: 18 August 2025).

The Royal Swedish Academy of Sciences (2024) *Press release: The Nobel Prize in Chemistry 2024*, *NobelPrize.org*. Available at: https://www.nobelprize.org/prizes/chemistry/2024/press-release/ (Accessed: 25 September 2025).

Thi, V.L.D. *et al.* (2012) 'Characterization of Hepatitis C Virus Particle Subpopulations Reveals Multiple Usage of the Scavenger Receptor BI for Entry Steps *', *Journal of Biological Chemistry*, 287(37), pp. 31242–31257. Available at: https://doi.org/10.1074/jbc.M112.365924.

Thursz, M. and Fontanet, A. (2014) 'HCV transmission in industrialized countries and resource-constrained areas', *Nature Reviews Gastroenterology & Hepatology*, 11(1), pp. 28–35. Available at: https://doi.org/10.1038/nrgastro.2013.179.

Timofeev, V. and Samygina, V. (2023) 'Protein Crystallography: Achievements and Challenges', *Crystals*, 13(1), p. 71. Available at: https://doi.org/10.3390/cryst13010071.

Tong, X. *et al.* (2014) 'In vivo emergence of a novel mutant L159F/L320F in the NS5B polymerase confers low-level resistance to the HCV polymerase inhibitors mericitabine and sofosbuvir', *The Journal of Infectious Diseases*, 209(5), pp. 668–675. Available at: https://doi.org/10.1093/infdis/jit562.

Tong, Y. *et al.* (2017) 'Functional Analysis of Hepatitis C Virus (HCV) Envelope Protein E1 Using a trans-Complementation System Reveals a Dual Role of a Putative Fusion Peptide of E1 in both HCV Entry and Morphogenesis', *Journal of Virology*, 91(7), pp. e02468-16. Available at: https://doi.org/10.1128/JVI.02468-16.

Torre, P. *et al.* (2024) 'Elimination of HCV Infection: Recent Epidemiological Findings, Barriers, and Strategies for the Coming Years', *Viruses*, 16(11), p. 1792. Available at: https://doi.org/10.3390/v16111792.

Torrents de la Peña, A. *et al.* (2022) 'Structure of the hepatitis C virus E1E2 glycoprotein complex', *Science*, 378(6617), pp. 263–269. Available at: https://doi.org/10.1126/science.abn9884.

Trickey, A. *et al.* (2019) 'The contribution of injection drug use to hepatitis C virus transmission globally, regionally, and at country level: a modelling study', *The Lancet. Gastroenterology & Hepatology*, 4(6), pp. 435–444. Available at: https://doi.org/10.1016/S2468-1253(19)30085-8.

Tscherne, D.M. *et al.* (2006) 'Time- and Temperature-Dependent Activation of Hepatitis C Virus for Low-pH-Triggered Entry', *Journal of Virology*, 80(4), pp. 1734–1741. Available at: https://doi.org/10.1128/JVI.80.4.1734-1741.2006.

Tzarum, N., Wilson, I.A. and Law, M. (2018) 'The Neutralizing Face of Hepatitis C Virus E2 Envelope Glycoprotein', *Frontiers in Immunology*, 9, p. 1315. Available at: https://doi.org/10.3389/fimmu.2018.01315.

US Preventive Services Task Force (2020) 'Screening for Hepatitis C Virus Infection in Adolescents and Adults: US Preventive Services Task Force

Recommendation Statement', *JAMA*, 323(10), pp. 970–975. Available at: https://doi.org/10.1001/jama.2020.1123.

Varadi, M. *et al.* (2022) 'AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models', *Nucleic Acids Research*, 50(D1), pp. D439–D444. Available at: https://doi.org/10.1093/nar/gkab1061.

Velázquez-Moctezuma, R. *et al.* (2019) 'Hepatitis C Virus–Escape Studies for Human Monoclonal Antibody AR4A Reveal Isolate-Specific Resistance and a High Barrier to Resistance', *The Journal of Infectious Diseases*, 219(1), pp. 68–79. Available at: https://doi.org/10.1093/infdis/jiy481.

Veldt, B.J. *et al.* (2007) 'Sustained virologic response and clinical outcomes in patients with chronic hepatitis C and advanced fibrosis', *Annals of Internal Medicine*, 147(10), pp. 677–684. Available at: https://doi.org/10.7326/0003-4819-147-10-200711200-00003.

Vieyres, G. *et al.* (2010) 'Characterization of the Envelope Glycoproteins Associated with Infectious Hepatitis C Virus', *Journal of Virology*, 84(19), pp. 10159–10168. Available at: https://doi.org/10.1128/JVI.01180-10.

Villalaín, J. (2025) 'Membrane fusion by dengue virus: The first step', *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1867(1), p. 184400. Available at: https://doi.org/10.1016/j.bbamem.2024.184400.

Wagih, O. (2017) 'ggseqlogo: a versatile R package for drawing sequence logos', *Bioinformatics*, 33(22), pp. 3645–3647. Available at: https://doi.org/10.1093/bioinformatics/btx469.

Wahid, A. *et al.* (2013) 'Disulfide Bonds in Hepatitis C Virus Glycoprotein E1 Control the Assembly and Entry Functions of E2 Glycoprotein', *Journal of Virology*, 87(3), pp. 1605–1617. Available at: https://doi.org/10.1128/JVI.02659-12.

Wakita, T. *et al.* (2005) 'Production of infectious hepatitis C virus in tissue culture from a cloned viral genome', *Nature medicine*, 11(7), pp. 791–796. Available at: https://doi.org/10.1038/nm1268.

Wang, H. *et al.* (2014) 'Oxysterol-binding protein is a phosphatidylinositol 4-kinase effector required for HCV replication membrane integrity and cholesterol trafficking', *Gastroenterology*, 146(5), pp. 1373-1385.e1–11. Available at: https://doi.org/10.1053/j.gastro.2014.02.002.

Wang, S. *et al.* (2022) 'CavitySpace: A Database of Potential Ligand Binding Sites in the Human Proteome', *Biomolecules*, 12(7), p. 967. Available at: https://doi.org/10.3390/biom12070967.

Wang, W. *et al.* (2014) 'Alanine scanning mutagenesis of hepatitis C virus E2 cysteine residues: Insights into E2 biogenesis and antigenicity', *Virology*, 448, pp. 229–237. Available at: https://doi.org/10.1016/j.virol.2013.10.020.

Ward, S. *et al.* (2002) 'Cellular immune responses against hepatitis C virus: the evidence base 2002', *Clinical and Experimental Immunology*, 128(2), pp. 195–203. Available at: https://doi.org/10.1046/j.1365-2249.2002.01840.x.

Wasilewski, L.N., Ray, S.C. and Bailey, J.R. (2016) 'Hepatitis C virus resistance to broadly neutralizing antibodies measured using replication-competent virus and pseudoparticles', *Journal of General Virology*, 97(11), pp. 2883–2893. Available at: https://doi.org/10.1099/jgv.0.000608.

Weiner, A.J. *et al.* (1991) 'Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins', *Virology*, 180(2), pp. 842–848. Available at: https://doi.org/10.1016/0042-6822(91)90104-j.

Weng, Y. *et al.* (2022) 'Identification of Potential WSB1 Inhibitors by AlphaFold Modeling, Virtual Screening, and Molecular Dynamics Simulation Studies', *Evidence-Based Complementary and Alternative Medicine: eCAM*, 2022, p. 4629392. Available at: https://doi.org/10.1155/2022/4629392.

White, J.M. *et al.* (2023) 'Viral Membrane Fusion: A Dance Between Proteins and Lipids', *Annual review of virology*, 10(1), pp. 139–161. Available at: https://doi.org/10.1146/annurev-virology-111821-093413.

Wickham, H. (2016) 'ggplot2: Elegant Graphics for Data Analysis'. Springer-Verlag New York. Available at: https://ggplot2.tidyverse.org.

Wohlwend, J. *et al.* (2025) 'Boltz-1 Democratizing Biomolecular Interaction Modeling'. bioRxiv, p. 2024.11.19.624167. Available at: https://doi.org/10.1101/2024.11.19.624167.

Wong, K.A. *et al.* (2013) 'Characterization of Hepatitis C virus resistance from a multiple-dose clinical trial of the novel NS5A inhibitor GS-5885', *Antimicrobial Agents and Chemotherapy*, 57(12), pp. 6333–6340. Available at: https://doi.org/10.1128/AAC.02193-12.

World Health Organisation (2017) *Current status on blood safety and availability in the WHO African Region — report of the 2013 survey*. WHO. Regional Office for Africa. Available at: https://iris.who.int/handle/10665/254656 (Accessed: 19 August 2025).

World Health Organisation (2022) *Updated recommendations on simplified service delivery and diagnostics for hepatitis C infection*. Available at: https://www.who.int/publications-detail-redirect/9789240052697 (Accessed: 28 June 2022).

World Health Organisation (2025) *Hepatitis C - Fact Sheet*. Available at: https://www.who.int/news-room/fact-sheets/detail/hepatitis-c (Accessed: 18 August 2025).

Wozniak, A.L. *et al.* (2010) 'Intracellular Proton Conductance of the Hepatitis C Virus p7 Protein and Its Contribution to Infectious Virus Production', *PLOS Pathogens*, 6(9), p. e1001087. Available at: https://doi.org/10.1371/journal.ppat.1001087.

Wrapp, D. *et al.* (2020) 'Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation', *Science (New York, N.Y.)*, 367(6483), pp. 1260–1263. Available at: https://doi.org/10.1126/science.abb2507.

Wu, X., Kriz, A.J. and Sharp, P.A. (2014) 'Target specificity of the CRISPR-Cas9 system', *Quantitative biology*, 2(2), pp. 59–70. Available at: https://doi.org/10.1007/s40484-014-0030-x.

Xie, Q. *et al.* (2020) 'Self-training with Noisy Student improves ImageNet classification'. arXiv. Available at: https://doi.org/10.48550/arXiv.1911.04252.

Xu, R. and Wilson, I.A. (2011) 'Structural Characterization of an Early Fusion Intermediate of Influenza Virus Hemagglutinin', *Journal of Virology*, 85(10), pp. 5172–5182. Available at: https://doi.org/10.1128/JVI.02430-10.

Xue, W. *et al.* (2012) 'Molecular modeling study on the resistance mechanism of HCV NS3/4A serine protease mutants R155K, A156V and D168A to TMC435', *Antiviral Research*, 93(1), pp. 126–137. Available at: https://doi.org/10.1016/j.antiviral.2011.11.007.

Yagnik, A.T. *et al.* (2000) 'A model for the hepatitis C virus envelope glycoprotein E2', *Proteins*, 40(3), pp. 355–366. Available at: https://doi.org/10.1002/1097-0134(20000815)40:3%3C355::aid-prot20%3E3.0.co;2-k.

Yamamoto, H. *et al.* (2015) 'Molecular architecture of the ribosome-bound Hepatitis C Virus internal ribosomal entry site RNA', *The EMBO Journal*, 34(24), pp. 3042–3058. Available at: https://doi.org/10.15252/embj.201592469.

Yamamoto, S. *et al.* (2016) 'Lipoprotein Receptors Redundantly Participate in Entry of Hepatitis C Virus', *PLOS Pathogens*, 12(5), p. e1005610. Available at: https://doi.org/10.1371/journal.ppat.1005610.

Yao, H. and Ye, J. (2008) 'Long Chain Acyl-CoA Synthetase 3-mediated Phosphatidylcholine Synthesis Is Required for Assembly of Very Low Density Lipoproteins in Human Hepatoma Huh7 Cells*', *Journal of Biological Chemistry*, 283(2), pp. 849–854. Available at: https://doi.org/10.1074/jbc.M706160200.

You, S. and Rice, C.M. (2008) '3' RNA elements in hepatitis C virus replication: kissing partners and long poly(U)', *Journal of Virology*, 82(1), pp. 184–195. Available at: https://doi.org/10.1128/JVI.01796-07.

Yu, C.-C. *et al.* (2022) 'Sequential Phosphorylation of Hepatitis C Virus NS5A Protein Requires the ATP-Binding Domain of NS3 Helicase', *Journal of Virology*, 96(7), p. e0010722. Available at: https://doi.org/10.1128/jvi.00107-22.

Yuhashi, K. *et al.* (2014) 'In vitro selection of the 3'-untranslated regions of the human liver mRNA that bind to the HCV nonstructural protein 5B', *Virology*, 450–451, pp. 13–23. Available at: https://doi.org/10.1016/j.virol.2013.11.036.

Zahid, M.N. *et al.* (2013) 'The Postbinding Activity of Scavenger Receptor Class B Type I Mediates Initiation of Hepatitis C Virus Infection and Viral Dissemination', *Hepatology*, 57(2), p. 492. Available at: https://doi.org/10.1002/hep.26097.

Zapatero-Belinchón, F.J. *et al.* (2021) 'Interdependent Impact of Lipoprotein Receptors and Lipid-Lowering Drugs on HCV Infectivity', *Cells*, 10(7), p. 1626. Available at: https://doi.org/10.3390/cells10071626.

Zeuzem, S. *et al.* (2011) 'Telaprevir for retreatment of HCV infection', *The New England Journal of Medicine*, 364(25), pp. 2417–2428. Available at: https://doi.org/10.1056/NEJMoa1013086.

Zhang, Y. *et al.* (2022) 'Benchmarking Refined and Unrefined AlphaFold2 Structures for Hit Discovery'. ChemRxiv. Available at: https://doi.org/10.26434/chemrxiv-2022-kcn0d-v2.

Zhao, Z. *et al.* (2014) 'A Neutralization Epitope in the Hepatitis C Virus E2 Glycoprotein Interacts with Host Entry Factor CD81', *PLoS ONE*, 9(1), p. e84346. Available at: https://doi.org/10.1371/journal.pone.0084346.

Zhong, J. *et al.* (2005) 'Robust hepatitis C virus infection in vitro', *Proceedings of the National Academy of Sciences of the United States of America*, 102(26), pp. 9294–9299. Available at: https://doi.org/10.1073/pnas.0503596102.

Zhou, D. *et al.* (2019) 'Unexpected mode of engagement between enterovirus 71 and its receptor SCARB2', *Nature Microbiology*, 4(3), pp. 414–419. Available at: https://doi.org/10.1038/s41564-018-0319-z.

Zhou, T. *et al.* (2018) 'NS3 from Hepatitis C Virus Strain JFH-1 Is an Unusually Robust Helicase That Is Primed To Bind and Unwind Viral RNA', *Journal of Virology*, 92(1), pp. e01253-17. Available at: https://doi.org/10.1128/JVI.01253-17.

Zimmerman, B. *et al.* (2016) 'Crystal Structure of a Full-Length Human Tetraspanin Reveals a Cholesterol-Binding Pocket', *Cell*, 167(4), pp. 1041-1051.e11. Available at: https://doi.org/10.1016/j.cell.2016.09.056.

Zufferey, R. *et al.* (1997) 'Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo', *Nature Biotechnology*, 15(9), pp. 871–875. Available at: https://doi.org/10.1038/nbt0997-871.

Zwanzig, R., Szabo, A. and Bagchi, B. (1992) 'Levinthal's paradox.', *Proceedings of the National Academy of Sciences*, 89(1), pp. 20–22. Available at: https://doi.org/10.1073/pnas.89.1.20.