Adaptive Immune Evasion in Clinically Latent HIV Infection



Sophie Marie Andrews

St Edmund Hall

University of Oxford

Professors Sarah Rowland-Jones & Tao Dong

A thesis submitted for the degree of

Doctor of Philosophy

in

Infection, Immunology and Translational Medicine

Medical Sciences Doctoral Training Centre

Trinity Term 2016

Adaptive Immune Evasion in Clinically Latent HIV Infection

Sophie Marie Andrews, St Edmund Hall

DPhil Thesis, Trinity Term 2016

HIV is a master of immune evasion, utilising a range of different techniques to not only survive the human immune system, but also mediate its eventual catastrophic decline. Understanding how HIV evades the adaptive immune response is paramount to developing effective treatments and vaccines. This thesis aimed to investigate three key ways in which HIV-1 and HIV-2 mediate immune evasion in the context of clinically latent infection. Chapter three summarises a study into selection pressure and mutation in the gp120 envelope gene in a narrow-source HIV-1 cohort of former plasma donors (FPDs) from China. This study further characterised the cohort, and identified specific mutations in the gene consistent with antibody and CTL-driven selection pressure. Chapter four describes an investigation into the downregulation of HLA-I mediated by primary isolates of HIV-1 and HIV-2 Nef. Nef-mediated HLA-I downregulation contributes to the evasion of CTL responses. In stark contrast to previous reports, no evidence for differential downregulation of HLA-A and HLA-B was detected, but primary isolates invariably showed reduced activity relative to laboratory-adapted and consensus variants. In performing this study, a number of limitations came to light regarding how bifurcate analyses are used to interpret flow cytometric data collected in studies of receptor modulation. A novel technique - SWARM - was therefore developed to address these limitations, and is described in chapter five. Chapter six aimed to address why CTL responses against HIV-2 Nef are rare, despite HIV-1 Nef being highly immunogenic. A series of *in vitro* (immuno)proteasomal processing assays revealed HIV-2 Nef is more extensively digested than HIV-1 Nef, but further experimentation is required to explain the difference in response. Finally, chapter seven briefly summarises a successful collaborative attempt to resolve the first ever crystal structure of HIV-2 Nef.

Acknowledgements

When I embarked on this journey four years ago, I never anticipated that my DPhil studies would afford me the opportunities that they have. Having travelled to six countries on three continents, performed exciting research, and met lifelong friends all over the world, I can confidently say that my time has been well spent. However, the path has been far from smooth, and completing this thesis would not have been possible without the support and knowledge of a wealth of individuals who must be duly acknowledged.

First and foremost, I would like to express my gratitude to the patients, without whom this research would quite simply not be possible. Thank you so much for providing your samples and consent: your generosity in trying times is instrumental, and so this thesis is dedicated to you.

I would next like to acknowledge my supervisors, Sarah Rowland-Jones and Tao Dong. Sarah, thank you for your unwavering support and endless patience: you are an inspiration to me on so many levels, and I have felt honoured to be a part of your lab family. Truly, you are one of the coolest ladies I know. Tao, thank you for your stimulating discussions and passion for science: no one knows their way around a T cell quite like you!

My heartfelt thanks also go to Joakim Esbjörnsson - King of Phylogenetics; Simon Brackenridge - Lord of Cloning; and Kat James - Statistician Extraordinaire (and wonderful friend). I am indebted to you all, and can honestly say that without your help, I would not have a thesis to submit!

Benedikt Kessler and Nicola Ternette, thank you for your insight into proteasomal processing and immunopeptidomes: your knowledge of these fields is staggering. Thank you also to Marie-Laëtitia Thézénas for your expertise in mass spectrometry, and to Paul Sopp for yours in flow cytometry. Sunetra Gupta and Paul Wikramaratna - thank you for involving me in what ultimately became a very exciting collaborative project; and to Thushan de Silva - thank you for your sample collection and support which has made the HIV-2 work possible.

My warmest thanks also go to Katsumi Maenaka and Kimiko Kuroki for hosting me at the University of Hokkaido in Japan: your hospitality and kindness was exceptional. I am particularly grateful to Kengo Hirao, for allowing me to work alongside him in the lab, and Toyoyuki Ose for his guidance in structural biology. I would also like to thank Chisato Yamada, Mizuki Saito, Ami Takahashi, XinXin Jiang, and many others for their kindness and friendship. Similarly, I would like to thank Yonghong Zhang for hosting me in his lab at Beijing You'an Hospital in China, and the members of his group for being so welcoming and supportive: in particular, JinHua Liu and Ling Qin.

Also deserving of acknowledgement are Seph Borrow and Mary Carrington. Seph, thank you for always being an approachable and understanding graduate adviser: I am grateful for our coffee chats, and I always feel much calmer having spoken with you. Mary, thank you for your insightful guidance and your passion: the lab is not the same in your absence.

Louis-Marie Yindom, thank you for your calm wisdom, your humour, and your friendship. Singe Sibeko - you are perhaps the warmest and most thoughtful person I know, and I am so grateful to call you my friend. Henry Laurenson-Schafer and Xi Li - my Dongsterfam - thank you for legit bantz, and for keeping me sane A.F.! Glenn Wong, Yanchun Peng, Chi Zhang, Kerry Kite, Megat Abd Hamid, Xuan Yao, Miguel Garcia-Knight, Linghang Wang, Mohammad Ali Rai, Brittany Shepherd, Hani Abidi, Lili Wang, Yan Zhao, and the numerous other SRJ and Dong lab members, past and present: my time in Oxford would not have been the same without you, so thank you for being a part of it and my apologies for such a brief and insufficient acknowledgement.

Karen Poxon and Laura Nevay, thank you so much for all you have done (and continue to do) to keep both groups running smoothly! Thanks also to the administrators and facilities staff, without whom all research would grind to a halt. In particular, Craig Waugh; Kathryn Robson; Alastair Waugh; Becky Horsefall; Julia McCann; Matt Husband; Keith Bourne; Tiph Bouriez-Jones; Omar Rivas Rosiles; Leo Dempsey; Ling Jinks; Ross MacRae; Rosie Sharp and Steph Yoxall.

I would next like to acknowledge the Wellcome Trust as the source of funding for this research, and the Infection, Immunology and Translational Medicine DPhil programme directors for allowing me the opportunity to be here. I'd also like to thank

the British Society for Immunology (BSI), the International Antiviral Society USA (IAS-USA), and the League of European Research Universities (LERU) for providing travel grants to attend various conferences and workshops.

Lara Spieker, Meike Aßmann, Zoë Monkley, and Imogen Schofield - thank you so much for continuing to put up with this nonsense. Your friendship has been a gift.

And last but definitely not least, I would like to give my biggest thanks to my wonderful family, who have supported me right from the beginning ("Ask me a question, any question!"). To my mum (who taught me everything I know about germs) and to my dad (who has always encouraged me to smash down all the boundaries): I am the person I am today because of you. Thank you so much for the wonderful upbringing you have given me, and for instilling in me the desire to try harder, do better, and be more. To my sisters, Elicea, Lily and Maria, thank you for always being there when I need you most. To Sara, thank you for always believing in me, and to grandad, thank you for being the voice of reason throughout this frantic time.

Finally, I would like to thank my fiancé and best friend, Matthew. You've been a source of infinite love and encouragement, and have selflessly given so much to support me. I can't even begin to put into words how grateful I am to have you in my life. I love you, and I'm hooked on a feeling.

Statement of Originality

I declare that the work presented within this thesis entitled "Adaptive Immune Evasion in Clinically Latent HIV Infection" is my own. Contributions made by collaborators have been clearly acknowledged. I have not submitted this thesis in whole or in part for any other degree or qualification at The University of Oxford or elsewhere.

Sophie Marie Andrews
Trinity Term 2016

Authored Publications

1. Andrews, S. M. & Pollard, A. J. A vaccine against serogroup B *Neisseria meningitidis*: dealing with uncertainty. *The Lancet Infectious Diseases* **14,** 426–434 (2014)

2. Andrews, S. M. *et al.* in *Encyclopedia of AIDS* (eds Hope, T. J. *et al.*) 1–10 (Springer New York, 2015)

3. Andrews, S. M. *et al. Does Reduced HLA-I Downregulation by HIV-2 Nef Contribute to Functional Cure? [CROI Abstract 253]* in *Special Issue: Abstracts From the 2016 Conference on Retroviruses and Opportunistic Infections.* **24(e-1):391** (Topics in Antiviral Medicine, 2016)

Grants and Awards

1. Awarded a four year stipendiary DPhil studentship (£159,340 GBP) by the Wellcome Trust; October 2012.

2. Selected to attend the 19$^{\text{th}}$ International Bioinformatics Workshop on Virus Evolution and Molecular Epidemiology (VEME) in Rome (IT); September 2014.

3. Awarded a Travel Grant (£700 GBP) by the British Society for Immunology (BSI) to attend the 2015 Society for Molecular Biology and Evolution (SMBE) Annual Meeting in Vienna (AT); July 2015.

4. Awarded a Young Investigator Scholarship ($2,200 USD) by the IAS-USA / CROI Foundation to attend the 2016 Conference on Retroviruses and Opportunistic Infections (CROI) in Boston, Massachusetts (US); February 2016.

5. Selected to attend and awarded a full scholarship (€600 EUR) for the 2016 League of European Research Universities (LERU) Doctoral Summer School on *Data Stewardship for Scientific Discovery and Innovation* in Leiden, The Netherlands (NL); July 2016.

Contents

1 Introduction **1**

1.1 The Human Immunodeficiency Virus 1

1.1.1 The Origins and History of HIV 1

1.1.2 Evolution and Diversity 3

1.1.3 Epidemiology 6

1.1.4 Transmission 7

1.1.5 Natural History of Infection 9

1.1.6 Life Cycle and Replication 14

1.1.7 The Structure of a Mature Virion 19

1.1.8 The Viral Genome 21

1.2 The Immune Response to HIV Infection 33

1.2.1 Mucosal Barriers to Infection 33

1.2.2 Innate Immune Response 35

1.2.3 Antigen Presentation and the Adaptive Response 46

1.2.4 Cellular Adaptive Response 50

1.2.5 Humoral Adaptive Response 53

1.3 Summary and Study Objectives 57

2 Materials and Methods **61**

2.1 Study Cohorts 61

2.1.1 HIV-1 (SM cohort) 61

2.1.2 HIV-2 (TD cohort) 63

2.2 Materials and Methods for Chapter 3 66

2.2.1 Selection of Patient Isolates 66

2.2.2 Extraction of Viral RNA 68

2.2.3 Amplification of Gp120 *Env* Fragment 68

2.2.4 Cloning of PCR Products for Sequencing 70

2.2.5 Colony PCR 70

2.2.6 Sequencing of Colony PCR Products . . . 72
2.2.7 Sequence Analysis and Statistics . . . 72
2.3 Materials and Methods for Chapters 4 & 5 . . . 78
2.3.1 Generation of Consensus HIV-1 and HIV-2 *Nef* Genes . . . 78
2.3.2 Selection of Patient Isolates . . . 78
2.3.3 Amplification of Full-Length *Nef* Gene . . . 78
2.3.4 Cloning of Full-Length *Nef* Gene . . . 81
2.3.5 Large-Scale Purification of Plasmid DNA . . . 84
2.3.6 Cell Culture and Maintenance . . . 85
2.3.7 Transfection by Electroporation . . . 86
2.3.8 Cell Staining . . . 87
2.3.9 Flow Cytometry . . . 87
2.3.10 Analysis and Statistics . . . 89
2.3.11 Generation of Flexible Expression Vector . . . 89
2.3.12 SWARM Development . . . 90
2.4 Materials and Methods for Chapter 6 . . . 91
2.4.1 *In Vitro* (Immuno)Proteasomal Digestion Assay . . . 91
2.4.2 Analysis and Statistics . . . 92
2.5 Materials and Methods for Chapter 7 . . . 94
2.5.1 Statement of Contribution . . . 94
2.5.2 Gene Cloning & Protein Expression . . . 94
2.5.3 Protein Purification . . . 96
2.5.4 Crystallisation & Structure Determination . . . 98
2.6 Reagents and Buffers . . . 99

3 Selection and Mutation in Gp120 Envelope in Clinically Latent HIV Infection 103
3.1 Introduction . . . 103
3.1.1 Aims . . . 105
3.2 Results . . . 106
3.2.1 Epitope Mapping of Envelope . . . 106
3.2.2 SM Cohort Characterisation . . . 108
3.2.3 Evolution and Selection Pressure . . . 116
3.3 Discussion . . . 131

4 Nef-Mediated Class-I HLA Downregulation in HIV-1 and HIV-2 Infection **137**
4.1 Introduction . . . 137
4.1.1 Mechanism of Downregulation . . . 138
4.1.2 Functional Motifs Associated with Downregulation . . . 140
4.1.3 Differential Downregulation of HLA-A and HLA-B . . . 142
4.1.4 Aims . . . 143
4.2 Results . . . 144
4.2.1 Cohort Characterisation . . . 144
4.2.2 Experimental Optimisation . . . 149
4.2.3 Nef-Mediated HLA-I Downregulation . . . 152
4.3 Discussion . . . 168

5 A Novel Technique for the Analysis of Flow Cytometric Data **177**
5.1 Introduction . . . 177
5.1.1 Limitations of Bifurcate Analyses . . . 178
5.1.2 Proposals for an Improved Analysis . . . 184
5.1.3 Aims . . . 185
5.2 Results . . . 186
5.2.1 SWARM Workflow . . . 186
5.2.2 Vector Optimisation . . . 188
5.2.3 Development of Staggered Window Approach . . . 193
5.2.4 Curve Fitting . . . 197
5.2.5 Nef-Mediated HLA-I Downregulation . . . 199
5.3 Discussion . . . 207

6 Proteasomal Processing and the Antigenicity of HIV-2 Nef **213**
6.1 Introduction . . . 213
6.1.1 Aims . . . 215
6.2 Results . . . 216
6.2.1 Mapping of Nef CTL Epitopes . . . 216
6.2.2 Characterisation of Digestion Products . . . 217
6.2.3 Identification & Analysis of Optimally Cleaved CTL Epitopes 222
6.2.4 HLA-I Types Associated with the Epitopes Generated . . . 224
6.2.5 Comparisons of Corresponding Digestion Products . . . 225
6.2.6 Identification and Analysis of Epitope Precursors . . . 227
6.3 Discussion . . . 231

7 Structural Biology of the HIV-2 Nef Protein **237**
7.1 Introduction 237
7.1.1 Aims 238
7.2 Results 239
7.3 Discussion 242

8 Conclusions and Future Directions **245**

References **255**

Appendix A Supplementary Tables **293**

List of Figures

1.1 Natural History of Untreated HIV-1 Infection 10
1.2 Life-Cycle and Replication of HIV 15
1.3 Structure of a Mature Virion 20
1.4 Genomes of HIV-1 and HIV-2 22
1.5 The Genital Mucosa in Health and Disease 34
1.6 Activation Pathways of the Complement Cascade 43
1.7 Structures of Class I and II Human Leucocyte Antigen 48
1.8 The Cellular Adaptive Response 51
1.9 The Humoral Adaptive Response 54
1.10 The Big Picture: The Immune Response Against HIV 57

2.1 Henan Province, China 62
2.2 Cacheu Region, Guinea-Bissau 63
2.3 Gating Strategy 88

3.1 Epitope Map of the Gp120 Envelope Protein 107
3.2 Phylogenetic Subtyping of SM Cohort Samples 109
3.3 Alignments of Inferred Founder Sequence 113
3.4 Viral Population Dynamics 114
3.5 Clock Rate of Time-Stamped Samples 116
3.6 Clock Rate and Clinical Parameters 118
3.7 Codon-Specific dN/dS Estimates in Gp120: Part 1 120
3.8 Codon-Specific dN/dS Estimates in Gp120: Part 2 121
3.9 Codon-Specific dN/dS Estimates Across Patients 122
3.10 Sites Under Significant Positive Selection Pressure 123
3.11 Characterisation of Sequence Variants 124
3.12 Unique Variants 126
3.13 Major Variant Frequency 127
3.14 Major Variants by Patient 128

3.15 Major Variants Mapped to Structure 129

4.1 Proposed Mechanisms of Nef-Mediated HLA-Downregulation 140
4.2 Baseline Median Fluorescent Intensity of Chimeric HLA Molecules . . 149
4.3 Proportion of HIV-1 and HIV-2 Nef Positive Cells 150
4.4 Proportion of Nef Positive Cells and HLA-I Downregulation 151
4.5 HLA-A and HLA-B are Not Differentially Downregulated 152
4.6 HLA-I is not Differentially Downregulated by HIV-1 and HIV-2 Nef . 153
4.7 Plots of Fold Downregulation by Patient Nef Isolate 155
4.8 Frequency of HLA-I Groups in SM and TD Cohorts 156
4.9 Relationship Between HLA-I Downregulation & Clinical Parameters . 158
4.10 HLA-I Downregulation Mediated by SF2 Nef 159
4.11 Most Patient Nef Isolates are Less Efficient than SF2 Nef 160
4.12 HLA-I Downregulation Mediated by Consensus HIV-1 & HIV-2 Nef . 161
4.13 Patient Nef Isolates are Less Efficient than Consensus Nef 162
4.14 Sequence Variation in Nef Mapped to Epitope Density 164
4.15 Structural Modelling of HLA Cytoplasmic Tail Variants 166
4.16 Baseline HLA-I Levels and Fold Downregulation 167

5.1 Idealised and Actual Population Distributions in Transfected Cells . . 178
5.2 The Importance of Appropriate Gating in Bifurcate Analyses 180
5.3 The Effect of Transfection Efficiency on Bifurcate Analyses 181
5.4 The Effect of Expression Levels on Bifurcate Analyses 183
5.5 SWARM Workflow 187
5.6 Linker Length Affects HIV-1 Nef Folding in Fusion 189
5.7 Linker Length Affects HIV-2 Nef Folding in Fusion 189
5.8 Generation of Flexible Expression Vector 191
5.9 Expression of Nef from Flexible Expression Vector 192
5.10 Sliding Windows are Robust 194
5.11 Density of SWARM Bins 196
5.12 Verification of Bin Parameters 197
5.13 Optimisation of SWARM Best Fit Line 198
5.14 SWARM Analysis Grouped by Nef Isolate 200
5.15 SWARM Analysis Grouped by HLA Type 201
5.16 Slopes of HLA-I Downregulation Mediated by Nef 202
5.17 Interquartile Range of Downregulation by Bin 203
5.18 Surface Expression of HLA-I Following Transfection With Nef 204

6.1 Epitope Map of the Nef Protein 216
6.2 Digestion Products from Peptide Pairs 1 & 2 218
6.3 Digestion Products from Peptide Pairs 3 & 4 219
6.4 Location of Cleavage Sites in Digested Peptides 221
6.5 Number of Optimally Cleaved Nef CTL Epitopes 223
6.6 Venn Diagrams of Overlapping Digestion Products 226
6.7 Frequency of N-Terminal Cleavage 228
6.8 Frequency of C-Terminal Cleavage 229
6.9 Proportion of Cleavage Sites Corresponding to Epitope Termini . . . 230

7.1 Crystal Structure of HIV-2 Nef 240
7.2 HIV-2 Nef C-Terminal α-Helix 241

List of Tables

1.1 Comparison of HIV-1 and HIV-2 5
1.2 Risk of HIV Infection by Exposure Route 9
1.3 AIDS-Defining Illnesses 13
1.4 Immune Evasion Mechanisms of HIV 58

2.1 Demographic Characteristics of Selected SM Cohort Patients 66
2.2 Clinical Characteristics of Selected SM Cohort Patients 67
2.3 PCR Reagents and Cycling Parameters to Amplify *Env* Gp120 69
2.4 Colony PCR Reagents and Cycling Parameters 71
2.5 Programs used for Sequence Analysis 75
2.6 Clinical Characteristics of Selected SM Cohort Patients 79
2.7 PCR Reagents and Cycling Parameters to Amplify *Nef* Genes 80
2.8 PCR Reagents and Cycling Parameters to Adhere Restriction Sites . 82
2.9 Overlapping Nef Peptides 91
2.10 Buffer and Media Composition 99

4.1 HLA Typing of the HIV-1 SM Cohort 145
4.2 HLA Typing of the HIV-2 TD Cohort 146
4.3 Functional Domains of Selected Nef Isolates 148

5.1 Surface Expression of HLA-I Following Transfection With Nef 206

6.1 (Immuno)Proteasomal Digestion Product Summary 217
6.2 HLA Types Associated with Optimally Cleaved Epitopes 225

7.1 Statistics of Crystallographic Data Collection and Refinement 239

A.1 PCR Primers 293
A.2 Inferred Founder Sequence 295
A.3 HLA Types Sharing Cytoplasmic Tail 296
A.4 Cytoplasmic Tails of Chimeric HLA-I Molecules 298

A.5 Synthetic Genes Used for Construction of Modified Vectors 299
A.6 Binning Profile Used in SWARM Analysis 301
A.7 Nef Digestion Products Corresponding to Known CTL Epitopes . . . 302
A.8 Accession Numbers 307

List of Abbreviations

aa	amino acid
Ab	antibody
ABC	ATP-binding cassette
ACE	angiotensin-converting enzyme
ADCC	antibody-dependent cell-mediated cytotoxicity
ADP	adenosine diphosphate
AFU	arbitrary fluorescent units
AIDS	acquired immune deficiency syndrome
AP	alkaline phosphatase
APC	antigen-presenting cell
APOBEC3	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
APS	ammonium persulphate
Arf1	ADP-ribosylation factor 1
Arf6	ADP-ribosylation factor 6
ARNO	ARF nucleotide-binding site opener
ARS	acute retroviral syndrome
ART	antiretroviral therapy
BCL	B cell line
BEAST	Bayesian evolutionary analysis sampling trees
BEAUti	Bayesian evolutionary analysis utility
bNAb	broadly neutralising antibody
CA	capsid
cDC	conventional dendritic cell
cDNA	complementary DNA
Chk1	checkpoint kinase 1
CLIP	class II-associated invariant chain peptide
CLR	C-type lectin-like receptor
CNS	central nervous system
COP-I	coat protein I
CRL4	cullin ring E3 ligase
CRF	circulating recombinant form
cryo-EM	cryo-electron microscopy
CST	cytometer setup and tracking
CTL	cytotoxic lymphocyte
DC	dendritic cell
DC-SIGN	dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin
DISC	death-inducing signalling complex

DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DRiP	defective ribosomal product
dsDNA	double-stranded DNA
DTT	dithiothreitol
EBNA1	Epstein-Barr nuclear antigen 1
EC	elite controller
EDTA	ethylenediaminetetraacetic acid
ELISpot	enzyme-lined immuno-spot
eME	effective maximum expression
env	envelope
EPS	effective population size
ER	endoplasmic reticulum
ERAP	endoplasmic reticulum aminopeptidase
ESS	effective sample size
ExoI	exonuclease I
FADD	fas-associated death domain
FasL	fas ligand
FasR	fas receptor
FCS	foetal calf serum
fH	factor H
FPD	former plasma donor
gag	group-specific antigen
GRID	gay-related immune deficiency
GTPase	guanosine triphosphatase
GTR	generalised time-reversible
GUI	graphical user interface
HAART	highly active antiretroviral treatment
HCV	hepatitis C virus
HEK	human embryonic kidney
HEPES	N-2-hydroxyethylpiperazine-N-2-ethane sulphonic acid
HIV	human immunodeficiency virus
HLA	human leucocyte antigen
HLA-I	class I human leucocyte antigen
HPD	highest posterior density
HTS	high-throughput sampler
ICS	intracellular cytokine staining
IDE	integrated development environment
IDO	indoleamine 2,3-dioxygenase
IDU	injection drug user
IFN	interferon
IL	interleucin
IN	integrase
INSTI	integrase strand transfer inhibitor
IPTG	isopropyl β-D-1-thiogalactopyranoside
IQR	interquartile range

IRES	internal ribosome entry site
IRF	interferon regulatory factor
IS	immunological synapse
KIR	killer cell immunoglobulin-like receptor
KS	Kaposi's sarcoma
LANL	The Los Alamos National Laboratory
LB	Luria-Bertani
LC-Q-ToF-MS	liquid chromatography quadrupole time-of-flight mass spectrometry
LOESS	local regression
LRT	likelihood ratio test
LTNP	long-term non-progressor
LTR	long terminal repeat
MA	matrix
MAC	membrane attack complex
MALT	mucosa-associated lymphoid tissue
MAPK	mitogen-activated protein kinase
MCC	maximum clade credibility
MCMC	Markov chain Monte Carlo
MCS	multiple cloning site
ME	maximum expression
MEGA	molecular evolutionary genetics analysis
MES	2-(N-morpholino)ethanesulfonic acid
MFI	mean fluorescence intensity
MHC	major histocompatibility complex
MIIC	major histocompatibility complex class-II compartment
MRC	Medical Research Council
MRCA	most recent common ancestor
mRNA	messenger RNA
MSM	men who have sex with men
MTCT	mother-to-child transmission
MUSCLE	multiple sequence comparison by log-expectation
NAb	neutralising antibody
NC	nucleocapsid
nef	negative factor
NF	nuclear factor
Ni-NTA	nickel-nitrilotriacetic acid
NK	natural killer
NLR	NOD protein-like receptor
NLS	nuclear localisation signal
NNRTI	non-nucleotide reverse transcriptase inhibitor
NOD	nucleotide-binding oligomerisation domain
NRTI	nucleotide reverse transcriptase inhibitor
ORF	open reading frame
OxTREC	Oxford Tropical Research Ethics Committee
PACS	phosphofurin acidic cluster sorting
PAMP	pathogen-associated molecular pattern
PBMC	peripheral blood mononuclear cell

PBS	phosphate buffered saline
PCP	*Pneumocystis jirovecii* pneumonia
PCR	polymerase chain reaction
PDB	Protein Data Bank
pDC	plasmacytoid dendritic cell
PEG	polyethylene glycol
PI	protease inhibitor
PIC	pre-integration complex
PIP3	phosphatidylinositol-(3,4,5)-trisphosphate
PI3K	phosphoinositide 3-kinase
PM	plasma membrane
PML	progressive multifocal leucoencephalopathy
PMT	photomultiplier tube
pol	polymerase
PPT	polypurine tract
PR	protease
PRR	pathogen recognition receptor
P-TEFb	positive transcription elongation factor b
PVDF	polyvinylidene difluoride
RCSB	Research Collaboratory for Structural Bioinformatics
rev	regulator of expression of virion proteins
RIG-1	retinoic acid-inducible gene 1
RLR	RIG-1-like receptor
RNA	ribonucleic acid
RP	rapid progressor
RPM	rotations per minute
RPMI	Roswell Park Memorial Institute medium
RRE	rev-response element
RT	reverse transcriptase
RTC	reverse transcription complex
SAIDS	simian acquired immune deficiency syndrome
SAM	sterile alpha motif
SAMHD1	SAM domain and HD domain-containing protein 1
SAP	shrimp alkaline phosphatase
SD	standard deviation
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SERINC	serine incorporator
SEVI	semen-derived enhancer of virus infection
SFK	src family kinase
SFU	spot-forming unit
SH3	Src homology region 3
SIV	simian immunodeficiency virus
SLPI	secretory leucocyte protease inhibitor
SM	Shuang Miao
SOC	super optimal broth with catabolite repression
SP	spacer peptide

ssRNA	single-stranded RNA
SU	surface protein gp120
SWARM	staggered window approach to receptor modulation
TAE	tris-acetate-EDTA
TAP	transporter associated with antigen processing
tat	transactivator of transcription
TB	terrific broth
TBST	tris-buffered saline with tween
TCEP	tris(2-carboxyethyl)phosphine
TCR	T cell receptor
TE	tris-EDTA
TEMED	tetramethylethylenediamine
TFA	trifluoroacetic acid
TGN	trans-Golgi network
Th	T helper
TLR	toll-like receptor
TM	transmembrane protein gp41
TNF	tumour necrosis factor
Treg	T regulatory lymphocyte
TRIM5	tripartate-motif containing protein 5
Tris	trisaminomethane
tRNA	transfer RNA
URF	unique recombinant form
vif	viral infectivity factor
VL	viral load
vpr	viral protein regulatory
VPS	vacuolar protein sorting protein
vpu	viral protein unique
vpx	viral protein X
WebPSSM	web position-specific scoring matrices
WFDC1	whey acidic protein four-disulphide core domain 1
ZAP-70	zeta-chain-associated protein kinase 70

Chapter 1

Introduction

1.1 The Human Immunodeficiency Virus

1.1.1 The Origins and History of HIV

The emergence of the human immunodeficiency virus (HIV) was arguably one of the most catastrophic and influential disease outbreaks of the late 20$^{\text{th}}$ century. HIV is the virus responsible for acquired immune deficiency syndrome (AIDS), and is estimated to have caused over 36 million deaths globally.

AIDS was first identified in the early 1980s, after an "unexpected cluster" of homosexual men in the San Francisco and New York City areas presented with an unusual wasting disease, extremely rare illnesses such as Kaposi's sarcoma (KS) and *Pneumocystis jirovecii* pneumonia (PCP), and an increased susceptibility to opportunistic infections suggestive of an immune disorder [1–4]. In response, the term "gay-related immune deficiency" (GRID) was coined to describe the syndrome. Before long, GRID was being reported in other parts of the United States and in other demographic groups such as immigrating Haitians, injection drug users (IDUs) and haemophiliacs [5].

Due largely to this demographic expansion, the term GRID was soon retired, and replaced with "acquired immune deficiency syndrome" (AIDS) which is still in use today. The aetiological agent of AIDS was at this point unclear, and little progress was made in understanding the disease until the isolation of the virus itself in 1983 by Barré-Sinoussi and colleagues [6].

Whilst HIV garnered attention as a result of cases in the United States, the virus was introduced into the human population in Sub-Saharan Africa, where it still exerts its greatest impact in terms of infection, morbidity, and mortality. HIV entered humans through a series of cross-species transmissions of simian immunodeficiency viruses (SIVs) from non-human primates.

The initial route of zoonotic transmission is poorly understood and a matter of hot debate, but several hypotheses have been proposed, and phylogenetic studies tend to place the date of transmission between the late 19th and early 20th century [7–9]. A possible explanation for the origin of HIV is the "bushmeat theory" [10–12], which proposes that whilst hunters are frequently exposed to SIV-contaminated blood while procuring and preparing bushmeat [11], a small number of individuals became infected with a virus that subsequently underwent recombination and acquired mutations that led to its evolution into HIV in its current form. The evolved virus then spread within the hunter's population and further afield. Although SIVs have been in existence for at least 32,000 years [13], and bushmeat hunting is an ancient practice, the rapid emergence of multiple epidemic groups of HIV within a short time are the result of a number of social and political factors arising in regions of Africa at the time.

Colonialism has been suggested as a likely cause of increased exposure, spread, and adaptation of the virus [10]. European colonisation of Africa triggered many social

changes, which corresponded with increased sexual promiscuity and prostitution, establishment of forced labour camps with poor conditions, non-sterile vaccination [14], and an increase in the prevalence of health problems classed as risk-factors, such as genital ulcer disease [15].

1.1.2 Evolution and Diversity

HIV consists of two distinct lineages of virus: HIV-1 and HIV-2 which arose from separate zoonotic transmissions from chimpanzees (*Pan troglodytes troglodytes*) & gorillas (*Gorilla gorilla*) and sooty mangabeys (*Cercocebus atys*), respectively. Furthermore, HIV-1 can be divided into four groups on the basis of independent transmission events: major (M), outlier (O) and the newer groups N (non-major/non-outlier) and P (pending). Group M, which constitutes the majority of HIV-1 infections, contains nine different subtypes: A, B, C, D, F, G, H, J and K.

Further variation exists in the form of recombinants which arise when different subtypes combine within an infected cell to form a hybrid virus: hundreds of unique recombinant forms (URFs) have been described, and some successful recombinants spread within populations to become circulating recombinant forms (CRFs). The prevalence of each individual subtype varies dramatically by geographic region, with subtype B dominant in Europe [16]. Recently however, there has been an increase in diagnoses of infection with viruses non-endemic to the area, likely due to increases in international travel [17].

Conversely, HIV-2 has arisen from at least eight individual zoonotic transmissions [18], with the descendants classified into groups A-H. Of these, only group A and B have resulted in significant numbers of infections, with groups C-H having each been identified in only one or two individuals. Unlike HIV-1, HIV-2 has not disseminated

widely and the majority of cases have been reported in West Africa. However, additional clusters of infections exist in Europe - particularly in France and Portugal - where introduction of the virus has been attributed to the Portuguese Colonial War of 1961-1974 [19].

Owing to the notoriously error-prone nature of reverse transcriptase (section 1.1.8.3), both HIV-1 and HIV-2 have some of the highest described rates of evolution, and exhibit extensive genetic diversity. In HIV-1, the introduction of base-substitution errors has been estimated to occur between $^1/_{2000}$ to $^1/_{4000}$ bases, which equates to approximately five to ten errors per genome per round of viral replication [20]. Furthermore, mutational hotspots have been described in which error rates could be as high as $^1/_{70}$ bases [21]. The rapid rate of evolution is responsible for the striking degree of viral diversity seen both within individual patients (quasispecies diversity), and between infected individuals (interpatient diversity). A strong positive correlation between diversity and viral fitness has been reported in the literature [22].

The combination of distinct lineages and rapid evolution has led to a staggering degree of viral heterogeneity. Whilst HIV-1 and HIV-2 are capable of causing indistinguishable end-stage disease, there are a number of significant genomic, structural and function differences between the viruses, summarised in table 1.1.

Table 1.1 Comparison of the genetic, structural and functional differences between HIV-1 and HIV-2

	HIV-1	HIV-2	Ref
Zoonosis			
Species of origin	Chimpanzee (M & N)	Sooty Mangabey	[23, 24]
	Pan troglodytes troglodytes	*Cercocebus atys*	
	Gorilla (O & P)		
	Gorilla gorilla		
Phylogenetic classification	Four groups: M, N, O & P	Eight groups: A-H	[18, 25][1]
	M subgroups: A, B, C, D, F, G, H, J & K		
	CRFs prevalent	CRFs rare	[26, 27]
Transmission date	1930 (±20yr; Group M)	1940 (±16yr, Group A)	[7, 8]
		1945 (±14yr, Group B)	
Epidemiology[2]			
Distribution	Pandemic	Restricted endemic	[28, 29]
Prevalence	Global: 36.7 million	Global: 1-2 million	
	UK: 103,700	UK: 150	
	Africa: 25.5 million (70% of global cases)	West Africa: 1-2 million	
	Stable/Increasing	Decreasing	
MTCT Risk	15-45% risk	0-4% risk	[30–33]
Natural History & Progression[3]			
$CD4^+$ decline	Equivalent in patients matched for viral load		[34]
Proviral load	Equivalent in patients matched for $CD4^+$ count		[35]
Plasma viral load	High	30-fold lower	[35–37]
Five-year survival	Poor	Good	[38]
Prognosis with HAART	Equivalent survival and immunological/virological response		[39]
Life cycle			
Co-receptors	Narrow usage	Broad usage	[40–43]
	CCR5 & CXCR4	CCR1, CCR2b, CCR3, CCR4, CCR5, CCD8, CXCR4, CXCR6 & GPR15	
TCR-CD3 modulation	No Nef-mediated downregulation	Nef-mediated downregulation	[44]

Continued on next page

Table 1.1 – *Continued from previous page*

	HIV-1	**HIV-2**	**Ref**
Genome			
Size	9.5kb	10.5kb	[45]
Accessory proteins	Vif, Vpr, Vpu & Nef	Vif, Vpr, Vpx & Nef	[45]
Immunological Response			
Immune activation	Systemic	Reduced	[46, 47]
$CD4^+$ response to Gag	Moderate	More prevalent and polyfunctional	[48, 49]
TCR usage	Limited	Promiscuous	[50]
CTL response	Equivalent in frequency and magnitude		[51, 52]
	HIV-2 responses may be more polyfunctional		[49]
NAb response	Limited neutralisation	bNAbs generated	[53, 54]
NK cell response	Progressive loss of function	Functionally preserved	[55, 56][1]

[1] Reference listed is a review article as the primary source material is extensive.
[2] All epidemiological data are approximate values taken from the most recently published sources.
[3] In untreated individuals, except where specified.

1.1.3 Epidemiology

Since its emergence, HIV-1 has reached pandemic levels, with an estimated 36.7 million people worldwide living with the virus in 2015 (approximately 0.5% of the global population) [58]. However, the greatest burden of disease is concentrated amongst low to middle income countries [59] - especially in Sub-Saharan Africa - which is home to 70% of HIV-positive people globally and where one in twenty adults are believed to be infected. Some Sub-Saharan countries including Botswana, Lesotho, and Swaziland have exceedingly high prevalence rates, with recent estimates being 18.5%, 23%, and 31% of the adult population, respectively [60–62].

In the United Kingdom, the picture is very different. Approximately 103,700 people are believed to be HIV positive, with around 17% unaware that they are infected. 6,151 new cases were diagnosed in 2014: a number that has been increasing gradually since 2011 [63]. The demographics of HIV patients in the UK have also shifted dramatically since the onset of the epidemic, with an increase in the mean age of patients in treatment. However, key risk groups remain men who have sex with men (MSM), black Africans, and injection drug users (IDUs) [64].

The majority of HIV infections worldwide are attributable to HIV-1, with HIV-2 being found almost exclusively in West Africa where one to two million people are infected and numbers are currently decreasing [65]. HIV-2 establishes at lower viral loads [35] and is less infectious [66], meaning that transmissibility is greatly reduced.

Due largely to the development of highly active antiretroviral treatment (HAART) and availability of barrier methods, the spread of the HIV pandemic is slowing in many countries. However, access to antiretroviral drugs is often restricted for many patients in developing countries due to poverty or political unrest, accounting for their significantly higher transmission and steady increase in new infections.

1.1.4 Transmission

The major route of HIV infection globally is sexual transmission, with the receptive partner typically being at higher risk per act. Men who have sex with men (MSM) as well as heterosexual couples who practice penetrative anal sex are at high risk due to the increased likelihood of traumatic injury to mucocutaneous tissues, although transmission risk is also significant through vaginal sex. Transmission through non-penetrative sexual acts is also possible although the risk of transmission is greatly reduced. Mother-to-child transmission is an important route of infection in developing

countries, where transmission rates range from around 15-45% for HIV-1 and 0-4% for HIV-2 [30–33, 67]. The risk increases with extended breastfeeding, but with limited resources available, mothers often have few options. Mother-to-child transmission has been all but eradicated in the Western world, where treatment with antiretroviral drugs has dramatically cut the transmission risk, but access to such treatment is not readily available in Sub-Saharan Africa.

Injection drug users (IDUs) are at greatly increased risk of HIV acquisition as a result of needle-sharing practices which may lead to contact with infected blood. Others at risk of infection through contact with contaminated blood or blood-products are medical professionals, who face occupational hazards such as exposure to the virus through needle-stick injury or mucocutaneous splashes with infected material. The risk of infection through blood-transfusion has greatly reduced since the onset of the epidemic as a result of more stringent procedures regarding donor screening.

A list of exposure routes and their associated risks is shown in table 1.2. It is important to note that whilst the risk of infection per exposure is typically low, the cumulative effect of multiple exposures may be substantial.

Table 1.2 Risk of HIV infection by exposure route. Listed per 10,000 exposures and per single exposure. List adapted from NAM aidsmap [68]

Exposure route and risk per incident	Risk per 10,000 exposures	Risk per single exposure
Sexual routes		
Vaginal sex (female-to-male, high-income countries)	4	0.04%
Vaginal sex (male-to-female, high-income countries)	8	0.08%
Vaginal sex (female-to-male, low-income countries)	38	0.38%
Vaginal sex (male-to-female, low-income countries)	30	0.30%
Vaginal sex (source partner asymptomatic)	7	0.07%
Vaginal sex (source partner has late-stage disease)	56	0.55%
Receptive anal sex (MSM, partner unknown status)[1]	27	0.27%
Receptive anal sex (MSM, partner HIV positive)[1]	81	0.82%
Receptive anal sex (MSM, partner unknown status, condom)[1]	18	0.18%
Insertive anal sex (MSM, partner unknown status)[1]	6	0.06%
Insertive anal sex (MSM, partner unknown status, condom)[1]	4	0.04%
Receptive fellatio	0 to 4	0.00 to 0.04%
Parenteral routes		
Mother-to-child (mother untreated)	3,000	30.00%
Mother-to-child (mother on antiretroviral therapy)	80	0.80%
Mother-to-child (mother on combination therapy, VL <50)	10	0.10%
Injecting drug use	68 to 244	0.63 to 2.4%
Needlestick injury (no other risk factors)	13	0.13%
Blood transfusion with contaminated blood	9,250	92.50%

[1] The risk per penetrative anal event is higher in heterosexual couples as well as men who have sex with men. However, studies of heterosexual transmission typically do not make the distinction between vaginal and anal sex, as defining the exact transmission event is not possible.

1.1.5 Natural History of Infection

HIV-1 infection in untreated individuals almost invariably progresses through well-characterised and distinct phases: acute infection, clinical latency, the symptomatic phase, and AIDS (figure 1.1).

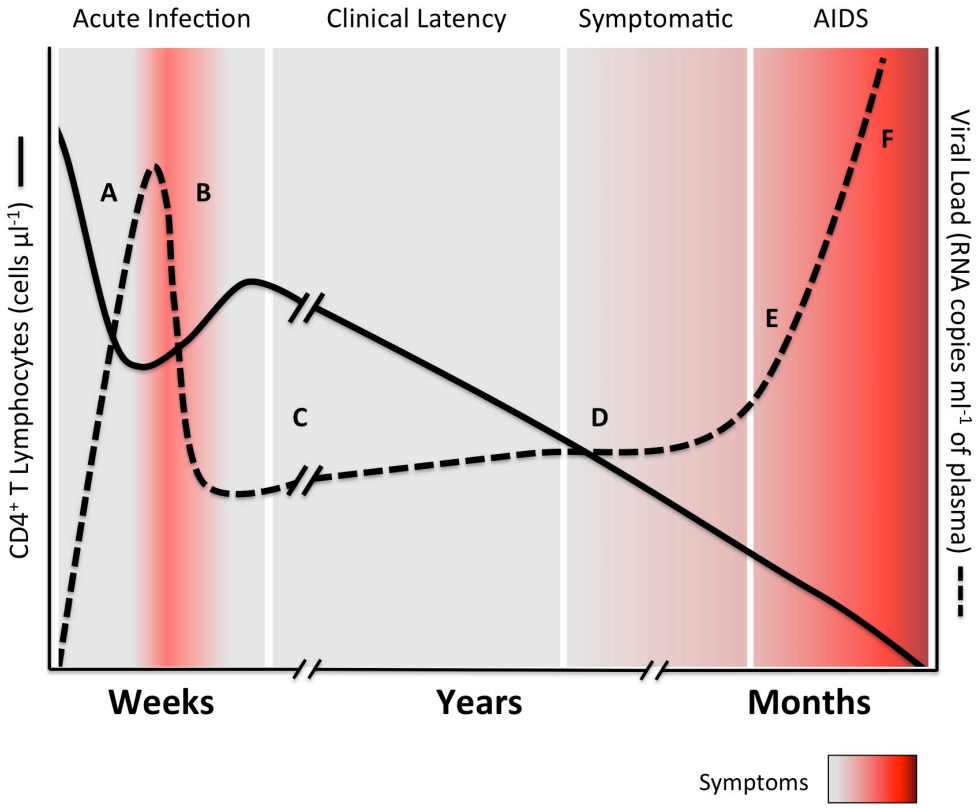


Figure 1.1: The natural history of untreated HIV-1 infection A) During acute (or primary) infection, the virus disseminates rapidly throughout the bloodstream, leading to high levels of circulating virus (dashed line), whilst the number of $CD4^+$ lymphocytes plummets (solid line). Symptoms may present in some patients (red shading). **B)** The immune system regains partial control, leading to suppression of the virus and restoration of $CD4^+$ lymphocyte levels. **C)** A viral set point is established, which may be maintained for up to a decade. **D)** Control of the virus begins to break down and the $CD4^+$ lymphocyte count steadily decreases, and constitutional symptoms begin to emerge. **E)** Viral dissemination continues as opportunistic infections take hold and the patient's $CD4^+$ lymphocyte count declines further. The patient now has clinically-defined AIDS **F)** Death soon follows. Figure re-drawn and adapted from Pantaleo *et al.* [69] and Fauci *et al.* [70].

1.1.5.1 Acute Infection

Immediately following infection with HIV, a patient enters the acute (or primary) phase (figure 1.1A). For approximately the first ten days, viral RNA is not yet detectable in the patient's plasma, and so this period is often referred to as the eclipse. Meanwhile, the virus replicates rapidly and disseminates throughout the

bloodstream [71]. Here, the patient is highly infectious [72–74]. The proliferating virus causes extensive damage to the $CD4^+$ T lymphocyte population, and large numbers of these cells die. At around 2-4 weeks, many patients present with an assortment of symptoms, collectively referred to as acute retroviral syndrome (ARS). ARS manifests in the form of flu-like symptoms [75] including fever, pharyngitis, myalgia, malaise and lymphadenopathy. Around half of patients will also develop a maculopapular rash.

The viral load soon peaks, and begins to decline as the immune system retaliates and regains a degree of control (figure 1.1B) [76]. This is due largely to the production of cytotoxic $CD8^+$ lymphocyte responses against the virus [77]. At this point, viral expansion is suppressed and the $CD4^+$ T lymphocyte population is partially restored, although rarely to pre-infection levels. During the later stages of acute infection, anti-HIV antibodies are generated in a process referred to as seroconversion.

1.1.5.2 Clinical Latency

Following acute infection, a set-point is reached, where both the viral load and $CD4^+$ count are maintained in relatively stable equilibrium (figure 1.1C). This marks the onset of the phase known as clinical latency (or chronic infection). The viral load established at the set-point is an important determinant of disease progression, and varies enormously by patient [78]. Lower set-points are associated with slower disease progression as a result of significantly reduced $CD4^+$ lymphocyte destruction [79, 80]. Clinical latency is characteristically asymptomatic, and its duration is around a decade on average although highly variable [81]. The virus continues to replicate during this time, but does not expand to the levels seen in acute infection [69]. The $CD4^+$ lymphocyte population continues to decline gradually.

1.1.5.3 Symptomatic Phase

Marking the end of clinical latency and signalling the breakdown of viral control is a short-lived symptomatic phase (figure 1.1D). This describes the transitional grey area leading up to the onset of clinically-defined AIDS. Patients present with non-specific constitutional symptoms such as fever, diarrhoea and night-sweats, as well as minor opportunistic infections.

1.1.5.4 AIDS

For reasons that are not yet fully understood, the immune system eventually loses control of the infection, and rapid dissemination of the virus occurs, coupled with a further loss of $CD4^+$ lymphocytes (figure 1.1E). As the $CD4^+$ lymphocyte population becomes increasingly damaged, the patient becomes highly susceptible to infections and neoplasms that would not pose a risk to healthy individuals, such as *Pneumocystis jirovecii* pneumonia (PCP) and Kaposi's sarcoma. In fact, it was the emergence of these highly unusual illnesses that first alerted medical professionals to the emergence of HIV in the Western hemisphere in the 1980s (section 1.1.1).

A patient is considered to have progressed to clinically-defined AIDS when their $CD4^+$ lymphocyte count has dropped below 200 cells μl^{-1} (or 14% of the total lymphocyte population); or they have acquired a recognised AIDS-defining illness [82]. A list of AIDS-defining illnesses is included in table 1.3. Without medical intervention, the median time of survival is 6-19 months for individuals in resource-poor settings who have progressed to AIDS (figure 1.1F) [83]. Survival time is also heavily influenced by external factors, such as age, nutrition, and the cumulative pathology of opportunistic illnesses. Owing to the very high levels of circulating virus at this point, patients with end-stage disease are highly infectious [74]. However, the availability of HAART means that few people in middle- to high-income countries progress to this stage.

Table 1.3 AIDS-Defining Illnesses List adapted from Schneider *et al.* [82]

Candidiasis (of airways)	Invasive cervical cancer	Coccidioidomycosis
Cryptococcosis	Cryptosporidiosis (chronic)	Cytomegalovirus disease
Encephalopathy	Herpes simplex (chronic)	Histoplasmosis
Isosporiasis (chronic)	Kaposi's sarcoma	Lymphoma (several forms)
Tuberculosis	*M. avium* complex (MAC)	Pneumonia (PCP / recurrent)
Leucoencephalopathy (PML)	Salmonella septicaemia	Toxoplasmosis (of brain)

1.1.5.5 Atypical Progression

Whilst the majority of HIV patients progress as described, outliers do exist. Rapid progressors (RPs) are usually described as HIV patients who exhibit the symptoms of end-stage disease or AIDS within two to three years from infection. Conversely, long-term non-progressors (LTNPs) are asymptomatic patients who maintain a stable $CD4^+$ lymphocyte count greater than 500-600 cells μl^{-1} whilst antiretroviral-naïve, despite a detectable viral load [84, 85]. Approximately 3% of people living with HIV are LTNPs [86, 87] - although this varies greatly (0.5%-25%), dependent on the cohort and definition used [88]. Many do not know that they are infected. Elite controllers (ECs) are the even rarer population of HIV-positive individuals who maintain an undetectable viral load (<50 RNA copies ml^{-1}) in the absence of any retroviral treatment [89].

Whilst the vast majority of untreated HIV-1-infected individuals eventually progress to AIDS, a substantial number of HIV-2 patients maintain immunological control over the virus, and do not progress further than clinical latency. Those who do may progress more slowly, but manifest clinically indistinguishable disease to HIV-1 progressors [90]. The proportion of HIV-2 patients who control the virus varies dramatically by cohort studied. This may perhaps be an artefact of recruitment

differences - for example from the clinic or community - or differences intrinsic to the population under study.

1.1.6 Life Cycle and Replication

The viral life cycle is dependent on infection of and replication within host cells, and progresses through defined steps: attachment, entry, uncoating, reverse transcription, integration, production of viral proteins, virion budding, and maturation (figure 1.2). Many of these steps are targeted by antiretroviral drugs and host restriction factors.

1.1.6.1 Attachment and Co-Receptor Usage

Circulating HIV attaches to target cells through the heteromeric envelope glycoprotein (section 1.1.8.4) spikes on the virion's surface. The envelope gp120 subunit binds to CD4 on the surface of T lymphocytes and macrophages, as well as a co-receptor: canonically the chemokine receptors CCR5 or CXCR4 (figure 1.2A). This triggers a conformational change that exposes and inserts the fusion peptide of the amino terminus of gp41 into the cell's plasma membrane [92, 93]. CCR5 inhibitors, and the fusion inhibitor class of antiretroviral drugs target this step of the viral life cycle.

Co-receptor usage determines cell tropism, and CCR5 is usually required for the establishment of primary infection. Over the course of infection, co-receptor usage may change, and there is a switch towards CXCR4 in a large proportion of patients [94]. The presence of CXCR4-tropic viruses in a patient is indicative of late-stage disease and faster progression [95]. Co-receptor usage by HIV-2 is considerably broader than that of HIV-1, and use of the chemokine receptors CCR1, CCR2b, CCR3, CCR4, CCR5, CCR8, CXCR4, CXCR6, and GPR15 has been recorded [41–43].

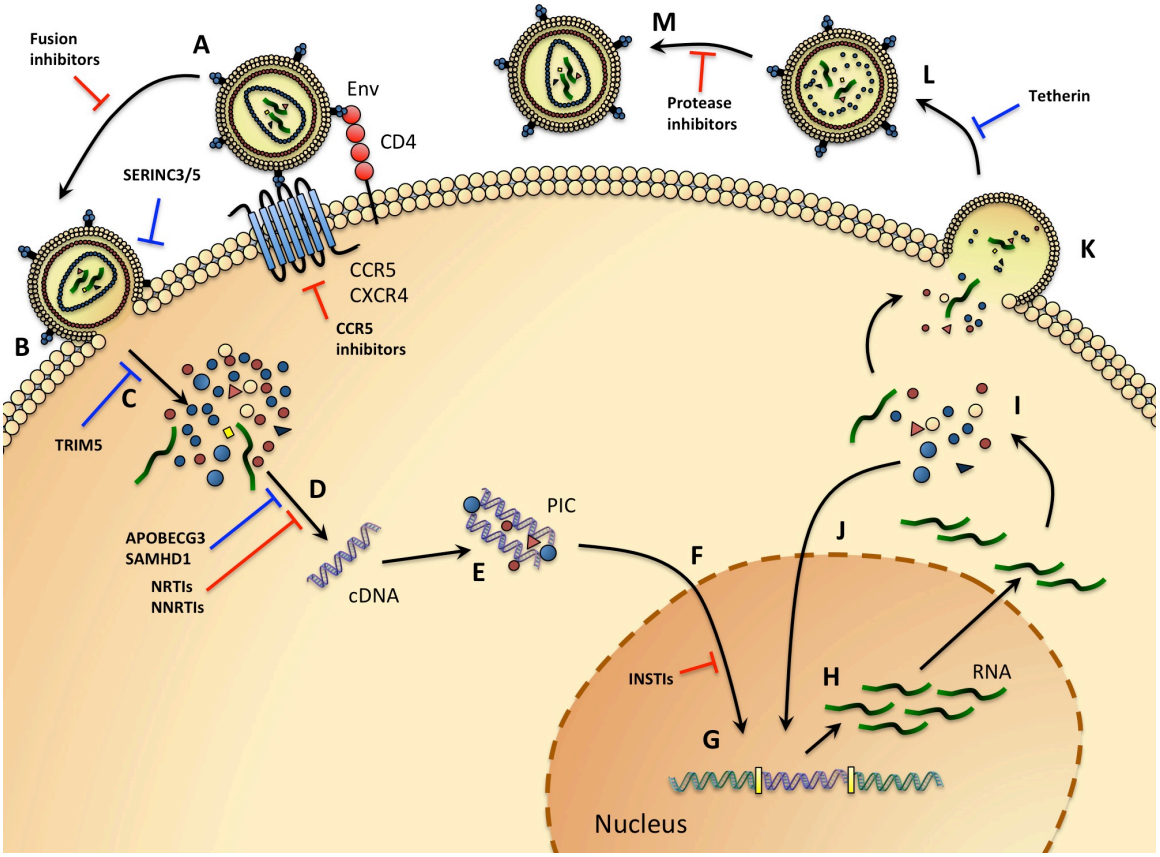


Figure 1.2: The life cycle and replication of HIV A) The viral envelope glycoprotein binds with CD4 on the surface of the target cell, as well as either the transmembrane CCR5 or CXCR4 co-receptor. **B)** This permits fusion of the virion with the cell's plasma membrane. **C)** Viral uncoating releases the contents of the virion into the cell's cytoplasm. **D)** The viral RNA is here reverse transcribed into cDNA. **E)** The viral cDNA is incorporated into the pre-integration complex PIC, **F)** which is transported into the cell's nucleus through nuclear pores. **G)** Once inside the nucleus, the cDNA is incorporated into the host cell's own DNA forming the proviral reservoir. **H)** The integrated viral genome is then transcribed - along with host genes - by host RNA polymerase II. **I)** The synthesised RNA is exported from the nucleus, and acts as mRNA to be translated into viral proteins, **J)** some of which are subsequently transported back to the nucleus, where they play a role in orchestrating further gene expression. **K)** Full-length viral RNA as well as the newly-synthesised viral proteins are incorporated into budding virions, **L)** which are released from the cell surface, **M)** and undergo further protease-mediated modifications to form mature and infectious virions. Steps targeted by antiretroviral drugs (red) and host restriction factors (blue) are indicated. Figure re-drawn and adapted from Engelman & Cherepanov [91].

1.1.6.2 Entry and Uncoating

The conformational change of the fusion peptide permits membrane fusion and entry of the virus into the target cell (figure 1.2B). Once inside, the viral core - composed of the capsid protein (section 1.1.8.2) - is disassembled, releasing the contents of the virion into the cytoplasm (figure 1.2C). This uncoating process is facilitated by dynein and kinesin 1 [96, 97], as well as phosphorylation of the capsid protein [98]. It is also inhibited by the primate restriction factor TRIM5α (section 1.2.2.7).

1.1.6.3 Reverse Transcription

The released RNA is reverse-transcribed into DNA by the viral enzyme reverse transcriptase (section 1.1.8.3), following formation of a reverse transcription complex (RTC) during the uncoating step (figure 1.2D) [99]. Reverse transcription is a multistep process, which takes place in the cytoplasm of the infected cell. It is initiated through priming of the viral RNA with a transfer RNA (tRNA) molecule that attaches to a primer binding site in the U5 long terminal repeat (LTR). Reverse transcriptase then catalyses initiation and extension of single-stranded cDNA from the binding site, replicating the LTR. At this point, the replicated fragment of RNA is degraded by the RNAse H domain, permitting dissociation of the newly synthesised cDNA.

The free cDNA then effectively jumps to the 3' LTR where it re-attaches to the RNA and continues extending in the 5' direction to synthesise the – strand. RNAse H again degrades the replicated RNA, with the exception of a small region in the polypurine tract (PPT) which is then used as an RNA primer in the synthesis of the complementary + strand. As RNAse H degrades the RNA, the products form further primers which adhere to various regions of the – strand DNA allowing much of the genome to be synthesised. All primers are removed, and the DNA circularises

owing to the complementary ends. Finally, reverse transcriptase repairs and ligates any gaps in the + strand and completes synthesis of the LTRs. The circular DNA linearises, giving rise to the complete viral genome. This process is reviewed in detail by Sarafianos *et al.* [100].

The reverse transcription step is targeted by two classes of antiretroviral drugs: nucleotide reverse transcriptase inhibitors (NRTIs) and non-nucleotide reverse transcriptase inhibitors (NNRTIs). NRTIs mimic dNTPs but lack a 3'-hydroxyl group. They are incorporated into the nascent DNA, effectively acting as chain terminators. NNRTIs allosterically inhibit reverse transcriptase. The reverse transcription step is also inhibited by the host restriction factors APOBEC3G & SAMHD1 (section 1.2.2.7).

1.1.6.4 Nuclear Import and Integration of Provirus

Once the DNA has been synthesised, it is incorporated into the pre-integration complex (PIC), along with matrix (section 1.1.8.2), integrase (1.1.8.3), viral protein regulatory (Vpr; 1.1.8.8), and host proteins (figure 1.2E). The PIC is then actively transported through nuclear pores to the cell's nucleus by a process that is cell-division-independent [101], and requires the nuclear localisation signal (NLS) of Vpr [102] (figure 1.2F).

The viral DNA is then incorporated into the target cell's DNA, mediated by the enzyme integrase. Integrase catalyses cleavage of the LTRs to expose 3'-hydroxyl groups, which in turn are used by integrase to sever chromosomal DNA. The corresponding ends are joined together, and then ligated by the target cell's own repair mechanisms [91] (figure 1.2G). The integrated "proviral DNA" is highly stable, and tends to be incorporated at sites of active transcription.

The integration step is targeted by the integrase inhibitor class of antiretrovirals, including the multifunctional integrase strand transfer inhibitors (INSTIs).

1.1.6.5 Production of Viral Proteins

Integrated provirus functions as a cellular gene that can be expressed and regulated by normal nuclear machinery, and as such, can be either active or latent: this latency is one of the major challenges in the successful eradication of HIV. Viral transcription is initiated from the U3 promoter of the LTR (section 1.1.8.1) by RNA polymerase II (figure 1.2H), and occurs in two distinct phases. Initially, spliced transcripts of varying lengths are produced at levels roughly proportional to the local basal transcriptional activity, the smallest of which are exported from the nucleus and are translated to produce the early viral proteins including negative factor (Nef; section 1.1.8.7), transactivator of transcription (Tat; 1.1.8.5) and regulator of expression of virion proteins (Rev; 1.1.8.6; figure 1.2I).

Once sufficiently accumulated in the cell, Tat and Rev are transported to the nucleus (figure 1.2J) where they orchestrate further gene expression and allow production of the late stage proteins. Tat recruits positive transcription elongation factor b (P-TEFb) [103], increases the rate of RNA synthesis, and is essential for efficient elongation of mRNA allowing full-length transcripts to be synthesised; whilst Rev facilitates export of larger partially-spliced mRNA transcripts [104].

1.1.6.6 Budding, Release and Maturation

The newly-synthesised viral (poly)proteins - as well as two copies of the full-length genomic RNA - are assembled into virions (figure 1.2K), in a process mediated by the p55 group-specific antigen (Gag) precursor polypeptide assemblin (section 1.1.8.2) [105] which promotes membrane-targeting and evagination. Budding requires cellular

class E vacuolar protein sorting proteins (VPSs) to pinch off the nascent particles from the plasma membrane [91] (figure 1.2L). This step is inhibited by the host restriction factor tetherin (section 1.2.2.7).

Immature virions must finally undergo protease-mediated maturation to become infectious viral particles (figure 1.2M). This step involves the proteolysis of p55 assemblin and the Gag/Pol polyprotein [106]. The viral enzymes are now able to undergo structural remodelling to fold into their mature functional forms; and core assembly can proceed [107], so as to produce a stable unit.

The protease inhibitor (PI) class of antiretrovirals prevent virion maturation by competitively binding and occupying the active site of the protease enzyme. Resistance is often developed against these drugs, but they can be used effectively as part of combination therapy with other antiretrovirals such as NRTIs [108].

1.1.7 The Structure of a Mature Virion

HIV is an enveloped RNA virus of the *Retroviridae* family, measuring approximately 145nm in diameter [109]. Two copies of single-stranded RNA (ssRNA) comprise the viral genome (section 1.1.8). These are packaged within a fullerene conical core of hexameric (approximately 200-250) and pentameric (exactly 12) capsid protein units [110]. The capsid is contained within a lipid bilayer, derived from the plasma membrane of the host cell from which it was budded. Immediately below the membrane is a thin layer of the viral protein matrix, and studded throughout the membrane are approximately 14 heteromeric spikes composed of the envelope glycoproteins gp41 and gp120 [111]. The structure of the mature virion is shown in figure 1.3.

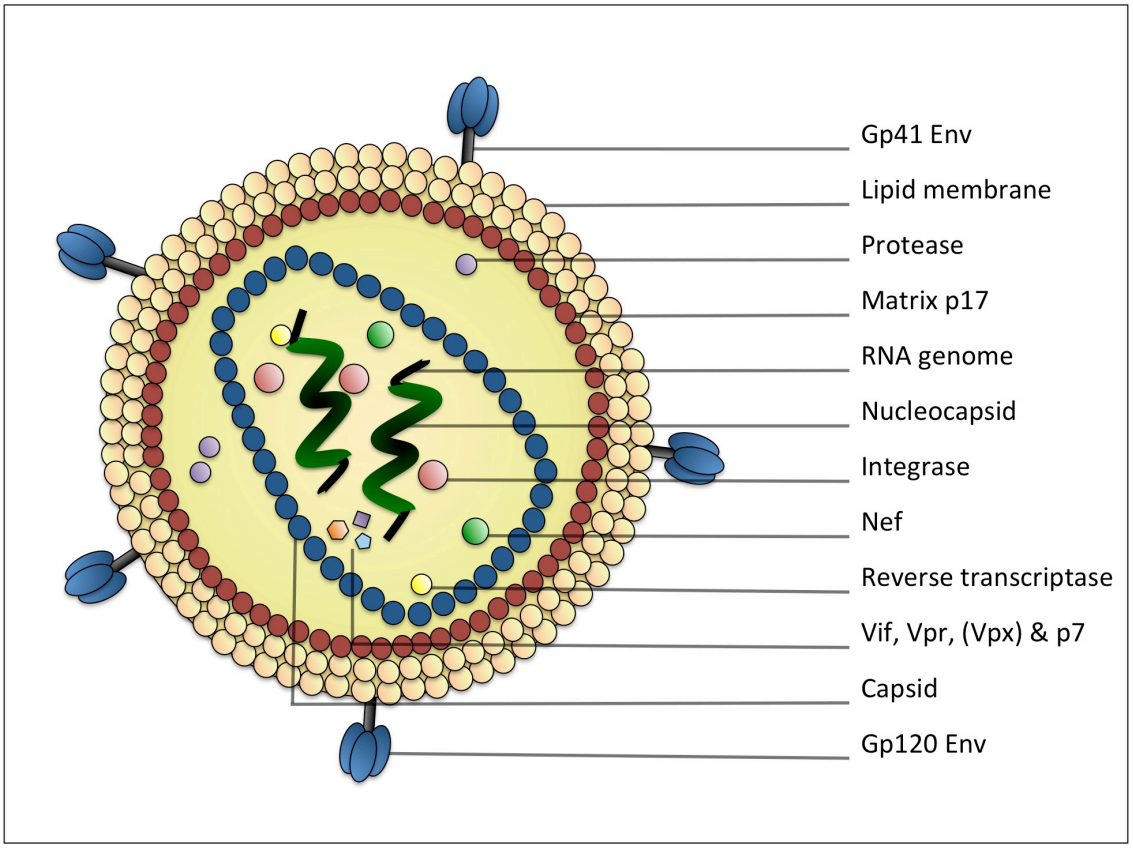


Figure 1.3: The structure of a mature HIV virion The viral particle consists of a fullerene conical core composed of the capsid protein which encases two copies of the single-stranded RNA genome, and a number of viral proteins, including nucleocapsid, which is tightly associated with the RNA. The capsid core is itself encased in a host-derived lipid membrane, lined with a thin layer of matrix protein. On its surface are a number of 'spikes', composed of the envelope glycoproteins gp41 and gp120. The mature virion also contains the enzymes protease, integrase, and reverse transcriptase; as well as the accessory proteins Nef, Vif, Vpr, and p7. Additionally, HIV-2 virions contain Vpx.

In addition to the structural proteins, a number of other viral proteins are encased within the virion, including nucleocapsid (NC) - which is in tight association with the genomic RNA; and the enzymes protease (PR), integrase (IN) & reverse transcriptase (RT). Several accessory proteins are also incorporated, including negative factor (Nef), viral infectivity factor (Vif), viral protein regulatory (Vpr), and p7; whilst HIV-2 virions also contain viral protein X (Vpx). Studies have not demonstrated the presence of the remaining mature viral proteins and peptides in mature virions in-

cluding transactivator of transcription (Tat), regulator of expression of virion proteins (Rev) and viral protein unique (Vpu) in HIV-1. Instead, these exert their function within the intracellular phase of the viral life cycle.

During the budding process, various host proteins are taken up and incorporated into the maturing virion by both HIV-1 and HIV-2. Although differences have been found between the protein repertoires acquired by each of these viruses [112], they are conserved across HIV-1 virions budded from different cell types [113].

1.1.8 The Viral Genome

For such a complex virus, the genome of HIV is surprisingly compact, with each copy of its RNA genome being around 9,500 bases in length for HIV-1, and 10,500 for HIV-2. There are nine open reading frames (ORFs), flanked by LTRs at each end (figure 1.4), and the virus encodes fifteen mature proteins. The LTRs play an important role in the control of gene expression. The genome encodes three large polyproteins conserved across the *Retroviridae*: Gag, Pol, and Env, which are cleaved and processed into their respective mature functional proteins. Additionally, the virus encodes six accessory proteins: Vif, Vpr, Nef, Tat, Rev, and Vpu/Vpx; with various roles in increasing viral fitness and survival. The genomes of HIV-1 and HIV-2 are structurally similar; the major difference being that HIV-1 expresses the accessory protein Vpu whilst HIV-2 expresses Vpx.

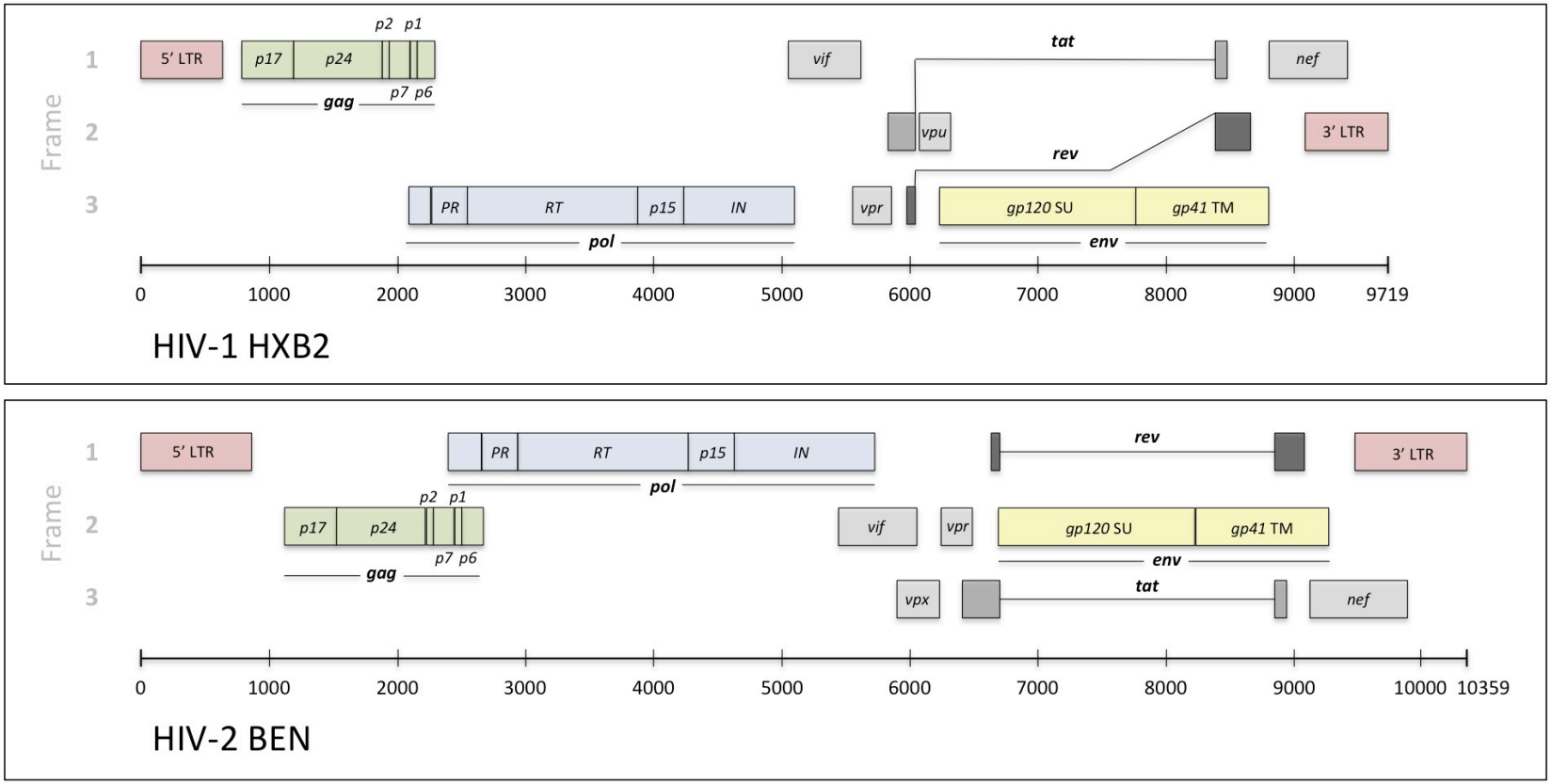


Figure 1.4: The genomes of HIV-1 and HIV-2 The reference strains HXB2 and BEN have been used for HIV-1 and HIV-2 respectively. Open reading frames are denoted by rectangles, and size corresponds to gene length in bases. The spliced (*tat* and *rev*) exons are indicated by corresponding shaded blocks. Figure re-drawn and adapted from The Los Alamos National Laboratory (LANL) [45].

1.1.8.1 Long Terminal Repeats

The LTRs of HIV are synthesised as a result of the retroviral reverse transcription process. The LTRs consist of a U3, R and U5 region. Both U3 and U5 contain a large number of transcription factor binding sites, and can also interact with some virus-encoded proteins. The LTRs play a part in numerous processes, including regulating transcription and integrating viral double-stranded DNA (dsDNA) into the host genome. The 5'-LTR acts as a promoter for RNA polymerase II in the transcription of the retrovirus; whilst the 3'-LTR is important for transcriptional termination and polyadenylation [114]. Disease progression is associated with alterations in LTR promoter activity in some individuals [115].

1.1.8.2 *Gag*

The group-specific antigen (*gag*) gene - found in all retroviruses - encodes structural proteins, and is indispensable for viral viability. It codes for a 55 kDa myristoylated precursor protein - p55 (assemblin) - which is processed by viral protease into the intermediate proteins p41 and p15. P41 is further processed into p24 capsid (CA) and p17 matrix (MA); whilst p15 is processed into p7 nucleocapsid (NC), p6 and p1 [116].

Capsid, the largest of the mature *gag* proteins, acts as the building block of the viral core, forming hexamers and pentamers, which assemble into a fullerene cone structure, surrounding the viral genome [110]. The matrix protein is also a structural component of HIV, forming a shell that lines the inner surface of the virion membrane. Both CA and MA also have functional roles, and are important for viral assembly and maturation [116]. The basic nucleocapsid protein p7 has several functions in the viral life cycle: it is particularly important in packaging genomic RNA, and in facilitating reverse transcription and integration due to its nucleic acid chaperone activity [117].

It has been proposed that although NC is not essential for successful infection, it vastly enhances retroviral survival and productivity [118]. The p6 protein is important for release of mature infectious particles, and also acts as an important docking site for binding by cellular and viral factors [119]. P6 is the most variable *gag* product in length and sequence, and is highly mutable with the exception of the LXXLF motif required for incorporating Vpr into HIV-1 virions [120], and the P(T/S)APP motif important in viral budding [121]. Finally, the spacer peptides (SPs) p1 and p2 supposedly play a role in viral maturation, but their functions are not currently well understood. It has been proposed that the spacer peptides may regulate proteolytic processing by retroviral protease [106].

1.1.8.3 *Pol*

The *pol* region of the genome encodes three important viral enzymes: protease (PR), integrase (IN), and reverse transcriptase (RT). The first of which - protease - is a homodimeric aspartyl enzyme essential for the specific cleavage of newly synthesised viral precursor proteins such as the Gag and Gag/Pol polyproteins, and is required for formation of functional infectious virus [122, 123]. Cleavage of Gag/Pol is required for the conformational changes underpinning viral maturation to occur. Protease-deficient viruses are non-infectious [122], and so protease has been extensively studied as a drug target, with protease inhibitors forming an important arm of antiretroviral drug therapy.

Reverse transcriptase is another retroviral enzyme contained within the *pol* region. Its function is to catalyse the reverse transcription of viral RNA into double-stranded DNA, which can then be integrated into the host cell genome. It is exceedingly error prone as it lacks exonucleolytic proof-reading ability (section 1.1.2). It also contains

a C-terminal RNase H domain (p15), which is required for cleavage of the DNA/RNA hybrid intermediates formed during reverse transcription [124].

The final enzyme encoded within HIV *pol* is integrase. The role of integrase, as its name suggests, is to mediate integration of viral cDNA into host cell chromosomes. Integrase is present at about 40-100 copies per virus particle [125], and is indispensable for viral replication as HIV must be incorporated into host DNA for transcription to occur. Integration can occur throughout the host genome, but is most favoured at sites of active transcription - especially those containing transcription-associated histone modifications [126] - which promotes high levels of viral gene expression.

1.1.8.4 *Env*

The envelope (*env*) region of the genome - much like *gag* - encodes structural components of the virus, this time in the form of surface-expressed glycoprotein spikes. A 160 kDa glycoprotein (gp160) is produced which is proteolytically cleaved into two smaller proteins: the 120 kDa surface protein gp120 (SU), and the 41 kDa transmembrane protein gp41 (TM). Gp120 and gp41 are bound noncovalently into trimers and localise to the surface of the virion, where their primary function is to mediate entry into human cells. This is achieved through binding to the CD4 receptor and chemokine co-receptors including CCR5 and CXCR4. As an exposed protein complex, envelope is accessible to antibodies and undergoes rapid mutation to avoid recognition [127], but is also heavily glycosylated. Furthermore, changes in the glycosylation of the protein aid evasion of the immune system, without impacting on receptor binding [128].

1.1.8.5 *Tat*

The *tat* gene encodes the transactivator of transcription protein, which is important for efficient transcription of HIV dsDNA [129]: indeed it increases the production of viral mRNAs by around 100-fold [130, 131]. This is achieved by recruiting host cellular factors, which promote the formation of a highly efficient transcription complex, leading to enhanced RNA polymerase II processivity. Additionally, Tat blocks premature termination of transcription, encouraging the production of full-length transcripts [131].

1.1.8.6 *Rev*

The *rev* gene encodes the regulator of expression of virion proteins: an RNA-binding protein that is involved in the export of viral RNA from the nucleus to the cytoplasm [132]. It is expressed early on, and results in a three to twelve fold increase in the level of cytoplasmic RNA [133]. This process is essential, as it permits un- or partially-spliced RNA to act as a template for gene expression, expanding the number of proteins that can be produced from such a compact genome. It effectively acts as a selective shuttle to export intron-containing RNA in the late phase of the viral life cycle. Export is dependent on recognition of and binding to a cis-acting Rev-response element (RRE) [104], which is encoded within the *env* sequence of unspliced/partially spliced RNA. Binding reveals a nuclear export signal [134], triggering the formation of a complex with various nuclear export factors. Once in the cytoplasm, translation can occur, giving rise to the full repertoire of viral proteins.

1.1.8.7 *Nef*

Negative factor (Nef) is a small (27-35 kDa), myristoylated, non-structural accessory protein encoded at the 3' end of the retroviral genome, overlapping the LTR. Early studies of the *nef* gene showed that deletions resulted in highly cytopathic viruses,

with approximately five-fold higher levels of replication, suggesting an important role in attenuating viral replication [135, 136]. The name negative factor was coined shortly after.

However, further studies *in vitro* not only failed to support this finding [137, 138], but also showed that the presence of functional Nef proteins promoted more efficient growth in certain cell lines [139]. Nef is now known to accelerate disease progression through a number of mechanisms and makes an important contribution to immune evasion. Indeed, *in vivo* studies have shown that it is important in viral replication, and defective or absent *nef* is associated with long-term non-progression in some HIV-1 infected individuals [140–143].

The Nef protein is expressed at high levels early in infection [144, 145], with evidence of transcription from pre-integrated DNA [146]. It forms stable homo-dimers and homo-trimers in solution, which are important for many of its functions [147]. Within the cell, the protein is localised to the cytoplasm, nucleus, and paranuclear region [148].

Nef is arguably the most multifunctional retroviral protein, with several roles described, including mediation of cell activation and signalling, regulation of apoptosis, enhancement of viral infectivity, and modulation of the expression of host receptors on the surface of the cell. In addition, SIV - but not HIV-1 - Nef proteins function as antagonists of the host restriction factor tetherin (section 1.2.2.7) [149].

A core function of Nef is the manipulation of host cell machinery to modulate signalling pathways in infected cells. This brings about a series of changes that alter the intracellular environment to one which is more supportive of viral growth. Modulation

of the T cell receptor (TCR) signalling pathway appears to be particularly important, with studies showing that Nef enhances TCR signalling through mechanisms which include modulation of the immunological synapse (IS), actin remodelling, and dysregulation of signalling molecule trafficking [150–154]. As well as this, myristoylated Nef enhances the early activation of nuclear factor kappa B (NF-κB), which is required for transcription of the integrated retroviral genome [155].

Nef also triggers bystander cell death by promoting the expression of the fas ligand on the surface of infected cells [156, 157]. This interacts with the fas receptor in neighbouring cells and forms the death-inducing signalling complex (DISC), which then activates the apoptotic signalling cascade. It has more recently been found that Nef is also released in exosomes, which can themselves trigger apoptosis [158]. Taken together, Nef-induced bystander cell death is understood to be an important immune evasion mechanism, and the resulting depletion of $CD4^+$ bystander cells has been proposed to contribute directly to disease progression [159].

The enhancement of infectivity is a more poorly understood function of Nef. Studies have shown that the efficiency by which a virus establishes productive infection is higher in viruses encoding functional Nef protein, relative to those where it is absent or defective. Enhancement of infectivity is independent of viral co-receptor tropism [160], and the extent of the enhancement is variable: typically reported as around a four to forty fold increase in infectivity per single round of infection [161–164]. The effect appears dependent on the presence of Nef during the production of nascent virions. This has been demonstrated in studies where the infectivity of *nef*-defective viruses was restored by the ectopic expression of Nef in the infected cells, only when the defective virion had itself been produced in the presence of functional Nef [163, 165, 166]. Two recent studies have shed light on this enigma by reporting that Nef

counters the host restriction factors SERINC3 and SERINC5 (section 1.2.2.7) which inhibit penetration of infecting virions into permissive cells [167, 168]. Nef prevents incorporation of SERINC3/5 into nascent virions forming at the plasma membrane of infected cells.

Additionally, lentiviral Nef plays an important role in mediating the downregulation of a wide array of surface-expressed receptors, including but not limited to CD4, HLA-I, CXCR4, CD3 (HIV-2 only), CD28, SERINC3/5, and CD8$\alpha\beta$ [169–173]. By far the most extensively studied of these receptors are CD4 and HLA-I. Nef also upregulates the invariant chain Ii [366].

The CD4 receptor is used by primate lentiviruses to gain entry into cells. Following infection, Nef downregulates CD4 from the plasma membrane to prevent further viruses binding and entering, which would lead to super-infection. Super-infection unduly harms the cell and is of no benefit to the virus [174]. CD4 downregulation also prevents the inhibition of budding virion release at the cell surface [175]. Nef-mediated CD4-downregulation is important for viral pathogenicity, and viruses defective in this function show greatly reduced fitness [176, 177]. Indeed, Nef from late infection mediates more efficient CD4 downregulation than does early Nef, suggesting that this property is selected for over the course of disease progression [178].

Nef achieves CD4 downregulation by functioning alongside adaptor proteins to hijack host signalling pathways. However, the specific mechanism depends on the type of cell infected as basal trafficking of CD4 differs between lymphoid and myeloid cells. In myeloid cells, CD4 is constitutively internalised from the cell surface and rapidly recycled, whilst in lymphoid cells, the lymphocyte-specific tyrosine kinase $p56_{lck}$ retains CD4 at the cell surface [179–184]. Therefore, in myeloid cells, Nef

promotes intracellular retention and degradation of CD4, whilst in lymphoid cells, it must also actively drive endocytosis [185–187].

Finally, Nef mediates downregulation of class I human leucocyte antigen (HLA-I). These molecules present viral peptides (amongst others) to cytotoxic $CD8^+$ lymphocytes and natural killer cells, leading to the recognition and destruction of infected cells. As such, Nef-mediated downregulation of HLA-I is a particularly important immune evasion mechanism, and will be discussed in detail in chapter 4.

1.1.8.8 *Vpr*

Viral protein regulatory (Vpr) is a multifunctional retroviral accessory protein with a number of important roles in establishing infection of macrophages and $CD4^+$ lymphocytes. Its major role is regulating the import of the PIC into the nucleus, permitting integration of the viral genome into the host cell DNA, establishing the proviral reservoir [188]. This is particularly important in the context of non-dividing macrophages, where functional Vpr is essential for the establishment of productive infection [189, 190]. Additionally, Vpr co-activates - in combination with other regulatory elements - regions of the LTR to reactivate latently infected cells. Although the mechanism by which this occurs is currently unclear, multiple LTR binding sites have been identified as potentially important, as reviewed in Kogan & Rappaport [191].

Further roles of Vpr include induction of G_2 cell cycle arrest by activating phosphorylation of checkpoint kinase 1 (Chk1) [192], and regulation of NF-κB activity, which has a downstream suppressive effect on cytokine production, apoptosis, and immune activation [193].

1.1.8.9 *Vpu*

Viral protein unique (Vpu) is an accessory protein found in HIV-1 and closely related simian retroviruses such as SIV_{cpz} which infects chimpanzees. It is absent from the HIV-2 genome, which instead contains the *vpx* gene (section 1.1.8.10). Vpu is a membrane-associated phosphoprotein and contributes significantly to increased viral pathogenesis.

Much like Nef, Vpu reduces the density of important receptors on the surface of the infected cell: including HLA-I [194], CD4 [195], and potentially a slew of others including tetraspanins [173]. The benefits to the virus of reducing CD4 expression are at least threefold: it enhances the release of virions by alleviating CD4-mediated inhibition at the surface; it prevents viral superinfection of the cells; and it relieves gp120 retention in the endoplasmic reticulum (ER) by newly synthesised CD4. Vpu-mediated HLA-I downregulation helps the virus to evade cytotoxic lymphocytes and - in contrast to Nef - Vpu also selectively downregulates HLA-C [196]. Also in contrast to Nef, Vpu inhibits NF-κB activation to suppress innate immune responses later in infection [197].

The activity of tetherin (section 1.2.2.7) is also antagonised by Vpu [198]. Vpu mediates the degradation of tetherin by promoting ubiquitin-mediated proteasomal processing [199].

1.1.8.10 *Vpx*

Viral protein X (Vpx) is encoded by HIV-2 and most SIVs. It is a paralogue of Vpr, and is believed to have arisen from duplication of the *vpr* gene in an ancestral virus [200]. Whilst it shares a high degree of structural homology with Vpr, functional differences have arisen since its divergence. For example, Vpx has lost the ability to

induce G_2 cell-cycle arrest, but has gained an important role in the counteraction of protective host factors. Vpx recruits SAMHD1 to the cullin ring E3 ligase (CRL4) complex, targeting it for ubiquitination and proteasomal degradation [201]. It also interacts with APOBEC3A and the interferon regulatory factor 5 (IRF5) [202] to inhibit their functions, although the mechanisms by which these occur are not well understood.

1.1.8.11 *Vif*

The primary function of viral infectivity factor (Vif) is to counter the host restriction factor APOBEC3G (section 1.2.2.7). Vif targets APOBEC3G for ubiquitin-dependent degradation [203, 204], and also interferes with its expression [205] and packaging into nascent virions [206].

1.2 The Immune Response to HIV Infection

The immune system has played an important role in shaping the life cycles and genomes of retroviruses: multiple viral accessory proteins have evolved to mediate immune evasion (section 1.1.8) in the arms race which underlies the unique pathology of HIV infection. Whilst the overwhelming majority of untreated HIV-1 infections ultimately result in AIDS, the initial suppression of the virus following acute infection, and the characteristic extended period of clinical latency are attributable to a well-orchestrated immune response mounted by the host. Why this response eventually breaks down is not yet known, but understanding the processes involved in HIV control and the mechanisms used by the virus to escape is imperative to developing appropriate vaccines and treatments.

1.2.1 Mucosal Barriers to Infection

The first line of defence against sexually-acquired infection is the genital mucosa. Acidity, mucus, and the structural integrity of epithelial junctions form a physical protective barrier against infection, reviewed by Borrow *et al.* [207]. Few virions are able to traverse this barrier, as demonstrated by the finding that a single sexual exposure is often insufficient for infection to occur, and that quasispecies isolated from patients with acute productive infection can often be traced back to a single transmitting founder virus [208]. Similar findings have been reported in the macaque model: on intrarectal inoculation with SIV_{sm} or SIV_{mac}, invasive infection is established from a single founder virus [209], and whilst several small foci of infection can form beneath the mucosa, very few become systemic [210].

Secreted by the mucosa are a number of small proteins and cationic peptides such as elafin, secretory leucocyte protease inhibitor (SLPI), and select defensins which

exhibit antiviral properties and provide an added layer of defence [211–213]. However, there is also increasing evidence that further secreted molecules - such as semen-derived enhancer of virus infection (SEVI), and whey acidic protein four-disulphide core domain 1 (WFDC1) - may have a role in enhancing viral infection [214–216]. Such peptides act through various mechanisms to exert their permissive or restrictive effects, including activation of signalling pathways, competitive binding of cellular receptors, regulation of cytokine production, and direct interaction with the virion itself. The genital mucosa is a complex microenvironment: a simple schematic overview of which is shown in figure 1.5. Delineating the mechanisms by which the mucosa is able to either support or restrict infection should be of high priority.

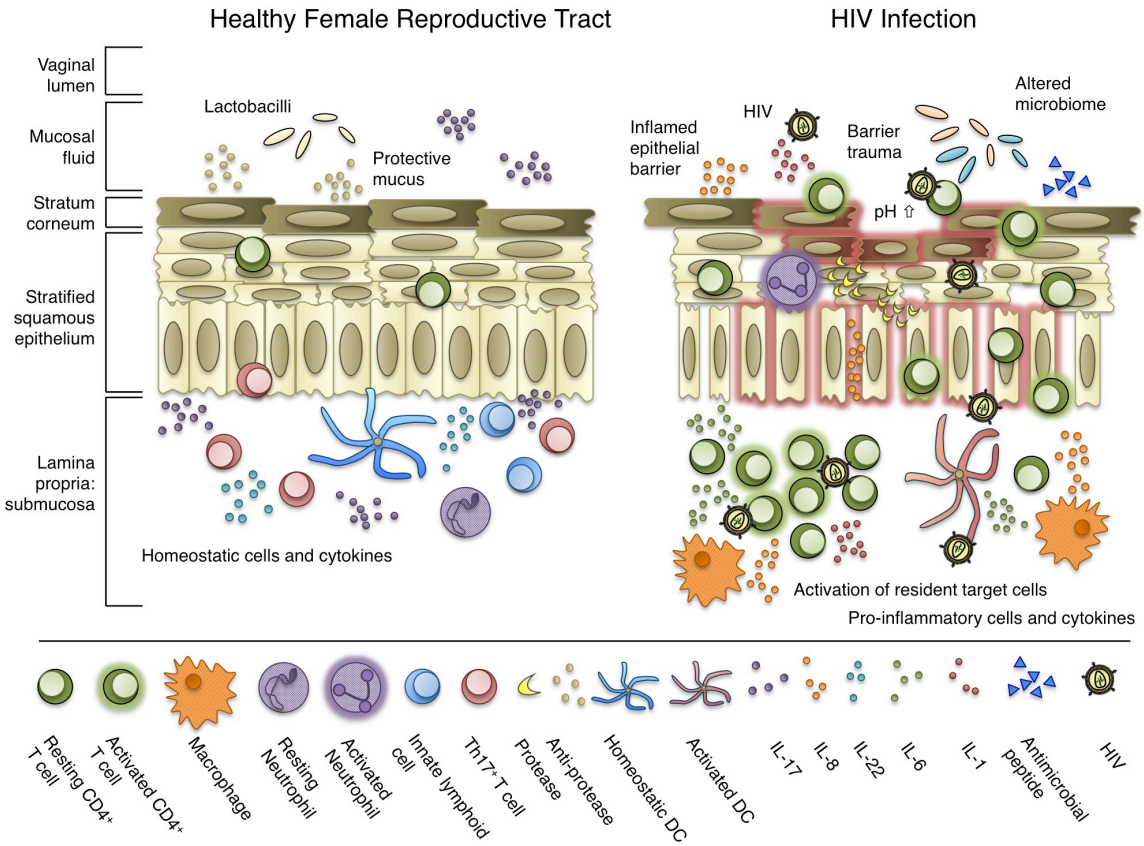


Figure 1.5: The mucosa of the female genital tract in health and in HIV infection. Figure re-drawn and adapted from Burgener *et al.* [217].

1.2.2 Innate Immune Response

Robust early responses against HIV infection are essential for regaining control of the virus, establishing the viral set-point, and are predictive of clinical prognosis and disease progression. Once the virus has successfully breached the mucosal barrier and established foci of infection, the innate arm of the immune system becomes a particularly important line of defence.

1.2.2.1 Pattern Recognition Receptors

The innate response is rapid, memoryless, and antigen non-specific. It functions instead through recognition of pathogen-associated molecular patterns (PAMPs), which are highly-conserved molecular motifs specific to pathogenic organisms, but notably absent from healthy host cells. The motifs are found in structures essential to the pathogen's survival or infectivity, with common examples being lipopolysaccharides, glycoproteins, and proteoglycans. PAMPs are recognised by pathogen recognition receptors (PRRs): pathogen-sensing proteins expressed by cells of the innate immune system, particularly antigen-presenting cells (APCs).

Human APCs such as dendritic cells and macrophages express a broad repertoire of PRRs, including the Toll-like receptors (TLRs), nucleotide-binding oligomerisation domain (NOD) protein-like receptors (NLRs), retinoic acid-inducible gene 1 (RIG-1)-like receptors (RLRs), and C-type lectin-like receptors (CLRs). Each type of receptor has a related but fundamentally distinct role in protecting the cell from pathogens, reviewed by Akira *et al.* [218]. Whilst there is much controversy surrounding the specific PRRs involved in HIV immunology, the most important are currently understood to be the TLRs, and the CLR dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin (DC-SIGN) [219].

TLRs are transmembrane receptors that can be found both on the cell surface and within the endosomal membrane. Humans express ten known TLRs, which are differentially expressed and recognise disparate pathogen motifs. The TLRs canonically associated with recognition of viral structures are TLR3 (dsRNA), TLRs 7 & 8 (ssRNA), TLR9 (DNA), and TLRs 2 & 4 (surface glycoproteins). Many TLRs have been implicated as having a role in priming the immune response against HIV-1 (especially TLRs 7 & 8), but findings are often contradictory [219–222].

The downstream effects of TLR triggering are manifold, resulting in the activation of pro-inflammatory signalling pathways involving nuclear factor kappa B (NF-κB), mitogen-activated protein kinase (MAPK), and interferon regulatory factors (IRFs) 3 & 7. This culminates in the release of chemokines and cytokines including tumour necrosis factor alpha (TNF-α), interleucin 1 beta (IL-1β), and a variety of interferons (IFNs), reviewed by Mogensen *et al.* [223]. The consequences of the release of these factors can be either antiviral and immunopathological.

In addition to TLRs, DC-SIGN is present on dendritic cells and macrophages. DC-SIGN binds mannose-containing glycoproteins, such as gp120 envelope and triggers phagocytosis [224, 225].

1.2.2.2 Dendritic Cells

Whilst "innate" and "adaptive" responses are frequently presented as distinct concepts for clarity, the reality is much more complex. Many cells plays both innate and adaptive roles in immunity, with professional antigen-presenting cells (APCs) being key examples.

Dendritic cells (DCs) are APCs which detect and capture antigens before presenting them on their surface to cells of the adaptive immune system. They are able to prime naïve T lymphocytes and as such are an important link between the innate and adaptive immune response. DCs are widely distributed throughout the body, and are broadly divisible into two groups based on their lineage and surface markers: conventional dendritic cells (cDCs) and plasmacytoid dendritic cells (pDCs) [226]. Langerhans cells are immature, specialised cDCs found in the genital mucosa and skin, and are of particular interest as some of the earliest cells to encounter HIV on sexual exposure [227].

Antigen presentation by cDCs is a well-characterised process. Immature, circulating, bone-marrow-derived precursors migrate to peripheral tissues - such as the mucosa - where they reside until they detect and capture antigens via phagocytosis, macropinocytosis, and various PRRs (section 1.2.2.1). Antigen capture leads to activation of intracellular signalling cascades promoting maturation of the DC. Maturation triggers large changes to the cell's morphology and receptor expression profile favouring antigen presentation, reviewed by Reis e Sousa [228]. Key changes include downregulation of phagocytic receptors, upregulation of the major histocompatibility complex (MHC) and costimulatory molecules, secretion of IL-12, and upregulation of chemotactic receptors such as CCR7 which triggers migration to the spleen and lymph nodes. Here, the DC can induce IL-12-mediated Th1 lymphocyte differentiation (section 1.2.4).

Whilst cDCs are essential for priming the adaptive immune response against HIV, they have also been shown to promote viral dissemination by facilitating transinfection of uninfected $CD4^+$ lymphocytes. The DCs capture viral particles through the CLR

DC-SIGN, and transmit them to the $CD4^+$ cells either at the site of infection or in the lymphoid tissue following maturation [229, 230].

pDCs are rarer than cDCs, making up only 0.3-0.5% of peripheral blood mononuclear cells (PBMCs) [231]. In addition to antigen presentation, they can mediate further innate responses by producing high levels of type-I interferon (IFN) [232]. pDCs recognise viral nucleic acids through TLRs 7 & 9 expressed in the endosomal compartment [233, 234], and have been shown both to be activated and suppressed by HIV [235–237]. pDCs also release indoleamine 2,3-dioxygenase (IDO) on activation which promotes differentiation of T regulatory lymphocytes (Tregs), potentially suppressing the adaptive response against HIV [238, 239].

Whether or not pDC activation is beneficial to the host remains controversial [240, 241], and the absence of viral mechanisms to counter this suggests activation supports HIV survival or replication. In an attempt to resolve apparent contradictions in recent studies, it has been hypothesised that initial pDC responses may be host-protective, followed by a switch to immunopathogenicity with disease progression [207].

1.2.2.3 Macrophages

Macrophages are monocyte-derived phagocytic cells that both engulf virus and function as professional APCs. They are also a major target of infection by HIV, and may play an important role in viral transmission, spread and pathology. On exposure to HIV, mucosa-resident macrophages contribute to the activation of the adaptive immune response by presenting viral antigens through the major histocompatibility complex (MHC).

However, macrophages of the genital mucosa are also permissive to HIV, and a small proportion may become productively infected [242]. The infected macrophages then facilitate the spread of virus through at least two mechanisms. Firstly, by releasing inflammatory chemokines and cytokines they recruit further immune cells to the site of infection, thereby dramatically increasing the local pool of susceptible cells [243–245]. Secondly, infected macrophages transmit the virus to target lymphocytes through direct cell-cell contact through the formation of a virological synapse [246], supposedly during MHC-mediated antigen presentation [247].

During clinically latent infection, macrophages also function as viral reservoirs, retaining HIV in intracellular compartments for up to several weeks [248]. These compartments protect the virus from antibodies and other immune effectors, allowing them to evade detection until release and further spread.

1.2.2.4 Natural Killer Cells

Natural killer (NK) cells comprise an important arm of the innate immune response. They are cytotoxic lymphocytes that recognise and kill virally infected cells through an antigen- and PRR-independent mechanism. Cytotoxicity is regulated by the receptors on the surface of NK cells, which can be either activating or inhibitory. The killer cell immunoglobulin-like receptors (KIRs) - which recognise host MHC molecules - are of particular importance in the context of HIV infection. Activating receptors help contain HIV replication, and have been associated with slower disease progression [249, 250].

NK cells are broadly divisible into cytotoxic and regulatory subsets based on their expression of the cell surface markers CD16 and CD56 ($CD16^{+}$ $CD56^{dim}$, and $CD16^{+/-}$ $CD56^{bright}$, respectively) [251]. Cytotoxic NK cells are the predominant form in

peripheral blood, whilst the regulatory form typically constitute no more than 10% and secrete large quantities of chemokines and cytokines, including IFNγ. In addition, a third rare subset of functionally anergic cells ($CD16^+$ $CD56^-$) are significantly over-represented in HIV viraemic patients, whilst the cytolytic subset is greatly reduced [252–254].

NK cells contribute to the control of HIV through cytolytic and noncytolytic mechanisms. Cytotoxicity is mediated directly through fas-mediated apoptosis and the release of perforin and granzyme-containing granules, or instead through antibody-dependent cell-mediated cytotoxicity (ADCC) [255, 256]. In ADCC, virally infected cells are lysed when NK cells recognise the Fc domain of bound antibodies via their FcR receptor, triggering degranulation [257]. Noncytolytic control is mediated through the release of chemokines and cytokines, which limit replication in infected cells, and block entry of the virus into target cells by limiting co-receptor access through competitive binding [258, 259].

The natural killer T (NKT) cells are developmentally and functionally distinct from NK cells. NKT cells are rare, innate-like lymphoid cells that express some NK cell markers in addition to (semi-)invariant T cell receptors. NKT cells are broadly divisible into CD4 positive and negative subsets: $CD4^+$ NKT cells secrete Th1 and Th2 cytokines (section 1.2.4.3); whilst $CD4^-$ NKT cells secrete primarily Th1 cytokines. The secretory profiles of these cells imparts function, with the $CD4^+$ cells having an immunoregulatory role and $CD4^-$ cells being more cytolytic, reviewed by Li & Xu [260]. The exact role of NKT cells in HIV infection is unclear. However, peripheral $CD4^+$ NKT cells are massively depleted during early infection at a faster rate than the greater peripheral $CD4^+$ lymphocyte pool, implicating a role in viral pathogenesis [261].

Over the course of HIV-1 infection, the virus subverts the NK response by: down-regulating class-I HLA from the surface of the infected cell by Nef and Vpu, blocking the synthesis of HLA-I, expressing HLA-I homologues, expressing cytokine-binding proteins and cytokine receptor antagonists, inhibiting NK cell activating receptors, and the direct infection of NK cells; reviewed in detail by Funke *et al.* [55].

1.2.2.5 Gamma Delta T Cells

Gamma-delta ($\gamma\delta$) T cells are thymus-derived lymphocytes that express on their surface an unusual T cell receptor composed of a γ and a δ chain as opposed to the α and β chains that make up the receptors on conventional T cells. Whilst systemically rare, $\gamma\delta$ T cells comprise up to half of the cells of the mucosa-associated lymphoid tissue (MALT), including the reproductive mucosa [262].

Much like APCs, $\gamma\delta$ T cells sit at the interface between the innate and adaptive immune response, although for different reasons. They are considered adaptive in that both the TCRγ and TCRδ loci undergo V(D)J somatic recombination to generate adaptive antigen-specific responses, similar to the α and β loci during development of the $\alpha\beta$ TCR [263]. There are two main classes of $\gamma\delta$ T cell, expressing either the first (Vδ1) or second (Vδ2) variable region of the delta locus. $\gamma\delta$ T cells are also considered to be innate-like in that tissue-resident cells express limited diversity [264] and their TCRs essentially function as PRRs, recognising PAMPs and activating rapid effector functions [265].

$\gamma\delta$ T cells are functionally diverse in the context of infection: they lyse infected cells, secrete immunoregulatory chemokines and cytokines, help B lymphocytes, prime $\alpha\beta$ T cells, and trigger DC maturation, reviewed by Vantourout & Hayday [266]. However, $\gamma\delta$ T cells appear to be of limited use in the control of HIV infection. In fact,

numerous studies have shown disruptions in the optimal ratios of different $\gamma\delta$ cell types in HIV-infected individuals relative to healthy controls [267–271]: largely due to expansion of Vδ1 cells and massive depletion of circulating Vδ2 cells. Vδ2 cell counts have previously been shown to correlate with $CD4^+$ lymphocyte levels, and inversely with viral load [272], and surviving Vδ2 cells become functionally anergic on infection with HIV [273].

1.2.2.6 Complement

The complement system consists of a tightly-regulated network of over 30 different small proteins and glycoproteins which play a major role in supporting - or complementing - the immune response against infections. Whilst the components that make up the complement system are themselves innate and non-adaptive, they functionally support both innate and adaptive immune processes [274], and they are often considered an arm of the humoral response. Broadly, the roles of complement in terms of antiviral activity are: the recruitment of inflammatory and phagocytic cells; the opsonisation of virions; and the lysis of virions & infected cells through formation of the membrane attack complex (MAC) - a lytic transmembrane pore [275].

Activation of the complement cascade can occur either through the classical pathway (mediated by binding of the C1q component to the pathogen itself or antigen-bound antibodies); through the alternative pathway (which is antibody-independent and occurs constitutively at low levels through hydrolysis of the C3 component); or through the mannose-binding lectin pathway (figure 1.6). All three pathways trigger the enzymatic cleavage of C3 into the fragments C3a and C3b, which themselves trigger further downstream events culminating in the antiviral effects described above. The classical activation pathway has been most heavily implicated as important in the context of HIV [277–280].

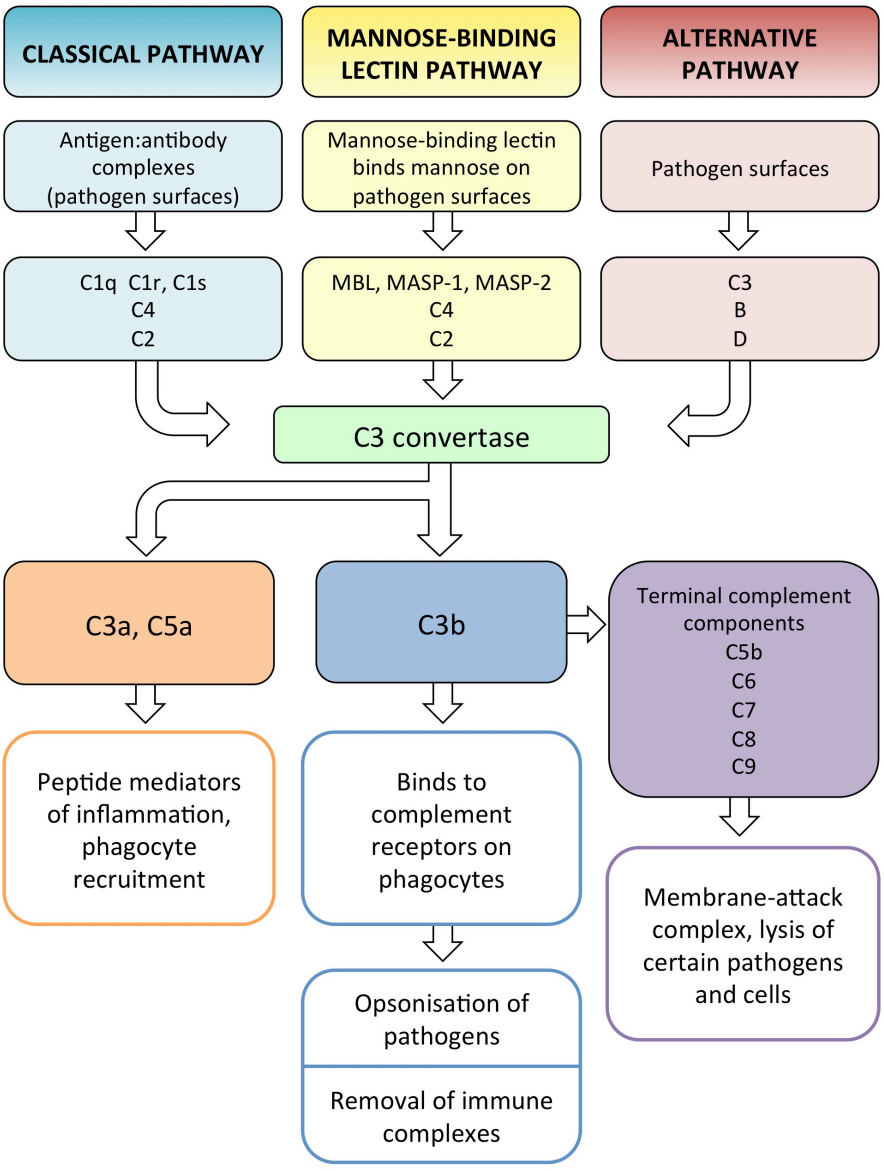


Figure 1.6: The activation pathways of the complement cascade. Figure re-drawn and adapted from figure 2-19 of Janeway *et al.* [276].

Despite evidence to support functional complement-mediated virolysis as well as neutralising antibody (NAb) enhancement *in vivo* [281, 282], efficiency in the control of HIV infection is restricted as the virus is able to subvert the complement system to suit its own purposes. The binding of complement components such as C3 & C5a to the virion surface promotes recruitment and interaction of HIV particles with uninfected target cells expressing corresponding complement receptors, including DCs and macrophages [283]. This leads to enhanced infectivity and transfer to $CD4^+$ lymphocytes [284]. In addition, the virus is able to evade virolysis by the MAC through recruiting complement regulators such as CD55 and CD59 during budding [285, 286], and also by binding factor H (fH) to its surface [287, 288].

1.2.2.7 Host Restriction Factors

Restriction factors are germline-encoded, constitutively-expressed host proteins that comprise an important first-line of defence against infecting virus, reviewed by Malim & Bieniasz [289]. They can be distinguished from other antiviral effectors on the basis that they are functionally independent and do not require cellular co-factors. Restriction factors are IFN-inducible and demonstrate broad protective activity against disparate species as a result of numerous historical interactions with countless independent pathogens. As such, their genes show evidence of strong positive selection in the form of high dN/dS ratios. Viruses necessarily develop antagonistic measures against restriction factors, and due to their substantially faster rate of evolution, it is not uncommon for restriction factors to be defective against contemporaneous virus [290].

Six restriction factors have so far been identified as potent inhibitors of primate retroviruses. Namely, tripartate-motif containing protein 5 (TRIM5) α, apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like G (APOBEC3G), SAM domain

and HD domain-containing protein 1 (SAMHD1), tetherin, serine incorporator 3 (SERINC3), and SERINC5.

TRIM5α is a cytoplasmic protein, and is the restriction factor that acts earliest in the viral life cycle, recognising and rapidly disassembling retroviral capsid [291]. This prevents controlled uncoating, abrogates reverse transcription and ultimately terminates replication. The efficiency of TRIM5α-mediated viral inhibition is highly variable and heavily species-specific. Generally, TRIM5α is unable to restrict viruses that naturally infect the species from which it is derived. For example, whilst simian TRIM5α is generally highly effective at restricting HIV-1, human TRIM5α only weakly inhibits HIV-1 *in vitro* [292, 293]. Conversely, HIV-2 shows much increased susceptibility to human TRIM5α [57], which could partially explain the disparity in pathology between HIV-1 and HIV-2 infection.

Both APOBEC3G and more recently SAMHD1 have been shown to restrict HIV through interference with the process of reverse transcription, albeit through different mechanisms. APOBEC3G possesses cytidine deaminase activity, and exerts an antiviral effect by catalysing the post-synthetic conversion of cytidine residues to uridine in the retroviral cDNA – strand during reverse transcription [294, 295]. This leads to catastrophic guanosine to adenosine hypermutation in the resulting viral genome. In addition, APOBEC3G interferes with the necessary transfer RNA priming [296] & elongation [297]. Conversely, SAMHD1 functions by hydrolysing deoxynucleoside triphosphates (dNTPs), in turn depleting the intracellular pool available for use by retroviral reverse transcriptases [298].

The interferon-inducible antiviral protein tetherin inhibits viral replication by retaining budding viral particles on the surface of the cell, preventing their release

[198, 299]. Tetherin associates with the plasma membrane of the infected cell, and becomes incorporated into the viral envelope of budding virions, forming a physical bridge bringing the two together [300].

Finally, both SERINC3 and SERINC5 have recently been shown to restrict HIV-1 infection [167, 168]. Although the mechanism by which this occurs is not yet well understood, it has been shown that SERINC3/5 are incorporated into budding virions at the cell surface, and subsequently interfere with infection of target cells.

Unsurprisingly, retroviral accessory proteins have evolved to actively inhibit the antiviral effects of host restriction factors, obscuring their contribution to viral control. APOBEC3G is antagonised by Vif (section 1.1.8.11) [301]; SAMHD1 by the HIV-2 protein Vpx (1.1.8.10) [302, 303]; tetherin by Vpu, Nef, or Env - dependent on the viral lineage (1.1.8.9, 1.1.8.7 & 1.1.8.4) [149, 198, 304], and SERINC3/5 by Nef (1.1.8.7). Conversely, the lack of evident retroviral countermeasure against TRIM5α indicates that this protein exerts greatly reduced selection pressure and may be of limited use against HIV. Considering these complex interactions, the overall contribution of restriction factors to the control of HIV infection is not well understood, and furthermore, it is widely believed that additional restriction factors remain to be identified.

1.2.3 Antigen Presentation and the Adaptive Response

Shortly after infection with HIV, strong cellular and humoral adaptive responses are elicited against the virus. The cellular response is driven by T lymphocytes; whilst B lymphocytes orchestrate humoral immunity. As adaptive processes, both of these are initiated in response to antigen. Two primary classes of molecules comprise the major histocompatibility complex (MHC), whose role is to present processed antigens to T

lymphocytes. Class I (MHC-I) plays an important role in presentation of endogenous antigens [305] - including those of both self and non-self origin - to $CD8^+$ cytotoxic lymphocytes (CTLs); whilst class II (MHC-II) present exogenous antigens to helper T cells. In humans, MHC is referred to as human leucocyte antigen (HLA).

HLA-I molecules are heterodimeric glycoproteins expressed on the surface of nucleated cells (figure 1.7). They contain a highly polymorphic membrane-spanning α-chain (α1, α2 and α3 domains), and a non-covalently linked constant β-2-microglobulin (β2M) chain: high levels of these in the plasma serve as a marker of immune activation. The membrane-adjacent region contains the β2m and α3 domain; whilst the distal region contains the α1 and α2 domains, which house the antigen-binding site. α1 and α2 are stabilised by hydrogen bonds, and form a crevice large enough to accommodate 8- to 10-mer peptides, or up to a 20-residue α-helix. The structure of HLA-II is similar, but it consists of one α and one β chain, each containing two domains.

Before peptides can be loaded into the HLA-I binding site, intracellular proteins must first be degraded by the proteasome: a 2,000 kDa cylindrical catalytic complex. Human 26S proteasome contains two regulatory 19S caps, surrounding a 20S core particle of four stacked heptameric rings encasing a proteolytic core. Seven α subunits comprise the outer rings, whilst the inner rings are composed of β subunits including β1, β2 & β5 which each perform a specific proteolytic process within the catalytic cavity. In response to stimulation with IFNγ or TNFα, these three subunits are replaced with the β1i, β2i & β5i inducible subunits, respectively. This inducible form of the proteasome is referred to as the immunoproteasome, and it generates peptides with preferred C-termini for HLA-I binding through altering the cleavage pattern - reviewed by Basler *et al.* [306].

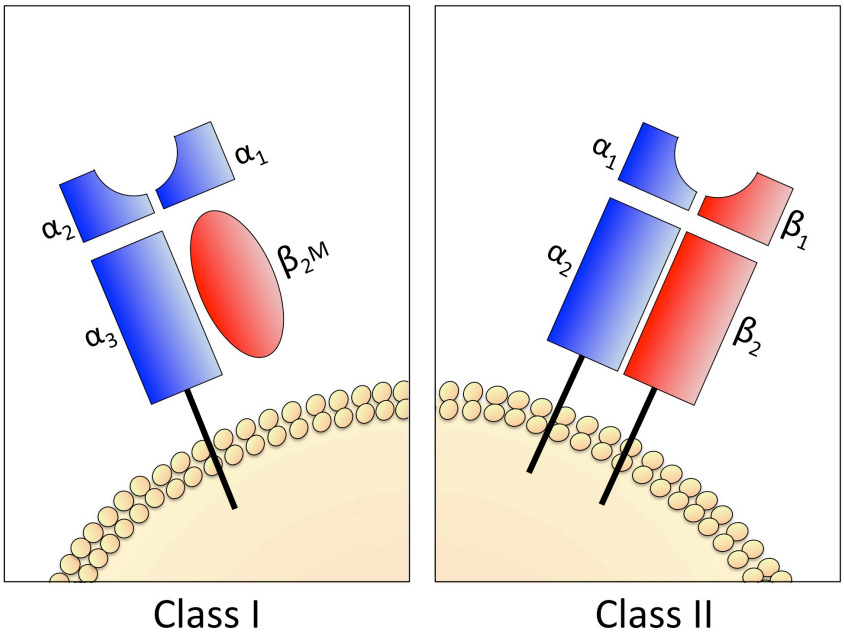


Figure 1.7: The structure of class I and class II HLA molecules HLA-I is made up of a highly polymorphic membrane-spanning α chain, consisting of three domains, as well as a non-covalently linked constant β-2-microglobulin (β2M) chain. HLA-II consists of membrane-spanning α and β chains, each containing two domains. Both classes bind peptides via the peptide-binding domain at the distal end of the molecule.

Proteins - which can be either viral or host-derived such as defective ribosomal products (DRiPs) - are first polyubiquitinated by a cascade of the ubiquitin-activating, conjugating, and ligating enzymes E1, E2 & E3 respectively. This permits recognition by the 19S regulatory elements, and the protein then (partially) unfolds before being translocated through the proteolytic core of the complex where it is cleaved. The generated peptides are released into the cytosol, and are then translocated into the ER by the transporter associated with antigen processing (TAP) [307] - a member of the ATP-binding cassette (ABC) family of transporters - which is localised in the ER and Golgi-apparatus, and consists of the homologous subunits TAP1 and TAP2.

Studies have shown that proteasome may generate optimally cleaved HLA-I epitopes *in vitro* - examples reviewed by [308] - but they more typically contain N- or C-terminal extensions. Post-proteasomal trimming of the epitope precursors is per-

formed primarily by aminopeptidases such as the endoplasmic reticulum aminopeptidases ERAP1 and ERAP2 [309]; but also by carboxypeptidases such as angiotensin-converting enzyme (ACE) [310].

Once the epitope has been correctly generated, the HLA-I:peptide complex is formed with the help of TAP and the chaperone proteins calnexin [311], tapasin [312] and ERp57 [313, 314]. The complex is trafficked to the cell surface via the secretory pathway, where it presents the peptide to $CD8^+$ lymphocytes.

HLA-II antigen processing is fundamentally different. HLA-II is expressed primarily on APCs and is used to present extracellular antigens to $CD4^+$ lymphocytes. HLA-II is constitutively synthesised in the ER, and is exported via the Golgi apparatus to major histocompatibility complex class-II compartments (MIICs) where it is functionally inhibited by the bound class II-associated invariant chain peptide (CLIP). Exogenous antigens that have been taken up and enzymatically processed through the cell's endosomal pathway displace CLIP and are loaded into the HLA-II complex, regulated by the histocompatibility proteins HLA-DO and HLA-DM. The loaded complex is transported to the cell surface for peptide presentation, reviewed by Kobayashi & van den Elsen [315]. Immature $CD4^+$ lymphocytes then recognise and bind the complex through their TCR and - following the interaction of further receptors such as CD28 and CD80/CD86 (signal 2) - become activated to form T helper (Th) cells which can themselves perform numerous antiviral activities such as activation of B cells and CTLs.

The genes that encode the HLA complex exist as a cluster on chromosome 6 (locus 6p21.3), and constitute one of the most highly polymorphic regions of the human genome. HLA type has been strongly linked with disease progression in the context

of HIV, with some types (such as HLA-B27 and HLA-B57) being associated with a good prognosis; whilst others (such as HLA-B35) are predictive of a much poorer outcome [316–318].

1.2.4 Cellular Adaptive Response

The cellular response plays an important role in both the control and pathogenesis of HIV infection. $CD8^+$ T lymphocytes respond to HIV via both cytotoxic and non-cytotoxic mechanisms. The role of the cellular response against HIV-1 is well established, and forms the basis of this overview. For a detailed review[1] of the cellular response in HIV-2 infection, please see my chapter on this topic published in *The Encyclopedia of AIDS* (Springer, 2015) [319].

1.2.4.1 Cytotoxic $CD8^+$ T Lymphocytic Response

The cytotoxic $CD8^+$ response is briefly summarised in figure 1.8. At the cell surface, the HLA-I:peptide complex is presented to CTLs, expressing on their surface unique TCRs and the CD8 co-receptor. If the TCR binds the HLA-I:peptide complex with sufficiently high affinity the immunological synapse is formed [320, 321], leading to polarised release of lytic granules. The lytic granule contents include perforin [322], serine esterase 'granzymes' [323], and fas ligand (FasL) [324]. The release of perforin creates transient pores in the membrane of the infected cell [325, 326], allowing access by the granzymes, which cleave cellular substrates [327]. The delivery of FasL to the infected cell also induces apoptosis through multimerisation of the fas receptor (FasR) [328], causing a conformational change in the intracellular domain [329], which recruits the fas-associated death domain (FADD) [330]. This leads to assembly of the

[1]Sections titled "$CD4^+$ virus-specific T-lymphocyte responses: HIV-1 vs HIV-2" and "T-lymphocyte responses in HIV-2: controllers vs progressors" were contributed by co-author G. Wong.

death-inducing signalling complex (DISC) [331], subsequently initiating a cascade of caspases and resulting in cell death [332].

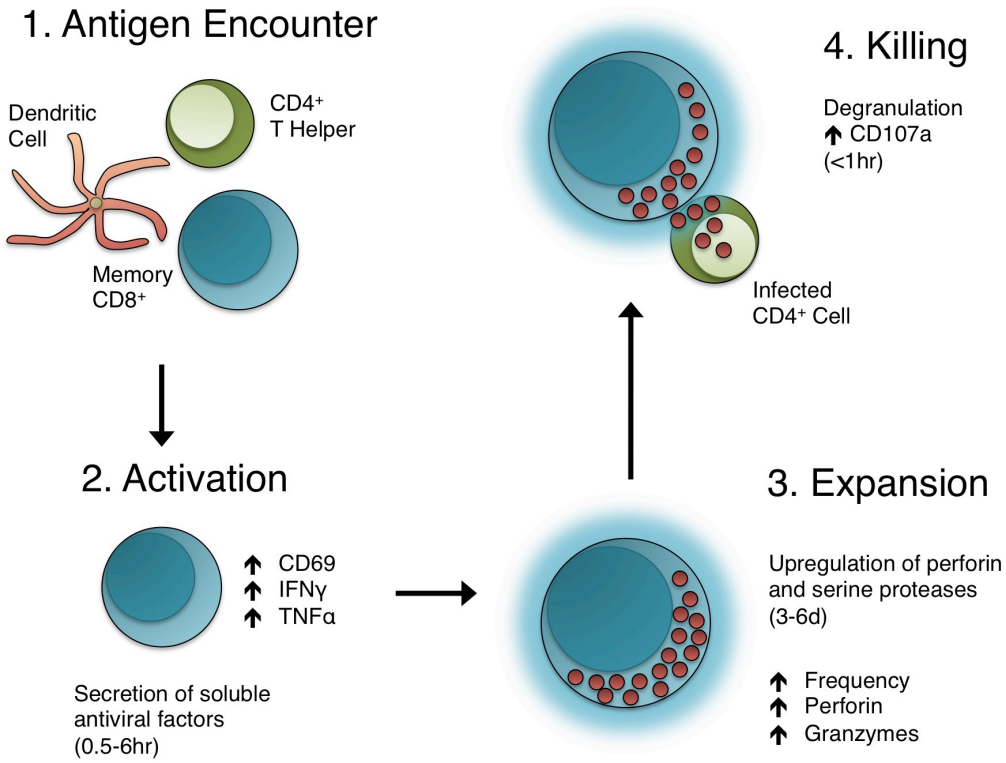


Figure 1.8: The cytotoxic cellular adaptive immune response against HIV infected cells. Figure re-drawn and adapted from Migueles & Connors [333].

Such antigen-specific CTL responses are heavily involved in the control of viraemia during acute infection [334], with rapid expansion of HIV-specific CTLs coinciding with a rapid decrease in circulating plasma viral load (section 1.1.5) [77]. Studies in the macaque model have also demonstrated that individuals present with greatly increased progression rates and an inability to control viraemia on depletion of the $CD8^+$ T lymphocyte population by targeted monoclonal antibodies [335, 336]. However, whilst the CTL response regains a degree of immunological control, it fails to eradicate the virus entirely [337].

The emergence of escape mutations is the leading mechanism by which HIV evades the cytotoxic T cellular response. Due to the enormous selection pressure exerted by these cells, and the error-prone nature of viral replication, escape mutations can be detected soon after infection [338] in and around epitopes associated with specific HLA types. These escape mutations may or may not revert on transmission to non-HLA-matched individuals dependent on their associated fitness costs [339, 340]. In addition to rapid mutation, HIV evades the cytotoxic response by mediating the downregulation of HLA-I from the surface of infected cells, via the accessory protein Nef (chapter 4). The HIV-1 unique accessory protein Vpu (section 1.1.8.9) can also mediate a reduction in surface HLA-I and resultant cytotoxicity by retaining nascent HLA-I molecules in the endoplasmic reticulum; but no functionally equivalent protein is found in HIV-2.

1.2.4.2 Non-Cytotoxic $CD8^+$ T Lymphocytic Response

While the protective effect of $CD8^+$ lymphocytes is largely attributed to their role in mediating cell killing, they are also able to control infection by non-cytotoxic means [341, 342]. The production of high levels of antiviral β-chemokines such as MIP-1α (CCL3), MIP-1β (CCL4), and RANTES (CCL5) by $CD8^+$ cells plays a role in the prevention and/or control of HIV-1 and SIV infections [343–345], by inhibiting binding of macrophage-tropic virus to the co-receptor CCR5 on entry [346, 347].

1.2.4.3 The Role of $CD4^+$ T Lymphocytes

Activated $CD4^+$ T helper (Th) lymphocytes are important in the orchestration of cellular and humoral adaptive immune responses. These cells are divisible into type 1 (Th1) or type 2 (Th2) helper cells on the basis of their cytokine secretion profile [348]. Th1 cells trigger cellular responses and phagocyte-dependent inflammation through the release of IFNγ, IL-2 and TNFβ; whilst Th2 cells trigger antibody responses

and eosinophil accumulation through the release of IL-4, IL-5, IL-6, IL-9, IL-10, and IL-13.

$CD4^+$ lymphocytes are themselves the targets of HIV infection, and this molecule is used by the virus to gain entry into the cell (section 1.1.6.1). Progressive loss of such $CD4^+$ cells is a key marker of disease progression and the rate of decline in untreated patients is indicative of prognosis [349]. Productively infected $CD4^+$ cells expressing viral epitopes on their surface are targeted for CTL-mediated killing, but more insidiously, non-permissive bystander cells in lymphoid tissue are also destroyed by pyroptosis [350]. This aggressive process promotes the release of pro-inflammatory cytokines which recruit further quiescent $CD4^+$ lymphocytes in a self-perpetuating cycle of cell death. The resulting chronic immune activation promotes an environment of viral replication and further depletion of T cells.

1.2.5 Humoral Adaptive Response

On recognition of antigen and additional activation (usually from mature $CD4^+$ Th2 cells), B cells proliferate and differentiate into plasma cells which then secrete large amounts of antigen-specific antibody. Antiviral antibodies have many functions, including activation of the complement cascade (section 1.2.2.6), activation of antibody-dependent cell-mediated cytotoxicity (ADCC) (section 1.2.2.4), and viral neutralisation. These functions are briefly summarised in figure 1.9.

The HIV-specific humoral response develops relatively early in infection: immune complexes can be seen approximately eight days after detectable plasma viraemia, with free antibodies against gp41 appearing after thirteen days, and against gp120 after twenty seven days [352]. However, these antibodies are unable to neutralise

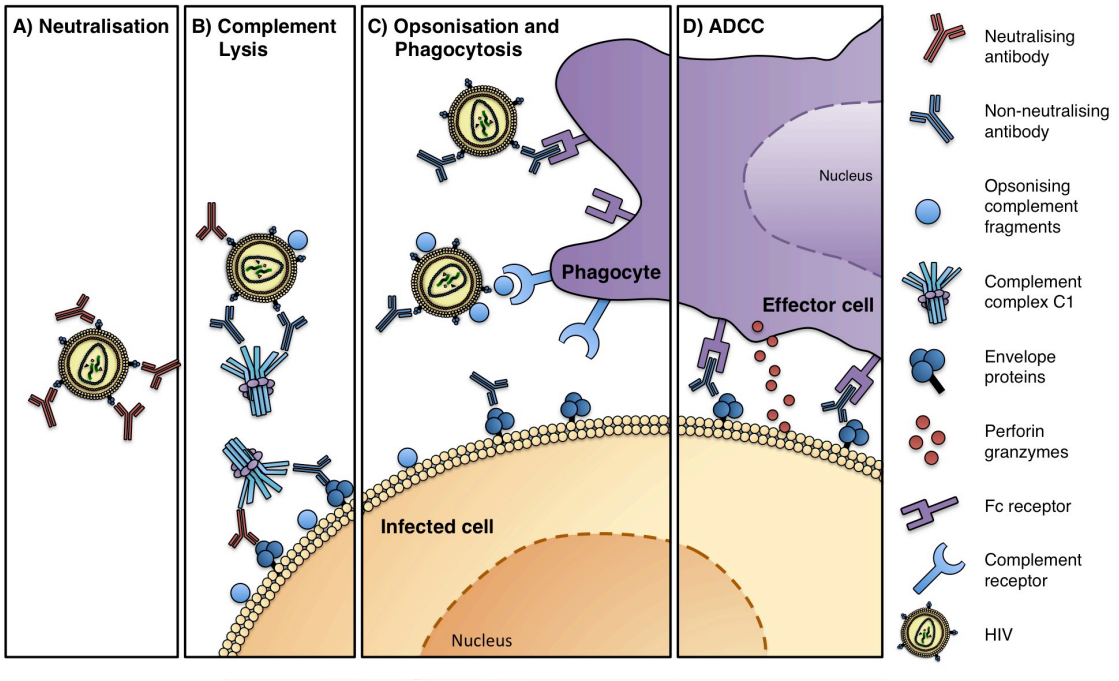


Figure 1.9: The humoral adaptive immune response, which functions via **A)** virus neutralisation, **B)** activation of the complement cascade, **C)** opsonisation and phagocytosis, and **D)** activation of ADCC. Figure re-drawn and adapted from Huber & Trkola [351].

live virus, and do not exert measurable selection pressure on viral epitopes [208]. Neutralising antibodies (NAbs) first emerge after several more months [353–355].

NAbs are antibodies that bind free virions as opposed to infected cells and can promote opsonisation and phagocytosis, as well as inhibit transmission and spread of virus by - for example - blocking key receptors. Many NAbs bind to regions of the envelope glycoprotein gp120. The contribution of neutralising antibodies to the control of HIV infection is subject to intense debate: titres of NAbs have not been found to correlate with disease progression, and in elite controllers they are typically low [356]. Furthermore, the breadth and potency of HIV-specific antibodies are generally poor, making them relatively strain-specific and ineffective against divergent virus.

Neutralising antibodies drive extensive viral escape in acute infection [127, 128]. Escape mutations are particularly prevalent in the surface-exposed variable regions of gp120: V1-V5, which shield the conserved protein core. However, whilst escape mutations appear rapidly and consistently, highly neutralisation-sensitive viruses have been detected *in vivo* in later infection [357]. The ability of these NAb-sensitive variants to emerge suggests that they have evolved within protective compartments or immunoprivileged sites such as the central nervous system (CNS), where the antibodies are unable to reach them. Evasion of the humoral response in this way presumably allows the virus to develop structural changes to vulnerable functional regions improving fitness.

In addition to rapid mutation and compartmentalisation, glycosylation has long been considered a mechanism of immune evasion. HIV-1 gp120 is extensively glycosylated, with approximately 50% of its molecular mass being contributed by N-glycans [358]. Glycosylation cloaks viral epitopes, making them inaccessible to antibodies [359]. The glycosylation pattern can also rapidly shift due to mutation of the underlying residues, creating a flexible and responsive glycans shield [128, 360]. However, whilst glycosylation shields viral epitopes, it is itself immunogenic owing to an excess of underprocessed oligomannose-type glycans which appear as "non-self motifs", reviewed by Doores [361]. Highly potent broadly neutralising antibodies (bNAbs) can be raised against such motifs, and as such, they are becoming targets for vaccine design.

In a subset of individuals, a class of bNAbs develop after an extended period of infection (usually several years) which are able to cross-neutralise heterologous virus, reviewed in depth by Kwong *et al.* [362]. It is believed that bNAbs may be more resistant to viral escape mutants due to their greater breadth of recognition. Furthermore, studies in animal models have shown that passive transfer of bNAbs provides

full protection against the establishment of novel infection, for example Hessell *et al.* [363]. As such, they have become the focus of intense scrutiny by those in the field of HIV vaccine development.

1.3 Summary and Study Objectives

The immune response against HIV is a highly complex, tightly regulated network of multiple cellular, humoral, and molecular effectors, as summarised in figure 1.10.

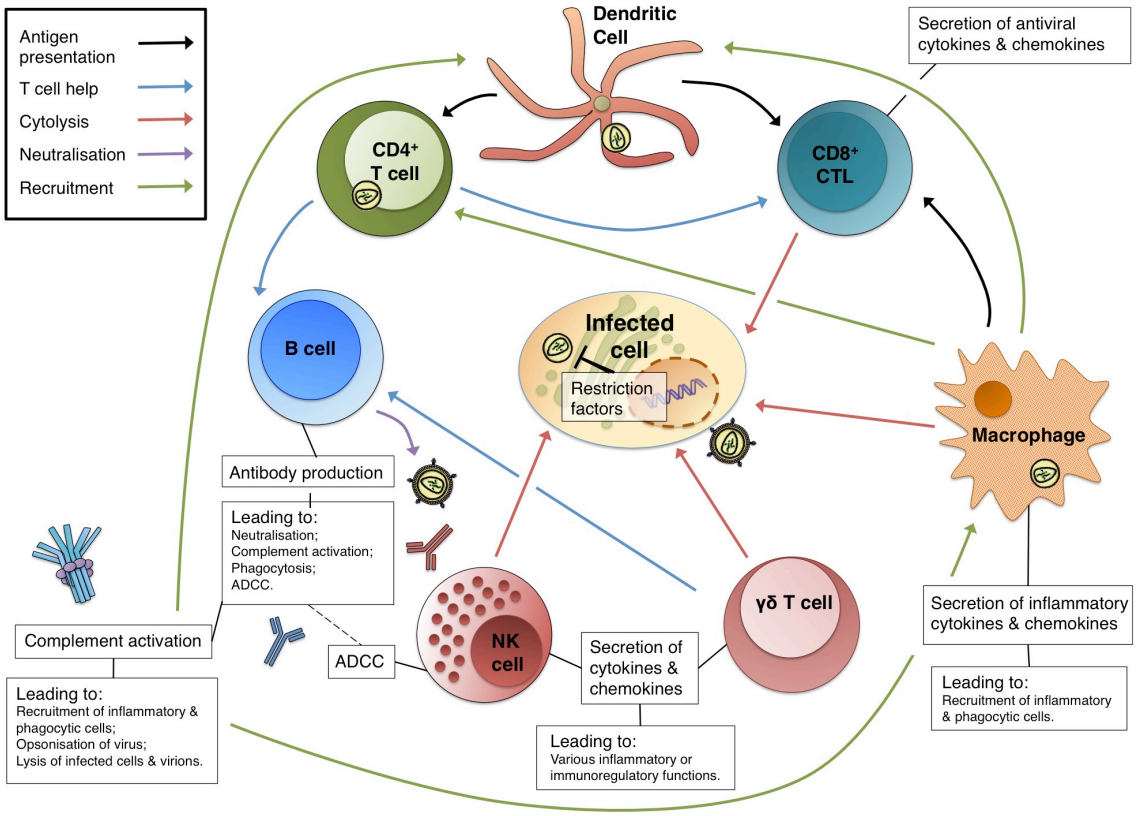


Figure 1.10: The Big Picture: key interactions in the complex network that constitutes the immune response against HIV. Arrows denote the nature and direction of interactions. Permissive cells are marked with a virus particle. Figure loosely based on Walker & Yu [364].

Yet despite its small genome, simple life cycle, and recent origin, HIV has developed a number of mechanisms that allow it to not only evade this response, but also launch an effective counteroffensive which leads to the eventual catastrophic failure of the immune system. A selection of key immune evasion mechanisms are summarised in table 1.4, adapted from Kirchhoff [366].

Table 1.4 Key immune evasion mechanisms used by HIV to counter specific host defences. Table adapted and updated from Kirchhoff [366].

Immune Response	Host Defence	Antiviral Effect	Evasion Mechanism	Protein or Property
Innate	NK cells	Lysis of infected cells	Selective downregulation of HLA-A & HLA-B but not HLA-C & HLA-E	Nef
	Complement	Recruitment of inflammatory and phagocytic cells; opsonisation; lysis of virions/infected cells	Recruitment of complement regulators and fH binding	gp120; gp41
Intrinsic	APOBEC3G	Hypermutation; interference with tRNA priming & elongation	Polyubiquitination and degradation	Vif
	TRIM5α	Premature uncoating	Variation in capsid	High variability
	Tetherin	Blocks virion release	Sequestration from the site of virion budding	Vpu; Nef; Env
	SAMHD1	Depletes dNTP pool	Polyubiquitination and degradation	Vpx
	SERINC3/5	Interference with infection of target cells	Prevention of incorporation into nascent virions	Nef
Adaptive	$CD8^+$ T cells	Lysis of infected cells; inhibitory cytokines	HLA-I downregulation; escape mutations; latent infection	Nef; high variability
	$CD4^+$ T cells	Helper function to promote antibody and CTL responses	Destruction by infection or bystander apoptosis; downregulation of CD4, CD3, CD28, and CXCR4	Nef; Vpu; viral cytopathicity
	B cells, antibodies	Neutralisation	Antigenic variation; glycosylation; shielding of functional epitopes; inhibition of IgG2; IgA switching	High variability; N-linked glycosylation sites; gp120 structure; Nef
	Antigen-presenting cells	Viral antigen presentation; helper T cell activation	Upregulation of Ii surface expression	Nef

To explore all of the ways in which this unique organism derails the host immune response would be a monumental task, and well beyond the scope of this thesis. Instead, this thesis will fill several gaps in the existing body of knowledge pertaining to three specific adaptive immune evasion mechanisms implemented during clinically latent HIV infection.

The first chapter reports on a longitudinal study into the intrapatient and interpatient mutation of the envelope protein gp120: a surface-expressed glycoprotein and major target of the antibody response. Whilst longitudinal sequence analysis of envelope mutation has been performed before, this study has utilised a narrow-source cohort of former plasma donors from China (SM cohort). The unique qualities of the SM cohort enable mutations to be considered in the absence of many factors which potentially confound similar studies. The specific aims of this project were therefore to characterise gp120 *env* sequences in the cohort, and to subsequently use these sequences to qualitatively and quantitatively investigate escape mutations in response to immunological selection pressure.

The following study sheds light on the role of primary Nef isolates in mediating the differential downregulation of HLA-A and HLA-B molecules. Nef-mediated HLA-I downregulation is an important mechanism of immune evasion in that it reduces the visibility of infected cells to cytotoxic lymphocytes, thereby rendering them more resistant to killing. In light of recent evidence that HLA-B may be more resistant to downregulation than HLA-A mediated by Nef isolated from laboratory-adapted strains of virus, the objective of this study was to determine whether primary patient isolates of both HIV-1 and HIV-2 Nef also mediate differential downregulation of HLA-I molecules *in vitro*.

To fully address these questions, a novel method was developed for the analysis of flow cytometric data, allowing HLA-I downregulation to be studied at higher resolution than previously possible. The primary purpose of developing an original analytical technique was to overcome many of the limitations associated with interpreting the results of bifurcate analyses. Presented is the reasoning behind the novel approach, as well as a step-by-step explanation of the analytical pipeline, and a demonstration by direct comparison of how this technique may outperform standard methods.

Following on from this study is an investigation into a further inconsistency between HIV-1 and HIV-2 Nef in the context of clinically latent infection. Whilst HIV-1 Nef is highly immunogenic and elicits robust T cell responses from very early in infection, HIV-2 Nef is often not recognised by CTLs and remains poorly antigenic throughout the course of disease. By performing a series of *in vitro* studies using putatively antigenic peptides, the (immuno)proteasomal processing of HIV-1 and HIV-2 Nef was characterised to determine whether this step in antigen preparation accounts in whole or in part for the observed differences in immunogenicity.

The final data chapter summarises the outcome of a successful collaborative attempt to resolve the crystal structure of HIV-2 Nef. Knowledge of the structure of HIV-1 Nef has provided essential insight into the role of this protein in infection and immune evasion, and the HIV-2 Nef structure presented herein may prove equally informative.

Whilst HIV immunology remains an active field of biomedical research, much is still to be learned about this interesting virus, and substantial gaps remain in the existing body of knowledge. By focusing on these three key areas, this thesis will further current understanding of how the virus escapes detection and killing, which will support informed attempts to disrupt these processes in future.

Chapter 2

Materials and Methods

2.1 Study Cohorts

2.1.1 HIV-1 (SM cohort)

The HIV-1 work described in this thesis was performed using samples collected from a community cohort of clinically latent former plasma donors from Shuang Miao (SM) village in Henan province, East Central China (figure 2.1). The most recent report - released by the National Health and Family Planning Commission of the People's Republic of China in May 2015 - estimated that 501,000 people were living with HIV in China by the end of 2014 [368]. The prevalence is high in Henan, where rates are amongst the highest in the country. Shuang Miao is a close-knit rural community in Henan province with a population of approximately 3,000. The villagers, mostly agricultural workers, belong to the Han Chinese clan, and share similar genetic backgrounds due to traditional marital practices that promote union within the community.

Between 1993-1995, over four hundred people were infected with a narrow-source of virus while participating in an illegal paid plasma donation scheme. Mobile units were set up in the village, where blood was repeatedly taken from patients, separated



Figure 2.1: Geographic location of Henan province, China. Figure adapted from "China EDCP Location Map" by Uwe Dedering, licensed under CC BY-SA 3.0.

to remove the plasma, and the cellular content transfused back to the donors. The resulting infections are believed to be due to contact with contaminated equipment and/or blood-products. 149 premature deaths consistent with AIDS-related diseases were reported in the village by 2004, leading to large-scale testing. 258 HIV positive individuals were identified through this screening, and subsequently gave informed consent to be recruited for further study in 2005 [369]. Further recruitment over subsequent years expanded the cohort to over 500 patients.

All patients within the cohort are believed to have been infected with a similar (purportedly single) strain of virus, through the same route, within a narrow time-frame. Study of the cohort therefore permits the unique opportunity to study the evolution of virus in response to the human immune system, with minimal confounding factors resulting from a broad spectrum of infecting viral strains and host ethnicity.

Another unique characteristic of the cohort is that although all patients remained antiretroviral naïve until at least 2004, many exhibited slow disease progression, and approximately 20% of the patients are long-term non-progressors: substantially higher than typically reported (section 1.1.5.5). This affords the opportunity to study

differences between - for example - progressors and LTNPs within a single cohort infected with essentially the same virus strain.

The anonymised samples are identified by a unique three-digit patient reference number prefixed with the letters SM. The reference numbers will also be used on occasion to refer to viral genes or proteins isolated from that patient.

Ethical approval for the use of the SM cohort samples was granted from the Oxford Tropical Research Ethics Committee (OxTREC).

2.1.2 HIV-2 (TD cohort)

The HIV-2 work described within this thesis was undertaken using samples collected from patients within a well-characterised community cohort from Caió in the Cacheu region of North-Western Guinea Bissau (figure 2.2). The Caió cohort was established in 1989 by the Medical Research Council (MRC) Gambia Unit, and the Bandim Health Project following observations from Dominique Ricard that a large proportion of HIV positive sex workers attending genitourinary medicine clinics in Ziguinchor (Southern Senegal) had originated from the village.

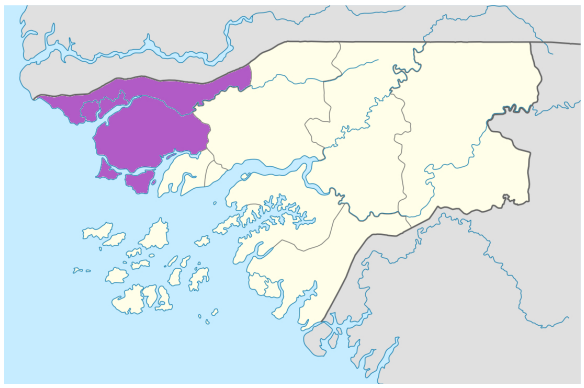


Figure 2.2: Geographic location of the Cacheu region of Guinea-Bissau, where Caió is situated. Figure adapted from "Guinea-Bissau Location Map" by NordNordWest, licensed under CC BY 3.0.

Caió is a rural community with a population of approximately ten thousand. Whilst Guinea-Bissau is an ethnically diverse country, inhabitants of Caió belong predominantly to the Manjako tribe. The Manjako are an indigenous people, with their own language, and an animistic belief system based around earth spirits. Inhabitants customarily marry within the community, and as a result the population is demographically and genetically highly homogeneous, and distinct from other ethnicities.

The village economy is built primarily on subsistence farming, with the major harvests being rice, cashew nuts, palm oil and palm wine owing to the location amongst rice fields, cashew forests, and palm groves [370]. Inhabitants also travel outside of the village to find work, with common destinations for men being Senegal, Portugal and France, whilst women often travel to Bissau - the capital city of Guinea-Bissau - or bordering countries for paid domestic work or to engage in commercial sex work [371, 372].

For more than twenty-five years, the village has been subject to numerous immunological and epidemiological studies of HIV-1, HIV-2 and HTLV-1, conducted from a small field station that used to serve as a local shop. Sero-surveys and demographic censuses were initiated between 1989 and 1991 to determine HIV prevalence and clinical impact within the village [373]. 8.3% of the population were found to be HIV-2 positive at this time, and infected individuals were then paired with uninfected controls based on demographic factors such as age, sex and location. The cohort has been studied since [374], and is so successful - at least in part - because the community setting allows patients to be followed-up without necessarily disclosing HIV status.

Further sero-surveys were performed in 1996-1998 & 2006-2007 [370, 371], and additional patients identified through these studies have since been recruited to the cohort.

Whilst HIV-1 prevalence has increased from 0.5% to 3.6% since the establishment of the cohort, HIV-2 prevalence within Caió has decreased markedly: most recently measured at 4.7% in 2007 [370]. Indeed, HIV-2 has been predicted to become extinct in Caió during the second half of the 21st century [375].

Access to medical care is provided free of charge to all cohort participants through the project's physician. Furthermore, ART was made available in Guinea-Bissau in 2007.

The subset of patient samples used in these studies were collected in 2010 by collaborating researcher Dr Thushan de Silva. The anonymised samples are identified by a unique three-digit patient reference number prefixed with the letters TD, accordingly. As with the SM cohort, the unique patient reference numbers will also be used on occasion to refer to viral genes or proteins isolated from that patient.

Ethical approval for the use of Caió cohort samples in this project was sought and obtained from the Gambian Government/MRC Joint Ethics Committee, from the Republic of Guinea-Bissau Ministry of Health, and from OxTREC.

2.2 Materials and Methods for Chapter 3

2.2.1 Selection of Patient Isolates

Twelve patients from the SM cohort were selected for this study. The demographic and clinical characteristics of these patients are listed in tables 2.1 & 2.2. These patients were selected as longitudinally collected samples were available, permitting the calculation of evolutionary rates and identification of sites under selection. The patients were also largely untreated at the time the study was initiated in 2013[1].

Table 2.1 Demographic characteristics of selected SM cohort patients

Patient	Date[a]	Sex	HLA Type					
			A.i	A.ii	B.i	B.ii	C.i	C.ii
SM007	1996	♀	A01	A02	B40	B57	C06	C08
SM021	1995	♀	A02	A29	B44	B38	C04	C07
SM039	1995	♀	A02	A11	B49	B39	C07	C07
SM176	1994/5	♀	A11	A32	B38	B51	C07	C14
SM209	1993/4	♂	A24	A31	B51	B54	C01	C14
SM335	1994/5	♂	A30	A31	B13	B40	C06	C03
SM358	1995	♀	A02	A11	B49	B50	C03	C06
SM446	2001	♂	A02	A26	B38	B51	C12	C14
SM505	1994/5	♂	A11	A02	B35	B52	C12	C12
SM514	1994/5	♀	A02	A31	B40	B40	C03	C15
SM536	1994	♀	A11	A30	B13	B15	C06	C07
SM538	1994	♂	A02	A24	B13	B40	C03	C03

[a] Date refers to the self-reported estimated date of infection.

[1]Owing to the extended duration of infection, antiretroviral treatments were being used by some patients at later sample collection dates as indicated in table 2.2. These patients were excluded from certain analyses where appropriate and specified.

Table 2.2 Clinical characteristics of selected SM cohort patients

	2010				2011				2012				2014			
	T	**VL**	**CD8 (%)**	**CD4 (%)**	**T**	**VL**	**CD8 (%)**	**CD4 (%)**	**T**	**VL**	**CD8 (%)**	**CD4 (%)**	**T**	**VL**	**CD8 (%)**	**CD4 (%)**
SM007	N	7388	1197 (76)	362 (23)	N	1445	917 (74)	300 (24)	N	2884	949 (60)	253 (16)	Y	8733	1033 (63)	253 (15)
SM021	N	35827	1181 (80)	266 (18)	-	-	-	-	N	14782	397 (39)	206 (20)	-	-	-	-
SM039	N	1612	1068 (64)	492 (30)	N	791	814 (65)	346 (28)	Y	1732	884 (49)	319 (18)	-	-	-	-
SM176	N	43569	887 (62)	429 (30)	N	435669	N/R	N/R	N	27348	166 (5)	95 (3)	Y	<50	632 (35)	495 (28)
SM209	N	44584	648 (63)	334 (33)	N	1195	750 (62)	354 (32)	N	4034	971 (44)	478 (22)	Y	11922	910 (55)	242 (14)
SM335	N	N/R	482 (59)	294 (36)	N	7751	478 (61)	257 (33)	Y	71263	723 (52)	228 (16)	Y	88620	491 (55)	59 (6)
SM358	N	<50	1224 (66)	469 (25)	N	1025	977 (68)	390 (27)	N	1716	887 (57)	387 (25)	-	-	-	-
SM446	N	N/R	638 (68)	255 (27)	N	7226	750 (62)	374 (31)	N	3217	394 (52)	187 (25)	-	-	-	-
SM505	N	8871	1958 (80)	478 (19)	N	175519	775 (69)	337 (30)	N	210735	1085 (64)	402 (24)	Y	180	1054 (54)	593 (30)
SM514	N	30403	3070 (87)	389 (11)	N	N/R	609 (33)	1233 (66)	N	5166	2545 (81)	216 (7)	-	-	-	-
SM536	N	4812	1234 (67)	567 (31)	N	23956	1239 (74)	397 (24)	N	12637	1598 (60)	351 (13)	-	-	-	-
SM538	N	409157	807 (87)	92 (10)	N	<50	524 (84)	73 (12)	Y	<50	290 (48)	52 (9)	-	-	-	-

VL: viral load; T: antiretroviral treatment (Y/N: yes/no); N/R: not recorded on collection; dash (-): no sample as patient unavailable for bleeding, or deceased.

2.2.2 Extraction of Viral RNA

RNA extraction was performed from 200 µl of frozen patient plasma, using a QIAamp® UltraSens® Virus Kit (QIAGEN®), following the manufacturer's protocol. The extracted RNA was eluted into a final volume of 20 µl of nuclease-free water, and reverse-transcribed using the SuperScript® III reverse transcriptase system (Invitrogen™). Remaining RNA was immediately stored at -80℃.

2.2.3 Amplification of Gp120 *Env* Fragment

Nested primer pairs were designed to amplify an approximately 900 bp fragment of gp120 *env* spanning C2-V5, based on the results of epitope mapping demonstrating a clear clustering of epitopes within the V3 variable loop (section 3.2.1). 512 subtype B' *env* sequences from similar HIV$^+$ Chinese former plasma donors (FPDs) were downloaded from the LANL HIV Sequence Database [376] as a reference dataset and aligned in Geneious® v9.0.5 [377], implementing the multiple sequence comparison by log-expectation (MUSCLE) algorithm [378]. Relatively conserved stretches within the alignment were identified as appropriate primer binding sites, and the optimal sequence was then determined using the most frequent base in each position. Accession numbers for sequences used are listed in appendix table A.8.

The selected fragment was amplified from cDNA by nested touchdown polymerase chain reaction (PCR) in a Hain Lifescience GTQ-cycler 96 thermal cycler. The Advantage HF®2 PCR Kit (Clontech TaKaRa) was used, broadly following the manufacturer's protocol. The details of each PCR reagent and the cycling parameters used are listed in table 2.3. Primers used in the first round of PCR were FPD_Env_OF and FPD_Env_OR, and for the second round were FPD_Env_IF and FPD_Env_IR (sequences listed in appendix table A.1).

Table 2.3 PCR reagents and cycling parameters used to amplify the *env* gp120 fragment.

(a) PCR reagents

Reagent	**Quantity**
10X Advantage HF®2 PCR Buffer	5 µl
10 mmol dNTP mix	1 µl
Sense primer (10 µmol)	4 µl
Anti-sense primer (10 µmol)	4 µl
50X Advantage-HF®2 Polymerase Mix	1 µl
Template (cDNA[1]; first round product[2])	10-200 ng[1]; 10 µl[2]
Nuclease-free water	to 50 µl total volume

(b) Cycling parameters

Step	**Temperature (°C)**	**Time (S)**
Denaturation	95	60
40 cycles of:		
Denaturation	95	30
Annealing	56-52	30
Extension	68	90[1]; 60[2]
Final Extension	68	90[1]; 60[2]

[1] First round of nested PCR
[2] Second round of nested PCR

The PCR products were run on a 1% agarose tris-acetate-EDTA (TAE) gel, stained with 0.5 µg ml^{-1} of ethidium bromide, and visualised using ultraviolet transillumination. Products of the expected size were excised and purified using the MinElute® Gel Extraction Kit (QIAGEN®), eluting into a final volume of 12 µl of nuclease-free water.

2.2.4 Cloning of PCR Products for Sequencing

Purified PCR products were ligated into a pCR™4-TOPO® sequencing vector using the TOPO® TA Cloning® Kit for Sequencing (Invitrogen™). Briefly, 4 µl of each PCR product was added to 1 µl of pCR™4-TOPO® vector and 1 µl of salt solution. The mixture was incubated at room temperature for 20 minutes.

Chemically competent One Shot® MAX Efficiency® DH5α™ *E. coli* (Invitrogen™) were then transformed with the ligated plasmids. Briefly, 5 µl of the ligation mixture was added to 200 µl of cells, mixed gently, and incubated on ice for 30 minutes. The cells were then heat-shocked at 42℃ for 30 seconds, and returned to ice for 2 minutes. 250 µl of super optimal broth with catabolite repression (SOC) media was added to each vial, prior to incubation at 37℃ for one hour in a horizontally shaking incubator programmed to 225 RPM. Finally, 75 µl of each transformation product was spread onto selective warmed LB-agar plates containing 50 µg mL^{-1} of kanamycin, and incubated at 37℃ overnight.

2.2.5 Colony PCR

Colony PCR was performed to amplify the cloned PCR products from a selection of the colonies generated for sequencing. The number of colonies selected per sample was determined as a compromise between power to resolve major variants (defined herein as contributing greater than 15% of the quasispecies pool), and the financial constraints of the project. It was agreed that sufficient colonies must be selected to resolve all major variants to within a confidence interval of 95%. With this in mind, a sample size of 18 colonies per patient per time point was considered optimal. The power calculation used is shown in equation 2.1.

$$n = \frac{\ln(P)}{\ln(1 - X)} \tag{2.1}$$

Where n is the number of colonies to be selected, X is prevalence of variant x in the quasispecies pool, and P is probability of missing variant x.

In anticipation of occasional empty vectors or off-target amplification, 24 colonies were selected for colony PCR in the first instance. Individual colonies were selected using a clean pipette tip, and transferred to PCR reaction mixtures as described in table 2.4. The reactions were performed in 96-well plates in a Hain Lifescience GTQ-cycler 96 thermal cycler.

Table 2.4 Colony PCR reagents and cycling parameters used to amplify the gp120 *env* fragment from individual *E. coli* colonies.

(a) PCR reagents

Reagent	**Quantity**
10X Advantage HF®2 PCR Buffer	2 µl
10 mmol dNTP mix	0.4 µl
Sense primer (M13F; 10 µmol)	0.4 µl
Anti-sense primer (M13R; 10 µmol)	0.4 µl
50X Advantage-HF®2 Polymerase Mix	0.4 µl
Nuclease-free water	to 20 µl total volume

(b) Cycling parameters

Step	**Temperature (°C)**	**Time (S)**
Denaturation	95	60
30 cycles of:		
Denaturation	95	30
Annealing	55	30
Extension	72	30
Final Extension	72	60

PCR products were run on large-scale 1% agarose TAE gels, stained with 0.5 µg ml^{-1} of ethidium bromide, and visualised using ultraviolet transillumination.

2.2.6 Sequencing of Colony PCR Products

Where the PCR produced clean, single bands, the products were purified for sequencing using an in-house modified ExoSAP protocol. Exonuclease I (ExoI) removes residual single-stranded DNA including primers, and shrimp alkaline phosphatase (SAP) removes remaining dNTPs. Briefly, 0.025 µl ExoI and 0.25 µl SAP were incubated with 9.725 µl PCR product and 22 µl of RNAse-free water at 37℃ for 30 minutes, before being heat-inactivated at 85℃ for 15 minutes. Where the PCR generated additional off-target products which were unable to be adequately reduced with further optimisation of the reaction parameters, the QIAquick® Gel Extraction Kit (QIAGEN®) was used instead to extract and purify the band of interest.

The purified products were diluted to 1 ng μl^{-1} per 100 bp in nuclease-free water and sent for Sanger sequencing by Source BioScience plc with the primer pair used for amplification (M13F & M13R; appendix table A.1).

2.2.7 Sequence Analysis and Statistics

Reads generated were quality-checked by examination of the electropherograms for resolution, noise, and miscalled bases. High-quality sequences were then imported into Geneious® (v9.0.5) [377], where the forward and reverse reads were compiled by pairwise alignment. Disagreements between contiguous sequences were resolved by selection of the residue at that position in the highest-quality electropherogram.

Multiple sequence alignments were then compiled within Geneious® implementing the MUSCLE algorithm [378]. These were manually edited and refined in MEGA

v6.06 [380] to preserve codon integrity, and promote construction of an evolutionarily sound alignment. Sequences were computationally translated, and those containing premature stop codons within their open reading frame were excluded as "evolutionary dead ends". Further patient-specific or consensus alignments were generated from the universal alignment, as described in the text.

For phylogenetic studies, an appropriate nucleotide substitution model was selected for each particular dataset using jModeltest v2.0 [381, 382]. Log likelihood was calculated for four major substitution models: JC69 [383], F81 [384], HKY85 [385], and generalised time-reversible (GTR) [386]. These values were then directly compared using the likelihood ratio test (LRT) to determine the most appropriate model for that study. The selected model for each analysis is stated in the text, along with the program and parameters used to generate the associated phylogeny.

Where Bayesian Markov chain Monte Carlo (MCMC) phylogenetic inference was implemented, the graphical user interface (GUI) Bayesian evolutionary analysis utility (BEAUti) v1.8.2 was used to assign parameters to analyses run through the Bayesian evolutionary analysis sampling trees (BEAST) v1.8.2 program [387]. Unless otherwise specified, the Markov chain was run for 100,000,000 iterations to allow adequate mixing, with posterior samples extracted every 10,000 generations. Outputs were then viewed and analysed using Tracer v1.6 [388] and convergence of posterior probabilities was assumed once the effective sample size (ESS) of each parameter exceeded 200. A burn-in of 10% of the posterior distribution was applied. Further input parameters are listed alongside each analysis. 95% highest posterior density (HPD) is stated where relevant.

A number of both open-source and licensed programs were used for sequence analysis. Table 2.5 contains a list of programs used, including a brief description and relevant referencing. Programs which have also been used in additional chapters are indicated for clarity.

Custom analytical scripts developed for particular analyses were written in R (v3.3.1) implemented through the RStudio (v.0.98.1103) integrated development environment (IDE).

Table 2.5 List of programs used for the sequence analysis described within this thesis.

Program	Reference	Version	Description	Chapter
BEAST	Drummond *et al.* [387]	1.8.2	The Bayesian Evolutionary Analysis Sampling Trees (BEAST) package is a program used for the analysis of sequences using Bayesian Markov chain Monte Carlo (MCMC) methods. It is used for inferring time-measured rooted phylogenies with strict or relaxed molecular clock models, as well as testing of evolutionary hypotheses without the restraints of conditioning on a single tree topology. BEAST averages over tree space using MCMC to weigh each tree proportionally to its posterior probability. Multiple different trait associations can be analysed within the BEAST framework simultaneously.	3
BEAUti	Drummond *et al.* [387]	1.8.2	Bayesian evolutionary analysis utility (BEAUti) is a graphical user interface used for assigning the conditions and generating the necessary files to run an analysis in BEAST.	3
FigTree	Rambaut [389]	1.4.0	FigTree is a program used for viewing and formatting publication-quality phylogenetic trees.	3,4
FinchTV	Geospiza Inc. [390]	1.5.0	FinchTV is a program for visualising and editing sequencing electropherograms. This program was used extensively throughout chapters 3, 4, and 5 for quality-checking sequencing reads, and resolving base ambiguities.	3,4,5

Continued on next page

Table 2.5 – *Continued from previous page*

Program	Reference	Version	Description	Chapter
Geneious®	Kearse *et al.* [377]	9.0.5	Geneious® is an integrated bioinformatics software platform used for performing sequence alignments, assembly, mapping, annotation and analysis. It includes built-in packages for the construction of pylogenetic trees as well as the design of primers and cloning experiments, amongst others. Geneious® was the main platform used for undertaking the sequence analysis work described in this thesis.	3-7
jModeltest	Darriba *et al.* [381], Guindon & Gascuel [382]	2.0	jModeltest is a program used for selecting the appropriate nucleotide substitution model for a given set of data to infer the most likely phylogeny. It utilises five different model-selection frameworks to identify the model of best-fit.	3
LogCombiner	Drummond *et al.* [387]	1.8.2	LogCombiner is a tool for combining MCMC log and tree files from independent runs of BEAST.	3
MEGA	Tamura *et al.* [380]	5.2.2	Molecular evolutionary genetics analysis (MEGA) is an integrated software platform for sequence alignment, phylogenetic tree inference, estimation of divergence, database mining, estimation of molecular evolutionary rates, and testing of evolutionary hypotheses. MEGA was used in this thesis for the manual editing of aligned sequences as well as data visualisation.	3

Continued on next page

Table 2.5 – *Continued from previous page*

Program	Reference	Version	Description	Chapter
PyMOL	Schrödinger, LLC [391]	1.7.4	PyMOL is a graphical user interface for the visualisation and analysis of three-dimensional molecular structures including proteins.	3,4,7
SWISS-MODEL	Arnold *et al.* [392], Kiefer *et al.* [393], Biasini *et al.* [394], Guex *et al.* [395]		SWISS-MODEL is a web-server used for building three-dimensional homology models from sequence data, using experimental protein structures as templates.	3,4
Tracer	Rambaut *et al.* [388]	1.6	Tracer is a graphical user interface for visualisation, analysis, and diagnostics of the outputs generated by MCMC programs such as BEAST. Demographic history reconstruction including SkyGrid may be implemented through Tracer.	3
WebPSSM	Jensen *et al.* [396], Brumme *et al.* [397], Jensen *et al.* [398]		WebPSSM is a web tool used for the prediction of HIV-1 co-receptor usage, based on the primary sequence of the V3 hypervariable loop.	3

2.3 Materials and Methods for Chapters 4 & 5

2.3.1 Generation of Consensus HIV-1 and HIV-2 *Nef* Genes

Unambiguous HIV-1 and HIV-2 consensus *nef* sequences were generated from data available in the LANL HIV Sequence Database [376] as well as sequences generated by a collaborating group (Professor Judy Breuer, personal communication). 3,086 HIV-1 subtype B and 356 HIV-2 *nef* sequences were separately aligned in Geneious® v9.0.5 [377], implementing the MUSCLE algorithm [378] with a limit of one sequence per patient applied. The consensus nucleotide sequences for HIV-1 and HIV-2 *nef* are listed in appendix table A.5. These were synthesised by Biomatik on request.

2.3.2 Selection of Patient Isolates

A stratified sample of ten HIV-1 and ten HIV-2 mono-infected patients were selected from the SM and TD cohorts, respectively. These patients were stratified on population-level HLA-I frequencies, and they also exhibited a broad range of clinical outcomes (table 2.6). Untreated patients were used where possible but individuals receiving antiretroviral treatment were excluded from certain analyses where specified.

2.3.3 Amplification of Full-Length *Nef* Gene

Nested primer pairs were designed to individually amplify full length HIV-1 and HIV-2 *nef*. Primers used in the first round of PCR were HIV1_Nef_OF & HIV1_Nef_OR for HIV-1, and HIV2_Nef_OF & HIV2_Nef_OR for HIV-2. For the second round, they were HIV1_Nef_IF & HIV1_Nef_IR and HIV2_Nef_IF & HIV2_Nef_IR, respectively. The sequences for these primers are listed in appendix table A.1.

Table 2.6 Clinical characteristics of selected SM cohort patients for the *nef* study

(a) HIV-1 Cohort (SM)

Patient	Date[1]	HLA Type						T[2]	VL[3]	CD4[4]
		A.i	A.ii	B.i	B.ii	C.i	C.ii			
SM046	1994	A02	A30	B27	B15	C03	C15	N	46543	179
SM335	1994/5	A30	A31	B13	B40	C06	C03	N	131292	294
SM339	1990	A02	A24	B40	B40	C07	C15	N	<50	581
SM434	1995	A11	A30	B13	B13	C03	C06	N	<50	1029
SM526	1993/5	A31	A68	B27	B40	C02	C03	N	N/R[5]	N/R
SM528	1993/5	A11	A24	B15	B37	C04	C06	Y	<50	351
SM547	1995	A02	A30	B13	B40	C06	C08	Y	853	N/R
SM550	1995	A02	A11	B13	B46	C01	C06	N	<50	810
SM551	1995	A02	A11	B13	B13	C03	C06	N	12996	486
SM569	1994	A02	A30	B13	B13	C06	C06	N	74	943

(b) HIV-2 Cohort (TD)

Patient	Date[6]	HLA Type						T[2]	VL[3]	CD4[4]
		A.i	A.ii	B.i	B.ii	C.i	C.ii			
TD001	2007	A23	A30	B07	B58	C03	C17	N	<50	629
TD014	2007	A23	A74	B49	B53	C04	C06	N	61952	178
TD027	1997	A23	A30	B35	B49	C04	C07	N	<50	435
TD037	2007	A26	A34	B08	B35	C03	C04	N	708	293
TD041	1997	A26	A33	B08	B15	C03	C03	N	20956	213
TD045	1997	A23	A74	B49	B53	C04	C07	N	<50	596
TD052	1989	A26	A30	B08	B50	C03	C06	N	<50	321
TD053	2007	A02	A33	-	-	C14	C14	N	528354	575
TD055	1989	A26	A33	B08	B82	C03	C03	N	484	407
TD062	1989	A26	A33	B08	B13	C03	C08	N	139519	497

[1] Date refers to the estimated date of infection.
[2] T refers to whether or not the patient was receiving antiretroviral treatment when sampled.
[3] VL refers to viral load in copies ml^{-1} in the year of sampling (2010).
[4] CD4 refers to $CD4^+$ T cell count in cells μl^{-1} in the year of sampling (2010).
[5] N/R refers to data not recorded when sampled.
[6] Date refers to the date of diagnosis.

The genes were amplified from proviral DNA by nested touchdown PCR in a Hain Lifescience GTQ-cycler 96 thermal cycler. The Advantage HF®2 PCR Kit (Clontech TaKaRa) was used, broadly following the manufacturer's protocol. The details of each PCR reagent and the cycling parameters used are listed in table 2.7.

Table 2.7 PCR reagents and cycling parameters used to amplify full-length HIV-1 and HIV-2 *nef* genes

(a) PCR reagents

Reagent	**Quantity**
10X Advantage HF®2 PCR Buffer	2 µl
10 mmol dNTP mix	0.4 µl
Sense primer (10 µmol)	0.4 µl
Anti-sense primer (10 µmol)	0.4 µl
50X Advantage-HF®2 Polymerase Mix	0.4 µl
Template (gDNA[1]; first round product[2])	10-200 ng[1]; 5 µl[2]
Nuclease-free water	to 20 µl total volume

(b) Cycling parameters

Step	**Temperature (°C)**	**Time (S)**
Denaturation	95	60
40 cycles of:		
Denaturation	95	30
Annealing	55-52	30
Extension	68	90
Final Extension	68	90

[1] First round of nested PCR
[2] Second round of nested PCR

The PCR products were run on a 1% agarose TAE gel, stained with 0.5 µg ml^{-1} of ethidium bromide, and visualised using ultraviolet transillumination. Products of

the expected size were excised and purified using the QIAquick® Gel Extraction Kit (QIAGEN®), eluting into a final volume of 30 µl of nuclease-free water.

Samples of the purified bulk products were sent for Sanger sequencing by Source BioScience plc with the inner primer pair used for amplification. Remaining products were cloned into a pCR™4-TOPO® sequencing vector using the TOPO® TA Cloning® Kit for Sequencing (Invitrogen™). Chemically competent One Shot® TOP10 *E. coli* (Invitrogen™) were then transformed with the prepared plasmids (section 2.2.4). Ten colonies were selected for PCR (section 2.2.5), and the products were again sent for sequencing (section 2.2.6) to provide a crude estimate of quasispecies diversity.

2.3.4 Cloning of Full-Length *Nef* Gene

The clonal sequence most closely matching - and in most cases identical to - the bulk sequence from each patient was selected for functional studies. This was presumed the likely dominant variant in the quasispecies pool. In addition to the primary isolates, laboratory-adapted SF2 *nef* (Accession number: K02007), a non-functional SF2 variant with a G2A substitution, and consensus HIV-1 and HIV-2 *nef* (section 2.3.1) were also cloned.

To allow ligation into a mammalian expression vector, primers were designed manually to adhere restriction sites to the 5' and 3' ends of each selected *nef* sequence using PCR. Unless specified, a BamHI restriction site (`GGATCC`) was added to the 5' end of *nef*, and EcoRI (`GAATTC`) to the 3' end. In addition to the restriction sites, a Kozak sequence (`ACCACC`) was added to the sense primer to encourage strong expression, and supplementary bases were added to the antisense primer to maintain frame continuity where appropriate. Six further bases were added to the overhanging ends of each primer to allow attachment of the restriction endonucleases. Where internal

ribosome entry site (IRES) containing plasmids were used, a stop codon (`TGA`) was incorporated into the antisense primer. The sequences of a representative primer pair (HIV1_Nef_EcoRI_F_g1 & HIV1_Nef_EcoRI_R_g1) are listed in appendix table A.1.

Reactions took place in a Lifescience GTQ-cycler 96 thermal cycler, using the Advantage HF®2 PCR kit (Clontech TaKaRa). The details of each PCR reagent and the cycling parameters used are listed in table 2.8. Following completion of the PCR, products were purified with a QIAquick® PCR Purification Kit (QIAGEN®) following the manufacturer's protocol.

Table 2.8 PCR reagents and cycling parameters used to adhere restriction sites to full-length HIV-1 and HIV-2 *nef* genes

(a) PCR reagents

Reagent	**Quantity**
10X Advantage HF®2 PCR Buffer	2.5 µl
10 mmol dNTP mix	0.5 µl
Sense primer (10 µmol)	2 µl
Anti-sense primer (10 µmol)	2 µl
50X Advantage-HF®2 Polymerase Mix	0.5 µl
Template (*nef* gene of interest)	10 ng
Nuclease-free water	to 25 µl total volume

(b) Cycling parameters

Step	**Temperature (°C)**	**Time (S)**
Denaturation	95	60
25 cycles of:		
Denaturation	95	30
Annealing	55	30
Extension	68	60
Final Extension	68	60

500 ng of the prepared *nef* inserts and 500 ng of the expression vector were sequentially digested with 10 U of the appropriate restriction enzymes at 37℃ for one hour. The vectors used were pAcGFP1-N1 and pIRES2-ZsGreen1 (both Clontech TaKaRa), and the custom pAcGFP1-N1-flexi vector, as specified in text. The restriction enzymes used were EcoRI & BamHI (both New England BioLabs® Inc.), unless stated otherwise.

Digestion products were purified by gel extraction (described previously) and ligated using T4 DNA ligase (Promega) with a molar ratio of 1:1 vector to insert. The reaction was performed at room temperature for two hours, using 10U of ligase in a 20 µl total reaction volume.

Chemically competent One Shot® MAX Efficiency® DH5α^{TM} *E. coli* (Invitrogen^{TM}) were transformed with the prepared plasmids as described in section 2.2.4, and five colonies per plate were tested for the presence of the correct insert by colony PCR (section 2.2.5) using the pAcGFP1N1 or pIRES2ZsG1 primer pairs in appendix table A.1. Successful colonies were expanded overnight at 37℃ with 225 RPM horizontal rotation in 5 ml of LB broth supplemented with 50 µg mL^{-1} of kanamycin. Plasmid DNA was purified from the cultures using the QIAprep® Spin Miniprep Kit (QIAGEN®), following the manufacturer's protocol and eluting in 30 µl of nuclease-free water. Sequence integrity was confirmed by Sanger sequencing (Source BioScience plc). Cultures containing each cloned plasmid were cryopreserved in 25% glycerol solution at -80℃.

2.3.5 Large-Scale Purification of Plasmid DNA

One litre of bacterial culture was expanded in terrific broth (TB) supplemented with 50 µg mL^{-1} of kanamycin, from individual colonies containing each cloned plasmid. The plasmids were purified on a large scale using a caesium chloride gradient.

Briefly, cells were harvested by centrifugation at 2,000 x g in a Beckman Avanti® J-25 centrifuge (Beckman Coulter Inc.) at 4℃. The pellet was re-suspended in 50 ml of buffer P1[2]. 50 ml of lysis buffer P2 was then added and incubated at room temperature for five minutes followed by neutralisation with 50 ml of buffer P3. The precipitate was removed by further centrifugation, and the supernatant was filtered. Nucleic acid was precipitated for one hour at room temperature by the addition of 150 ml of isopropanol. The DNA was pelleted by centrifugation and air-dried before being re-suspended with tris-EDTA (TE) to a final mass of four grams. Four grams of caesium chloride and 375 µl of 10 mg mL^{-1} ethidium bromide were added, and mixed by vortexing. Flocculant was removed by centrifugation at 10,000 x g in a Heraeus™ Biofuge for 20 minutes and the clarified sample was then transferred to a 4.7 ml OptiSeal™ tube (Beckman Coulter Inc.). The sample was banded by ultracentrifugation at 400,000 x g in a Beckman TLA 100.4 rotor (Beckman Coulter Inc.) for 24 hours at 24℃.

The plasmid band was aspirated and the ethidium bromide extracted with water-saturated butanol. DNA was precipitated in two volumes of 100% ethanol, and washed twice with 70% isopropanol. Finally, the clean pellet was re-dissolved in 500 µl of TE buffer.

[2]Buffer composition is listed in table 2.10

2.3.6 Cell Culture and Maintenance

A series of seventeen transgenic lymphoblastic cell lines were generated and described previously by Dr Ushani Rajapaksa [399]. Each Jurkat (JJK subline[3]) derivative line stably expresses a chimeric HLA-I molecule on its surface composed of the extracellular domain of HLA-A2 and a variable cytoplasmic tail representative of a range of different HLA-I molecules. The panel of cell lines spans all known diversity of the HLA-A, -B, and -C cytoplasmic tail, and HLA types which share each cytoplasmic tail are listed in appendix table A.3. These lines - in addition to conventional JJK cells - were individually cultured in R10 media, consisting of Roswell Park Memorial Institute medium (RPMI) 1640 (Sigma) with 10% foetal calf serum (FCS), 2 mM of L-glutamine, 50 U ml^{-1} of penicillin and 50 mg ml^{-1} of streptomycin. Cell lines were maintained at approximately $0.5x10^6$ cells per millilitre, grown at 37℃ with 5% CO_2, and were split every four days or when the media turned yellow.

For long-term storage, certified mycoplasma-free cells were prepared for cryopreservation by fully replacing the media 24 hours before freezing. Cells were harvested by spinning at 300 x g in a Heraeus™ Biofuge before being re-suspended in foetal calf serum (FCS) containing 10% dimethyl sulphoxide (DMSO) at a final concentration of 3 million cells per millilitre. 1.5 ml aliquots were transferred to labelled cryovials, and placed in a Mr. Frosty™ freezing container (Thermo Scientific™) containing isopropyl alcohol according to the manufacturer's recommendations. The unit was transferred to a -80℃ freezer overnight. The cryovials were then removed and stored in the vapour phase of a liquid nitrogen storage unit.

[3]Generously gifted by Professor Dan Littman, Skirball Institute, New York, NY, USA

When required, cryovials containing the frozen cells were placed in a 37℃ water bath and thawed rapidly by gentle swirling. When a thin sliver of frozen cells remained, the cells were transferred drop-wise to an excess of pre-warmed growth media to wash away the cryopreservatives. The cells were pelleted at 300 x g, and re-suspended in R10 media for expansion at 37℃ with 5% CO_2.

2.3.7 Transfection by Electroporation

The chimeric cell lines were transiently transfected with the purified expression vectors (section 2.3.5) using electroporation. Twenty four hours prior to transfection, cells were washed with RPMI-1640 and re-suspended in fresh R10 media. Cell density was calculated using a haemocytometer, and aliquots of $3x10^6$ cells were pelleted per transfection. The pellets were washed twice with RPMI-1640, and all residual media was removed. The pellets were then re-suspended in 200 µl of RPMI-1640. 15 µg of plasmid was added and gently mixed. The samples were transferred to each 0.2 cm well of a 96-well electroporation plate, and placed in the plate holder of a GenePulser MXcell^{TM} Electroporation System (Bio-Rad Laboratories). The cells were shocked at 250 V, 1050 µF, 50 Ω. Immediately after, the cells were gently transferred to 24-well plates containing 1 ml pre-warmed R10, and incubated at 37℃ with 5% CO_2 overnight.

Where transfection was performed on a smaller scale, a GenePulser Xcell^{TM} Electroporation System (Bio-Rad Laboratories) was used[4]. All volumes were doubled as standard, and 0.4 cm electroporation cuvettes were used. Shock parameters remained the same.

[4]For example, during optimisation of protocols and verification of plasmids. Data obtained following electroporation with the MXcell^{TM} and Xcell^{TM} systems were never compared directly.

2.3.8 Cell Staining

Cells were stained for flow cytometry using the LIVE/DEAD® Fixable Dead Cell Stain Kit in violet (Thermo Fisher Scientific™), and an ALEXA FLUOR® 647 conjugated mouse IgG2b antibody (clone BB7.2) against human HLA-A2 (Bio-Rad Laboratories). Staining was performed in 96-well plates, 24 hours after transfection to allow for a sufficient period of recovery. Briefly, 500 µl of cells were pelleted and washed twice with phosphate buffered saline (PBS) before being incubated on ice in the dark for 15 minutes with 0.2 µl of the LIVE/DEAD stain. The cells were washed twice in flow cytometry wash buffer and then incubated on ice in the dark for 20 minutes with 1 µl of the HLA-A2 antibody. The cells were again washed twice to remove unbound antibody before being re-suspended in 200 µl of flow cytometry fix buffer. Buffer compositions are listed in table 2.10.

Unstained and single-stained controls were prepared for each stain used in the panel. Flow cytometry was preferentially performed on the same day; however, if this could not be achieved the samples were wrapped and stored at 4℃ to be run within 48 hours.

2.3.9 Flow Cytometry

Flow cytometry was performed on a Becton Dickinson LSR Fortessa™ Analyser running BD FACSDiva™ software. A 96-well high-throughput sampler (HTS) attachment was used for large-scale experiments, whilst smaller runs were performed manually. Consistency of measurements was ensured by daily evaluation of the equipment by cytometer setup and tracking (CST) runs, and regular maintenance and servicing arranged by the flow cytometry facility team.

Compensation matrices were established in FACSDiva™ during data acquisition where required. Data generated were imported into FlowJo 10.1r5 for analysis. Cells were gated using the strategy shown in figure 2.3.

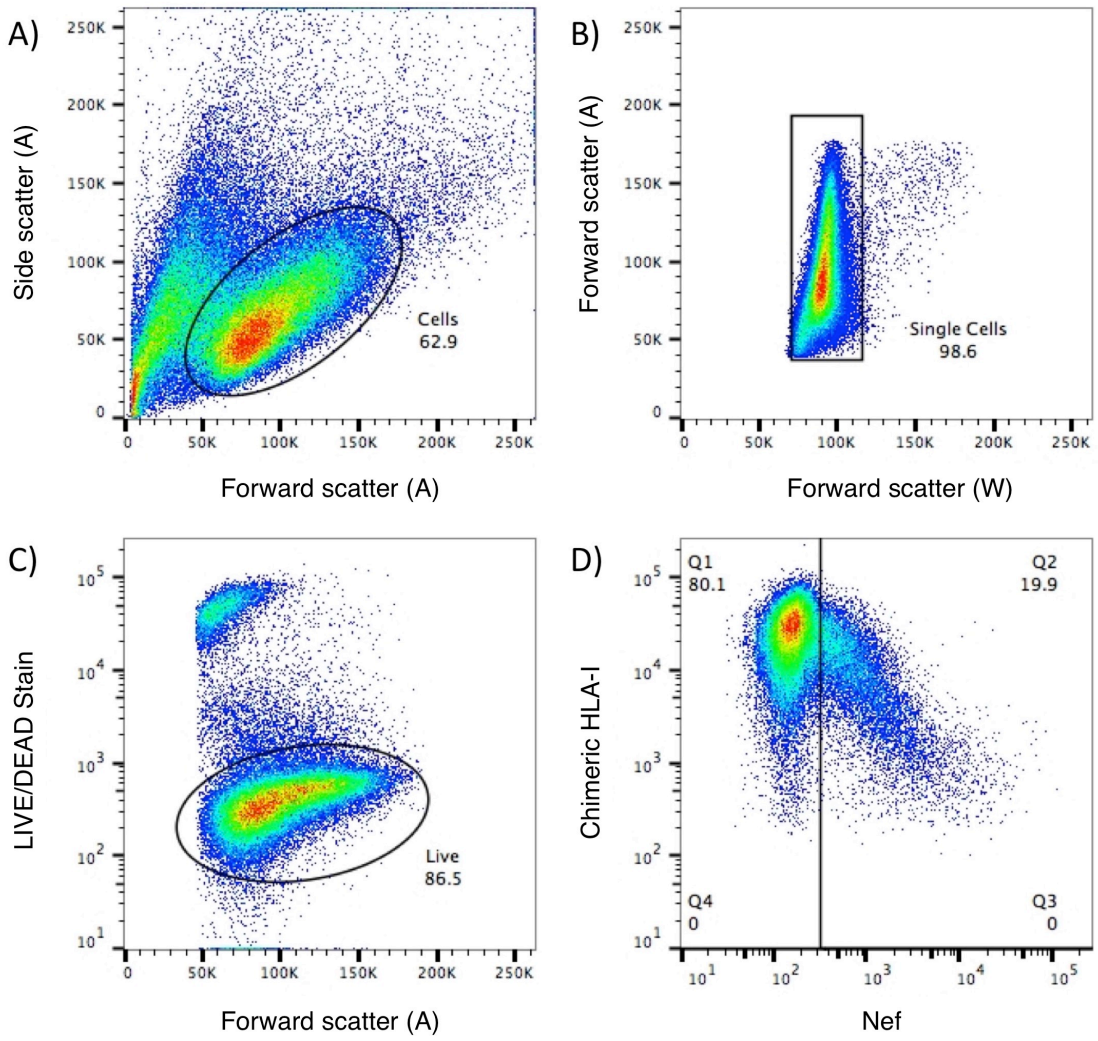


Figure 2.3: Gating strategy used to define populations of live, single cells following transfection. Plots show transgenic JJK-derived cell lines stably expressing chimeric HLA-I molecules, transfected by electroporation with fluorescently-labelled Nef isolates. Cells have been stained with the LIVE/DEAD® Fixable Dead Cell Stain Kit in violet (Thermo Fisher Scientific™), and an ALEXA FLUOR® 647 conjugated mouse IgG2b antibody (clone BB7.2) against human HLA-A2 (Bio-Rad Laboratories). Plot D differentiates between Nef negative (Q1) and positive (Q2) cells.

2.3.10 Analysis and Statistics

Gated flow cytometric data were exported for downstream analysis using R (v3.3.1) implemented through the RStudio (v.0.98.1103) IDE. Fold downregulation was calculated using equation 2.2.

$$\text{Fold Downregulation} = \frac{MFI\left(HLA_{Nef^{-}\text{ population}}\right)}{MFI\left(HLA_{Nef^{+}\text{ population}}\right)} \tag{2.2}$$

Homology modelling was performed and annotated using SWISS-MODEL and PyMOL, respectively, as described in table 2.5.

2.3.11 Generation of Flexible Expression Vector

A flexible expression vector was generated by modifying the pAcGFP1-N1 vector (Clontech TaKaRa). Briefly, the region spanning the EcoRI and AgeI restriction sites was removed by performing a digest with the appropriate enzymes. A custom synthetic gene was then designed to combine a Kozak sequence, a *nef* sequence (consensus HIV-1 or HIV-2, as described in section 2.3.1), a highly flexible linker based on that described by Waldo *et al.* [400], and a series of restriction sites for subcloning (EcoRI, BamHI, AgeI). The custom gene was ordered from Biomatik, digested with EcoRI and AgeI, and cloned into the vector as described in section 2.3.4.

The Kozak sequence was included to drive efficient translation and promote strong expression. The Waldo linker, which was initially developed to aid the folding of proteins in fusion with GFP [400], was modified to increase its length further and to remove existing restriction sites that would be incompatible with the pAcGFP1-N1 vector. The inclusion of BamHI between the synthetic *nef* gene and the flexible

linker allows different genes of interest (here, *nef* variants) to be cloned into the vector, replacing the existing gene without disturbing the integrity of the linker sequence.

2.3.12 SWARM Development

Limitations of bifurcate analyses are supported by examples using experimentally-collected flow cytometric data from the transfection of transgenic Jurkat-derived (JJK) cell lines with Nef:pAcGFP1-N1 constructs. Development and optimisation of SWARM was performed with training subsets of randomly selected samples. The samples used are indicated in the text.

The full dataset analysed in SWARM was a subset of that used in chapter 4. Namely, *nef* genes isolated from patients SM046, SM339, and TD001 (table 2.6), as well as SF2, G2A and consensus HIV-1 and HIV-2 *nef* (section 2.3.1). These isolates were cloned into the pAcGFP1-N1 expression vector (Clontech TaKaRa) and used to transfect the full panel of transgenic HLA-A- and HLA-B-expressing cell lines (appendix table A.4) as described previously (section 2.3.7). An empty pAcGFP1-N1 expression vector was also used as a negative control.

Samples were stained and measured by flow cytometry as described previously (sections 2.3.8 & 2.3.9). Live, single cells were gated in FlowJo 10.1r5 using the strategy shown in figure 2.3 A-C. The data were then exported for analysis.

The SWARM program was written in R (v3.3.1) implemented through RStudio (v.0.98.1103).

2.4 Materials and Methods for Chapter 6

2.4.1 *In Vitro* (Immuno)Proteasomal Digestion Assay

A series of four overlapping 30mer peptides spanning the central region of HIV-1 and HIV-2 Nef were designed using the consensus sequences described in section 2.3.1 as template. The peptides overlap by six amino acids at each end. These were synthesised by Biomatik on request at a purity of >99%. The amino acid sequences of these eight peptides are listed in table 2.9.

Table 2.9 Overlapping HIV-1 and HIV-2 Nef peptides synthesised for the (immuno)proteasomal digestion assay.

(a) HIV-1

Peptide	Sequence
HIV-1-p1	EAQEEEEVGFPVRPQVPLRPMTYKGALDLS
HIV-1-p2	GALDLSHFLKEKGGLEGLIYSQKRQDILDL
HIV-1-p3	QDILDLWVYHTQGYFPDWQNYTPGPGIRYP
HIV-1-p4	PGPGIRYPLTFGWCFKLVPVEPEKVEEANE

(b) HIV-2

Peptide	Sequence
HIV-2-p1	DSDDDDLVGVPVTPRVPLRAMTYKLAVDMS
HIV-2-p2	LAVDMSHFIKEKGGLEGMFYSERRHRILDI
HIV-2-p3	HRILDIYLEKEEGIIPDWQNYTHGPGIRYP
HIV-2-p4	PGIRYPMFFGWLWKLVPVDVPQEGEDNETH

Each peptide was incubated with either 20S human proteasome or 20Si human immunoproteasome (both Enzo Life Sciences Ltd.) at a concentration ratio of 5:1. 2.5 µg of peptide was incubated with 0.5 µg of enzyme in 50 µl of digestion buffer (table 2.10)

at 37 ℃ for 24 hours before termination by the addition of 5 µl formic acid. A zero time point was taken, as well as a negative control in which the enzyme was excluded.

The samples were then desalted using ZipTip® C18 pipette tips (Merck Millipore) according to the manufacturer's instructions. Briefly, samples were acidified with the addition of trifluoroacetic acid (TFA) to a final concentration of 0.1%. The ZipTip® was equilibrated by washing with buffer B followed by buffer A by repeated aspiration. The samples were then loaded to the ZipTip®. The ZipTip® was washed five times with buffer A, before the peptide was eluted into 10 µl of buffer B. Finally, the sample was dried in a Savant™ SpeedVac™ for one hour, before being re-suspended in 20 µl of buffer A. Buffer compositions are listed in table 2.10.

The prepared samples were then analysed by nano liquid chromatography quadrupole time-of-flight mass spectrometry (LC-Q-ToF-MS)[5]. Peptides with a Mascot MOWSE score greater than the standard significance threshold of 20 were retained for analysis.

2.4.2 Analysis and Statistics

Analysis was performed using custom scripts written in R (v3.3.1), implemented through the RStudio (v.0.98.1103) IDE.

2.4.2.1 Data Binarisation

(Immuno)proteasomal digestion data were binarised using a custom R script. All products were first aligned to HXB2 Nef in Geneious® v9.0.5 [377], implementing the

[5]The mass spectrometer was operated by Ms Marie-Laëtitia Thézénas, who also assisted with the data analysis.

MUSCLE algorithm [378]. Amino acids were then converted to 1 whilst all spaces in the alignment were converted to 0. For example:

-----EEVGFPVKPQV-------

to

00000111111111110000000

The full list of Nef CTL epitopes reported in the LANL Immunology Database [401] were also aligned to HXB2, and binarised in the same way. The LANL alignment was then truncated to correspond to the 102 aa region of study covered by the four overlapping peptides. The HXB2 sequence was removed to leave only the known epitopes. The data were then looped as follows:

```
For each alignment from each digestion:
  For each reported CTL epitope (i):
    For each digestion product (j):
      If i match j: 1 else 0
```

A series of binary matrices were generated whereby `1` indicated a perfect match and `0` a mismatch. The matrices were searched and all matches returned.

2.5 Materials and Methods for Chapter 7

2.5.1 Statement of Contribution

The work described in chapter 7 was a collaborative effort between researchers at the University of Hokkaido and the Rowland-Jones group at the University of Oxford. All steps - with the exception of the initial amplification and cloning of HIV-2 *nef* - were performed at the Faculty of Pharmaceutical Sciences, University of Hokkaido, Japan.

During a one month placement, I was specifically involved in: protein expression, His-tag purification, size exclusion chromatography, and preliminary crystallisation screening. All work during this placement was performed under constant supervision by and direct involvement of Mr Kengo Hirao: a student with whom primary responsibility of the structural project lies. Following my departure, successful crystallisation was achieved, and X-ray crystallography and structure determination was performed by Mr Kengo Hirao and Associate Professor Toyoyuki Ose. Throughout the placement, supervision, technical training, and guidance was provided by Professor Katsumi Maenaka, and Associate Professors Kimiko Kuroki & Toyoyuki Ose.

2.5.2 Gene Cloning & Protein Expression

The full-length sequence of Nef isolated from patient TD062 of the Caió cohort was amplified from proviral DNA and sequenced as described previously (section 2.3.3). Site-directed mutagenesis was performed to insert a C193Y substitution, and the core region (residues 90-256) was then subcloned into the pET-16b (Novagen®) expression vector. Residues 1-89 were excluded based on the inherent flexibility of the N-terminus, and on successful early crystallisation of N-terminally truncated HIV-1 Nef

[402, 403]. The C193Y substitution was introduced to prevent unwanted interactions between molecules that would interfere with crystallisation.

Rosetta™ 2 *E. coli* (Novagen®) were transformed with the prepared construct, and expression of the protein was induced with 0.1 mM of IPTG at 16℃ for 16 hours with 225 RPM shaking. Following incubation, 1 mL aliquots of bacterial suspension were taken, and the remaining cells were harvested by centrifugation for ten minutes at 4℃ at 5,000 RPM. The supernatant was discarded and the pellet re-suspended in a residual volume of 30 mL. The re-suspended pellet was concentrated by spinning down as before, and the clarified supernatant was aspirated.

The concentrated pellet was re-suspended in 30 mL of buffer P1, and sonicated for 15 minutes at duty cycle 40% with a three second pulse. The cells were pelleted by centrifugation for 15 minutes at 4℃ at 8,000 RPM. The supernatant was separated by decanting and the pellet re-suspended in 30 mL of buffer P1. 15 µl each of the re-suspended pellet, supernatant, and primary total cell suspension were then analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) to determine the solubility of the protein.

Briefly, the samples were mixed with 3 µl of 5X loading buffer and boiled at 95℃ for 5 minutes to denature the protein. 200 mM of dithiothreitol (DTT) was added, and the samples were loaded to two 12.5% SDS-PAGE gels alongside a protein standard. Electrophoresis was performed at 30 mA for 70 minutes. All recipes are listed in section 2.6. One gel was then stained with 0.1% Coomassie Brilliant Blue R-250 staining solution for one hour, before de-staining overnight in deionised water.

The other gel was used for a Western Blot. Briefly, a polyvinylidene difluoride (PVDF) membrane was hydrated in 100% methanol and soaked in transfer buffer. The gel and hydrated membrane were then assembled within the blotting cassette and placed into the buffer tank. The Western Blot was performed at $2\,\mathrm{mA\,cm^{-2}}$ for 30 minutes. The membrane was then submerged in blocking buffer for one hour, before being incubated with HisProbe$^{\mathrm{TM}}$-HRP Conjugate (Thermo Fisher Scientific$^{\mathrm{TM}}$) at a 5,000:1 ratio in tris-buffered saline with tween (TBST) for 1 hour at room temperature. The membrane was washed five times in TBST before the substrate reaction was performed: Amersham$^{\mathrm{TM}}$ ECL$^{\mathrm{TM}}$ Prime Western Blotting Detection Reagent (GE Healthcare Life Sciences) was added at a 1:1 ratio of luminol solution to peroxide solution, and incubated for five minutes at room temperature. All incubations were performed on a bench-top orbital shaker.

Chemiluminescence was then detected using an ImageQuant$^{\mathrm{TM}}$ LAS 4000 CCD Imager (GE Healthcare). The majority of the protein was found to be in the soluble fraction.

2.5.3 Protein Purification

The soluble protein was purified by a two-step process: first by nickel-nitrilotriacetic acid (Ni-NTA) agarose purification, and then by size-exclusion chromatography.

2.5.3.1 Affinity Chromatography

250 µl of Ni-NTA agarose resin was washed in 1.75 mL of re-suspension buffer (both QIAGEN®) and spun down at 13,000 RPM for 1 minute at 4°C in a bench-top centrifuge. The supernatant was decanted and the wash repeated. The resin was then re-suspended to a final volume of 2 mL in re-suspension buffer. The resin was added to 30 mL of cleared supernatant containing the protein, and imidazole was

added to a final concentration of 10 mM. The sample was incubated for one hour at 4℃ on a bench-top orbital shaker.

A column was then loaded with re-suspension buffer, followed by the resin, which was washed five times with 1 mL of 10 mM imidazole. The sample was then eluted with increasing concentrations of imidizole ranging from 50 mM to 1 M. All fractions were collected and analysed by SDS-PAGE (section 2.5.2).

The His-Tag was cleaved from the purified protein by factor Xa at 20℃ overnight, by Mr Kengo Hirao.

2.5.3.2 Size Exclusion Chromatography

The protein was secondarily purified by size exclusion chromatography on a Superdex 75 10/300 GL column attached to an ÄKTA™ chromatography system, running UNICORN™ 6.3 (all GE Healthcare Life Sciences).

Briefly, the chromatography system and sample loop were de-aerated and washed with purification buffer (section 2.6). The column was attached, and the system programmed to a flow rate of $0.5\,\mathrm{mL\,min^{-1}}$ with fractionation every 0.5 mL. The sample was filtered through a $0.22\,\mu\mathrm{m}$ Ultrafree®-CL centrifugal filter (Merck Millipore), before being injected into the sample loop. The sample was then run to completion, and fractions containing the eluted protein and no observable contamination were retained.

The eluted protein was concentrated by centrifugation in an Amicon Ultra-15 Centrifugal Filter (Merck Millipore) at 4℃ at 4,500 RPM for 45 minutes. The purity of the protein was again verified by SDS-PAGE (section 2.5.2).

2.5.4 Crystallisation & Structure Determination

Crystallisation screening was performed using the PACT Suite, PEGs Suite (both QIAGEN®), and Crystal Screen™ 1 and 2 Kits (Hampton Research). Sitting drop vapour diffusion was set-up using a mosquito® Crystal liquid handler (TTP Labtech).

Crystallisation was achieved by Mr Kengo Hirao following my departure from Japan (10-20% PEG, 10.9 mg mL^{-1} Nef, 0.1 mM MES pH 6.5). The crystal was cryoprotected, and X-ray crystallography was performed using the PF BL5A Beamline at the Structural Biology Research Center in Tsukuba, Japan. Diffraction intensities were measured using a ADSC QUANTUM 315r detector.

Structure determination was performed by molecular replacement of HIV-1 (RCSB PDB accession number 1AVV) and SIV Nef (3IK5) structures, by Mr Kengo Hirao and Associate Professor Toyoyuki Ose (manuscript in preparation).

2.6 Reagents and Buffers

Manufacturer details of all kits and non-generic reagents have been listed in the text where appropriate. The composition of buffers and media prepared in-house are described in table 2.10.

Table 2.10 The composition of buffers and media prepared in-house.

Name	Reagent	Amount	Chapter
Buffers			
Plasmid extraction buffers			4,5
P1	Trisaminomethane (Tris; pH 7.4)	50 mM	
	Ethylenediaminetetraacetic acid (EDTA)	10 mM	
P2	Sodium hydroxide	200 mM	
	Sodium dodecyl sulphate (SDS)	1%	
P3	Potassium acetate (pH 5.5)	3 M	
Molecular biology			3,4,5
TE	Tris (pH 8.0)	10 mM	
	EDTA	1 mM	
TBST	Tris (pH 7.5)	20 mM	
	Sodium chloride	150 mM	
	Tween 20	0.1%	
Flow cytometry			4,5
Wash	Phosphate buffered saline (PBS)		
	Foetal calf serum (FCS)	2%	
	Sodium azide	0.1%	
Fix	Wash buffer (above)		
	Formaldehyde	1%	
Purification buffer			7
	Tris (pH 7.4)	20 mM	
	Sodium chloride	200 mM	

Continued on next page

Table 2.10 – *Continued from previous page*

Name	Reagent	Amount	Chapter
	tris(2-carboxyethyl)phosphine (TCEP)	0.3 mM	
(Immuno)proteasomal digestion & desalting buffers			6
Digest	HEPES (pH 7.8)	20 mM	
	Magnesium chloride	5 mM	
	Dithiothreitol (DTT)	2 mM	
A	Acetonitrile	2%	
	Trifluoroacetic acid (TFA)	0.1%	
B	Acetonitrile	65%	
	Trifluoroacetic acid (TFA)	0.1%	
Media			
Microbiological growth media			3,4,5
LB	Tryptone	$10\,\mathrm{g\,L^{-1}}$	
	Sodium chloride	$10\,\mathrm{g\,L^{-1}}$	
	Yeast extract	$5\,\mathrm{g\,L^{-1}}$	
TB	Tryptone	$12\,\mathrm{g\,L^{-1}}$	
	Yeast extract	$24\,\mathrm{g\,L^{-1}}$	
	Dipotassium phosphate	$9.4\,\mathrm{g\,L^{-1}}$	
	Monopotassium phosphate	$2.2\,\mathrm{g\,L^{-1}}$	
	Glycerol	0.8%	
Mammalian cell growth media			4,5
R10	RPMI 1640		
	Foetal calf serum (FCS)	10%	
	L-glutamine	2 mM	
	Penicillin-Streptomycin	50 U $\mathrm{mL^{-1}}$	
H10	RPMI 1640		
	Human serum	10%	
	L-glutamine	2 mM	

Continued on next page

Table 2.10 – *Continued from previous page*

Name	Reagent	Amount	Chapter
	Penicillin-Streptomycin	50 U mL^{-1}	
SDS-PAGE			
SDS-PAGE Resolving Gel			7
	30% Acrylamide 0.8% Bisacrylamide	3.3 mL	
	1.5 M Tris (pH 8.8)	2.0 mL	
	10% SDS	100 µL	
	Tetramethylethylenediamine (TEMED)	20 µL	
	10% Ammonium persulphate (APS)	60 µL	
	Nuclease-free water	2.6 mL	
SDS-PAGE Stacking Gel			7
	30% Acrylamide 0.8% Bisacrylamide	480 µL	
	0.5 M Tris (pH 6.8)	1.2 mL	
	10% SDS	48 µL	
	TEMED	5 µL	
	10% APS	50 µL	
	Nuclease-free water	1.26 mL	
SDS-PAGE 5X Sample Buffer			7
	0.125 M Tris (pH 6.5)	0.75 g	
	25% Glycerol	12.5 mL	
	5% SDS	2.5 g	
	0.25% Bromophenol blue	0.125 g	
	2-mercaptoethanol	5%	
	Nuclease-free water	to 500 mL	
SDS-PAGE 10X Running Buffer			7
	0.25 M Tris (pH 6.5)	30.25 g	
	1.92 M Glycine	144 g	
	10% SDS	10 g	

Continued on next page

Table 2.10 – *Continued from previous page*

Name	Reagent	Amount	Chapter
	Nuclease-free water	to 1 L	
Western Blotting			
Transfer Buffer			7
	Tris-HCl	25 mM	
	Glycine	192 mM	
	Methanol	20%	
Blocking Buffer			7
	TBST		
	Skimmed milk powder	5%	

Chapter 3

Selection and Mutation in Gp120 Envelope in Clinically Latent HIV-1 Infection

3.1 Introduction

HIV-1 gp120 is a 120 kDa surface-expressed glycoprotein that is essential for viral entry into the cell (section 1.1.8.4). It is encoded by the *env* gene, and consists of five variable regions (V1-V5) interspersed between conserved regions (C1-C5) [404]. Gp120 forms heterodimers with gp41 and the complexes subsequently trimerise and stud the viral membrane at a density of around fourteen copies per virion [111]. As the glycoprotein is exposed on the surface of the virion, it is a key target of the adaptive immune response. Indeed, the majority of neutralising antibodies against HIV-1 are targeted against gp120 [405].

Gp120 is therefore under substantial pressure to escape recognition and killing. It is a hotspot of immune evasion, utilising many different techniques to hide from the adaptive response. Rapid genetic mutation resulting from the combined effects of error-prone reverse transcriptase (section 1.1.8.3) and a short generation time leads to a staggering degree of diversity at both the interpatient and intrapatient level. This rapid evolution generates "escape mutations" in viral epitopes.

Gp120 also exhibits a degree of structural flexibility. Revertible conformational changes are required for membrane fusion and viral entry to the cell (section 1.1.6.1), but have also been proposed to facilitate immune evasion by shielding conserved motifs from antibody recognition: particularly at functional sites where escape mutations would generate defective virus [406]. Indeed, recognition of receptor-binding-sites by targeted antibodies has been shown to induce conformational changes in gp120 [407], consistent with the masking of these sites.

Additionally, gp120 is masked by a "glycan shield" consisting of host-derived N-linked glycans which make up approximately half of its molecular mass [408]. The role of the glycan shield is complex: initially it was believed to act as a cloaking device, hindering access of antibodies to viral epitopes [128]. As the glycans are produced by host cell machinery and are thus "self" molecules, they are in theory poorly immunogenic [405]. Despite this, there is some recent evidence that viral glycans may also function as a target of broadly neutralising antibodies [409]. The glycan shield is highly dynamic, exhibiting a shifting pattern of glycan attachment that evolves over time [128].

In this chapter, the role of genetic mutation and viral evolution in the context of clinically latent HIV-1 infection will be assessed. Whilst this topic has been previously studied (key examples being Shankarappa *et al.* [410], Markham *et al.* [411], Lemey *et al.* [412], amongst others), a potentially confounding factor of such studies has been the extensive degree of interpatient viral diversity. As cohorts typically consist of patients infected at different times with genetically divergent virus, inference of the contribution of immunological pressure on viral diversity is difficult to detect conclusively over underlying variation resulting from artefacts of the nature of infection. Indeed, such factors may explain the apparent contradictions in findings

of these studies - particularly those pertaining to evolutionary rates and correlates of disease progression - reviewed by Esbjörnsson [413].

What distinguishes this study from previous works is the use of the SM cohort from Henan province, China (section 2.1.1). The patients of this cohort were intravenously infected with a narrow (purportedly single) source of virus through exposure to contaminated blood products and/or equipment during a paid plasma donation scheme in the mid-1990s. Owing to the unusual homogeneous route and source of infection, and the narrow time-frame during which patients were exposed to virus, the cohort is a useful resource for studying the effects of individual patient immune responses with fewer conventional confounding factors.

Whilst this cohort is a valuable resource, the infecting virus has not yet been comprehensively characterised. In 2011, Dong *et al.* [414] described sequence diversity in *gag*, *pol* and *nef*, but *env* was not include in their study. This study will therefore characterise gp120 *env*, use the obtained sequences to further investigate the unique nature of the cohort, and assess the influence of immunological pressure on sequence evolution.

3.1.1 Aims

The specific aims of this project were to:

- Characterise the intrapatient and interpatient sequence diversity of a gp120 *env* fragment in a cohort of clinically latent HIV-1 patients.
- Investigate gp120 *env* mutation in this cohort to map regions under immunological selection pressure.

3.2 Results

3.2.1 Epitope Mapping of Envelope

As gp120 is a hotspot of immunological activity, to narrow down this study the protein was first mapped to determine which regions were the most epitope-dense. The Los Alamos National Laboratory (LANL) maintains an up-to-date database of CTL, antibody, and T-helper cell epitopes reported in the literature [415–417]. Full epitope alignments mapped to HXB2 were downloaded from the LANL Immunology Database, and heatmaps were constructed, whereby colour intensity denotes the absolute number of epitopes spanning each position in the alignment (figure 3.1).

In terms of antibodies (figure 3.1A), the most epitope-dense region of gp120 is contained within the V3 loop, with substantially lower levels of activity elsewhere in the protein. CTL epitopes were less clearly clustered than antibody epitopes (figure 3.1B), but again, a peak in activity could be detected within the V3 variable loop. In addition, a further cluster of CTL epitopes can be detected towards the N-terminus of the protein. Outside of these clusters, epitopes are relatively evenly distributed through the conserved regions, but notably absent in the remaining variable loops: particularly V1 and V4. Lastly, the T-helper cell epitopes were also prevalent in the V3 loop (figure 3.1C), but a more intense cluster of activity flanked V2 C-terminally. Further clusters could be found at the N-terminus of the protein, and C-terminally to V4 and V5.

These data demonstrate that numerous epitopes - particularly antibody and CTL epitopes - cluster within V3, and so this region of gp120 was selected as the central focus area for the study.

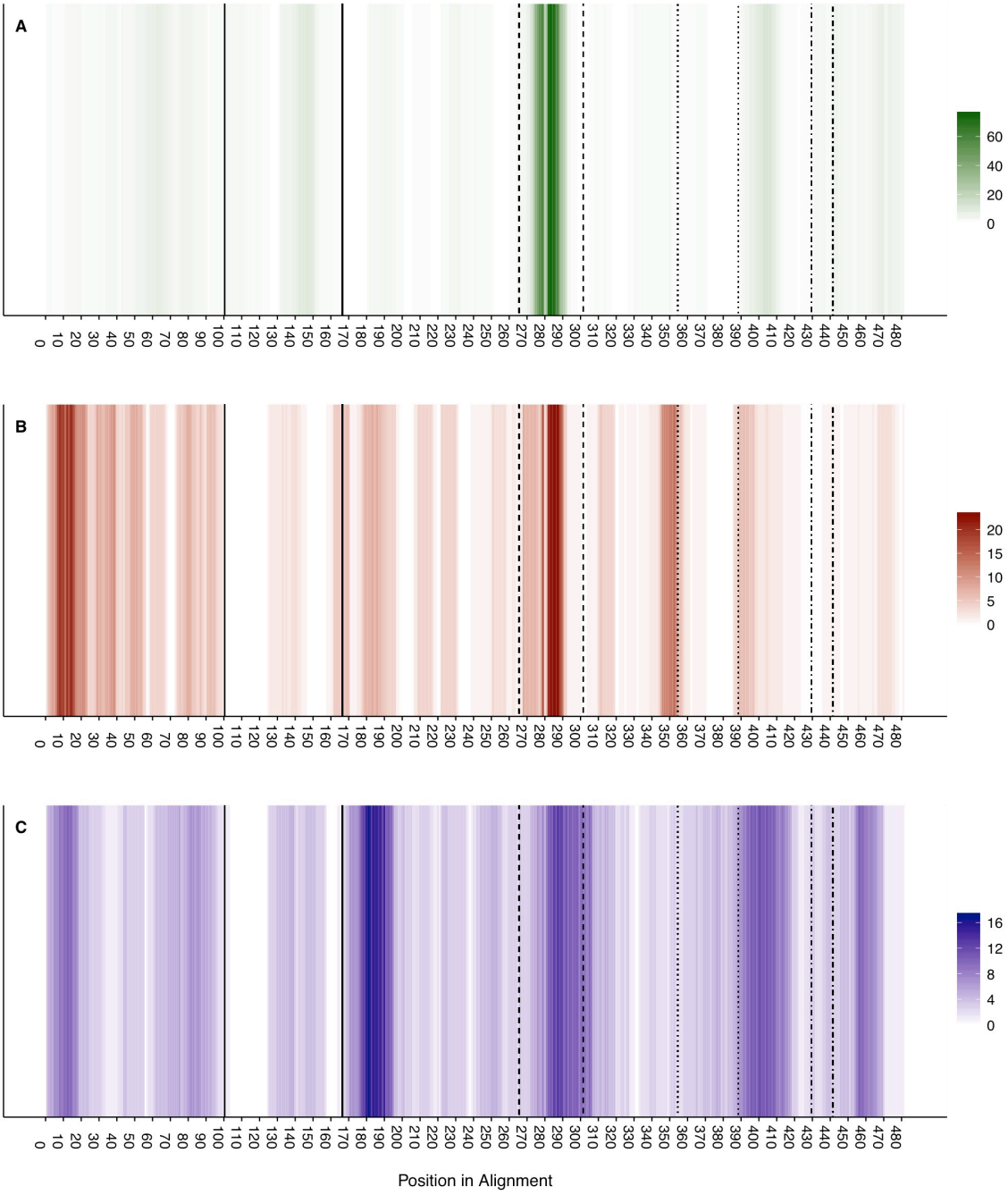


Figure 3.1: Map of the known A) antibody, B) CTL, and C) T-helper cell epitopes of the gp120 envelope protein. The location of each epitope is defined by its relative amino acid position in the reference strain HXB2. Colour intensity denotes number of epitopes spanning that residue reported in the LANL database. Variable loops are flanked by black lines: V1V2 solid, V3 dashed, V4 dotted, and V5 dot-dashed. Figure constructed from data made available by The Los Alamos National Laboratory (LANL) [415–417].

3.2.2 SM Cohort Characterisation

Amplification and sequencing of the gp120 *env* fragment - described in section 2.2 - generated a final dataset of 588 sequences. PCR amplification was successful for 28 patient time-points of the 40 listed in table 2.2, and success was heavily dependent on viral load. Of the twelve patient time-points from which product could not be obtained, viral load was <50 for four samples (SM358 2010, SM538 2011, SM538 2012, SM176 2014), <200 for another sample (SM505 2014), and unrecorded for a final sample (SM446 2010). For two further samples (SM007 2011, SM209 2012), plasma had run out prior to initiation of the study and so no material was available to work with. The remaining four samples did not yield product after three attempts at amplification (SM176 2011, SM335 2011, SM514 2010, SM514 2012), perhaps indicating substantial mutation within the primer-binding sites in these patients, or degradation of plasma resulting from prior handling.

Sequences were obtained from nine patients sampled in 2010, seven patients in 2011, nine patients in 2012, and three patients in 2014. Six patients gave rise to sequences from three time points; four patients yielded two time points, and the remaining two patients yielded only a single usable time-point.

3.2.2.1 Viral Subtyping

Viral suptyping was performed using a phylogenetic approach. Patient-specific alignments were generated as described, and an unambiguous consensus sequence was produced for each. The sequences were then aligned to the 2005 LANL gp120 reference dataset (accession numbers: appendix table A.8), and the phylogenetic relationship was determined by Bayesian Markov chain Monte Carlo (MCMC) inference. Based on the results of jModelTest 2.1.7, the generalised time-reversible (GTR) nucleotide substitution model plus invariant sites and γ-distributed rate heterogeneity (GTR+I+G)

was selected. A constant size coalescent, estimated base frequencies, and a strict molecular clock were used.

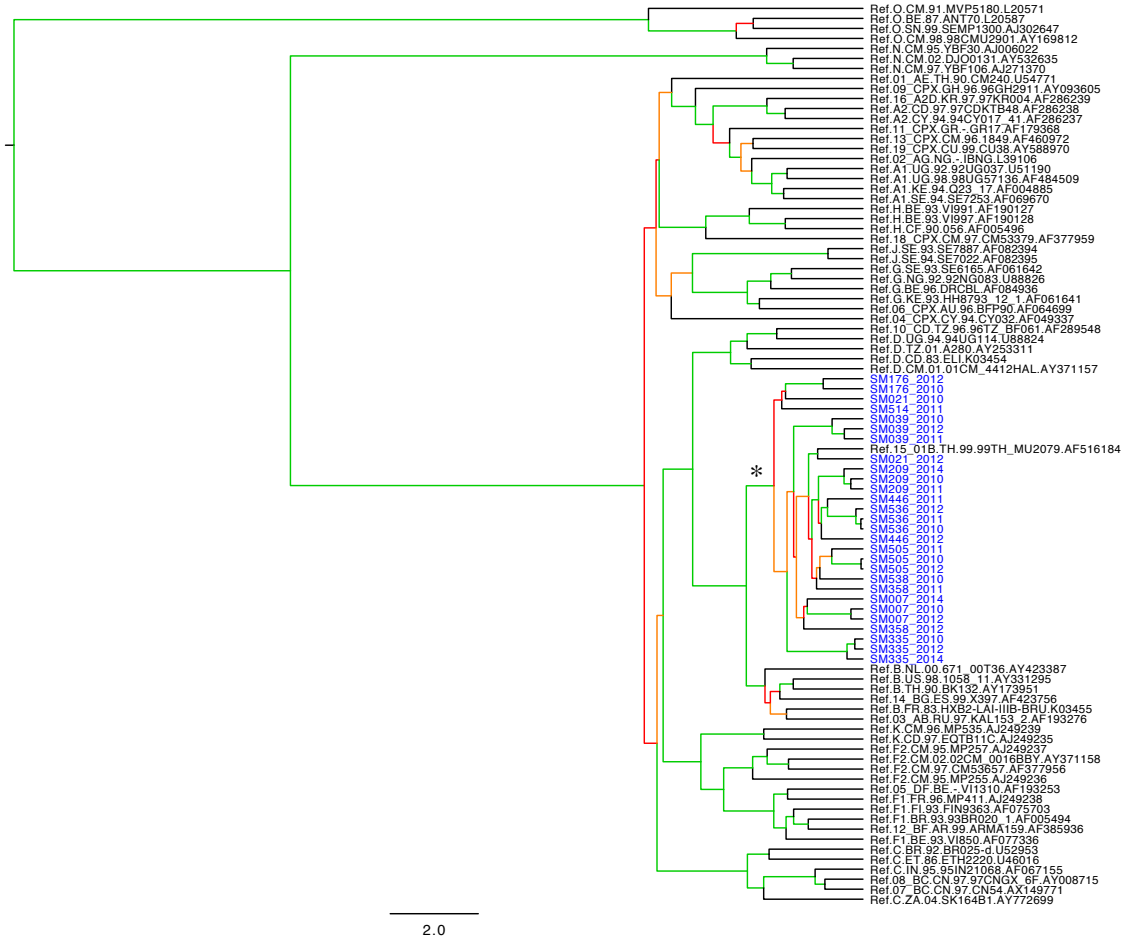


Figure 3.2: Phylogenetic subtyping of SM cohort viral sequences represented as a midpoint-rooted maximum clade credibility (MCC) tree of intrapatient consensus sequences (blue) aligned to the 2005 Los Alamos National Laboratory (LANL) gp120 reference dataset (black) [376]. Support (posterior probability) for internal nodes is indicated by branch colour. Support of $\geq$0.95 is green; $\geq$0.75 is orange; and $<$0.75 is red. Support for the node associating all SM cohort sequences with CRF 15-01B (marked by an asterisk) is 1.

A burn in of 2,500 trees (25% of the posterior distribution) were excluded, and TreeAnnotator v1.8.2 was used to determine the maximum clade credibility (MCC) tree from those remaining. This was annotated in FigTree v1.4.2, and is presented in figure 3.2. All patient time points clustered strongly (maximal branch support 1)

with the sequence from CRF 15-01B (accession number: AF516184): a circulating recombinant virus initially reported in Thailand and composed of CRF01-AE with a subtype B envelope.

3.2.2.2 Predicted Co-Receptor Usage

A switch in co-receptor usage from CCR5 (R5) to CXCR4 (X4) occurs in the virus of many patients during late-stage disease. This is associated with accelerated $CD4^+$ cell decline and increased rates of disease progression. The proportion of patients in which this switch occurs is heavily influenced by the subtype of the virus, and may be as high as 76% [94]. As an unusually high proportion of patients within the SM cohort are slow progressors (section 2.1.1), it was hypothesised that virus within these patients would be less likely to have undergone co-receptor switching.

V3 is an important determinant of co-receptor usage [418, 419], and the web position-specific scoring matrices (WebPSSM) tool can be used to predict co-receptor usage from amino acid sequences [396]. Running the V3 segments of each consensus for each patient and time-point through WebPSSM revealed that only the virus from patient SM514 (sampled 2011) had undergone an R5 to X4 shift. Whilst it was unfortunately not possible to amplify further sequences from patient SM514 in 2010 and 2012 to support this prediction more robustly, it is notable that patient SM514 did indeed undergo a rapid decline in $CD4^+$ lymphocytes following the 2011 bleed (66% to 7%; table 2.2).

3.2.2.3 Estimation of Infection Date

The patients of the SM cohort are believed to have become infected between 1993 and 1995. To substantiate this claim, Bayesian phylogenetics were implemented to estimate the date the virus was introduced to the cohort, by inferring the time

to the most recent common ancestor (MRCA). One time-point was selected per patient - with an approximately even distribution of time-points across the sampling period - and sequences were then aligned in a step-wise manner. Patient SM007 was excluded because it was not possible to conclusively rule out dual- or superinfection as preliminary data exploration revealed that sequences did not resolve monophyletically. The alignment was manually edited to ensure codon integrity, and was then stripped to remove gaps (final length of 759 nucleotides).

A consensus sequence was generated from each patient time-point using the LANL Ambiguity Consensus Maker [420], with an ambiguity threshold of 10%. The divergence time was then estimated in BEAST v.1.8.2, using the SRD06 nucleotide substitution model [421] with tip-dating, and without codon partitioning. An uncorrelated lognormal relaxed molecular clock with a rate prior of 0.001 was implemented with a Bayesian SkyGrid tree prior [422]. A burn-in of 10-35% was applied, dependent on when convergence was reached.

The results of three runs were combined in LogCombiner v1.8.2, and the mean root height was calculated at 15.77 years (95% HPD 7.88-25.98), placing the date of convergence at around January 1999: slightly later than the window of plasma donation. However, the 95% HPD spanned October 1988 to November 2006, comfortably accommodating the period in question.

3.2.2.4 Inference of Founder Virus

As the SM cohort patients were infected with a narrow (purportedly single) source of virus, the most recent common ancestor (MRCA) of the phylogeny is likely a good estimation of the founder viral sequence. The reconstructed ancestral states at the tree root from each run following burn-in were therefore extracted, and an ambiguity

consensus [420] was constructed from the alignment (26,300 sequences). The inferred founder sequence is listed in appendix table A.2.

Aligning the inferred founder to a consensus of all 588 sequences generated in this study (figure 3.3A) confirms that this approach has generated a sensible viral sequence. Eight sites in the alignment could not be matched due to ambiguity, seven sites had been excluded due to the gap-stripping process, and three mismatches exist: one within each variable loop.

The inferred founder was next aligned to a consensus generated from the 512 Henan FPDs sequences from the LANL Sequence Database described in section 2.2.3 (accession numbers: appendix table A.8). The sequences were found to be remarkably similar (figure 3.3B). Again, seven sites could not be matched due to ambiguity, and gaps are present due to stripping, but in the alignment of 250 amino acids, mismatches occur at only seven positions. Four of these positions fall within the hypervariable loops V4 and V5, and the remaining three mismatches all occur towards the N-terminus of the amplicon.

As many previous studies have focused on a smaller stretch of gp120, coverage of this region in the alignment is particularly low (figure 3.3C), falling to just 91 sequences in places, decreasing the certainty with which the consensus residue can be identified. However, the similarity between the inferred founder virus and the consensus sequence generated from geographically distinct FPD cohorts in Henan province is striking, and is consistent with a shared source of virus. To highlight the extent of the sequence similarity, aligning any two random sequences (isolated from two distinct patients) from within the SM cohort generates a substantial degree of mismatch (figure 3.3D).

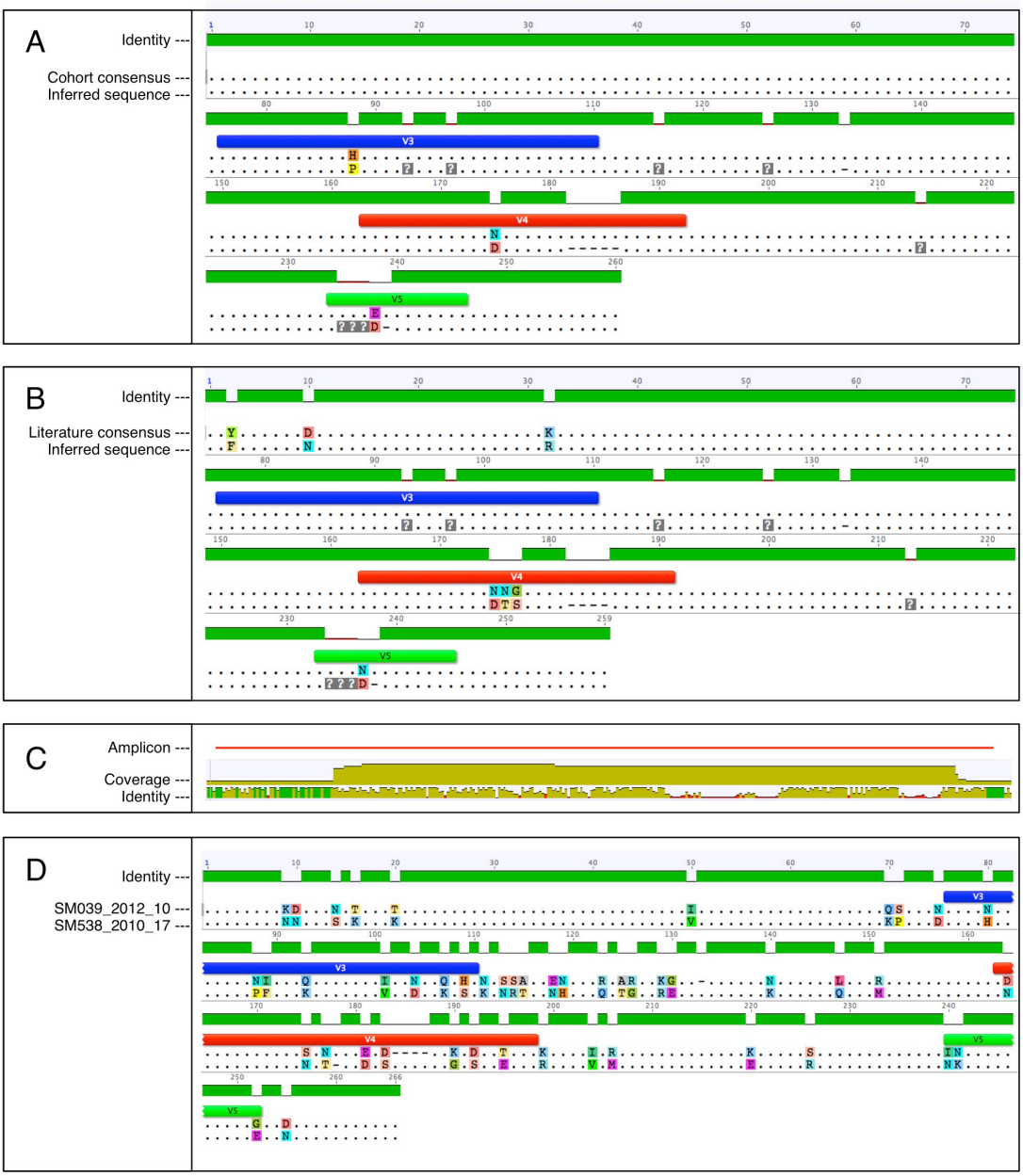


Figure 3.3: Alignment of the inferred founder sequence with **A)** the SM cohort consensus (512 sequences), **B)** and the literature consensus generated from 512 Henan FPD sequences. **C)** Coverage map denoting the number of sequences from which the literature consensus was generated across each residue. **D)** Alignment of two randomly selected sequences from within the SM cohort.

3.2.2.5 Viral Population Dynamics

The effective population size (EPS) trajectory within the cohort was estimated by generating a non-parametric Bayesian SkyGrid plot [422]. The EPS is the smallest idealised population that can evolve in the same way as an actual population under observation. Changes in the EPS over time provide a crude estimate of the relative number of new transmission events occurring, and the rate of spread of a virus within a population [423]. A SkyGrid reconstruction was implemented through Tracer v.1.6.0 from the three BEAST runs described in section 3.2.2.3 (figure 3.4).

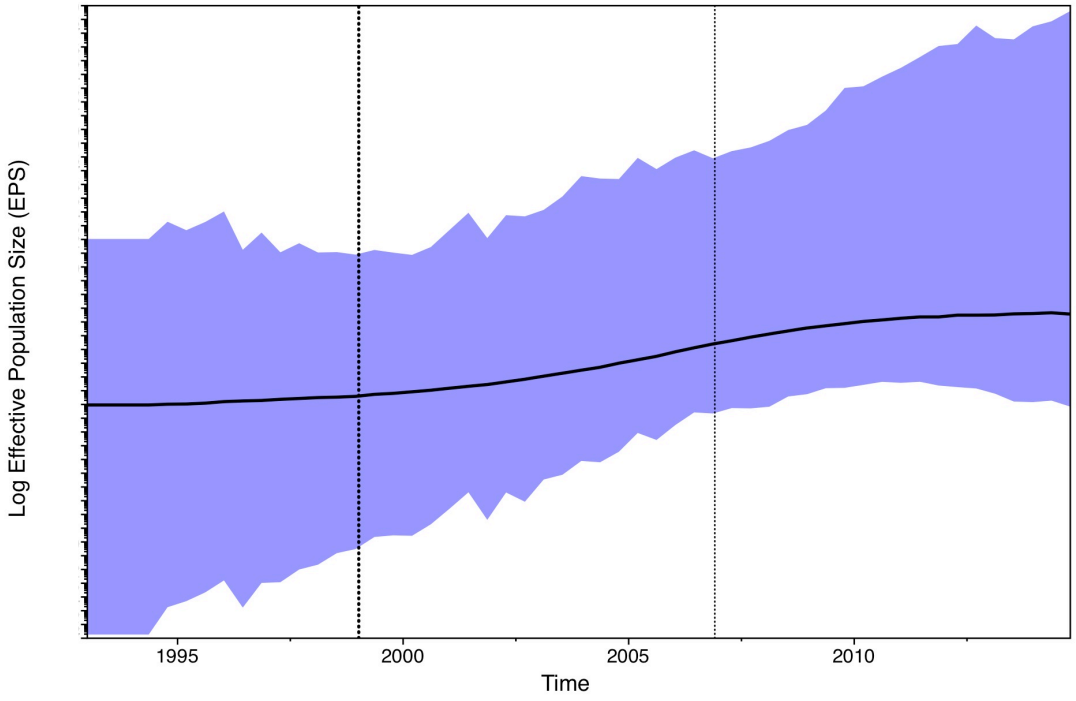


Figure 3.4: Viral population dynamics within the SM cohort. Bayesian SkyGrid estimates of past population dynamics showing the effective population size trajectory. The mean time to MRCA is marked by the leftmost dotted horizontal line, with the faint rightmost line marks the upper limit of the 95% HPD. Actual transmission events are believed to have occurred between 1993 and 1995.

The SkyGrid plot does not show evidence of extensive spread within the population following the initial outbreak of the infection in the mid 1990s, which is consistent with the limited extent of spread between sexual partners reported in this cohort (Dong, T., personal communication).

3.2.3 Evolution and Selection Pressure

Having been used to further characterise the SM cohort, the sequences were next used to investigate evolution and selection pressure.

3.2.3.1 Intrapatient Evolutionary Rates

For each of the ten patients from whom serially sampled viral sequences were obtained (section 3.2.2), viral evolutionary rates were estimated by Bayesian inference in BEAST v.1.8.2. Individual patient alignments were first constructed with MUSCLE implemented in Geneious® v6.1.6. The HKY85 nucleotide substitution model [385] with invariant sites and γ-distributed rate heterogeneity (HKY85+I+G) was used. This was the least parameter-rich model selected by JModelTest 2.1.7 for any of the sequence alignments subset by patient, and as direct comparison of clock rate is desired, the same model was applied to all. A strict molecular clock and codon partitioning ((1,2)3) were employed, and longitudinally collected samples were time-stamped. The nucleotide substitution rates are shown in figure 3.5.

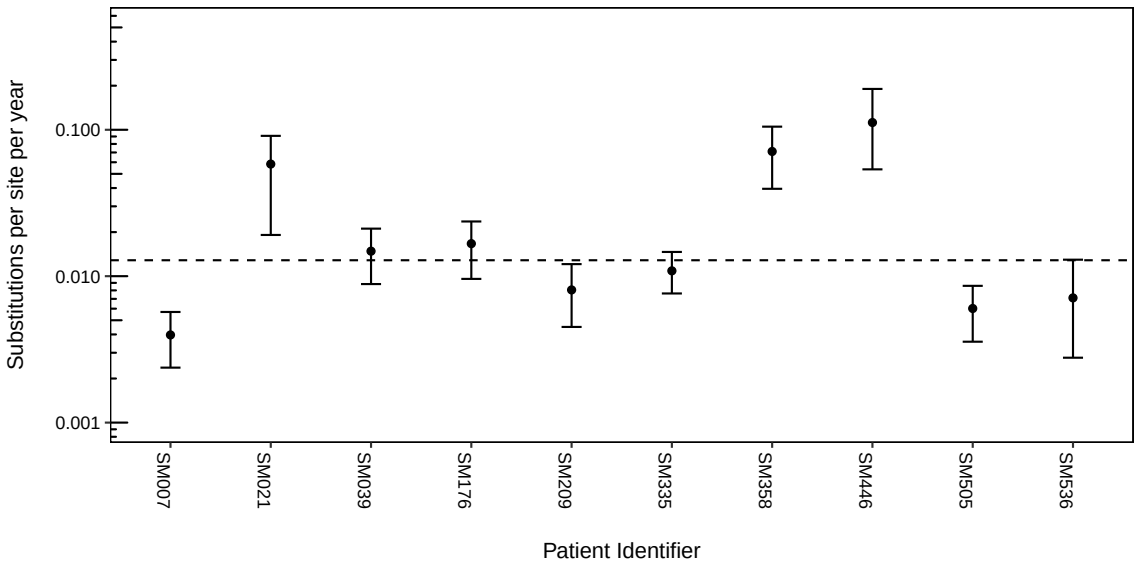


Figure 3.5: Evolutionary rates of virus from the ten patients with serially collected samples in substitutions per site per year. Error bars denote ± 95% HPD. Median rate is indicated by the dashed line.

The median nucleotide substitution rate was calculated at 12.9×10^{-3} substitutions per site per year (IQR 7.35×10^{-3} - 47.9×10^{-3}), similar to reported rates of evolution in gp120 [424]. As four of the patients (SM007, SM039, SM209 & SM335) had begun antiretroviral treatment by the latest bleeds, the samples were stratified accordingly, but no statistically significant difference between the treated and untreated groups was identified by a Wilcoxon rank sum test (W = 18, $p = 0.257$).

As substantial differences in substitution rate were observed between patients, the association between rate and a range of clinical parameters was next measured. Stratifying the patients by the highest recorded viral load revealed no statistically significant difference in rate between those above and below the median (W = 11, $p = 0.841$), nor did stratifying according to whether the viral load was increasing or decreasing during the period of study (W = 19, $p = 0.171$).

No statistically significant association was identified by computing Spearman's rank correlation coefficients between substitution rate and: peak CD4% (figure 3.6A; r_S = -0.086, $p = 0.813$), rate of CD4% decline (figure 3.6B; r_S = -0.140, $p = 0.699$), peak viral load (figure 3.6C; r_S = -0.467, $p = 0.174$), or change in viral load over the period of study (figure 3.6D; r_S = -0.491 $p = 0.150$).

Codon partitioning also allowed the relative evolutionary rate ratio of positions 1+2 to position 3 to be calculated. A 1+2:3 ratio greater than 1 indicates positive selection whilst a ratio less than 1 indicates purifying selection. The median 1+2:3 rate ratio for the amplicon was 0.861:1 (IQR 0.779 - 0.892) indicating that the fragment is on average under weak purifying selection pressure. The ratio was <1 for virus from all patients except SM021, which had a ratio of 1.306:1. Notably, this patient had a reasonably high viral load in both 2010 and 2012 (35,827 and 14,782 copies ml^{-1},

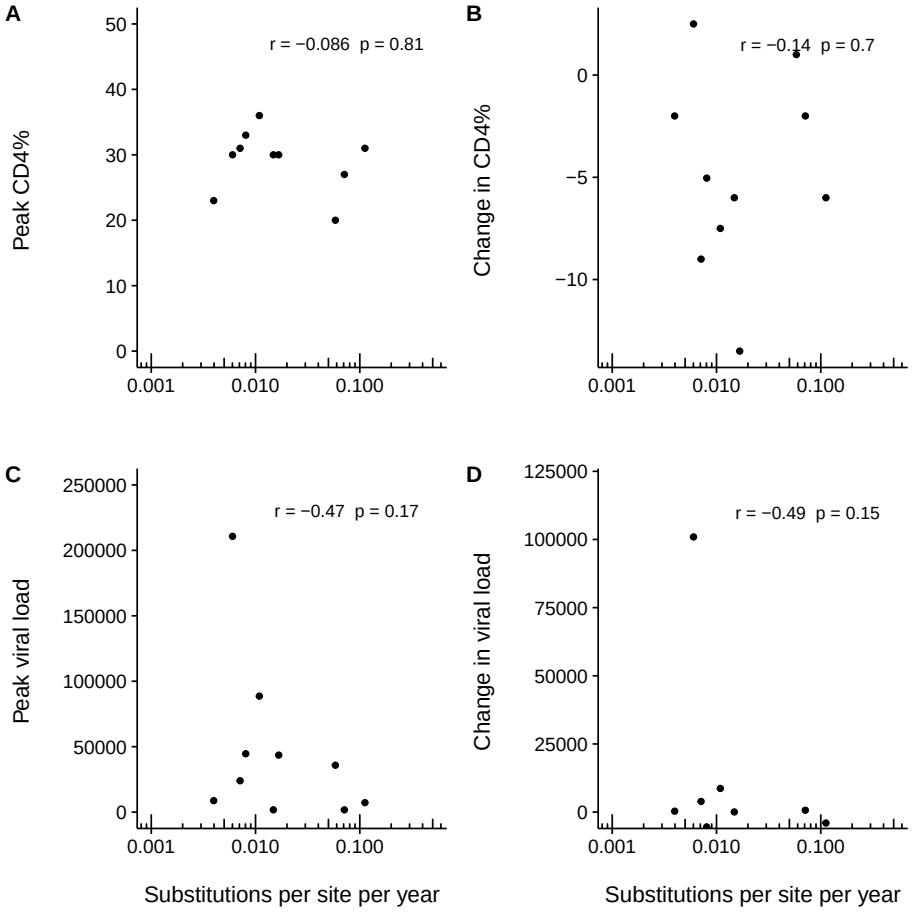


Figure 3.6: Clock rate and clinical parameters. No statistically significant association was identified between substitution rate and **A)** peak CD4%, **B)** CD4% decline, **C)** peak viral load, or **D)** change in viral load over the period of study.

respectively) despite no decline in CD4% over the period of study (18% to 20%) in the absence of treatment, perhaps driving more extensive positive selection.

3.2.3.2 Selection Pressure

To map selection pressure more precisely, the ratio of non-synonymous (dN) to synonymous (dS) mutations along the length of the alignment was estimated for each individual codon. A dN/dS score greater than 1 indicates positive selection as non-synonymous mutations are accumulating at a faster rate than synonymous mutations. Conversely, a score of less than 1 indicates purifying selection. BEAST v.1.8.2 was used to implement renaissance counting [425, 426] to estimate dN/dS for each codon in an alignment for each patient from whom longitudinal sequences were collected. The HKY85 nucleotide substitution model [385] was selected, with three-site codon partitioning, a strict molecular clock, and tip-dating of time stamped sequences. The resulting dN/dS estimates were extracted and are plotted for each patient in figures 3.7 & 3.8.

Residues were classed as under significant selection if the range of the 95% HPD did not encompass 1. If the 95% HPD encompassed 1, neutral evolution is assumed. The overwhelming majority of amino acids in the gp120 amplicon were under significant positive or purifying selection in seven of the ten patients. Consistent with the 1+2:3 codon rate ratios, only patient SM021 appeared to have more residues under significant positive than purifying selection, and in all patients selection is weak (approaching 1) in most positions.

To identify specific residues under significant selection pressure across the cohort, a step-wise alignment was constructed from these data - due to differences in the lengths of the variable loops across individuals - and the dN/dS estimates were combined across each position. From these data, the proportion of patients with virus showing evidence of significant selection pressure in each codon was calculated (figure 3.9).

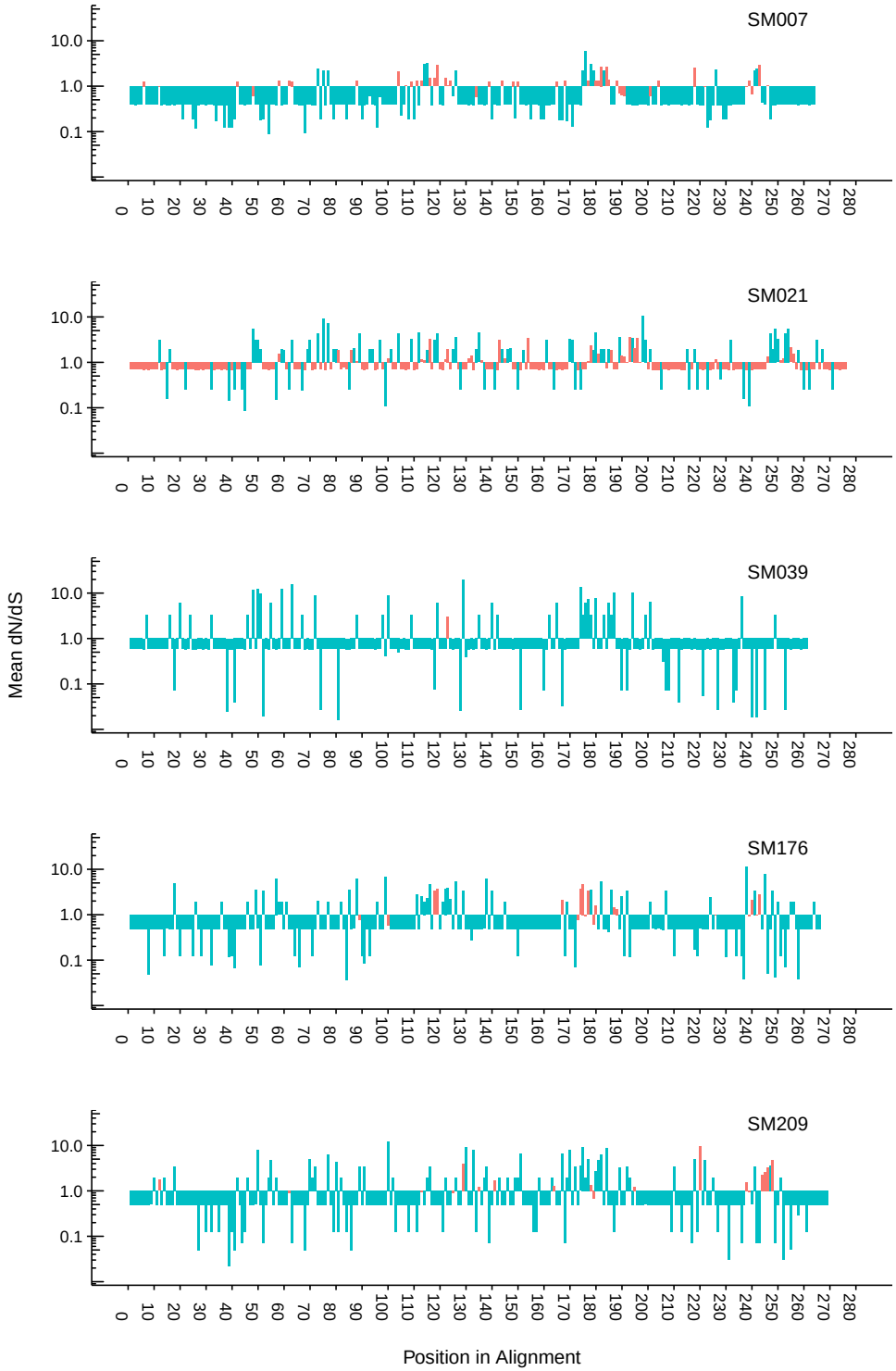


Figure 3.7: Mean codon-specific dN/dS estimates in the gp120 amplicon for each patient - part 1. An estimate greater than 1 indicates positive selection whilst an estimate of less than 1 indicates purifying selection. Sites under significant selection are shown in blue whilst sites lacking significance are shown in red and are assumed neutral.

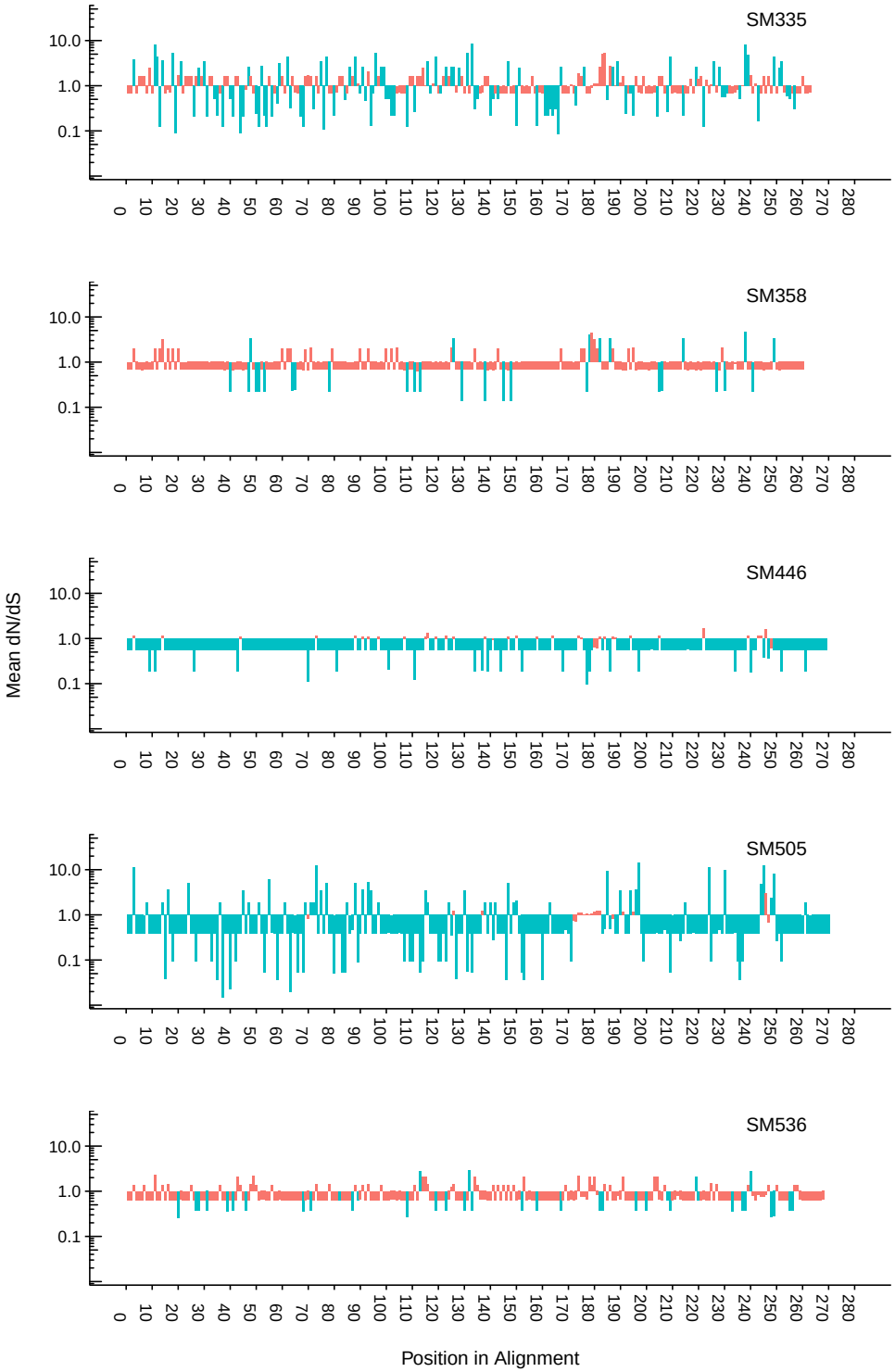


Figure 3.8: Mean codon-specific dN/dS estimates in the gp120 amplicon for each patient - part 2. An estimate greater than 1 indicates positive selection whilst an estimate of less than 1 indicates purifying selection. Sites under significant selection are shown in blue whilst sites lacking significance are shown in red and are assumed neutral.

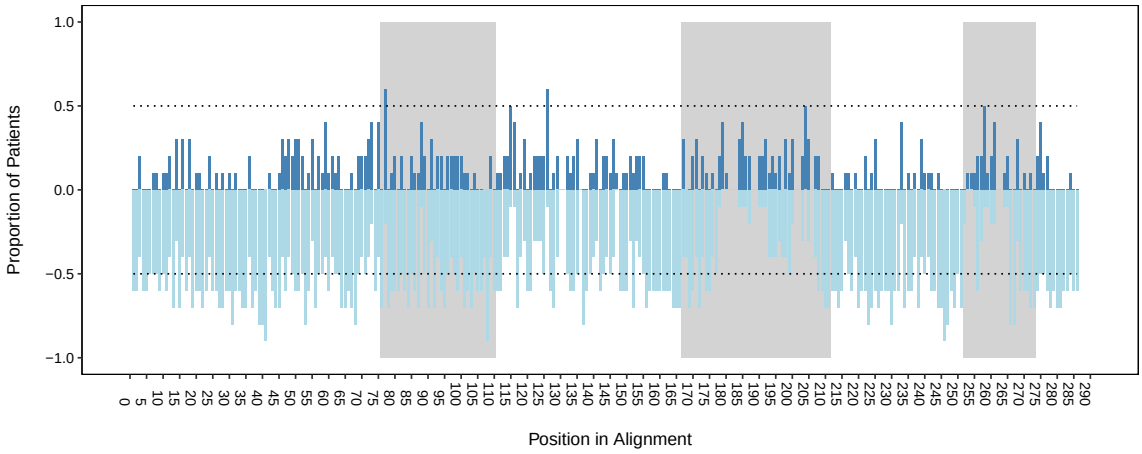


Figure 3.9: The proportion (absolute) of patients showing evidence of significant positive or purifying selection across each codon of the gp120 amplicon. Sites under significant positive purification are shown in dark blue, whilst those under significant purifying selection are shown in light blue. Variable loops V3, V4 and V5 are shaded grey. The dashed line indicates 50% of patients.

Two sites in the alignment were under significant positive selection in the majority (60%) of patients (corresponding to T297 and S348 in HXB2 gp120) whilst three further sites were selected in half of the patient sampled (corresponding to A337, D415 and S468). Within the variable loops, substantial purifying selection could be seen in V3, consistent with some degree of conservation owing to its defined role in co-receptor usage. Conversely, both V4 and V5 showed a dearth of purifying selection indicating that these loops may be more rapidly evolving.

To put this selection pressure into a biological context, homology modelling was implemented through SWISS-MODEL [392–395] to map the SM cohort gp120 consensus sequence to a cryo-electron microscopy (cryo-EM) crystal structure of a glycosylated HIV-1 envelope trimer (RCSB PDB 3J5M). The structure was then annotated in PyMOL v.1.7.4, marking on the variable loops and the five specific amino acids identified as being under significant positive selection (figure 3.10).

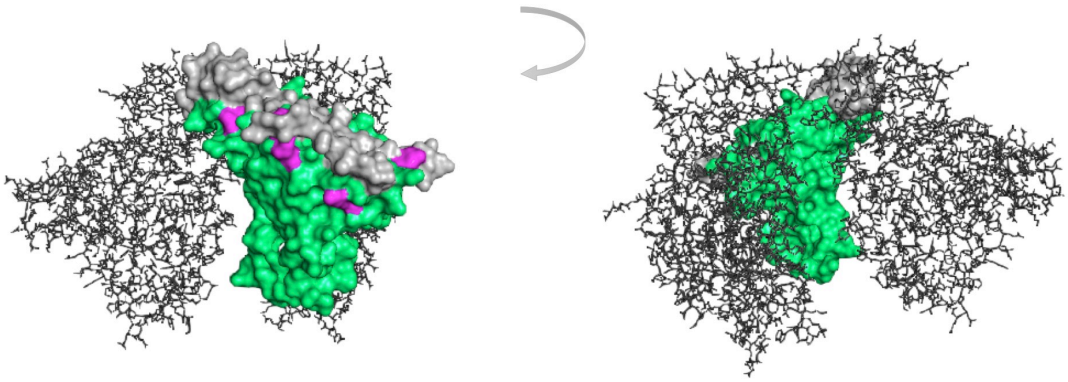


Figure 3.10: Homology modelled structure of the SM cohort consensus gp120 sequence in surface representation, with variable loops V3, V4, and V5 in grey and sites under significant positive selection in magenta. Structure has been modelled on a glycosylated HIV-1 gp120 trimer (RCSB PDB 3J5M) [427]. For clarity, two molecules in the trimer are shown in line representation in grey.

All five of the sites under significant positive selection in 50% or more of the patients clustered in or around the variable loops on the outermost exposed edge of the protein.

3.2.3.3 Characterisation and Mapping of Sequence Variation

Having estimated evolutionary rates and selection pressure, the next aim was to establish where and in what way mutation was occurring in the amplicon. Using an alignment of 575 translated sequences[1], variants were characterised as sites deviating (substitutions or deletions) from the inferred founder sequence (section 3.2.2.4) across each position. Owing to length discrepancies and extensive sequence variation, the hypervariable loops were conservatively stripped from this alignment (final length 179aa).

[1]Thirteen sequences from patient SM358 were removed from this study due to an abnormally large deletion rendering them directly incomparable.

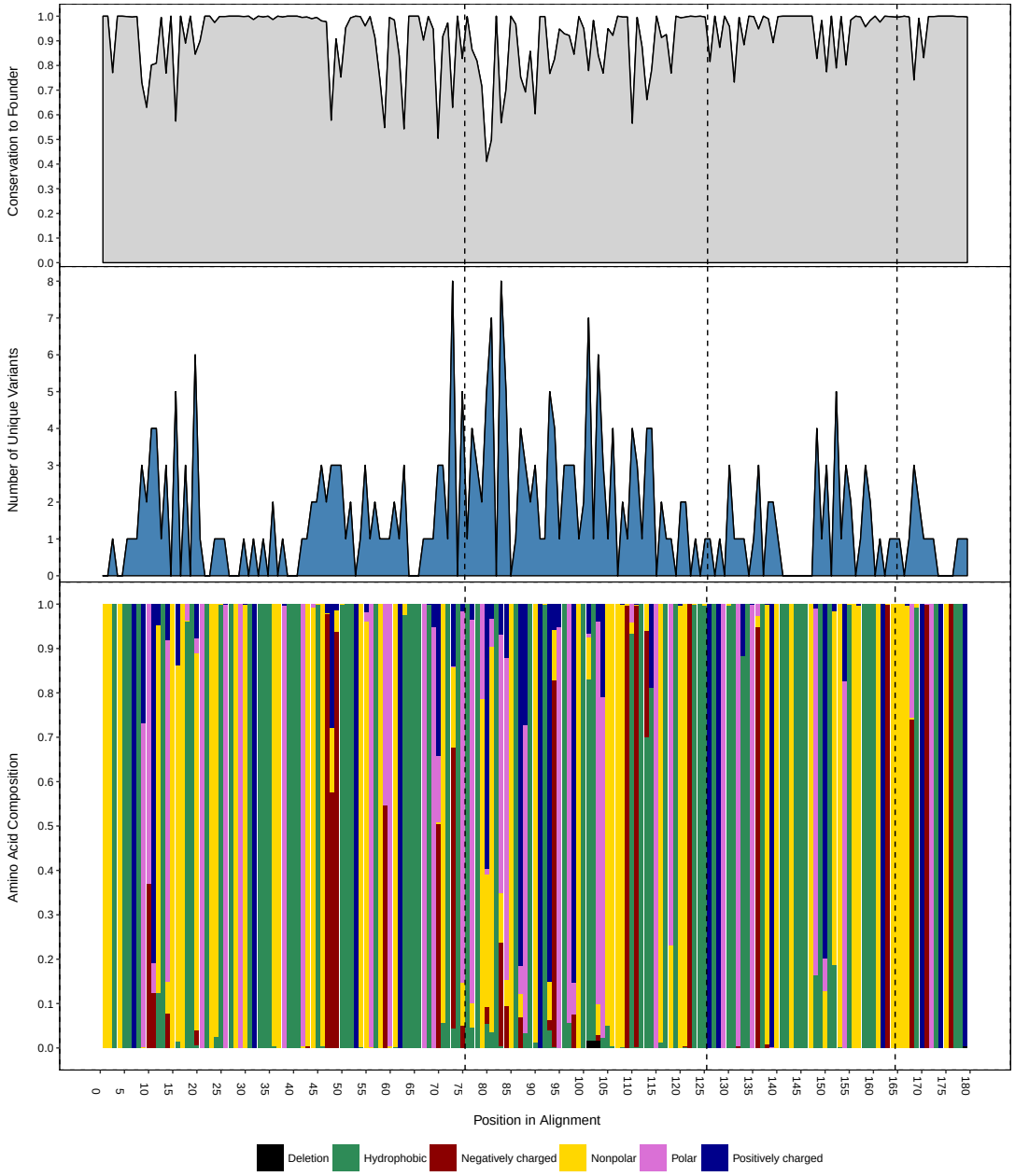


Figure 3.11: Characterisation of sequence variants. Upper panel) Conservation of each position in the alignment to that of the inferred founder sequence. **Middle panel)** Number of unique variants found in each position along the alignment. **Lower panel)** Amino acid composition across each position, showing the proportion of amino acids with a particular physicochemical property, coloured according to the Lesk format [428]. The vertical dashed lines denote the positions of V3, V4 & V5 which were conservatively stripped for this analysis.

In total, 288 unique variants were identified across 128 variable sites, whilst the remaining 51 positions were completely invariant (28.5%): substantially lower than the 77% reported across the full-length genome [429]. Plotting the degree of conservation across the alignment (figure 3.11 upper panel) revealed that a greater proportion of sequences deviated from the inferred founder towards the centre of the amplicon, particularly flanking the V3 loop (first dashed line). Counting the number of different amino acids that occur in each position (figure 3.11 middle panel) also revealed that this region showed a high degree of amino acid diversity.

Characterising the amino acids found at each position in terms of their physicochemical properties (figure 3.11 lower panel) revealed this region showed a marked degree of molecular heterogeneity relative to the beginings and ends of the amplicon where particular properties were more conserved.

Of the variants identified, many were present in a single or small collection of sequences and are therefore of potentially limited biological relevance. "Major variants" were therefore resolved, defined as mismatches from the inferred founder sequence present in at least 15% of the total sequences. Whilst major variants are often defined as those present at a frequency of greater than 5% - although no recognised definition is agreed upon - this value was conservatively tripled as the amplicon was approximately three-times more variable than the full-length genome. To prevent over-representation by particular patients where more time-points were successfully sampled, the alignment was first stripped to one time-point per patient, with time-points selected to ensure an approximately even distribution of samples collected each year (2010: three; 2011: three; 2012: four; 2014: two samples; final alignment of 248 sequences). 200 unique variants were identified in the stripped alignment (figure 3.12), with 31 of these meeting the definition of "major" variants.

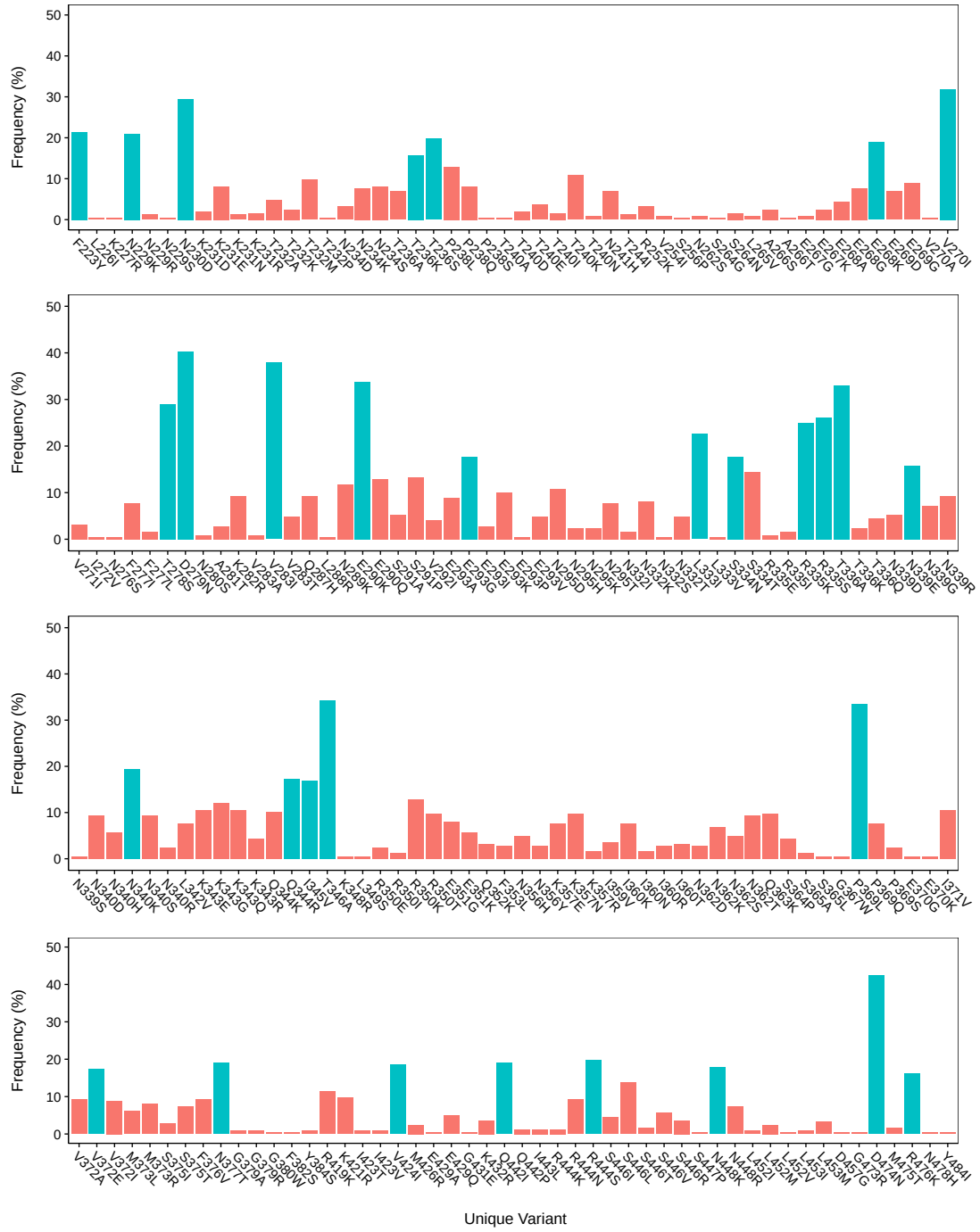


Figure 3.12: Unique variants identified in the SM cohort gp120 envelope amplicon. Major variants are shown in blue whilst minor variants are shown in red. Position is relative to HXB2 gp120.

The frequency of these variants in the SM cohort were next compared with similar FPD cohorts in Henan. The frequency of each major variant was calculated in an alignment of the translated FPD sequences (accession numbers: appendix table A.8). Coverage was between 88 and 474 sequences at each position. Of the 31 major variants identified in the SM cohort, 14 were also major variants within the FPD sequences from other cohorts (figure 3.13). The remaining 17 were major variants in the SM cohort alone.

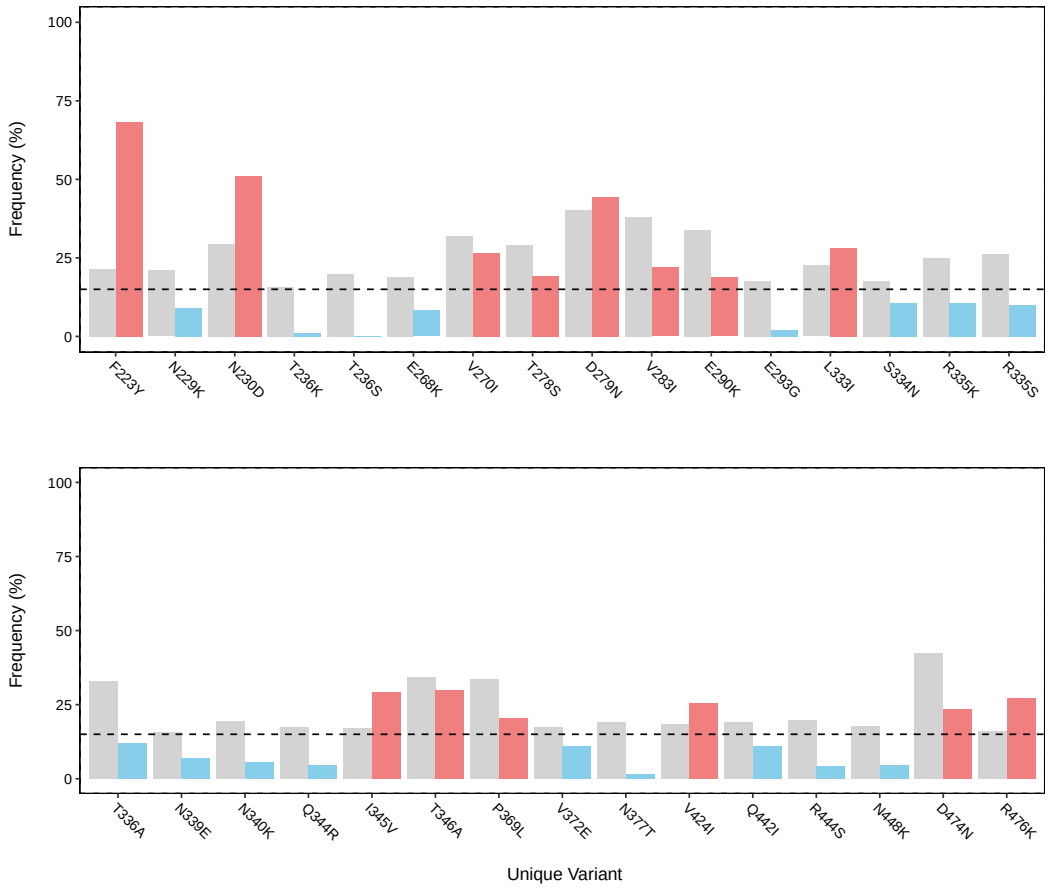


Figure 3.13: Frequency of major variants identified in the SM cohort in both the SM cohort (grey) and similar Chinese FPD cohorts. Variants that are also found at greater than 15% in the non-SM cohort sequences are shown in red whilst those found at less than 15% are in blue. Position is relative to HXB2 gp120.

An interpatient comparison was next performed to check for the presence of patient-specific variant "fingerprints" (figure 3.14). In patients with serially sampled sequences, some major variants were present in all time-points. Conversely, other major variants are not as temporally conserved within patients and appear or disappear over time. Comparing the variants classed as major in the SM cohort alone with those also classed as major variants in the non-SM cohort sequences revealed no apparent difference in their distribution.

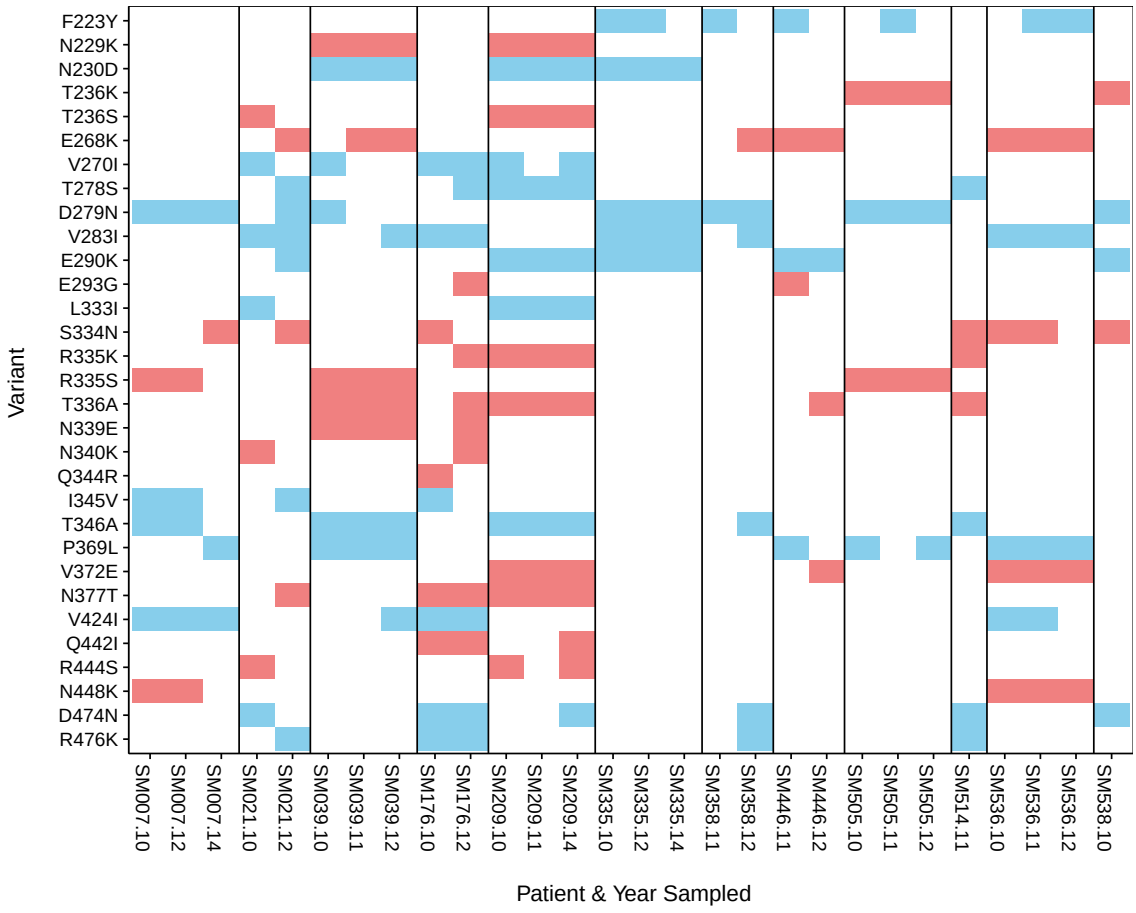


Figure 3.14: Presence of major variants in individual patients of the SM cohort at each time-point sampled. Variants that are major only in the SM cohort are shown in blue, whilst those that are also major variants in similar Chinese FPD sequences from the LANL database are shown in red.

Finally, all major variants identified in this study as well as completely invariant sites were mapped to the homology modelled structure of the SM cohort consensus gp120 sequence described previously (figure 3.15).

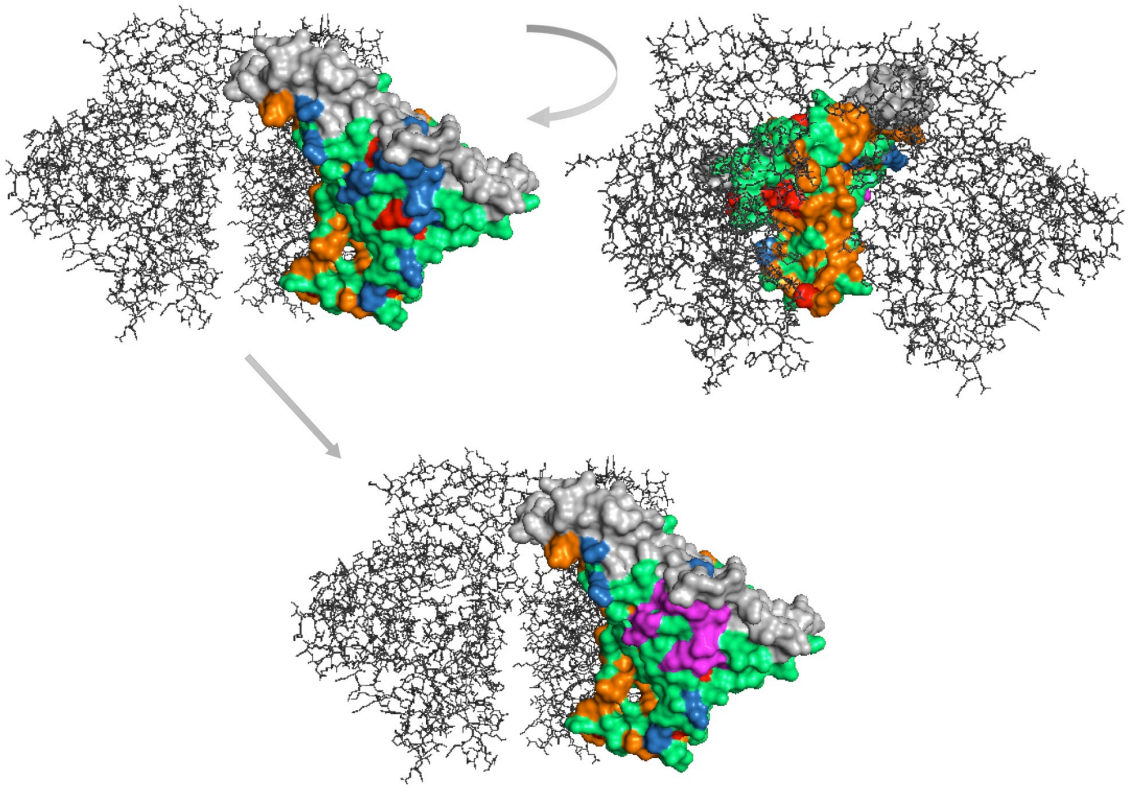


Figure 3.15: Homology modelled structure of the SM cohort consensus gp120 sequence in surface representation, with variable loops V3, V4, and V5 in grey and invariant sites in orange. Major variants are marked either in red if they are present at $>$ 15% in the SM cohort and LANL sequences, or blue if in the SM cohort only. Structure has been modelled on a glycosylated HIV-1 gp120 trimer (RCSB PDB 3J5M) [427]. For clarity, the other two molecules in the trimer are shown in line representation in grey. Below, a continuous cluster of major variants has been marked in magenta, which is composed of residues found at intermittent positions.

The overwhelming majority of the invariant sites (orange) identified in this study were located on the inner face of the structure, shielded by the other two molecules in the gp120 trimer (figure 3.15). Conversely, major variants (red and blue) were distributed around the outside of the protein. Indeed, several of these sites clustered together on

the three-dimensional structure, despite being located far apart in the primary amino acid sequence. For example, a large continuous cluster of major variants on the outer surface is composed of amino acids corresponding to positions 290, 293, 333, 334, 336, 339, 340, 344, and 448 in HXB2 gp120.

3.3 Discussion

The surface-exposed HIV envelope protein gp120 is a major target of the adaptive immune response, against which the virus must develop effective countermeasures. This chapter describes an investigation into the role of such immune evasion processes against the gp120-targeted response. Specifically, the aims of this study were to characterise gp120 in a unique narrow-source cohort of FPDs from Henan province in China, and to investigate how and in what way selection pressure drives mutation within this gene fragment.

By first mapping the density of gp120 epitopes reported in the literature, it was shown that antibody epitopes in particular were mostly clustered within the V3 hypervariable loop. This may be due to a genuine peak in immune activity in this area, a bias in reporting due to the extensive study of V3 for the purpose of vaccine design, or a combination of the two factors.

The immunogenicity of V3 is somewhat controversial. Whilst numerous antibodies against V3 have been described, the cross-neutralisation potential of these antibodies is generally low, reviewed by Hartley *et al.* [430]. Glycosylation, sequence variation, masking by V1V2, and the specific amino acid make-up of the loop may contribute to this, reviewed by Pantophlet *et al.* [431]. However, some monoclonal and polyclonal antibodies specific to epitopes within V3 have been demonstrated to neutralise diverse HIV strains *in vitro* [432–435], and of the five variable loops in gp120, V3 is the most conserved with amino acid variation restricted to approximately 20% of the loop's residues [436]. Additionally, the length of V3 is relatively conserved owing to functional constraints as V3 plays an important role in co-receptor binding [418, 419] and deletion abrogates viral infectivity [437]. Due to the large number of epitopes

described in this region and evidence for functional immunological activity, V3 was selected as the focus point for the study.

Having amplified from twelve patients with longitudinally-collected samples a fragment of gp120 encompassing the C2-V5 region, the clonal sequences generated were used for comprehensive characterisation of the SM cohort. Viral subtyping confirmed that all patients sampled were infected with a virus that clustered strongly with a circulating recombinant from Thailand, consistent with cohort's reported history of infection with Thai B virus (also referred to as B') [414]. The cohort's history was further substantiated by Bayesian inference of the time to convergence placing the date of infection within a window that encompasses the time-frame during which patients were active blood donors. Furthermore, a SkyGrid analysis demonstrated limited spread within the population consistent with clinical observations in this cohort.

Of note, the inferred founder sequence for the SM cohort was strikingly similar to the consensus sequence of a series of FPD viruses from nearby cities obtained from the LANL HIV Sequencing Database, consistent with a shared or related source of virus in other FPD cohorts. The similarity of these sequences has severe sociopolitical implications, as it highlights the homogeneity of infection and absence of protective countermeasures which could have halted the spread of infection before it swept through the population as it did.

In addition to simply characterising the SM cohort, the sequences obtained in this study were used to provide insight into the contribution of mutation to immune evasion. The intrapatient viral evolutionary rates were determined, but were not found to correlate with any clinical parameter. Next, codon partitioning permitted the

relative evolutionary rate ratio of codon positions 1+2 to position 3 to be calculated, which demonstrated that on the whole, gp120 was under weak purifying selection pressure in these patients.

Investigating selection pressure to higher resolution through the calculation of site-specific dN/dS ratios revealed that - in many patients - particular sites were under significant positive selection. Aligning and summarising these sites on a interpatient level showed that five positions were under significant positive selection in the majority of individuals. Mapping these sites to a three-dimensional homology-modelled structure revealed that all five sites are located on the outermost distal edge of the gp120 protein, either flanking or contained within the hypervariable loops.

This finding is arguably more likely to be due to antibody-mediated than CTL-mediated selection pressure. Traditionally, CTL epitopes in HIV have been described as clustering in hydrophobic regions [438] and are heavily influenced by flanking hydrophobic residues [439], making it unlikely that they would nucleate in surface-exposed regions of the protein. More recent evidence suggests that CTL epitope distribution in HIV may in fact be random with apparent clustering an artefact of study design [440]. Under this assumption, it is even less likely that such clear congregation would occur were the selection detected the result of CTL- rather than antibody-mediated pressure.

This is the first (and at present, only) instance whereby sites identified through renaissance counting as being under significant selection pressure have been mapped to a three-dimensional protein structure. Mapping the sites in this way has not only provided sound reasoning for the hypothesis that selection in these sites is

mediated by antibodies; but has also provided independent testing and verification of the renaissance counting framework.

Through the characterisation of sequence variants within gp120, it was also shown that the regions flanking the V3 loop are particularly variable. The amino acids in these regions are less conserved and have substantially greater physicochemical heterogeneity than the remainder of the amplicon. This is of note as V3 is the most conserved of the five hypervariable loops, perhaps suggesting that the additional flanking heterogeneity is a compensatory measure for the purifying selection (also shown in figure 3.9) associated with V3 and its role in co-receptor usage.

Plotting the presence of each major variant by patient revealed that some variants were present in all time-points sampled. This might suggest that the substitution was made early on and has been retained. This is more consistent with CTL-mediated escape mutations than antibody-mediated pressure. Supporting this, patient SM007 - for example - maintains both the D279N and V424I variants throughout the course of the study, and both of these substitutions fall within regions that contain known CTL epitopes corresponding to the HLA-I types of the patient (HLA-A02, HLA-C06). However, there is not a clear relationship between the location of major variants and known epitopes corresponding to the HLA-I types of the patients on the whole. This is likely because gp120 epitopes have been less comprehensively mapped than, for example, Gag epitopes.

Whilst these retained major variants are suggestive of CTL escape signatures, some variants - for example many of those detected in patient SM176 - are not found in both time-points sampled, indicating that they may have arisen in response to fluctuating waves of antibody pressure. Such fingerprints may provide insight into whether

a particular mutation is the result of either antibody- or CTL-mediated selection pressure, but this novel approach requires substantial further verification in additional cohorts before firm conclusions can be drawn. For example, it is possible that some major variants may have arisen as a result of random drift early on and have been retained due to a lack of purifying selection in that residue. Whilst random drift may seem unlikely to contribute on a large-scale to a virus with such a compact genome, further analysis will be required to establish the false discovery rate associated with using such a technique.

By mapping major variants to the homology modelled structure, it was shown that many appear in close proximity to each other in the tertiary structure despite appearing at disjointed positions along the primary amino acid sequence. This clustering again may suggest that this region of the protein is under substantial pressure to mutate in response to antibodies. This seems parsimonious as the clustering occurs primarily on the outer face of the protein, whilst the inner face is almost wholly invariant.

The use of the SM cohort helps to overcome many of the potential confounding factors associated with previous studies wherein patients were infected through disparate routes with divergent virus. However, it is not without its own challenges, and one key limitation of this study is that the SM cohort is by its very nature not particularly representative. The next step should therefore be to try to apply some of the findings and techniques developed in this study to other applicable cohorts in order to further investigate selection pressure and immune evasion in gp120.

To conclude, this study has further characterised a unique narrow-source HIV-1 cohort, and has provided evidence that adaptive immunological pressure plays an

important role in driving sequence mutation in the gp120 *env* gene. Specific sites under significant selection pressure have been mapped, which may have important translational implications by demonstrating where and how the surface of the virus evolves in response to the host immune system.

This chapter has described the role of gp120 envelope mutation in the evasion of the adaptive immune response, of which CTLs comprise an important faction. The next two chapters will further investigate how the virus counters CTL responses through mediating the downregulation of class-I human leucocyte antigen (HLA-I) via the retroviral accessory protein Nef.

Chapter 4

Nef-Mediated Downregulation of Class-I Human Leucocyte Antigen in HIV-1 and HIV-2 Infection

4.1 Introduction

The retroviral accessory protein Nef is highly multifunctional and contributes to immune evasion in several ways, described in section 1.1.8.7. This multifunctionality is due largely to the ability of Nef to behave as an adaptor protein, interacting with several cellular factors through a series of highly conserved motifs - reviewed by Arhel & Kirchhoff [441]. Of particular interest is the ability to mediate downregulation of class-I HLA molecules in virally infected cells [442, 443]. Nef selectively reduces the amount of HLA-A and HLA-B expressed on the cell surface, whilst leaving HLA-C and HLA-E unaffected. Downregulation of HLA-A and HLA-B facilitates escape from immune surveillance by preventing recognition and killing by CTLs and NK cells [444], whilst retention of HLA-C and HLA-E prevents lysis by NK cells by acting as an inhibitory "self" signal [445]. Emergence of this selective downregulation of HLA-A and HLA-B preceded the zoonotic transmission of SIV_{cpz} from chimpanzees to humans, and is conserved amongst a range of primate lentiviral lineages [446].

The ability of patient Nef isolates to downregulate HLA-I *in vitro* has been found to correlate with resistance to viral suppression by CTLs [447], and the extent of downregulation mediated by transmitted founder variants does not differ significantly from clinically latent variants, indicating that downregulation may be important throughout the course of infection [448]. However, functional activity may be reduced on progression to AIDS [449]. Whilst most functional studies have been performed *in vitro*, evidence has been put forward for a role of Nef-mediated HLA-I downregulation *in vivo*. Studies in the macaque have shown that HLA-I downregulation provides a selective advantage for SIV replication [450] and that mutations in SIV Nef eliminating only HLA-I downregulation revert to restore full function within only four weeks of infection, highlighting the importance of this mechanism to viral fitness.

4.1.1 Mechanism of Downregulation

The process of Nef-mediated downregulation of HLA-I has been extensively studied, but remains controversial and conflicting models of the mechanism exist. Early studies proposed a mechanism in which Nef induced endocytosis of surface-expressed HLA-I [442, 451–453]. Here, retrograde transport of surface-expressed HLA-I is initiated through binding of Nef to the acidic cluster sorting protein PACS-2, which localises to the trans-Golgi network (TGN) and complexes with the src family kinase (SFK) Hck. Hck activates zeta-chain-associated protein kinase 70 (ZAP-70) which itself activates phosphoinositide 3-kinase (PI3K), leading to accumulation of phosphatidylinositol-(3,4,5)-trisphosphate (PIP3) at the plasma membrane (PM). PIP3 recruits the guanine nucleotide exchange factor ARF nucleotide-binding site opener (ARNO), which activates ADP-ribosylation factor 6 (Arf6), triggering endocytosis of HLA-I. Subsequently, Nef recruits the clathrin adaptor protein AP-1 to the endosome, and redirects it to the TGN. This process is reviewed in detail by Wonderlich *et al.* [454].

Later studies began to favour a mechanism by which Nef disrupts trafficking of nascent HLA-I in the secretory pathway [455–458]. Indeed, many viral immune escape strategies that target antigen presentation do so at steps preceding surface presentation of HLA-I [459]. Nascent HLA-I molecules are available in the ER for loading of viral antigens, whilst molecules on the surface of the cell are likely to be pre-bound to cellular antigens, and as such may pose the virus less of an immediate threat in newly-infected cells [454].

Disruption of HLA-I trafficking is dependent on AP-1. AP-1 is first activated at the TGN by the small guanosine triphosphatase (GTPase) ADP-ribosylation factor 1 (Arf1). In the presence of cargo such as HLA-I or tetherin, Nef triggers the trimerisation of activated AP-1:Arf1 complexes, to which further AP-1:Arf1 dimers are able to dock, forming a hexagonal ring that structurally resembles clathrin in both size and symmetry. Formation of the hexagonal ring templates the clathrin cage and triggers vesicle formation [460], which may then be delivered to the lysosome for degradation.

A simple schematic representation of the proposed mechanisms of Nef-mediated HLA-I downregulation is shown in figure 4.1.

It has also been suggested that Nef targets both existing and newly synthesised HLA-I by temporally switching from the retrograde transport of surface expressed molecules ("signalling" mode) early in infection, to disruption of trafficking of nascent molecules ("stoichiometric" mode) at around 48 hours post-infection [461]. This model resolves a number of seemingly contradictory *in vitro* observations, and may explain some of the inconsistencies observed between cell lines/types which might preferentially utilise different pathways. However, its relevance *in vivo* is currently not well defined.

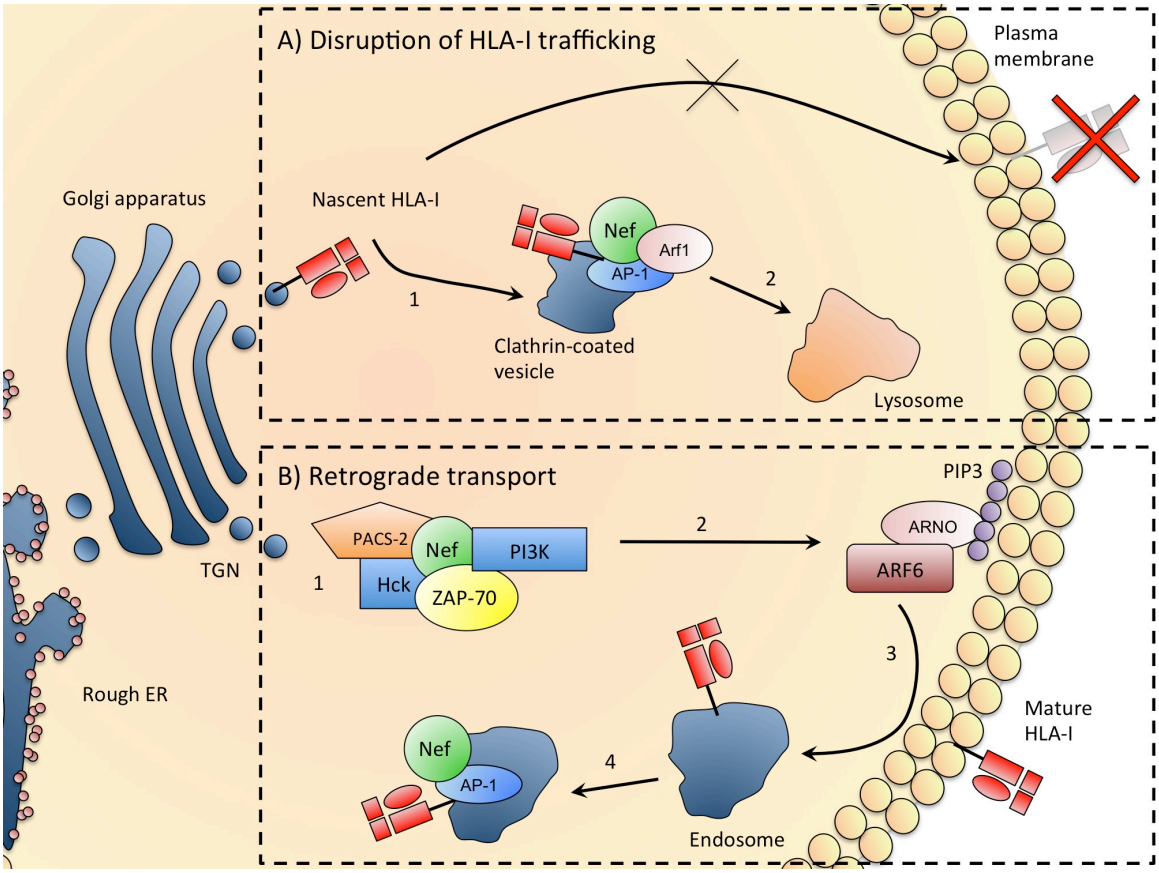


Figure 4.1: Proposed mechanisms of Nef-mediated HLA-downregulation. Nef-mediated downregulation of HLA-I may occur through **A)** disruption of intracellular trafficking to prevent nascent HLA-I molecules from reaching the plasma membrane; or **B)** retrograde transport of surface expressed HLA-I molecules through ARF6-dependent endocytosis. Temporal switching between the two mechanisms may occur.

4.1.2 Functional Motifs Associated with Downregulation

The crystal structure of Nef in direct complex with the cytoplasmic domain of HLA-I and AP-1 was published in 2012 [462]. Complex formation is dependent on the μ1 chain of AP-1, which interacts with the tyrosine residues of integral membrane proteins such as HLA-I [463]. Several intact Nef functional motifs[1] are required for correct complex formation and subsequent downregulation.

[1] All such motifs and residues in this chapter will be identified by their relative amino acid positions in HXB2 Nef.

The myristoylation site at position 2 permits Nef to bind the inner leaflet of the plasma membrane [464] and is important for the membrane localisation and subcellular targeting associated with many of its functions. As such, it is under strong purifying selection. The N-terminal α-helix of Nef [465] at positions 17-26 is also important for downregulation of HLA-I [466]. In particular, an arginine-rich region of the helix at positions 17-19 mediates binding to βCOP - a component of coat protein I (COP-I) - which coats and targets the molecules for degradation [467]. Also within the N-terminal α-helix is a methionine residue at position 20, which plays a (currently poorly-defined) role in mediating HLA-I downregulation but does not interact with the HLA-I cytoplasmic tail directly [468].

The acidic cluster found at positions 62-65 is a highly conserved motif which interacts with phosphofurin acidic cluster sorting (PACS) proteins 1 and 2 [452]. Tryptophan at position 113 is also required for this interaction [469]. PACS1 and PACS2 are putative sorting proteins which are believed to play a role in the transport of proteins between intracellular membranes: for example, endosome to TGN trafficking [470]. PACS-2 in particular has been implicated in the retrograde transport route of HLA-I downregulation [471].

At positions 69-81 there is a proline-rich repeat with a known role in HLA-I downregulation [472]. This region contains a tetra-proline motif interspersed with less-conserved residues ($_{69}$PxxPxxPxxP$_{78}$). The final proline found in position 78 is crucial for HLA-I downregulation mediated by HIV-1 Nef [473]. The proline-rich repeat region forms an SH3-binding motif which facilitates formation of a kinase complex and initiation of the retrograde transport pathway of downregulation.

Aspartic acid at position 123 has long been considered an important residue for oligomerisation of Nef [474], which is required for numerous functions. However, disruption of D123 has been shown not to interfere with retrograde transport *in vitro* [475], and its role in homodimer formation has since been called into some doubt. Residues R105, R106 and C206 have since been implicated instead [476].

Whilst SIV Nef proteins are also able to downregulate MHC-I molecules, the mechanism through which this occurs differs, and the functional motifs associated are not necessarily the same [477].

4.1.3 Differential Downregulation of HLA-A and HLA-B

It has recently been demonstrated that HLA-B may be more resistant to downregulation mediated by Nef than HLA-A [399]. Several HLA-B types are strongly associated with better protection against HIV-1 and slower progression to AIDS [478], and of these, HLA-B27 and HLA-B57 are considered strongly protective [317, 479]. The increased resistance of the HLA-B cytoplasmic domains to downregulation may directly contribute to the substantially improved control of HIV infections by some HLA-B-restricted CTLs [399]. However, the study describing this phenomenon was performed with a single Nef isolate derived from the laboratory-adapted viral strain SF2. It therefore remains unclear whether this observation of selective differential downregulation is generalisable to Nef isolated from primary patient isolates. Furthermore, whilst HIV-2 Nef also downregulates HLA-I [480], it has not yet been established whether this protein also mediates differential downregulation of HLA-A and HLA-B. The purpose of this project was therefore to fill these gaps in knowledge.

4.1.4 Aims

The specific aims of this project were to:

- Explore the differential downregulation of HLA-A and HLA-B molecules mediated by Nef isolated from primary patient samples, and compare these data with those from SF2 Nef and consensus HIV-1 & HIV-2 Nef;
- Explore whether the extent of downregulation of HLA-A and HLA-B molecules correlates with clinical parameters and disease progression in HIV-1 and HIV-2 infected patients.

4.2 Results

4.2.1 Cohort Characterisation

4.2.1.1 HLA-I Characterisation of SM & TD Cohorts

The SM (HIV-1) cohort comprises Han Chinese patients from Henan province in China; whilst the TD (HIV-2) cohort patients from Caió in Guinea-Bissau belong primarily to the Manjako tribe. The frequency of individual HLA types vary substantially by geography and ethnicity, and so to permit stratified sampling, the class I HLA types of patients within each cohort were characterised.

The HLA types of 343 patients from the SM cohort were resolved to two digits by Mr Tim Rostron and Dr Louis-Marie Yindom. From these data, the frequency of each type was calculated. 14 HLA-A, 27 HLA-B, and 11 HLA-C types were present in the SM cohort (table 4.1). HLA-A02 was the most frequent HLA-A type; HLA-B40 was the most frequent HLA-B type; and HLA-C03 was the most frequent HLA-C type.

The number of individuals carrying each HLA type in the TD cohort were taken from data published previously (table 4.2) [481]. 8 HLA-A, 11 HLA-B, and 9 HLA-C types were recorded. The cohort exhibited a lower degree of HLA-I diversity than the SM cohort, particularly in HLA-B. HLA-A33 was the most frequent HLA-A type recorded; HLA-B15 was the most frequent HLA-B type; and HLA-C03 was the most frequent HLA-C type.

From these data, a stratified sample of ten representative patients was selected from each cohort, as described in section 2.3.2.

Table 4.1 Frequency of HLA types in the HIV-1 SM cohort.

Type	Freq. (%)
HLA-A02	30.32
HLA-A24	14.74
HLA-A11	13.57
HLA-A33	10.89
HLA-A30	10.22
HLA-A01	5.70
HLA-A03	5.19
HLA-A31	2.85
HLA-A32	2.68
HLA-A26	1.68
HLA-A29	1.51
HLA-A68	0.34
HLA-A23	0.17
HLA-A74	0.17

Type	Freq. (%)
HLA-B40	15.42
HLA-B13	11.53
HLA-B51	8.81
HLA-B15	8.64
HLA-B44	6.95
HLA-B46	6.10
HLA-B58	5.42
HLA-B35	4.41
HLA-B52	4.24
HLA-B07	4.07
HLA-B27	3.39
HLA-B57	3.22
HLA-B48	2.88
HLA-B38	2.71
HLA-B54	2.54
HLA-B39	2.20
HLA-B08	1.69
HLA-B37	1.69
HLA-B55	1.02
HLA-B50	0.68
HLA-B18	0.51
HLA-B49	0.51
HLA-B67	0.51
HLA-B14	0.34
HLA-B05	0.17
HLA-B47	0.17
HLA-B56	0.17

Type	Freq. (%)
HLA-C03	19.10
HLA-C07	16.84
HLA-C06	15.10
HLA-C08	11.63
HLA-C01	10.42
HLA-C14	6.42
HLA-C12	6.08
HLA-C04	5.56
HLA-C15	5.56
HLA-C02	2.08
HLA-C05	1.22

Table 4.2 Frequency of HLA types in the HIV-2 TD cohort.

Type	Freq. (%)
HLA-A33	23.22
HLA-A23	19.19
HLA-A02	13.42
HLA-A30	11.81
HLA-A68	9.26
HLA-A26	8.72
HLA-A01	8.05
HLA-A74	6.31

Type	Freq. (%)
HLA-B15	25.65
HLA-B53	12.58
HLA-B58	12.10
HLA-B08	11.61
HLA-B35	11.13
HLA-B49	8.23
HLA-B14	5.32
HLA-B07	3.71
HLA-B13	3.39
HLA-B44	3.23
HLA-B18	3.06

Type	Freq. (%)
HLA-C03	25.24
HLA-C07	20.50
HLA-C04	17.57
HLA-C02	11.16
HLA-C08	8.65
HLA-C06	5.30
HLA-C01	4.46
HLA-C16	3.77
HLA-C05	3.35

4.2.1.2 Characterisation of the Functional Motifs of Nef Isolates

The full-length *nef* gene was sequenced from the proviral DNA of each patient and translated. Several functional motifs in Nef are known to be important in HLA-I downregulation (section 4.1.2), and so these were aligned, compared, and are presented in table 4.3.

All Nef isolates maintained an intact myristoylation site (G2), with limited variation existing in the C-terminal flanking region. One isolate from HIV-1 patient SM528 exhibited a G3N substitution, and all isolates from the HIV-2 patients exhibited a G3A substitution.

The N-terminal α-helix (15-28) was relatively conserved in most isolates. The HIV-1 consensus N-terminal α-helix was identical to that present in HXB2 Nef, and all bar one patient isolates contained between zero and three substitutions in this region. M20 was conserved in all HIV-1 isolates except for SM547 which contained an M20I substitution. The N-terminal α-helix of this patient was broadly different to the other SM cohort patients. Typing the full-length SM547 Nef sequence with the REGA HIV-

1 Subtyping Tool (v.3.0) [482, 483] revealed this virus belonged to subtype C, rather than B' which is characteristic of the SM cohort as a whole. In all HIV-2 patients, Methionine at position 20 was replaced with leucine.

The acidic cluster (60-67) was reasonably conserved amongst the HIV-1 isolates. Six of the ten patient isolates contained a run of four glutamic acid residues. In the HIV-2 isolates, glutamic acid was replaced with aspartic acid. Eight of the ten isolates had an intact tetra-aspartic acid motif, whilst two contained a D64N mutation. In both HIV-1 and HIV-2, the flanking regions of this domain were almost completely conserved within each virus.

The tetra-proline motif (67-83) was invariant across all HIV-1 isolates, and very limited variation was present in the interspersed residues. The proline-rich repeat region was more variable in HIV-2, with all isolates containing a substitution in at least one of the proline residues. Despite P78 being the most critical of the proline residues in HIV-1 [473], in HIV-2 P78 was universally substituted with either alanine, threonine or glutamic acid. The only proline residue that remained unchanged across all HIV-2 isolates was P72.

The PACS-binding motif (112-114) was wholly conserved within HIV-1, but HIV-2 contained tyrosine in all isolates. Limited variation was present in the flanking regions for both viruses. However, due to the variation between HIV-1 and HIV-2, the strength of the alignment in this region is poor. Of the remaining three (putative) oligomerisation domains, domains 2 & 3 (122-124; 206-207) were the most conserved in both HIV-1 and HIV-2. Domain 1 (104-107) was somewhat more variable, particularly in the flanking regions.

Table 4.3 Functional motifs and flanking regions of selected Nef isolates. Functional motifs are shown in red for HXB2. Consensus sequences constructed from the Nef isolates of each cohort are included above the respective sequences, and are labelled as HIV-1 and HIV-2 accordingly. Sites identical to the respective consensus are marked with a dot.

	Myristoylation Site	N-Terminal α-Helix	Acidic Cluster	Proline-Rich Repeat	Oligomerisation (putative 1)	PACS-Binding	Oligomerisation (putative 2)	Oligomerisation (putative 3)
	1-3	15-28	60-67	67 - 83	104-107	112-114	122-124	205-207
HXB2	MGG	TVRERMRRAEPAAD	AQE--EEEV-G	GFPVTPQVPLRPMTYKA	QRRQ	LWI	PDW	KNC
SF2	MGG	AIRERMRRAEPRAE	AQE--EEEV-G	GFPVRPQVPLRPMTYKA	QRRQ	LWI	PDW	KDC
HIV-1	MGG	RVRERMRRAEPAAD	AQE--EEEV-G	GFPVRPQVPLRPMTYKG	QKRQ	LWV	PDW	KDC
SM046	•••	••••••QHT•••••	••••••••••••	••••••••••••••••••	••••	•••	•••	•••
SM335	•••	S•••••Q•••••••	••K•••••••	••••••••••••••••••	••••	•••	•••	•••
SM339	•••	••••••••••••••	••••••••••••	••••K••••••••••A	••••	••I	•••	•••
SM434	•••	A••••••••••••••	••••••••••••	••••T••••••••••A	•R••	••I	•••	•N•
SM526	•••	S•••••K•••••••	•••••••G•••	••••K••••••••••••	••••	•••	•••	•••
SM528	••N	S•••••K•T••-••	••••••••••••	••••••••••••••••••	••••	•••	•••	•••
SM547	•••	EI•••I••TPA••E	••GE••GD•••	••••••••••••••••••	RR••	•••	•••	•••
SM550	•••	A••••••••••••••	••••••••••••	••••T••••••••••A	•R••	••I	•••	•N•
SM551	•••	••••••E•••••••	•••EN••••••	•••I••••••••••••••	•••R	•••	•••	•••
SM569	•••	••••••E•••••••	••••••••••••	••••••••••••••••A	••••	•••	•••	•••
HIV-2	MGA	GLQERLLRARAGXC	SDD--DDLV-G	GVPVTPRVPLRAMTYRL	ERRH	IYL	PDW	RGI
TD001	•••	•••••••QTSGET•	••••••••••••	••••••••Q•••••••	••••	•••	A••	•••
TD014	•••	••••••••••••T•	••••••••LT•	••A••••••••T••P••	••••	•••	A••	•••
TD027	•••	•••••••••CGET•	••••••••••••	•••••••TQ•••••••	••••	•••	•••	•••
TD037	•••	••••••••••••A•	•••••••MI••	•••••••T••••••H••	••••	V••	•••	•••
TD041	•••	••R•••••••••TY	•••••••QI••	••••••••••••••H••	•••R	V••	•••	•••
TD045	•••	••R•••••••••A•	••••••••••••	••S•••K••••••S•K•	••••	•••	•••	•••
TD052	•••	••••••••T•••A•	••••••N••••	•••••••A••••••••M	••••	•••	•••	•••
TD053	•••	E•H•••••••••A•	••••••N••••	•••••••A•••E•••••	A•••	•••	A••	•••
TD055	•••	••••••••••••T•	•••••••Q•••	••S•••••Q•••••••	•K••	•••	•••	•••
TD069	•••	••R••••Q••••A•	••••••••••••	•••••••TQ••E•••••	••••	•••	A••	•••

4.2.2 Experimental Optimisation

4.2.2.1 Baseline HLA-I Expression Levels

The selected *nef* isolates were cloned into the pIRES2-ZsGreen1 expression vector, and used to transfect a series of transgenic lymphoblastic cell lines stably expressing a chimeric HLA-I molecule composed of the extracellular domain of HLA-A2 and variable cytoplasmic tails that encompass all known sequence variants. The HLA types which share each cytoplasmic tail are listed in appendix table A.3, and the sequences of the tails are listed in appendix table A.4.

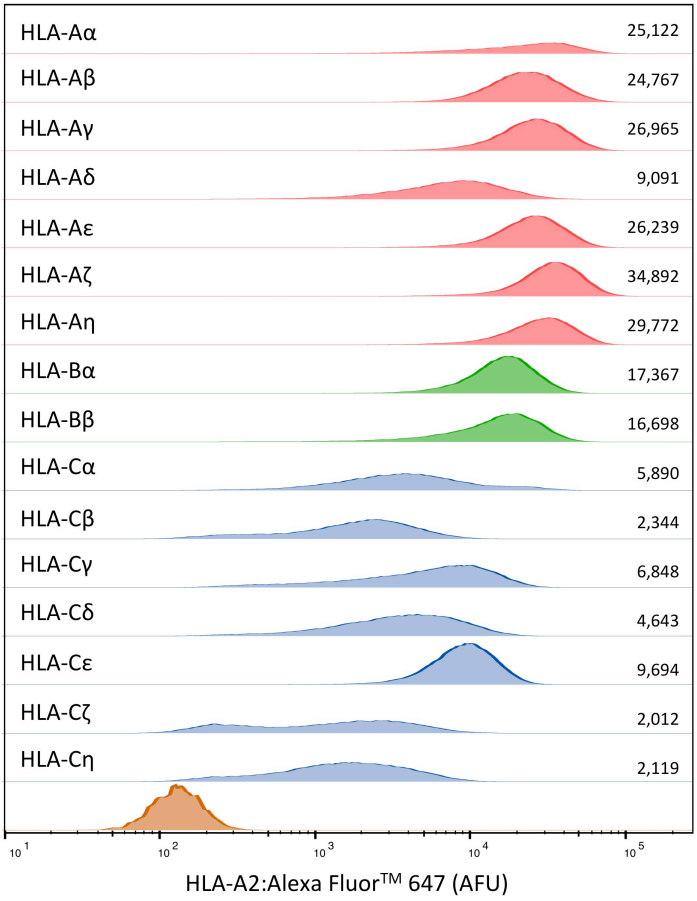


Figure 4.2: Baseline median fluorescent intensity of chimeric HLA molecules in transgenic JJK-derived cell lines. HLA-A in red, HLA-B in green, and HLA-C in blue. Wild-type JJK cells are shown in orange as a negative control. Values are in AFU to the right of each plot.

Prior to transfection, the baseline surface expression level of each chimeric HLA-I molecule was measured using flow cytometry. A large degree of variability was observed in the expression of the different chimeric HLA-I molecules across the cell lines (figure 4.2). The HLA-C molecules were consistently expressed at much lower levels than were the HLA-A and HLA-B molecules, and amongst HLA-A, the lowest expressing chimera was HLA-Aδ at 9,091 AFU compared with the other molecules that were consistently above 20,000 AFU.

4.2.2.2 Proportion of HIV-1 & HIV-2 Nef Positive Cells

The proportion of HIV-2 Nef positive cells following transfection was consistently lower than that of HIV-1 Nef (figure 4.3). A Welch's t-test was conducted to compare the proportion of HIV-1 Nef positive ($\bar{x}$ 0.062) and HIV-2 Nef positive ($\bar{x}$ 0.025) cells, and the difference was found to be highly statistically significant (t(630.25) = 25.587, p <0.0001).

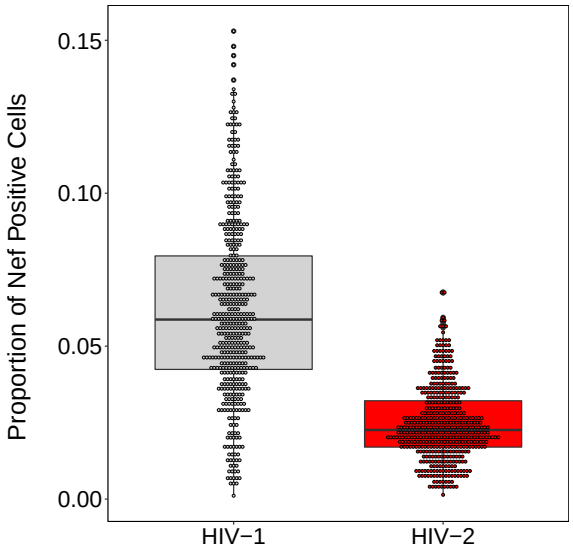


Figure 4.3: Proportion of HIV-1 & HIV-2 Nef positive cells following transfection of transgenic JJK-derived cell lines with the selected Nef isolates. HIV-1 is shown in grey; HIV-2 is shown in red. The box denote interquartile range, and the median is marked by the central horizontal line.

To determine whether this may confound analysis (as observed differences between the two viruses could potentially result from variable expression rather than the functional efficiency of Nef), the relationship between downregulation and proportion of positive cells was assessed by computing a Pearson's product-moment correlation coefficient. Fold change was calculated using equation 2.2 from data collected in triplicate, and HLA-C values were excluded as these molecules are not downregulated by Nef.

No statistically significant correlation was identified between the two variables (figure 4.4; $r(177) = -0.064$, $p = 0.397$), demonstrating that differences in the proportion of positive cells following transfection will not confound the analysis.

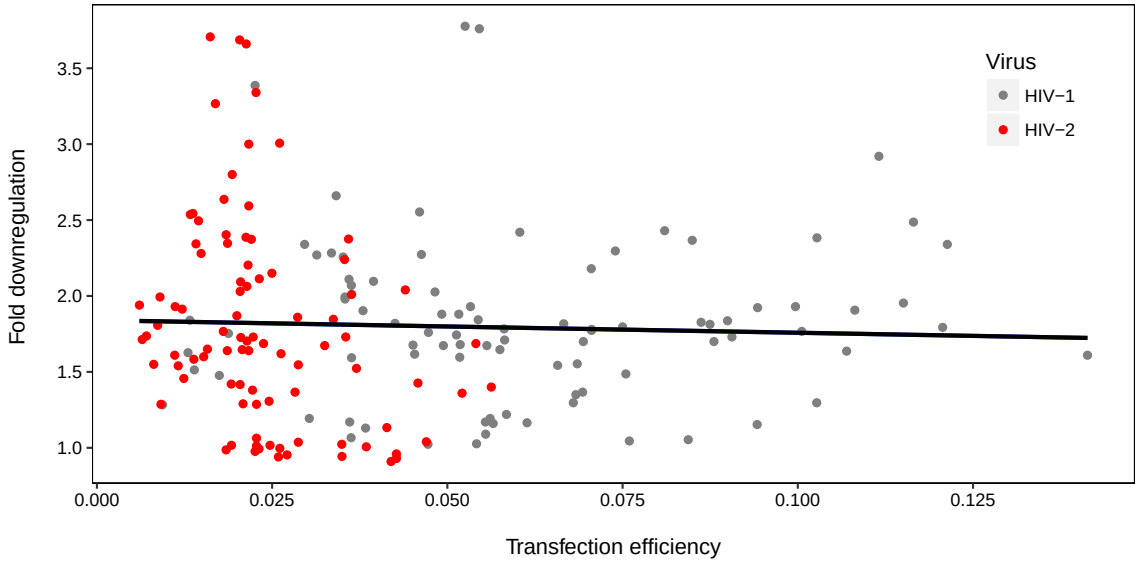


Figure 4.4: Relationship between the proportion of Nef positive cells following transfection and downregulation of HLA-I. HIV-1 is shown in grey; HIV-2 is shown in red. HLA-C excluded.

Having quality-checked the data in this way, analysis was performed on the full dataset of twenty Nef isolates from ten HIV-1 and ten HIV-2 patients transfected into each of the sixteen transgenic cell lines.

4.2.3 Nef-Mediated HLA-I Downregulation

4.2.3.1 Primary Isolates of HIV-1 & HIV-2 Nef do not Differentially Downregulate HLA-A & HLA-B

A key aim of this study was to identify whether Nef isolates from primary patient samples mediate differential downregulation of HLA-A and HLA-B molecules, as is reported to be the case with laboratory-adapted SF2 Nef [399]. In contrast to the SF2 findings, a Wilcoxon rank sum test indicated no statistically significant difference in the downregulation of HLA-A and HLA-B molecules mediated by primary isolates (figure 4.5). This was consistent for both HIV-1 ($W = 825$, $p = 0.186$) and HIV-2 Nef ($W = 842.5$, $p = 0.168$). HLA-C was not downregulated by either HIV-1 or HIV-2 Nef.

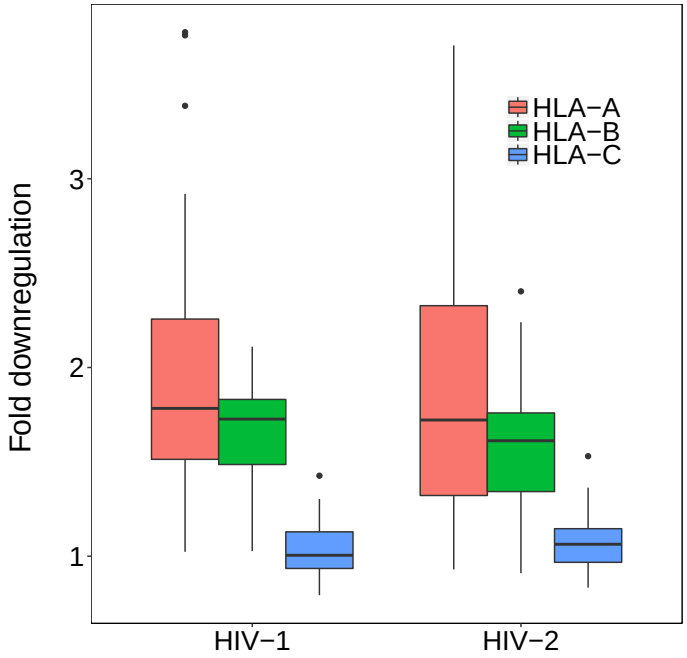


Figure 4.5: HLA-A and HLA-B molecules are not differentially downregulated by primary patient isolates of Nef. Data grouped by virus. The boxes span the upper and lower quartile, and medians are marked by the central horizontal lines. Whiskers denote ± 1.5 IQR, and dots show outliers.

4.2.3.2 HIV-1 & HIV-2 Nef do not Differentially Downregulate HLA-I

The ability of primary HIV-1 and HIV-2 Nef isolates to mediate HLA-I downregulation were next compared. HLA-C was again excluded from analysis. A Wilcoxon rank sum test failed to detect a statistically significant difference in the downregulation of HLA-I mediated by HIV-1 and HIV-2 patient Nef isolates (figure 4.6; $W = 4329.5$, $p = 0.350$). This was also true when HLA-A ($W = 2558$, $p = 0.548$) and HLA-B ($W = 236$, $p = 0.337$) were considered independently of each other.

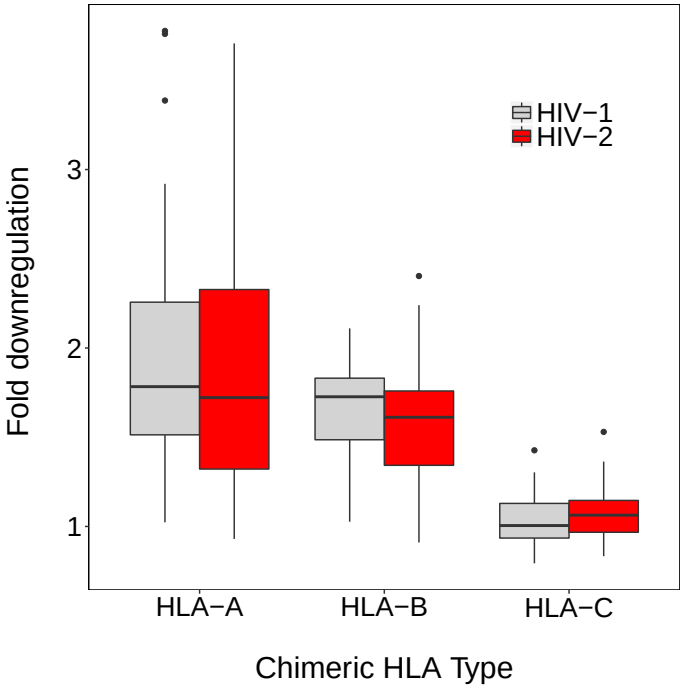


Figure 4.6: HIV-1 and HIV-2 Nef downregulate HLA-I to an equivalent extent. The boxes span the upper and lower quartile, and medians are marked by the central horizontal lines. Whiskers denote ± 1.5 IQR, and dots show outliers.

4.2.3.3 Nef Isolate-Specific Patterns of HLA-I Downregulation

Whilst Nef-mediated HLA-I downregulation did not differ between HIV-1 and HIV-2, the pattern and extent of functionality differed on an individual isolate level (figure 4.7). For example, isolates from patients SM335 and SM339 both downregulated chimeric HLA-Aα to the highest levels, but the difference between the individual HLA-I types was more pronounced in SM335. Furthermore, both SM046 and SM551 Nef most strongly downregulated HLA-Aη, although to similar levels. Several of the HIV-2 patient isolates downregulated HLA-Aα and HLA-Aγ slightly more effectively than other molecules. In all cases, amongst the HLA-A molecules, the extent of HLA-Aδ downregulation was consistently very low. Four patient isolates (SM528, SM547, TD014 & TD055) failed to downregulate any of the chimeric HLA-I molecules tested.

As a number of the patient Nef isolates mediated superior downregulation of a particular HLA type, it was next investigated whether the isolates more efficiently downregulate autologous HLA: i.e. the chimeric HLA molecules containing the cytoplasmic domains that represent the HLA types of the patients from which the Nef was isolated. However, whilst Nef from patient SM339 (who carries HLA-A2/A24) most strongly downregulated chimeric HLA-Aα (which shares its cytoplasmic domain with HLA-A2; appendix table A.3) there was no evidence for this in other isolates.

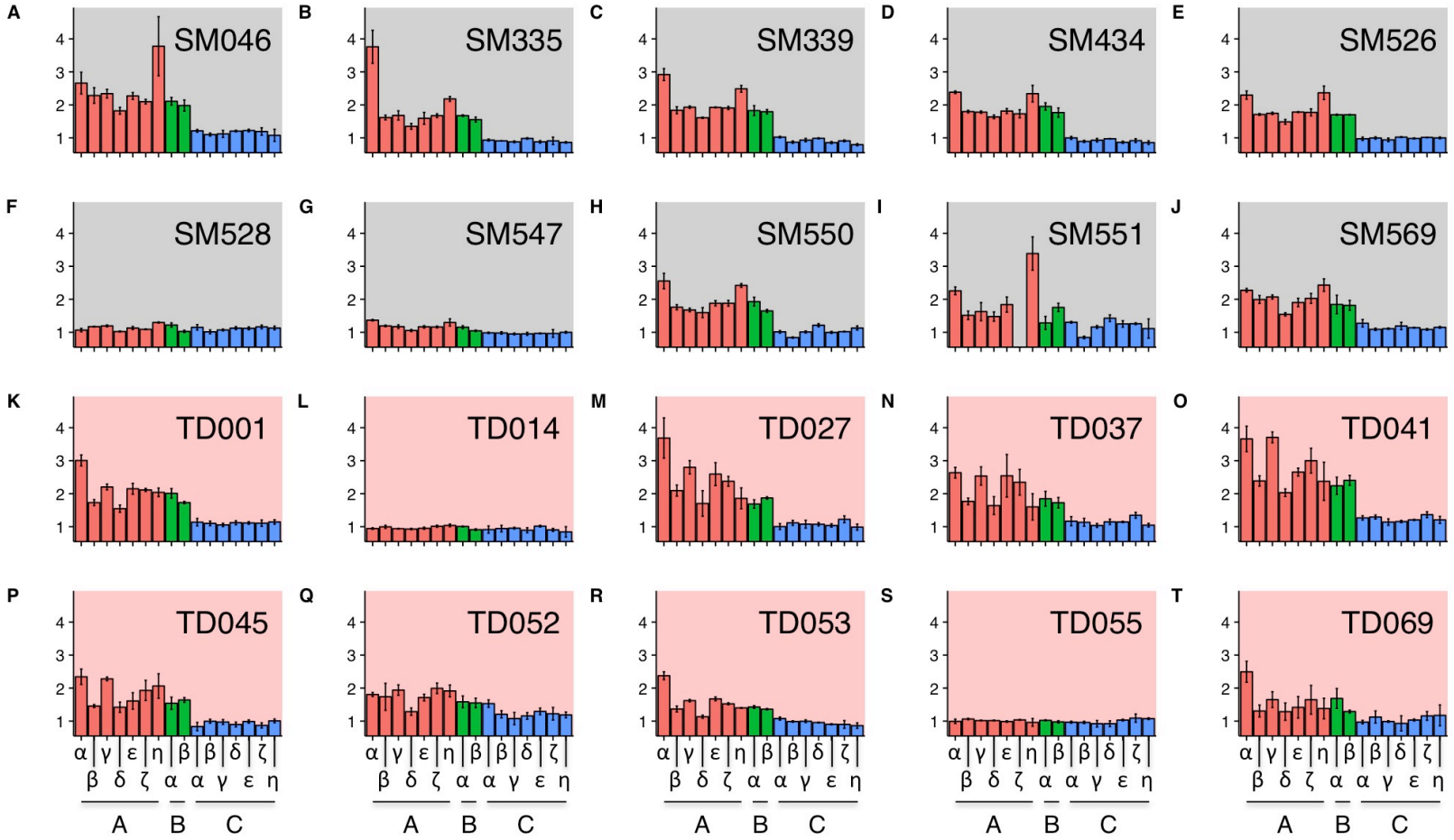


Figure 4.7: Plots showing HLA-I fold downregulation mediated by each primary patient Nef isolate. Plots correspond to patients: HIV-1 patients are shaded in grey and HIV-2 patients in red. Coloured bars denote HLA-I type: HLA-A in red, HLA-B in green, and HLA-C in blue. For clarity, axis labels have been omitted; The X axis shows chimeric HLA type; and the Y axis shows fold downregulation. Error bars are $\pm$ 1 SD.

Associations with autologous HLA were next assessed on a population level. In both the HIV-1 and HIV-2 cohorts, there is clear frequency hierarchy of HLA-I types (tables 4.1 & 4.2). HLA types were therefore grouped by the cytoplasmic tail that they express and the corresponding chimeric HLA-I molecule (figure 4.8). HLA-C was excluded as differences between groups could not be resolved with two-digit typing alone.

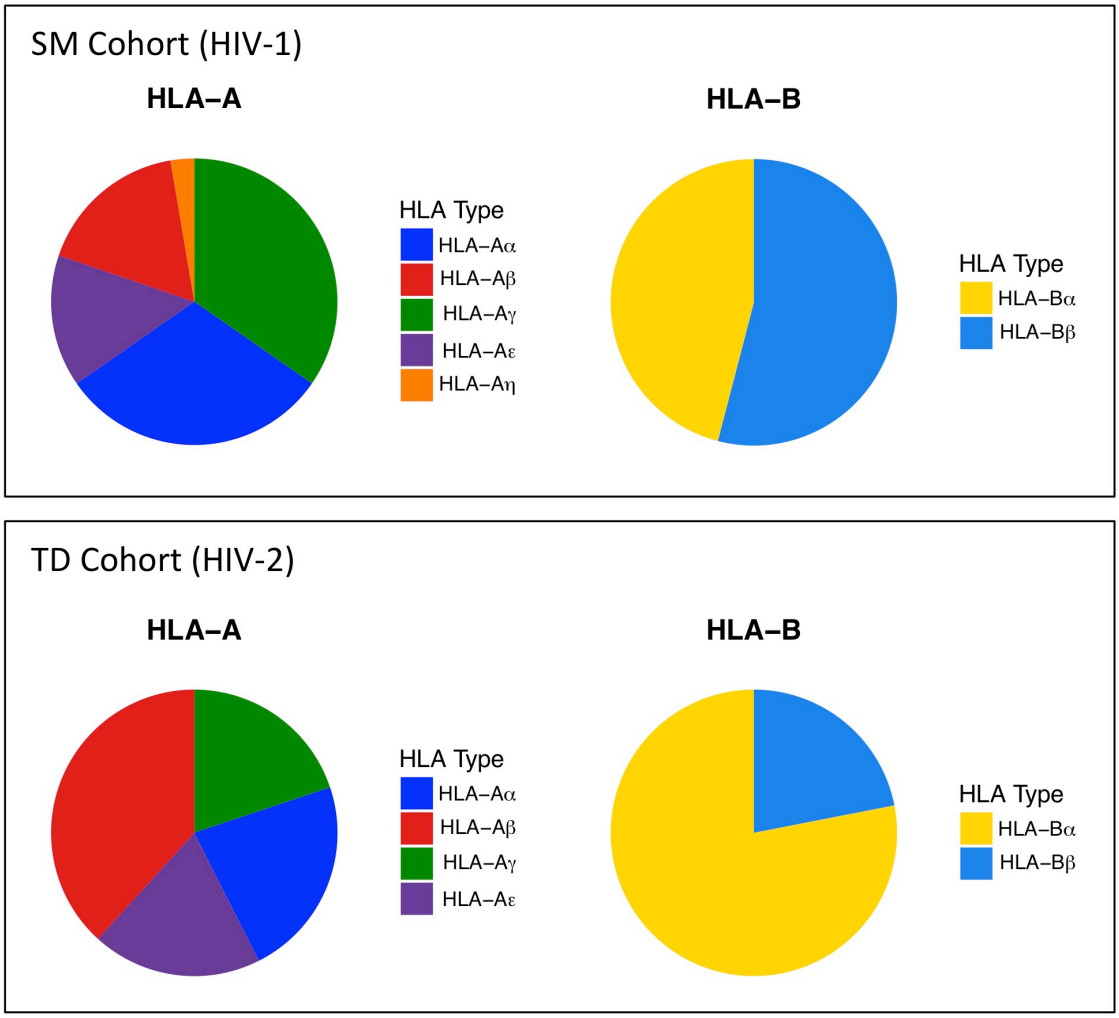


Figure 4.8: Frequency of chimeric HLA-A and HLA-B types in the HIV-1 (SM) and HIV-2 (TD) cohorts.

Within the HIV-1 SM cohort, the most frequent HLA-A types corresponded to HLA-Aα and HLA-Aγ, whilst the HLA-B molecules were split relatively evenly. In the HIV-2 TD cohort, HLA-Aβ-like HLA types were most frequent, and HLA-B types fell predominantly into the HLA-Bα group.

Whilst HLA-Aη was strongly downregulated by several of the patient isolates (figure 4.7), few patients within either cohort had HLA types corresponding to this group. Indeed, HLA-Aη-like HLA types were found in only a small fraction of the HIV-1 cohort, and did not occur at all in the HIV-2 cohort. Conversely, whilst HLA-Bα-like HLA types are present in a large majority of HIV-2 patients, there is no functional difference between downregulation of HLA-Bα and HLA-Bβ mediated by any of these patients' Nef isolates. These observations are not consistent with preferential downregulation of autologous HLA by primary Nef isolates.

4.2.3.4 Clinical Correlates of HLA-I Downregulation

As Nef-mediated HLA-I downregulation differed by patient isolate, the relationship between the extent of downregulation and patient clinical progression markers ($CD4^+$ count and viral load) was next assessed by computing Spearman's rank correlation coefficients. As CD4 percentage was not available for some patients, absolute $CD4^+$ cell count was used in all instances. Patients on antiretroviral treatment were excluded from analysis, and as clinical data from patient SM526 were not collected at the time of sample acquisition, this patient was also excluded. As HLA-C is not downregulated by Nef, only HLA-A and HLA-B were considered so as not to bias the analysis.

No statistically significant relationship was identified between the extent of HLA-I downregulation and viral load (figure 4.9A; r_S = -0.287, p = 0.264), or between HLA-I downregulation and absolute CD4 count (figure 4.9B; r_S = -0.034, p = 0.896).

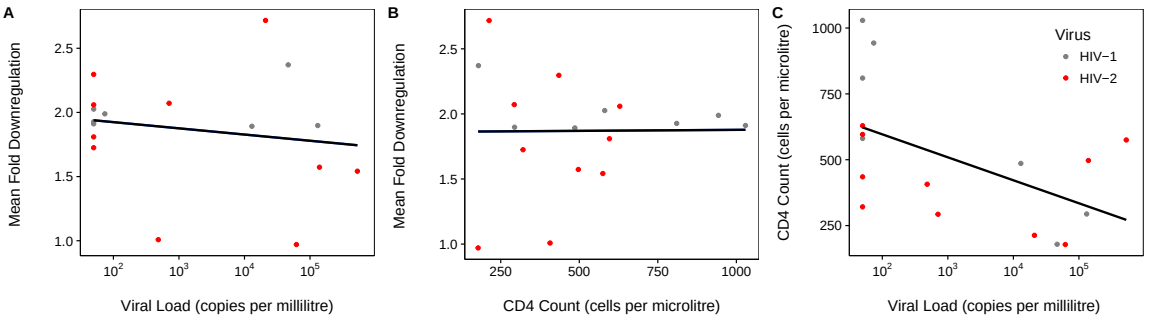


Figure 4.9: Relationship between downregulation of HLA-A and HLA-B molecules and patient clinical parameters. A) Mean fold downregulation and viral load; **B)** Mean fold downregulation and absolute CD4 count; **C)** Absolute CD4 count and viral load.

A moderate but statistically significant negative correlation was observed between absolute CD4 count and viral load (figure 4.9C; r_S = -0.531, p = 0.028).

4.2.3.5 Laboratory-Adapted SF2 Nef does not Differentially Downregulate HLA-A and HLA-B

Having failed to observe differential Nef-mediated downregulation of HLA-A and HLA-B alleles by primary patient Nef isolates (section 4.2.3.1), the original study reported by Rajapaksa *et al.* [399] was repeated in an attempt to verify the initial findings. The SF2 *nef* sequence was cloned into pAcGFP1-N1 (as used in the original study) and used to transfect each of the transgenic cell lines. Identical conditions, reagents, and equipment to those originally reported were applied.

No clear difference was observed between HLA-A and HLA-B (figure 4.10): the mean fold downregulation of HLA-A was 2.568 (SD 0.560); whilst for HLA-B was 2.312 (SD 0.087). Indeed, much larger differences were observed between the various HLA-A molecules, than between HLA-A and HLA-B. Due to the sample size (HLA-B: n=2), unequal variance, and small effect size, meaningful inferential statistics cannot be

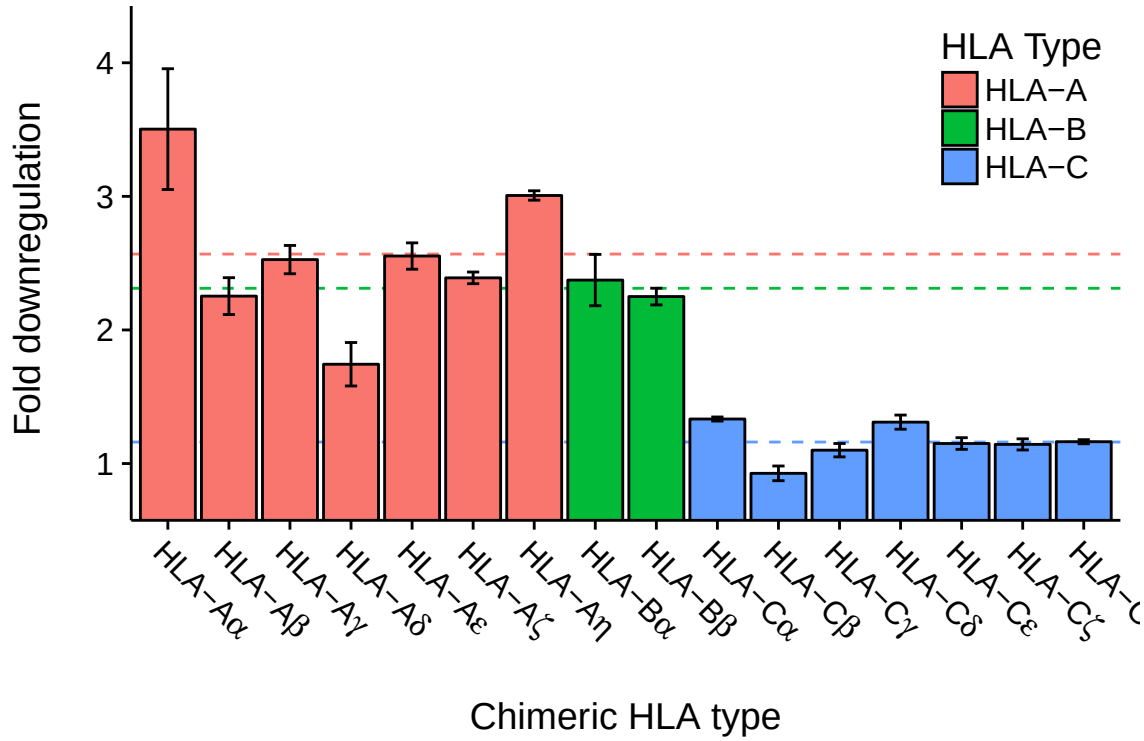


Figure 4.10: Laboratory-adapted SF2 Nef does not mediate differential downregulation of HLA-A and HLA-B molecules. Colour denotes HLA-I type. Dashed horizontal lines mark the mean fold downregulation value of each group of molecules. Error bars are ± 1 SD.

applied to this dataset [484]. However, it is worth noting that the same limitation applies to the original published study, from which the conclusion was initially drawn.

4.2.3.6 Most Primary Isolates of HIV-1 and HIV-2 Nef are Less Efficient at Mediating HLA-I Downregulation than SF2 Nef

The degree of downregulation mediated by the patient Nef isolates was next directly compared with that of SF2 Nef for each chimeric HLA type to determine relative activity (figure 4.11). The four Nef isolates that did not downregulate any of the chimeric HLA molecules (SM528, SM547, TD014, TD055) were excluded, as were data pertaining to HLA-C.

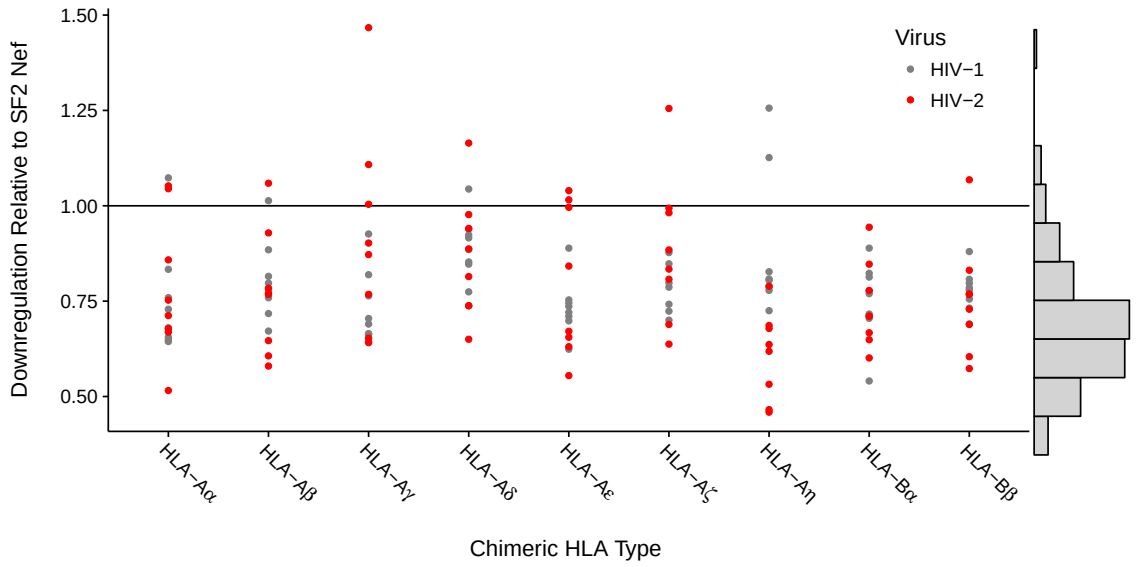


Figure 4.11: Most patient Nef isolates mediate downregulation of HLA-A and HLA-B less efficiently than does SF2 Nef. HIV-1 isolates are shown in grey and HIV-2 isolates are shown in red. The marginal histogram denotes count data for the relative downregulation values.

The patient isolates downregulated the HLA-I molecules less effectively than did SF2 Nef in the majority of cases (figure 4.11). A one-sample t test was conducted to compare the relative activity of the primary isolates to SF2 Nef ($\bar{x}$ 0.792:1), and the difference was found to be highly statistically significant (t(142) = -77.640, $p <$ 0.0001).

4.2.3.7 Consensus HIV-1 & HIV-2 Nef do not Differentially Downregulate HLA-A and HLA-B

Patient Nef isolates were next compared with corresponding consensus HIV-1 and HIV-2 Nef (section 2.3.1). Downregulation of each chimeric HLA-I molecule mediated by the consensus HIV-1 and HIV-2 Nef proteins was plotted (figure 4.12). Again no evidence for differential downregulation of HLA-A and HLA-B molecules was observed for either HIV-1 (HLA-A: μ 3.869 SD 0.607 & HLA-B: μ 3.487 SD 0.462) or HIV-2 (HLA-A: μ 3.710 SD 0.973 & HLA-B: μ 3.580 SD 0.434) .

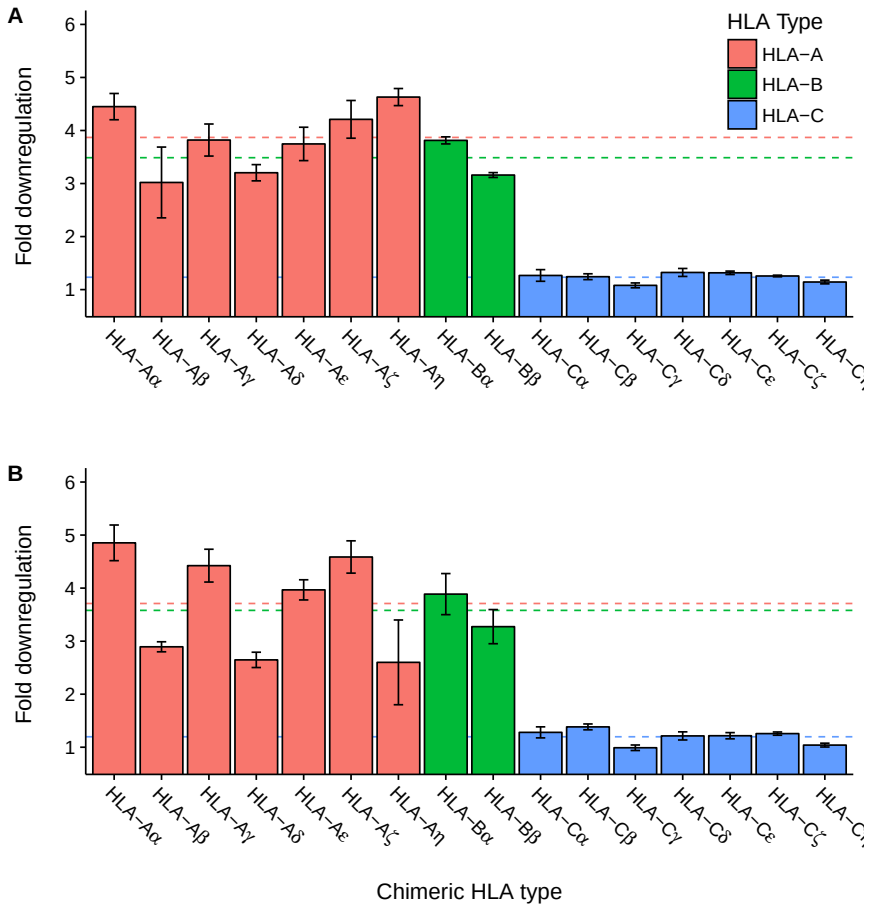


Figure 4.12: Fold downregulation of HLA-I mediated by consensus Nef. **A)** HIV-1; **B)** HIV-2. Colour denotes HLA-I type. Dashed horizontal lines mark the mean fold downregulation value of each group of molecules. Error bars are ± 1 SD.

4.2.3.8 Primary Isolates of Nef are Less Efficient at Mediating HLA-I Downregulation than Consensus HIV-1 & HIV-2 Nef

The data from consensus HIV-1 and HIV-2 Nef were then directly compared with the patient Nef-mediated downregulation data and a clear difference was observed. All primary HIV-1 and HIV-2 isolates mediated downregulation of HLA-A and HLA-B to a lesser extent than did the corresponding consensus sequence, without exception (figure 4.13). Here, the four patients that did not downregulate any of the HLA-I alleles were again excluded as described in section 4.2.3.6.

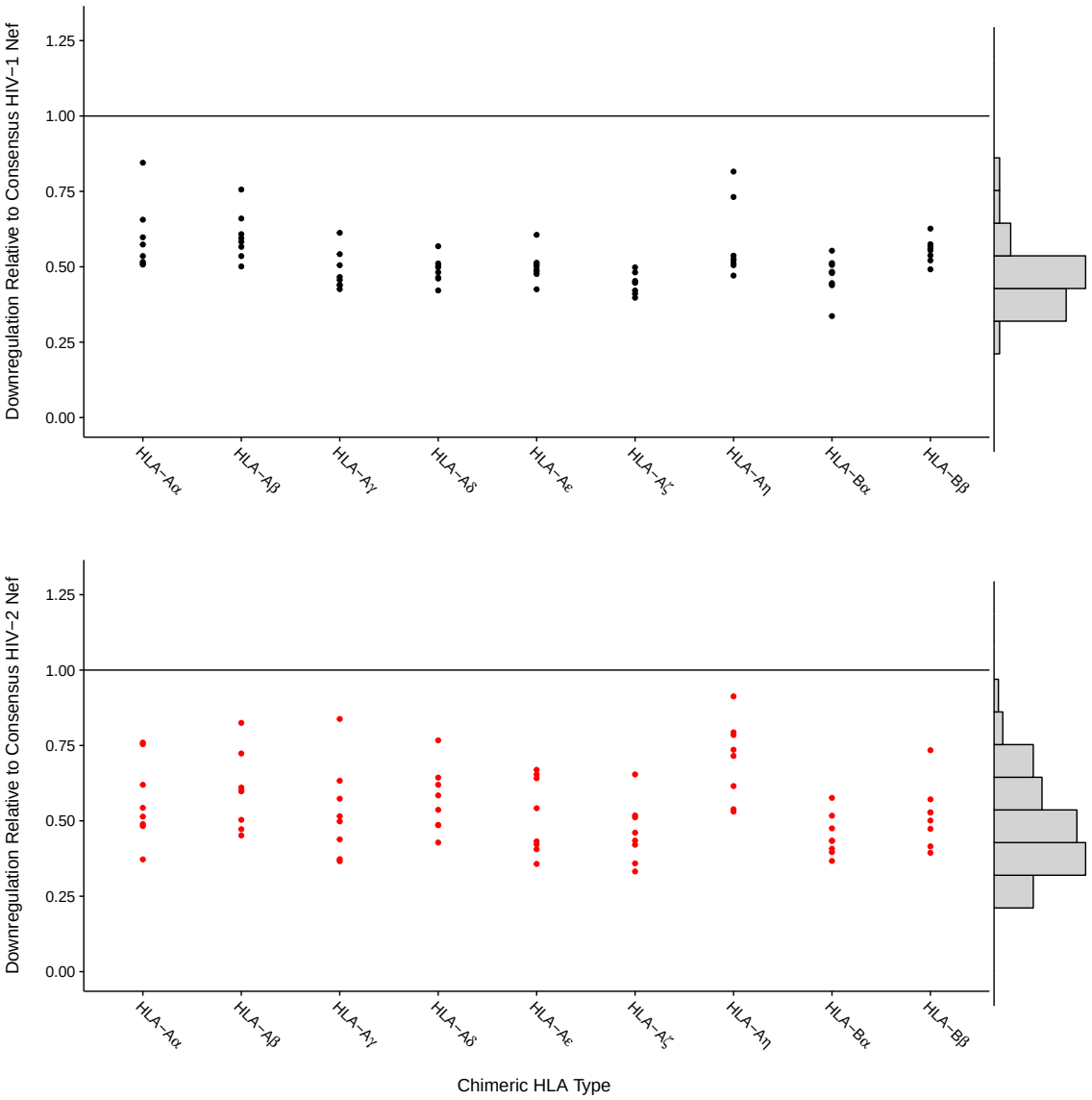


Figure 4.13: Patient Nef isolates mediate downregulation of HLA-A and HLA-B less efficiently than do the corresponding HIV-1 and HIV-2 consensus Nef proteins. HIV-1 isolates are shown in grey and HIV-2 isolates are shown in red. The marginal histogram denotes count data for the relative downregulation values.

On average, HLA-I downregulation mediated by the primary HIV-1 Nef isolates was approximately half that of consensus HIV-1 Nef ($\bar{x}$ 0.525:1 SD 0.090). A one-sample t test found that this difference was highly statistically significant (t(70) = -44.696, $p < 0.0001$). The HIV-2 Nef isolates behaved similarly ($\bar{x}$ 0.546:1 SD 0.135), and again, the difference was highly statistically significant (t(71) = -28.627, $p < 0.0001$).

4.2.3.9 Mapping of Nef Sequence Variation

As all patient isolates mediated HLA downregulation less efficiently than did the corresponding consensus Nef, mismatches to the HIV-1 consensus sequence in all HIV-1 patient sequences were mapped to identify variable regions. These were then overlaid to the functional motifs associated with HLA-I downregulation, as well as all known antibody, CTL, and T helper cell epitopes reported in the LANL Immunology Database (figure 4.14) [401, 485, 486]. The equivalent analysis was not performed with the HIV-2 sequences as epitopes are much less extensively mapped in this virus [319].

Most of the mismatches from the HIV-1 consensus sequence exist outside of the functional motifs (figure 4.14A), and especially towards the C-terminus. Of the eight functional motifs, the N-terminal α-helix (light green) had diverged most substantially from the HIV-1 consensus (for motif functions and positions, see table 4.3).

Few (28) antibody epitopes are recorded for HIV-1 Nef as it is (predominantly) an intracellular protein. These epitopes are found throughout the protein with a slightly higher distribution towards the N- and C-termini (figure 4.14B). Whilst the Nef isolates also exhibited more variability towards the ends of the protein compared with the middle, there did not appear to be a direct relationship between variability and the location of antibody epitopes.

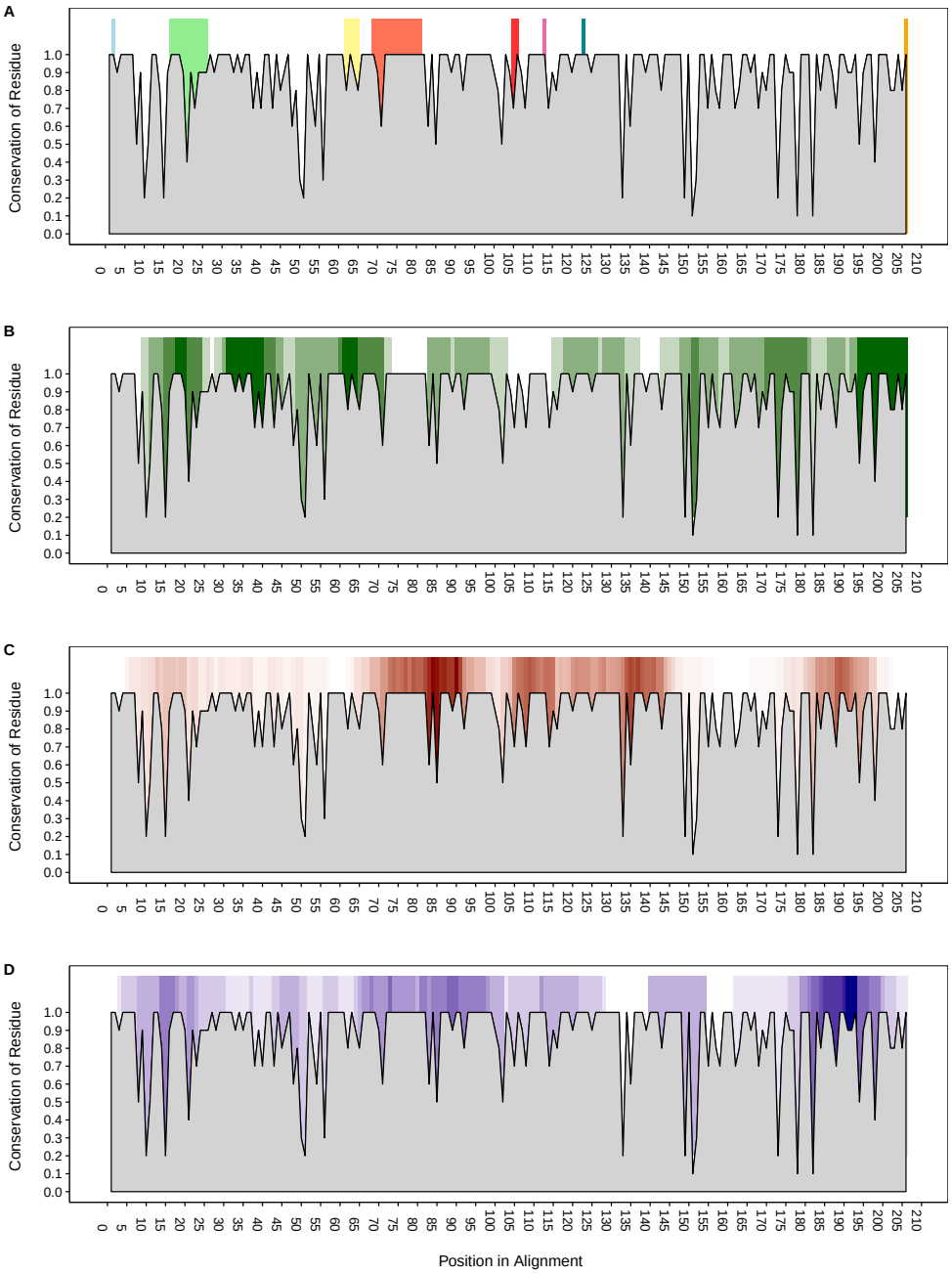


Figure 4.14: Patient Nef sequence variation mapped to **A)** the functional motifs associated with HLA-I downregulation (see table 4.3 for positions and functions), as well as **B)** antibody epitopes, **C)** CTL epitopes, and **D)** T helper cell epitopes. The number of known epitopes mapped to a position is indicated by colour intensity. Amino acid conservation is denoted by the grey box. Sites that are identical to the HIV-1 consensus across all ten HIV-1 patient sequences are recorded as 1; those that differ from the consensus in all patients are recorded as 0. Figure constructed from data made available by The Los Alamos National Laboratory (LANL) [401, 485, 486].

Currently there are 235 CTL epitopes in Nef recorded in the LANL database. In contrast to the antibody epitopes, these cluster much more centrally in the protein (figure 4.14C). There is no clear relationship between the location of these epitopes and amino acid conservation within the samples, and indeed the most variable regions exist predominantly in stretches of the protein that have sparse coverage of CTL epitopes.

Lastly, forty T helper cell epitopes have been described in HIV-1 Nef (figure 4.14D). These are found throughout the protein with a denser cluster towards the C-terminus. Again, no relationship is apparent between their location and sequence variability in these patients.

4.2.3.10 Structural Modelling of Sites of Interaction Between Nef and the Cytoplasmic Domain of HLA-I

As the data presented have consistently shown no differential downregulation between HLA-A and HLA-B, the amino acid substitutions that distinguish these two groups of molecules were mapped to a published structure (RCSB PDB Accession 4EMZ [462]) of the partial HLA-I cytoplasmic tail (positions: 314-332; magenta) in complex with HIV-1 Nef (cyan) and the μ1 chain of the adaptor protein AP-1 (green). This is shown in figure 4.15.

The only sites that consistently differ between the two groups of cytoplasmic tails are positions 315 and 316 (HLA-A: D315 R316; HLA-B: G315 G316) as shown in appendix table A.4. However, mapping these to the structure reveals that these two residues (dark blue) do not appear to come into direct contact with the Nef protein. Furthermore, an additional substitution at position 326 (yellow) distinguishes HLA-Bβ-like molecules (S326) from HLA-Bα-like molecules (C326). However, despite this

substitution occurring right at the interface of the HLA-I cytoplasmic tail with both Nef and AP-1, it does not confer any reported functional difference between the groups of molecules.

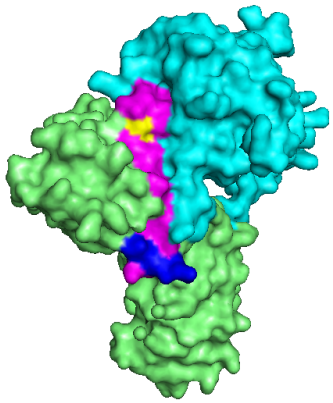


Figure 4.15: Structural modelling of the HLA cytoplasmic tail amino acid mismatches which distinguish HLA-A and HLA-B. The HLA-I cytoplasmic domain is shown in magenta; Nef is shown in cyan; the μ1 chain of AP-1 is in green and wraps around the rear of HLA-I. Positions 315 and 316 which distinguish HLA-A from HLA-B are shown in dark blue; whilst position 326 which distinguishes HLA-Bβ and HLA-Bα is shown in yellow. Substitutions have been mapped to the Jia *et al.* [462] structure under RCSB PDB accession number 4EMZ.

4.2.3.11 HLA-I Downregulation Correlates with Expression Level

Throughout this investigation, certain chimeric HLA-I types were consistently downregulated at lower levels than others. For example, whilst HLA-Aα was often amongst the most strongly downregulated, HLA-Aβ and HLA-Aδ were downregulated poorly (for example, figures 4.7, 4.10, 4.12). This was not the result of improved downregulation of autologous HLA-I (section 4.2.3.4) and so the cause of this difference was of great interest.

The relationship between baseline expression levels of the HLA-I chimeric molecules (section 4.2.2.1) and the extent of downregulation was assessed for all Nef isolates, excluding those that did not downregulate any of the HLA-I molecules. HLA-C

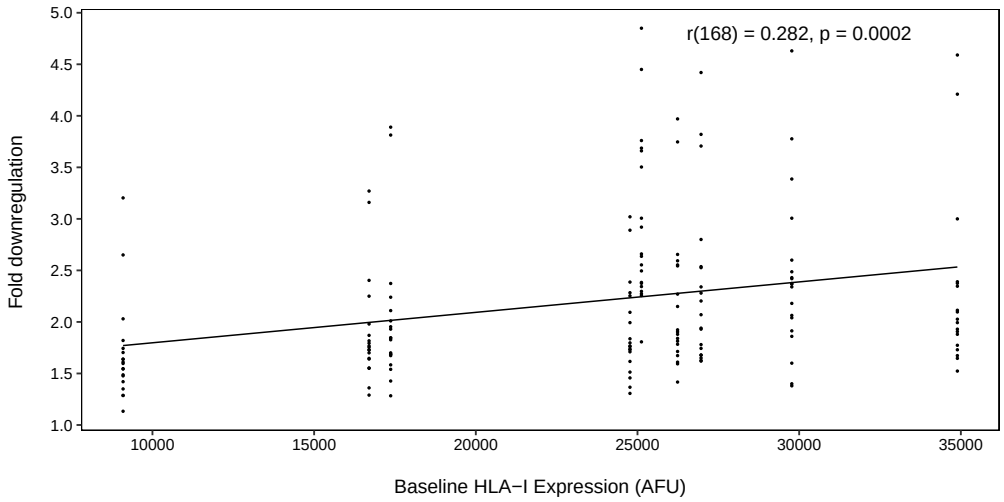


Figure 4.16: Correlation between baseline HLA-I levels and fold downregulation A weak yet statistically significant positive linear relationship was detected.

was again excluded. A weak positive linear relationship was detected (figure 4.16). Computing a Pearson's product-moment correlation coefficient demonstrated that the relationship was statistically significant ($r(168) = 0.282$, $p = 0.0002$).

Therefore, the fold downregulation of a molecule is affected - at least in part - by the amount of that molecule initially found on the surface of the cell: specifically, molecules with high resting expression levels have higher fold downregulation than those with low expression levels.

4.3 Discussion

In light of recent evidence that SF2 Nef may differentially downregulate HLA-A and HLA-B molecules, the main aim of this chapter was to investigate this phenomenon in Nef isolates from primary patient samples. Within this context, both HIV-1 and HIV-2 Nef were studied, and the functionality of these proteins were compared.

For the purposes of stratified sampling, HLA types were first characterised and patients were selected based on the frequencies observed. The distributions of HLA types in each cohort were broadly in keeping with population-level frequencies, matched by ethnicity [487].

Having selected an appropriate subset of patients, functional studies were performed. Following observations that HIV-1 consistently achieved higher transfection efficiency than did HIV-2 Nef isolates, the relationship between transfection efficiency and the extent of HLA-I downregulation was assessed for the purpose of quality-control. No correlation was detected, indicating that any differences in downregulation are unlikely to be an artefact of the transfection efficiency, and therefore can be used to infer the functional activity of Nef.

The consistently higher transfection efficiency of HIV-1 compared with HIV-2 Nef indicates that these two proteins must differ in some key way. HIV-2 Nef is substantially longer than HIV-1 Nef (257 aa compared with 208 aa, on average), and as the size of plasmids has been shown to correlate inversely with transfection efficiency [488], size alone could account for this difference.

However, when cloned into the pIRES2-ZsGreen1 backbone, the size ratio of the prepared plasmids is only 1:1.02 (HIV-1:HIV-2). With fifteen nanograms per trans-

fection, this equates to a difference in copy number of 2.35×10^9 to 2.29×10^9 which is not proportional to the 2.52-fold reduction in transfection efficiency observed. This strongly indicates that a biological rather than experimental factor may underlie the observed difference: for example differences in toxicity, stability or expression level may exist.

In contrast to previous reports, this study found no evidence for differential downregulation of HLA-A and HLA-B molecules. Indeed, this was consistent across the board for both HIV-1 and HIV-2. No statistically significant difference was detected between HLA-A and HLA-B downregulation mediated by the primary isolates of Nef, and whilst inferential statistics could not be appropriately applied to individual Nef data - specifically, SF2 and consensus Nef - descriptive statistics provided no circumstantial support either. It is also of note that whilst inferential statistics are inappropriate for the analysis of individual Nef isolates - owing to the small sample size (HLA-B n=2), unequal variance, and small effect size [484] - the initial conclusions drawn by Rajapaksa *et al.* [399] were corroborated by such statistics, and should thus be weighted accordingly.

In a similar study recently published by Mahiti *et al.* [489], the authors attempted to assess whether naturally-occurring HIV-1 Nef sequences mediate differential downregulation of HLA-A24, HLA-B35, and HLA-C04. Were these HLA-I molecules to be categorised according to appendix table A.3, they would be classed as HLA-Aϵ, HLA-Bα, and HLA-Cα/γ, respectively. Whilst the authors did find a statistically significant difference between HLA-A24 and HLA-B35 downregulation, the effect size was small, with only around 0.8 fold mean difference between them. Indeed, this difference is substantially smaller than the observed variation within the HLA-A molecules alone in this presented study. Furthermore - as highlighted by data

demonstrating that amongst the groups of molecules, substantial variation can be seen - a single representative molecule from HLA-A and HLA-B cannot adequately be used as a surrogate for all HLA-A and all HLA-B molecules. I.e. one should not infer from the statement "HLA-A23 and HLA-B35 are differentially downregulated" that "HLA-A and HLA-B are differentially downregulated".

Mapping of the cytoplasmic tail has provided further circumstantial evidence against differential HLA-A and HLA-B downregulation as the only residues consistently differing between HLA-A and HLA-B molecules sit outside of the interface with Nef. However, a recognised limitation of this study is that it has been performed using a crystal structure from only one type of HLA-A cytoplasmic tail. It cannot be ruled out that these differences in sequence may confer a degree of structural change that could, in principle, bring the residues of interest into closer proximity with the interface. However, due to the orientation of the molecules and the similar tail lengths, this seems improbable.

Furthermore, if a difference did exist between the downregulation of HLA-A and HLA-B - which is not supported by these data - it would be important to consider the functional implications. Namely, whether a minimal difference such as those presented here ($\approx$ 0.24 fold mean difference), and by Mahiti *et al.* [489] ($\approx$ 0.8 fold mean difference) is functionally significant and corresponds to an appreciable difference in visibility of the infected cells to CTLs and subsequent killing.

In a similar line of thinking, one of the purported implications of the original study was that differential downregulation of HLA-A and HLA-B may explain the improved control of HIV infection associated with HLA-B restricted CTLs. However, one could argue that the logic of this statement is flawed, as whilst some HLA-B types are

associated with improved prognosis and better viral control (HLA-B27 and HLA-B57 as examples), this is not a universal feature of HLA-B in general: indeed, all humans carry two copies, including rapid progressors. A more convincing argument could have been put forward if a difference was seen between the downregulation of the two groups of HLA-B molecules. Had one of these groups contained all (or many) of the protective HLA-B molecules and had exhibited increased resistance to downregulation relative to the other HLA-B group, an argument could be put forward that resistance to downregulation of protective HLA-B alleles may contribute to better control of infection and improved prognosis. However, again, this was not supported by the data.

Moving on from differential downregulation of HLA-A and HLA-B, this study also found no evidence for a difference between HIV-1 and HIV-2 Nef-mediated HLA-I downregulation. This was consistent for both HLA-A and HLA-B, and the data do not support a contributing role of Nef-mediated HLA-I downregulation in the prognostic differences observed between HIV-1 and HIV-2 infection.

Whilst no evidence for differential downregulation was found overall between HLA-A and HLA-B or between HIV-1 and HIV-2, there was some evidence that individual patient isolates more efficiently downregulate particular HLA-I types than others. It was hypothesised that Nef isolates may more efficiently downregulate HLA-I types associated with the patient from which they were isolated, as a result of viral adaptations in response to immunological selection pressure. However, no relationship was identified between autologous HLA-I types and the extent of downregulation, either on a patient or population level for either cohort.

Furthermore, some patient isolates were identified that did not downregulate any of the HLA types. Whilst the sample size is too small to perform a comprehensive analysis of the residues responsible, both of the HIV-1 isolates that did not downregulate HLA-I contained interesting mutations in known functional domains or flanking regions of Nef. Specifically, the only Nef isolate exhibiting a G3N mutation in the C-terminal flanking region of the myristoylation site was isolated from patient SM528. Whilst glycine is small and non-polar, asparagine contains a relatively large polar side chain which could interfere with myristoylation and membrane association in this variant. SM547 Nef is also the only isolate which contains a highly divergent N-terminal α-helix in which M20 is notably absent (M20I substitution)[2]. Performing site-directed mutagenesis to restore the G3N and M20I substitutions followed by the functional studies described would determine whether these mutations were responsible for the reduction in function.

In addition to the two HIV-1 isolates, two HIV-2 isolates also failed to downregulate HLA-I. However, inference of the residues responsible is not realistic as the functional domains of HIV-2 are poorly mapped, and do not necessarily equate to those identified in HIV-1.

The extent of downregulation by each Nef isolate was also found not to correlate with patient clinical parameters. The absence of a relationship between the extent of HLA-I downregulation and either $CD4^+$ count or viral load in these two clinically latent cohorts is in standing with published findings reported in acute infection that

[2]As an aside, the SM547 Nef protein was found to belong to subtype C rather than subtype B' virus, indicating that this patient became infected with the virus independently of the narrow-source outbreak associated with this cohort (section 2.1.1). Whilst this has limited impact on the current study, this patient should be removed from the SM cohort for phylogenetic studies in particular.

HLA-I downregulation does not correlate with viral set point [490]. It is also possible that the less active forms of Nef could result from the variants being isolated from proviral DNA rather than circulating plasma virus.

Another key finding of this study was that primary Nef isolates were less functional than both laboratory-adapted and consensus Nef sequences in the context of HLA-I downregulation. In terms of the SF2 Nef, it was hypothesised that this may be the result of a lack of immunological pressure *in vitro*. SF2 Nef is a relatively laboratory-adapted strain and has been grown extensively in culture [491]. In the absence of an effective immune response, selection pressure may be reduced, allowing the virus to accumulate mutations that restore functions of Nef which had been previously affected by escape mutations. Whilst HLA downregulation is of little functional use *in vitro*, other functions of Nef may be more strongly selected for, and overlap may exist in the motifs responsible.

Similarly, as consensus sequences present the most prevalent amino acid at a position in a protein over a population level, the effects of immunological pressure within individual patients - particularly in CTL epitopes - may be masked. Therefore, it was assumed that consensus Nef may behave as a functionally "ideal" variant without the constraints of immunological compromise. As each primary isolate downregulated HLA-I less efficiently than did the corresponding consensus and SF2 Nef, this strongly indicates that functional activity is to some extent compromised *in vivo*. It was hypothesised that CTL pressure could be the driving factor.

To test this hypothesis, sequence variation within the Nef protein was next mapped and these data were overlaid with the functional motifs, as well as CTL, antibody and T helper cell epitope maps. In all cases, no direct relationship could be established.

However, a key limitation of the analysis is the small sample size resulting from having only ten patients from which sequences and functional data were obtained. Inflating the sample size with sequences available in the literature would be of limited value without corresponding functional data; and functional data obtained from disparate studies is not directly comparable where different experimental variables come into play such as the expression vector and method of quantification used. Furthermore, the demographic underpinning of each cohort has resulted in few HLA types being studied (tables 4.1 & 4.2). Indeed, whilst it would be preferable to narrow the investigation down to known epitopes of the most frequent HLA types in each cohort, the small number of both patients and described epitopes would make it impossible to detect signal over noise.

Additionally, whilst mapping known epitopes gives an indication of immunological activity against a particular site, it would be naïve to assume that all epitopes against all regions of Nef have been described: particularly for the less common HLA-I types. Therefore, whilst this study has not conclusively identified an immunological cause for the reduced functional activity of primary isolates of Nef, this hypothesis should not yet be rejected owing to the current limitations of the study.

It was also shown that the variable extent of Nef-mediated downregulation of particular HLA types was not due to improved function against autologous HLA. Instead, downregulation was found to correlate with HLA-I expression levels. Whilst these cell lines are highly artificial, the HLA-A molecules are consistently the most highly expressed (with the exception of HLA-Aδ), the HLA-C molecules are consistently low, and the HLA-B molecules are found to be intermediate. A similar pattern can be detected in the figures published by Mahiti *et al.* [489] - despite unsubstantiated

assertions that expression levels were equivalent - and so it is plausible that expression may also confound the conclusions drawn from this study.

The relative levels presented herein are similar to those reported in peripheral blood lymphocytes measured by flow cytometry and mass spectrometry [492]. Furthermore, convincing evidence has been put forward that HLA-I expression levels play an important role in the control and pathogenesis of HIV-1 infection [493].

Finally, this investigation has generated novel data that conflicts with recently published studies proposing differential Nef-mediated downregulation of HLA-A and HLA-B molecules. The findings presented herein have been consistent and reproducible, and the interpretations have been made in light of balanced arguments. As a result of the broadly different conclusion drawn from this study, it is proposed that substantial further investigation should be made not into differential downregulation of HLA-A and HLA-B, but rather differential expression levels of HLA types and the associations they may have with control of infection. Research groups are already beginning to fill these gaps in knowledge, and data from such studies may revolutionise how the field views the adaptive immune response against HIV.

In performing this study, several limitations have been identified in the way that surface molecule internalisation is measured and analysed. Such limitations are often difficult or impossible to control for, yet may dramatically impact the interpretation of data and the conclusions drawn. The next chapter will discuss some of these limitations and their implications, before detailing the development of a novel technique for the analysis of receptor modulation measured by flow cytometry.

Chapter 5

A Novel Technique for the Analysis of Flow Cytometric Data: Quantifying Receptor Modulation with SWARM

5.1 Introduction

Interpretation of flow cytometric data can be fraught with potential pitfalls: many of which arise as the result of overly-simplistic bifurcate analyses, which categorise cells into two populations: those that are negative and those that are positive for a fluorophore of interest.

Perhaps the simplest and most widely used bifurcate analysis for receptor modulation is fold change: for examples, see Rose *et al.* [187], Rajapaksa *et al.* [399], and Feldmann *et al.* [494] amongst others. Fold change is calculated by comparing the mean fluorescence intensity (MFI) of a particular fluorophore (hereafter, "fluorophore Y") in a population of treated cells with that of an untreated, control population. To calculate fold upregulation, the MFI of the treated population of cells is divided by that of the untreated cells; whereas the reverse is the case in fold downregulation.

However, there are a number of inherent limitations to bifurcate analyses - which will be discussed - and which may lead to misinterpretation of data in some experimental contexts. This chapter describes the development of a novel technique - the staggered window approach to receptor modulation (SWARM) - which aims to address many of these limitations. SWARM can be applied across different cell lines and has a broad range of applications, making it highly transferable and of high potential value to the wider scientific community.

5.1.1 Limitations of Bifurcate Analyses

5.1.1.1 Bifurcate Analyses Assume a Discrete Population Distribution

As bifurcate analyses compare positive and negative populations of cells, the interpretation of these data is dependent on the assumption that the two populations are clearly defined and distinct, represented by the idealised plot in figure 5.1A.

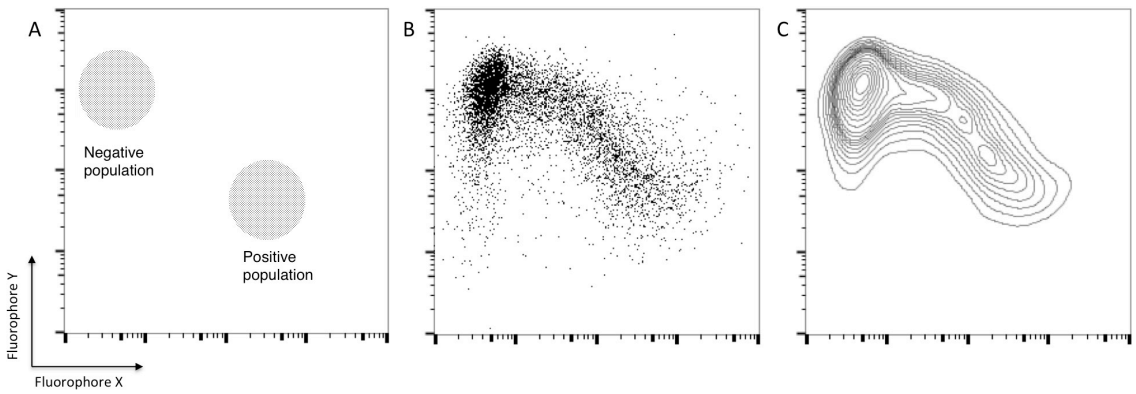


Figure 5.1: Idealised and actual distributions of positive and negative populations of transfected cells A) Idealised plot where cells positive and negative for fluorophore X form distinct, non-overlapping populations. Gating of these populations would be unambiguous. **B)** Dot plot of experimentally-collected flow cytometric data from the transfection of transgenic JJK-derived cell lines with selected Nef variants. **C)** The data presented as a contour plot (level 5%) and outliers excluded, demonstrating the highly smeared nature of the population distribution.

However, in reality populations are rarely so well defined, and plots typically represent a continuum, as demonstrated in figures 5.1 B & C. Here there is no clear distinction between the cells that are positive or negative for a fluorophore of interest (hereafter, "fluorophore X"), making gating highly subjective.

In cases where fluorophore X has been introduced by transfection of a plasmid encoding it, negative cells can be distinguished by gating on an sample of untransfected cells. However, as transfection can alter the state of cells [495], it may be preferable to gate instead on cells transfected with a dummy vector that does not encode the fluorophore. Conversely, if fluorophore X is introduced by staining with a conjugated antibody against a molecule of interest, it may be preferable gate on an unstained control. Whilst such gates separate out truly negative cells, they are not particularly informative as they do not distinguish between cells producing high levels of the fluorophore and those expressing marginally more than background fluorescence.

Occasionally, studies attempt to distinguish between low-, medium- and high-levels of expression in cells by manually assigning additional gates, either visually based on the distribution of the data, or by using log decades: for an example, see Mann *et al.* [496]. The issue here is that the positioning of such gates can and does significantly impact the result of the analysis. Figure 5.2 demonstrates this by showing that the same dataset can generate very different fold change values, based solely on the manual positioning of gates.

Furthermore, whilst data can be analysed if gating is consistent within an experiment, it may not be possible to directly compare data across different experiments, let alone independent studies.

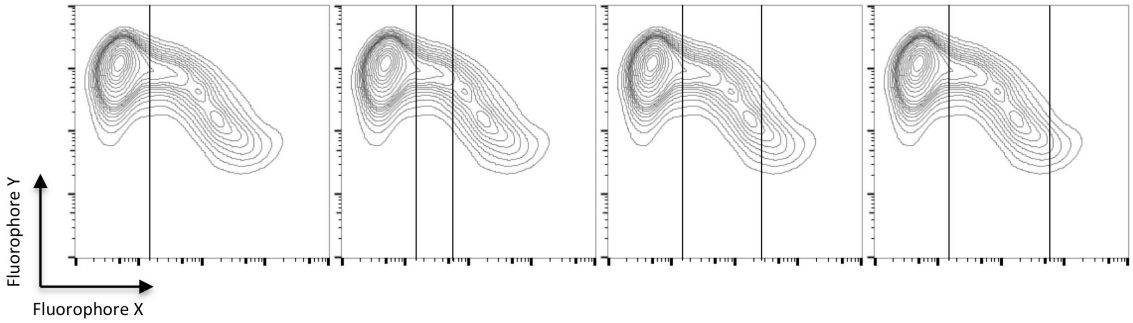


Figure 5.2: The importance of appropriate gating in bifurcate analyses. The same data can be interpreted differently, and give very different fold change values dependent entirely on the manual positioning of gates. Here, one dataset generates fold downregulation values of 2.24, 3.89, 6.60 & 7.18 when "high-expressing" cells are gated as shown in each panel. The gate defining the negative population has been fitted to untransfected cells and does not change.

5.1.1.2 Bifurcate Analyses are not Robust Against Experimental Variation

Experimental variation can affect the values generated by bifurcate analyses. Whilst every effort is made to ensure consistency, error will be introduced, and the rate is proportional to the size of the experiment.

Variation in transfection efficiency is introduced by numerous factors including the method of transfection, the health of the cells, and even the stage of the cell cycle during which the transfection is performed [497]: as such, even replicates performed on the same day under the same conditions are likely to differ to some degree. Additionally, with an increasing number of samples comes increased manual handling time, which may result in treatments - such as transfection and staining - being completed at staggered intervals. There may then be a high degree of disparity in the number (and expression level) of positive cells in the subsequent analysis which are actually attributable to experimental variation rather than true physiological differences.

Such experimental variation - whilst common - is difficult to control for, and can have a measurable impact on the results of bifurcate analyses. An example of this is illustrated in figure 5.3, showing two independent technical replicates of a single transfection. Despite the transfections being performed under identical conditions, different transfection efficiencies were achieved. The first replicate achieved a transfection efficiency of 27.5%, whilst the second was less efficient at 15.9%.

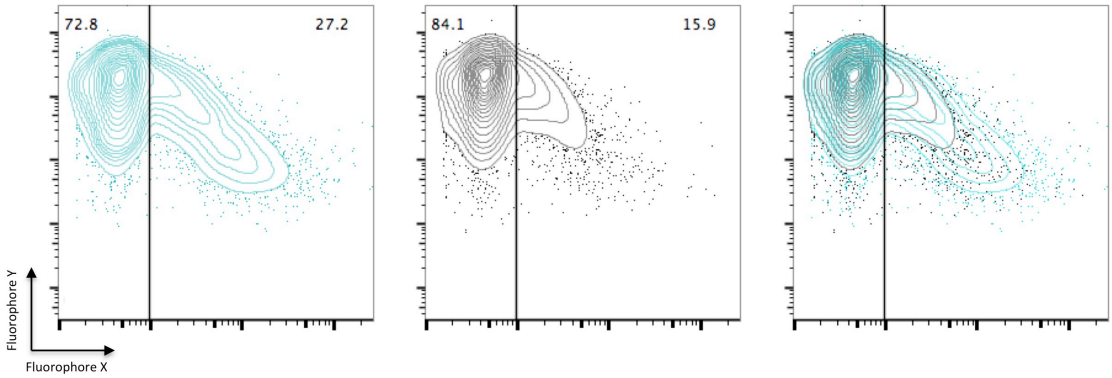


Figure 5.3: The effect of transfection efficiency on bifurcate analyses. Two technical replicates of a transfection of chimeric JJK-derived cell lines with selected Nef, in which the efficiencies achieved were 27.2% & 15.9%. Overlaying these data demonstrates an equivalent slope of HLA-I downregulation, but the fold change values generated are 2.66 and 1.98, respectively.

More efficient transfection results in a greater proportion of the cells expressing high levels of fluorophore X. This increases the MFI of fluorophore X within the positive population of cells, which in turn brings down the MFI of fluorophore Y, independent of any physiological difference. The plots in figure 5.3 generated fold downregulation values of 2.66 and 1.98 respectively. In such cases, the difference in fold change is an artefact of fewer high-expressing cells in samples that have achieved less efficient transfection rather than a product of true biological variation. This is supported by the overlay, which shows equivalent slopes of downregulation in each replicate, despite differences in expression levels.

Controlling for confounding factors in experiments can be problematic, as variation can be introduced in so many ways. As standardisation is often unrealistic, it is important to perform an analysis that can account and correct for these inconsistencies.

5.1.1.3 Bifurcate Analyses may not be Informative if Initial Expression Varies

A major argument for the use of particular bifurcate analyses such as fold measurements is that they emphasise change rather than raw values, allowing standardisation of data when required. Whilst this is a very useful quality in some analyses, it can be a major drawback when consideration of absolute values is important for understanding the physiological relevance of a phenomenon. This is particularly salient when considering, for example, HLA-I downregulation (introduced in chapter 4) as the absolute surface expression level of some HLA types has been shown to correlate with the magnitude of the CTL response, which in turn directly affects HIV control [493].

This concept is illustrated in figure 5.4 showing the transfection of two cell lines. Here, despite similar measures of fold downregulation, there is approximately three times less fluorophore Y remaining in the fluorophore X positive population of cells in the first plot than in the second plot (231 versus 699 AFU respectively). This is primarily the result of the difference (dY) in the initial expression level of fluorophore Y in each cell line. Whilst dY looks small, it is important to note that the logarithmic scale somewhat obscures the large difference in expression (865 AFU) between the two starting populations, highlighting the importance of using statistical analysis rather than basing initial interpretation on the visual appearance of the plots.

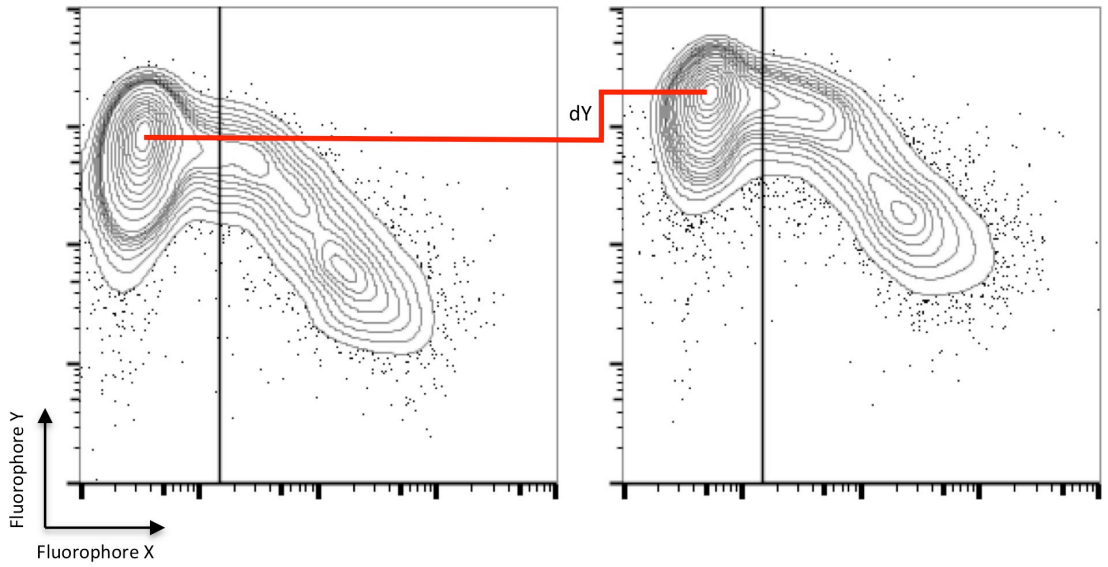


Figure 5.4: The effect of varying initial expression levels on bifurcate analyses. The transfection of two transgenic JJK-derived cell lines exhibiting different initial expression levels of fluorophore Y. The MFI of fluorophore Y in the fluorophore X negative populations are 677 and 1542 (dY = 865) for each plot, and in the positive populations are 699 and 231 respectively, despite similar fold change values (difference 0.73).

Here, if fluorophore Y represents HLA-I, and fluorophore X represents Nef, figure 5.4 implies that the Nef positive cells in the first plot are less visible to $CD8^+$ lymphocytes than those in the second plot. This may be missed as the difference in fold downregulation is small, but it could in part contribute to the difference observed between particular protective and high risk HLA-I types. Whilst this is a hypothetical example, the conclusion to draw from this is that important physiological consequences may be masked if absolute values are ignored in preference of data standardisation. The raw values generated in these studies can be highly informative but are often excluded due to incompatibility with comparative analytical techniques.

5.1.2 Proposals for an Improved Analysis

Building on the limitations described, an improved analysis should:

- Account for the continuous distribution of cell populations measured with flow cytometry;
- Be robust against experimental variation such as differences in transfection efficiency, peak expression of constructs and incubation times;
- Provide comparable data for molecules that exhibit different initial expression levels;
- Provide informative data that give insight into the physiological relevance of receptor modulation.

A program was designed to implement a simple sliding window-inspired algorithm to bin flow cytometric data before analysis. Binning the data in this way overcomes several of the issues inherent to bifurcate analyses by treating cells with a range of expression levels independently, without relying on subjective manual gating. This overcomes issues relating to experimental variation - especially in large studies where variables become harder to control for - and also limits opportunities for human error and bias by minimising decision-making in gate placement.

The program also accounts for background fluorescence, and reports receptor modulation in terms of percentage - rather than fold - change. Whilst converting between the two is mathematically simple, the interpretation of percentage change is more intuitive when considering the remaining portion of receptors still present on the cell's surface, and the physiological relevance of this regarding visibility to CTLs.

This chapter will outline the development of the SWARM program, and present a working example of where this approach may outperform standard bifurcate analysis.

5.1.3 Aims

The specific aims of this project were to:

- Develop, optimise and validate a novel analytical technique for flow cytometric data;
- Apply the novel technique to an experimental dataset;
- Compare and contrast the information generated through the novel and traditional analyses.

5.2 Results

5.2.1 SWARM Workflow

SWARM analyses flow cytometric data by performing a series of tasks according to the following workflow:

1. Assign all cells to bins by the fluorescence intensity of fluorophore X;
2. Calculate the MFI of fluorophore Y for each bin;
3. Calculate the mean percentage of downregulation of fluorophore Y for each bin, using calculation 5.1;

$$\%\ \text{internalisation} = \left\{1 - \frac{MFI\left(Y_{X+}\right) - MFI\left(Y_{background}\right)}{MFI\left(Y_{X-}\right) - MFI\left(Y_{background}\right)}\right\} \times 100 \qquad (5.1)$$

4. Plot the raw data, and overlay the calculated mean values for verification;
5. Generate a best fit line to the calculated mean values for data visualisation;
6. Combine values from experimental triplicates;
7. Plot the percentage of fluorophore Y internalised in each bin;
8. Export the resulting values.

As SWARM was developed with a particular research question in mind - *"Does Nef mediate differential downregulation of different HLA-I molecules?"* - the work described in this chapter will hereafter invariably refer to fluorophore X as Nef, and fluorophore Y as HLA-I molecules. An example of this workflow in action is shown in figure 5.5.

Plots can then be overlaid to demonstrate differences between experimental variables such as cell lines and isolates used. In addition, as the algorithm ensures that bin

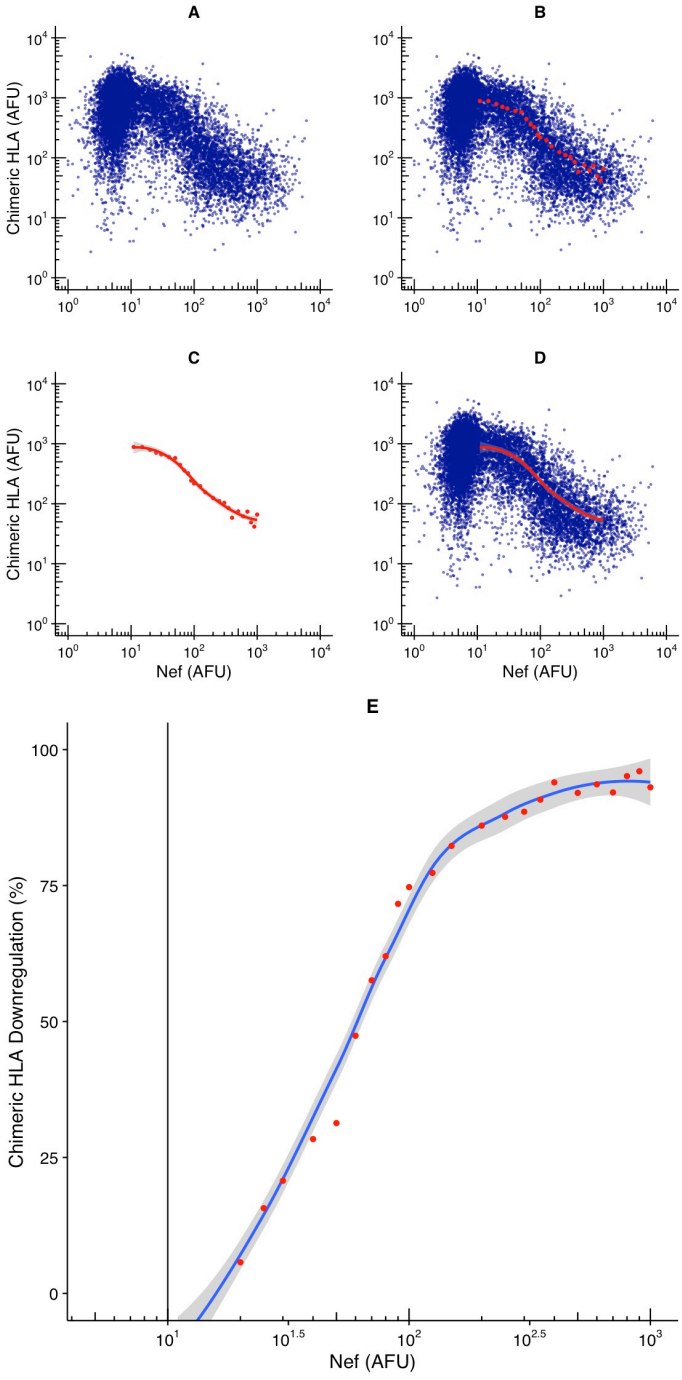


Figure 5.5: The SWARM workflow. A) Flow cytometric data are imported and the raw values (blue dots) plotted. **B)** Cells are binned by expression of fluorophore X (here, Nef), and the MFI of fluorophore Y (here, chimeric HLA) within each bin is calculated (red dots). **C)** A best fit line (red) is plotted. **D)** The best fit line is superimposed onto the raw data to verify fit. **E)** The percentage of fluorophore Y downregulation in cells within each bin is calculated and the mean is taken from experimental replicates (triplicate) for plotting (red dots; blue line). Representative plot of the cell line expressing chimeric HLA-Aα transfected with SM046 Nef.

positioning is consistent across the dataset, the bins can be directly compared using standard descriptive and inferential statistics. The development and optimisation of the workflow will next be described.

5.2.2 Vector Optimisation

Binning and relative quantification of data by SWARM requires a 1:1 ratio of the protein of interest (here, Nef) to the fluorescent marker with which it is being detected. Whilst the pIRES2-ZsGreen1 vector was sufficient for chapter 4, an IRES vector does not drive the expression of both encoded proteins at equal levels, and the upstream protein is typically produced at higher levels [498]. Therefore, for this study a fusion vector was required.

Owing to its tertiary structure, appropriate construction of a Nef fusion protein is paramount. Fusing a fluorophore to the N-terminus of Nef would interfere with the myristoylation domain impacting membrane association and function, and so fusions must be made to the C-terminus. However, the length of the linker region separating Nef and the attached fluorphore must too be optimised.

By cloning primary HIV-1 Nef isolates into an expression vector with a 3 aa stretch separating it from the fluorophore (AcGFP1), expression was detected in only a small fraction of cells, and in these cells, expression was occurring at low levels. Cloning further upstream in the multiple cloning site (MCS) to increase the linker to 7 aa allowed the same Nef isolate to be correctly expressed in a higher number of cells at substantially higher levels. This is demonstrated in figure 5.6. This observation was consistent across all HIV-1 Nef isolates in all transgenic cell lines, and in the absence of other modified variables, the rescue effect is congruous with the increased linker length permitting correct folding and stability of the protein [499].

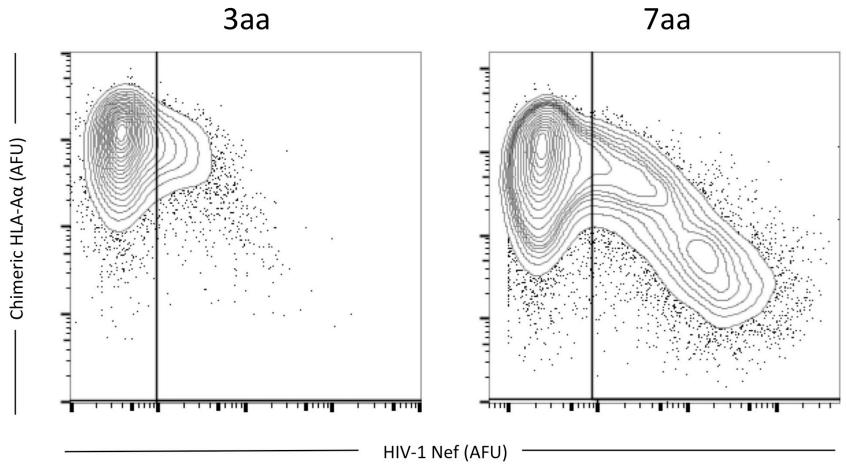


Figure 5.6: Linker length affects the folding and function of HIV-1 Nef in fusion. Representative plot of a cell line stably expressing chimeric HLA-Aα, transfected with a primary HIV-1 Nef isolate, with either a 3 aa or a 7 aa linker separating Nef from AcGFP1. Strong expression could be detected with the 7 aa linker, but expression was poor with a 3 aa linker. All other variables were controlled, and this observation was consistent across HIV-1 variants and cell lines.

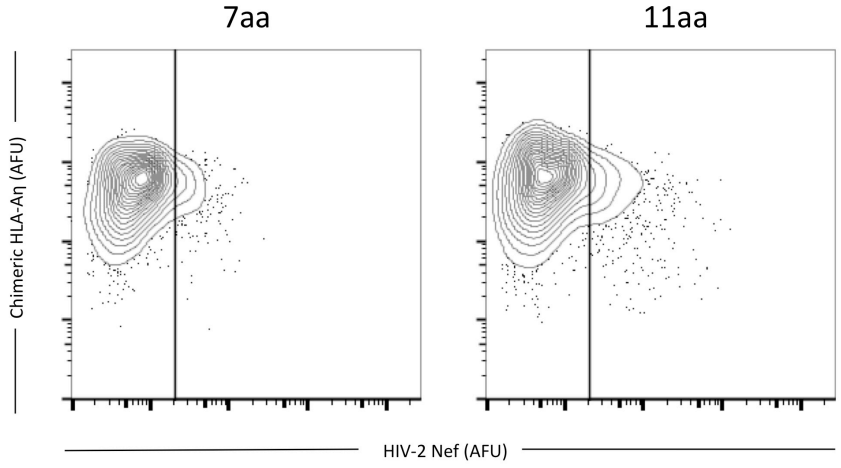


Figure 5.7: Linker length affects the folding and function of HIV-2 Nef in fusion. Representative plot of a cell line stably expressing chimeric HLA-Aη, transfected with a primary HIV-2 Nef isolate, with either a 7 aa or an 11 aa linker separating Nef from the AcGFP1. Poor expression was achieved. All other variables were controlled, and this observation was consistent across all HIV-2 variants and cell lines.

However, this issue was more pronounced with HIV-2 Nef, which was not resolved by inserting a 7 aa linker sequence. Cloning further upstream the MCS to increase the linker to 11 aa elicited only a modest improvement in expression (figure 5.7). Again, this observation was consistent across HIV-2 primary Nef isolates and across cell lines. The inability of the 7 aa linker to rescue folding and function of HIV-2 Nef could indicate a structural difference between HIV-1 and HIV-2 Nef at the C-terminus.

Rather than continue cloning further upstream to attempt to restore expression, an existing fluorescent fusion vector was modified to generate an *in situ* highly flexible linker region between Nef and the fluorophore that would not need to be regenerated for each variant cloned. The benefit of this approach is that whilst cloning further upstream generates a linker composed of the amino acids encoded by the MCS, generating a linker *de novo* allows the amino acids to be selected based on biochemical properties, for example, short side-chains to increase flexibility.

Briefly, the pAcGFP1-N1 vector was modified according to the schematic workflow described in figure 5.8 to replace the region spanning the EcoRI and AgeI restriction sites with a custom synthetic gene. The full nucleotide sequences of the synthetic genes used for the generation of this vector are listed in appendix table A.5.

The modified flexible vector permitted correct folding and strong expression of both HIV-1 and HIV-2 and was a marked improvement on previous constructs. This was true for both the consensus sequences and primary patient isolates (figure 5.9). Having optimised the expression vector, data was collected from all selected samples (section 2.3.12) for the development of the SWARM analysis.

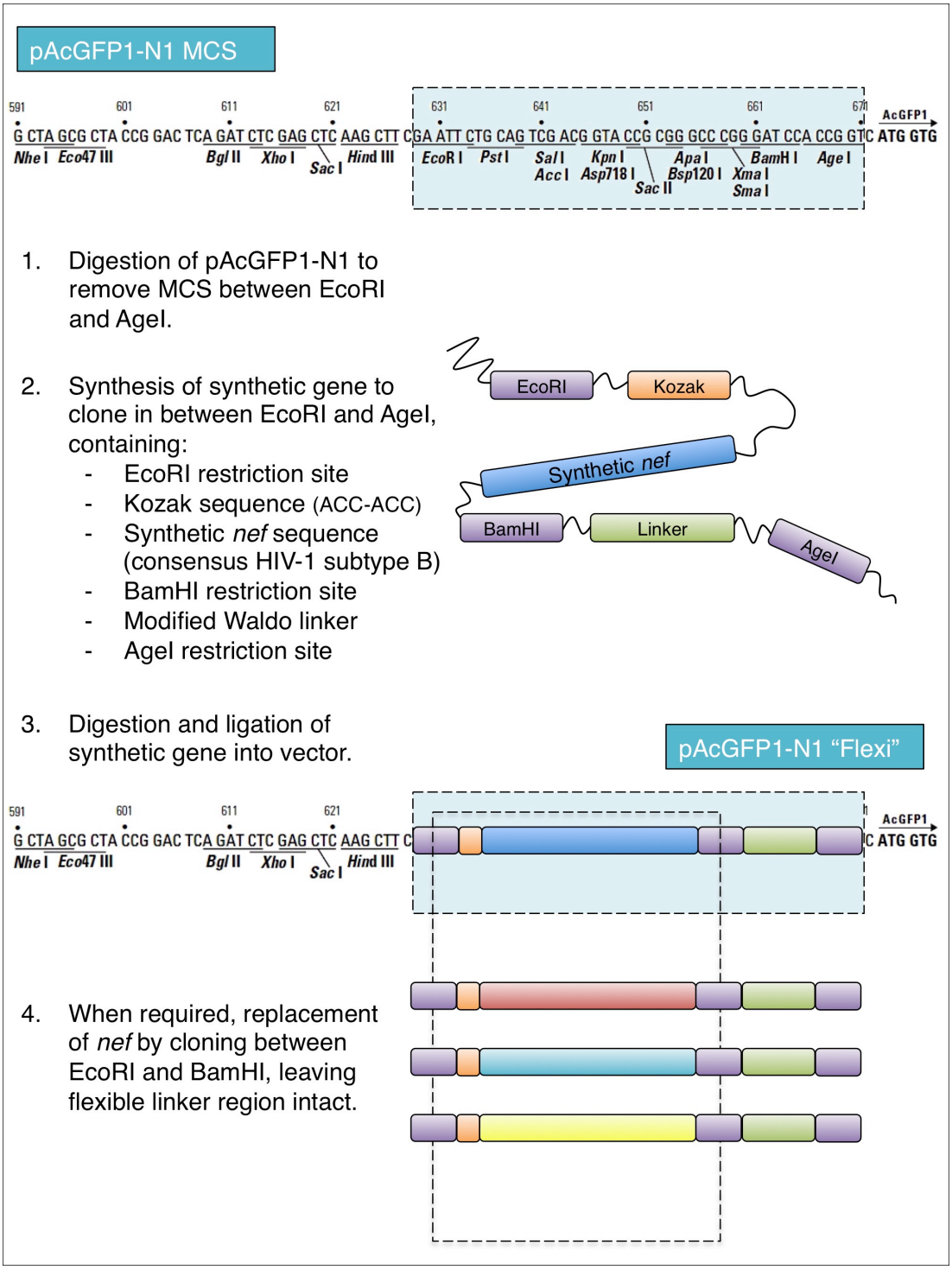


Figure 5.8: Schematic representation of the generation of a modified expression vector from pAcGFP1-N1. The expression vector contains a highly flexible linker adapted from Waldo *et al.* [400], which will not be cleaved out with subsequent re-cloning of different genes of interest between the BamHI and EcoRI restriction sites. Figure constructed using some elements of vector maps published by Clontech TaKaRa.

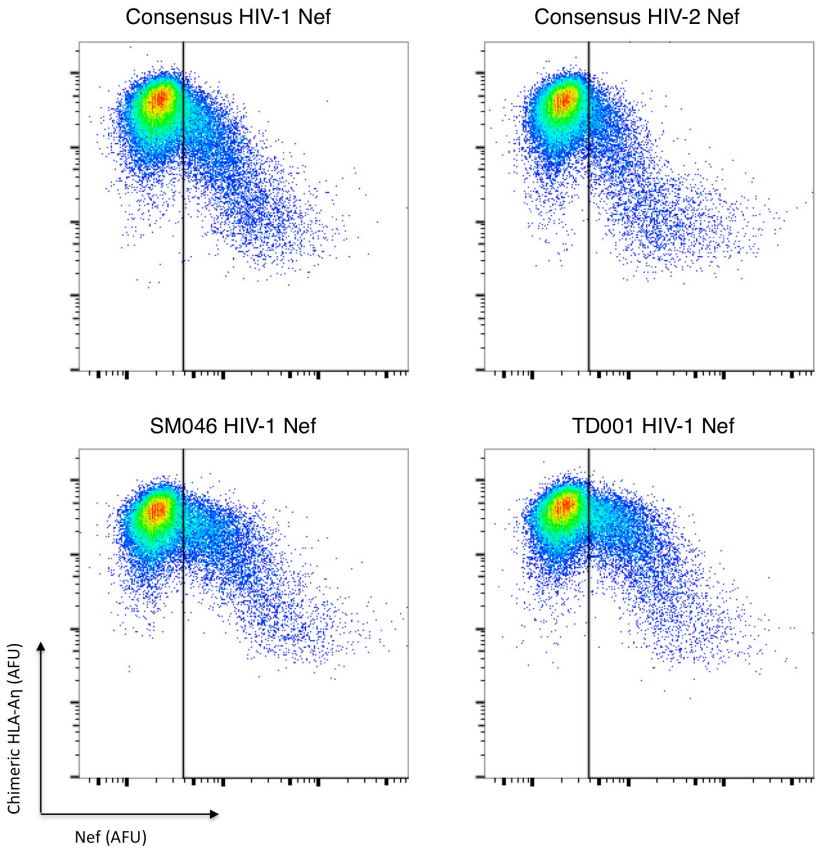


Figure 5.9: Representative example of Nef expression in the modified flexible fusion expression vector. Consensus HIV-1 (upper left) and HIV-2 Nef (upper right), as well primary Nef isolates from patients SM046 (HIV-1; lower left) and TD001 (HIV-2, lower right). Transfections performed in the cell line stably expressing chimeric HLA-Aη are shown.

5.2.3 Development of Staggered Window Approach

A sliding window algorithm is a simple and efficient way to determine if and how a relationship exists between the levels of two molecules in a cell: here, whether increasing concentrations of Nef affect the extent of HLA-I downregulation. Traditional sliding window algorithms iterate across a dataset by a defined window length and step size. For example, in an array of integers from 1 to 10, a sliding window with a length of 4 and a step size of 2 would group the data accordingly: (1,2,3,4), (3,4,5,6), (5,6,7,8), (7,8,9,10). When the number of points in a dataset is fixed, this permits direct comparison of bins.

To demonstrate that sliding windows are robust ways of identifying trends in data such as these, a series of plots were generated using a range of window lengths and step sizes (figure 5.10). Neither altering the window length (figure 5.10A-C; overlaid 5.10D) nor the step size (figure 5.10E-G; overlaid 5.10H) appreciably altered the shape of the curve fitted to the resulting points.

However, the limitation of a sliding window in this study is that the number of cells is not fixed. The three million cells that are counted out for transfection are in reality not precise, and countless additional variables such as transfection efficiency, cell viability, and pipetting error contribute to further differences. This means that whilst plotting the values generated by such a sliding window would give a strong estimation of the trends in downregulation, the values within each window cannot be directly compared with corresponding windows of a different dataset (for example, where a different Nef isolate has been used; or even experimental replicates) as the cells contained within the windows are not equivalent in terms of expression.

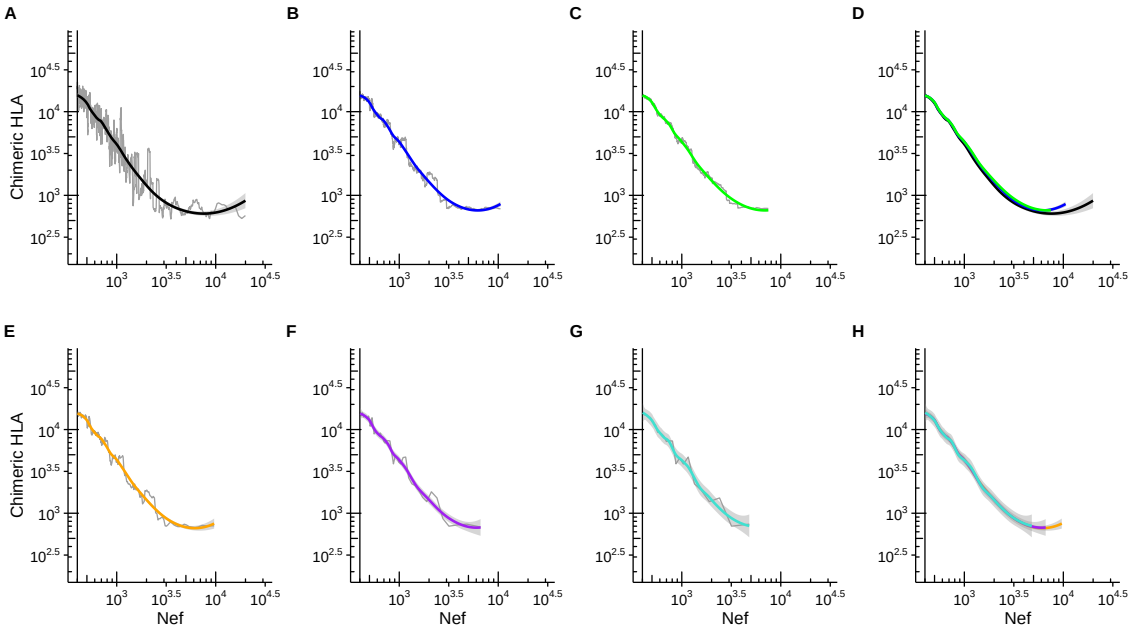


Figure 5.10: Sliding windows can be robust ways of identifying trends in datasets of this size and type. Increasing the window length from **A)** 20, to **B)** 100, to **C)** 200 with a fixed step size of 2 did not alter the profile of the curve fitted to these data when overlaid **(D)**. Increasing the step size from **E)** 10, to **F)** 50, to **G)** 100 with a fixed window length of 100 did not alter the profile of the curve either **(H)**. Example uses data from chimeric HLA-Aγ-expressing cells transfected with consensus HIV-2 Nef, but pattern consistently observed across samples.

As direct comparison of different Nef isolates and chimeric HLA types is a key component of this study, a modified "staggered window" approach was used. Here, instead of iterating across the data by defined windows and steps, the data were instead assigned to bins of fixed width and position, each containing cells expressing Nef at levels within a given range.

5.2.3.1 Optimisation of Bin Parameters

To ensure that bins are assigned objectively and by fixed rules, a simple program was written to determine the position and widths using a randomly selected training subset of the data. The program loops across each dataset and performs the following:

1. Segregates true negative cells from Nef positive cells;
2. Determines the maximum expression level (ME) of Nef;
3. Calculates an effective maximum expression level (eME) of ME-15%;
4. Selects twenty five points evenly spaced on a $\log_{10}$ scale between the negative cells and the eME;
5. Constructs bins around each of these points containing cells that fall within $\pm$ 8% of the value of the bin's midpoint.

The eME rather than the ME was used as the centre of the uppermost bin because any bin centred on the ME by definition contains no cells above its midpoint. A bin width of $\pm$ 8% of the midpoint was selected as this value generates bins that do not overlap but do increase in width as the expression of Nef increases. This is important as there are always fewer cells expressing high levels of Nef due to the drop-off in transfection efficiency and these bins would otherwise contain too few cells from which to extract a reliable estimate of the mean.

The bin positions and widths were then compared across the samples, and further rounds of optimisation were performed. In order to improve resolution at lower levels of expression where the greatest differences in downregulation were being observed, the process was repeated, capping the eME at increasingly lower levels until a suitable bin profile was achieved across all samples. The final bin profile contained 21 bins with the upper bin midpoint capped at 20,000 AFU. The values used are listed in appendix table A.6.

To verify that each bin contained sufficient cells from which to obtain a reliable mean to be used for the calculation of HLA-I downregulation, the number of cells within each bin using this profile was plotted for the five selected datasets (figure 5.11). If the bin contained fewer than 0.24% of the Nef positive cells, it was excluded. The cut

off of 0.24% was selected as with 21 bins, and all being equal, each bin may contain no more than 4.76% of the total population of positive cells.

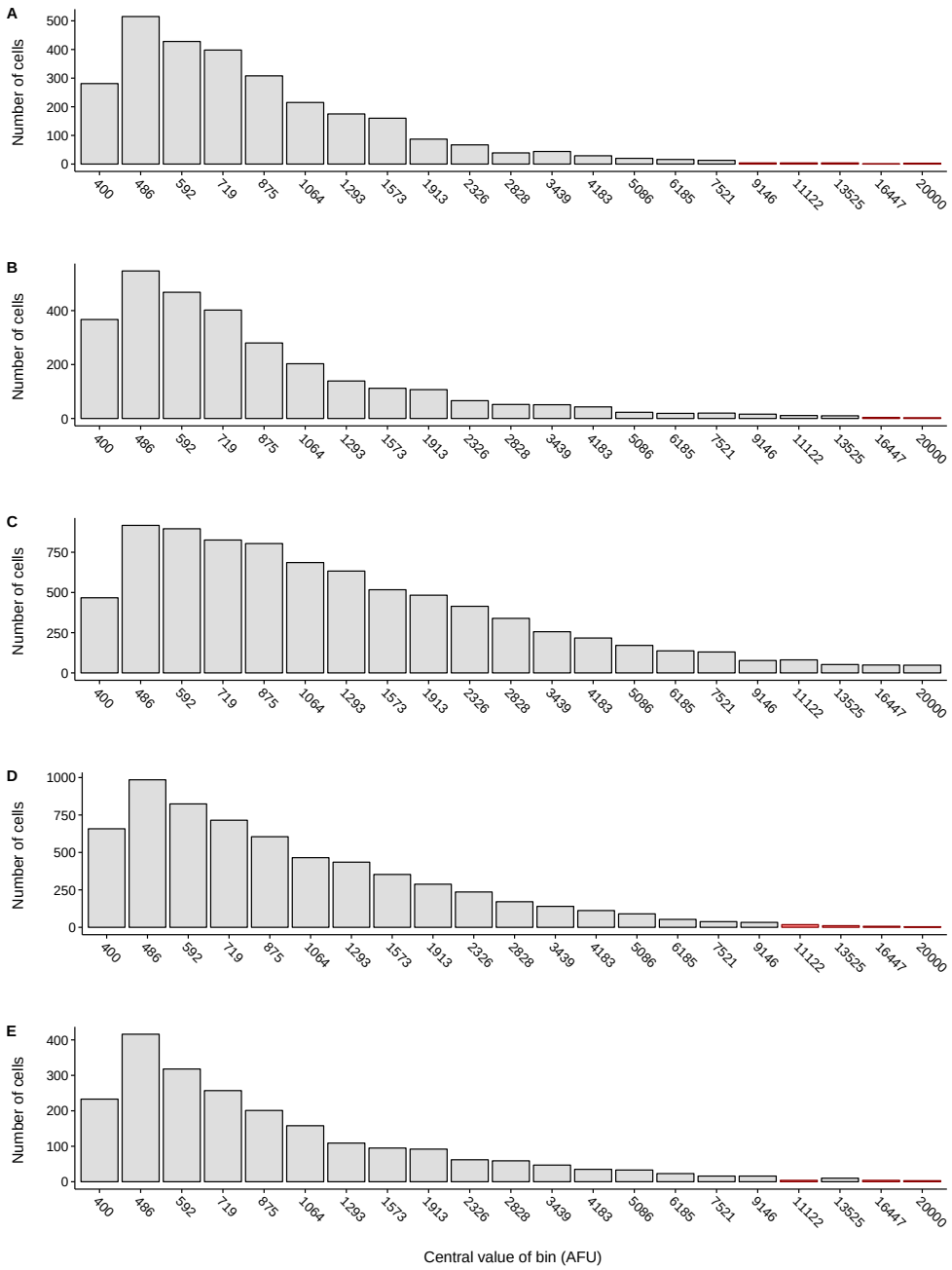


Figure 5.11: Density of SWARM bins. The number of cells found within each bin for the five datasets selected: **A)** SM339 Nef in cells expressing chimeric HLA-Aγ, **B)** TD001 Nef in HLA-Aγ, **C)** SF2 Nef in HLA-Aδ, **D)** Consensus HIV-1 Nef in HLA-Aη and **E)** Consensus HIV-2 Nef in HLA-Aγ. Bins containing fewer than 0.24% of the Nef positive population are marked red for exclusion.

Bins containing cells that express low levels of Nef are expected to be larger owing to transfection efficiency, and so the higher bins will necessarily contain a smaller proportion of cells. The lowest acceptable number of cells in a bin was defined by the commonly used cut off of 5%, and 5% of 4.76% is 0.24%. As a result of the drop-off across bins, a quality-checking step was incorporated into SWARM to exclude insufficiently large bins from plotting and statistical analysis, and henceforth all plots describe quality-checked data.

The bin parameters were finally verified by plotting the mean value within each bin overlaid to the equivalent data from a true sliding window algorithm (figure 5.12) This demonstrates that with a properly selected bin profile, a level of information and accuracy can be generated in a staggered window analysis which is comparable with that of a traditional sliding window approach.

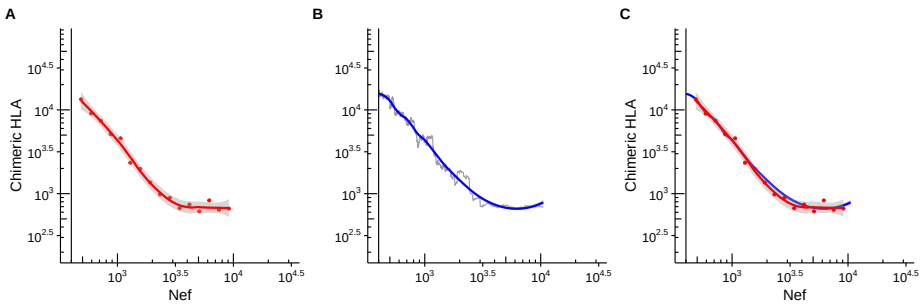


Figure 5.12: Verification of bin parameters. **A)** The bin profile described. Dots represent the MFI of HLA-I within each bin. **B)** A corresponding true sliding window (window length 100, step size 2; figure 5.10B). **C)** The two plots overlaid. Example uses data from chimeric HLA-Aγ-expressing cells transfected with consensus HIV-2 Nef.

5.2.4 Curve Fitting

Applying a best fit line to the raw flow cytometric data and the percentage of down-regulation permits visual comparison between the response of a particular chimeric HLA-I type to different Nef isolates.

As the relationship was neither linear nor exponential (HLA-I downregulation appeared to plateau faster in some samples than others) a global function could not be applied. Instead, a non-parametric local regression (LOESS) curve was fit [500]. LOESS builds a curve to the data by establishing a fit at x, using data points in the vicinity of x, weighted by their distance from x. The smoothness of the curve is determined by the span of the area surrounding x which is used for curve fitting. The span - or smoothing parameter f - should be selected on the basis of the experimental data and should be large enough to minimise noise in the data, whilst not so large as to mask patterns in the data. Through successive trials of a range of f values, $f = 0.6$ generated the closest fitting line to the data (figure 5.13C).

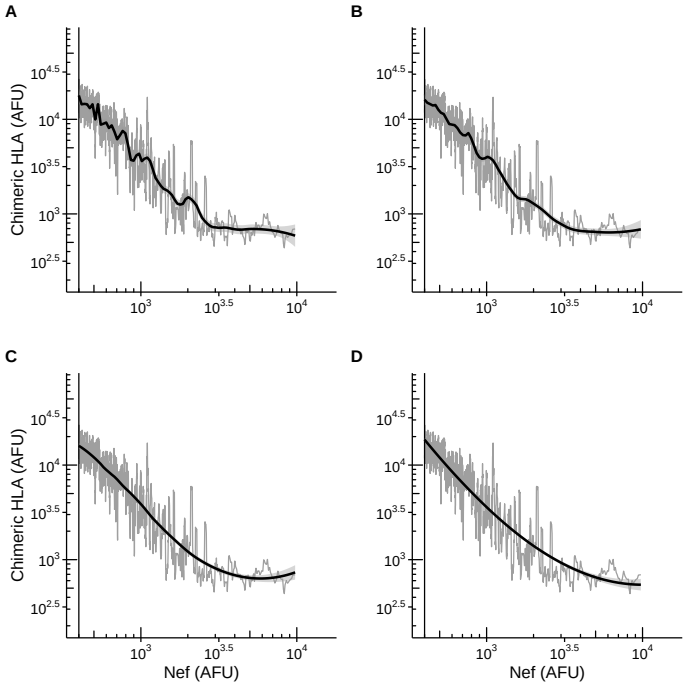


Figure 5.13: Optimisation of SWARM best fit line. $f = 0.6$ was identified from a range of values as generating the most optimal best fit lines for both the raw data and the fold downregulation values. Plots show representative LOESS curves of raw flow cytometric data from the transfection of cell lines expressing chimeric HLA-Aγ with consensus HIV-2 Nef. The smoothing parameter f was set to **A)** 0.1, **B)** 0.2, **C)**, 0.6, **D)** 2.0.

5.2.5 Investigating HLA-I Downregulation with SWARM

Having optimised SWARM, this technique was used to perform high resolution analysis of HIV-1 and HIV-2 Nef-mediated HLA-I downregulation. Six Nef isolates (section 2.3.12) and the full panel of transgenic cell lines expressing chimeric HLA-A and HLA-B were used.

5.2.5.1 HLA-A and HLA-B are not Differentially Downregulated

Further to the findings of chapter 4, no consistent difference was observed in the downregulation of HLA-A and HLA-B molecules by any of the Nef isolates tested (figure 5.14). Indeed, HLA-Bα (purple) and HLA-Bβ (pink) molecules appeared to fall intermediate to the various chimeric HLA-A types in all cases, but the differences between the chimeric HLA-I types were small overall.

HLA-Aα (red) or HLA-Aη (blue) were the most efficiently downregulated molecules by all viral isolates. HLA-Aα was most efficiently downregulation by SF2, SM339 and TD001 Nef, whilst HLA-Aη was most efficiently downregulated by SM046. This hierarchy of HLA-I downregulation was consistent with that detected using fold change calculations (figures 4.7 & 4.10). Differences between the downregulation of HLA molecules were less clear in the context of consensus Nef using both SWARM and fold change (figure 4.12).

5.2.5.2 HIV-2 Nef Downregulates HLA-I More Efficiently than HIV-1

Whilst the extent of HLA-I downregulation mediated by a particular Nef isolate showed limited variability, greater differences could be seen between viral isolates (figure 5.15). Considering the semi-log scale, these differences were substantial. Consensus HIV-2 Nef (yellow) most efficiently downregulated all HLA-I types, followed by either TD001 Nef (pink) or consensus HIV-1 Nef (red). These data suggest that

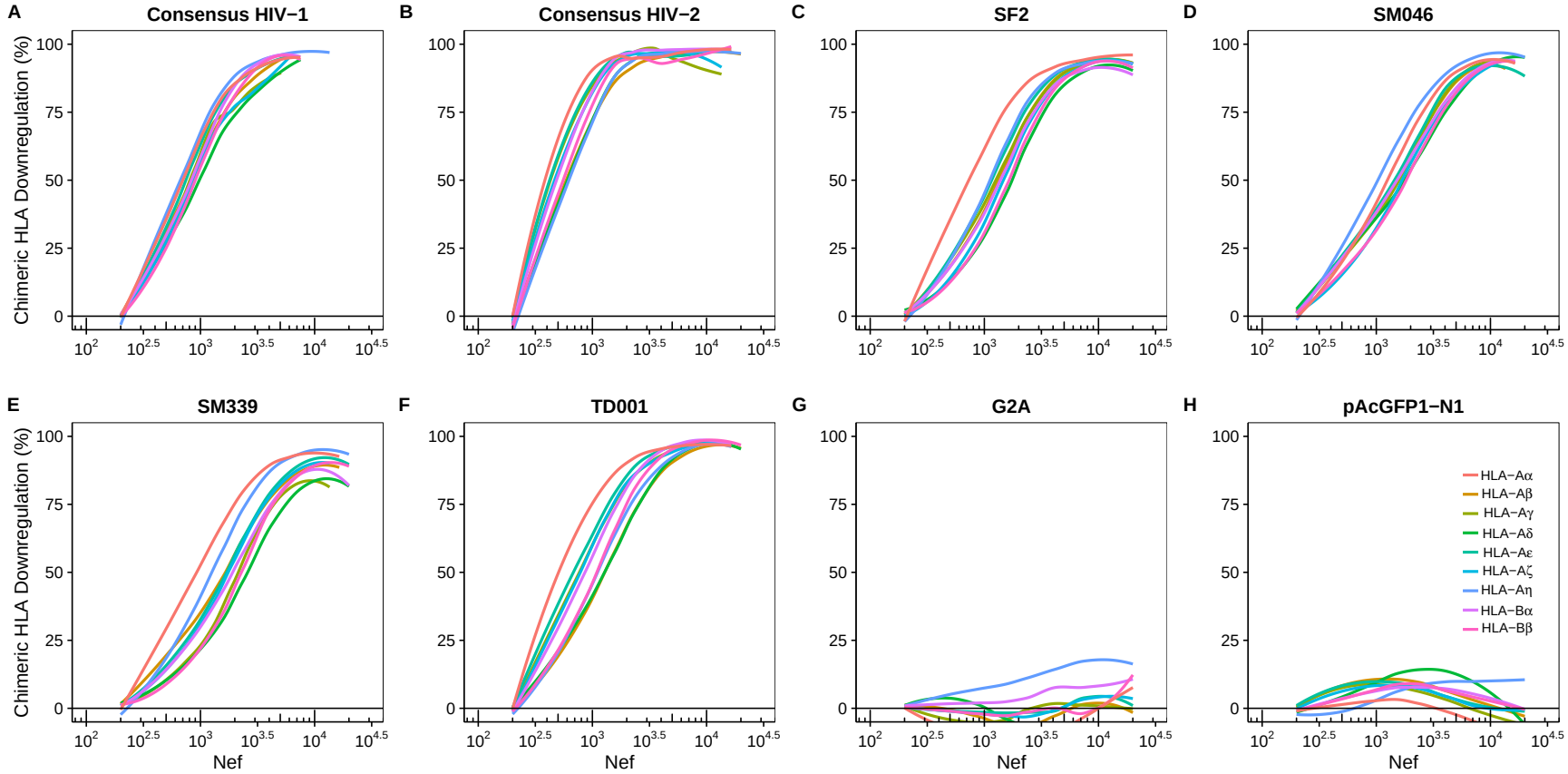


Figure 5.14: SWARM analysis of Nef-mediated HLA-I downregulation grouped by Nef isolate, showing the percentage of each chimeric HLA-I molecule internalised with increasing amounts of Nef (AFU). G2A and pAcGFP1-N1 included as controls.

HIV-2 Nef is more efficient than HIV-1 Nef at downregulating HLA-I. Consistent with chapter 4, these data also demonstrate that the primary Nef isolates are less efficient at downregulating HLA-I than their corresponding viral type consensus. For most HLA-I types, SF2 Nef (green) sat between the consensus HIV-1 Nef and the primary patient isolates (turquoise and blue).

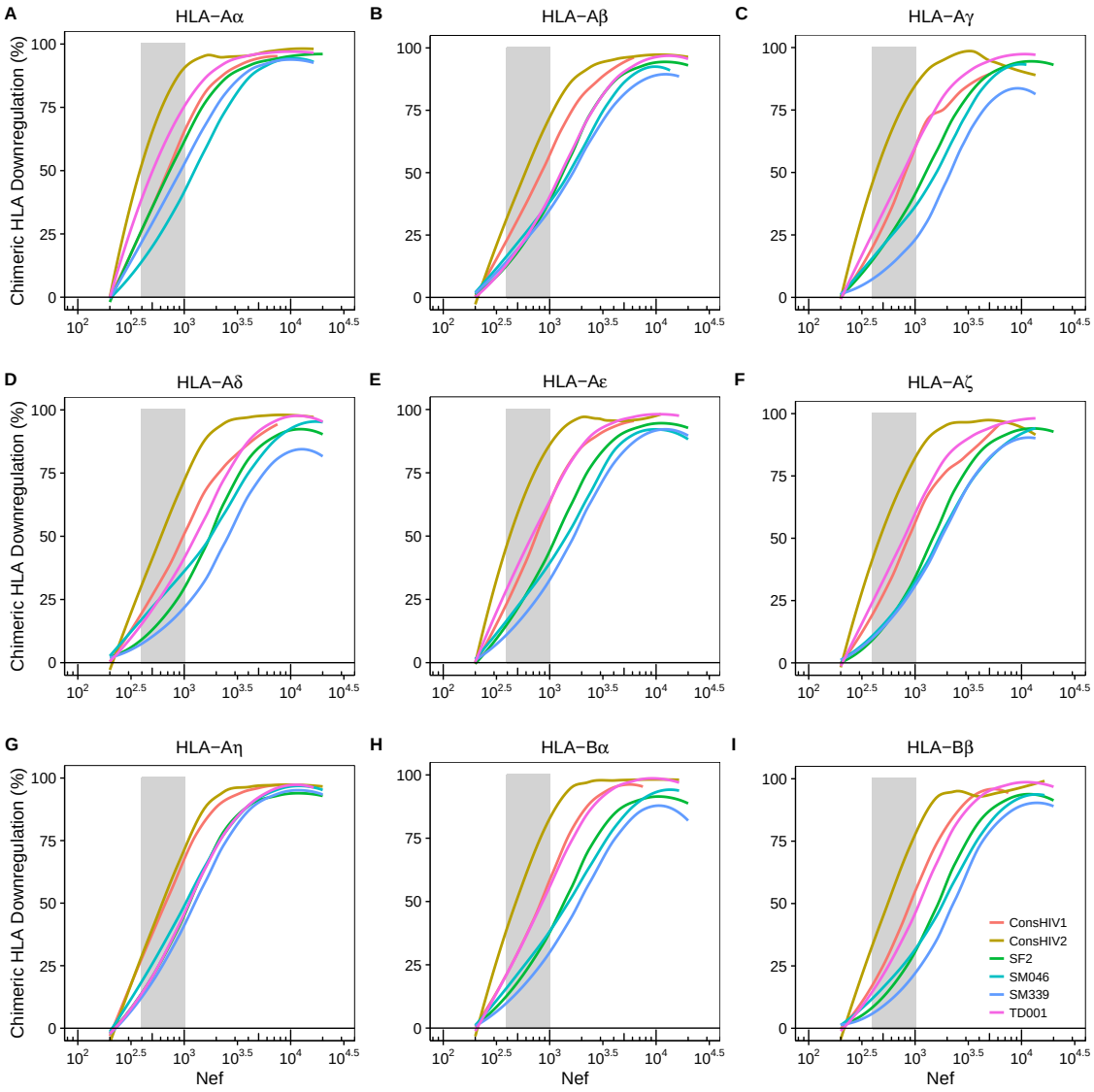


Figure 5.15: SWARM analysis of Nef-mediated HLA-I downregulation grouped by HLA-I type, showing the percentage of each chimeric HLA-I molecule internalised with increasing amounts of Nef (AFU). The shaded area defines the linear range from which the slopes were calculated.

To assess the differences quantitatively, the slope was calculated of the raw data before downregulation plateaued at high concentrations of Nef. As the plateau is reached in some isolates faster than others, the data were sliced between 400 to 1000 AFU of Nef which encompassed only Nef-expressing cells within the linear range of HLA-I downregulation on a $\log_{10}$ scale (shaded grey in figure 5.15). The slopes (absolute) were then plotted, grouped by Nef isolate (figure 5.16).

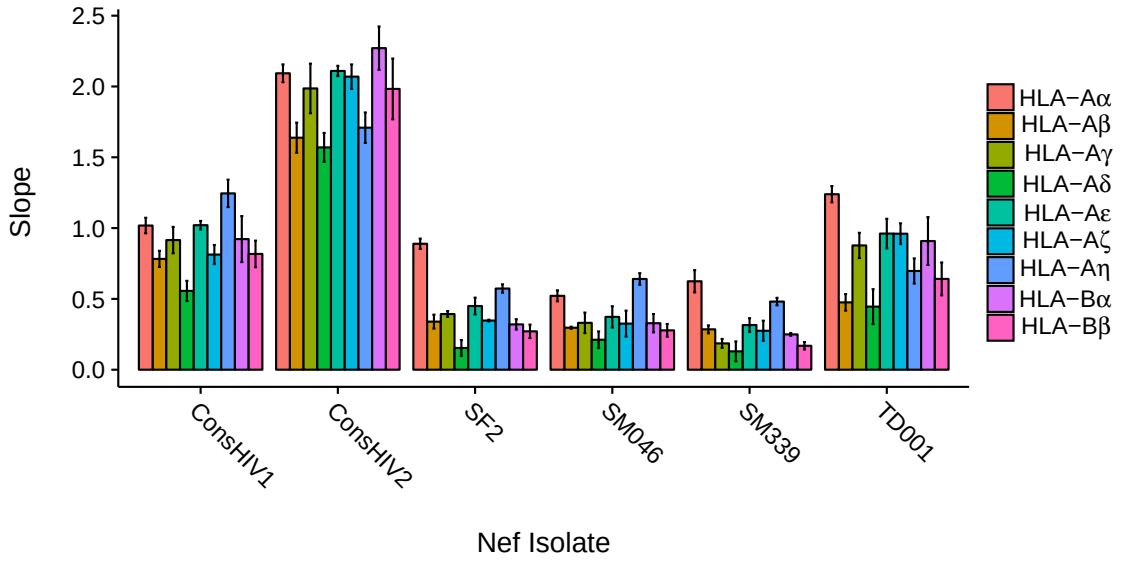


Figure 5.16: The absolute slopes of HLA-I downregulation mediated by each Nef isolate. Colour denotes HLA-I type; error bars are ± 1 SD.

A two-way ANOVA was conducted that showed there was a statistically significant interaction between the effects of the Nef isolate and HLA-I type on the slope ($F(40,108) = 8.085$, $p < 0.0001$). Pairwise t-tests were then performed with Bonferroni correction for multiple comparisons. The slopes of downregulation mediated by consensus HIV-2 Nef were by far the steepest (μ = -1.94), and the difference between these slopes and each other slope by every Nef isolate was statistically significant (all $p < 0.0001$). The mean slope of downregulation mediated by consensus HIV-1 Nef was also significantly steeper than that of each of the primary HIV-1 isolates (all $p < 0.0001$).

Whilst the bifurcate fold change calculation failed to identify a difference between consensus HIV-1 and HIV-2 Nef (chapter 4), the SWARM analysis has revealed that when the amount of Nef is taken into account, HIV-2 Nef is significantly more efficient at downregulating HLA-I than is HIV-1 Nef. This was observed with both the consensus variant and primary patient isolates.

5.2.5.3 Intermediate Nef Concentrations Exert Functional Differences

Although differences can be detected in the efficiency of HLA-I downregulation by particular Nef isolates, these differences converge when high levels of Nef are expressed in the cell. Here, the data plateaus somewhere around 95% HLA-I downregulation (figure 5.15). Therefore, it was of interest where the biggest differences in downregulation between Nef isolates and between HLA-I types were occurring.

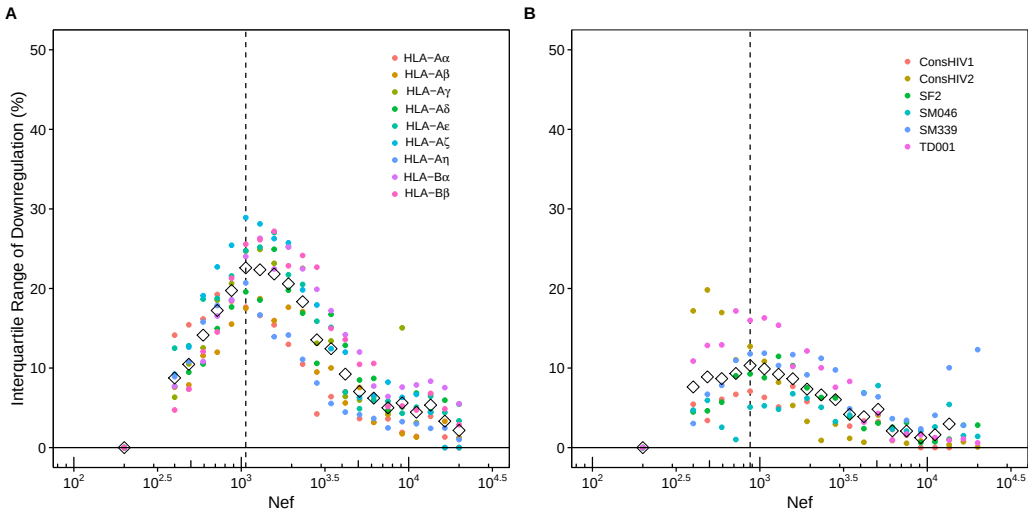


Figure 5.17: Interquartile range of HLA-I downregulation by bin. A) IQR of Nef isolates grouped by HLA-I type, and **B)** IQR of HLA-I types grouped by Nef isolate. Data for each subset are distinguishable by colour. The mean IQR for each bin within each plot is denoted by a diamond, and the bin containing the largest IQR is indicated by the dashed line.

To test this, the SWARM outputs were analysed grouped by either Nef isolate, or by HLA-I type, and the interquartile range (IQR) for the values within each bin were

calculated. The IQRs were then plotted (figure 5.17), the mean taken, and the bin containing the largest mean IQR was identified.

For the data subset by HLA-I type, bin 7 (centred on 1064 AFU) contained the largest IQR (figure 5.17A), and for the data subset by Nef isolate, bin 6 (centred on 875 AFU) contained the largest IQR (figure 5.17B). Notably, the IQR for both subsets increased towards intermediate Nef expression levels, before converging again at high levels of Nef. The variability was subtantially higher between Nef isolates than between HLA-I types.

5.2.5.4 Surface Retention of HLA-I Across Bins

The amount of HLA-I remaining on the surface of the cells was plotted for the Nef-negative population (figure 5.18A), cells containing intermediate levels of Nef (bin 7 - centred on 1064 AFU; figure 5.18B) and cells containing high levels of Nef (bin 17 - centred on 7521 AFU; figure 5.18C).

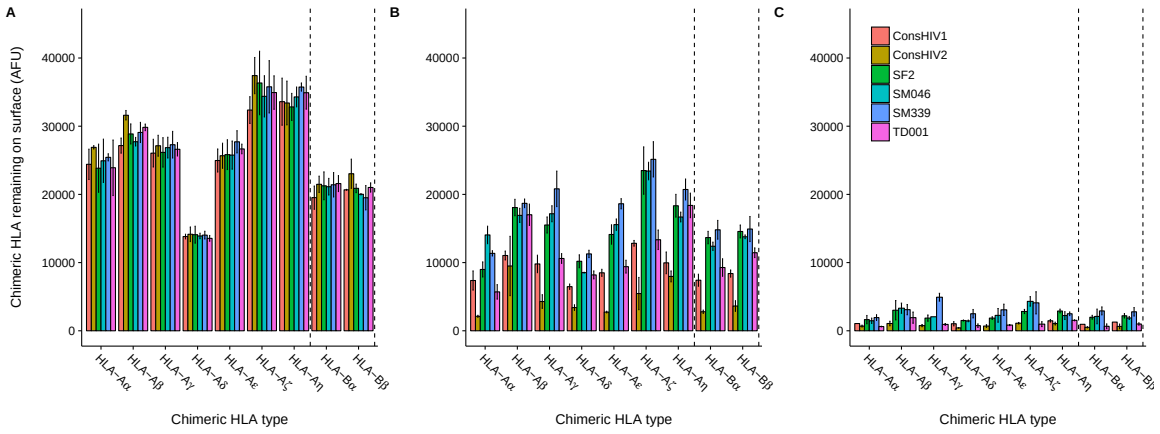


Figure 5.18: Surface expression of chimeric HLA-I molecules in A) Nef negative cells, **B)** cells expressing intermediate levels of Nef, and **C)** cells expressing high levels of Nef. Colour denotes Nef isolate; error bars are ± 1 SD.

In the untransfected populations, there are clear differences in the resting levels of HLA-I on the surface of each transgenic cell line. These differences are consistent with those reported in figure 4.2, and the differences between the majority of molecules are highly statistically significant when compared by pairwise t tests with Bonferroni correction for multiple comparisons (table 5.1a). However, on transfection with Nef, the differences in expression between the HLA-I molecules diminish in cells expressing both intermediate (table 5.1b) and high levels of Nef (table 5.1c). There are no statistically significant differences between any of the molecules in these bins.

Table 5.1 Comparison of surface expression of chimeric HLA-I molecules following transfection with Nef, expressed at intermediate and high levels. Numbers denote *p* value from pairwise *t* test comparing each chimeric HLA-I type, with Bonferroni correction for multiple comparisons.

(a) Nef Negative Population

	Aα	Aβ	Aγ	Aδ	Aϵ	Aζ	Aη	Bα
HLA-Aβ	<0.0001	-	-	-	-	-	-	-
HLA-Aγ	0.31	0.03	-	-	-	-	-	-
HLA-Aδ	<0.0001	<0.0001	<0.0001	-	-	-	-	-
HLA-Aϵ	1	0.002	1	<0.0001	-	-	-	-
HLA-Aζ	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	-	-	-
HLA-Aη	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	1	-	-
HLA-Bα	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	-
HLA-Bβ	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	1

(b) Intermediate Nef Expression (Bin 7, centred on 1064 AFU)

	Aα	Aβ	Aγ	Aδ	Aϵ	Aζ	Aη	Bα
HLA-Aβ	0.86	-	-	-	-	-	-	-
HLA-Aγ	1	1	-	-	-	-	-	-
HLA-Aδ	1	0.70	1	-	-	-	-	-
HLA-Aϵ	1	1	1	1	-	-	-	-
HLA-Aζ	0.14	1	1	0.11	1	-	-	-
HLA-Aη	0.77	1	1	0.62	1	1	-	-
HLA-Bα	1	1	1	1	1	0.68	1	-
HLA-Bβ	1	1	1	1	1	1	1	1

(c) High Nef Expression (Bin 17, centred on 7521 AFU)

	Aα	Aβ	Aγ	Aδ	Aϵ	Aζ	Aη	Bα
HLA-Aβ	1	-	-	-	-	-	-	-
HLA-Aγ	1	1	-	-	-	-	-	-
HLA-Aδ	1	1	1	-	-	-	-	-
HLA-Aϵ	1	1	1	1	-	-	-	-
HLA-Aζ	0.96	1	1	1	1	-	-	-
HLA-Aη	1	1	1	1	1	1	-	-
HLA-Bα	1	1	1	1	1	1	1	-
HLA-Bβ	1	1	1	1	1	1	1	1

5.3 Discussion

The work described in this chapter arose out of a need to address fundamental issues with the way that flow cytometric data are analysed. As has been shown, several limitations are associated with bifurcate analyses, which are arguably the most widely used way of dealing with these kinds of datasets. Such limitations include infrequently satisfied assumptions about the data, and a lack of robustness against common experimental variation.

The SWARM analysis attempts to address these limitations by categorising data not into binary groups - "positive" and "negative" - but into a series of bins selected with minimal decision-making on the part of the investigator. Bins are instead selected primarily computationally, taking into account the underlying distribution of the data.

The use of bins permits direct comparison of cells expressing equivalent levels of fluorophore X, and also allows slopes to be calculated, providing a measure of the strength of a relationship between molecules labelled by fluorophore X and fluorophore Y. Whilst bifurcate analyses may be able to identify the presence of such a relationship, inferring the degree of association is much less straight-forward because of experimental variation that is difficult to control for.

Bins help to overcome many of the issues related to experimental variation. For example, in a large experiment where manual handling results in samples reaching the incubator at staggered intervals, then some samples may contain a greater proportion of cells expressing high levels of fluorophore X. If the samples were to be analysed using bifurcate analyses, then differences may be seen that are actually artefacts of the incubation times. However, were these samples to be analysed using SWARM,

the uppermost bin could be centred on the effective maximum expression level of the cells which were incubated for a shorter duration. This would allow direct comparison of the samples whilst controlling for the experimental variation.

Having developed and optimised SWARM, this technique was used to perform a high-resolution analysis of Nef-mediated HLA-I downregulation.

In agreement with the findings of chapter 4, SWARM analysis also revealed no evidence of differential downregulation of HLA-A and HLA-B mediated by any of the isolates of Nef: downregulation of the HLA-B molecules appeared to fall intermediate to the various HLA-A molecules across the full spectrum of Nef expression. Also in agreement with chapter 4, particular Nef isolates downregulated HLA-Aα or HLA-Aη most efficiently. As this pattern has been detected by two different analyses using independently collected datasets, the difference is likely biological rather than an artefact of either method of analysis. However, limited variation in downregulation is seen between the various HLA-I types overall.

Conversely, large differences were detected between the different Nef isolates. As demonstrated in chapter 4, consensus HIV-1 and HIV-2 Nef proteins consistently downregulated HLA-I more efficiently than did the corresponding primary isolates. However, in contrast to the previous bifurcate analysis, SWARM also revealed that when the amount of Nef is controlled for, HIV-2 Nef isolates are significantly more efficient at downregulating HLA-I than are HIV-1 Nef isolates. The marked and consistent difference in efficiency of HIV-1 and HIV-2 Nef was unexpected. As a smaller proportion of HIV-2 patients progress to AIDS than do HIV-1 patients (section 1.1.5.5), it would be more intuitive if HIV-2 Nef were less efficient at down-

regulating HLA-I than HIV-1 Nef, as this would correspond to a more efficient CTL response and a greater degree of viral control.

Although these findings were consistent across all HLA types and the differences were highly statistically significant, a limitation to the interpretation of these data is that the number of Nef isolates tested was small (four HIV-1 Nef isolates, two HIV-2 Nef isolates). Performing the SWARM analysis on a wider range of Nef isolates would allow broader interpretation of these data.

The SWARM analysis also showed that whilst large differences could be seen in the rate of HLA-I downregulation mediated by different Nef isolates expressed at intermediate levels, in cells expressing high levels of Nef, downregulation appears to plateau at approximately 95%. These data have important implications as they suggest that although some Nef isolates may be more efficient at downregulating HLA-I than others, the functional outcome is dependent on the amount of Nef in the cell. With very high levels of expression, the differences between the isolates diminish to the point where all isolates behave similarly.

This observation may explain why measures of fold downregulation often do not correlate with markers of disease progression [490, 496] (also, see section 4.2.3.4). The effect observed *in vitro* may not be applicable at physiologically relevant levels of Nef. Performing a SWARM analysis on data collected from virally infected cells - as opposed to transfected cells - would provide more insight by demonstrating the precise amount of Nef expressed in true infection. However, this is not straightforward. Generating a live virus that endogenously expresses fluorescently tagged Nef is problematic for a number of reasons. As Nef contains an N-terminal myristorylation motif, adhering the fluorescent tag to the N-terminus would again affect the ability

of the protein to attach to membranes and perform associated functions. However, as the *nef* gene also overlaps with the LTR at the 3' end of the genome, attaching the fluorophore to Nef C-terminally could affect the ability of the virus to replicate properly, and may result in the fluorophore being disrupted or excised completely during reverse transcription (section 1.1.6.3).

Whilst a transgenic virus could be generated in which Nef and a fluorophore are co-expressed by the same promoter under the control of an IRES, such a system would not express both proteins at equivalent levels [498], and the fluorophore would therefore not provide an accurate read-out of physiological levels of Nef expression. Conversely, an unlabelled virus could be used to infect the cells, with Nef then being stained post-transfection with an appropriate antibody. However, the suitability of the antibody for staining a range of HIV-1 and HIV-2 Nef isolates would need to be rigorously verified, and at the time of writing, no such pan-HIV Nef antibody is available.

Although it has not been possible to precisely measure physiological Nef expression levels in virally-infected cells within the scope of the current project, this should remain a key priority for future study. It is worth noting that Nef is expressed from early in infection [501], and can even be detected extracellularly in patient plasma at levels as high as $10\,\mathrm{ng\,ml^{-1}}$ [502] in both viraemic and non-viraemic patients on ART [503]. This suggests that expression levels may be high at least in the context of HIV-1 infection. However, as CTL responses can still be detected against HIV epitopes, almost complete downregulation of HLA-I seems unlikely in primary infected cells: particularly as responses are sufficiently effective to drive CTL escape mutations.

Consequently, the susceptibility of cells to CTL-mediated cytotoxicity should also be considered. As recognition of viral peptide-bound HLA-I triggers killing by CTLs, the absolute level of HLA-I remaining on the surface of the cell may arguably be more important than the proportional extent of its downregulation. Indeed, HLA-I expression has been shown to correlate with recognition of melanoma cells by CTLs [504], and more recently, the role of HLA-C expression levels in the control of HIV has been demonstrated [493]. This is potentially of great importance as the expression levels of some HLA-I types differ substantially [492, 505].

The expression levels of the chimeric HLA-I molecules in the stable transgenic cell lines also differed extensively, which may or may not be reflective of the physiological levels of these families *in vivo*. Lending support to this, HLA-C was always expressed at lower levels than HLA-A and HLA-B (figure 4.2), consistent with previous reports [492, 505]. However, the highly artificial nature of the system makes generalisation about the expression levels of different HLA-I families risky.

Keeping in mind this caveat, the relative amount of HLA-I remaining on the surface of each transgenic cell line following transfection with Nef was considered as a crude surrogate for the relative visibility and susceptibility of these cells to CTL killing. In untransfected cells, the expression levels of the different chimeric HLA-I types matched the distribution seen in section 4.2.2.1. However, the variation diminished to the point that no significant differences were seen when intermediate or high levels of Nef were present in the cell. This may indicate that with intracellular Nef, the cells expressing different HLA-I types are equally susceptible to recognition and killing by CTLs. However, supporting data from functional killing assays would be required to test this hypothesis.

Whilst the SWARM analysis was conceived and developed to address Nef-mediated HLA-I downregulation, this technique could be applied to different datasets to address other research questions. For example, SWARM could putatively be used for high-resolution assessment of other virus-host relationships (e.g. Ebola glycoprotein-mediated downregulation of $\beta 1$ integrins [506]), or further afield (e.g. in assessing relationships between hormones and the expression of their associated receptors in health and disease [507]). SWARM may therefore be of use in other scientific fields such as cancer research.

There are few restrictions to the use of SWARM, assuming that the molecules of interest can be appropriately tagged or otherwise labelled. Similar to bifurcate analyses, flow cytometric data to be analysed should be collected from the same instrument - which must be well-calibrated - and should ideally be collected in one session to minimise confounding variables. If more than one cell line is to be used, they should also be rigorously tested to ensure that the degree of autofluorescence emitted under experimental conditions is equivalent. Fundamentally, the results of any analysis will only be as good as the data that are used to generate them, and care should be taken to ensure that appropriate controls, conditions, and parameters are in place throughout the course of the experiment.

To summarise, in addition to developing a novel technique for the analysis of flow cytometric data, this chapter has shown that substantial differences may exist between the ability of HIV-1 and HIV-2 Nef to downregulate HLA-I. The next chapter will continue to investigate differences between HIV-1 and HIV-2 Nef by considering their relative antigenicity.

Chapter 6

Proteasomal Processing and the Antigenicity of HIV-2 Nef

6.1 Introduction

HIV-1 Nef is a highly immunogenic viral protein that is heavily targeted by CTLs from acute infection onwards. Indeed, it is one of the key targets of the early specific T-cellular response [508–511]. The substantial immunogenicity of HIV-1 Nef has led to intense interest in the field of HIV vaccinology. However, Nef-containing vaccine candidates have performed disappointingly - for example Gray *et al.* [512] - which may be related to ineffectual targeting of full-length protein by CTLs resulting in rapid escape mutations in immunodominant yet variable regions [513]. However, there is also evidence that Nef-specific CTL responses to more conserved epitopes may contribute at least in part to viral suppression [514].

Whilst HIV-1 Nef is highly immunogenic, responses against HIV-2 Nef are comparably rare. Despite equivalent frequency and magnitude of HIV-specific T-cell responses in HIV-1 and HIV-2 infection, responses specifically against Nef are detected in around 2.5-fold fewer HIV-2 patients [52, 515]. It is unusual for an expressed viral protein not to be a target of the T-cellular response, and in the few examples where this occurs there is a clear mechanism through which interference of antigen presentation

is achieved: for a review of these mechanisms see Yewdell & Hill [516]. A key example of this phenomenon is the Epstein-Barr viral protein Epstein-Barr nuclear antigen 1 (EBNA1), which does not induce CTL responses as it contains a characteristic and long (≈200 aa) glycine-alanine rich repeat that interferes with translation and renders it insusceptible to proteasomal processing [517–519]. Proteasomal processing is an important step in antigen preparation and the generation of epitopes which trigger a CTL response.

Why HIV-2 Nef-specific responses are less prevalent than HIV-1 Nef-specific responses is not currently known, but the homologous proteins exhibit a number of differences in both sequence and function which may in principle impact immunogenicity in a multitude of ways. In terms of sequence, consensus HIV-1 and HIV-2 Nef share only 37.4% pairwise identity [1], and so it was of interest whether such differences could interfere with correct proteasomal processing of viral epitopes in a similar manner to that achieved by EBNA1.

Although HIV-2 Nef does not contain any motifs as strikingly unusual as the extended glycine-alanine repeat of EBNA1, more subtle differences have been shown to impact both the quantity and quality of HIV-2 epitopes generated via proteasomal processing. The presence of flanking prolines surrounding the Gag CTL epitope DA9 - which are themselves significantly associated with low viral load *in vivo* [520] - corresponds to both the efficiency of proteasomal processing *in vitro* and the strength of CTL responses generated against the epitope in patients infected with HIV-2 strains containing these prolines [439].

[1] Calculated from consensus sequences generated from alignments constructed of 3,086 HIV-1 subtype B and 356 HIV-2 viral sequences, as described in section 2.3.1.

As described in section 1.2.3, surplus to 26S proteasome, 26S immunoproteasome is induced by inflammatory cytokines such as IFNγ and is considered to improve both the quality and quantity of CTL epitopes produced by altering the cleavage pattern to generate preferred C-termini for HLA-I binding. This chapter reports on the *in vitro* processing of HIV-1 and HIV-2 Nef by both proteasome and immunoproteasome to establish firstly whether HIV-1 and HIV-2 Nef are digested equivalently in terms of both quantity and quality of epitope precursors generated; and secondly whether the digestion products from the two different types of enzymes complex are comparable. The purpose of this study was to demonstrate whether abrogated antigen processing as a result of differential (immuno)proteasomal processing could - in whole or in part - explain the observed difference in CTL response against the two viral Nef proteins.

6.1.1 Aims

The specific aims of this project were to:

- Compare and contrast the (immuno)proteasomal digestion products of consensus HIV-1 and HIV-2 Nef peptides;
- Compare and contrast the products of the digestion with proteasome and immunoproteasome;
- Determine whether the efficiency of (immuno)proteasomal processing is consistent with the observed difference in CTL response against HIV-1 and HIV-2 Nef.

6.2 Results

6.2.1 Mapping of Nef CTL Epitopes

To identify a suitable region of Nef for this study, CTL epitope density was considered. In 2003, Lucchiari-Hartz *et al.* [438] demonstrated that viral CTL epitopes predominantly cluster within hydrophobic regions, and that within HIV-1 Nef these clusters arise primarily towards the centre of the protein. However, in the last thirteen years, many more CTL epitopes have been reported, and so the density of all CTL epitopes listed in the LANL Immunology Database [401] were mapped to confirm whether this finding still holds.

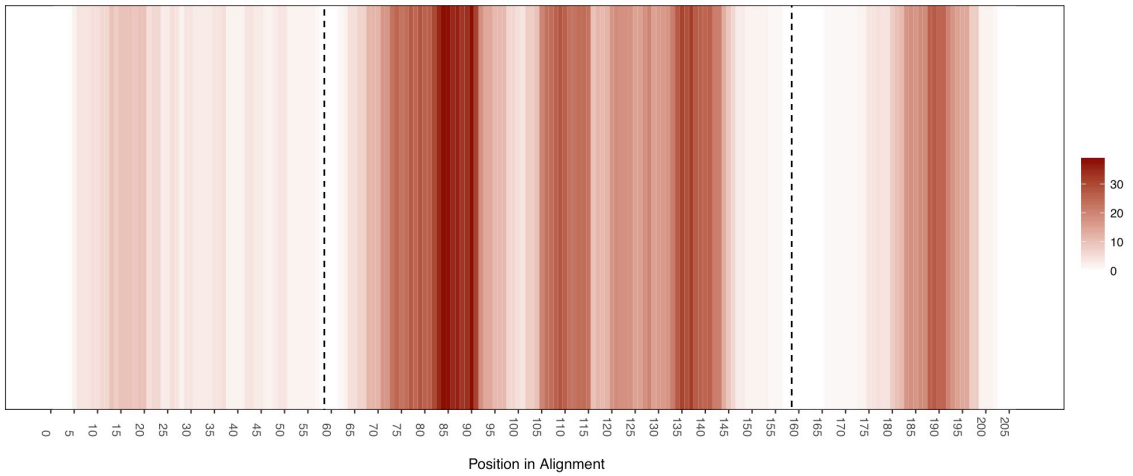


Figure 6.1: Map of the known CTL epitopes of the HIV-1 Nef protein. The location of each epitope is defined by its relative amino acid position in the reference HXB2 sequence. Colour intensity denotes number of epitopes spanning that residue reported in the LANL Immunology Database. The region selected for study is contained between the two vertical dashed lines. Figure constructed from data made available by The Los Alamos National Laboratory (LANL) [401].

The majority of CTL epitopes did indeed cluster towards the centre of the protein (figure 6.1), with a secondary cluster occurring towards the C-terminus. Based on these data, a fragment of 102 amino acids was selected as the region for study - shown

between the two vertical dashed lines - from which four overlapping peptides were generated for both HIV-1 and HIV-2 (table 2.9).

6.2.2 Characterisation of Digestion Products

Each peptide was digested independently with immunoproteasome and proteasome. All digestion products are summarised in table 6.1 and presented individually in figures 6.2 & 6.3.

All eight peptides had generated a series of fragments following a 24 hour incubation with both proteasome and immunoproteasome, but substantial variations were found in the pattern and extent of digestion. With the exception of peptide pair 4, all HIV-2 peptides generated more unique products than did the HIV-1 peptides on digestion with each enzyme complex.

Table 6.1 Summary of the (immuno)proteasomal digestion products generated from HIV-1 and HIV-2 Nef peptides.

	№ Digestion Products				Mean Length (SD)				Product Ratio			
	IP[2]		P[3]		IP		P		HIV-1:2		IP:P	
	HIV-1	HIV-2	HIV-1	HIV-2	HIV-1	HIV-2	HIV-1	HIV-2	IP	P	HIV-1	HIV-2
P1[1]	12	30	21	25	20.3 (6.3)	22.7 (6.7)	19.2 (6.4)	19.3 (5.9)	0.40:1	0.84:1	0.57:1	1.20:1
P2	8	59	8	56	16.3 (6.8)	13.9 (4.4)	13.5 (5.4)	12.8 (4.2)	0.14:1	0.14:1	1.00:1	1.05:1
P3	18	56	20	71	13.6 (5.3)	16.0 (5.9)	10.5 (4.8)	15.4 (5.6)	0.32:1	0.28:1	0.90:1	0.79:1
P4	34	31	30	24	11.6 (5.7)	11.7 (4.2)	11.4 (5.2)	12.4 (3.1)	1.10:1	1.25:1	1.13:1	1.29:1

[1] P1-P4: peptides 1 to 4. [2] IP: immunoproteasome. [3] P: proteasome.

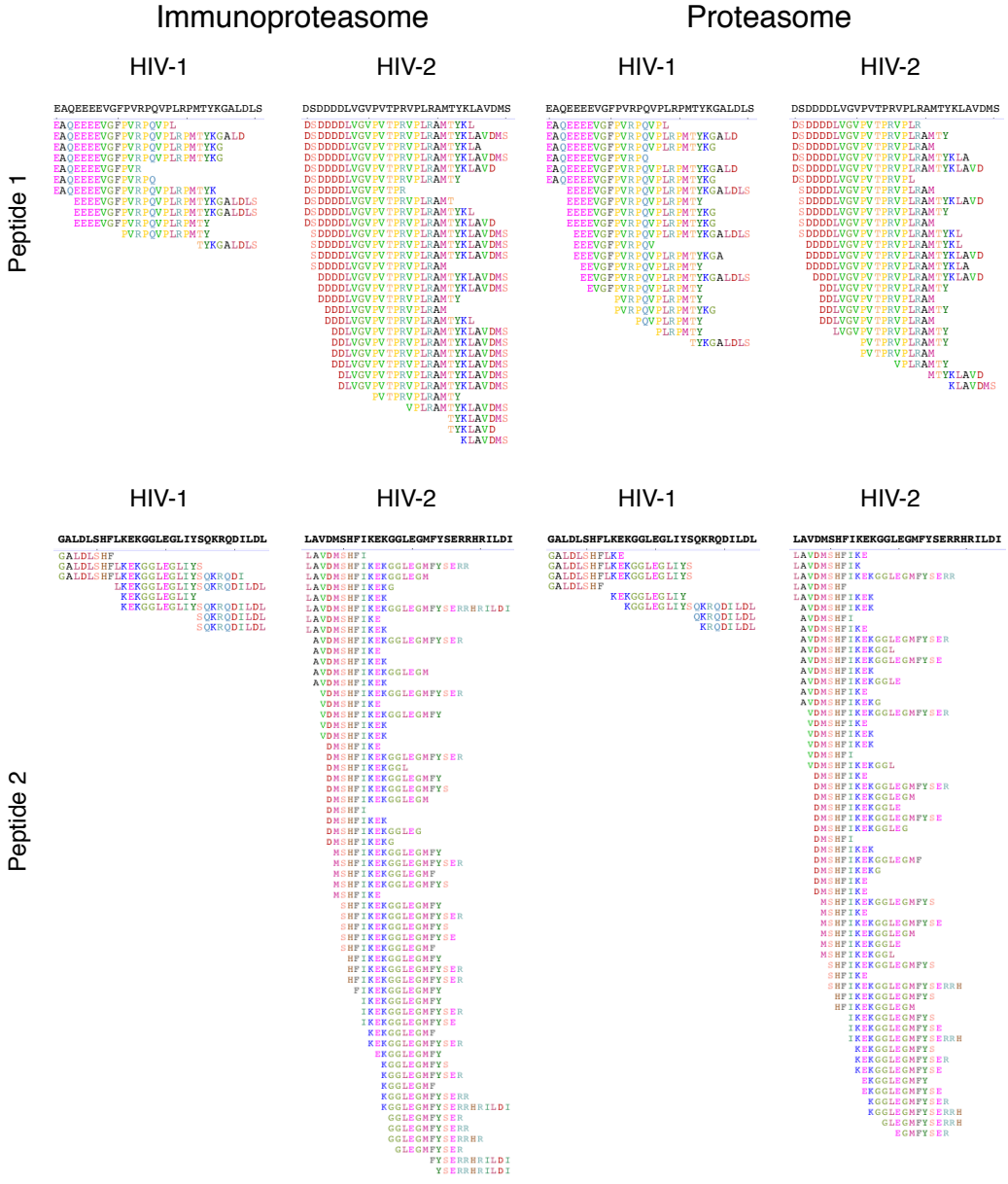


Figure 6.2: Digestion products from peptide pairs 1 and 2 following a 24 hour incubation with immunoproteasome and proteasome, and analysed by liquid chromatography quadrupole time-of-flight mass spectrometry (LC-Q-ToF-MS). The full parent peptide is shown in black, and the digestion products are listed below, coloured by amino acid.

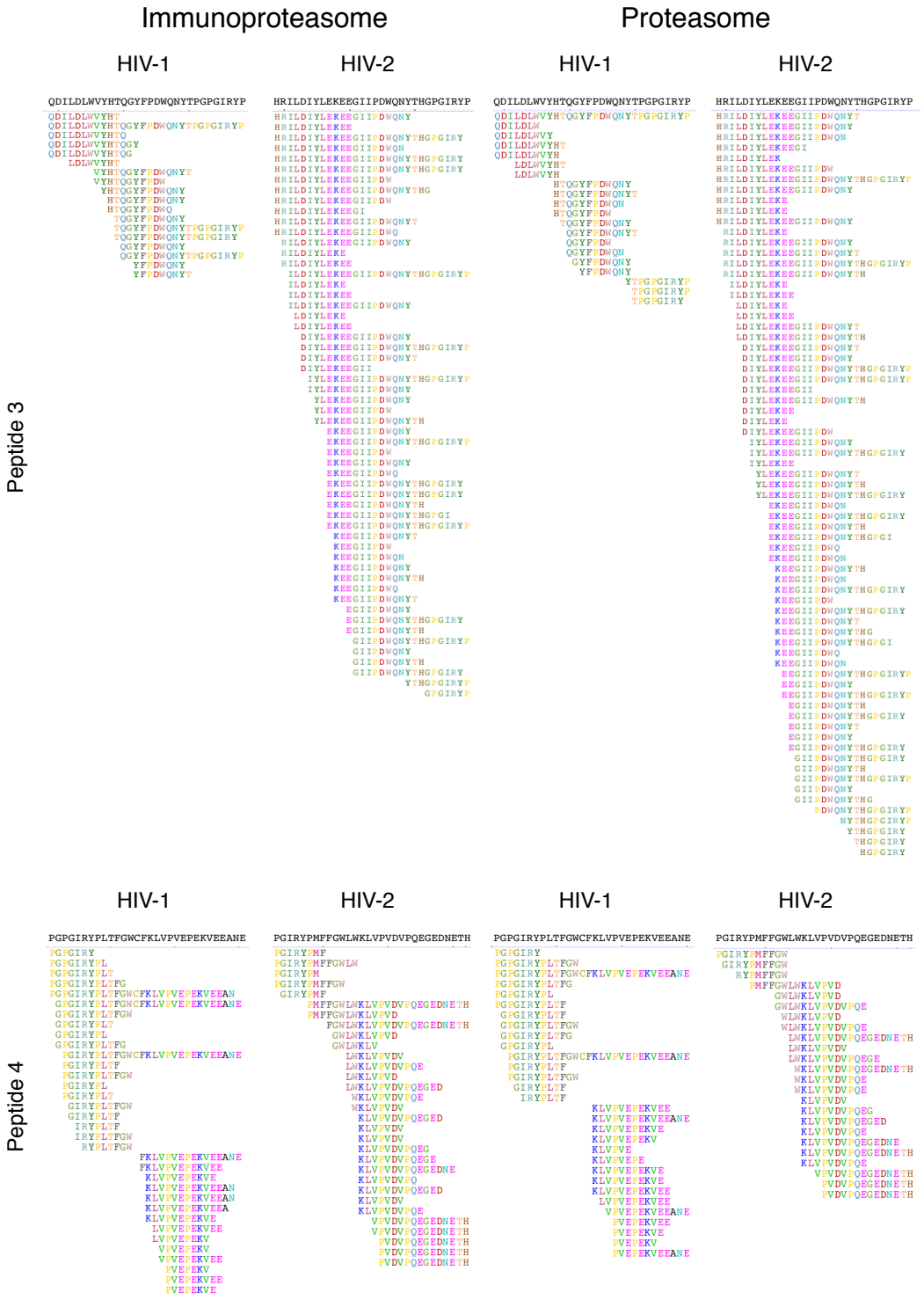


Figure 6.3: Digestion products from peptide pairs 3 & 4 following a 24 hour incubation with immunoproteasome and proteasome, and analysed by liquid chromatography quadrupole time-of-flight mass spectrometry (LC-Q-ToF-MS). The full parent peptide is shown in black, and the digestion products are listed below, coloured by amino acid.

This was most pronounced in peptide 2, where the number of HIV-1 digestion products generated by both immunoproteasome and proteasome was only around 14% of the corresponding HIV-2 peptide digestion. Marked differences were also found in the digestion of peptide 3, where HIV-1 generated around a third of the number of products than did HIV-2. The number of products generated from the digestion of a particular peptide with either immunoproteasome or proteasome were generally more comparable.

In terms of the quality of the digestion products, there were no substantial differences in the lengths of the fragments produced from HIV-1 and HIV-2 (table 6.1), but as can be seen in figures 6.2 & 6.3, the patterns of cleavage appeared quite different. To investigate this further, the frequency of cleavage sites across each position in each digested peptide were plotted, and the data stacked for HIV-1 and HIV-2 (figure 6.4).

In peptides 2 and 3, the number and diversity of cleavage sites in the HIV-2 peptides far exceeded those found in the HIV-1 peptides, by both immunoproteasomal and proteasomal digestion. This was particularly striking in peptide 2 where HIV-1 exhibited cuts in only 5 (immunoproteasome) or 6 (proteasome) positions at clustered sites approximately one third and two thirds along the length of the peptide. Conversely, the corresponding HIV-2 peptide exhibited a large number of dispersed cut sites which were well-distributed across the length of the peptide. The HIV-1 and HIV-2 peptide 1 and 4 pairs exhibited more similar cleavage in terms of pattern and frequency.

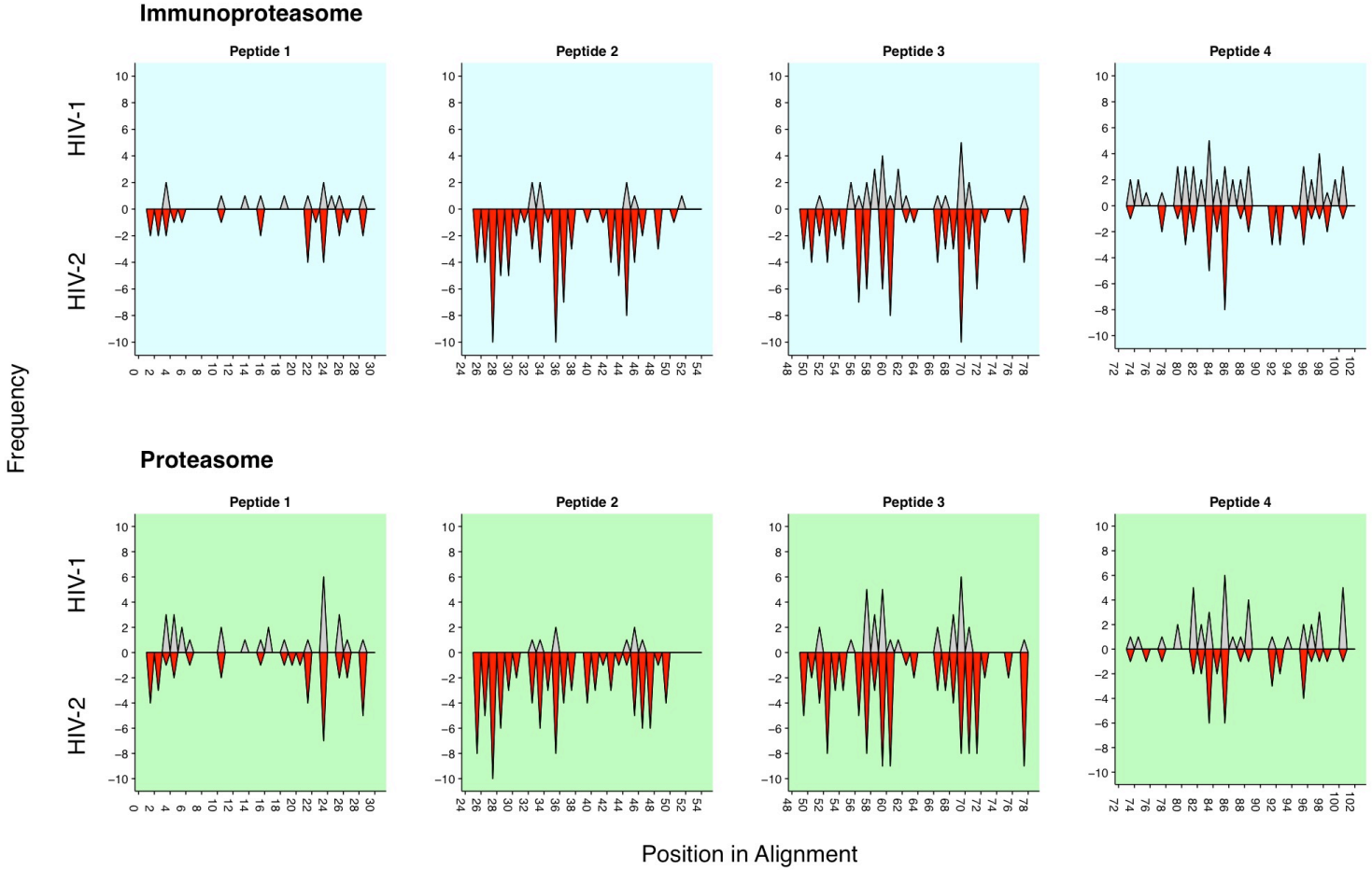


Figure 6.4: Location of cleavage sites in the digested HIV-1 and HIV-2 Nef peptides. Upper panel (blue) shows immunoproteasomal digestion products and lower panel (green) shows proteasomal digestion products. HIV-1 (grey) is shown on the positive axis, with HIV-2 (red) on the negative axis (absolute frequency).

6.2.3 Identification & Analysis of Optimally Cleaved CTL Epitopes

As differences were observed between the HIV-1 and HIV-2 digestion patterns, the next consideration was whether or not the fragments generated correspond to known CTL epitopes reported in the LANL Immunology Database. However, the inherent amino acid mismatches between corresponding HIV-1 and HIV-2 peptide pairs posed a substantial challenge to this analysis, as irrespective of the efficiency of digestion, a "perfect" epitope will never be generated from a peptide which doesn't contain the sequence of interest in the first place. For example, a blanket search for the HLA-B35 restricted VY8 epitope (VPLR**P**MTY) in the digestion products of HIV-2 peptide 1 would be a fruitless task as the parent peptide instead contains VPLR**A**MTY. To overcome this issue, the products from each digestion were binarised (section 2.4.2.1). Binarising allows direct determination as to whether a product has been generated in the digestion of either HIV-1 or HIV-2 that corresponds to the specific location of a reported epitope - most often from HIV-1 - which may or may not have the same primary sequence.

The binarised digestion products and CTL epitopes from the LANL database were cross-compared and 77 fragments were identified that corresponded to the positions of known epitopes (appendix table A.7). The number generated from each digestion was then tallied and plotted, grouped by either virus (figure 6.5 A & B) or by enzyme (figure 6.5 C & D).

As in the case of the bulk digestion products, proteasome generally produced a greater number of products corresponding to known epitopes (hereafter, "optimally cleaved epitopes") than did immunoproteasome (figure 6.5 A & B). The one exception was HIV-1 peptide 4 in which the reverse was observed. Immunoproteasome failed to

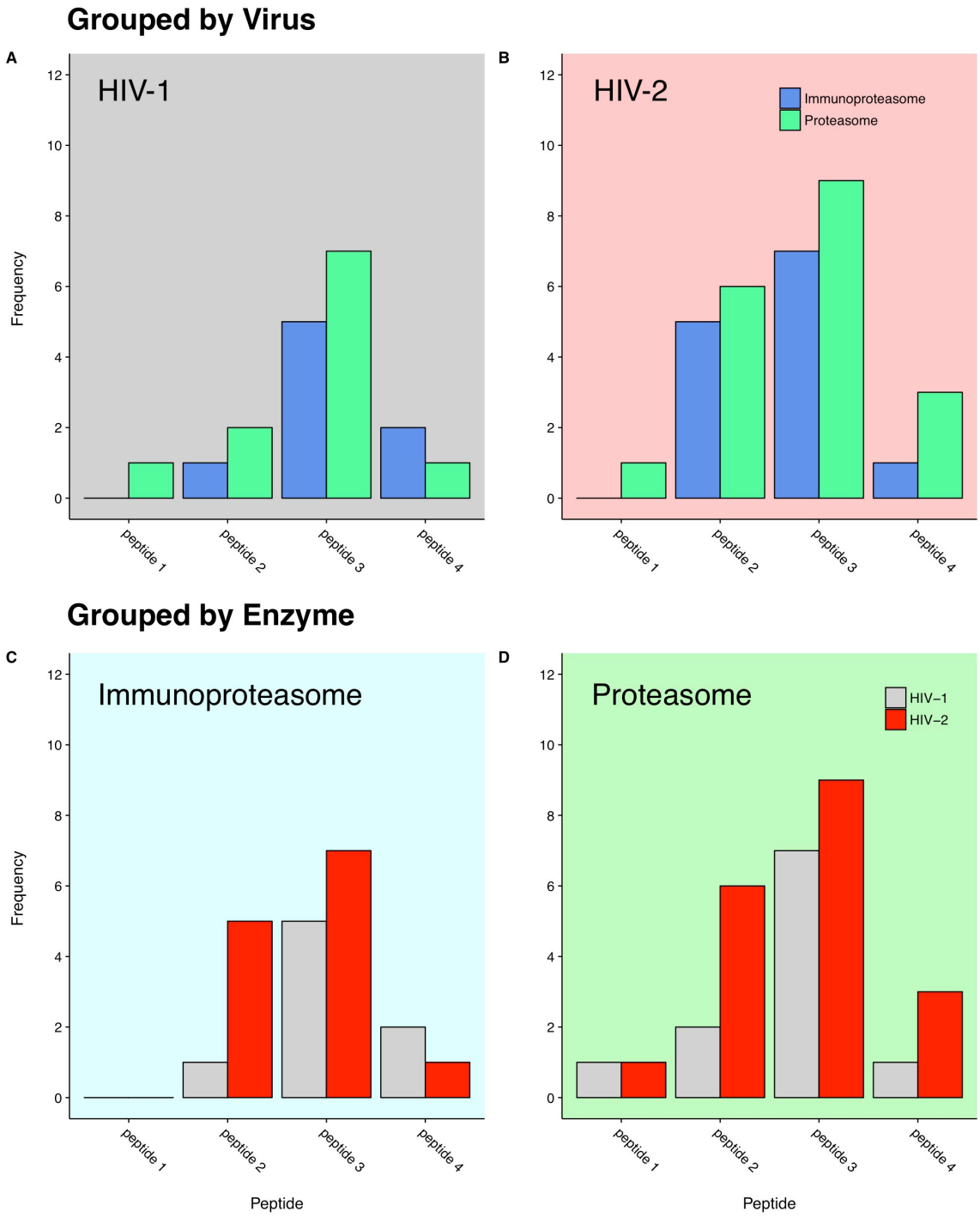


Figure 6.5: Number of optimally cleaved Nef CTL epitopes generated by the digestion of either **A)** HIV-1 or **B)** HIV-2 peptides with either **C)** immunoproteasome or **D)** proteasome.

produce any optimally cleaved epitopes from peptide 1 for either HIV-1 or HIV-2. More optimally cleaved epitopes were also generated from HIV-2 than from HIV-1, consistent with the bulk digestion product data. The one exception for this was again in peptide 4, where more products were generated from HIV-1 than HIV-2 by digestion with immunoproteasome.

To establish whether optimally cleaved epitopes were generated more efficiently from HIV-2 than HIV-1, or whether the increased frequency of epitopes generated was an artefact of the increased number of products generally, the proportion of the total products that were optimally cleaved was calculated. The median percentage of digestion products corresponding to known epitopes was 9.19% (IQR 4.40 - 25.69%) for HIV-1, and 9.59% (IQR 3.81 - 12.50%) for HIV-2. The difference was not statistically significant, as determined by a Wilcoxon rank sum test ($W = 31$, $p = 0.441$).

6.2.4 HLA-I Types Associated with the Epitopes Generated

As particular HLA-I types are associated with better control of HIV infection (section 1.2.3), the HLA types associated with each optimally cleaved epitope were identified to determine whether relatively efficient generation of these epitopes may in part contribute to the improved CTL response. The list of optimally cleaved epitopes was cross-checked with the associated HLA types reported in the LANL Immunology Database. Where more than one allele was listed within an allelic group (e.g. HLA-B57:01 & HLA-B57:03), only one was counted to prevent overrepresentation resulting from variation in the resolution of sequencing. The data were tabulated (table 6.2) and any HLA types contributing more than 5% were highlighted. The most frequent HLA types were HLA-B57, HLA-B58, HLA-B35, HLA-B62 and HLA-A29.

Table 6.2 HLA types associated with optimally cleaved CTL epitopes generated through the digestion of HIV-1 and HIV-2 Nef peptides with immunoproteasome or proteasome. HLA types contributing more than 5% are indicated in bold.

Type	Freq. (%)
HLA-B57	**7.34**
HLA-B58	**7.34**
HLA-B35	**5.96**
HLA-B62	**5.50**
HLA-A29	**5.05**
HLA-C14	4.59
HLA-A11	4.13
HLA-B15	4.13
HLA-A02	3.67
HLA-B53	3.67
HLA-A03	3.21
HLA-B07	3.21
HLA-B18	3.21
HLA-A24	2.75
HLA-B51	2.75

Type	Freq. (%)
HLA-A30	2.29
HLA-B27	2.29
HLA-B63	2.29
HLA-B81	2.29
HLA-B49	1.83
HLA-B55	1.83
HLA-C03	1.83
HLA-A23	1.38
HLA-A31	1.38
HLA-A66	1.38
HLA-B17	1.38
HLA-B42	1.38
HLA-C02	1.38
HLA-C07	1.38
HLA-C15	1.38

Type	Freq. (%)
HLA-A33	0.92
HLA-A68	0.92
HLA-C06	0.92
HLA-C08	0.92
HLA-A01	0.46
HLA-B37	0.46
HLA-B44	0.46
HLA-B60	0.46
HLA-B78	0.46
HLA-C01	0.46
HLA-C04	0.46
HLA-C12	0.46
HLA-C17	0.46

When the HLA-I types associated with the HIV-1 and HIV-2 digestion products were considered independently, the most common HIV-1 associated HLA-I type was HLA-A29, whilst the most common HIV-2 associated type was HLA-A11. Other than this, the results were broadly similar with HLA-B35, HLA-B57, HLA-B58 and HLA-B62 occurring frequently in both contexts.

6.2.5 Comparisons of Corresponding Digestion Products

The degree of similarity in products generated from corresponding digestions of each peptide was next determined. I.e., whether the digestion of an HIV-1 and HIV-2 peptide by a particular enzyme generates products with equivalent cleavage sites; and conversely whether digestion of a particular viral peptide by immunoproteasome and proteasome generates products with equivalent cleavage sites. Venn diagrams representing the extent of overlap in all of the products of each of these corresponding pairs of digestions were constructed (figure 6.6).

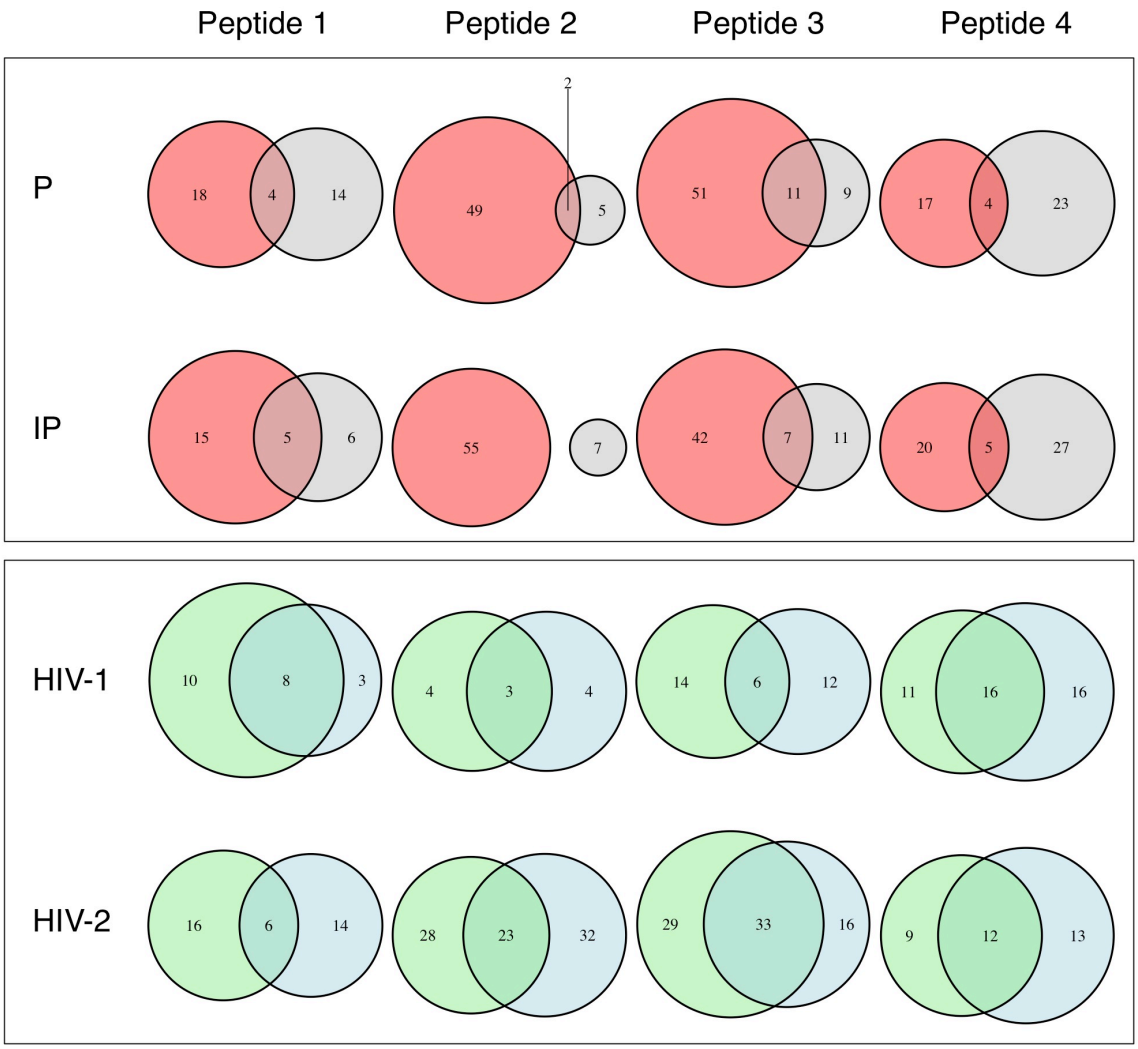


Figure 6.6: Venn diagrams showing the extent of overlap in the products of corresponding digestion pairs. The upper two panels show the overlap between HIV-1 (grey) and HIV-2 (red) peptide digestion products, whilst the lower two panels show the overlap between digests using immunoproteasome (blue) and proteasome (green). Circle size corresponds to the relative number of products found in each pair.

A greater degree of overlap was observed between the digestion products of the same peptide by either immunoproteasome or proteasome (figure 6.6 lower panel) than between digestion of HIV-1 and HIV-2 corresponding peptides by the same enzyme (figure 6.6 upper panel). This was again most extreme in peptide 2 where of the 55 products generated from the digestion of HIV-2 Nef by immunoproteasome, not one was found in the corresponding digestion of the HIV-1 peptide.

6.2.6 Identification and Analysis of Epitope Precursors

The work described thus far has considered the bulk products of digestion, and also the generation of optimally cleaved epitopes. However, epitope precursors generated by the (immuno)proteasome often undergo further processing by downstream enzymes - particularly aminopeptidases such as ERAP prior to loading and presentation (section 1.2.3). For this reason, digestion products containing known epitopes but also N- or C-terminal extensions (hereafter "correct epitope precursors") were next investigated.

To map the digestion products containing correct N- and C-termini, the frequencies of `01` and `10` transitions, respectively, were calculated at each position along the binarised alignment. The corresponding transitions were also calculated for the binarised LANL data, and the resulting frequencies were then recorded by position. The stacked plots for the N- and C-termini are shown in figures 6.7 & 6.8, respectively.

It is evident from figure 6.7 that many cleavage sites observed in the experimental data correspond to the N-termini of known epitopes in the database. However, cleavage sites that do not correspond to reported epitopes are also present in the data. For the digestion of HIV-1, the proportion of N-terminal cut sites that correspond to known epitopes is 60.5% for immunoproteasome, and 70.0% for proteasome. For HIV-2, they are 61.4% and 59.2% respectively.

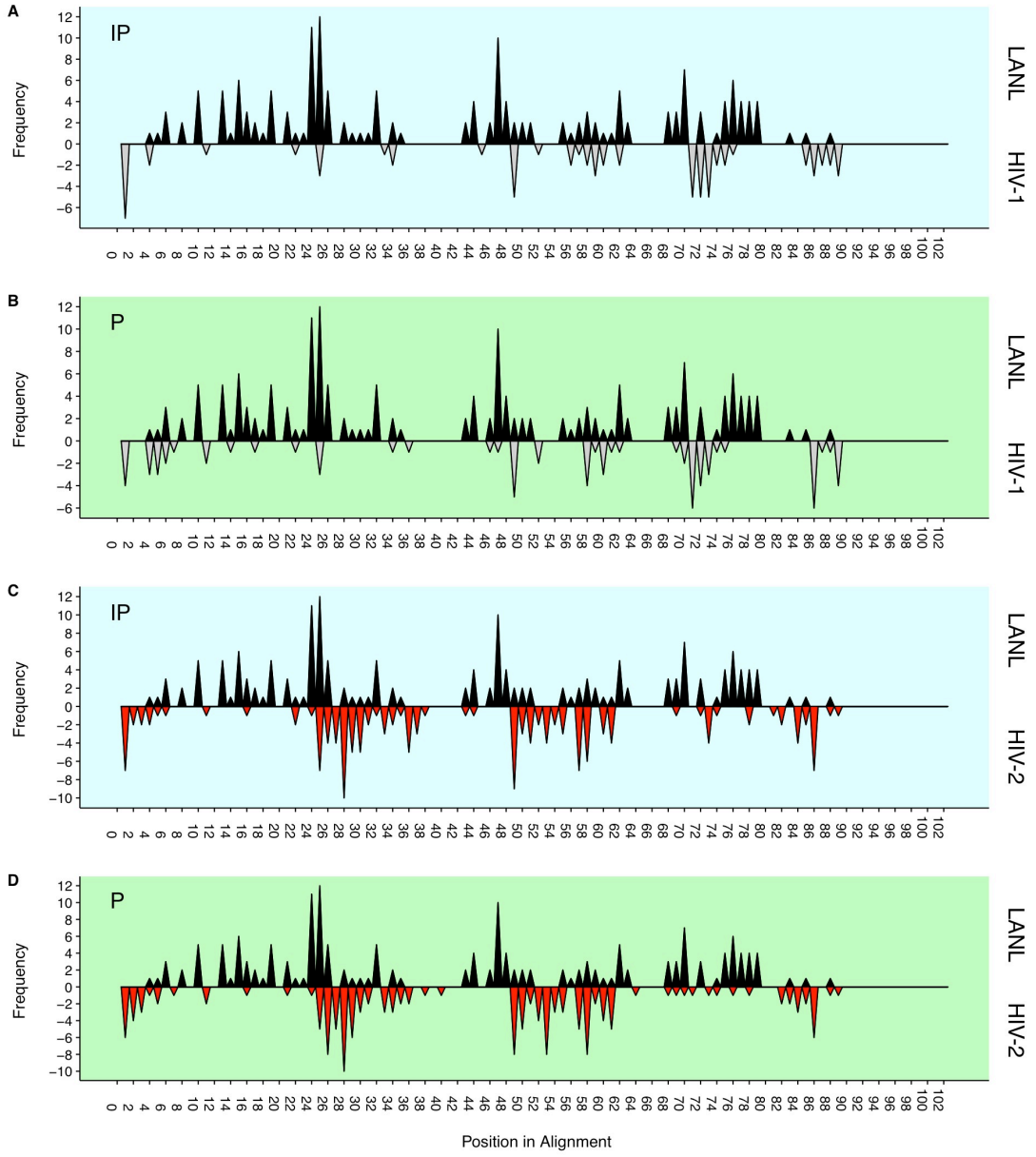


Figure 6.7: Frequency of N-terminal cleavage at each position along the alignment. N-termini of known epitopes in the LANL database are shown on the positive axis (black), whilst the negative axis (absolute frequency) shows the experimental data (HIV-1: grey; HIV-2: red). **A)** HIV-1 immunoproteasome (blue); **B)** HIV-1 proteasome (green); **C)** HIV-2 immunoproteasome (blue); **D)** HIV-2 proteasome (green).

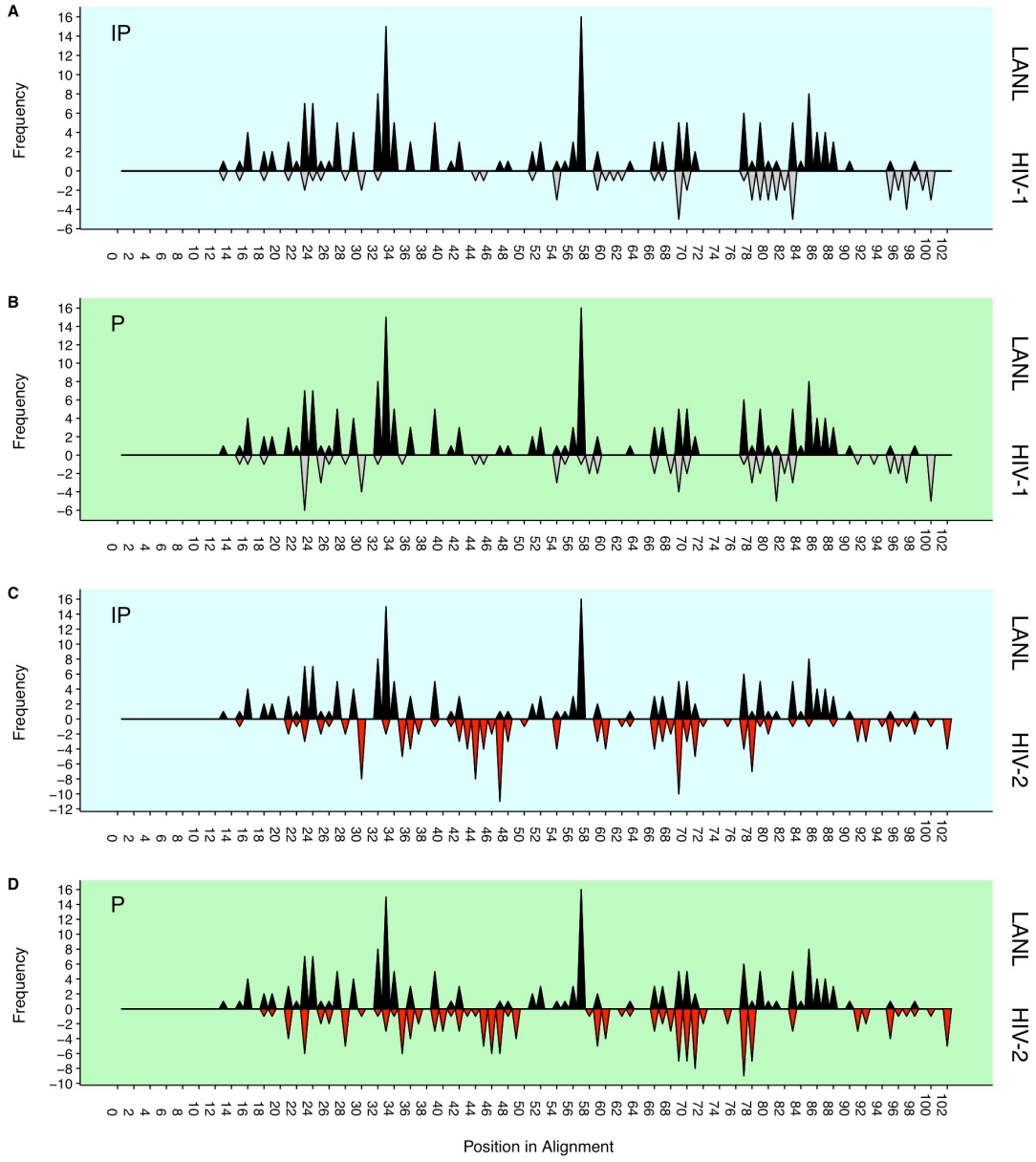


Figure 6.8: Frequency of C-terminal cleavage at each position along the alignment. C-termini of known epitopes in the LANL database are shown on the positive axis (black), whilst the negative axis (absolute frequency) shows the experimental data (HIV-1: grey; HIV-2: red). **A)** HIV-1 immunoproteasome (blue); **B)** HIV-1 proteasome (green); **C)** HIV-2 immunoproteasome (blue); **D)** HIV-2 proteasome (green).

In figure 6.8, several cleavage sites in the experimental data correspond to the C termini of known epitopes, as well as sites that fall outside of those reported. For the digestion of HIV-1, the proportion of cut sites that correspond to known epitopes is 66.5% for immunoproteasome, and 60.1% for proteasome. For HIV-2, they are 60.6% and 55.3% respectively.

Finally, plotting the proportion of N- or C-terminal cleavage sites corresponding to known CTL epitopes for each peptide reveals that there is no consistent difference between HIV-1 and HIV-2 or proteasome and immunoproteasome (figure 6.9).

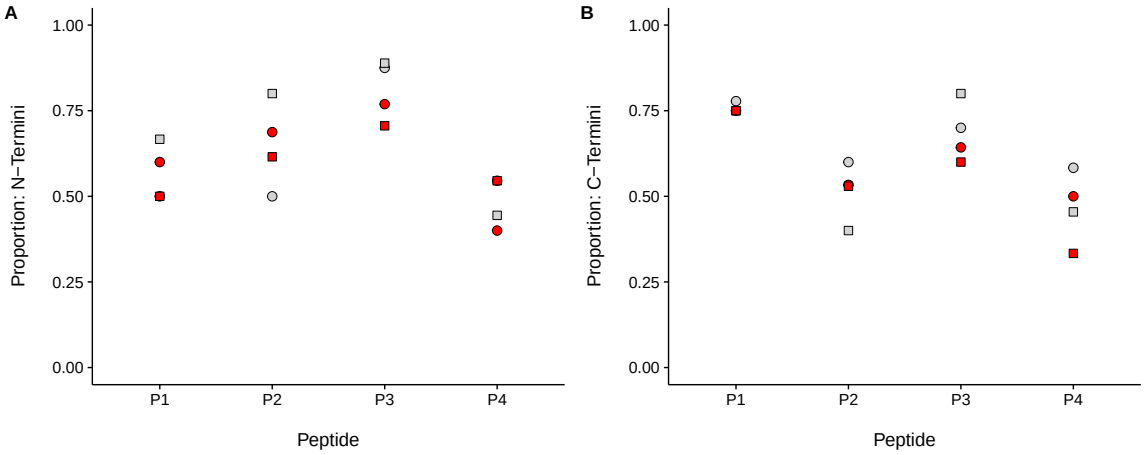


Figure 6.9: Proportion of (immuno)proteasomal cleavage sites corresponding to known CTL epitope A) N- and B) C-termini. HIV-1 is shown in grey, and HIV-2 in red. Circles denote immunoproteasome and squares denote proteasome. P1-P4 refer to peptides 1-4, respectively.

Indeed, there was no statistically significant difference between HIV-1 and HIV-2 in the proportion of products with optimally cleaved N-termini ($W = 35.5$, $p = 0.752$), or C-termini ($W = 39.5$, $p = 0.461$) as determined by a Wilcoxon rank sum test. Therefore, there is no evidence of more efficient generation of correct termini from either virus.

6.3 Discussion

Whilst HIV-1 Nef is a major target of the T-cellular response from early in infection, responses against HIV-2 Nef are notably scarce. The aim of the work described in this chapter was to determine whether such differences in response could be to any extent attributed to the efficiency of (immuno)proteasomal digestion during antigen processing.

Similarly to chapter 3, the CTL epitope density of Nef was first mapped using data reported in the literature. The majority of CTL epitopes were found to cluster towards the centre of the protein, which is purportedly the result of differential proteasomal processing of hydrophobic and hydrophilic regions [438]. With this in mind, a series of four overlapping peptides were designed to span this region for both HIV-1 and HIV-2.

The peptides were digested with both immunoproteasome and proteasome, and the resulting products were characterised. For each corresponding pair of peptides - with the exception of peptide 4 - HIV-2 generated more products following digestion under a given condition. This was particularly striking in peptides 2 and 3 where the number of HIV-2 products exceeded HIV-1 products by seven to one. Therefore, whilst the middle of Nef is most immunogenic, the two peptides overlapping the centre of this region also generated the greatest number of digestion products; but whilst HIV-1 Nef is more immunogenic than HIV-2 Nef, HIV-2 Nef generated a substantially higher number of products.

In vitro proteasomal digestions have previously been shown to generate degradation products that share structural similarities to HLA-I binding peptides, indicating that epitopes may be generated by the proteasome without a requirement for further

proteolysis [521]. Therefore, the digestion products were next scoured for "optimally cleaved" CTL epitopes.

In addition to generating a greater number of total digestion products, HIV-2 Nef peptides also generated more optimally cleaved CTL epitopes than did HIV-1 under most conditions. However, whilst the frequency of such epitopes was higher, this did not correlated with an increase in the proportion of products generated which were optimally cleaved. The same pattern was observed with regards to both the N- and C-terminally extended epitope precursors.

An increase in the extent of digestion - including generation of optimally cleaved CTL epitopes and epitope precursors - that is not coupled with an increase in efficiency of CTL killing raises the question of whether the (immuno)proteasome may in fact be over-digesting HIV-2 Nef. Within the cytoplasm, over 99% of intracellular peptides are destroyed through proteolytic cleavage by endo- and aminopeptidases [522], and so it is plausible that by generating so many different products, the enzyme is not generating any single epitope or epitope precursor at levels sufficiently high to survive this enzymatic attack before being loaded into HLA-I for presentation to CTLs.

One way to test this hypothesis would be to perform a series of experiments comparing both the generation and surface expression of HLA-I bound epitopes that are known to be efficiently generated by the digestion of both HIV-1 and HIV-2 Nef peptides. For example, products corresponding in size and position to the HLA-B62 associated epitope AF7 (HXB2 83-90) can be generated from both HIV-1 and HIV-2 Nef peptide 2 on digestion with proteasome (appendix table A.7).

Assessing the relative amount of each epitope generated is complicated by the fact that there are differences in the primary sequences of the HIV-1 and HIV-2 peptides. Sticking with AF7 as an example, ALDLSHF and AVDMSHF are present in the two peptides, respectively. The different amino acids confer different times of flight to the peptides, making the analysis of the existing mass spectrometry data only semi-quantitative. To measure the relative production of these two products, equal molar amounts of isotopically labelled samples of each peptide could be run alongside the digestion products on the mass spectrometer as internal controls against which the products could be quantified.

An immunopeptidome pull-down assay could then be performed to identify which epitopes are bound to HLA-I and presented on the surface of cells. Cells expressing the appropriate HLA-I type (in this example, HLA-62) would be transfected with expression vectors containing either full-length consensus HIV-1 or HIV-2 Nef. Following incubation, the HLA-I molecules and their associated peptides would be eluted from the cell, washed, and subsequently analysed by mass spectrometry. A key example of this technique in action is Ternette *et al.* [523], wherein HLA-I associated peptides were identified following infection of a cell line with the vaccine candidate MVA.HIVconsv.

It may then be informative to perform *in vitro* killing assays to determine whether the HLA-I bound epitopes identified in the immunopeptidome pull down assay are indeed functionally recognised by CTLs. This is particularly important as this study has relied on a definition of "optimally cleaved" which is something of a misnomer: whilst the epitopes studied correspond in length and position to those reported in the literature, there is no evidence that those products generated would be recognised by CTLs, especially as substantial amino acid mismatches may exist in key posi-

tions. Additionally, the context of the epitope may be important in its processing, presentation or recognition. To address these questions, constructs could be prepared containing the epitope of interest flanked with either a HIV-1 or HIV-2 Nef backbone. Cells expressing a suitable HLA-I type would then be transfected with the constructs before being exposed to appropriately-restricted CTLs. The extent of the cytotoxic response would then be measured using an enzyme-lined immuno-spot (ELISpot) assay or intracellular cytokine staining (ICS) and flow cytometry.

Through performing these further experiments, a clearer picture could be built up as to whether antigen processing, antigen presentation or antigen recognition is disrupted, rendering HIV-2 Nef less immunogenic than HIV-1 Nef.

For the results of such experiments to be generalisable, appropriate HLA-I types should be selected for study which are both associated with generated epitopes, and also prevalent enough to be of relevance. Investigating the HLA-I types associated with the optimally cleaved epitopes generated in this study revealed that five HLA types were particularly frequent: HLA-A29, HLA-B35, HLA-B57, HLA-B58 and HLA-B62. HLA-B restriction is surprisingly frequent in these data, which is perhaps interesting as some HLA-B types are associated with strong CTL responses and an improved prognosis.

Protective alleles include HLA-B57 and HLA-B58, which are the two most highly represented types in this analysis. One could speculate that the improved viral control by HLA-B57 and HLA-B58 positive patients could in part be due to superior processing of viral epitopes associated with these molecules. However, HLA-A29 and HLA-B35 are also frequent yet are permissive alleles associated with more rapid progression [524, 525]. As such, epitopes for these HLA types could simply be

overrepresented in the LANL Immunology Database owing to a greater number of studies on these molecules which are associated with atypical progression.

As HIV-2 Nef-specific CTL responses are less common than HIV-1 Nef-specific responses, one may also expect to find fewer signatures of HLA-associated selection pressure within the sequence of the protein. This could be assessed through large-scale sequence analysis and associated killing assays, to look at evolution within and between epitopes, stratifying patients based on the presence and magnitude of the CTL response elicited.

To conclude, this study represents an essential first step in unravelling how HIV-2 Nef is capable of evading the CTL response so commonly observed in HIV-1 infection. Whilst it has not yet conclusively been shown why HIV-2 Nef-specific CTL responses are less abundant than those specific to HIV-1 Nef, this study has demonstrated quantitative and qualitative differences in the products of the (immuno)proteasomal processing of the immunogenic central regions of these proteins. Indeed, such differences may have important implications in downstream antigen presentation and recognition which warrants substantial further investigation.

As this chapter has presented further discrepancies between HIV-1 and HIV-2 Nef, the following chapter will briefly describe the results of a collaborative effort to resolve the crystal structure of HIV-2 Nef, in an attempt to identify a structural basis for the differences observed thus far.

Chapter 7

Structural Biology of the HIV-2 Nef Protein

7.1 Introduction

Elucidating the crystal structure of HIV-1 Nef has provided key insights into how this highly multifunctional accessory protein contributes to infection and immune evasion [526]. Numerous crystallographic studies have been performed using HIV-1 Nef isolates since the mid 1990s, giving rise to a portfolio of structures showing Nef in complex or in isolation.

The structure of the conserved core was first crystallised in 1996 by Lee *et al.* [402], in complex with a Src-family SH3 domain. Soon after, a comparison of SH3-bound and unbound Nef structures was published, indicating further the motifs required for binding [403]. By 1999, the structure of the anchor domain had been resolved, revealing conformational differences between myristoylated and non-myristoylated versions of the protein [527]. More recent studies have been able to resolve the crystal structures of Nef in complex with the adaptor protein AP-2 [528] and in complex with both AP-1 and the cytoplasmic domain of HLA-I [462], demonstrating the structural basis through which Nef mediates downregulation of CD4 and HLA-I, respectively.

Whilst substantial progress has been made in understanding the structural biology of HIV-1 Nef, the crystal structure of HIV-2 Nef has not yet been resolved. Obtaining this structure could help to explain some of the fundamental differences between HIV-1 and HIV-2 Nef, and perhaps go some way towards understanding the characteristic prognostic differences of these two viruses.

The work presented in this thesis thus far has described a number of discrepancies between HIV-1 and HIV-2 Nef, including - but not limited to - the relative efficiency of HLA-I downregulation mediated by these two proteins (section 5.2.5.2), and the quantitative and qualitative differences in the products of (immuno)proteasomal digestion (section 6.2.5). Both of these discrepancies affect functions that contribute to evasion of the adaptive immune response, and so understanding why the two proteins behave differently could inform attempts to disrupt the processes involved. Comparing and contrasting the structures of HIV-1 and HIV-2 Nef may therefore help to answer such questions.

This very brief chapter summarises the outcome of a project performed in collaboration with researchers at the University of Hokkaido to resolve the crystal structure of HIV-2 Nef. The data described herein were collected and analysed by a number of different researchers, and their individual contributions are declared in chapter 2.

7.1.1 Aims

The specific aim of this project was to:

- Resolve the crystal structure of HIV-2 Nef.

7.2 Results

An N-terminally truncated fragment of unbound HIV-2 Nef containing a C193Y substitution was expressed, purified, and crystallised by vapour diffusion. The structure was then determined to a resolution of 2.07 Å by X-ray crystallography. The data collection and model refinement statistics associated with this structure are listed in table 7.1. The Nef variant used was isolated from patient TD062 of the Caió cohort.

Table 7.1 Statistics of crystallographic data collection and refinement.

Statistic	**Value**
Data collection	
X-ray source	PF BL5A
Resolution (Å)	50.0-2.07 (2.11-2.07)
Wavelength (Å)	1.0000
Space group	$P2_12_12_1$
Cell dimensions	a = 43.57 Å, b = 44.77 Å, c = 103.63 Å
Total reflections	90,611
Unique reflections	12,958 (619)
I / $\sigma(I)$	23.7 (2.9)
Redundancy	7.0 (6.4)
Completeness (%)	100.0 (99.7%)
R_{merge}	0.06 (0.53)
Model refinement	
R_{work} (%)	19.2
R_{free} (%)	23.7
Number of protein residues	145
RMSD bonds (Å)	0.019
RMSD angles (°)	1.86
Ramachandran favoured (%)	97.8
Ramachandran allowed (%)	1.4
Ramachandran outlier (%)	0.7
Average B factor (Å^2)	43.7

The obtained crystal structure of HIV-2 Nef is presented in figure 7.1B. Overlaying the structure with that of an HIV-1 Nef isolate (Arold *et al.* [403]; RCSB PDB accession number 1AVV; figure 7.1C) reveals that despite only 46% sequence identity, a substantial degree of structural overlap exists between the two proteins.

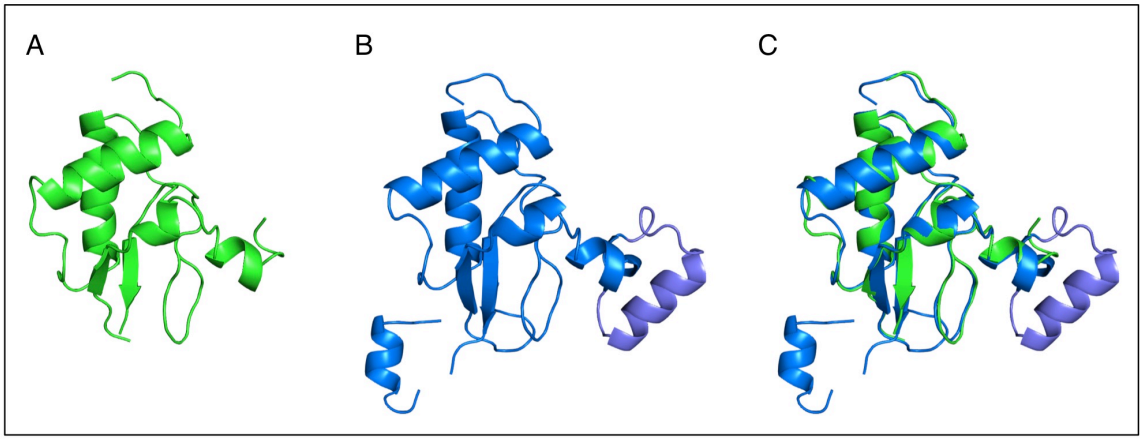


Figure 7.1: Cartoon representation of the crystal structure of HIV-2 Nef. A) Crystal structure of HIV-1 Nef (green); **B)** Crystal structure of HIV-2 Nef (blue), with C-terminal α-helix (purple); **C)** Overlaid structures. The HIV-1 structure presented was published by Arold *et al.* [403], and is listed under RCSB PDB accession number 1AVV.

Although the two proteins exhibit a remarkable degree of structural similarity, the HIV-2 Nef protein structure reveals the presence of a previously unresolved C-terminal α-helix (figure 7.1A & B; purple).

Along the length of the helix are a series of three evenly-spaced lysine residues (figure 7.2A). By aligning 160 HIV-2 Nef sequences collected from the Caió cohort and LANL Sequencing Database (accession numbers listed in appendix table A.8) [376] that contain full-coverage of the C-terminal α-helix, it was apparent that lysine is highly over-represented within the helix, contributing 14.6% of all recorded residues compared with 4.8% in the remainder of the protein. Whilst no equivalent helix is

present in HIV-1, the proportion of lysine residues across the entire length of HIV-1 Nef is 5.5%.

From the constructed alignment of HIV-2 Nef sequences, three further sites were identified which contained lysine in a large proportion of the sequences, but not in the TD062 Nef sequence from which the structure was obtained. These positions are highlighted in yellow in figure 7.2B.

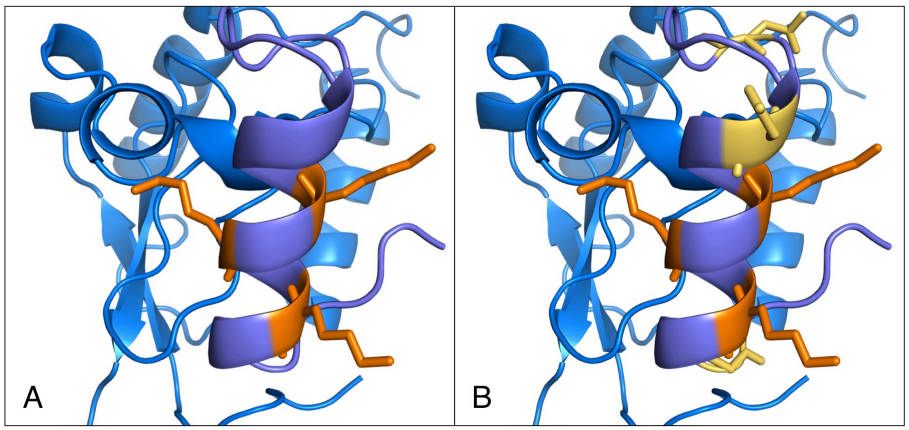


Figure 7.2: Cartoon representation of the C-terminal α-helix of HIV-2 Nef. A) Distribution of lysine residues along the α-helix (orange) in stick representation; **B)** Positions of further lysine residues (yellow) present in a large proportion of HIV-2 Nef sequences from the LANL Sequence Database [376].

7.3 Discussion

This brief chapter has summarised a successful collaborative attempt to resolve the crystal structure of HIV-2 Nef. Elucidation of the structure and comparison with HIV-1 Nef has demonstrated that despite limited sequence identity, HIV-1 and HIV-2 Nef share a relatively conserved tertiary structure. This is consistent with the broadly similar repertoires of functions that can be performed by the two proteins; albeit to different extents, as has been shown for Nef-mediated HLA-I downregulation (chapter 5).

However, HIV-2 Nef also contains a C-terminally extended α-helix, for which no equivalent structure is present in the HIV-1 protein. Whilst analysis of the amino acid sequences of similar primate retroviruses reveals extensions also exist in closely related species, the C-terminus has not been resolved in SIV Nef structures, likely due to intentional truncation to aid crystallisation. The presence of this α-helix may help to explain some of the discrepancies observed in the projects presented in this thesis. For example, as shown in section 5.2.2, expression and detection of fluorescently-labelled Nef was dependent on the presence of a flexible linker of adequate length. HIV-2 Nef required a substantially longer linker than did HIV-1 Nef in order to function correctly (figures 5.6 & 5.7). This observation is consistent with the highly structured and inflexible C-terminal α-helix interfering with correct folding of both Nef itself and the AcGFP1 fluorophore. This is supported by the finding that increasing the length and flexibility of the linker alone restored detection of the fluorophore and function of the Nef protein without further changes being made to the proteins. The role of flexible linkers in correct folding of fusion proteins is reviewed by Chen *et al.* [499].

The over-representation of lysine in the C-terminal α-helix of HIV-2 Nef may also contribute to the differences between HIV-1 and HIV-2 Nef. Lysine residues are

located along the length of the helix, and project their side-chains outwards. Lysine is the amino acid to which ubiquitin is attached in the process of polyubiquitination: an important step in the targeted degradation of proteins by the (immuno)proteasome, reviewed by Lecker *et al.* [529]. One could therefore hypothesise that the over-representation of lysine residues on an exposed and accessible helix may result in more efficient polyubiquitination and targeting of HIV-2 Nef for (immuno)proteasomal degradation.

Whilst purely speculative, this could resolve the apparent paradox of why the higher efficiency of HLA-I downregulation mediated by HIV-2 Nef - relative to HIV-1 Nef - is not coupled with reduced efficiency of CTL killing and higher resulting viral loads (chapter 5). Whilst a particular amount of HIV-2 Nef may downregulate HLA-I more efficiently than the equivalent amount of HIV-1 Nef, increased turnover of the protein by way of (immuno)proteasomal degradation may mean that HIV-2 Nef establishes at lower levels. Increased Nef turnover could therefore result in more HLA-I molecules remaining on the surface of the infected cell, which could explain why robust CTL responses can still be detected against some viral proteins such as Gag in the context of HIV-2 infection [515].

Low amounts of HIV-2 Nef in the infected cell may also help to explain why CTL responses are rarely detected against this particular protein [52, 515]. A combination of efficient polyubiquitination and the more extensive cleavage of HIV-2 Nef by (immuno)proteasome (chapter 6) could mean that HIV-2 Nef is over-processed, preventing accumulation of correct epitopes or epitope precursors at sufficiently high levels to be loaded into HLA-I and presented on the surface of the cell. However, polyubiquitination of the C-terminal α-helix lysine residues alone cannot explain

the relatively increased degradation of HIV-2 peptides seen in chapter 6, as the overlapping peptides used did not span the region in question.

Quantifying the amount of the HIV-1 and HIV-2 Nef proteins in virally infected cells is therefore of high priority, as discussed previously in section 5.3. In addition, measuring the amount of each protein in both the presence and absence of a proteasome inhibitor would clarify whether any differences are the result of the rate of degradation.

Were this hypothesis to stand up to empirical testing, it may also help to explain prognostic differences between HIV-1 and HIV-2 patients. As Nef is a highly multifunctional protein, more rapid turnover may mean that insufficient levels of the protein are available to fulfil its many roles. As Nef enhances viral infectivity, triggers bystander cell death, and mediates modulation the cellular environment through various means (section 1.1.8.7), the intracellular levels of the protein could arguably have a substantial impact on prognosis.

To summarise, this chapter presents the first crystal structure of HIV-2 Nef, which demonstrates the presence of a previously unresolved C-terminal α-helix. This helix may putatively contribute to some of the observed differences between the HIV-1 and HIV-2 Nef proteins, but substantial testing of this hypothesis must be performed. Based on the extensive role that HIV-1 Nef structures have played in understanding this protein's function, the structure presented herein may prove equally informative with time.

Chapter 8

Conclusions and Future Directions

It has been more than three decades since the onset of the HIV pandemic, and yet no successful prophylactic vaccine or cure is available. Intensive biomedical research on a global scale has generated a vast pool of knowledge surrounding this virus, but promising vaccine candidates trialled to date have been met with very limited success, and despite a well-orchestrated immune response, HIV infection is almost invariably fatal in untreated patients. A key difficulty in clearing HIV infection - both naturally and chemotherapeutically - is the ability of HIV to successfully circumvent the many arms of the adaptive immune response. Despite its small genome, HIV is a master of immune evasion, and employs a number of different techniques to survive and flourish within a hostile environment.

A comprehensive understanding of how HIV evades the adaptive immune system is fundamental to the development of effective treatments and vaccines. Indeed, as many prospective vaccine candidates aim to induce antibody and $CD8^+$ cellular responses [530], understanding and accounting for immune evasion is a necessary step in ensuring that appropriate measures are in place to prevent viral escape.

Assiduously investigating all of the ways in which HIV evades the immune response during clinical latency would be a task well beyond the scope of reason, and so this

thesis has focussed on three key areas in this field: namely, mutation of the gp120 *env* gene in response to antibody and CTL-mediated selection pressure; Nef-mediated HLA-I downregulation in the context of HIV-1 and HIV-2 infection; and finally the differential antigenicity of the HIV-1 and HIV-2 Nef proteins.

As the surface-expressed gp120 envelope protein is a major target of the adaptive response against HIV infection - undergoing extensive glycosylation and rapid mutation to escape antibodies targeted against it - this protein represents a logical starting point for a study into adaptive immune evasion. The specific aims of this phase of the study were to characterise intrapatient and interpatient sequence diversity of the gp120 *env* gene and to utilise the sequences obtained to map regions under immunological selection pressure in a narrow-source cohort of clinically latent HIV-1 patients.

This was accomplished by performing a series of phylogenetic analyses on quasispecies sequences isolated from longitudinally-collected plasma samples, from patients of a narrow-source cohort of former plasma donors in China. The analyses broadly confirmed the reported history of the cohort, in so much as the patients sampled were all infected with subtype B' (Thai B) virus within a time-frame encompassing the known dates of plasma donation, with limited spread within the population following the initial outbreak. Additionally, few viruses had undergone CCR5 to CXCR4 co-receptor switching, consistent with the slow rate of disease progression observed in these patients.

As the outbreak stemmed from a narrow source of virus, Bayesian inference was used to estimate the most likely gp120 *env* sequence of the infecting founder virus in the SM cohort. Doing so revealed a striking degree of similarity with the consensus sequence of

similar FPD viruses from neighbouring regions in Henan province. Tens of thousands of people were infected with HIV during commercial plasma donations in China as a result of unsanitary conditions and the redistribution of pooled red blood cells. The overwhelming majority of those infected were poor rural farmers, and approximately 95% of infections occurred in Henan province and neighbouring regions [531]. Owing to the scale, nature, and connotations of this regrettable outbreak, HIV infection in former plasma donors is a highly politically sensitive issue in China. The finding that the SM cohort founder virus so closely resembles the virus present in neighbouring regions will therefore have serious sociopolitical implications, and diplomatic handling of this controversial finding will therefore be paramount.

Returning to the topic of immune evasion, this phase of the study also revealed a number of interesting findings. Viral evolutionary rate estimates and dN/dS ratios indicated that gp120 *env* was primarily under purifying selection, but mutational hotspots could be identified that were under significant positive selection pressure. Five sites were identified that were under positive selection in at least half of the patients sampled. When these sites were modelled to the three-dimensional structure of a gp120 trimer, all five sites were found to be within or adjacent to the hypervariable loops on the outermost face of the protein complex indicating that the selection pressure is likely to be driven by antibody responses. This study is particularly novel in that this is the first recorded instance of sites identified through renaissance counting being considered in light of the quaternary - or even tertiary - structure of a protein.

Expanding from this point, the narrow source of the cohort was next used to identify and characterise variation from the inferred founder virus. Thirty-one "major variants" were identified in this study, as well as fifty-one invariant sites. Again, mapping

these sites to the three-dimensional structure revealed that the vast majority of invariant sites clustered within the core of the trimer complex, whilst the major variants were dispersed throughout the outer face of the protein. Determining the presence of each variant within individual patients over the longitudinally-collected samples gave an indication of whether the variants were putatively driven by antibody- or CTL-mediated selection pressure, although rigorous testing and verification of this novel technique for assessing selection pressure will be required before firm conclusions can be drawn.

More broadly, this study generated a novel dataset of almost 600 viral gp120 *env* sequences from largely untreated patients. Due to the widespread availability of antiretroviral drugs and the ethical constraints of working with human subjects, such datasets are not common and the narrow-source of the outbreak in this cohort makes this particular dataset an unusual and potentially highly informative resource. An appropriate data stewardship plan will be drawn up for the dataset in question, and the sequences will subsequently be made available for use by the wider scientific community. The study of such sequences can provide key insight into the evolutionary dynamics of HIV in response to specific host immune responses, and so it is anticipated that this recourse may help to pave the way to the development of disruptive therapeutic or prophylactic drugs. Indeed, the sequences are already being used by a collaborating group led by Professor Sunetra Gupta at the Department of Zoology, University of Oxford. This study will investigate the role of antibody-mediated selection pressure in clinically latent HIV-1 infection, in light of a recent model implicating an important role for antibody responses in determining progression to AIDS [532].

The second phase of the study moved away from escape mutations, and towards a more mechanistic view of immune evasion. The HIV accessory protein Nef downregulates HLA-I from the plasma membrane, to reduce the visibility of infected cells to cytotoxic lymphocytes. In light of a recent study indicating that SF2 Nef may differentially downregulate HLA-A and HLA-B [399], the specific aim of this project was to investigate whether this phenomenon was also mediated by primary isolates of Nef from both HIV-1 and HIV-2 infected patients.

A series of flow cytometric studies were performed to measure HLA-I downregulation mediated by ten Nef isolates sampled from both HIV-1 and HIV-2 mono-infected patients. Three key conclusions were drawn from this study. Firstly, in stark contrast to the initial report [399], and a recent further publication describing the differential downregulation of HLA-A24 and HLA-B35 [489], there was no evidence to support differential downregulation of HLA-A and HLA-B by any Nef isolates tested, including the SF2 Nef protein from which the original conclusions were drawn. Whilst many factors could in part account for the lack of reproducibility of the initial findings, differing expression levels of various HLA-I molecules and the prior use of suboptimal statistical techniques are a likely explanation for the discrepancy observed.

This study also revealed that primary isolates of Nef downregulate HLA-I less efficiently than do laboratory-adapted or consensus Nef variants. As laboratory-adapted virus is cultured *in vitro*, and consensus variants represent a purportedly "ideal" sequence - both of which are effectively shaped in the absence of immunological selection pressure - it was hypothesised that primary isolates of Nef may be less efficient as a result of incorporated escape mutations that compromise function. If this hypothesis is correct, there is an inherent irony to the fact that one immune evasion mechanism (escape mutation) may negatively impact a second immune evasion mechanism (HLA-

I downregulation). If anything, this simply highlights the complex interplay between host and virus that drives the evolutionary arms race.

In obtaining these data, a number of substantial limitations became apparent with the way that flow cytometric data are canonically analysed in the context of receptor modulation. To address these limitations, a novel technique was developed: the staggered window approach to receptor modulation (SWARM). Rather than categorising data simply by whether they relate to cells that are positive or negative for a fluorophore of interest, SWARM uses bins that are minimally affected by decision-making on the part of the researcher. This allows data to be considered in the absence of many factors that confound bifurcate analyses, and permits higher-resolution analysis of relationships between molecules.

This newly-developed technique revealed the novel and highly unexpected finding that HIV-2 Nef may be more efficient at downregulating HLA-I than HIV-1 Nef when a given concentration of the protein is present within the cell. Bifurcate fold change analyses failed to detect this disparity, but the SWARM data were consistent and the difference was found to be highly statistically significant. However, when high concentrations of Nef were present in the cell, differences between the Nef isolates converged to the point where all behaved similarly. This may resolve the apparent paradox that the extent of HLA-I downregulation mediated by Nef rarely correlates with markers of disease progression [490, 496], as physiological expression levels may not be equivalent to those achieved *in vitro*.

Whilst SWARM was developed for the analysis of Nef-mediated HLA-I downregulation, this technique could be used to address numerous biological questions. There are few restrictions associated with its application, and as flow cytometry is a commonly

used technique in many biomedical fields, SWARM may be of potential interest to the wider scientific community. As such, all scripts developed will shortly be made freely available, subject to an appropriate data stewardship plan.

The third phase of this study continued with the theme of CTL cytotoxicity, but redirected the focus towards antigen processing. Whilst HIV-1 Nef is highly immunogenic and CTL responses against it can be detected from early in infection, HIV-2 Nef rarely elicits a response [52, 515]. Inspired by studies showing that differential cleavage by proteasomal processing directly impacts the generation of epitopes and the resulting CTL response [439, 517–519], it was hypothesised that differential proteasomal processing of HIV-1 and HIV-2 Nef may contribute to the difference in CTL response as well.

A series of *in vitro* (immuno)proteasomal digestions were performed using overlapping HIV-1 and HIV-2 Nef peptides. Unexpectedly - and perhaps counter-intuitively - HIV-2 Nef was digested to a much greater extent than HIV-1 Nef by both proteasome and immunoproteasome. This is in stark contrast to conventional wisdom that states that efficient processing leads to efficient epitope generation and higher magnitude CTL responses. Indeed, studies have demonstrated that antigen processing rate determines the immunogenicity of immunoproteasome-generated epitopes in a murine model [533].

Whilst HIV-2 Nef generated more digestion products than did HIV-1 Nef, the proportion of products corresponding to reported CTL epitopes was no higher, indicating that there was no additional specificity. As such, it is hypothesised that HIV-2 Nef may in fact be over-digested, resulting in a pool of products, amongst which no

epitopes or epitope precursors are present at sufficiently high levels to reliably be presented and recognised by CTLs.

However, to test this hypothesis, a series of further experiments must be performed. The recommended approach would therefore be to perform an immunopeptidome pull-down assay to determine exactly what is bound to surface-expressed HLA-I following HIV-1 and HIV-2 infection, or alternatively, following transfection with the full-length proteins of interest. Such an approach has been successfully used before to characterise the immunopeptidome following exposure to an HIV vaccine candidate [523]. If both HIV-1 and HIV-2 Nef peptides are found to be correctly expressed, one could then assess the ability of CTLs to recognise the presented epitope by performing killing assays measured by an ELISpot assay, or ICS and flow cytometry. Performing these supplementary studies should produce a clearer picture of exactly why HIV-2 Nef is relatively non-immunogenic.

In performing these studies, it became apparent that a number of substantial differences exist between HIV-1 and HIV-2 Nef. As elucidating the structure of HIV-1 Nef has been fundamental to understanding the functions of this protein [526], a collaborative study was initiated with the University of Hokkaido in a successful attempt to resolve the crystal structure of HIV-2 Nef. The crystal structure of HIV-2 Nef is presented for the first time in this thesis. Resolving the structure has revealed a C-terminal α-helix which is not present in the homologous HIV-1 Nef protein. Whilst the function of the helix is at this stage purely speculative, it seems to be universally present in HIV-2 Nef, and contains an over-representation of lysine residues which are hypothesised to increase the efficiency of polyubiquitination and subsequent (immuno)proteasomal processing of the protein.

The presence of the helix may explain a number of the discrepancies described within this thesis, but may also have implications further afield. Identification of this previously unreported structure opens the door for numerous potential studies to discover its function and determine whether or not it contributes to the broadly different prognosis of HIV-1 and HIV-2 infection. Indeed, the collaborating research group at the University of Hokkaido has already begun attempting to identify putative binding partners of this protein.

In conclusion, this thesis has contributed to the body of knowledge surrounding three key areas of HIV immune evasion. A comprehensive understanding of immune evasion in the context of clinically latent HIV infection will support the successful development of antiretroviral vaccines and treatments by ensuring that sufficient measures are in place to "plug any gaps" in the intended adaptive response. Only once all possible viral escape scenarios are considered and accounted for will a vaccine be fundamentally successful.

Approximate word count: 43,000

References

1. Gottlieb, M. S. *et al.* Pneumocystis-carinii pneumonia and mucosal candidiasis in previously healthy homosexual men - evidence of a newly acquired cellular immunodeficiency. *New England Journal of Medicine* **305,** 1425–1431 (1981).
2. Masur, H. *et al.* An outbreak of community-acquired pneumocystis carinii pneumonia - initial manifestation of cellular immune dysfunction. *New England Journal of Medicine* **305,** 1431–1438 (1981).
3. Hymes, K. B. *et al.* Kaposi's sarcoma in homosexual men - a report of eight cases. *The Lancet* **2,** 598–600 (1981).
4. Centers for Disease Control and Prevention. Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men. New York City and California. *MMWR. Morbidity and mortality weekly report* **30,** 305–8 (1981).
5. Jaffe, H. W., Bregman, D. J. & Selik, R. M. Acquired immune deficiency syndrome in the United States: the first 1,000 cases. *The Journal of Infectious Diseases* **148,** 339–345 (1983).
6. Barré-Sinoussi, F. *et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune-deficiency syndrome (AIDS). *Science* **220,** 868–871 (1983).
7. Salemi, M. *et al.* Dating the common ancestor of SIVcpz and HIV-1 group M and the origin of HIV-1 subtypes using a new method to uncover clock-like molecular evolution. *The FASEB Journal* **15,** 276–278 (2001).
8. Lemey, P. *et al.* Tracing the origin and history of the HIV-2 epidemic. *Proceedings of the National Academy of Sciences of the United States of America* **100,** 6588–6592 (2003).
9. Wertheim, J. O. & Worobey, M. Dating the age of the SIV lineages that gave rise to HIV-1 and HIV-2. *PLoS Computational Biology* **5,** e1000377 (2009).
10. Chitnis, A., Rawls, D. & Moore, J. Origin of HIV type 1 in colonial French Equatorial Africa? *AIDS Research and Human Retroviruses* **16,** 5–8 (2000).
11. Peeters, M. *et al.* Risk to human health from a plethora of Simian immunodeficiency viruses in primate bushmeat. *Emerging Infectious Diseases* **8,** 451–457 (2002).
12. Wolfe, N. D. *et al.* Naturally acquired simian retrovirus infections in central African hunters. *The Lancet* **363,** 932–937 (2004).

13. Worobey, M. *et al.* Island biogeography reveals the deep history of SIV. *Science* **329,** 1487 (2010).

14. Headrick, R. *Colonialism, health and illness in French Equatorial Africa, 1885-1935* (African Studies Association Press, Atlanta, GA, 1994).

15. Muller, V., De Sousa, J. D., Vandamme, A. M. & Lemey, P. High GUD incidence in the early 20th century created a particularly permissive time window for the origin and initial spread of epidemic HIV strains. *Figshare* **1** (2010).

16. Abecasis, A. B. *et al.* HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics. *Retrovirology* **10,** e1–13 (2013).

17. Fox, J. *et al.* Epidemiology of non-B clade forms of HIV-1 in men who have sex with men in the UK. *AIDS* **24,** 2397–2401 (2010).

18. Sharp, P. M. & Hahn, B. H. Origins of HIV and the AIDS pandemic. *Cold Spring Harbor Perspectives in Medicine* **1,** a006841 (2011).

19. Poulsen, A. G., Aaby, P., Jensen, H. & Dias, F. Risk factors for HIV-2 seropositivity among older people in Guinea-Bissau. A search for the early history of HIV-2 infection. *Scandinavian Journal of Infectious Diseases* **32,** 169–175 (2000).

20. Preston, B., Poiesz, B. & Loeb, L. Fidelity of HIV-1 reverse transcriptase. *Science* **242,** 1168–1171 (1988).

21. Roberts, J., Bebenek, K. & Kunkel, T. The accuracy of reverse transcriptase from HIV-1. *Science* **242,** 1171–1173 (1988).

22. Troyer, R. M. *et al.* Changes in Human Immunodeficiency Virus Type 1 fitness and genetic diversity during disease progression. *Journal of Virology* **79,** 9006–9018 (2005).

23. Gao, F. *et al.* Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature* **397,** 436–441 (1999).

24. Hirsch, V. M., Olmsted, R. A., Murpheycorb, M., Purcell, R. H. & Johnson, P. R. An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature* **339,** 389–392 (1989).

25. Taylor, B. S., Sobieszczyk, M. E., McCutchan, F. E. & Hammer, S. M. The challenge of HIV-1 subtype diversity. *New England Journal of Medicine* **358,** 1590–1602 (2008).

26. Ibe, S. *et al.* HIV-2 CRF01 AB first circulating recombinant form of HIV-2. *Journal of Acquired Immune Deficiency Syndrome* **54,** 241–247 (2010).

27. Los Alamos National Laboatory (LANL). *HIV Circulating Recombinant Forms (CRFs)* <`http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html`> (Published Nov 2014, Accessed Mar 2015).

28. *HIV-2* NAM aidsmap (2015, Accessed Mar 2015). <`http://www.aidsmap.com/HIV-2/page/1322993/`>.

29. World Health Organization. *Number of people (all ages) living with HIV - Estimates by country* <http://apps.who.int/gho/data/node.main.620?lang=en> (Published 2015, Accessed Jun 2016).

30. Adjorlolo-Johnson, G. *et al.* Prospective comparison of mother-to-child transmission of HIV-1 and HIV-2 in Abidjan, Ivory Coast. *Journal of the American Medical Association* **272,** 462–466 (1994).

31. Lehman, D. A. & Farquhar, C. Biological mechanisms of vertical human immunodeficiency virus (HIV-1) transmission. *Reviews in Medical Virology* **17,** 381–403 (2007).

32. O'Donovan, D. *et al.* Maternal plasma viral RNA levels determine marked differences in mother-to-child transmission rates of HIV-1 and HIV-2 in The Gambia. MRC/Gambia Government/University College London Medical School working group on mother-child transmission of HIV. *AIDS* **14,** 441–448 (2000).

33. Burgard, M. *et al.* Mother-to-child transmission of HIV-2 infection from 1986 to 2007 in the ANRS French Perinatal Cohort EPF-CO1. *Clinical Infectious Diseases* **51,** 833–843 (2010).

34. Gottlieb, G. S. *et al.* Equal plasma viral loads predict a similar rate of CD4 T cell decline in Human Immunodeficiency Virus (HIV) type 1- and HIV-2-infected individuals from Senegal, West Africa. *Journal of Infectious Diseases* **185,** 905–914 (2002).

35. Popper, S. J. *et al.* Lower human immunodeficiency virus (HIV) type 2 viral load reflects the difference in pathogenicity of HIV-1 and HIV-2. *Journal of Infectious Diseases* **180,** 1116–1121 (1999).

36. Berry, N. *et al.* Low peripheral blood viral HIV-2 RNA in individuals with high CD4 percentage differentiates HIV-2 from HIV-1 infection. *Journal of Human Virology* **1,** 457–468 (1998).

37. Andersson, S. *et al.* Plasma viral load in HIV-1 and HIV-2 singly and dually infected individuals in Guinea-Bissau, West Africa: significantly lower plasma virus set point in HIV-2 infection than in HIV-1 infection. *Archives of Internal Medicine* **160,** 3286–3293 (2000).

38. Marlink, R. *et al.* Reduced rate of disease development after HIV-2 infection as compared to HIV-1. *Science* **265,** 1587–1590 (1994).

39. Peterson, I. *et al.* Mortality and immunovirological outcomes on antiretroviral therapy in HIV-1 and HIV-2-infected individuals in the Gambia. *AIDS* **25,** 2167–2175 (2011).

40. Clapham, P. R. & McKnight, A. HIV-1 receptors and cell tropism. *British Medical Bulletin* **58,** 43–59 (2001).

41. McKnight, A. *et al.* A broad range of chemokine receptors are used by primary isolates of human immunodeficiency virus type 2 as coreceptors with CD4. *Journal of Virology* **72,** 4065–4071 (1998).

42. Blaak, H. *et al.* CCR5, GPR15, and CXCR6 are major coreceptors of human immunodeficiency virus type 2 variants isolated from individuals with and without plasma viremia. *Journal of Virology* **79,** 1686–1700 (2005).

43. Calado, M. *et al.* Coreceptor usage by HIV-1 and HIV-2 primary isolates: The relevance of CCR8 chemokine receptor as an alternative coreceptor. *Virology* **408,** 174–182 (2010).

44. Schindler, M. *et al.* Nef-mediated suppression of T cell activation was lost in a lentiviral lineage that gave rise to HIV-1. *Cell* **125,** 1055–1067 (2006).

45. Los Alamos National Laboatory (LANL). *Landmarks of the HIV-1, HIV-2, and SIV genomes* <`http://www.hiv.lanl.gov/content/immunology/pdf/2000/intro/GenomeMaps.pdf`> (Published Dec 2000, Accessed Jan 2015).

46. Michel, P. *et al.* Reduced immune activation and T cell apoptosis in Human Immunodeficiency Virus Type 2 compared with Type 1: correlation of T cell apoptosis with beta2 microglobulin concentration and disease rvolution. *Journal of Infectious Diseases* **181,** 64–75 (2000).

47. Hanson, A. *et al.* Distinct profile of T cell activation in HIV type 2 compared to HIV type 1 infection: Differential mechanism for immunoprotection. *AIDS Research and Human Retroviruses* **21,** 791–798 (2005).

48. Duvall, M. G. *et al.* Maintenance of HIV-specific CD4+ T cell help distinguishes HIV-2 from HIV-1 infection. *Journal of Immunology* **176,** 6973–6981 (2006).

49. Duvall, M. G. *et al.* Polyfunctional T cell responses are a hallmark of HIV-2 infection. *European Journal of Immunology* **38,** 350–363 (2008).

50. Lopes, A. R. *et al.* Greater CD8+ TCR heterogeneity and functional flexibility in HIV-2 compared to HIV-1 infection. *The Journal of Immunology* **171,** 307–316 (2003).

51. Jaye, A. *et al.* No differences in cellular immune responses between asymptomatic HIV type 1- and type 2-infected Gambian patients. *Journal of Infectious Diseases* **189,** 498–505 (2004).

52. Zheng, N. N. *et al.* Comparison of human immunodeficiency virus (HIV)-specific T-cell responses in HIV-1- and HIV-2-infected individuals in Senegal. *Journal of Virology* **78,** 13934–13942 (2004).

53. Bjorling, E. *et al.* Autologous neutralizing antibodies prevail in HIV-2 but not in HIV-1 infection. *Virology* **193,** 528–530 (1993).

54. Shi, Y. *et al.* Evolution of human immunodeficiency virus type 2 coreceptor usage, autologous neutralization, envelope sequence and glycosylation. *Journal of General Virology* **86,** 3385–3396 (2005).

55. Funke, J., Durr, R., Dietrich, U. & Koch, J. Natural killer cells in HIV-1 infection: a double-edged sword. *AIDS Reviews* **13,** 67–76 (2011).

56. Nuvor, S. V., van der Sande, M., Rowland-Jones, S., Whittle, H. & Jaye, A. Natural killer cell function is well preserved in asymptomatic human immunodeficiency virus type 2 (HIV-2) infection but similar to that of HIV-1 infection when CD4 T-cell counts fall. *Journal of Virology* **80,** 2529–2538 (2006).

57. Takeuchi, J. S. *et al.* High level of susceptibility to human TRIM5alpha conferred by HIV-2 capsid sequences. *Retrovirology* **10,** e1–11 (2013).

58. World Health Organisation. *HIV/AIDS Fact Sheet 2016* <http://www.who.int/mediacentre/factsheets/fs360/en/> (Updated Jul 2016, Accessed Jul 2016).

59. Ortblad, K. F., Lozano, R. & Murray, C. J. The burden of HIV: insights from the Global Burden of Disease Study 2010. *AIDS* **27,** 2003–2017s (2013).

60. *Botwsana global AIDS response report* Republic of Botswana National AIDS Coordinating Agency (UNAIDS, June 2015, Accessed Jun 2016). <http://www.unaids.org/sites/default/files/country/documents/BWA_narrative_report_2015.pdf>.

61. *Lesotho global AIDS response progress report* Lesotho Ministry of Health (UNAIDS, 2015, Accessed Jun 2016). <http://www.unaids.org/sites/default/files/country/documents/LSO_narrative_report_2015.pdf>.

62. *Swaziland global AIDS response progress reporting* The Kingdom of Swaziland (UNAIDS, Mar. 2014, Accessed Jun 2016). <http://www.unaids.org/sites/default/files/country/documents/SWZ_narrative_report_2014.pdf>.

63. Public Health England. *HIV in the United Kingdom - Situation Report 2015* <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/477702/HIV_in_the_UK_2015_report.pdf/> (Published 2015, Accessed Jun 2016).

64. *HIV in the United Kingdom: 2014 Report* Public Health England (Nov. 2014, Accessed Jan 2015). <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/377194/2014_PHE_HIV_annual_report_1911_2014.pdf>.

65. Gottlieb, G. S. *et al.* A call for randomized controlled trials of antiretroviral therapy for HIV-2 infection in West Africa. *AIDS* **22,** 2069–2072 (2008).

66. Kanki, P. J. *et al.* Slower heterosexual spread of HIV-2 than HIV-1. *The Lancet* **343,** 943–946 (1994).

67. De Cock, K. *et al.* Prevention of mother-to-child HIV transmission in resource-poor countries: Translating research into policy and practice. *Journal of the American Medical Association* **283,** 1175–1182 (2000).

68. *Estimated risk per exposure* NAM aidsmap (2014, Accessed Jan 2015). <http://www.aidsmap.com/Estimated-risk-per-exposure/page/1324038/>.

69. Pantaleo, G. *et al.* HIV-infection is active and progressive in lymphoid-tissue during the clinically latent stage of disease. *Nature* **362,** 355–358 (1993).

70. Fauci, A., Pantaleo, G., Stanley, S. & Weissman, D. Immunopathogenic mechanisms of HIV infection. *Annals of Internal Medicine* **124,** 654–663 (1996).

71. Tindall, B. & Cooper, D. Primary HIV infection - Host responses and intervention strategies. *AIDS* **5,** 1–14 (1991).

72. Clark, S. J. *et al.* High titers of cytopathic virus in plasma of patients with symptomatic primary HIV-1 infection. *New England Journal of Medicine* **324,** 954–960 (1991).

73. Hyman, J. M., Li, J. & Stanley, E. A. The differential infectivity and staged progression models for the transmission of HIV. *Mathematical Biosciences* **155,** 77–109 (1999).

74. Hollingsworth, T. D., Anderson, R. M. & Fraser, C. HIV-1 transmission, by stage of infection. *Journal of Infectious Diseases* **198,** 687–693 (2008).

75. Schacker, T., Collier, A. C., Hughes, J., Shea, T. & Corey, L. Clinical and epidemiologic features of primary HIV infection. *Annals of Internal Medicine* **125,** 257–264 (1996).

76. Daar, E. S., Moudgil, T., Meyer, R. D. & Ho, D. D. Transient high levels of viremia in patients with primary human immunodeficiency virus type 1 infection. *New England Journal of Medicine* **324,** 961–964 (1991).

77. Borrow, P., Lewicki, H., Hahn, B. H., Shaw, G. M. & Oldstone, M. B. A. Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type-1 infection. *Journal of Virology* **68,** 6103–6110 (1994).

78. Fraser, C., Hollingsworth, T. D., Chapman, R., de Wolf, F. & Hanage, W. P. Variation in HIV-1 set-point viral load: Epidemiological analysis and an evolutionary hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 17441–17446 (2007).

79. Mellors, J. W. *et al.* Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* **272,** 1167–1170 (1996).

80. Goujard, C. *et al.* CD4 cell count and HIV DNA level are independent predictors of disease progression after primary HIV type 1 infection in untreated patients. *Clinical Infectious Diseases* **42,** 709–715 (2006).

81. Collaborative Group on AIDS Incubation and HIV Survival. Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis. *The Lancet* **355,** 1131–1137 (2000).

82. Schneider, E. *et al.* Revised surveillance case definitions for HIV infection among adults, adolescents, and children aged less than 18 months and for HIV infection and AIDS among children aged 18 months to less than 13 years. United States, 2008. *MMWR Recommendations and Reports - Centers for Disease Control* **57,** 1–12 (2008).

83. Zwahlen, M. & Egger, M. *Progression and mortality of untreated HIV-positive individuals living in resource-limited settings: update of literature review and evidence synthesis* UNAIDS: Geneva. 2006. <http://data.unaids.org/pub/periodical/2006/zwahlen_unaids_hq_05_422204_2007_en.pdf>.

84. Pantaleo, G. *et al.* Studies in subjects with long-term nonprogressive human immunodeficiency virus infection. *New England Journal of Medicine* **332,** 209–216 (1995).

85. Lefrere, J. J. *et al.* Even individuals considered as long-term nonprogressors show biological signs of progression after 10 years of human immunodeficiency virus infection. *Blood* **90,** 1133–1140 (1997).

86. Petrucci, A. *et al.* How many HIV-infected individuals may be defined as long-term nonprogressors? A report from the Italian seroconversion study. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* **14,** 243–248 (1997).

87. Grabar, S. *et al.* Prevalence and comparative characteristics of long-term non-progressors and HIV controller patients in the French Hospital Database on HIV. *AIDS* **23,** 1163–1169 (2009).

88. Strathdee, S. A. *et al.* Lack of consistency between five definitions of nonprogression in cohorts of HIV-infected seroconverters. *AIDS* **10,** 959–965 (1996).

89. Olson, A. D. *et al.* An evaluation of HIV elite controller definitions within a large seroconverter cohort collaboration. *PLoS ONE* **9,** e86719 (2014).

90. Rowland-Jones, S. L. & Whittle, H. C. Out of Africa: what can we learn from HIV-2 about protective immunity to HIV-1? *Nature Immunology* **8,** 329–331 (2007).

91. Engelman, A. & Cherepanov, P. The structural biology of HIV-1: mechanistic and therapeutic insights. *Nature Reviews Microbiology* **10,** 279–290 (2012).

92. Freed, E. O., Myers, D. J. & Risser, R. Characterization of the fusion domain of the human immunodeficiency virus type-1 envelope glycoprotein GP41. *Proceedings of the National Academy of Sciences of the United States of America* **87,** 4650–4654 (1990).

93. Gallo, S. A. *et al.* The HIV Env-mediated fusion reaction. *Biochimica Et Biophysica Acta-Biomembranes* **1614,** 36–50 (2003).

94. Fenyö, E. M., Esbjörnsson, J., Medstrand, P. & Jansson, M. Human immunodeficiency virus type 1 biological variation and coreceptor use: from concept to clinical significance. *Journal of Internal Medicine* **270,** 520–531 (2011).

95. Lengauer, T., Sander, O., Sierra, S., Thielen, A. & Kaiser, R. Bioinformatics prediction of HIV coreceptor usage. *Nature Biotechnology* **25,** 1407–1410 (2007).

96. Pawlica, P. & Berthoux, L. Cytoplasmic dynein promotes HIV-1 uncoating. *Viruses* **6,** 4195–4211 (2014).

97. Lukic, Z., Dharan, A., Fricke, T., Diaz-Griffero, F. & Campbell, E. M. HIV-1 uncoating is facilitated by dynein and kinesin 1. *Journal of Virology* **88,** 13613–13625 (2014).

98. Dochi, T. *et al.* Phosphorylation of human immunodeficiency virus type 1 capsid protein at serine 16, required for peptidyl-prolyl isomerase-dependent uncoating, is mediated by virion incorporated extracellular signal-regulated kinase 2. *Journal of General Virology* **95,** 1156–1166 (2014).

99. Warrilow, D., Tachedjian, G. & Harrich, D. Maturation of the HIV reverse transcription complex: putting the jigsaw together. *Reviews in Medical Virology* **19,** 324–337 (2009).

100. Sarafianos, S. G. *et al.* Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *Journal of Molecular Biology* **385,** 693–713 (2009).

101. Bukrinsky, M. I. *et al.* Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. *Proceedings of the National Academy of Sciences of the United States of America* **89,** 6580–6584 (1992).

102. Heinzinger, N. K. *et al.* The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. *Proceedings of the National Academy of Sciences of the United States of America* **91,** 7311–7315 (1994).

103. Ott, M., Geyer, M. & Zhou, Q. The control of HIV transcription: keeping RNA polymerase II on track. *Cell Host & Microbe* **10,** 426–435 (2011).

104. Malim, M. H., Hauber, J., Le, S. Y., Maizel, J. V. & Cullen, B. R. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral messenger RNA. *Nature* **338,** 254–257 (1989).

105. Gheysen, D. *et al.* Assembly and release of HIV-1 precursor PR55 gag virus-like particles from recombinant baculovirus infected insect cells. *Cell* **59,** 103–112 (1989).

106. Pettit, S. C. *et al.* The p2 domain of human immunodeficiency virus type-1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *Journal of Virology* **68,** 8017–8027 (1994).

107. Von Schwedler, U. K. *et al.* Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly. *The EMBO Journal* **17,** 1555–1568 (1998).

108. Wensing, A. M., van Maarseveen, N. M. & Nijhuis, M. Fifteen years of HIV protease inhibitors: raising the barrier to resistance. *Antiviral Research* **85,** 59–74 (2010).

109. Briggs, J. A., Wilk, T., Welker, R., Kräusslich, H.-G. & Fuller, S. D. Structural organization of authentic, mature HIV-1 virions and cores. *The EMBO Journal* **22,** 1707–1715 (2003).

110. Pornillos, O. *et al.* X-ray structures of the hexameric building block of the HIV capsid. *Cell* **137,** 1282–1292 (2009).

111. Zhu, P. *et al.* Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* **441,** 847–852 (2006).

112. Cantin, R., Methot, S. & Tremblay, M. J. Plunder and stowaways: incorporation of cellular proteins by enveloped viruses. *Journal of Virology* **79,** 6577–6587 (2005).

113. Linde, M. E. *et al.* The conserved set of host proteins incorporated into HIV-1 virions suggests a common egress pathway in multiple cell types. *Journal of Proteome Research* **12,** 2045–2054 (2013).

114. Krebs, F. C., Hogan, T. H., Quiterio, S., Gartner, S. & Wigdahl, B. Lentiviral LTR-directed expression, sequence variation, and disease pathogenesis. *HIV sequence compendium 2001. Los Alamos (New Mexico): Theoretical Biology and Biophysics, Los Alamos National Laboratory,* 29–70 (2001).

115. Hiebenthal-Millow, K. *et al.* Alterations in HIV-1 LTR promoter activity during AIDS progression. *Virology* **317,** 109–118 (2003).

116. Freed, E. O. HIV-1 Gag proteins: Diverse functions in the virus life cycle. *Virology* **251,** 1–15 (1998).

117. Levin, J. G., Guo, J. H., Rouzina, I. & Musier-Forsyth, K. in *Progress in Nucleic Acid Research and Molecular Biology, Vol 80* (ed Moldave, K.) 217–286 (Elsevier Academic Press Inc, San Diego, 2005).

118. Thomas, J. A. & Gorelick, R. J. Nucleocapsid protein function in early infection processes. *Virus Research* **134,** 39–63 (2008).

119. Solbak, S. M. O. *et al.* HIV-1 p6 - a structured to flexible multifunctional membrane-interacting protein. *Biochimica Et Biophysica Acta-Biomembranes* **1828,** 816–823 (2013).

120. Kondo, E. & Gottlinger, H. G. A conserved LXXLF sequence is the major determinant in p6(Gag) required for the incorporation of human immunodeficiency virus type 1 Vpr. *Journal of Virology* **70,** 159–164 (1996).

121. Huang, M. J., Orenstein, J. M., Martin, M. A. & Freed, E. O. P6(Gag) is required for particle-production from full-length human-immunodeficiency-virus type-1 molecular clones expressing protease. *Journal of Virology* **69,** 6810–6818 (1995).

122. Kohl, N. E. *et al.* Active human immunodeficiency virus protease is required for viral infectivity. *Proceedings of the National Academy of Sciences of the United States of America* **85,** 4686–4690 (1988).

123. Kaplan, A. H., Manchester, M. & Swanstrom, R. The activity of the protease of human-immunodeficiency-virus type-1 is initiated at the membrane of infected-cells before the release of viral-proteins and is required for release to occur with maximum efficiency. *Journal of Virology* **68,** 6782–6786 (1994).

124. Beilhartz, G. L. & Gotte, M. HIV-1 ribonuclease H: structure, catalytic mechanism and inhibitors. *Viruses* **2,** 900–926 (2010).

125. Pommier, Y., Johnson, A. A. & Marchand, C. Integrase inhibitors to treat HIV/AIDS. *Nature Reviews Drug Discovery* **4,** 236–248 (2005).

126. Wang, G. P., Ciuffi, A., Leipzig, J., Berry, C. C. & Bushman, F. D. HIV integration site selection: Analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Research* **17,** 1186–1194 (2007).

127. Frost, S. D. W. *et al.* Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 18514–18519 (2005).

128. Wei, X. P. *et al.* Antibody neutralization and escape by HIV-1. *Nature* **422,** 307–312 (2003).

129. Ruben, S. *et al.* Structural and functional-characterization of human immunodeficiency virus Tat protein. *Journal of Virology* **63,** 1–8 (1989).

130. Jones, K. A. & Peterlin, B. M. Control of RNA initiation and elongation at the HIV-1 promoter. *Annual Review of Biochemistry* **63,** 717–743 (1994).

131. Frankel, A. D. & Young, J. A. T. HIV-1: Fifteen proteins and an RNA. *Annual Review of Biochemistry* **67,** 1–25 (1998).

132. Pollard, V. W. & Malim, M. H. The HIV-1 Rev protein. *Annual Review of Microbiology* **52,** 491–532 (1998).

133. Blissenbach, M., Grewe, B., Hoffmann, B., Brandt, S. & Ueberla, K. Nuclear RNA export and packaging functions of HIV-1 Rev revisited. *Journal of Virology* **84,** 6598–6604 (2010).

134. Wen, W., Meinkoth, J. L., Tsien, R. Y. & Taylor, S. S. Identification of a signal for rapid export of proteins from the nucleus. *Cell* **82,** 463–473 (1995).

135. Terwilliger, E., Sodroski, J. G., Rosen, C. A. & Haseltine, W. A. Effects of mutations within the 3' ORF open reading frame region of human T-cell lymphotropic virus type-III (HLTV-III/LAV) on replication and cytopathogenicity. *Journal of Virology* **60,** 754–760 (1986).

136. Luciw, P. A., Chengmayer, C. & Levy, J. A. Mutational analysis of the human-immunodeficiency-virus: the orf-B region down-regulates virus-replication. *Proceedings of the National Academy of Sciences of the United States of America* **84,** 1434–1438 (1987).

137. Hammes, S. R., Dixon, E. P., Malim, M. H., Cullen, B. R. & Greene, W. C. Nef protein of human immunodeficiency virus type-1 - evidence against its role as a transcriptional inhibitor. *Proceedings of the National Academy of Sciences of the United States of America* **86,** 9549–9553 (1989).

138. Lu, Y. C., Stenzel, M., Sodroski, J. G. & Haseltine, W. A. Effects of long terminal repeat mutations on human immunodeficiency virus type-1 replication. *Journal of Virology* **63,** 4115–4119 (1989).

139. Kim, S. Y., Ikeuchi, K. J., Byrn, R., Groopman, J. & Baltimore, D. Lack of a negative influence on viral growth by the Nef gene of human immunodeficiency virus type-1. *Proceedings of the National Academy of Sciences of the United States of America* **86,** 9544–9548 (1989).

140. Kirchhoff, F., Greenough, T. C., Brettler, D. B., Sullivan, J. L. & Desrosiers, R. C. Brief report - absence of intact Nef sequences in a long-term survivor with nonprogressive HIV-1 infection. *New England Journal of Medicine* **332,** 228–232 (1995).

141. Mariani, R. *et al.* High frequency of defective nef alleles in a long-term survivor with nonprogressive human immunodeficiency virus type 1 infection. *Journal of Virology* **70,** 7752–7764 (1996).

142. Salvi, R. *et al.* Grossly defective Nef gene sequences in a human immunodeficiency virus type 1-seropositive long-term nonprogressor. *Journal of Virology* **72,** 3646–3657 (1998).

143. Gorry, P. R. *et al.* Pathogenicity and immunogenicity of attenuated, Nef-deleted HIV-1 strains in vivo. *Retrovirology* **4,** e1–13 (2007).

144. Robertguroff, M. *et al.* Structure and expression of Tat-specific, Rev-specific, and Nef-specific transcripts of human immunodeficiency virus type-1 in infected lymphocytes and macrophages. *Journal of Virology* **64,** 3391–3398 (1990).

145. Klotman, M. E. *et al.* Kinetics of expression of multiply spliced RNA in early human immunodeficiency virus type-1 infection of lymphocytes and monocytes. *Proceedings of the National Academy of Sciences of the United States of America* **88,** 5011–5015 (1991).

146. Wu, Y. T. & Marsh, J. W. Selective transcription and modulation of resting T cell activity by preintegrated HIV DNA. *Science* **293,** 1503–1506 (2001).

147. Poe, J. A. & Smithgall, T. E. HIV-1 Nef dimerization is required for Nef-mediated receptor downregulation and viral replication. *Journal of Molecular Biology* **394,** 329–342 (2009).

148. Kienzle, N., Bachmann, M., Muller, W. E. G. & Mullerlantzsch, N. Expression and cellular-localization of the Nef protein from human immunodeficiency virus-1 in stably transfected B-cells. *Archives of Virology* **124,** 123–132 (1992).

149. Zhang, F. *et al.* Nef proteins from simian immunodeficiency viruses are tetherin antagonists. *Cell Host and Microbe* **6,** 54–67 (2009).

150. Schrager, J. A. & Marsh, J. W. HIV-1 Nef increases T cell activation in a stimulus-dependent manner. *Procedings of the National Academy of Sciences of the United States of America* **96,** 8167–8172 (1999).

151. Wang, J. K., Kiyokawa, E., Verdin, E. & Trono, D. The Nef protein of HIV-1 associates with rafts and primes T cells for activation. *Procedings of the National Academy of Sciences of the United States of America* **97,** 394–399 (2000).

152. Simmons, A., Aluvihare, V. & McMichael, A. Nef triggers a transcriptional program in T cells imitating single-signal T cell activation and inducing HIV virulence mediators. *Immunity* **14,** 763–777 (2001).

153. Haller, C. & Fackler, O. T. HIV-1 at the immunological and T-lymphocytic virological synapse. *Biological Chemistry* **389,** 1253–1260 (2008).

154. Neri, F., Giolo, G., Potesta, M., Petrini, S. & Doria, M. The HIV-1 Nef protein has a dual role in T cell receptor signaling in infected CD4+ T lymphocytes. *Virology* **410,** 316–326 (2011).

155. Herbein, G. *et al.* Nef and TNFalpha are coplayers that favor HIV-1 replication in monocytic cells and primary macrophages. *Current HIV Research* **6,** 117–129 (2008).

156. Katsikis, P. D., Wunderlich, E. S., Smith, C. A., Herzenberg, L. A. & Herzenberg, L. A. Fas antigen stimulation induces marked apoptosis of T lymphocytes in human immunodeficiency virus-infected individuals. *Journal of Experimental Medicine* **181,** 2029–2036 (1995).

157. Xu, X. N. *et al.* Evasion of cytotoxic T lymphocyte (CTL) responses by nef-dependent induction of Fas ligand (CD95L) expression on simian immunodeficiency virus-infected cells. *Journal of Experimental Medicine* **186,** 7–16 (1997).

158. Lenassi, M. *et al.* HIV Nef is secreted in exosomes and triggers apoptosis in bystander CD4+ T cells. *Traffic* **11,** 110–122 (2010).

159. Azad, A. A. Could nef and vpr proteins contribute to disease progression by promoting depletion of bystander cells and prolonged survival of HIV-infected cells? *Biochemical and Biophysical Research Communications* **267,** 677–685 (2000).

160. Papkalla, A., Munch, J., Otto, C. & Kirchhoff, F. Nef enhances human immunodeficiency virus type 1 infectivity and replication independently of viral coreceptor tropism. *Journal of Virology* **76,** 8455–8459 (2002).

161. Miller, M. D., Warmerdam, M. T., Gaston, I., Greene, W. C. & Feinberg, M. B. The human immunodeficiency virus-1 Nef gene-product - a positive factor for viral-infection and replication in primary lymphocytes and macrophages. *Journal of Experimental Medicine* **179,** 101–113 (1994).

162. Chowers, M., Pandori, M., Spina, C., Richman, D. & Guatelli, J. The growth advantage conferred by HIV-1 nef is determined at the level of viral DNA formation and is independent of CD4 downregulation. *Virology* **212,** 451–457 (1995).

163. Miller, M. D., Warmerdam, M. T., Page, K. A., Feinberg, M. B. & Greene, W. C. Expression of the human immunodeficiency virus type 1 (HIV-1) nef gene during HIV-1 production increases progeny particle infectivity independently of gp160 or viral entry. *Journal of Virology* **69,** 579–584 (1995).

164. Qi, M. & Aiken, C. Nef enhances HIV-1 infectivity via association with the virus assembly complex. *Virology* **373,** 287–297 (2008).

165. Aiken, C. & Trono, D. Nef stimulates human-immunodeficiency-virus type-1 proviral DNA-synthesis. *Journal of Virology* **69,** 5048–5056 (1995).

166. Welker, R., Kottler, H., Kalbitzer, H. R. & Krausslich, H. G. Human immunodeficiency virus type 1 Nef protein is incorporated into virus particles and specifically cleaved by the viral proteinase. *Virology* **219,** 228–236 (1996).

167. Usami, Y., Wu, Y. & Gottlinger, H. G. SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. *Nature* **526,** 218–223 (2015).

168. Rosa, A. *et al.* HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation. *Nature* **526,** 212–217 (2015).

169. Garcia, J. V. & Miller, A. D. Serine phosphorylation independent down regulation of cell-surface CD4 by Nef. *Nature* **350,** 508–511 (1991).

170. Stove, V. *et al.* Human immunodeficiency virus Nef induces rapid internalization of the T-cell coreceptor CD8alphabeta. *Journal of Virology* **79,** 11422–11433 (2005).

171. Hrecka, K., Swigut, T., Schindler, M., Kirchhoff, F. & Skowronski, J. Nef proteins from diverse groups of primate lentiviruses downmodulate CXCR4 to inhibit migration to the chemokine stromal derived factor 1. *Journal of Virology* **79,** 10650–10659 (2005).

172. Heigele, A. *et al.* Down-modulation of CD8 alpha beta is a fundamental activity of primate lentiviral Nef proteins. *Journal of Virology* **86,** 36–48 (2012).

173. Haller, C. *et al.* HIV-1 Nef and Vpu are functionally redundant broad-spectrum modulators of cell surface receptors, including tetraspanins. *Journal of Virology* **88,** 14241–14257 (2014).

174. Pauza, C. D., Galindo, J. E. & Richman, D. D. Reinfection results in accumulation of unintegrated viral-DNA in cytopathic and persistent human-immunodeficiency-virus type-1 infection of CEM cells. *Journal of Experimental Medicine* **172,** 1035–1042 (1990).

175. Benson, R. E., Sanfridson, A., Ottinger, J. S., Doyle, C. & Cullen, B. R. Downregulation of cell-surface CD4 expression by simian immunodeficiency virus Nef prevents viral super infection. *Journal of Experimental Medicine* **177,** 1561–1566 (1993).

176. Watkins, R. L. *et al.* In vivo analysis of highly conserved Nef activities in HIV-1 replication and pathogenesis. *Retrovirology* **10,** e1–16 (2013).

177. Hanna, Z., Priceputu, E., Hu, C. Y., Vincent, P. & Jolicoeur, P. HIV-1 Nef mutations abrogating downregulation of CD4 affect other Nef functions and show reduced pathogenicity in transgenic mice. *Virology* **346,** 40–52 (2006).

178. Arganaraz, E. R., Schindler, M., Kirchhoff, F., Cortes, M. J. & Lama, J. Enhanced CD4 down-modulation by late stage HIV-1 nef alleles is associated with increased env incorporation and viral replication. *Journal of Biological Chemistry* **278,** 33912–33919 (2003).

179. Laguette, N., Benichou, S. & Basmaciogullari, S. Human immunodeficiency virus type 1 Nef incorporation into virions does not increase infectivity. *Journal of Virology* **83,** 1093–1104 (2009).

180. Pelchenmatthews, A., Armes, J. E., Griffiths, G. & Marsh, M. Differential endocytosis of CD4 in lymphocytic and nonlymphocytic cells. *Journal of Experimental Medicine* **173,** 575–587 (1991).

181. Pelchenmatthews, A., Boulet, I., Littman, D. R., Fagard, R. & Marsh, M. The protein tyrosine kinase p56(Lck) inhibits CD4 endocytosis by preventing entry of CD4 into coated pits. *Journal of Cell Biology* **117,** 279–290 (1992).

182. Pelchenmatthews, A., Parsons, I. J. & Marsh, M. Phorbol ester-induced down-regulation of CD4 is a multistep process involving dissociation from p56(Lck), increased association with clathrin-coated pits, and altered endosomal sorting. *Journal of Experimental Medicine* **178,** 1209–1222 (1993).

183. Sleckman, B. P. *et al.* Disruption of the CD4-p56(Lck) complex is required for rapid internalization of CD4. *Proceedings of the National Academy of Sciences of the United States of America* **89,** 7566–7570 (1992).

184. Yoshida, H., Koga, Y., Nakamura, K., Kimura, G. & Nomoto, K. A lymphocyte-specific protein tyrosine kinase, p56 Lck, regulates the PMA-induced internalization of CD4. *Biochimica Et Biophysica Acta* **1137,** 321–330 (1992).

185. Aiken, C., Konner, J., Landau, N. R., Lenburg, M. E. & Trono, D. Nef induces CD4 endocytosis - requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain. *Cell* **76,** 853–864 (1994).

186. Kim, Y. H., Chang, S. H., Kwon, J. H. & Rhee, S. S. HIV-1 Nef plays an essential role in two independent processes in CD4 down-regulation: Dissociation of the CD4-p56(lck) complex and targeting of CD4 to lysosomes. *Virology* **257,** 208–219 (1999).

187. Rose, J. J. *et al.* CD4 down-regulation by HIV-1 and simian immunodeficiency virus (SIV) Nef proteins involves both internalization and intracellular retention mechanisms. *Journal of Biological Chemistry* **280,** 7413–7426 (2005).

188. Popov, S. *et al.* Viral protein R regulates nuclear import of the HIV-1 pre-integration complex. *The EMBO Journal* **17,** 909–917 (1998).

189. Connor, R. I., Chen, B. K., Choe, S. & Landau, N. R. Vpr is required for efficient replication of human immunodeficiency virus type-1 in mononuclear phagocytes. *Virology* **206,** 935–944 (1995).

190. Hattori, N. *et al.* The human immunodeficiency virus type-2 vpr gene is essential for productive infection of human macrophages. *Proceedings of the National Academy of Sciences of the United States of America* **87,** 8080–8084 (1990).

191. Kogan, M. & Rappaport, J. HIV-1 accessory protein Vpr: relevance in the pathogenesis of HIV and potential for therapeutic intervention. *Retrovirology* **8,** e1–20 (2011).

192. Li, G., Park, H. U., Liang, D. & Zhao, R. Y. Cell cycle G2/M arrest through an S phase-dependent mechanism by HIV-1 viral protein R. *Retrovirology* **7,** e1–18 (2010).

193. Ayyavoo, V. *et al.* HIV-1 Vpr suppresses immune activation and apoptosis through regulation of nuclear factor kappa B. *Nature Medicine* **3,** 1117–1123 (1997).

194. Kerkau, T. *et al.* The human immunodeficiency virus type 1 (HIV-1) vpu protein interferes with an early step in the biosynthesis of major histocompatibility complex (MHC) class I molecules. *Journal of Experimental Medicine* **185,** 1295–1305 (1997).

195. Willey, R. L., Maldarelli, F., Martin, M. A. & Strebel, K. Human immunodeficiency virus type-I vpu protein induces rapid degradation of CD4. *Journal of Virology* **66,** 7193–7200 (1992).

196. Apps, R. *et al.* HIV-1 Vpu Mediates HLA-C Downregulation. *Cell Host and Microbe* **19,** 686–695 (2016).

197. Sauter, D. *et al.* Differential regulation of NF-kappaB-mediated proviral and antiviral host gene expression by primate lentiviral Nef and Vpu proteins. *Cell Reports* **10,** 586–599 (2015).

198. Neil, S. J. D., Zang, T. & Bieniasz, P. D. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* **451,** 425–430 (2008).

199. Douglas, J. L. *et al.* Vpu directs the degradation of the human immunodeficiency virus restriction factor BST-2/tetherin via a beta TrCP-dependent mechanism. *Journal of Virology* **83,** 7931–7947 (2009).

200. Tristem, M., Marshall, C., Karpas, A. & Hill, F. Evolution of the primate lentiviruses - evidence from vpx and vpr. *The EMBO Journal* **11,** 3405–3412 (1992).

201. Ahn, J. *et al.* HIV/simian immunodeficiency virus (SIV) accessory virulence factor Vpx loads the host cell restriction factor SAMHD1 onto the E3 ubiquitin ligase complex CRL4(DCAF1). *Journal of Biological Chemistry* **287,** 12550–12558 (2012).

202. Cheng, X. & Ratner, L. HIV-2 Vpx protein interacts with interferon regulatory factor 5 (IRF5) and inhibits its function. *Journal of Biological Chemistry* **289,** 9146–9157 (2014).

203. Yu, X. *et al.* Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **302,** 1056–1060 (2003).

204. Sheehy, A. M., Gaddis, N. C. & Malim, M. H. The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nature Medicine* **9,** 1404–1407 (2003).

205. Mercenne, G. *et al.* HIV-1 Vif binds to APOBEC3G mRNA and inhibits its translation. *Nucleic Acids Research* **38,** 633–646 (2010).

206. Kao, S. *et al.* The human immunodeficiency virus type 1 Vif protein reduces intracellular expression and inhibits packaging of APOBEC3G (CEM15), a cellular inhibitor of virus infectivity. *Journal of Virology* **77,** 11398–11407 (2003).

207. Borrow, P., Shattock, R. J., Vyakarnam, A. & Grp, E. W. Innate immunity against HIV: a priority target for HIV prevention research. *Retrovirology* **7,** e1–17 (2010).

208. Keele, B. F. *et al.* Identification and characterisation of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America* **105,** 7552–7557 (2008).

209. Keele, B. F. *et al.* Low-dose rectal inoculation of rhesus macaques by SIVsm E660 or SIVmac 251 recapitulates human mucosal infection by HIV-1. *Journal of Experimental Medicine* **206,** 1117–1134 (2009).

210. Haase, A. T. Targeting early infection to prevent HIV-1 mucosal transmission. *Nature* **464,** 217–223 (2010).

211. Iqbal, S. M. *et al.* Elevated elafin/trappin-2 in the female genital tract is associated with protection against HIV acquisition. *AIDS* **23,** 1669–1677 (2009).

212. Jana, N. K., Gray, L. R. & Shugars, D. C. Human immunodeficiency virus type I stimulates the expression and production of secretory leukocyte protease inhibitor (SLPI) in oral epithelial cells: a role for SLPI in innate mucosal immunity. *Journal of Virology* **79,** 6432–6440 (2005).

213. Nakashima, H., Yamamoto, N., Masuda, M. & Fujii, N. Defensins inhibit HIV replication in vitro. *AIDS* **7,** 1129–1129 (1993).

214. Muench, J. *et al.* Semen-derived amyloid fibrils drastically enhance HIV infection. *Cell* **131,** 1059–1071 (2007).

215. Roan, N. R. *et al.* The cationic properties of SEVI underlie its ability to enhance human immunodeficiency virus infection. *Journal of Virology* **83,** 73–80 (2009).

216. Alvarez, R. *et al.* WFDC1/ps20 is a novel innate immunomodulatory signature protein of human immunodeficiency virus (HIV)-permissive CD4(+) CD45RO(+) memory T cells that promotes infection by upregulating CD54 integrin expression and is elevated in HIV type 1 infection. *Journal of Virology* **82,** 471–486 (2008).

217. Burgener, A., McGowan, I. & Klatt, N. R. HIV and mucosal barrier interactions: consequences for transmission and pathogenesis. *Current Opinion in Immunology* **36,** 22–30 (2015).

218. Akira, S., Uematsu, S. & Takeuchi, O. Pathogen recognition and innate immunity. *Cell* **124,** 783–801 (2006).

219. Gringhuis, S. I. *et al.* HIV-1 exploits innate signaling by TLR8 and DC-SIGN for productive infection of dendritic cells. *Nature Immunology* **11,** 419–426 (2010).

220. Sanders, C. M., Cruse, J. M. & Lewis, R. E. Toll-like receptors, cytokines and HIV-1. *Experimental and Molecular Pathology* **84,** 31–36 (2008).

221. Heil, F. *et al.* Species-specific recognition of single-stranded RNA via Toll-like receptor 7 and 8. *Science* **303,** 1526–1529 (2004).

222. Bochud, P. Y. *et al.* Polymorphisms in Toll-like receptor 9 influence the clinical course of HIV-1 infection. *AIDS* **21,** 441–446 (2007).

223. Mogensen, T. H., Melchjorsen, J., Larsen, C. S. & Paludan, S. R. Innate immune recognition and activation during HIV infection. *Retrovirology* **7,** e1–19 (2010).

224. Curtis, B. M., Scharnowske, S. & Watson, A. J. Sequence and expression of a membrane-associated C-type lectin that exhibits CD4-independent binding of human immunodeficiency virus envelope glycoprotein gp120. *Proceedings of the National Academy of Sciences of the United States of America* **89,** 8356–8360 (1992).

225. Geijtenbeek, T. B. H., Engering, A. & van Kooyk, Y. DC-SIGN, a C-type lectin on dendritic cells that unveils many aspects of dendritic cell biology. *Journal of Leukocyte Biology* **71,** 921–931 (2002).

226. Cella, M., Sallusto, F. & Lanzavecchia, A. Origin, maturation and antigen presenting function of dendritic cells. *Current Opinion in Immunology* **9,** 10–16 (1997).

227. Hu, J., Gardner, M. B. & Miller, C. J. Simian immunodeficiency virus rapidly penetrates the cervicovaginal mucosa after intravaginal inoculation and infects intraepithelial dendritic cells. *Journal of Virology* **74,** 6087–6095 (2000).

228. Reis e Sousa, C. Dendritic cells in a mature age. *Nature Reviews Immunology* **6,** 476–483 (2006).

229. Geijtenbeek, T. B. H. *et al.* DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell* **100,** 587–597 (2000).

230. Arrighi, J. F. *et al.* DC-SIGN-mediated infectious synapse formation enhances X4 HIV-1 transmission from dendritic cells to T cells. *Journal of Experimental Medicine* **200,** 1279–1288 (2004).

231. Reizis, B., Bunin, A., Ghosh, H. S., Lewis, K. L. & Sisirak, V. Plasmacytoid dendritic cells: recent progress and open questions. *Annual Review of Immunology* **29,** 163–183 (2011).

232. Liu, Y. J. IPC: professional type 1 interferon-producing cells and plasmacytoid dendritic cell precursors. *Annual Review of Immunology* **23,** 275–306 (2005).

233. Hornung, V. *et al.* Sequence-specific potent induction of IFN-alpha by short interfering RNA in plasmacytoid dendritic cells through TLR7. *Nature Medicine* **11,** 263–270 (2005).

234. Ito, T., Wang, Y. H. & Liu, Y. J. Plasmacytoid dendritic cell precursors/type I interferon-producing cells sense viral infection by Toll-like receptor (TLR) 7 and TLR9. *Springer Seminars in Immunopathology* **26,** 221–229 (2005).

235. Beignon, A. S. *et al.* Endocytosis of HIV-1 activates plasmacytoid dendritic cells via toll-like receptor-viral RNA interactions. *Journal of Clinical Investigation* **115,** 3265–3275 (2005).

236. Martinelli, E. *et al.* HIV-1 gp120 inhibits TLR9-mediated activation and IFN-alpha secretion in plasmacytoid dendritic cells. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 3396–3401 (2007).

237. Mandl, J. N. *et al.* Divergent TLR7 and TLR9 signaling and type I interferon production distinguish pathogenic and nonpathogenic AIDS virus infections. *Nature Medicine* **14,** 1077–1087 (2008).

238. Boasso, A. *et al.* HIV inhibits CD4(+) T-cell proliferation by inducing indoleamine 2,3-dioxygenase in plasmacytoid dendritic cells. *Blood* **109,** 3351–3359 (2007).

239. Manches, O. *et al.* HIV activated human pDCs induce Tregs through an indoleamine 2,3 dioxygenase dependent mechanism. *Journal of Clinical Investigation* **118,** 3431–3439 (2008).

240. O'Brien, M., Manches, O. & Bhardwaj, N. Plasmacytoid dendritic cells in HIV infection. *Advances in Experimental Medicine and Biology* **762,** 71–107 (2013).

241. Li, G. *et al.* Plasmacytoid DCs suppress HIV-1 replication but contribute to HIV-1 induced immunopathogenesis in humanized mice. *PLoS Pathogens* **10,** e1004291 (2014).

242. Shen, R. *et al.* Macrophages in vaginal but not intestinal mucosa are monocyte-like and permissive to human immunodeficiency virus type 1 infection. *Journal of Virology* **83,** 3258–3267 (2009).

243. Swingler, S. *et al.* HIV-1 Nef mediates lymphocyte chemotaxis and activation by infected macrophages. *Nature Medicine* **5,** 997–1003 (1999).

244. Swingler, S. *et al.* HIV-1 Nef intersects the macrophage CD40L signalling pathway to promote resting-cell infection. *Nature* **424,** 213–219 (2003).

245. Herbein, G., Gras, G., Khan, K. A. & Abbas, W. Macrophage signaling in HIV-1 infection. *Retrovirology* **7,** e1–13 (2010).

246. Groot, F., Welsch, S. & Sattentau, Q. J. Efficient HIV-1 transmission from macrophages to T cells across transient virological synapses. *Blood* **111,** 4660–4663 (2008).

247. Koppensteiner, H., Brack-Werner, R. & Schindler, M. Macrophages and their relevance in Human Immunodeficiency Virus Type I infection. *Retrovirology* **9,** e1–11 (2012).

248. Sharova, N., Swingler, C., Sharkey, M. & Stevenson, M. Macrophages archive HIV-1 virions for dissemination in trans. *The EMBO Journal* **24,** 2481–2489 (2005).

249. Martin, M. P. *et al.* Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nature Genetics* **31,** 429–434 (2002).

250. Alter, G. *et al.* Differential natural killer cell-mediated inhibition of HIV-1 replication based on distinct KIR/HLA subtypes. *Journal of Experimental Medicine* **204,** 3027–3036 (2007).

251. Cooper, M. A., Fehniger, T. A. & Caligiuri, M. A. The biology of human natural killer-cell subsets. *Trends in Immunology* **22,** 633–640 (2001).

252. Hu, P. F. *et al.* Natural killer cell immunodeficiency in HIV disease is manifest by profoundly decreased numbers of CD16+CD56+ cells and expansion of a population of CD16dimCD56- cells with low lytic activity. *Journal of Acquired Immune Deficiency Syndrome and Human Retrovirology* **10,** 331–340 (1995).

253. Tarazona, R. *et al.* Selective Depletion of CD56dim NK Cell Subsets and Maintenance of CD56bright NK Cells in Treatment-Naive HIV-1-Seropositive Individuals. *Journal of Clinical Immunology* **22,** 176–183 (2002).

254. Mavilio, D. *et al.* Characterization of CD56(-)/CD16(+) natural killer (NK) cells: A highly dysfunctional NK subset expanded in HIV-infected viremic individuals. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 2886–2891 (2005).

255. Bandyopadhyay, S. *et al.* Natural killer cell-mediated lysis of T cell lines chronically infected with HIV-1. *Clinical and Experimental Immunology* **79,** 430–435 (1990).

256. Ahmad, A. & Menezes, J. Antibody-dependent cellular cytotoxicity in HIV infections. *The FASEB Journal* **10,** 258–266 (1996).

257. Chung, A., Rollman, E., Johansson, S., Kent, S. J. & Stratov, I. The utility of ADCC responses in HIV infection. *Current HIV Research* **6,** 515–519 (2008).

258. Fehniger, T. A. *et al.* Natural killer cells from HIV-1(+) patients produce C-C chemokines and inhibit HIV-1 infection. *Journal of Immunology* **161,** 6433–6438 (1998).

259. Oliva, A. *et al.* Natural killer cells from human immunodeficiency virus (HIV)-infected individuals are an important source of CC-chemokines and suppress HIV-1 entry and replication in vitro. *Journal of Clinical Investigation* **102,** 223–231 (1998).

260. Li, D. & Xu, X.-N. NKT cells in HIV-1 infection. *Cell Research* **18,** 817–822 (2008).

261. Fernandez, C. S., Kelleher, A. D., Finlayson, R., Godfrey, D. I. & Kent, S. J. NKT cell depletion in humans during early HIV infection. *Immunology & Cell Biology* **92,** 578–590 (2014).

262. Carding, S. R. & Egan, P. J. Gamma Delta T cells: functional plasticity and heterogeneity. *Nature Reviews Immunology* **2,** 336–345 (2002).

263. Kazen, A. R. & Adams, E. J. Evolution of the V, D, and J gene segments used in the primate gamma delta T-cell receptor reveals a dichotomy of conservation and diversity. *Proceedings of the National Academy of Sciences of the United States of America* **108,** E332–E340 (2011).

264. Asarnow, D. M. *et al.* Limited diversity of gamma delta antigen receptor genes of thy-1+ dendritic epidermal cells. *Cell* **55,** 837–847 (1988).

265. Girardi, M. Immunosurveillance and Immunoregulation by gamma delta T Cells. *Journal of Investigative Dermatology* **126,** 25–31 (2006).

266. Vantourout, P. & Hayday, A. Six-of-the-best: unique contributions of gammadelta T cells to immunology. *Nature Review Immunology* **13,** 88–100 (2013).

267. Autran, B. *et al.* T cell receptor gamma/delta+ lymphocyte subsets during HIV infection. *Clinical and Experimental Immunology* **75,** 206–210 (1989).

268. Depaoli, P. *et al.* A subset of gamma-delta lymphocytes is increased during HIV-1 infection. *Clinical and Experimental Immunology* **83,** 187–191 (1991).

269. Demaria, A. *et al.* Selective increase of a subset of T cell receptor gamma-delta lymphocytes in the peripheral blood of patients with human immunodeficiency virus type-1 infection. *Journal of Infectious Diseases* **165,** 917–919 (1992).

270. Hermier, F. *et al.* Decreased blood TcR gamma delta+ lymphocytes in AIDS and p24-antigenemic HIV-1-infected patients. *Clinical Immunology and Immunopathology* **69,** 248–250 (1993).

271. Hinz, T. *et al.* T-cell receptor-gamma-delta repertoire in HIV-1-infected individuals. *European Journal of Immunology* **24,** 3044–3049 (1994).

272. Li, H. *et al.* Association between V gamma 2 / V delta 2 T cells and disease progression after infection with closely related strains of HIV in China. *Clinical Infectious Diseases* **46,** 1466–1472 (2008).

273. Poccia, F. *et al.* Peripheral V gamma 9/V delta 2 T cell deletion and anergy to nonpeptidic mycobacterial antigens in asymptomatic HIV-1-infected persons. *The Journal of Immunology* **157,** 449–461 (1996).

274. Ricklin, D., Hajishengallis, G., Yang, K. & Lambris, J. D. Complement: a key system for immune surveillance and homeostasis. *Nature Immunology* **11,** 785–797 (2010).

275. Hirsch, R. L. The complement system: its importance in the host response to viral infection. *Microbiological Reviews* **46,** 71–85 (1982).

276. Janeway, C., Travers, P., Walport, M. & Shlomchik, M. *Immunobiology* (Garland Science, 2005).

277. Senaldi, G. *et al.* Activation of the complement system in human immunodeficiency virus infection: relevance of the classical pathway to pathogenesis and disease severity. *Journal of Infectious Diseases* **162,** 1227–1232 (1990).

278. Ebenbichler, C. F. *et al.* Human immunodeficiency virus type 1 activates the classical pathway of complement by direct C1 binding through specific sites in the transmembrane glycoprotein gp41. *Journal of Experimental Medicine* **174,** 1417–1424 (1991).

279. Haurum, J. S. *et al.* Complement activation upon binding of mannan-binding protein to HIV envelope glycoproteins. *AIDS* **7,** 1307–1314 (1993).

280. Aasa-Chapman, M. I. *et al.* Detection of antibody-dependent complement-mediated inactivation of both autologous and heterologous virus in primary human immunodeficiency virus type 1 infection. *Journal of Virology* **79,** 2823–2830 (2005).

281. Sullivan, B. L. *et al.* Susceptibility of HIV-1 plasma virus to complement-mediated lysis. Evidence for a role in clearance of virus in vivo. *Journal of Immunology* **157,** 1791–1798 (1996).

282. Huber, M. *et al.* Complement lysis activity in autologous plasma is associated with lower viral loads during the acute phase of HIV-1 infection. *PLoS Medicine* **3,** e441 (2006).

283. Bajtay, Z., Speth, C., Erdei, A. & Dierich, M. P. Cutting edge: productive HIV-1 infection of dendritic cells via complement receptor type 3 (CR3, CD11b / CD18). *Journal of Immunology* **173,** 4775–4778 (2004).

284. Bouhlal, H. *et al.* Opsonization of HIV with complement enhances infection of dendritic cells and viral transfer to CD4 T cells in a CR3 and DC-SIGN-dependent manner. *Journal of Immunology* **178,** 1086–1095 (2007).

285. Schmitz, J. *et al.* Antibody-dependent complement-mediated cytotoxicity in sera from patients with HIV-1 infection is controlled by CD55 and CD59. *Journal of Clinical Investigation* **96,** 1520–1526 (1995).

286. Saifuddin, M. *et al.* Human immunodeficiency virus type 1 incorporates both glycosyl phosphatidylinositol-anchored CD55 and CD59 and integral membrane CD46 at levels that protect from complement-mediated destruction. *Journal of General Virology* **78,** 1907–1911 (1997).

287. Stoiber, H., Schneider, R., Janatova, J. & Dierich, M. P. Human complement proteins C3b, C4b, factor H and properdin react with specific sites in gpl20 and gp4l, the envelope proteins of HIV1. *Immunobiology* **193,** 98–113 (1995).

288. Stoiber, H., Pinter, C., Siccardi, A. G., Clivio, A. & Dierich, M. P. Efficient destruction of human immunodeficiency virus in human serum by inhibiting the protective action of complement factor H and decay accelerating factor (DAF, CD55). *The Journal of Experimental Medicine* **183,** 307–310 (1996).

289. Malim, M. H. & Bieniasz, P. D. HIV restriction factors and mechanisms of evasion. *Cold Spring Harbour Perspectives in Medicine* **2,** a006940 (2012).

290. Duggal, N. K. & Emerman, M. Evolutionary conflicts between viruses and restriction factors shape immunity. *Nature Reviews Immunology* **12,** 687–695 (2012).

291. Stremlau, M. *et al.* Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5 restriction factor. **103,** 5514–5519 (2006).

292. Stremlau, M. *et al.* The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature* **427,** 848–853 (2004).

293. Battivelli, E. *et al.* Strain-specific differences in the impact of human TRIM5 alpha, different TRIM5 alpha alleles, and the inhibition of capsid-cyclophilin A interactions on the infectivity of HIV-1. *Journal of Virology* **84,** 11010–11019 (2010).

294. Harris, R. S. *et al.* DNA deamination mediates innate immunity to retroviral infection. *Cell* **113,** 803–809 (2003).

295. Mangeat, B. *et al.* Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424,** 99–103 (2003).

296. Guo, F., Cen, S., Niu, M., Saadatmand, J. & Kleiman, L. Inhibition of tRNA Lys primed reverse transcription by human APOBEC3G during human immunodeficiency virus type 1 replication. *Journal of Virology* **80,** 11710–11722 (2006).

297. Bishop, K. N., Verma, M., Kim, E.-Y., Wolinsky, S. M. & Malim, M. H. APOBEC3G inhibits elongation of HIV-1 reverse transcripts. *PLoS Pathogens* **4,** e1000231 (2008).

298. Lahouassa, H. *et al.* SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nature Immunology* **13,** 223–228 (2012).

299. Perez-Caballero, D. *et al.* Tetherin inhibits HIV-1 release by directly tethering virions to cells. *Cell* **139,** 499–511 (2009).

300. Hammonds, J., Wang, J.-J., Yi, H. & Spearman, P. Immunoelectron microscopic evidence for tetherin/BST2 as the physical bridge between HIV-1 virions and the plasma membrane. *PLoS Pathogens* **6,** e1000749 (2010).

301. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418,** 646–650 (2002).

302. Laguette, N. *et al.* SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature* **474,** 654–657 (2011).

303. Hrecka, K. *et al.* Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature* **474,** 658–661 (2011).

304. Gupta, R. K. *et al.* Simian immunodeficiency virus envelope glycoprotein counteracts tetherin/BST-2/CD317 by intracellular sequestration. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 20889–20894 (2009).

305. Yewdell, J. W. & Bennink, J. R. Cell biology of antigen-processing and presentation to major histocompatibility complex class-I molecule-restricted T-lymphocytes. *Advances in Immunology* **52,** 1–123 (1992).

306. Basler, M., Kirk, C. J. & Groettrup, M. The immunoproteasome in antigen processing and other immunological functions. *Current Opinion in Immunology* **25,** 74–80 (2013).

307. Elliott, T. Transporter associated with antigen processing. *Advances in Immunology* **65,** 47–109 (1997).

308. York, I. A. & Rock, K. L. Antigen processing and presentation by the class I major histocompatibility complex. *Annual Review of Immunology* **14,** 369–396 (1996).

309. Weimershaus, M., Evnouchidou, I., Saveanu, L. & van Endert, P. Peptidases trimming MHC class I ligands. *Current Opinion in Immunology* **25,** 90–96 (2013).

310. Shen, X. Z. *et al.* The carboxypeptidase ACE shapes the MHC class I peptide repertoire. *Nature Immunology* **12,** 1078–1085 (2011).

311. Jackson, M. R., Cohendoyle, M. F., Peterson, P. A. & Williams, D. B. Regulation of MHC class-I transport by the molecular chaperone, calnexin (p88, ip90). *Science* **263,** 384–387 (1994).

312. Sadasivan, B., Lehner, P. J., Ortmann, B., Spies, T. & Cresswell, P. Roles for calreticulin and a novel glycoprotein, tapasin, in the interaction of MHC class I molecules with TAP. *Immunity* **5,** 103–114 (1996).

313. Hughes, E. A. & Cresswell, P. The thiol oxidoreductase ERp57 is a component of the MHC class I peptide-loading complex. *Current Biology* **8,** 709–712 (1998).

314. Morrice, N. A. & Powis, S. J. A role for the thiol-dependent reductase ERp57 in the assembly of MHC class I molecules. *Current Biology* **8,** 713–716 (1998).

315. Kobayashi, K. S. & van den Elsen, P. J. NLRC5: a key regulator of MHC class I-dependent immune responses. *Nature Reviews Immunology* **12,** 813–820 (2012).

316. Carrington, M. *et al.* HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* **283,** 1748–1752 (1999).

317. Altfeld, M. *et al.* Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection. *AIDS* **17,** 2581–2591 (2003).

318. Peterson, T. A. *et al.* HLA class I associations with rates of HIV-1 seroconversion and disease progression in the Pumwani Sex Worker Cohort. *Tissue Antigens* **81,** 93–107 (2013).

319. Andrews, S. M., Wong, G. & Rowland-Jones, S. L. in *Encyclopedia of AIDS* (eds Hope, T. J., Stevenson, M. & Richman, D.) 1–10 (Springer New York, 2015).

320. Paul, W. E. & Seder, R. A. Lymphocyte-responses and cytokines. *Cell* **76,** 241–251 (1994).

321. Stinchcombe, J. C., Bossi, G., Booth, S. & Griffiths, G. M. The immunological synapse of CTL contains a secretory domain and membrane bridges. *Immunity* **15,** 751–761 (2001).

322. Podack, E. R., Young, J. D. E. & Cohn, Z. A. Isolation and biochemical and functional-characterization of perforin-1 from cytolytic T-cell granules. *Proceedings of the National Academy of Sciences of the United States of America* **82,** 8629–8633 (1985).

323. Masson, D. & Tschopp, J. A family of serine esterases in lytic granules of cytolytic lymphocytes-T. *Cell* **49,** 679–685 (1987).

324. Lowin, B., Hahne, M., Mattmann, C. & Tschopp, J. Cytolytic T-cell cytotoxicity is mediated through perforin and Fas lytic pathways. *Nature* **370,** 650–652 (1994).

325. Lopez, J. A. *et al.* Rapid and unidirectional perforin pore delivery at the cytotoxic immune synapse. *Journal of Immunology* **191,** 2328–2334 (2013).

326. Lopez, J. A. *et al.* Perforin forms transient pores on the target cell plasma membrane to facilitate rapid access of granzymes during killer cell attack. *Blood* **121,** 2659–2668 (2013).

327. Heusel, J. W., Wesselschmidt, R. L., Shresta, S., Russell, J. H. & Ley, T. J. Cytotoxic lymphocytes require granzyme-B for the rapid induction of DNA fragmentation and apoptosis in allogeneic target-cells. *Cell* **76,** 977–987 (1994).

328. Nagata, S. & Golstein, P. The Fas death factor. *Science* **267,** 1449–1456 (1995).

329. Scott, F. L. *et al.* The Fas-FADD death domain complex structure unravels signalling by receptor clustering. *Nature* **457,** 1019–1022 (2009).

330. Chinnaiyan, A. M., Orourke, K., Tewari, M. & Dixit, V. M. FADD, a novel death domain-containing protein, interacts with the death domain of FAS and initiates apoptosis. *Cell* **81,** 505–512 (1995).

331. Kischkel, F. C. *et al.* Cytotoxicity-dependent Apo-1 (Fas/CD95)-associated proteins form a death-inducing signaling complex (DISC) with the receptor. *The EMBO Journal* **14,** 5579–5588 (1995).

332. Cohen, G. M. Caspases: the executioners of apoptosis. *Biochemical Journal* **326,** 1–16 (1997).

333. Migueles, S. A. & Connors, M. Success and failure of the cellular immune response against HIV-1. *Nature Immunology* **16,** 563–570 (2015).

334. Koup, R. A. *et al.* Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *Journal of Virology* **68,** 4650–4655 (1994).

335. Schmitz, J. E. *et al.* Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes. *Science* **283,** 857–860 (1999).

336. Jin, X. *et al.* Dramatic rise in plasma viremia after CD8(+) T cell depletion in simian immunodeficiency virus-infected macaques. *Journal of Experimental Medicine* **189,** 991–998 (1999).

337. McMichael, A. J. & Rowland-Jones, S. L. Cellular immune responses to HIV. *Nature* **410,** 980–987 (2001).

338. Goonetilleke, N. *et al.* The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *The Journal of Experimental Medicine* **206,** 1253–1272 (2009).

339. Leslie, A. J. *et al.* HIV evolution: CTL escape mutation and reversion after transmission. *Nature Medicine* **10,** 282–289 (2004).

340. Fryer, H. R. *et al.* Cytotoxic T-lymphocyte escape mutations identified by HLA association favor those which escape and revert rapidly. *Journal of Virology* **86,** 8568–8580 (2012).

341. Walker, C. M., Erickson, A. L., Hsueh, F. C. & Levy, J. A. Inhibition of human-immunodeficiency-virus replication in acutely infected CD4+ cells by CD8+ cells involves a noncytotoxic mechanism. *Journal of Virology* **65,** 5921–5927 (1991).

342. Levy, J. A., Mackewicz, C. E. & Barker, E. Controlling HIV pathogenesis: The role of the noncytotoxic anti-HIV response of CD8(+) T cells. *Immunology Today* **17,** 217–224 (1996).

343. Cocchi, F. *et al.* Higher macrophage inflammatory protein (MIP)-1-alpha and MIP-1-beta levels from CD8(+) T cells are associated with asymptomatic HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America* **97,** 13812–13817 (2000).

344. Ahmed, R. K. S. *et al.* Beta-chemokine production in macaques vaccinated with live attenuated virus correlates with protection against simian immunodeficiency virus (SIVsm) challenge. *Journal of General Virology* **80,** 1569–1574 (1999).

345. Ahmed, R. K. S. *et al.* Spontaneous production of RANTES and antigen-specific IFN-gamma production in macaques vaccinated with SHIV-4 correlates with protection against SIVsm challenge. *Clinical and Experimental Immunology* **129,** 11–18 (2002).

346. Alkhatib, G. *et al.* CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* **272,** 1955–1958 (1996).

347. Cocchi, F. *et al.* The V3 domain of the HIV-1 gp120 envelope glycoprotein is critical for chemokine-mediated blockade of infection. *Nature Medicine* **2,** 1244–1247 (1996).

348. Romagnani, S. T-cell subsets (Th1 versus Th2). *Annals of Allergy, Asthma and Immunology* **85,** 9–18 (2000).

349. Easterbrook, P. J., Emami, J. & Gazzard, B. Rate of CD4 cell decline and prediction of survival in zidovudine-treated patients. *AIDS* **7,** 959–967 (1993).

350. Doitsh, G. *et al.* Cell death by pyroptosis drives CD4 T-cell depletion in HIV-1 infection. *Nature* **505,** 509–514 (2014).

351. Huber, M. & Trkola, A. Humoral immunity to HIV-1: neutralization and beyond. *Journal of Internal Medicine* **262,** 5–25 (2007).

352. Tomaras, G. D. *et al.* Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *Journal of Virology* **82,** 12449–12463 (2008).

353. Legrand, E. *et al.* Course of specific T lymphocyte cytotoxicity, plasma and cellular viral loads, and neutralizing antibody titers in 17 recently seroconverted HIV type 1-infected patients. *AIDS Research and Human Retroviruses* **13,** 1383–1394 (1997).

354. Moog, C., Fleury, H. J., Pellegrin, I., Kirn, A. & Aubertin, A. M. Autologous and heterologous neutralizing antibody responses following initial seroconversion in human immunodeficiency virus type 1-infected individuals. *Journal of Virology* **71,** 3734–3741 (1997).

355. Richman, D. D., Wrin, T., Little, S. J. & Petropoulos, C. J. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proceedings of the National Academy of Sciences of the United States of America* **100,** 4144–4149 (2003).

356. Pereyra, F. *et al.* Genetic and immunologic heterogeneity among persons who control HIV infection in the absence of therapy. *Journal of Infectious Diseases* **197,** 563–571 (2008).

357. Aasa-Chapman, M. M. I. *et al.* In vivo emergence of HIV-1 highly sensitive to neutralizing antibodies. *PLoS ONE* **6,** e23961 (2011).

358. Leonard, C. K. *et al.* Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *Journal of Biological Chemistry* **265,** 10373–10382 (1990).

359. Reitter, J. N., Means, R. E. & Desrosiers, R. C. A role for carbohydrates in immune evasion in AIDS. *Nature Medicine* **4,** 679–684 (1998).

360. Cole, K. S., Steckbeck, J. D., Rowles, J. L., Desrosiers, R. C. & Montelaro, R. C. Removal of N-linked glycosylation sites in the V1 region of simian immunodeficiency virus gp120 results in redirection of B-cell responses to V3. *Journal of Virology* **78,** 1525–1539 (2004).

361. Doores, K. J. The HIV glycan shield as a target for broadly neutralizing antibodies. *Federation of European Biochemical Societies Journal* **282,** 4679–4691 (2015).

362. Kwong, P. D., Mascola, J. R. & Nabel, G. J. Broadly neutralizing antibodies and the search for an HIV-1 vaccine: the end of the beginning. *Nature Reviews Immunology* **13,** 693–701 (2013).

363. Hessell, A. J. *et al.* Broadly neutralizing human anti-HIV antibody 2G12 is effective in protection against mucosal SHIV challenge even at low serum neutralizing titers. *PLoS Pathogens* **5,** e1000433 (2009).

364. Walker, B. D. & Yu, X. G. Unravelling the mechanisms of durable control of HIV-1. *Nature Reviews Immunology* **13,** 487–498 (2013).

365. Johnson, W. E. & Desrosiers, R. C. Viral persistence: HIV's strategies of immune system evasion. *Annual Review of Medicine* **53,** 499–518 (2002).

366. Kirchhoff, F. Immune evasion and counteraction of restriction factors by HIV-1 and other primate lentiviruses. *Cell Host & Microbe* **8,** 55–67 (2010).

367. Alqudah, M. A. Y., Yaseen, M. M. M. & Yaseen, M. M. S. HIV-1 strategies to overcome the immune system by evading and invading innate immune system. *HIV & AIDS Review* **15,** 1–12 (2016).

368. *2015 China AIDS Response Progress Report* National Health and Family Planning Commission of the Peoples Republic of China (UNAIDS, May May 2015, Accessed Jun 2016). <`http://www.unaids.org/sites/default/files/country/documents/CHN_narrative_report_2015.pdf`>.

369. Zhang, Y. *The cellular immune response and virological nature of Human Immunodeficiency Virus-1 in infected slow progressors plasma donors in China* PhD thesis (University of Oxford, 2008).

370. Van Tienen, C. *et al.* Two distinct epidemics: the rise of HIV-1 and decline of HIV-2 infection between 1990 and 2007 in rural Guinea-Bissau. *Journal of Acquired Immune Deficiency Syndromes* **53,** 640–647 (2010).

371. Van der Loeff, M. F. S. *et al.* Undetectable plasma viral load predicts normal survival in HIV-2-infected people in a West African village. *Retrovirology* **7,** e1–10 (2010).

372. Buckner, M. in *Vivre et Penser le Sida en Afrique* (eds Becker, C., Dozon, J.-P. & Obbo, C.) 389–404 (Karthala, Paris: Codesria, IRD, 1999).

373. Wilkins, A. *et al.* The epidemiology of HIV infection in a rural area of Guinea-Bissau. *AIDS* **7,** 1119–1122 (1993).

374. Ricard, D. *et al.* The effects of HIV-2 infection in a rural area of Guinea-Bissau. *AIDS* **8,** 977–982 (1994).

375. Fryer, H. R. *et al.* Predicting the extinction of HIV-2 in rural Guinea-Bissau. *AIDS* **29,** 2479–2486 (2015).

376. Los Alamos National Laboatory (LANL). *HIV Sequence Database* <`http://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html`> (Published Oct 2014, Accessed Feb 2015).

377. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28,** 1647–1649 (2012).

378. Edgar, R. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32,** 1792–1797 (2004).

379. Larkin, M. *et al.* ClustalW and ClustalX version 2. *Bioinformatics* **23,** 2947–2948 (2007).

380. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* **30,** 2725–2729 (2013).

381. Darriba, D., Taboada, G., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9,** 772 (2012).

382. Guindon, S. & Gascuel, O. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology* **52,** 696–704 (2003).

383. Jukes, T. & Cantor, C. *Evolution of protein molecules* (New York: Academic Press, 1969).

384. Felsenstein, J. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* **17,** 368–376 (1981).

385. Hasegawa, M., Kishino, H. & Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22,** 160–174 (1985).

386. Tavare, S. in *Lectures on Mathematics in the Life Sciences* Ed., 57–86 (1986).

387. Drummond, A., Suchard, M., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29,** 1969–1973 (2012).

388. Rambaut, A., Suchard, M., Xie, D. & Drummond, A. *Tracer* <`http://beast.bio.ed.ac.uk/Tracer`>.

389. Rambaut, A. *FigTree* <`http://tree.bio.ed.ac.uk/software/figtree/`>.

390. Geospiza Inc. *FinchTV* <`http://www.geospiza.com/ftvdlinfo.html/`>.

391. Schrödinger, LLC. *The PyMOL Molecular Graphics System, Version 1.8* Nov. 2015.

392. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS MODEL work space: a web based environment for protein structure homology modelling. *Bioinformatics* **22,** 195–201 (2006).

393. Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L. & Schwede, T. The SWISS MODEL Repository & associated resources. *Nucleic Acids Research* **37,** D387–D392 (2009).

394. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research* **42,** W252–W258 (2014).

395. Guex, N., Peitsch, M. C. & Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* **30 S1,** S162–173 (2009).

396. Jensen, M. A. *et al.* Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. *Journal of Virology* **77,** 13376–13388 (2003).

397. Brumme, Z. L. *et al.* Clinical and immunological impact of HIV envelope V3 sequence variation after starting initial triple antiretroviral therapy. *AIDS* **18,** F1–F9 (2004).

398. Jensen, M. A., Coetzer, M., van 't Wout, A. B., Morris, L. & Mullins, J. I. A reliable phenotype predictor for human immunodeficiency virus type 1 subtype C based on envelope V3 sequences. *Journal of Virology* **80,** 4698–4704 (2006).

399. Rajapaksa, U. S. *et al.* HLA-B may be more protective against HIV-1 than HLA-A because it resists negative regulatory factor (Nef) mediated downregulation. *Proceedings of the National Academy of Sciences of the United States of America* **109,** 13353–13358 (2012).

400. Waldo, G. S., Standish, B. M., Berendzen, J. & Terwilliger, T. C. Rapid protein-folding assay using green fluorescent protein. *Nature Biotechnology* **17,** 691–695 (1999).

401. Los Alamos National Laboatory (LANL). *Nef CTL Epitope Map* <`http://www.hiv.lanl.gov/content/immunology/maps/ctl/Nef.html`> (Published Apr 2016, Accessed Apr 2016).

402. Lee, C. H., Saksela, K., Mirza, U. A., Chait, B. T. & Kuriyan, J. Crystal structure of the conserved core of HIV-1 Nef complexed with a Src family SH3 domain. *Cell* **85,** 931–942 (1996).

403. Arold, S. *et al.* The crystal structure of HIV-1 Nef protein bound to the Fyn kinase SH3 domain suggests a role for this complex in altered T cell receptor signaling. *Structure* **5,** 1361–1372 (1997).

404. Wyatt, R. *et al.* The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* **393,** 705–711 (1998).

405. Pantophlet, R. & Burton, D. R. GP120: target for neutralizing HIV-1 antibodies. *Annual Review of Immunology* **24,** 739–769 (2006).

406. Myszka, D. G. *et al.* Energetics of the HIV gp120-CD4 binding reaction. *Proceedings of the National Academy of Sciences of the United States of America* **97,** 9026–9031 (2000).

407. Kwong, P. D. *et al.* HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* **420,** 678–682 (2002).

408. Zhu, X., Borchers, C., Bienstock, R. J. & Tomer, K. B. Mass spectrometric characterization of the glycosylation pattern of HIV-gp120 expressed in CHO cells. *Biochemistry* **39,** 11194–11204 (2000).

409. Burton, D. R. & Mascola, J. R. Antibody responses to envelope glycoproteins in HIV-1 infection. *Nature Immunology* **16,** 571–576 (2015).

410. Shankarappa, R. *et al.* Evolution of human immunodeficiency virus type 1 envelope sequences in infected individuals with differing disease progression profiles. *Virology* **241,** 251–259 (1998).

411. Markham, R. B. *et al.* Patterns of HIV-1 evolution in individuals with differing rates of CD4 T cell decline. *Proceedings of the National Academy of Sciences of the United States of America* **95,** 12568–12573 (1998).

412. Lemey, P. *et al.* Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. *PLoS Computational Biology* **3,** e29 (2007).

413. Esbjörnsson, J. *HIV-1 evolution, disease progression and molecular epidemiology of HIV-1 single and HIV-1 and HIV-2 dual-infected individuals in Guinea-Bissau* PhD thesis (Lund University, 2010).

414. Dong, T. *et al.* Extensive HLA-driven viral diversity following a narrow-source HIV-1 outbreak in rural China. *Blood* **118,** 98–106 (2011).

415. Los Alamos National Laboatory (LANL). *gp160 Ab Epitope Map* <`http://www.hiv.lanl.gov/content/immunology/maps/ab/gp160.html`> (Published Jan 2015, Accessed Jan 2015).

416. Los Alamos National Laboatory (LANL). *gp160 CTL Epitope Map* <`http://www.hiv.lanl.gov/content/immunology/maps/ctl/gp160.html`> (Published Apr 2016, Accessed Apr 2016).

417. Los Alamos National Laboatory (LANL). *gp160 T-Helper Epitope Map* <`http://www.hiv.lanl.gov/content/immunology/maps/helper/gp160.html`> (Published Mar 2016, Accessed Apr 2016).

418. Shioda, T., Levy, J. A. & Cheng-Mayer, C. Small amino acid changes in the V3 hypervariable region of gp120 can affect the T-cell-line and macrophage tropism of human immunodeficiency virus type 1. *Proceedings of the National Academy of Sciences of the United States of America* **89,** 9434–9438 (1992).

419. Cardozo, T. *et al.* Structural basis for coreceptor selectivity by the HIV type 1 V3 loop. *AIDS Research and Human Retroviruses* **23,** 415–426 (2007).

420. *LANL Ambiguity Consensus Maker* Los Alamos National Laboatory (LANL). <`http://www.hiv.lanl.gov/content/sequence/CONSENSUS/AmbigCon.html`> (Last Edited Jan 2016, Accessed May 2016).

421. Shapiro, B., Rambaut, A. & Drummond, A. J. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Molecular Biology and Evolution* **23,** 7–9 (2006).

422. Gill, M. S. *et al.* Improving Bayesian population dynamics inference: a coalescent based model for multiple loci. *Molecular Biology and Evolution* **30,** 713–724 (2013).

423. Frost, S. D. W. & Volz, E. M. Viral phylodynamics and the search for an effective number of infections. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **365,** 1879–1890 (2010).

424. Mild, M. *et al.* High intrapatient HIV-1 evolutionary rate is associated with CCR5-to-CXCR4 coreceptor switch. *Infection, Genetics and Evolution* **19,** 369–377 (2013).

425. O'Brien, J. D., Minin, V. N. & Suchard, M. A. Learning to count: robust estimates for labeled distances between molecular sequences. *Molecular Biology and Evolution* **26,** 801–814 (2009).

426. Lemey, P., Minin, V. N., Bielejec, F., Kosakovsky Pond, S. L. & Suchard, M. A. A counting renaissance: combining stochastic mapping and empirical Bayes to quickly detect amino acid sites under positive selection. *Bioinformatics* **28,** 3248–3256 (2012).

427. Lyumkis, D. *et al.* Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science* **342,** 1484–1490 (2013).

428. Lesk, A. M. *Introduction to Bioinformatics* ISBN: 0199251967, 9780199251964 (Oxford University Press, 2002).

429. Snoeck, J., Fellay, J., Bartha, I., Douek, D. C. & Telenti, A. Mapping of positive selection sites in the HIV-1 genome in the context of RNA and protein structural constraints. *Retrovirology* **8,** e1–8 (2011).

430. Hartley, O., Klasse, P. J., Sattentau, Q. J. & Moore, J. P. V3: HIV's switch-hitter. *AIDS Research and Human Retroviruses* **21,** 171–189 (2005).

431. Pantophlet, R., Wrin, T., Cavacini, L. A., Robinson, J. E. & Burton, D. R. Neutralizing activity of antibodies to the V3 loop region of HIV-1 gp120 relative to their epitope fine specificity. *Virology* **381,** 251–260 (2008).

432. Conley, A. J. *et al.* Neutralization of primary human immunodeficiency virus type 1 isolates by the broadly reactive anti-V3 monoclonal antibody, 447-52D. *Journal of Virology* **68,** 6994–7000 (1994).

433. Gorny, M. K. *et al.* Cross-clade neutralizing activity of human anti-V3 monoclonal antibodies derived from the cells of individuals infected with non-B clades of human immunodeficiency virus type 1. *Journal of Virology* **80,** 6865–6872 (2006).

434. Corti, D. *et al.* Analysis of memory B cell responses and isolation of novel monoclonal antibodies with neutralizing breadth from HIV-1-infected individuals. *PLoS ONE* **5,** e8805 (2010).

435. Hioe, C. E. *et al.* Anti-V3 monoclonal antibodies display broad neutralizing activities against multiple HIV-1 subtypes. *PLoS ONE* **5,** e10254 (2010).

436. Zolla-Pazner, S. & Cardozo, T. Structure-function relationships of HIV-1 envelope sequence-variable regions provide a paradigm for vaccine design. *Nature Reviews Immunology* **10,** 527–535 (2010).

437. Cao, J. *et al.* Replication and neutralization of human immunodeficiency virus type 1 lacking the V1 and V2 variable loops of the gp120 envelope glycoprotein. *Journal of Virology* **71,** 9808–9812 (1997).

438. Lucchiari-Hartz, M. *et al.* Differential proteasomal processing of hydrophobic and hydrophilic protein regions: Contribution to cytotoxic T lymphocyte epitope clustering in HIV-1-Nef. *Proceedings of the National Academy of Sciences of the United States of America* **100,** 7755–7760 (2003).

439. Jallow, S. *et al.* The presence of prolines in the flanking region of an immunodominant HIV-2 gag epitope influences the quality and quantity of the epitope generated. *European Journal of Immunology* **45,** 2232–2242 (2015).

440. Schmid, B. V., Kemir, C. & de Boer, R. J. The distribution of CTL epitopes in HIV-1 appears to be random, and similar to that of other proteomes. *BMC Evolutionary Biology* **9,** 1–15 (2009).

441. Arhel, N. J. & Kirchhoff, F. Implications of Nef: host cell interactions in viral persistence and progression to AIDS. *Current Topics in Microbiology and Immunology* **339,** 147–175 (2009).

442. Greenberg, M. E., Iafrate, A. J. & Skowronski, J. The SH3 domain-binding surface and an acidic motif in HIV-1 Nef regulate trafficking of class I MHC complexes. *The EMBO Journal* **17,** 2777–2789 (1998).

443. Schwartz, O., Marechal, V., Le Gall, S., Lemonnier, F. & Heard, J. M. Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein. *Nature Medicine* **2,** 338–342 (1996).

444. Collins, K. L., Chen, B. K., Kalams, S. A., Walker, B. D. & Baltimore, D. HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. *Nature* **391,** 397–401 (1998).

445. Cohen, G. B. *et al.* The selective downregulation of class I major histocompatibility complex proteins by HIV-1 protects HIV-infected cells from NK cells. *Immunity* **10,** 661–671 (1999).

446. Specht, A. *et al.* Selective downmodulation of HLA-A and -B by Nef alleles from different groups of primate lentiviruses. *Virology* **373,** 229–237 (2008).

447. Lewis, M. J. *et al.* Functional adaptation of nef to the immune milieu of HIV-1 infection in vivo. *The Journal of Immunology* **180,** 4075–4081 (2008).

448. Mlcochova, P. *et al.* Immune evasion activities of accessory proteins Vpu, Nef and Vif are conserved in acute and chronic HIV-1 infection. *Virology* **482,** 72–78 (2015).

449. Carl, S. *et al.* Modulation of different human immunodeficiency virus type 1 Nef functions during progression to AIDS. *Journal of Virology* **75,** 3657–3665 (2001).

450. Munch, J., Stolte, N., Fuchs, D., Stahl-Hennig, C. & Kirchhoff, F. Efficient class I major histocompatibility complex down-regulation by simian immunodeficiency virus Nef is associated with a strong selective advantage in infected rhesus macaques. *Journal of Virology* **75,** 10532–10536 (2001).

451. Le Gall, S. *et al.* Nef interacts with the mu subunit of clathrin adaptor complexes and reveals a cryptic sorting signal in MHC I molecules. *Immunity* **8,** 483–495 (1998).

452. Piguet, V. *et al.* HIV-1 Nef protein binds to the cellular protein PACS-1 to downregulate class I major histocompatibility complexes. *Nature Cell Biology* **2,** 163–167 (2000).

453. Blagoveshchenskaya, A. D., Thomas, L., Feliciangeli, S. F., Hung, C. H. & Thomas, G. HIV-1 Nef downregulates MHC-I by a PACS-1-and PI3K-regulated ARF6 endocytic pathway. *Cell* **111,** 853–866 (2002).

454. Wonderlich, E. R., Leonard, J. A. & Collins, K. L. HIV immune evasion: disruption of antigen presentation by the HIV nef protein. *Advances in virus research* **80,** 103–127 (2011).

455. Swann, S. A. *et al.* HIV-1 nef blocks transport of MHC class I molecules to the cell surface via a PI 3-kinase-dependent pathway. *Virology* **282,** 267–277 (2001).

456. Roeth, J. F., Williams, M., Kasper, M. R., Filzen, T. M. & Collins, K. L. HIV-1 nef disrupts MHC-I trafficking by recruiting AP-1 to the MHC-I cytoplasmic tail. *The Journal of Cell Biology* **167,** 903–913 (2004).

457. Roeth, J. F. & Collins, K. L. Human immunodeficiency virus type 1 Nef: adapting to intracellular trafficking pathways. *Microbiology and Molecular Biology Reviews* **70,** 548–563 (2006).

458. Leonard, J. A., Filzen, T., Carter, C. C., Schaefer, M. & Collins, K. L. HIV-1 nef disrupts intracellular trafficking of major histocompatibility complex class I, CD4, CD8, and CD28 by distinct pathways that share common elements. *Journal of Virology* **85,** 6867–6881 (2011).

459. Hansen, T. H. & Bouvier, M. MHC class I antigen presentation: learning from viral evasion strategies. *Nature Reviews Immunology* **9,** 503–513 (2009).

460. Shen, Q. T., Ren, X., Zhang, R., Lee, I. H. & Hurley, J. H. HIV-1 nef hijacks clathrin coats by stabilizing AP-1:Arf1 polygons. *Science* **350,** aac5137 (2015).

461. Dikeakos, J. D. *et al.* Small molecule inhibition of HIV-1-induced MHC-I down-regulation identifies a temporally regulated switch in nef action. *Molecular Biology of the Cell* **21,** 3279–3292 (2010).

462. Jia, X. *et al.* Structural basis of evasion of cellular adaptive immunity by HIV-1 Nef. *Nature Structural & Molecular Biology* **19,** 701–706 (2012).

463. Ohno, H. *et al.* Interaction of tyrosine-based sorting signals with clathrin associated proteins. *Science* **269,** 1872–1875 (1995).

464. Fackler, O. T., Luo, W., Geyer, M., Alberts, A. S. & Peterlin, B. M. Activation of Vav by Nef induces cytoskeletal rearrangements and downstream effector functions. *Molecular Cell* **3,** 729–739 (1999).

465. Barnham, K. J., Monks, S. A., Hinds, M. G., Azad, A. A. & Norton, R. S. Solution structure of a polypeptide from the N terminus of the HIV protein nef. *Biochemistry* **36,** 5970–5980 (1997).

466. Mangasarian, A., Piguet, V., Wang, J. K., Chen, Y. L. & Trono, D. Nef-induced CD4 and major histocompatibility complex class I (MHC-I) down-regulation are governed by distinct determinants: N-terminal alpha helix and proline repeat of Nef selectively regulate MHC-I trafficking. *Journal of Virology* **73,** 1964–1973 (1999).

467. Schaefer, M. R., Wonderlich, E. R., Roeth, J. F., Leonard, J. A. & Collins, K. L. HIV-1 Nef targets MHC-I and CD4 for degradation via a final common beta-COP-dependent pathway in T cells. *PLoS Pathogens* **4,** e1000131 (2008).

468. Singh, R. K., Lau, D., Noviello, C. M., Ghosh, P. & Guatelli, J. C. An MHC-I cytoplasmic domain/HIV-1 Nef fusion protein binds directly to the mu subunit of the AP-1 endosomal coat complex. *PLoS ONE* **4,** e8364 (2009).

469. Dikeakos, J. D. *et al.* An interdomain binding site on HIV-1 Nef interacts with PACS-1 and PACS-2 on endosomes to down-regulate MHC-I. *Molecular Biology of the Cell* **23,** 2184–2197 (2012).

470. Simmen, T. *et al.* PACS-2 controls endoplasmic reticulum-mitochondria communication and Bid-mediated apoptosis. *The EMBO Journal* **24,** 717–729 (2005).

471. Atkins, K. M. *et al.* HIV-1 Nef binds PACS-2 to assemble a multikinase cascade that triggers major histocompatibility complex class I (MHC-I) down-regulation - analysis using short interfering RNA and knock-out mice. *Journal of Biological Chemistry* **283,** 11772–11784 (2008).

472. Mangasarian, A. *et al.* The HIV-1 Nef protein acts as a connector with sorting pathways in the Golgi and at the plasma membrane. *Immunity* **6,** 67–77 (1997).

473. Yamada, T. *et al.* Proline 78 is crucial for human immunodeficiency virus type 1 Nef to down-regulate class I human leukocyte antigen. *Journal of Virology* **77,** 1589–1594 (2003).

474. Liu, L. X. *et al.* Mutation of a conserved residue (D123) required for oligomerization of human immunodeficiency virus type 1 Nef protein abolishes interaction with human thioesterase and results in impairment of Nef biological functions. *Journal of Virology* **74,** 5310–5319 (2000).

475. Johannes, L. *et al.* Effects of HIV-1 Nef on retrograde transport from the plasma membrane to the endoplasmic reticulum. *Traffic* **4,** 323–332 (2003).

476. Kwak, Y. T. *et al.* Self-association of the Lentivirus protein, Nef. *Retrovirology* **7,** e1–22 (2010).

477. Swigut, T., Iafrate, A. J., Muench, J., Kirchhoff, F. & Skowronski, J. Simian and human immunodeficiency virus Nef proteins use different surfaces to down-regulate class I major histocompatibility complex antigen expression. *Journal of Virology* **74,** 5691–5701 (2000).

478. Goulder, P. J. R. & Watkins, D. I. Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nature Reviews Immunology* **8,** 619–630 (2008).

479. Den Uyl, D., van der Horst-Bruinsma, I. E. & van Agtmael, M. Progression of HIV to AIDS: a protective role for HLA-B27? *AIDS Reviews* **6,** 89–96 (2004).

480. Munch, J. *et al.* Primary sooty mangabey simian immunodeficiency virus and human immunodeficiency virus type 2 nef alleles modulate cell surface expression of various human receptors and enhance viral infectivity and replication. *Journal of Virology* **79,** 10547–10560 (2005).

481. Yindom, L.-M. *et al.* Influence of HLA class I and HLA-KIR compound genotypes on HIV-2 infection and markers of disease progression in a Manjako community in West Africa. *Journal of Virology* **84,** 8202–8208 (2010).

482. De Oliveira, T. *et al.* An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics* **21,** 3797–3800 (2005).

483. Alcantara, L. C. *et al.* A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. *Nucleic Acids Research* **37,** W634–W642 (2009).

484. De Winter, J. Using the Students t-test with extremely small sample sizes. *Practical Assessment, Research & Evaluation* **18,** e1–10 (2013).

485. Los Alamos National Laboatory (LANL). *Nef Antibody Epitope Map* <`http://www.hiv.lanl.gov/content/immunology/maps/ab/Nef.html`> (Published Apr 2016, Accessed Apr 2016).

486. Los Alamos National Laboatory (LANL). *Nef T-Helper Epitope Map* <`http://www.hiv.lanl.gov/content/immunology/maps/helper/Nef.html`> (Published Apr 2016, Accessed Apr 2016).

487. Gonzalez-Galarza, F. *et al.* Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acid Research* **28,** D784–D788 (2015).

488. Yin, W., Xiang, P. & Li, Q. Investigations of the effect of DNA size in transient transfection assay using dual luciferase system. *Analytical Biochemistry* **346,** 289–294 (2005).

489. Mahiti, M. *et al.* Relative resistance of HLA-B to downregulation by naturally occurring HIV-1 nef sequences. *mBio* **7,** e01516–15 (2016).

490. Mann, J. K. *et al.* Nef-mediated down-regulation of CD4 and HLA class I in HIV-1 subtype C infection: association with disease progression and influence of immune pressure. *Virology* **468,** 214–225 (2014).

491. McCarthy, M., He, J. & Wood, C. HIV-1 strain-associated variability in infection of primary neuroglia. *Journal of Neurovirology* **4,** 80–89 (1998).

492. Apps, R. *et al.* Relative expression levels of the HLA class-I proteins in normal and HIV-infected cells. *Journal of Immunology* **194,** 3594–3600 (2015).

493. Apps, R. *et al.* Influence of HLA-C expression level on HIV control. *Science* **340,** 87–91 (2013).

494. Feldmann, J. *et al.* Downregulation of the T-cell receptor by human immunodeficiency virus type 2 Nef does not protect against disease progression. *Journal of Virology* **83,** 12968–12972 (2009).

495. Meulenberg, C. J. W., Todorovic, V. & Cemazar, M. Differential cellular effects of electroporation and electrochemotherapy in monolayers of human microvascular endothelial cells. *PLoS ONE* **7,** e52713 (2012).

496. Mann, J. K. *et al.* Ability of HIV-1 Nef to downregulate CD4 and HLA class I differs among viral subtypes. *Retrovirology* **10,** e1–16 (2013).

497. Marenzi, S. *et al.* Efficiency of expression of transfected genes depends on the cell cycle. *Molecular Biology Reports* **26,** 261–267 (1999).

498. Bouabe, H., Fässler, R. & Heesemann, J. Improvement of reporter activity by IRES-mediated polycistronic reporter system. *Nucleic Acids Research* **36,** e28 (2008).

499. Chen, X., Zaro, J. L. & Shen, W. C. Fusion protein linkers: property, design and functionality. *Advanced Drug Delivery Reviews* **65,** 1357–1369 (2013).

500. Cleveland, W. S. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* **74,** 829–836 (1979).

501. Kim, S. Y., Byrn, R., Groopman, J. & Baltimore, D. Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. *Journal of Virology* **63,** 3708–3713 (1989).

502. Olivetta, E., Arenaccio, C., Manfredi, F., Anticoli, S. & Federico, M. The contribution of extracellular Nef to HIV-induced pathogenesis. *Current Drug Targets* **17,** 46–53 (2016).

503. Lee, J. H. *et al.* HIV-Nef and ADAM17-containing plasma extracellular vesicles induce and correlate with immune pathogenesis in chronic HIV infection. *EBioMedicine* **6,** 103–113 (2016).

504. Rivoltini, L. *et al.* Quantitative correlation between HLA class-I allele expression and recognition of melanoma-cells by antigen-specific cytotoxic T-lymphocytes. *Cancer Research* **55,** 3149–3157 (1995).

505. Snary, D., Barnstable, C. J., Bodmer, W. F. & Crumpton, M. J. Molecular structure of human histocompatibility antigens: the HLA-C series. *European Journal of Immunology* **7,** 580–585 (1977).

506. Takada, A. *et al.* Downregulation of beta1 integrins by Ebola virus glycoprotein: implication for virus entry. *Virology* **278,** 20–26 (2000).

507. Smith, P., Rhodes, N. P., Ke, Y. & Foster, C. S. Upregulation of estrogen and androgen receptors modulate expression of FGF-2 and FGF-7 in human, cultured, prostatic stromal cells exposed to high concentrations of estradiol. *Prostate Cancer and Prostatic Diseases* **5,** 105–110 (2002).

508. Betts, M. R. *et al.* Analysis of total human immunodeficiency virus (HIV)-specific CD4+ and CD8+ T-cell responses: relationship to viral load in untreated HIV infection. *Journal of Virology* **75,** 11983–11991 (2001).

509. Addo, M. M. *et al.* Comprehensive epitope analysis of human immunodeficiency virus type 1 (HIV-1)-specific T-cell responses directed against the entire expressed HIV-1 genome demonstrate broadly directed responses, but no correlation to viral load. *Journal of Virology* **77,** 2081–2092 (2003).

510. Malhotra, U. *et al.* Enhanced detection of human immunodeficiency virus type 1 (HIV-1) nef-specific T cells recognizing multiple variants in early HIV-1 infection. *Journal of Virology* **81,** 5225–5237 (2007).

511. Mlotshwa, M. *et al.* Fluidity of HIV-1-specific T-cell responses during acute and early subtype C HIV-1 infection and associations with early disease progression. *Journal of Virology* **84,** 12018–12029 (2010).

512. Gray, G. E. *et al.* Recombinant adenovirus type 5 HIV gag/pol/nef vaccine in South Africa: unblinded, long-term follow-up of the phase 2b HVTN 503/Phambili study. *The Lancet Infectious Diseases* **14,** 388–396 (2014).

513. De La Cruz, J. *et al.* Ineffectual targeting of HIV-1 nef by cytotoxic T lymphocytes in acute infection results in no functional impairment or viremia reduction. *Journal of Virology* **88,** 7881–7892 (2014).

514. Adland, E. *et al.* Nef-specific CD8+ T cell responses contribute to HIV-1 immune control. *PLoS ONE* **8,** e73117 (2013).

515. Leligdowicz, A. *et al.* Robust Gag-specific T cell responses characterize viremia control in HIV-2 infection. *Journal of Clinical Investigation* **117,** 3067–3074 (2007).

516. Yewdell, J. W. & Hill, A. B. Viral interference with antigen presentation. *Nature Immunology* **3,** 1019–1025 (2002).

517. Levitskaya, J. *et al.* Inhibition of antigen-processing by the internal repeat region of the Epstein-Barr-virus nuclear antigen-1. *Nature* **375,** 685–688 (1995).

518. Blake, N. *et al.* Human CD8+ T cell responses to EBV EBNA1: HLA class I presentation of the (gly-ala)containing protein requires exogenous processing. *Immunity* **7,** 791–802 (1997).

519. Tellam, J. T., Lekieffre, L., Zhong, J., Lynn, D. J. & Khanna, R. Messenger RNA sequence rather than protein sequence determines the level of self-synthesis and antigen presentation of the EBV-encoded antigen, EBNA1. *PLoS Pathogens* **8,** e1003112 (2012).

520. Onyango, C. O. *et al.* HIV-2 capsids distinguish high and low virus load patients in a West African community cohort. *Vaccine* **28,** B60–B67 (2010).

521. Niedermann, G. *et al.* The proteolytic fragments generated by vertebrate proteasomes: structural relationships to major histocompatibility complex class I binding peptides. *Proceedings of the National Academy of Sciences of the United States of America* **93,** 8572–8577 (1996).

522. Yewdell, J. W., Reits, E. & Neefjes, J. Making sense of mass destruction: quantitating MHC class I antigen presentation. *Nature Reviews Immunology* **3,** 952–961 (2003).

523. Ternette, N. *et al.* Early kinetics of the HLA class I-associated peptidome of MVA.HIVconsv-infected cells. *Journal of Virology* **89,** 5760–5771 (2015).

524. Hendel, H. *et al.* New class I and II HLA alleles strongly associated with opposite patterns of progression to AIDS. *The Journal of Immunology* **162,** 6942–6946 (1999).

525. Itescu, S. *et al.* HLA-B35 is associated with accelerated progression to AIDS. *Journal of Acquired Immune Deficiency Syndrome* **5,** 37–45 (1992).

526. Geyer, M., Fackler, O. T. & Peterlin, B. M. Structure-function relationships in HIV-1 Nef. *EMBO Reports* **2,** 580–585 (2001).

527. Geyer, M., Munte, C. E., Schorr, J., Kellner, R. & Kalbitzer, H. R. Structure of the anchor-domain of myristoylated and non-myristoylated HIV-1 Nef protein. *Journal of Molecular Biology* **289,** 123–138 (1999).

528. Ren, X. F., Park, S. Y., Bonifacino, J. S. & Hurley, J. H. How HIV-1 Nef hijacks the AP-2 clathrin adaptor to downregulate CD4. *eLIFE* **3,** e01754 (2014).

529. Lecker, S. H., Goldberg, A. L. & Mitch, W. E. Protein degradation by the ubiquitin-proteasome pathway in normal and disease states. *Journal of the American Society of Nephrology* **17,** 1807–1819 (2006).

530. Schiffner, T., Sattentau, Q. J. & Dorrell, L. Development of prophylactic vaccines against HIV-1. *Retrovirology* **10,** e1–16 (2013).

531. Dou, Z. *et al.* HIV-infected former plasma donors in rural Central China: from infection to survival outcomes, 1985-2008. *PLoS ONE* **5,** e13737 (2010).

532. Wikramaratna, P. S., Lourenço, J., Klenerman, P., Pybus, O. G. & Gupta, S. Effects of neutralizing antibodies on escape from CD8+ T-cell responses in HIV-1 infection. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **370,** e20140290 (2015).

533. Deol, P., Zaiss, D. M., Monaco, J. J. & Sijts, A. J. Rates of processing determine the immunogenicity of immunoproteasome-generated epitopes. *Journal of Immunology* **178,** 7557–7462 (2007).

Appendix A

Supplementary Tables

Table A.1 PCR Primers used in the studies described in this thesis. Colour denotes a functional sequence. Blue: restriction site; red: Kozak sequence; green: frame conservation sequence. Primer binding sites are indicated in square brackets, and correspond to either the specified vector, or an appropriate full-length reference genome.

Name	Sequence	Target	Chapter
FPD_Env_OF	AAAGAGCAGAAGACAGTGGCAATG	HIV-1 *gp120* [HXB2 6204-6227]	3
FPD_Env_OR	AGCCTCCTACTAYCATTATGAAT	HIV-1 *gp120* [HXB2 8276-8298]	3
FPD_Env_IF	CGGCTGGTTTTGCGATTCTA	HIV-1 *gp120* [HXB2 6883-6902]	3
FPD_Env_IR	TATTTATATAATTCACTTCTCCA	HIV-1 *gp120* [HXB2 7659-7681]	3
M13F	GTAAAACGACGGCCAG	pCR®4-TOPO [355-370]	3,4
M13R	CAGGAAACAGCTATGAC	pCR®4-TOPO [205-221]	3,4
HIV1_Nef_OF	GTAGCTGAGGGGACAGATAG	HIV-1 *nef* [HXB2 8688-8707]	4,5
HIV1_Nef_OR	TGCTAGAGATTTTCCACAC	HIV-1 *nef* [HXB2 9701-9719]	4,5

Continued on next page

Table A.1 – *Continued from previous page*

Description	Sequence	Target	Chapter
HIV1_Nef_IF	GAAGAATAAGACAGGGCT	HIV-1 *nef* [HXB2 8758-8775]	4,5
HIV1_Nef_IR	AGGCTCAGATCTGGTCTAA	HIV-1 *nef* [HXB2 9551-9569]	4,5
HIV2_Nef_OF	AAGGGCTATAGGCCTGTTTTCTC	HIV-2 *nef* [BEN 8813-8835]	4,5
HIV2_Nef_OR	GTTCCCCACCAGCTCCTCCCA	HIV-2 *nef* [BEN 9992-10012]	4,5
HIV2_Nef_IF	AGAAACAGAAGAAGACGTTGGAG	HIV-2 *nef* [BEN 8899-8921]	4,5
HIV2_Nef_IR	TGCAGTCTCAGCTGTTTTCAGTA	HIV-2 *nef* [BEN 9931-9953]	4,5
HIV1_EcoRI_F1	GGAAAGGAATTCACCACCATGGGTGGCAAGTGGTC	HIV-1 *nef* [HXB2 8797-8813]	4,5
HIV1_BamHI_R1	GCATCGGGATCCCCGCAGTCTTTGTAGTACTCCGG	HIV-1 *nef* [HXB2 9394-9414]	4,5
pAcGFP1N1_F	GCAGAGCTGGTTTAGTGAACCGTCAGATCC	pAcGFP1-N1 [561-590]	4,5
pAcGFP1N1_R	CCGGTGAACAGCTCGGCGCCCTTGCTCACC	pAcGFP1-N1 [675-704]	4,5
pIRES2ZsG1_F	TTTAGTGAACCGTCAGATCC	pIRES2-ZsGreen1 [571-590]	4,5
pIRES2ZsG1_R	AGCGGCTTCGGCCAGT	pIRES2-ZsGreen1 [695-710]	4,5

Table A.2 Inferred founder sequence of the SM cohort gp120 fragment, determined by ancestral state reconstruction implemented through BEAST.

Description	Sequence
Nucleotide	GCTGGTTTTGCGATTCTAAAGTGTAACAATAAAACGTTCAATGGAACAG GACCATGTACAAATGTCAGTACAGTACAATGTACACATGGAATTAGGCC AGTAGTATCAACTCAACTGCTGTTAAATGGCAGTTTAGCAGAAGAAGAG GTAGTAATTAGATCTAGCAATTTCACRGACAATGCTAAAGTCATAATAG TACAGCTGAATGAATCTGTAGAAATTAATTGTACAAGACCCAACAACAA TACAAGAAAAAGTATACCTCTAGGACCAGGGCRAGCATGGTATRCAACA GGACAAATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAA CARAATGGAATAACACTTTAAAACAGATAACTRAAAAATTAAGAGAACA ATTTAACAAAACAATAATCTTTAATCAATCTTCAGGAGGGGACCCAGAG ATTGTAATGCACAGTTTTAATTGTGGAGGGGAATTTTTCTACTGTAATA CATCACAACTGTTTAATAGTACTTGGAATGATACTAGTACTTGGAATGA TACTATCACGCTCCCTTGCAGAATAAAACAAATTGTAAACATGTGGCAG GAAGTAGGGAAAGCAATGTATGCCCCTCCCATTRAAGGACAAATTAGAT GTTCATCAAATATTACAGGGTTGCTATTAACAAGAGATGGTGGTAATGA KARCMASGACACCGAGACCTTCAGACCTGGAGGAGGAGATATGAGAGAC AATTGGAGAAGTGAATTATATAAA
Amino Acid	AGFAILKCNNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEE VVIRSSNFXDNAKVIIVQLNESVEINCTRPNNNTRKSIPLGPGXAWYXT GQIIGDIRQAHCNLSRTXWNNTLKQITXKLREQFNKTIIFNQSSGGDPE IVMHSFNCGGEFFYCNTSQLFNSTWNDTSTWNDTITLPCRIKQIVNMWQ EVGKAMYAPPIXGQIRCSSNITGLLLTRDGGNXXXDTETFRPGGGDMRD NWRSELYK

Table A.3 List of HLA types that share the cytoplasmic tails of chimeric HLA expressed in each of the stable transgenic cell lines. Table adapted from the DPhil thesis of Dr Ushani Rajapaksa (University of Oxford, 2012).

Chimeric HLA	HLA Types Sharing Cytoplasmic Tail
HLA-Aα	A*0201, A*0202, A*0203, A*0204, A*0205, A*0206, A*0207, A*0209, A*0210, A*0211, A*0212, A*02124, A*0213, A*0216, A*0217, A*0218, A*0221, A*0222, A*02230, A*0224, A*02249, A*02251, A*02253, A*02259, A*02264, A*02265, A*02266, A*02269, A*02279, A*0230, A*0233, A*0248, A*0251, A*0257, A*0260, A*0264, A*0268, A*0277, A*0279, A*0281, A*0289, A*1105, A*6801, A*6802, A*6803, A*6808, A*6817, A*6822, A*6836, A*6871, A*6901
HLA-Aβ	A*0265, A*2501, A*2502, A*2601, A*2602, A*2603, A*2604, A*2605, A*2607, A*2608, A*2615, A*2650, A*2901, A*2902, A*2903, A*3101, A*3104, A*3132, A*3146, A*3201, A*3202, A*3205, A*3217, A*3301, A*3303, A*3307, A*3349, A*3401, A*3402, A*4301, A*6601, A*6602, A*7401, A*7402
HLA-Aγ	A*0101, A*0102, A*0103, A*0109, A*0114, A*0120, A*0125, A*0301, A*0302, A*0304, A*03134, A*0326, A*0339, A*0377, A*0378, A*1101, A*1102, A*1103, A*1104, A*1107, A*11110, A*1112, A*1125, A*1150, A*1156, A*1160, A*1161, A*1162, A*1174, A*1175, A*1177, A*3001, A*3002, A*3004, A*3008, A*3025, A*3601
HLA-Aδ	A*43
HLA-Aϵ	A*2301, A*2306, A*2309, A*2326, A*2402, A*2403, A*2404, A*2407, A*2408, A*2410, A*24128, A*24129, A*24142, A*24143, A*24152, A*2420, A*2422, A*2461
HLA-Aζ	A*8001
HLA-Aη	A*32

Continued on next page

Table A.3 – *Continued from previous page*

Chimeric HLA	**HLA Types Sharing Cytoplasmic Tail**
HLA-Bα	B*14, B*1402, B*15, B*15/41, B*1501, B*15011, B*1501101, B*1502, B*1504, B*1508, B*1510, B*1511, B*1512, B*1513, B*1514, B*1515, B*1516, B*1517, B*15170101, B*1518, B*1519, B*1520, B*1521, B*1522, B*1523, B*1525, B*1542, B*1554, B*18, B*1801, B*35, B*3502, B*3511, B*3512, B*3515, B*3541, B*3565, B*38, B*3801, B*390101, B*39011, B*3906, B*3913, B*46, B*49, B*51, B*510111, B*510113, B*510114, B*510115, B*510116, B*51022, B*5104, B*5108, B*53, B*5301, B*5401, B*5501, B*58, B*65, B*72, B*73, B*7301, B*78
HLA-Bβ	B*07, B*0701, B*07021, B*0703, B*0704, B*0705, B*0706, B*0707, B*07362, B*0750, B*0802, B*13, B*1301, B*1302, B*130203, B*27, B*27052, B*2706, B*2711, B*2732, B*3701, B*3702, B*4002, B*4007, B*41, B*42, B*4201, B*4202, B*4208, B*44, B*44.2, B*44021, B*4403, B*4407, B*4803, B*4805, B*57, B*5701, B*5706, B*5711, B*62, B*8101
HLA-Cα	C*0102, C*0103, C*0104, C*0106, C*0108, C*0114, C*0121, C*0130, C*0140, C*0202, C*0210, C*0211, C*0216, C*0220, C*0221, C*0223, C*0302, C*0303, C*0304, C*0305, C*0306, C*03100, C*0313, C*0340, C*0346, C*0361, C*0401, C*0403, C*0406, C*0442, C*0470, C*0471, C*0501, C*0503, C*0508, C*0509, C*0602, C*0624, C*0801, C*0802, C*0803, C*0804, C*0820, C*0821, C*0824, C*0825, C*0827, C*0831, C*0840, C*0841, C*1202, C*1203, C*1204, C*1205, C*1208, C*1213, C*1219, C*1222, C*1403, C*1423, C*1436, C*1502, C*1503, C*1504, C*1505, C*1513, C*1516, C*1517, C*1601, C*1602, C*1604, C*1801, C*1802
HLA-Cβ	C*0822
HLA-Cγ	C*0430
HLA-Cδ	C*0750
HLA-Cϵ	C*0711

Continued on next page

Table A.3 – *Continued from previous page*

Chimeric HLA	HLA Types Sharing Cytoplasmic Tail
HLA-Cζ	C*0701, C*0702, C*0704, C*07149, C*07170, C*07172, C*07208, C*07235, C*0726, C*0729, C*0730, C*0736, C*0742, C*0746, C*0749, C*0752, C*0756, C*0757, C*0759, C*0766, C*0767, C*0793
HLA-Cη	C*0706, C*0718, C*0719

Table A.4 Cytoplasmic tails of the chimeric HLA-I molecules. Table adapted from the DPhil thesis of Dr Ushani Rajapaksa (University of Oxford, 2012).

Chimeric HLA	Cytoplasmic Tail Sequence
HLA-Aα	^{310}RRKSSDRKGGSYSQAASSDSAQGSDVSLTACKV^{342}
HLA-Aβ	^{310}.....M.....^{342}
HLA-Aγ	^{310}.....T.....^{342}
HLA-Aδ	^{310}.....F.....M.....^{342}
HLA-Aϵ	^{310}..N.....^{342}
HLA-Aζ	^{310}.K...V.....^{342}
HLA-Aη	^{310}....L.....M.....^{342}
HLA-Bα	^{310}.....GG.....^{339}
HLA-Bβ	^{310}.....GG.....C.....^{339}
HLA-Cα	^{310}.....GG....C.....N.....E..I...A^{342}
HLA-Cβ	^{310}.....GG.....N.....E..I...A^{342}
HLA-Cγ	^{310}.....GG....C.....TN.....E..I...A^{342}
HLA-Cδ	^{310}.....GG....C.....N.....E..IT..A^{342}
HLA-Cϵ	^{310}.....GG....C....C.N.....E..I...A^{342}
HLA-Cζ	^{310}.....GG....C....C.N.....E..IT..A^{342}
HLA-Cη	^{310}.....GG....C..V.C.N.....E..IT..A^{342}

Table A.5 Synthetic genes used for the construction of modified flexible expression vectors. Restriction sites are shown in red, Kozak sequence in blue, and modified Waldo linker [400] in green.

Description	Nucleotide Sequence
Consensus HIV-1 *nef* and components for construction of flexible expression vector	GGA AAG GAA TTC ACC ACC ATG GGT GGC AAG TGG TCA AAA AGT AGT GTG GTT GGA TGG CCT GCT GTA AGG GAA AGA ATG AGA CGA GCT GAG CCA GCA GCA GAT GGG GTG GGA GCA GTA TCT CGA GAC CTG GAA AAA CAT GGA GCA ATC ACA AGT AGC AAT ACA GCA GCT AAC AAT GCT GAT TGT GCC TGG CTA GAA GCA CAA GAG GAG GAG GAG GTG GGT TTT CCA GTC AGA CCT CAG GTA CCT TTA AGA CCA ATG ACT TAC AAG GGA GCT TTA GAT CTT AGC CAC TTT TTA AAA GAA AAG GGG GGA CTG GAA GGG CTA ATT TAC TCC CAA AAA AGA CAA GAT ATC CTT GAT CTG TGG GTC TAC CAC ACA CAA GGC TAC TTC CCT GAT TGG CAG AAC TAC ACA CCA GGG CCA GGG ATC AGA TAT CCA CTG ACC TTT GGA TGG TGC TTC AAG CTA GTA CCA GTT GAG CCA GAG AAG GTA GAA GAG GCC AAT GAA GGA GAG AAC AAC AGC TTG TTA CAC CCT ATG AGC CAG CAT GGG ATG GAT GAC CCG GAG AAA GAA GTG TTA GTG TGG AAG TTT GAC AGC CGC CTA GCA TTT CAT CAC ATG GCC CGA GAG CTG CAT CCG GAG TAC TAC AAG GAC TGC GGA TCC GGA GGC GGA TCT GCT GGC TCC GCT GCT GGT TCT GGC GCA TTC TTA CCG GTC GAT GC

Continued on next page

Table A.5 – *Continued from previous page*

Description	**Nucleotide Sequence**
Consensus HIV-2 *nef* and components for construction of flexible expression vector	GGA AAG GAA TTC ACC ACC ATG GGT GCG AGT GGC TCC AAG AAG CGT TCC AAG CCT TCG CGA GGG CTA CGA GAG AGA CTC TTG CGA GCG CGT GGA GGG ACT TGT GGG GGG CAC TGC GAC GAA TCG GGA GGG GAA TAC TTG CAG TCC CAA GAA GAA TCA GGC AGG GAG CAG AAA TCG CCC TCC TGT GAG GGA CAG CGG TAT CAG CAG GGA GAC TTT ATG AAT ACC CCA TGG AGA ACC CCA GCA GCA GAA GGG GAG AAA AAA TTG TAC AAG CAA CAA AAT ATG GAT GAT GTA GAT TCA GAT GAT GAT GAC CTA GTA GGG GTC CCT GTC ACA CCA AGA GTA CCA CTA AGA GCA ATG ACA TAT AAA TTG GCA GTA GAT ATG TCA CAT TTT ATA AAA GAA AAG GGG GGA CTG GAA GGG ATG TTT TAC AGT GAG AGA AGA CAT AGA ATC TTA GAC ATA TAC TTA GAA AAG GAA GAA GGG ATA ATT CCA GAT TGG CAG AAC TAT ACT CAT GGG CCA GGA ATA AGA TAC CCA ATG TTC TTT GGG TGG CTA TGG AAG CTA GTA CCA GTA GAT GTC CCA CAA GAG GGG GAG GAC AAT GAG ACT CAC TGC TTA ATG CAC CCA GCA CAA ACA AGC AGG TTT GAT GAC CCG CAT GGG GAA ACA CTA GTC TGG AGG TTT GAC CCC ATG CTG GCC TAT GAG TAC AAG GCC TTT ATT CGA TAC CCA GAG GAA TTT GGG CAC AAG TCA GGC CTG CCA GAA GAA GAG TGG AAG GCA AGA CTG AAA GCA AGA GGG ATA CCA TTT AGT GGA TCC GGA GGC GGA TCT GCT GGC TCC GCT GCT GGT TCT GGC GCA TTC TTA CCG GTC GAT GC

Table A.6 Binning profile used in SWARM analysis. Values correspond to the expression levels of Nef in arbitrary fluorescent units (AFU).

Midpoint	Lower Limit	Upper Limit
400	400	432
486	448	525
592	544	639
719	662	777
875	805	945
1064	979	1149
1293	1190	1397
1573	1447	1699
1913	1760	2066
2326	2140	2512
2828	2602	3055
3439	3164	3715
4183	3848	4517
5086	4679	5493
6185	5690	6680
7521	6920	8123
9146	8414	9878
11122	10232	12012
13525	12443	14607
16447	15131	17763
20000	18400	21600

Table A.7 Nef (immuno)proteasomal digestion products corresponding to known CTL epitopes reported in the LANL Immunology Database [401]. Epitope position is mapped to HXB2 Nef.

	Epitope	Virus	Peptide	Enzyme	Product Number
1	Nef-62-71-FNLPPIVAKEI	HIV2	p4	P	036
2	Nef-62-71-FNLPPIVAKEI	HIV2	p4	P	043
3	Nef-62-71-FNLPPIVAKEI	HIV2	p4	IP	041
4	Nef-72-81-PQVPLRPMTY	HIV1	p1	P	015
5	Nef-74-81-VPLRPMTF	HIV2	p1	P	068
6	Nef-74-81-VPLRPMTY	HIV2	p1	P	068
7	Nef-83-90-AAVDLSHF	HIV2	p2	P	063
8	Nef-83-90-AAVDLSHF	HIV1	p2	P	013
9	Nef-83-90-AAVDLSHF	HIV1	p2	IP	010
10	Nef-83-91-AAFDLSFFL	HIV2	p2	IP	008
11	Nef-83-91-AALDLSHFL	HIV2	p2	IP	008
12	Nef-83-91-AALDMSHFL	HIV2	p2	IP	008
13	Nef-83-91-AAVDLSHFL	HIV2	p2	IP	008
14	Nef-83-91-GAFDLSFFL	HIV2	p2	IP	008
15	Nef-83-91-GALDLSFFL	HIV2	p2	IP	008
16	Nef-83-91-GALDLSHFL	HIV2	p2	IP	008
17	Nef-83-91-GAVDLSHFL	HIV2	p2	IP	008

Continued on next page

Table A.7 – *Continued from previous page*

	Epitope	Virus	Peptide	Enzyme	Product Number
18	Nef-83-92-AAVDLSHFLK	HIV2	p2	P	036
19	Nef-83-92-GAFDLSFFLK	HIV2	p2	P	036
20	Nef-83-94-AAVDLSHFLKEK	HIV2	p2	P	082
21	Nef-83-94-AAVDLSHFLKEK	HIV2	p2	IP	034
22	Nef-83-94-AAVDLSHFLKEK	HIV2	p2	IP	086
23	Nef-84-91-ALDLSHFL	HIV2	p2	P	043
24	Nef-84-91-AVDLSHFL	HIV2	p2	P	043
25	Nef-86-94-DLSFLKEK	HIV2	p2	P	062
26	Nef-86-94-DLSFLKEK	HIV2	p2	IP	092
27	Nef-86-94-DLSHFLKEK	HIV2	p2	P	062
28	Nef-86-94-DLSHFLKEK	HIV2	p2	IP	092
29	Nef-92-105-KEKGGLEGLVYSQK	HIV2	p2	P	070
30	Nef-92-105-KEKGGLEGLVYSQK	HIV2	p2	IP	084
31	Nef-104-112-QRRQDILDL	HIV1	p2	P	005
32	Nef-107-115-QDILDLWIY	HIV1	p3	P	010
33	Nef-107-115-QEILDLWVY	HIV1	p3	P	010
34	Nef-109-117-ILDLWIYHT	HIV2	p3	P	065

Continued on next page

Table A.7 – *Continued from previous page*

	Epitope	Virus	Peptide	Enzyme	Product Number
35	Nef-109-117-ILDLWIYHT	HIV2	p3	IP	081
36	Nef-109-117-ILDLWVYHT	HIV2	p3	P	065
37	Nef-109-117-ILDLWVYHT	HIV2	p3	IP	081
38	Nef-115-124-YHTQGYFPDW	HIV2	p3	IP	010
39	Nef-115-125-YHTQGYFPDWQ	HIV2	p3	P	100
40	Nef-115-125-YHTQGYFPDWQ	HIV2	p3	IP	035
41	Nef-116-124-HTQGFFPDW	HIV2	p3	P	034
42	Nef-116-124-HTQGFFPDW	HIV2	p3	IP	018
43	Nef-116-124-HTQGFFPDW	HIV1	p3	P	027
44	Nef-116-124-HTQGYFPDW	HIV2	p3	P	034
45	Nef-116-124-HTQGYFPDW	HIV2	p3	IP	018
46	Nef-116-124-HTQGYFPDW	HIV1	p3	P	027
47	Nef-116-125-HTQGYFPDWQ	HIV2	p3	P	066
48	Nef-116-125-HTQGYFPDWQ	HIV2	p3	IP	067
49	Nef-116-125-HTQGYFPDWQ	HIV1	p3	IP	031
50	Nef-117-127-TQGYFPDWQNY	HIV2	p3	P	082

Continued on next page

Table A.7 – *Continued from previous page*

	Epitope	Virus	Peptide	Enzyme	Product Number
51	Nef-117-127-TQGYFPDWQNY	HIV1	p3	P	026
52	Nef-117-127-TQGYFPDWQNY	HIV1	p3	IP	015
53	Nef-118-127-QGYFPDWQNY	HIV2	p3	P	030
54	Nef-118-127-QGYFPDWQNY	HIV2	p3	IP	032
55	Nef-118-127-QGYFPDWQNY	HIV1	p3	IP	016
56	Nef-119-127-GYFPDWQNY	HIV2	p3	P	044
57	Nef-119-127-GYFPDWQNY	HIV2	p3	IP	030
58	Nef-119-127-GYFPDWQNY	HIV1	p3	P	020
59	Nef-120-127-YFPDWQNY	HIV1	p3	P	016
60	Nef-120-127-YFPDWQNY	HIV1	p3	IP	014
61	Nef-120-128-FFPDWKNYT	HIV1	p3	IP	030
62	Nef-120-128-YFPDWQDYT	HIV1	p3	IP	030
63	Nef-120-128-YFPDWQNYT	HIV1	p3	IP	030
64	Nef-127-135-YTPGPGIRY	HIV2	p3	P	099
65	Nef-127-135-YTPGPGTRY	HIV2	p3	P	099
66	Nef-127-135-YTPGPGVRY	HIV2	p3	P	099

Continued on next page

Table A.7 – *Continued from previous page*

	Epitope	Virus	Peptide	Enzyme	Product Number
67	Nef-128-135-TPGPGVRY	HIV2	p3	P	069
68	Nef-128-135-TPGPGVRY	HIV1	p3	P	022
69	Nef-128-136-TPGPGVRYP	HIV1	p3	P	014
70	Nef-130-139-GPGVRYPLTF	HIV1	p4	P	011
71	Nef-133-141-GRFPLTFGW	HIV1	p4	IP	052
72	Nef-133-141-TRYPLTFGW	HIV1	p4	IP	052
73	Nef-133-141-VRYPLTFGW	HIV1	p4	IP	052
74	Nef-134-141-RFPLTFGW	HIV2	p4	P	038
75	Nef-134-141-RFPLTFGW	HIV1	p4	IP	046
76	Nef-134-141-RYPLTFGW	HIV2	p4	P	038
77	Nef-134-141-RYPLTFGW	HIV1	p4	IP	046

IP refers to immunoproteasome; P refers to proteasome.
Digestion products correspond to epitopes in terms of location, but not necessarily primary amino acid sequence.

Table A.8 Accession numbers of sequences used from published sources.

Use (n)	Accession Numbers
Generation of Chinese FPD *gp120* alignment and consensus sequence (512)	AY732040 AY732041 AY732042 AY732043 AY732044 AY732045 AY732046 AY732047 AY732048 AY732049 AY732050 AY732051 AY732052 AY732053 AY732054 AY732055 AY732056 AY732057 AY732058 AY732059 AY732060 AY732061 AY732062 AY732063 AY732064 AY732065 AY732066 AY732067 AY732068 AY732069 AY732070 AY732071 AY732072 AY732073 AY732074 AY732075 AY732076 AY732077 AY732078 AY732079 AY732080 AY732081 AY732082 AY732083 AY732084 AY732085 AY732086 AY732087 AY732088 AY732089 AY732090 AY732091 AY732092 AY732093 AY732094 AY732095 AY732096 AY732097 AY732098 AY732099 DQ007901 DQ007902 DQ007903 DQ493967 DQ493968 DQ493969 DQ493970 DQ493971 DQ493972 DQ493973 DQ493974 DQ493975 DQ493976 DQ493977 DQ493978 DQ493979 DQ493980 DQ493981 DQ493982 DQ493983 DQ493984 DQ493985 DQ493986 DQ493987 DQ493988 DQ493989 DQ493990 DQ493991 DQ493992 DQ493993 DQ493994 DQ493995 DQ493996 DQ493997 DQ493998 DQ493999 DQ494000 DQ494001 DQ494002 DQ494003 DQ494004 DQ494005 DQ494006 DQ494007 DQ494008 DQ494009 DQ494010 DQ494011 DQ494012 DQ494013 DQ494014 DQ494015 DQ494016 DQ494017 DQ494018 DQ494019 DQ494020 DQ494021 DQ494022 DQ494023 DQ494024 DQ494025 DQ494026 DQ494027 DQ494028 DQ494029 DQ494030 DQ494031 DQ494032 DQ494033 DQ494034 DQ494035 DQ494036 DQ494037 DQ494038 DQ494039 DQ494040 DQ494041 DQ494042 DQ494043 DQ494044 DQ494045 DQ494046 DQ494047 DQ494048 DQ494049 DQ494050 DQ494051 DQ494052 DQ494053 DQ494054 DQ494055 DQ494056 DQ494057 DQ494058 DQ494059 DQ494060 DQ494061 DQ494062 DQ494063 DQ494064 DQ494065 DQ494066 DQ494067 DQ494068 DQ494069 DQ494070 DQ494071 DQ494072 DQ494073 DQ494074 DQ494075 DQ494076 DQ494077 DQ494078 DQ494079 DQ494080 DQ494081 DQ494082 DQ494083 DQ494084 DQ494085 DQ494086 DQ494087 DQ494088 DQ494089 DQ494090 DQ494091 DQ494092 DQ494093 DQ494094 DQ494095 DQ494096 DQ494097 DQ494098 DQ494099 DQ494100 DQ494101 DQ494102 DQ494103 DQ494104 DQ494105 DQ494106 DQ494107 DQ494108 DQ494109 DQ494110 DQ494111 DQ494112 DQ494113 DQ494114 DQ494115 DQ494116 DQ494117 DQ494118 DQ494119 DQ494120 DQ494121 DQ494122 DQ494123 DQ494124 DQ494125 DQ494126 DQ494127 DQ494128 DQ494129 DQ494130 DQ494131 DQ494132 DQ494133 DQ494134 DQ494135 DQ494136 DQ494137 DQ494138 DQ494139 DQ494140 DQ494141 EU272190 EU272191 EU272192 EU272193 EU272194 EU272195 EU272196 EU272197 EU272198 EU272199 EU272200 EU272201 EU272202 EU272203 EU272204 EU272205 EU272206 EU272207 EU272208 EU272209 EU272210 EU272211 EU272212 EU272213 EU272214 EU272215 EU272216 EU272217 EU272218 EU272219 EU272220 EU272221 EU272222 EU272223 EU272224 EU272225 EU272226 EU272227 EU272228 EU272229 EU272230 EU272231 EU272232 EU272233 EU272234 EU272235 EU272236 EU272237 EU272238 EU272239 EU272240 EU272241 EU272242 EU272243 EU272244 EU272245 EU272246 EU272247 EU272248 EU272249 EU272250 EU272251 EU272252 EU272253 EU272254 EU272255 EU272256 EU272257 EU272258 EU272259 EU272260 EU272261 EU272262 EU272263 EU272264 EU272265 EU272266 EU272267 EU272268 EU272269 EU272270 EU272271 EU272272 EU272273 EU272274 EU272275 EU272276 EU272277 EU272278 EU272279 EU272280 EU272281 EU272282 EU272283 EU272284 EU272285 EU272286 EU272287 EU272288 EU272289 EU272290 EU272291 EU272292 EU272293 EU272294 EU272295 EU272296 EU272297 EU272298 EU272299 EU272300 EU272301 EU272302 EU272303 EU272304 EU272305 EU272306 EU272307 EU272308 EU272309 EU272310 EU272311 EU272312 EU272313 EU272314 EU272315 EU272316 EU272317 EU272318 EU272319 EU272320 EU272321 EU272322 EU272323 EU272324 EU272325 EU272326 EU272327 EU272328 EU272329 EU272330 EU272331 EU272332 EU272333 EU272334 HM215397 HM215398 HM215399 HM215400 HM215401 HM215402 HM215403 HM215404 HM215405 HM215406 HM215407 HM215408 HM215409 HM215410 HM215411 HM215412 HM215413 HM215414 HM215415 HM215416 HM215417 HM215418 HM215419 HM215420 HM215421 HM215422 HM215423 HM215424 HM215425 HM215426 HM215427 HM215428 HQ699949 HQ699950 HQ699951 HQ699952 HQ699953 HQ699954 HQ699955 HQ699956 HQ699957 HQ699958 HQ699959 HQ699960 HQ699961 HQ699962 HQ699963 HQ699964 HQ699965 HQ699966 HQ699967 HQ699968 HQ699969 HQ699970 HQ699971 HQ699972 HQ699973 HQ699974 HQ699975 HQ699976 HQ699977 HQ699978 HQ699979 HQ699980 HQ699981 HQ699982 HQ699983 HQ699984 HQ699985 HQ699986 HQ699987 HQ699988 HQ699989 HQ699990 HQ699991 HQ699992 HQ699993 HQ699994 HQ699995 HQ699996 HQ699997 HQ699998 HQ699999 HQ700000 HQ700001 HQ700002 HQ700003 HQ700004 JF773491 JF773492 JF773493 JF773494 JF773495 JF773496 JF773497 JF773502 JF773503 JF773504 JF773505 JF773506 JF773507 JF773508 JF773509 JF773518 JF773519 JF773520 JF773521

Continued on next page

Table A.8 – *Continued from previous page*

Use (n)	Accession Numbers
	JF773522 JF773523 JF773530 JF773531 JF773532 JF773533 JF773534 JF773535 JF773542 JF773543 JF773544 JF773545 JF773546 JF773547 JF773548 JF773549 JF773555 JF773556 JF773557 JF773558 JF773559 JF773560
Viral subtyping of SM cohort *gp120* sequences (62)	AF004885 AF005494 AF005496 AF049337 AF061641 AF061642 AF064699 AF067155 AF069670 AF075703 AF077336 AF082394 AF082395 AF084936 AF179368 AF190127 AF190128 AF193253 AF193276 AF286237 AF286238 AF286239 AF289548 AF377956 AF377959 AF385936 AF423756 AF460972 AF484509 AF516184 AJ006022 AJ249235 AJ249236 AJ249237 AJ249238 AJ249239 AJ271370 AJ302647 AX149771 AY008715 AY093605 AY169812 AY173951 AY253311 AY331295 AY371157 AY371158 AY423387 AY532635 AY588970 AY772699 K03454 K03455 L20571 L20587 L39106 U46016 U51190 U52953 U54771 U88824 U88826
Generation of HIV-1 subtype B *nef* consensus sequence (3,086)	A04321 AB012824 AB026242 AB034252 AB034257 AB034258 AB034263 AB034268 AB034272 AB034273 AB034279 AB034282 AB034286 AB034290 AB034295 AB034299 AB078005 AB097870 AB221005 AB221125 AB286955 AB287363 AB287364 AB287366 AB287368 AB287369 AB289587 AB289589 AB428551 AB428552 AB428553 AB428554 AB428555 AB428556 AB428557 AB428558 AB428559 AB428560 AB428561 AB428562 AB480692 AB480694 AB480696 AB480698 AB485638 AB485641 AB564744 AB564746 AB565478 AB565495 AB565497 AB565499 AB565501 AB604946 AB604948 AB604950 AB641836 AB731663 AB731665 AB731667 AB873602 AB873603 AB873604 AB873605 AB873606 AB873607 AB873608 AB873609 AB873610 AB873611 AB873612 AB873613 AB873614 AB873615 AB873616 AB873617 AB873618 AB873619 AB873620 AB873621 AB873622 AB873623 AB873624 AB873625 AB873626 AB873627 AB873628 AB873629 AB873630 AB873631 AB873632 AB873633 AB873634 AB873635 AB873636 AB873637 AB873638 AB873639 AB873640 AB873641 AB873642 AB873643 AB873644 AB873645 AB873646 AB873647 AB873648 AB873649 AB873650 AB873651 AB873652 AB873653 AB873654 AB873655 AB873656 AB873657 AB873658 AB873659 AB873660 AB873661 AB873662 AB873663 AB873664 AB873665 AB873666 AB873667 AB873668 AB873669 AB873670 AB873671 AB873672 AB873673 AB873674 AB873675 AB873676 AB873677 AB873678 AB873679 AB873680 AB873681 AB873682 AB873683 AB873684 AB873685 AB873686 AB873687 AB873688 AB873689 AB873690 AB873691 AB873692 AB873693 AB873694 AB873695 AB873696 AB873697 AB873698 AB873699 AB873700 AB873701 AB873702 AB873703 AB873704 AB873705 AB873706 AB873707 AB873708 AB873709 AB873710 AB873711 AB873712 AB873713 AB873714 AB873715 AB873716 AB873717 AB873718 AB873719 AB873720 AB873721 AB873722 AB873723 AB873724 AB873725 AB873726 AB873727 AB873728 AB873729 AB873730 AB873731 AB873732 AB873733 AB873734 AB873735 AB873736 AB873737 AB873738 AB873739 AB873740 AB873741 AB873742 AB873743 AB873744 AB873745 AB873746 AB873747 AB873748 AB873749 AB873750 AB873751 AB873752 AB873753 AB873754 AB873755 AB873756 AB873757 AB873758 AB873759 AB873760 AB873761 AB873762 AB873763 AB873764 AB873766 AB873767 AB873768 AB873769 AB873770 AB873771 AB873772 AB873773 AB873774 AB873775 AB873776 AB873777 AB873778 AB873779 AB873780 AB873781 AB873782 AB873783 AB873784 AB873785 AB873786 AB873787 AB873788 AB873789 AB873790 AB873791 AB873792 AB873793 AB873794 AB873795 AB873796 AB873797 AB873798 AB873799 AB873800 AB873801 AB873802 AB873803 AB873804 AB873805 AB873806 AB873807 AB873808 AB873809 AB873810 AB873811 AB873812 AB873813 AB873814 AB873815 AB873816 AB873817 AB873818 AB873819 AB873820 AB873821 AB873822 AB873823 AB873824 AB873825 AB873826 AB873827 AB873828 AB873829 AB873830 AB873831 AB873832 AB873833 AB873834 AB873835 AB873836 AB873837 AB873838 AB873839 AB873840 AB873841 AB873842 AB873843 AB873844 AB873845 AB873846 AB873847 AB873848 AB873849 AB873850 AB873851 AB873852 AB873853 AB873854 AB873855 AB873856 AB873857 AB873858 AB873859 AB873860 AB873861 AB873862 AB873863 AB873864 AB873865 AB873866 AB873867 AB873868 AB873869 AB873870 AB873871 AB873872 AB873873 AB873874 AB873875 AB873876 AB873877 AB873878 AB873879 AB873880 AB873881 AB873882 AB873883 AB873884 AB873885 AB873886 AB873887 AB873888 AB873889 AB873890 AB873891 AB873892 AB873893 AB873894 AB873895 AB873896 AB873897 AB873898 AB873899 AB873900 AB873901 AB873902 AB873903 AB873904 AB873905 AB873906

Continued on next page

Table A.8 – *Continued from previous page*

Use (n)	Accession Numbers
	AB873907 AF003887 AF004394 AF011469 AF011471 AF011473 AF011474 AF011477 AF011480 AF011481 AF011483 AF011486 AF011487 AF011488 AF011492 AF011493 AF011494 AF033567 AF033568 AF033569 AF033570 AF033571 AF033572 AF042100 AF042101 AF047080 AF047081 AF047082 AF047084 AF047085 AF047086 AF047087 AF049494 AF049495 AF063909 AF063910 AF063911 AF063912 AF063913 AF063914 AF063915 AF063916 AF063918 AF063919 AF063921 AF063922 AF063923 AF063924 AF063925 AF063926 AF063927 AF063928 AF063929 AF063931 AF063932 AF063933 AF064660 AF064661 AF064662 AF064663 AF064664 AF064665 AF064666 AF064667 AF064668 AF064669 AF064670 AF064671 AF064672 AF064673 AF064674 AF064675 AF064676 AF064677 AF064678 AF064679 AF064680 AF064681 AF069139 AF070521 AF086817 AF120745 AF120772 AF120789 AF120801 AF120811 AF120824 AF120830 AF120840 AF120851 AF120867 AF120876 AF120887 AF120898 AF120909 AF129333 AF129334 AF129335 AF129336 AF129337 AF129338 AF129339 AF129340 AF129342 AF129343 AF129344 AF129345 AF129346 AF129347 AF129348 AF129349 AF129350 AF129351 AF129352 AF129353 AF129354 AF129355 AF129356 AF129358 AF129359 AF129360 AF129361 AF129362 AF129363 AF129364 AF129365 AF129366 AF129367 AF129368 AF129369 AF129370 AF129372 AF129373 AF129375 AF129376 AF129377 AF129378 AF129379 AF129381 AF129382 AF129386 AF129387 AF129388 AF129389 AF129390 AF129391 AF129392 AF129393 AF129394 AF146728 AF147728 AF147729 AF147730 AF147731 AF147732 AF147733 AF147734 AF147735 AF147736 AF147737 AF203108 AF203111 AF203116 AF203117 AF203126 AF203137 AF203141 AF203145 AF203153 AF203157 AF203161 AF203165 AF203169 AF203172 AF203175 AF203180 AF203183 AF203188 AF203192 AF203194 AF203198 AF203200 AF219637 AF219644 AF219654 AF219662 AF219673 AF219678 AF219685 AF219691 AF219696 AF219704 AF219710 AF219715 AF219720 AF219726 AF219735 AF219741 AF219747 AF219751 AF219756 AF219761 AF219766 AF219771 AF219777 AF219782 AF219786 AF219792 AF219796 AF219801 AF219807 AF219813 AF219819 AF219831 AF219837 AF219846 AF219851 AF219857 AF224507 AF238268 AF238273 AF238274 AF238276 AF252897 AF252910 AF252926 AF252938 AF252949 AF252961 AF256205 AF272005 AF286365 AF397200 AF397201 AF462701 AF462703 AF462704 AF462705 AF462706 AF462707 AF462708 AF462714 AF462715 AF462720 AF462721 AF462726 AF462727 AF462730 AF462733 AF462740 AF462741 AF462742 AF462748 AF462749 AF462750 AF462751 AF462753 AF462754 AF462755 AF462758 AF462759 AF462760 AF462761 AF462763 AF462764 AF462765 AF462767 AF462768 AF462769 AF462771 AF462774 AF462778 AF462779 AF462782 AF462787 AF462788 AF462789 AF462791 AF462792 AF462793 AF538302 AF538303 AF538304 AF538305 AF538306 AF538307 AJ271445 AJ430646 AJ430661 AJ430664 AJ430667 AJ850861 AJ850870 AJ850871 AJ850883 AJ850885 AJ850886 AJ850906 AJ850907 AJ850908 AJ850909 AJ850911 AJ850913 AY005982 AY006064 AY037268 AY037269 AY037270 AY037282 AY093617 AY116676 AY116694 AY116714 AY116729 AY116734 AY116748 AY116754 AY116782 AY116794 AY116806 AY116818 AY116831 AY116843 AY121441 AY121450 AY121476 AY121477 AY173951 AY173952 AY173954 AY173955 AY180905 AY221668 AY221682 AY221687 AY221692 AY221696 AY221698 AY221705 AY221708 AY221711 AY221714 AY260782 AY260812 AY308760 AY314044 AY331282 AY331284 AY331285 AY331287 AY331289 AY331291 AY331292 AY331294 AY331296 AY352275 AY363309 AY363367 AY423381 AY444305 AY444307 AY444308 AY444310 AY444311 AY444312 AY444313 AY444314 AY444315 AY444316 AY444317 AY444318 AY496646 AY560107 AY560108 AY560109 AY560110 AY561236 AY561237 AY561238 AY561239 AY561240 AY561243 AY561245 AY586542 AY586543 AY682547 AY701254 AY701255 AY701256 AY701257 AY701262 AY701263 AY701264 AY701265 AY701266 AY701267 AY701268 AY701269 AY701270 AY701271 AY701272 AY701275 AY701276 AY701277 AY701278 AY701280 AY701281 AY701282 AY701283 AY701285 AY713408 AY713410 AY713411 AY713412 AY739026 AY739027 AY739029 AY739033 AY739034 AY739035 AY739036 AY739040 AY739041 AY739046 AY739047 AY739048 AY751406 AY751407 AY779550 AY779553 AY779557 AY781126 AY781127 AY786591 AY786601 AY786631 AY786641 AY786670 AY786680 AY786710 AY786724 AY786750 AY786760 AY795904 AY795905 AY818644 AY819715 AY835748 AY835753 AY835754 AY835755 AY835757 AY835759 AY835762 AY835766 AY835769 AY835770 AY835774 AY835777 AY835779 AY856759 AY856761 AY856762 AY856765 AY856766 AY856767 AY856768 AY856769 AY856770 AY856771 AY856772 AY856773 AY856775 AY856776 AY857022 AY857052 AY857127 AY857144 AY857165 AY857173 AY899356 AY899359 AY899382 AY905389 AY905390 AY905392 AY905394 AY905395 AY905396 AY905400

Continued on next page

Table A.8 – *Continued from previous page*

Use (n)	Accession Numbers
	AY970946 BD298114 BD493365 D10112 D86069 DQ007901 DQ007902 DQ007903 DQ121678 DQ121681 DQ121754 DQ121756 DQ121762 DQ121775
	DQ121801 DQ121819 DQ121822 DQ121870 DQ121881 DQ121912 DQ121927 DQ121928 DQ121931 DQ121957 DQ121963 DQ121965 DQ121984 DQ121986
	DQ121987 DQ121991 DQ121999 DQ122019 DQ122026 DQ122027 DQ122028 DQ122029 DQ122031 DQ122032 DQ122041 DQ122043 DQ122045 DQ122061
	DQ122069 DQ122075 DQ122088 DQ122093 DQ122096 DQ122098 DQ122100 DQ122102 DQ122108 DQ122110 DQ122111 DQ122112 DQ122113 DQ122114
	DQ122115 DQ122116 DQ122118 DQ122122 DQ127535 DQ127537 DQ127540 DQ127548 DQ207940 DQ207942 DQ207943 DQ242535 DQ322223 DQ322225
	DQ322227 DQ322233 DQ345398 DQ345402 DQ345404 DQ345405 DQ345407 DQ345408 DQ345409 DQ345410 DQ345411 DQ345412 DQ345413 DQ345414
	DQ345416 DQ345417 DQ345428 DQ354112 DQ354116 DQ354118 DQ354119 DQ357219 DQ357220 DQ358805 DQ358808 DQ358809 DQ358810 DQ383746
	DQ383748 DQ383749 DQ383750 DQ383751 DQ383752 DQ396398 DQ400965 DQ400967 DQ484085 DQ484090 DQ484096 DQ484098 DQ484101 DQ484102
	DQ484105 DQ484107 DQ484108 DQ484112 DQ484116 DQ484170 DQ484173 DQ484182 DQ484183 DQ484184 DQ484188 DQ484189 DQ484193 DQ484263
	DQ484266 DQ484267 DQ484272 DQ484281 DQ484283 DQ484284 DQ484287 DQ484288 DQ484289 DQ484291 DQ484339 DQ484345 DQ484349 DQ484350
	DQ484352 DQ484353 DQ484355 DQ484356 DQ484357 DQ484361 DQ484362 DQ484364 DQ484365 DQ484370 DQ484423 DQ484434 DQ484436 DQ484441
	DQ484444 DQ484445 DQ484451 DQ484452 DQ484457 DQ484460 DQ484462 DQ484463 DQ484523 DQ484526 DQ484529 DQ484532 DQ484536 DQ484574
	DQ484578 DQ484580 DQ484582 DQ484583 DQ484584 DQ484587 DQ484590 DQ484591 DQ484593 DQ484595 DQ484596 DQ484597 DQ484600 DQ484602
	DQ484604 DQ484605 DQ484609 DQ484613 DQ484614 DQ484615 DQ484618 DQ484622 DQ484623 DQ484624 DQ484625 DQ484628 DQ484629 DQ484630
	DQ484632 DQ484634 DQ484636 DQ484641 DQ484643 DQ484647 DQ484648 DQ484649 DQ484693 DQ484697 DQ484699 DQ484701 DQ484703 DQ484704
	DQ484706 DQ484707 DQ484711 DQ484714 DQ484716 DQ484717 DQ484718 DQ484719 DQ484721 DQ484722 DQ484724 DQ484725 DQ484729 DQ484730
	DQ484733 DQ484740 DQ484741 DQ484742 DQ484747 DQ484779 DQ484792 DQ484794 DQ484795 DQ484797 DQ484798 DQ484800 DQ484801 DQ484802
	DQ484804 DQ484807 DQ484810 DQ484811 DQ484814 DQ484815 DQ484816 DQ484817 DQ484818 DQ484819 DQ484820 DQ484821 DQ484822 DQ484823
	DQ484824 DQ484825 DQ484828 DQ484829 DQ484830 DQ484831 DQ484832 DQ484833 DQ484834 DQ484899 DQ484914 DQ484915 DQ484916 DQ484917
	DQ484918 DQ484919 DQ484920 DQ484921 DQ484924 DQ484925 DQ484927 DQ484929 DQ484930 DQ484934 DQ484955 DQ484994 DQ484996 DQ484997
	DQ484999 DQ485002 DQ485003 DQ485004 DQ485006 DQ485007 DQ485011 DQ485012 DQ485015 DQ485016 DQ485017 DQ485019 DQ485020 DQ485021
	DQ485023 DQ485053 DQ485072 DQ485073 DQ485074 DQ485076 DQ485077 DQ485078 DQ485079 DQ485080 DQ485082 DQ485083 DQ485084 DQ485087
	DQ485088 DQ485089 DQ485091 DQ485092 DQ485093 DQ485094 DQ485095 DQ485097 DQ485103 DQ487188 DQ487190 DQ659700 DQ659701 DQ659703
	DQ659706 DQ659710 DQ659725 DQ659731 DQ659733 DQ659737 DQ676879 DQ823362 DQ823363 DQ823364 DQ853427 DQ853455 DQ886031 DQ886032
	DQ886033 DQ886034 DQ886035 DQ886036 DQ886037 DQ990880 DQ996256 DQ996258 DQ996260 DQ996262 DQ996264 DQ996266 EF090282 EF090283
	EF090284 EF090285 EF090286 EF119608 EF121411 EF125652 EF178338 EF178358 EF363122 EF363124 EF363125 EF363127 EF462995 EF462996
	EF462997 EF462998 EF462999 EF463000 EF463001 EF463002 EF463003 EF463004 EF514697 EF514699 EF514700 EF514701 EF514702 EF514703
	EF514704 EF514705 EF514706 EF514707 EF514708 EF514710 EF514711 EF514712 EF637046 EF637047 EF637048 EF637049 EF637050 EF637051
	EF637053 EF637054 EF637056 EF637057 EF656966 EF656994 EF657007 EF657035 EF657063 EF657072 EF657107 EF694037 EU180773 EU180796
	EU312168 EU312169 EU312170 EU312171 EU312172 EU312173 EU312174 EU312176 EU312184 EU312185 EU327403 EU327409 EU327415 EU327419
	EU327428 EU327439 EU327445 EU327449 EU327486 EU410053 EU432531 EU432532 EU432533 EU432534 EU432535 EU432536 EU432537 EU432538
	EU432539 EU432540 EU432541 EU432542 EU432543 EU432544 EU432545 EU541617 EU547186 EU616640 EU786672 EU786674 EU786675 EU786676
	EU786677 EU786678 EU786679 EU786680 EU807775 EU807782 FB299025 FJ195086 FJ195088 FJ195089 FJ195090 FJ195091 FJ347134 FJ347135
	FJ423554 FJ423555 FJ423556 FJ423558 FJ423559 FJ423560 FJ423561 FJ423562 FJ423563 FJ423564 FJ423568 FJ423569 FJ430356 FJ430395
	FJ460499 FJ460500 FJ460501 FJ469682 FJ469683 FJ469684 FJ469685 FJ469686 FJ469687 FJ469688 FJ469689 FJ469690 FJ469691 FJ469692

Continued on next page

Table A.8 – *Continued from previous page*

Use (n)	Accession Numbers
	FJ469693 FJ469694 FJ469695 FJ469696 FJ469697 FJ469698 FJ469699 FJ469700 FJ469701 FJ469702 FJ469703 FJ469704 FJ469705 FJ469706 FJ469707 FJ469708 FJ469709 FJ469710 FJ469711 FJ469713 FJ469714 FJ469715 FJ469716 FJ469717 FJ469718 FJ469719 FJ469720 FJ469721 FJ469722 FJ469723 FJ469724 FJ469725 FJ469726 FJ469727 FJ469729 FJ469730 FJ469731 FJ469732 FJ469733 FJ469735 FJ469737 FJ469738 FJ469739 FJ469740 FJ469741 FJ469742 FJ469743 FJ469744 FJ469745 FJ469747 FJ469748 FJ469749 FJ469750 FJ469751 FJ469752 FJ469753 FJ469754 FJ469755 FJ469756 FJ469757 FJ469758 FJ469759 FJ469760 FJ469761 FJ469762 FJ469763 FJ469764 FJ469765 FJ469766 FJ469767 FJ469768 FJ469769 FJ469770 FJ469771 FJ469772 FJ471504 FJ471508 FJ471513 FJ471524 FJ495804 FJ495937 FJ496000 FJ496072 FJ496081 FJ496145 FJ496151 FJ496158 FJ496167 FJ647145 FJ670524 FJ670525 FJ670531 FJ694790 FJ798354 FJ798366 FJ798391 FJ798405 FJ798415 FJ798433 FJ798457 FJ798474 FJ798503 FJ798521 FJ798544 FJ822828 FJ822837 FJ822838 FJ822839 FJ822840 FJ822841 FJ822842 FJ822844 FJ822845 FJ822846 FJ822847 FJ822848 FJ822850 FJ822851 FJ822853 FJ822857 FJ822858 FJ822859 FJ822860 FJ822861 FJ822862 FJ822869 FJ822871 FJ822876 FJ853620 FJ853622 GQ256631 GQ256632 GQ372320 GQ372322 GQ372323 GQ372325 GQ372326 GQ372330 GQ372331 GQ372332 GQ372336 GQ372339 GQ372344 GQ372345 GQ372347 GQ372348 GQ372351 GQ372352 GQ372355 GQ372358 GQ372360 GQ372361 GQ372363 GQ372366 GQ372373 GQ372374 GQ372378 GQ372380 GQ372381 GQ372382 GQ372383 GQ372384 GQ372385 GQ372386 GQ372389 GQ372392 GQ372398 GQ372399 GQ372401 GQ372402 GQ372403 GQ372405 GQ372409 GQ372410 GQ372418 GQ372419 GQ372421 GQ372422 GQ372424 GQ372426 GQ372427 GQ372431 GQ372432 GQ372439 GQ372440 GQ372441 GQ372442 GQ372443 GQ372444 GQ372445 GQ372446 GQ372447 GQ372450 GQ372454 GQ372457 GQ372458 GQ372459 GQ372460 GQ372462 GQ372469 GQ372470 GQ372471 GQ372472 GQ372473 GQ372474 GQ372475 GQ372477 GQ372479 GQ372481 GQ372483 GQ372485 GQ372486 GQ372488 GQ372489 GQ372490 GQ372491 GQ372493 GQ372495 GQ372497 GQ372500 GQ372501 GQ372502 GQ372503 GQ372505 GQ372506 GQ372507 GQ372510 GQ372512 GQ372515 GQ372517 GQ372518 GQ372519 GQ372522 GQ372532 GQ372535 GQ372537 GQ372539 GQ372540 GQ372544 GQ372545 GQ372550 GQ372551 GQ372553 GQ372556 GQ372558 GQ372559 GQ372560 GQ372562 GQ372565 GQ372567 GQ372569 GQ372570 GQ372572 GQ372573 GQ372577 GQ372579 GQ372582 GQ372583 GQ372585 GQ372589 GQ372590 GQ372591 GQ372594 GQ372605 GQ372607 GQ372611 GQ372613 GQ372615 GQ372619 GQ372622 GQ372623 GQ372624 GQ372628 GQ372630 GQ372634 GQ372635 GQ372636 GQ372638 GQ372642 GQ372645 GQ372646 GQ372647 GQ372648 GQ372655 GQ372657 GQ372660 GQ372664 GQ372667 GQ372670 GQ372672 GQ372673 GQ372674 GQ372675 GQ372676 GQ372680 GQ372681 GQ372682 GQ372685 GQ372687 GQ372688 GQ372689 GQ372691 GQ372693 GQ372695 GQ372696 GQ372698 GQ372700 GQ372701 GQ372702 GQ372704 GQ372706 GQ372707 GQ372711 GQ372713 GQ372716 GQ372721 GQ372728 GQ372733 GQ372735 GQ372743 GQ372744 GQ372745 GQ372746 GQ372747 GQ372749 GQ372751 GQ372753 GQ372759 GQ372760 GQ372762 GQ372763 GQ372765 GQ372766 GQ372767 GQ372770 GQ372773 GQ372774 GQ372780 GQ372782 GQ372789 GQ372790 GQ372796 GQ372798 GQ372800 GQ372803 GQ372812 GQ372820 GQ372988 GQ372990 GQ398383 GU017575 GU017576 GU017579 GU017585 GU017587 GU017591 GU017593 GU017594 GU017596 GU017600 GU017601 GU017602 GU017604 GU017605 GU017619 GU017624 GU046566 GU046567 GU046568 GU046569 GU046570 GU046571 GU046572 GU046573 GU046574 GU046575 GU046576 GU046577 GU046578 GU046579 GU046580 GU046581 GU046582 GU046583 GU046584 GU046585 GU046586 GU046587 GU046588 GU046589 GU046590 GU046591 GU046592 GU046593 GU046594 GU046595 GU046596 GU046597 GU046598 GU046599 GU046600 GU046601 GU046602 GU046603 GU046604 GU177863 GU331096 GU331218 GU362881 GU362883 GU362885 GU362886 GU367444 GU367447 GU367449 GU367450 GU367455 GU367456 GU367457 GU367459 GU367460 GU457024 GU562001 GU562028 GU562058 GU562080 GU562105 GU562134 GU562153 GU562218 GU562249 GU562271 GU733713 HM030559 HM030560 HM030561 HM030562 HM030563 HM030564 HM030565 HM153284 HM237163 HM244483 HM244495 HM244499 HM244504 HM244509 HM244531 HM244533 HM244537 HM244541 HM244545 HM244554 HM244557 HM244560 HM244563 HM244566 HM244579 HM244584 HM244608 HM244612 HM244615 HM244618 HM244632 HM244635 HM244639 HM244642 HM244645 HM244649 HM244653 HM244657 HM244662 HM244673 HM244679 HM369898 HM586206 HM586210 HM586213 HM586451 HM586729 HQ174337 HQ174338 HQ174343 HQ174350 HQ215554 HQ215556 HQ448783 HQ846906 JF320002 JF320007

Continued on next page

Table A.8 – *Continued from previous page*

Use (n)	Accession Numbers
	JF320008 JF320011 JF320012 JF320017 JF320018 JF320028 JF320036 JF320038 JF320043 JF320044 JF320045 JF320048 JF320054 JF320059
	JF320072 JF320083 JF320099 JF320120 JF320131 JF320144 JF320145 JF320150 JF320160 JF320173 JF320179 JF320183 JF320184 JF320185
	JF320186 JF320189 JF320191 JF320196 JF320197 JF320198 JF320230 JF320241 JF320259 JF320276 JF320308 JF320315 JF320329 JF320356
	JF320361 JF320363 JF320375 JF320380 JF320387 JF320394 JF320411 JF320422 JF320427 JF320460 JF320482 JF320497 JF320520 JF320530
	JF320539 JF320559 JF320564 JF320570 JF320592 JF320613 JF320615 JF320630 JF689852 JF689854 JF689856 JF689857 JF689859 JF689860
	JF689862 JF689863 JF689864 JF689865 JF689866 JF689867 JF689868 JF689870 JF689871 JF689872 JF689873 JF689874 JF689875 JF689876
	JF689877 JF689879 JF689883 JF689884 JF689885 JF689886 JF689887 JF689889 JF689890 JF689892 JF689893 JF689894 JF689895 JF689896
	JF689897 JF932468 JF932469 JF932470 JF932471 JF932472 JF932473 JF932474 JF932475 JF932476 JF932477 JF932478 JF932479 JF932480
	JF932481 JF932482 JF932483 JF932484 JF932485 JF932486 JF932487 JF932488 JF932489 JF932490 JF932491 JF932492 JF932493 JF932494
	JF932495 JF932496 JF932497 JF932498 JF932499 JF932500 JN024100 JN024203 JN024274 JN024339 JN024349 JN024422 JN024450 JN024459
	JN024469 JN235958 JN235959 JN235965 JN248321 JN248329 JN248333 JN248335 JN248337 JN248343 JN248344 JN248345 JN248346 JN248347
	JN248348 JN248353 JN248354 JN251896 JN251901 JN251906 JN397362 JN397363 JN397364 JN397365 JN400469 JN599164 JN687739 JN687749
	JN687750 JN687758 JN687759 JN687760 JN687761 JN687762 JN687763 JN687773 JN687774 JN687847 JN692431 JN692432 JN692433 JN692435
	JN692436 JN692439 JN692443 JN692444 JN692445 JN692446 JN692447 JN692450 JN692451 JN692452 JN692453 JN692455 JN692457 JN692459
	JN692460 JN692461 JN692462 JN692463 JN692465 JN692467 JN692468 JN692470 JN692471 JN692472 JN692473 JN692474 JN692475 JN692476
	JN692479 JN692480 JN860769 JN944897 JN944911 JN944917 JN944930 JQ007433 JQ007434 JQ007441 JQ007443 JQ007444 JQ007451 JQ007452
	JQ007454 JQ007458 JQ007462 JQ007464 JQ007465 JQ007467 JQ007469 JQ007470 JQ007471 JQ007474 JQ007475 JQ007477 JQ007478 JQ007482
	JQ007483 JQ007484 JQ007487 JQ007488 JQ007489 JQ007490 JQ007492 JQ007494 JQ007496 JQ007498 JQ007499 JQ007500 JQ007501 JQ007502
	JQ007505 JQ255622 JQ255638 JQ255654 JQ255713 JQ255883 JQ255977 JQ256020 JQ256039 JQ256068 JQ256214 JQ256215 JQ256223 JQ256228
	JQ256235 JQ256258 JQ256268 JQ256301 JQ256339 JQ403019 JQ403021 JQ403022 JQ403023 JQ403024 JQ403025 JQ403026 JQ403029 JQ403030
	JQ403035 JQ403037 JQ403038 JQ403039 JQ403041 JQ403042 JQ403043 JQ403044 JQ403045 JQ403046 JQ403047 JQ403048 JQ403049 JQ403050
	JQ403056 JQ403058 JQ403059 JQ403060 JQ403061 JQ403062 JQ403064 JQ403066 JQ403067 JQ403068 JQ403069 JQ403070 JQ403071 JQ403074
	JQ403075 JQ403078 JQ403080 JQ403083 JQ403084 JQ403086 JQ403087 JQ403088 JQ403089 JQ403091 JQ403092 JQ403093 JQ403095 JQ403096
	JQ403099 JQ403100 JQ403101 JQ403102 JQ403103 JQ403104 JQ403105 JQ403106 JQ403107 JQ647929 JQ647930 JQ647931 JQ689607 JQ689608
	JQ689610 JQ990942 JQ990943 JX140652 JX140654 JX140656 JX140657 JX140658 JX140659 JX171201 JX171203 JX171215 JX171216 JX171217
	JX171218 JX171222 JX171227 JX171229 JX171230 JX171231 JX171232 JX171233 JX171234 JX171237 JX171238 JX171239 JX171242 JX440927
	JX440928 JX440930 JX440933 JX440934 JX440935 JX440937 JX440939 JX440940 JX440941 JX440942 JX440945 JX440946 JX440947 JX440948
	JX440949 JX440950 JX440951 JX440953 JX440955 JX440956 JX440957 JX440958 JX440961 JX440963 JX440964 JX440965 JX440967 JX440968
	JX440969 JX440970 JX446795 JX446817 JX447156 JX447795 JX448101 JX500707 JX500708 JX500709 JX503074 JX503075 JX972238 JX972390
	K02007 KC113004 KC113005 KC312367 KC312434 KC312467 KC312572 KC473824 KC473825 KC473826 KC473827 KC473828 KC473829 KC473830
	KC473831 KC473832 KC473833 KC473834 KC473835 KC473841 KC473846 KC596066 KC596067 KC797171 KC797215 KC899011 KC906900 KC906901
	KC906903 KC906907 KC906908 KC906910 KC906912 KC906913 KC906915 KC906916 KC906917 KC906918 KC906919 KC906920 KC906921 KC906922
	KC906923 KC906924 KC906925 KC906926 KC906927 KC906928 KC906929 KC906930 KC906931 KC906932 KC906933 KC906934 KC906936 KC906937
	KC906939 KC906940 KC906941 KC906942 KC906943 KC906944 KC906945 KC906947 KC906948 KC906949 KC906951 KC906952 KC906953 KC906954
	KC906955 KC906956 KC906957 KC906958 KC906959 KC906960 KC906962 KC906963 KC906964 KC906965 KC906966 KC906967 KC906968 KC906969

Continued on next page

Table A.8 – *Continued from previous page*

Use (n)	Accession Numbers
	KC906970 KC906971 KC906972 KC906973 KC906974 KC906976 KC906977 KC906978 KC906979 KC906981 KC906982 KC906983 KC906984 KC906987
	KC906988 KC906989 KC906990 KC935957 KC935958 KC935959 KF301953 KF301963 KF301973 KF301977 KF301981 KF301987 KF301992 KF301995
	KF302000 KF302004 KF302011 KF302017 KF302023 KF302025 KF302029 KF384802 KF384804 KF384808 KF384810 KF384811 KF384812 KF384813
	KF384814 KF526120 KF526172 KF526228 KF526261 KF526298 KF526318 KF701942 KF701944 KF701945 KF701946 KF701947 KF701948 KF701949
	KF701950 KF701951 KF701952 KF701953 KF701954 KF701955 KF701956 KF701957 KF701958 KF701959 KF701961 KF701963 KF701964 KF701966
	KF701967 KF701968 KF701969 KF701970 KF701971 KF701972 KF701973 KF701974 KF701975 KF701976 KF701977 KF701978 KF701979 KF701980
	KF701981 KF701982 KF701983 KF701984 KF701985 KF701986 KF701987 KF701988 KF701990 KF701991 KF701992 KF701993 KF701994 KF701995
	KF701996 KF701998 KF701999 KF702001 KF702003 KF702004 KF702005 KF702007 KF702008 KF702009 KF702010 KF702011 KF702012 KF702013
	KF702014 KF702016 KF702017 KF702020 KF702022 KF702023 KF702024 KF702025 KF702026 KF702027 KF702028 KF702029 KF702030 KF702031
	KF702032 KF702033 KF702034 KF702035 KF702036 KF702037 KF702038 KF702039 KF702040 KF702041 KF702043 KF702044 KF702045 KF702046
	KF702047 KF702048 KF702051 KF702053 KF702054 KF702055 KF702056 KF702057 KF702058 KF702059 KF702060 KF702061 KF702062 KF702063
	KF702064 KF702065 KF702066 KF702067 KF702068 KF702069 KF702070 KF702071 KF702073 KF702075 KF702077 KF702078 KF702079 KF702080
	KF702081 KF702082 KF702083 KF702084 KF702085 KF702086 KF702087 KF702088 KF702090 KF702091 KF702093 KF702094 KF702095 KF702097
	KF702098 KF702099 KF702100 KF702101 KF702102 KF702103 KF702104 KF702105 KF702107 KF702108 KF702110 KF702111 KF702112 KF702113
	KF702114 KF702115 KF702116 KF702117 KF702118 KF702119 KF702120 KF702121 KF702123 KF702124 KF702125 KF702126 KF702127 KF702129
	KF702130 KF702131 KF702133 KF702134 KF702135 KF702136 KF702138 KF702139 KF702140 KF702143 KF702144 KF702145 KF702146 KF702147
	KF702149 KF702150 KF702151 KF702152 KF702153 KF702154 KF702155 KF702156 KF702157 KF702159 KF702160 KF702161 KF702162 KF702163
	KF702164 KF702165 KF702166 KF702167 KF702168 KF702169 KF702170 KF702171 KF702172 KF702173 KF702174 KF702175 KF702176 KF702177
	KF702178 KF702181 KF702182 KF702183 KF702184 KF702185 KF702186 KF702187 KF702188 KF702189 KF702192 KF702193 KF702194 KF702195
	KF702196 KF702197 KF702198 KF702200 KF702201 KF702202 KF702203 KF702204 KF702205 KF702207 KF702209 KF702210 KF702211 KF702212
	KF702213 KF702214 KF702215 KF702216 KF702217 KF702219 KF702220 KF702221 KF702222 KF702223 KF702225 KF702226 KF702228 KF702229
	KF702230 KF702231 KF702232 KF702233 KF702234 KF702235 KF702236 KF702237 KF702238 KF702239 KF702240 KF702241 KF702243 KF702244
	KF702245 KF702246 KF702247 KF702248 KF702249 KF702250 KF702251 KF702252 KF702253 KF702254 KF702255 KF702256 KF702257 KF702258
	KF702259 KF702261 KF702262 KF702263 KF702264 KF702265 KF702266 KF702267 KF702268 KF702269 KF702270 KF702271 KF702272 KF702273
	KF702274 KF702275 KF702276 KF716494 KF716495 KF716496 KF716497 KF716498 KF990605 KF990608 KJ019215 KJ140251 KJ669921 KJ669925
	KJ849766 KJ849767 KJ849780 KJ849784 KJ849785 KJ849788 KJ849790 KJ849799 KJ849801 KJ849803 KJ849804 KJ849805 KJ849808 KJ849811
	KJ849812 KJ849814 KJ849815 KJ849817 KJ849818 KJ849825 KJ869736 KJ869737 KJ869738 KJ869739 KJ869740 KJ869741 KJ869742 KJ869743
	KJ869744 KJ869745 KJ869746 KJ869747 KJ869748 KJ869749 KJ869750 KJ869751 KJ869752 KJ869753 KJ869754 KJ869755 KJ869756 KJ869757
	KJ869759 KJ869760 KJ869761 KJ869762 KJ869763 KJ869767 KJ869768 KJ869769 KJ869770 KJ869772 KJ869773 KJ869774 KJ869775 KJ869776
	KJ869778 KJ869779 KJ869780 KJ869781 KJ869782 KJ869783 KJ869784 KJ869785 KJ869786 KJ869788 KJ869789 KJ869790 KJ869791 KJ869792
	KJ869793 KJ869794 KJ869795 KJ869796 KJ869797 KJ869798 KJ869799 KJ869801 KJ869802 KJ869803 KJ869804 KJ869805 KJ869806 KJ869807
	KJ869808 KJ869810 KJ869811 KJ869813 KJ869814 KJ869815 KJ869816 KJ869818 KJ869820 KJ869821 KJ869824 KJ869825 KJ869826 KJ869827
	KJ869828 KJ869829 KJ869830 KJ869831 KJ869853 KJ869856 KJ869858 KJ869859 KJ869860 KJ869861 KJ869862 KJ869863 KJ869864 KJ869865
	KJ869867 KJ869868 KJ869869 KJ869870 KJ869871 KJ869874 KJ869876 KJ869878 KJ869879 KJ869880 KJ869881 KJ869882 KJ869883 KJ869884
	KJ869885 KJ869888 KJ869890 KJ869891 KJ869892 KJ869893 KJ869894 KJ869896 KJ869897 KJ869898 KJ996063 KM025076 KM217584 KM217585

Continued on next page

Table A.8 – *Continued from previous page*

Use (n)	Accession Numbers
	KM217586 KM217587 KM217588 KM217589 KM217590 KM217591 KM217592 KM217593 KM217594 KM217595 KM217596 KM217597 KM217598 KM217599 KM217600 KM217601 KM217602 KM217603 KM217604 KM217605 KM217606 KM217607 KM217608 KM217609 KM217610 KM217611 KM217612 KM217613 KM217614 KM217662 KM217663 KM217664 KM217665 KM217666 KM217667 KM217668 KM217669 KM217670 KM217938 KM217939 KM217940 KM217941 KM217942 KM217943 KM217944 KM217945 KM217946 KM217947 KM217948 KM217949 KM217950 KM217951 KM217952 KM217953 KM217954 KM217955 KM217956 KM217957 KM217958 KM217959 KM217960 KM217961 KM217962 KM217963 KM217964 KM217965 KM217966 KM217967 KM217995 KM218138 KM218139 KM218140 KM218141 KM218142 KM218143 KM218144 KM218145 KM218146 KM218147 KM218148 KM218149 KM218150 KM218151 KM218152 KM218153 KM218154 KM218155 KM871790 KM871791 KM871792 KM871793 KM871794 KP109510 KP109511 KP109512 KP109514 KP109515 KP109518 KP411822 KP411823 KP411824 KP411825 KP411827 KP411828 KP411829 L02317 L07425 L15476 L15482 L15489 L15500 L15508 L15510 L15515 L15518 LC018135 LC018136 LC018138 LC018139 LC018140 LC018142 LC018143 LC018145 LC018146 LC018147 LC018148 LC018149 LC018150 LC018151 LC018152 LC018153 LC018154 LC018155 LC018156 LC018157 LC018158 LC018159 LC018160 LC018161 LC018162 LC018163 LC018164 LC018166 LC018167 LC018168 LC018169 LC018170 LC018171 LC018172 LC018173 LC018175 LC018176 LC018177 LC043169 LC043170 LC043172 LC043174 LC043176 LC043177 LC043179 LC043180 LC043182 LC043183 LC043185 LC043186 LC043187 LC043188 LC043189 LC043190 LC043192 LC043193 LC043198 LC043199 LC043200 LC043201 LC043202 LC043203 LC043207 LC043208 LC043209 LC043210 LC043211 M17449 M17450 M17451 M21098 M26727 M58173 M65024 M74426 M84627 M84628 M84629 M84630 M84631 M84632 M93258 U03295 U03299 U03338 U03343 U12055 U16863 U16869 U16875 U16893 U16898 U16909 U16915 U16921 U16934 U16940 U23487 U24455 U26062 U26069 U26087 U26110 U26119 U26138 U34603 U39362 U43096 U43106 U44444 U44450 U44462 U44465 U44468 U48917 U52465 U61778 U61782 U61785 U66543 U69584 U71182 U73339 U73341 U73370 U73372 U80224 U80225 X57462 X63040 X63041 X63042 X63043 X63044 X63045 X64745 X64746 X64747 X64748 X64750 X64752 X64754 Y15122 Y17886 Y17887 Y17888 Z98019 Z98022 Z98023 Z98024 Z98025 Z98027 Z98029 Z98030 Z98031 Z98032 Z98034
Generation of HIV-2 *nef* consensus sequence (317)	A03789 A03790 A04707 A09995 A14064 A14065 AB731742 AB731743 AF082339 AF170030 AF170031 AF170033 AF170035 AF170042 AF170044 AF170048 AF170049 AF170052 AF170057 AF170059 AF170060 AF170061 AF208027 AJ238995 AJ238996 AJ238997 AJ238998 AJ238999 AJ239000 AJ344369 AJ344370 AJ344371 AJ344372 AJ344373 AJ344374 AJ344375 AJ344376 AJ344377 AJ344378 AJ344379 AJ344380 AJ344381 AJ344382 AJ344383 AJ344384 AJ344385 AJ344386 AJ344387 AJ344388 AJ344389 AJ344390 AJ344391 AJ344392 AJ344393 AJ344394 AJ344395 AJ344396 AJ344397 AJ344398 AJ344399 AJ344400 AJ344401 AJ344402 AJ344403 AJ344404 AJ344405 AJ344406 AJ344407 AJ344408 AJ344409 AJ344410 AJ344411 AJ344414 AJ344415 AY509259 AY530889 AY622980 AY622981 AY622982 AY622983 AY622984 BD131284 BD413542 BD413545 CQ828033 CS230394 CS230398 D00835 D00835 DQ092764 DQ222472 DQ473391 DQ640016 DQ640017 DQ640018 E02138 EU650797 EU980602 FB293476 FB293480 FJ548891 GQ890616 GQ890617 GQ890618 GQ890619 GQ890620 GQ890621 GQ890622 GQ890623 GQ890624 GQ890625 GQ890626 GQ890627 GQ890628 GQ890629 GQ890630 GQ890631 GQ890632 GQ890633 GQ890634 GQ890635 GQ890636 GQ890638 GQ890639 GQ890640 GQ890641 GQ890642 GQ890643 GU983917 GU983918 GU983919 GU983920 GU983921 GU983922 GU983923 GU983924 GU983927 GU983928 GU983929 GU983930 GU983931 GU983932 GU983933 GU983934 GU983935 GU983936 GU983937 GU983938 GU983939 GU983940 GU983941 GU983942 GU983943 GU983944 GU983945 GU983946 GU983947 GU983948 GU983949 GU983950 GU983951 GU983952 GU983953 GU983954 GU983955 GU983956 J04498 J04542 JF683340 JN863892 JN863893 JN863894 JN863895 JN863896 JN863897 JN863898 JQ746624 JQ746625 JQ746627 JQ746628 JQ746629 JQ746630 JQ746633 JQ746634 JQ746635 JQ746636 JQ746637 JQ746638 JQ746639 JQ746642 JQ746643 JQ746644 JQ990946 JQ990948 JQ990949 JQ990950

Continued on next page

Table A.8 – *Continued from previous page*

Use (n)	Accession Numbers
	JX219591 JX219592 JX219595 JX219596 JX219597 JX219598 JX219599 JX219600 JX219601 JX219602 JX219603 JX219604 JX219605 JX219606 JX219607 JX219608 JX219609 JX219610 JX219611 JX219612 JX219613 JX219614 JX235895 JX235900 JX235901 JX235902 JX235903 JX235904 JX235905 JX235906 JX235907 JX235908 JX235909 JX235910 JX235911 JX235912 JX235913 JX235914 JX235915 JX235916 JX235917 JX235918 JX235919 JX235920 JX235921 JX235922 JX235923 JX235924 JX235925 KC693505 KC748535 KC748536 KC748537 KC748538 KC748539 KC748540 KC748541 KC748542 KC748543 KC748544 KC748547 KC748548 KC748549 KC748550 KC748551 KC748555 KC748556 KC748557 KC748558 KC748559 KC748560 KC748561 KC748562 KC748563 KC748564 KC748565 KC748566 KC748569 KC748570 KC748571 KC748572 KC748573 KC748574 KC748575 KC748576 KC748577 KC748578 KC748579 KC748580 KF156811 KF156813 KF156814 KF156815 KF156817 KF156818 L10637 L25445 L36874 M15390 M30502 M30895 M31113 M87117 U05350 U05351 U05352 U05353 U05354 U05355 U05356 U05357 U05358 U05359 U22047 U38293 U73757 U73758 U76642 X05291 X52223 Z48731
Laboratory-adapted SF2 *nef*	K02007
Alignment of the HIV-2 Nef C-terminal α-helix (133)	A09995 AB731742 AB731743 AF082339 AF208027 AJ344369 AJ344370 AJ344371 AJ344372 AJ344373 AJ344374 AJ344375 AJ344376 AJ344377 AJ344378 AJ344379 AJ344380 AJ344381 AJ344382 AJ344383 AJ344384 AJ344385 AJ344386 AJ344387 AJ344388 AJ344389 AJ344390 AJ344391 AJ344392 AJ344393 AJ344394 AJ344395 AJ344396 AJ344397 AJ344398 AJ344399 AJ344400 AJ344401 AJ344402 AJ344403 AJ344404 AJ344405 AJ344406 AJ344407 AJ344408 AJ344409 AJ344410 AJ344411 AJ344414 AJ344415 AY509259 AY530889 AY622980 AY622981 AY622982 AY622983 AY622984 BD131284 BD413542 D00835 D00835 DQ092764 DQ222472 DQ640016 DQ640017 E02138 EU650797 EU980602 FJ548891 GQ890616 GQ890617 GQ890618 GQ890619 GQ890620 GQ890621 GQ890622 GQ890623 GQ890624 GQ890625 GQ890626 GQ890627 GQ890628 GQ890629 GQ890630 GQ890631 GQ890632 GQ890633 GQ890634 GQ890635 GQ890636 GQ890638 GQ890639 GQ890640 GQ890641 GQ890642 GQ890643 J04498 J04542 JQ746624 JQ746625 JQ746627 JQ746628 JQ746629 JQ746630 JQ746633 JQ746634 JQ746635 JQ746636 JQ746637 JQ746638 JQ746639 JQ746642 JQ746643 JQ746644 JQ990946 JQ990948 JQ990949 JQ990950 KC693505 L36874 M15390 M30502 M30895 M31113 NC001722 U22047 U38293 U73757 U73758 U76642 X05291 X52223 Z48731