DRUG RESISTANCE IN HIV-1

BY

DAVID FRANK BIBBY

Department of Virology,
Royal Free and University College Medical School,
W1T 4JF

A thesis submitted to The University of London
in fulfilment for the degree of Doctor of Philosophy

2002
Abstract

Treatment of HIV-1 infection with 3’-azido-2’,3’-dideoxythymidine (AZT) frequently results in the evolution of viruses harbouring varying combinations of point mutations within the reverse transcriptase (RT) domain which confer increasing levels of resistance to AZT. It has been demonstrated that two such mutations (M41L and T215Y) reduce the replicative capacity of HIV-1 and are selected against in the absence of AZT. For this thesis, a patient undergoing AZT monotherapy was identified (A3), in which M41L and T215Y emerged only after therapy had been discontinued. It was proposed that rather than universally decreasing fitness, these mutations could increase fitness in the context of different RT genotypes.

A system was developed to generate recombinant HIV-1 stocks containing populations of sample-derived RT domains in a constant genetic background. The widely-used recombinant virus assay in which sample-derived sequences are inserted into a virus backbone by homologous recombination was adapted in order to prevent undefined and/or undesired sample-derived sequences being incorporated into recombinant viruses. Phenotypic differences between recombinant viruses could thus be ascribed solely to their differing RT domains. One of two silent point mutation markers was engineered into each recombinant, allowing the relative frequencies of mixed virus populations to be assayed by a highly-reproducible quantitative point mutation assay. Measuring the rate of change of these frequencies during pair-wise competition experiments enabled calculation of relative fitnesses.

Recombinant viruses were generated containing RT populations rescued from serial plasma samples taken from patient A3, and their relative fitness calculated from competitions against a reference virus. Significant differences between relative fitness values were observed. The onset of therapy resulted in lowered fitness, but over time, the RT population fitness returned to the pre-treatment level, concomitant with the emergence of M41L and T215Y, together with many other sequence changes observed throughout the duration of therapy.
Acknowledgements

Many thanks are owed to many people, without whom this thesis would never have happened. Firstly, thanks to my supervisors – Professor Richard Tedder, Dr Steve Kaye, and Dr Peter Balfe – for their help, guidance and support throughout my PhD. Secondly, my family deserve more than a little gratitude for their persistent faith, and for their unquestioning moral and financial assistance at the numerous critical points when their interventions single-handedly rescued me from succumbing to defeat. Also, my colleagues in the Department of Virology in which I carried out this work provided invaluable advice and support, including (but not limited to) Paul Grant, Phillip Tuke, Bridget Ferns, Julie Bennett, Ursula Ayliffe, Samreen Ijaz, Chris Perrons, Anna Hayman, Sara Bissett, Jim Waite, Stuart Kirk, Kevin Whitby, Sonia Turner, and Jeremy Garson. A special hand of gratitude is extended to the staff at the virology department in Utrecht, particularly Wilco Keulen, Monique Nijhuis, and Charles Boucher, for being so welcoming and accommodating during my six week visit to learn about the recombinant virus assay and fitness. I would also like to acknowledge the forbearance of the countless individuals who I have lived with during my PhD – with hindsight, their tolerance of my fluctuating stress levels is much appreciated. Finally, I would like to thank Jayne Woodcock for her unstinting love and support, without which it would never have seemed quite so worthwhile.

Thank you.
# Contents

Abstract ...................................................................................................................... 2
Acknowledgements .................................................................................................. 3
Contents .................................................................................................................... 4
Figures ..................................................................................................................... 7
Tables ........................................................................................................................ 13
Abbreviations .......................................................................................................... 15

## 1. Introduction ........................................................................................................ 17
   1.1. *Retroviridae* .................................................................................................. 19
      1.1.1. The Lentivirus Genus ............................................................................. 22
   1.2. HIV-1 Infection ............................................................................................. 24
      1.2.1. Transmission .......................................................................................... 25
      1.2.2. Aetiology ............................................................................................... 26
   1.3. The Life Cycle ................................................................................................ 30
      1.3.1. Transcription and Translation ................................................................ 35
      1.3.2. Virion Assembly and Release ................................................................. 36
      1.3.3. Fusion and Entry .................................................................................... 38
      1.3.4. Reverse Transcription .......................................................................... 39
      1.3.5. Integration and Latency ........................................................................ 45
   1.4. Antiretroviral Therapy ................................................................................... 46
      1.4.1. Nucleoside Analogues ........................................................................... 48
      1.4.2. Acyclic Nucleotide Analogues ................................................................ 48
      1.4.3. Non-Nucleoside Reverse Transcriptase Inhibitors ............................... 49
      1.4.4. Protease Inhibitors ................................................................................ 50
      1.4.5. Other Antiretroviral Strategies ............................................................... 50
   1.5. Resistance to Antiretroviral Compounds ......................................................... 52
   1.6. Evolution of HIV-1 ......................................................................................... 59
      1.6.1. RNA Virus Population Dynamics ........................................................... 59
      1.6.2. Sources of Genetic Variation in HIV-1 ................................................... 62
      1.6.3. The HIV-1 Quasispecies ....................................................................... 63
      1.6.4. Effects of Drug Resistance on the Evolution of HIV-1 ........................... 65
   1.7. Outline of the Thesis ....................................................................................... 70

## 2. Design and Application of the pHIVEC Vector System ....................................... 72
   2.1. Introduction .................................................................................................... 72
   2.2. Construction of pHIVEC .............................................................................. 76
   2.3. Methods ......................................................................................................... 83
   2.4. Results .......................................................................................................... 86
   2.5. Discussion ...................................................................................................... 91
10. Bibliography __________________________________________214

Appendix A: Materials and General Methods ____________________259

A.1 Plasma and Serum Samples ___________________________________259
A.2 Primer and Oligonucleotide Sequences _________________________260
A.3 *Escherichia coli* __________________________________________263
A.4 Mammalian Cell Lines and Growth Media ________________________264
A.5 Polymerase Chain Reaction ___________________________________265
  A.5.1 Ampli*Taq*-Mediated PCR _________________________________265
  A.5.2 Expand-Mediated PCR _____________________________________266
A.6 Reverse Transcription ________________________________________266
  A.6.1 The SuperScript II System _________________________________267
  A.6.2 The Titan RT-PCR System _________________________________267
A.7 Restriction Endonuclease Digestion _____________________________267
A.8 Nucleic Acid Purification _____________________________________268
  A.8.1 Purification of DNA from Agarose Gels, Restriction Digests and PCRs 268
  A.8.2 Preparation of Plasmid DNA from Bacterial Cultures _________268
  A.8.3 DNA Purification by Phenol-Chloroform Extraction _________269
  A.8.4 Purification of RNA from Plasma and Serum – the Boom Protocol 270
  A.8.5 Extraction of RNA from Culture Supernatants _______________271
A.9 Preparation of Competent Bacteria from Glycerol Stocks _________271
A.10 Transformation into *E. coli* of Plasmid DNA __________________272
A.11 Ligation __________________________________________________272
A.12 Hoescht Dye-Based Fluorometric DNA Quantification __________273
A.13 Kodak Gel-Based DNA Quantification _________________________273
A.14 DNA Dideoxy-Sequencing ____________________________________274
A.16 p24 Antigen ELISA __________________________________________276
A.17 Titration of Virus Stocks _____________________________________277
TABLE OF FIGURES

Figures

FIGURE 1-1: The general organisation of retrovirus genomes. Protein coding regions are shaded orange. The yellow U5 and U3 sections contain sequences involved in the regulation of transcription. The PBS (primer binding site), the polypurine tract (PPT) and the R (repeat) regions comprise short sequences involved in reverse transcription.

FIGURE 1-2: Phylogenetic tree (drawn to scale) based on Pol gene sequences showing the evolutionary relationships within the Primate Lentivirus Group, adapted from Hahn et al., 2000.

FIGURE 1-3: Schematic of the HIV-1 genome. The upper, shaded figure represents the organisation of the integrated, double-stranded DNA genome, below which is an arrow showing the sequences encompassed by the smaller, single-stranded RNA genome found in virions. The colour coding is identical to that of FIGURE 1-1 (page 21), with the additional components of the regulatory proteins (green) and accessory proteins (blue).


FIGURE 1-5: Schematic diagram of the process of reverse transcription (after Gotte et al., 1999).

FIGURE 1-6: A simplified two-axis representation of Fisher’s geometric model of evolution.

FIGURE 2-1: The principle of homologous recombination. The shaded blocks represent sequences where the derivation (either from the vector or from the sample virus) is undefinable.

FIGURE 2-2: Schematic outline of the pHIVEC vector system for the generation of recombinant HIV-1.

FIGURE 2-3: A schematic representation of plasmid pHXB2ART (~12.5kb), from which pHIVEC was derived.

FIGURE 2-4: The nucleotide sequence of the 3' terminus of HXB2 reverse transcriptase, showing the location of a potential silent SacII discovered using Vector NTI software.

FIGURE 2-5: The two-site method for ligation of PCR amplified RT inserts and the pHIVEC vector at the protease-RT junction.

FIGURE 2-6: Schematic diagram showing how the oligonucleotides in TABLE 2-1 align with the Pol domain of pHXB2a RT. The restriction sites introduced by these primers are also shown.

FIGURE 2-7: Digests of pHXB2ART. Lane 1 - pHXB2ART (uncut). Lane 2 - pHXB2ART digested with MscI. Lane 3 - pHXB2ART digested with Xmal. Lane 4 - pHXB2ART digested with both MscI and Xmal. Lane 5 - PCR product from pHXB2ART using primers Sac-Site and Mut-aSense. Lane 6 - PCR product from Lane 5 digested with both MscI and Xmal.

FIGURE 2-8: PCR amplification using primers 706M and Mut-aSense. Lanes 1-3 respectively - pMCS79, pHXB2ART, and a colony obtained from ligation between the bands purified from FIGURE 2-7.

FIGURE 2-9: Lanes 1 and 2 - SacII digestion of the amplicons shown in Lanes 2 and 3 of FIGURE 2-8 respectively.

FIGURE 2-10: BstBI digestion of amplicons from 17 colonies obtained by transforming ligations between SacII-Smal digested pHXB2ART SacII and the products of the annealing of oligonucleotides AD-1 and
AD-2. The expected amplicon of 1,050bp should cleave with BstBI to generate fragments of 907bp and 143bp. The arrows indicate the four products yielding correct amplification and digestion products. Lane 18 – amplicon from unmodified pHXB2ART SacII subjected to BstBI digestion.

FIGURE 2-11: Alignment of the 'Inner' PCR primer set on HIV-1 Pol domains (upper set – protease-RT junction, lower set – 3’ end of RT). Restriction sites introduced by the primers are highlighted in bold type.

FIGURE 2-12: Double restriction digests of pHIVEC. Lane 1 – digestion with both BstBI and BsaBI. Lane 2 – digestion with both SacII and Spel.

FIGURE 2-13: Results of gradient PCRs using primers Inner-1 and either Inner-2a (top picture) or Inner-2b (bottom picture). The temperature of the annealing steps ranged from 42°C to 56°C in 2°C increments.

FIGURE 2-14: Sample second-round PCRs of sample-derived first-round RT-PCR products using primers Inner-1 and Inner-2a (upper picture), and Inner-1 and Inner-2b (lower picture).

FIGURE 2-15: Lanes 1-12 – Sample PCR screen of ligations between pHIVEC and sample-derived RT inserts. Lane 13 – self-ligation (no insert). Lane 14 – negative control (pHIVEC). Lane 15 – positive PCR control (pHXB2). Lane 16 – negative PCR control (water). The band visible in lane 14 corresponds to plasmid template (pHIVEC) carried through the PCR.

FIGURE 2-16: Lanes 1-5 – Sample PCR screen of plasmids obtained from E. coli transformed with pHIVEC-RT insert ligations. Lane 6 – negative PCR control (water). Lane 7 – positive PCR control (pHXB2).

FIGURE 3-1: Schematic outline of the first shuttle vector system employed in the generation of recombinant HIV-1.

FIGURE 3-2: Schematic of pGEM-T Easy Vector (from Promega Corporation, Madison, WI).

FIGURE 3-3: Schematic of pGEM-TARD (3,965bp). Domains from pGEM-T are coloured blue, and the Pol domains from pHIVEC are coloured red.

FIGURE 3-4: Alignment of the 'IN' primer set on HIV-1 Pol domains (upper set – protease-RT junction, lower set – 3’ end of RT). Restriction sites introduced by the primers are highlighted in bold type.

FIGURE 3-5: Sample gel showing the high yield variability amongst second-round PCR amplifications of sample-derived RT domains using the primer set IN-1 and IN-2a, compared to the efficiency of amplification from plasmid-derived template. Lane 1 – pHXB2 as template. Lanes 2-8 – sample-derived templates.

FIGURE 3-6: Sample gel showing how some sample-derived material failed to amplify using primers IN-1 and IN-2b. Lanes 1-5 – amplifications of sample-derived material. Lane 6 – Negative PCR control (water). Lane 7 – Positive PCR control (pHXB2).

FIGURE 3-7: PCR screening of ligations between pGEM-TARD and RT inserts. Lanes 1-12 – using T7 and 469W as primers. Lanes 13-24 – using 111W and dSp6 as primers. In each row, the order of samples is as follows: eight ligations, self-ligation, negative PCR control (water) and positive PCR control (pHXB2).

FIGURE 3-8: Sample gel photo showing the plasmid DNA species extracted from bacteria transformed with ligations between pGEM-TARD and RT inserts.
Figure 3-9: PCR amplification of plasmid DNA migrating at a rate consistent with 5.6kb supercoils (as shown in Figure 3-8), using primers 706M and 708M. Lanes 1-4 - purified plasmid DNA as template. Lane 5 - Positive PCR control (pHXB2). Lane 6 - Negative PCR control (water).

Figure 3-10: PCR amplification of purified plasmid DNA migrating at a rate consistent with 5.6kb supercoils using primers S-HR and AS-HR. Lanes 1-5 - purified plasmid DNA as template. Lane 5 - Positive PCR control (pHXB2). Lane 6 - Negative PCR control (water).

Figure 3-11: Restriction endonuclease analysis of a plasmid obtained after transformation of a ligation between pGEM-TART and a plasmid template-derived RT insert. Lanes 1-3 - uncut supercoiled plasmid. Lane 4 - plasmid digested with Apal and SacII. Lane 5 - plasmid digested with Apal and EcoRV. Lane 6 - plasmid digested with Apal and EcoRI. Note the multiple digestion in products the final lane.

Figure 3-12: Plasmid DNA extracts from bacteria transformed with pGEM-TART-RT insert ligation reactions and separated by overnight electrophoresis. RT inserts were obtained from plasmid templates (e.g., pHXB2). The banding patterns are discussed in the text.

Figure 3-13: Electrophoresis profiles of plasmid extracts obtained from bacteria transformed with pGEM-TART-RT insert ligation reactions where the RT inserts were obtained from sample material, showing the presence of diverse DNA species smaller in size than pGEM-TART.

Figure 3-14: Plasmid DNA extracts from bacteria 're-transformed' with gel-purified DNA from putative insert-positive pGEM-TART constructs seen in Figure 3-13.

Figure 3-15: Lanes 1-5: Amplifications from purified insert-positive pGEM-TART constructs using primers S-HR and AS-HR. Lane 6 - Negative PCR control (water), Lane 7 - Positive PCR control (pHXB2).

Figure 3-16: ELISA data (p24 antigen concentration) from the successful transfection of amplified Pol sequences from pGEM-TART and the transfection of Pol sequence amplified directly from pHXB2. The cut-off for p24 detection was 100 pg/ml.

Figure 3-17: Lane 1 - RT-PCR product using recombinant virus pGEM-TART RNA as template and primers 111W and G5804. Lane 2 - SacII digestion of RT-PCR product in Lane 1. Lane 3 - identical amplification of recombinant virus generated using the positive PCR control product (HXB2).

Figure 4-1: HXB2 sequences at the 5' and 3' ends of RT showing potential silent restriction sites for use in a new shuttle vector. Mismatches are highlighted in bold type.

Figure 4-2: Comparison of the nucleotide sequences and restriction maps of the RT-deletion regions of pGEM-TART and pRTshut.

Figure 4-3: Schematic diagram of the RT-deletion region of pGEM-TART, showing the alignments of the mutagenic primers INT-X and MUN-1.

Figure 4-4: Restriction endonuclease digests of pGEM-TART and two mutagenic PCRs (see text). Lane 1 - pGEM-TART uncut (supercoiled plasmid). Lanes 2-4 - digests I-III detailed in Table 4-2. Lane 5 - T7-MUN-1-primed PCR digested with Apal and SacII. Lane 6 - INT-X-αSp6-primed PCR digested with SacII and EcoRI. The boxed bands were excised and purified.

Figure 4-5: Double restriction endonuclease digests on 'intermediate' plasmid constructs. Lane 1: pGEM-TART-XbaI digested with SacII and EcoRI, Lanes 2-5: pGEM-TART-MfiI A-D respectively, digested with Apal and SacII. Boxed bands were excised and purified.

Figure 4-6: Double restriction endonuclease digests on 'intermediate' plasmid constructs. Uppermost bands from each lane were purified. Lane 1: pGEM-TART-XbaI digested with SacII and Apal, Lanes 2-5:
pGEM-TART-Mfel A through D, digested with EcoRI and SacII. Boxed bands were excised and purified

FIGURE 4-7: Alignment of the 'Mut' primer set on HIV-1 Pol domains (upper set – protease-RT junction, lower set – 3' end of RT). Restriction sites introduced by the primers are highlighted in bold type. The alternate bases at the third base of the final RT codon in the lower alignment were introduced for the purposes of future qPMA analysis of mixed virus populations.

FIGURE 4-8: Gradient PCR using first round product amplified from a sample with low copy number RNA. Top row: Mut-5-Mut-3T, Bottom row: Mut-5-Mut-3C.

FIGURE 4-9: Schematic of the 1,717bp Mut-5-Mut-3T PCR product, showing the location of recognition sequences for restriction endonucleases XbaI, Mfel, BsrGI, and NlaIV.

FIGURE 4-10: Investigation into the activity of restriction endonucleases XbaI and Mfel on Mut-5 and Mut-3T/C-primed PCR products. Lanes 1-6 – reactions A-F respectively from TABLE 4-4.

FIGURE 4-11: Mfel-XbaI-digested Mut-5-Mut-3C PCR products from six sample-derived first round products (lanes 1-6), two plasmid templates (lanes 7 and 8), and pRTshut digested with Mfel and XbaI (lane 9).

FIGURE 4-12: Lane 1 – Insert-positive pGEM-TART plasmid from section 2. Lane 2 – undigested pRTshut. Lanes 3-6 – Plasmid DNA extracted from transformed pRTshut-RT insert ligations. Boxed bands represent the putative insert-positive constructs.

FIGURE 4-13: Lanes 1-3 – Three DNA extracts from bacteria 're-transformed' with bands corresponding to those boxed in FIGURE 4-12. Lanes 4-6. The same plasmids as in lanes 1-3 after digestion with XbaI and Mfel.

FIGURE 4-14: PCR screens of seven ligations using sample-derived RT domains. Top row: S-HR-469W-primed PCRs, Middle row: S-HR-AS-HR-primed PCRs, Bottom row: 111W-AS-HR-primed PCRs. In each row, lanes 1-7 – pRTshut-RT insert ligations, lane 8 – self-ligation, lane 9 – negative PCR control (water), lane 10 – positive PCR control (pHXB2).

FIGURE 4-15: Lanes 1-7 – electrophoretic profiles of seven DNA species extracted from bacterial cultures transformed with pRTshut ligated to sample-derived PCR-amplified RT domains. Lane 8 – DNA extracted from bacteria transformed with pRTshut self-ligation. The arrow indicates the row of bands that were excised.

FIGURE 4-16: Lanes 1-5 – The bands excised from FIGURE 4-15 after purification. Lane 6 – insert-positive pGEM-TART from section 03 as a transformation control. Lanes 7-11 – plasmid DNA derived from transformation of the material in lanes 1-5 respectively.

FIGURE 4-17: ‘Final’ PCRs on purified insert-positive bands using primers S-HR and AS-HR. Lanes 1-9 – bands purified as per those shown in FIGURE 4-15 as templates. Lane 10 – negative PCR control (water). Lane 11 – positive PCR control (pHXB2). The templates for the PCRs shown in lanes 3 and 4 were putatively insert-positive plasmids containing RT domains derived from pHXB2-based plasmids (see text). The templates for the other PCRs were plasmids whose RT domains would have been sample-derived.

FIGURE 5-1: Schematic of the restriction-endonuclease-ligation (REL) method of recombinant virus generation.

FIGURE 5-2: The 3'-terminal sequence of HXB2 protease showing potential 'silent' 6-base sequences with central 4-base palindromes.
TABLE OF FIGURES

FIGURE 5-3: Nucleotide manipulations involved in restriction-endonuclease-ligation cross-ligation at the protease-RT junction. 137

FIGURE 5-4: The 5’-terminal sequence of HXB2 integrase showing potential ‘silent’ 6-base sequences with central 4-base palindromes. The nucleotide sequence shown begins with the first base of integrase. 139

FIGURE 5-5: Nucleotide manipulations involved in restriction-endonuclease-ligation cross-ligation at the RT-integrase junction. 140

FIGURE 5-6: Alignment on HIV-1 Pol domains (upper set – protease-RT junction, lower set – 3’ end of RT) of the primer set used in this section. Restriction sites introduced by the primers are highlighted in bold type. 142

FIGURE 5-7: Repeated ‘gradient PCR’ using primers S-MT5 and AS-MT3B with anneal temperatures ranging from 52°C to 66°C in 2°C intervals. 146

FIGURE 5-8: Agarose gel allowing relative yield analysis of five sample-derived RT domains (lanes 1-5), two plasmid-derived RT domains (lanes 6 & 7), and the two flanker sequences (lanes 8 and 9). 146

FIGURE 5-9: Enzyme optimisation by lambda digestion (I). Lanes 1, 5, 8, and 15 – molecular weight markers. Lanes 2-4 – BspEI in NEB #3, NEB #4, and SUB respectively. Lanes 6-7 – BspEI and NgoMIV in NEB #4 and SUB respectively. Lanes 9-14 – Paired ApaLI and ApaLI & NgoMIV digests in NEB #3, NEB #4, and SUB respectively. 147

FIGURE 5-10: Enzyme optimisation by lambda digestion (I). Lanes 1, 5, 9, 12, and 16 – molecular weight markers. Lanes 2-4 – EcoRI in buffers U, NEB #4, and SUB respectively. Lanes 6-8 – Mfel in buffers NEB #3, NEB #4, and SUB respectively. Lanes 10-11 – EcoRI & Mfel double digests in NEB #4 and SUB respectively. Lanes 13-15 – BspEI, NgoMIV, EcoRI, & Mfel multiple digests in buffers NEB #3, NEB #4, and SUB respectively. 148

FIGURE 5-11: Experiment investigating the effect of flanker:insert ratio in REL reactions on the frequency of undesired ‘final’ PCR products. Lanes 1-4 – flanker:insert ratios of 1:3, 1:10, 1:30, and 1:100. Lane 5 – negative PCR control (water). Lane 6 – positive PCR control (pHXB2). 150

FIGURE 5-12: An example of Expand-mediated S-HR/206V-primed PCRs on REL reaction products. Lanes 1-4 – REL reaction products. Lane 5 – negative water PCR control. Lane 6 – positive PCR control using pHXB2 as template. 150

FIGURE 6-1: Schematic of the principle of the quantitative point mutation assay (qPMA). 157

FIGURE 6-2: Location within Pol of the primers detailed in TABLE 6-1. 158

FIGURE 6-3: Graphical representation of the data from TABLE 6-3, showing the calculated output values for each of the three series of standard curve data. 164

FIGURE 6-4: Graphical representation of the data shown in TABLE 6-3, showing the mean output frequencies at each input percentage, with standard error bars. A unit gradient line (dashed) is also presented. 164

FIGURE 6-5: Graph of the data in TABLE 6-4, with a line of equivalence (dashed). 165

FIGURE 6-6: Gel photo of the two controls from the experiment producing the data in TABLE 6-4. Lane 1: 100% ‘A’, Lane 2: 100% ‘G’. 165

FIGURE 8-1: Data from subject ‘A3’, showing RNA viral load pattern over time, together with PMA analysis of three AZT resistance-associated mutations (codons 41, 70 and 215) and duration of AZT therapy. Codons 67 and 219 were consistently wild-type (Kaye, 1997). 177
TABLE OF FIGURES

FIGURE 8-2: Graphical representation of the data given in TABLE 8-1, with a linear trend line. 180

FIGURE 8-3: Graphical representation of the data presented in TABLE 8-2 and 8-3 with linear trend lines for each of the three series. 183

FIGURE 8-4: Graphical representation of the data presented in TABLE 8-4, with the day at which each sample was taken against the mean rate of change of virus frequency in competition with HXB2. 185

FIGURE 8-5: Scatter plot of HIV-1 RNA viral load and relative RT fitness, from the A3 patient samples – r = 0.886, p = 0.023. 186

FIGURE 8-6: The relationship between relative fitness of RT populations in patient A3 and RNA load over the 192 day study period, with sequence changes from the baseline sample highlighted (changes to codons associated with AZT resistance are emboldened). 190

FIGURE A-1: Schematic of the 3' terminus of HXB2 Gag and the entire HXB2 Pol domain, showing the location of the binding sites for the primers listed in the upper part of TABLE A-1. 260

FIGURE A-2: Location of the binding sites within HXB2 Pol of the sequencing primers detailed in TABLE A-2. 262
Tables

TABLE 1-1: Retrovirus taxonomy. 21
TABLE 1-2: Lentivirus taxonomy. 21
TABLE 1-3: Clinical conditions indicative of Category B and Category C HIV-1 infections. 29
TABLE 1-4: Overview of the HIV-1 proteins. 32
TABLE 1-5: Table of nucleoside analogue compounds currently licensed for anti-HIV therapy. 49
TABLE 1-6: Table of mutations in RT associated with lowered susceptibility to NNRTIs. Mutations in bold type are those most frequently found in vivo (Deeks, 2001). 58
TABLE 2-1: Oligonucleotides used specifically for the construction of pHIVEC. 80
TABLE 2-2: Second-round PCR primers unique to this section. Inner-2a and Inner-2b differed at a single nucleotide position (boxed). 85
TABLE 3-1: Specific primers used in this section. 98
TABLE 4-1: Specific oligonucleotides used in the synthesis of pRTshut. 115
TABLE 4-2: The three restriction endonuclease combinations used to digest pGEM-TART, and the predicted band sizes thereof. 117
TABLE 4-3: Primers synthesised to replace the ‘IN’ series. 121
TABLE 4-4: Restriction endonuclease compositions of the six digest reactions. 123
TABLE 5-1: Analysis of the 4-base palindromes in FIGURE 5-2. Enzymes listed in parentheses leave overhangs upon cleavage rather than underhangs. 136
TABLE 5-2: Analysis of the sequences shown in FIGURE 5-4. Enzymes listed in parentheses leave overhangs upon cleavage rather than underhangs. 139
TABLE 5-3: Specific primers used in this section. 142
TABLE 5-4: Activities at 25°C of the enzymes used in the REL, relative to their activities at 37°C. 153
TABLE 6-1: Primer sequences used in the final qPMA protocol. 158
TABLE 6-2: Compositions of the buffers used in the qPMA protocol. 159
TABLE 6-3: Data from three qPMA experiments using the same standard curve samples. 163
TABLE 6-4: TopCount data and analysis from experiments in which five RNA extracts were subjected to duplicated second-round PCRs and qPMAs. 165
TABLE 7-1: qPMA data from a 10-day direct competition between rvHXB2Δ and rvHXB2-41L/215Y showing the raw data, and the method of calculating virus frequency. 170
TABLE 7-2: qPMA data from a 10-day direct competition between HXB2-41L/215YΔ and HXB2Δ showing the raw data, and the method of calculating virus frequency. 171
TABLE 8-1: qPMA data from the reciprocal virus experiment. 179
TABLE 8-2: qPMA data and analysis from parallel secondary competition assays competing rvA3-0^ against rvHXB2\textsuperscript{0}. 182

TABLE 8-3: qPMA data and analysis from a secondary competition assay competing rvA3-0^ against HXB2\textsuperscript{0}. 182

TABLE 8-4: Collected data from competition assays between A3 sample RT domains and HXB2 RT, showing the log\textsubscript{e} of the percentage change in frequency of the A3-derived virus per day over experimental periods of 5-7 days. Mean and standard error of the mean values for each of the five timepoint samples were calculated. 184

TABLE 8-5: p-values from two-tailed two-sample t tests assuming unequal variances between each permutation of paired data sets from those presented in TABLE 8.4. Values below 0.05 are highlighted in bold type. 186

TABLE 8-6: Summary of sequencing data from the A3 samples. Only RT codons varying from the HXB2 reference sequence are presented. Polymorphic amino acids reflect relative peak heights of ~50%. 189

TABLE 8-7: Expanded detail of the collected codon changes within RT observed in samples taken during the first 77 days of the study period. 199

TABLE 8-8: Expanded detail of the collected codon changes within RT observed in samples taken during the final 67 days of the study period. 199

TABLE A-1: General primer sequences 261

TABLE A-2: Details of the primer sequences used for sequencing RT domains 262

TABLE A-3: Buffer compositions for the Boom extraction protocol 271
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>adenosine</td>
</tr>
<tr>
<td>AIDS</td>
<td>Acquired immune deficiency syndrome</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>AZT</td>
<td>3’-azido-2’,3’-dideoxythymidine</td>
</tr>
<tr>
<td>bp</td>
<td>base pair(s)</td>
</tr>
<tr>
<td>C</td>
<td>cytidine</td>
</tr>
<tr>
<td>CDC</td>
<td>Communicable Diseases Center (Atlanta, USA)</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>ddC</td>
<td>2’,3’-dideoxycytidine</td>
</tr>
<tr>
<td>ddI</td>
<td>2’,3’-dideoxyinosine</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxynucleotide triphosphate</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylene diamine tetra-acetic acid</td>
</tr>
<tr>
<td>ELISA</td>
<td>enzyme-linked immunosorbent assay</td>
</tr>
<tr>
<td>FCS</td>
<td>foetal calf serum</td>
</tr>
<tr>
<td>G</td>
<td>guanosine</td>
</tr>
<tr>
<td>HAART</td>
<td>highly active antiretroviral therapy</td>
</tr>
<tr>
<td>HIV-1</td>
<td>human immunodeficiency virus type 1</td>
</tr>
<tr>
<td>HIV-2</td>
<td>human immunodeficiency virus type 2</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase pairs</td>
</tr>
<tr>
<td>LTR</td>
<td>long terminal repeat</td>
</tr>
<tr>
<td>NNRTI</td>
<td>non-nucleoside reverse transcriptase inhibitor</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide(s)</td>
</tr>
<tr>
<td>PBMC</td>
<td>peripheral blood mononuclear cell</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline / primer binding site</td>
</tr>
<tr>
<td>PI</td>
<td>protease inhibitor</td>
</tr>
<tr>
<td>PIC</td>
<td>pre-integration complex</td>
</tr>
</tbody>
</table>
PMA  point mutation assay
PPT  polypurine tract
qPMA quantitative PMA
RNA  ribonucleic acid
RT   reverse transcriptase
RTC  reverse transcription complex
RVA  recombinant virus assay
SIV  simian immunodeficiency virus
T    thymidine
Tris tris(hydroxyethyl) aminomethane
tRNA transfer RNA
3-TC 2',3'-dideoxy-3'-thiacytidine
U    Uridine
ZDV  zidovudine

Amino acid abbreviations

A, Ala  Alanine                  M, Met  Methionine
C, Cys  Cysteine                N, Asn  Asparagine
D, Asp  Aspartic Acid           P, Pro  Proline
E, Glu  Glutamic Acid           Q, Gln  Glutamine
F, Phe  Phenylalanine           R, Arg  Arginine
G, Gly  Glycine                 S, Ser  Serine
H, His  Histidine               T, Thr  Threonine
I, Ile  Isoleucine              V, Val  Valine
K, Lys  Lysine                  W, Trp  Tryptophan
L, Leu  Leucine                 Y, Tyr  Tyrosine
1. Introduction

In 1981, several incidences of opportunistic infections together with severe immune dysfunction were reported amongst the homosexual populations of Los Angeles, San Francisco and Manhattan (Hazuda et al., 2000). By 1982, hundreds of cases of this newly named ‘gay-related immunodeficiency syndrome’ (GRID) were identified. Intravenous drug users and haemophiliacs also began presenting with GRID symptoms, and the disease became known as ‘acquired immunodeficiency syndrome’ (AIDS) (CDC, 1982).

Early hypotheses suggested AIDS was caused by lifestyle and hyperstimulation of the immune system by foreign human antigens. Ultimately, these hypotheses failed to explain adequately certain transmission histories, and epidemiological analysis of the outbreak indicated an infectious agent (Francis et al., 1983). Initially, the herpesviruses cytomegalovirus and Epstein-Barr virus were considered, due to their previous associations with immunosuppression. However, strong evidence implicated CD4⁺ T-helper lymphocytes as the target cell and the then recent discovery of the first human retrovirus HTLV-1 and its T cell tropism led to suggestions that a similar agent was involved in the aetiology of AIDS. In 1983, several groups reported their detection of a retrovirus similar to HTLV in blood from AIDS patients (Gallo et al., 1983; Barre Sinoussi et al., 1983). Owing to its association with one of the classic symptoms of AIDS, it was initially named lymphadenopathy-associated virus (LAV). Evidence progressively mounted for LAV being the aetiological agent of AIDS (Gallo et al., 1984; Blattner et al., 1988), and in 1985, LAV and another AIDS-associated virus that workers had called human T-lymphotropic virus type III were shown to be closely related strains of the same virus – now known as human immunodeficiency virus type 1 (HIV-1) (Ratner et al., 1985).

Also in 1985, morphological and nucleic acid sequence comparisons were made between HIV-1 and other retroviruses. This led to its classification as a lentivirus, grouping it with previously described retroviruses of animals such as equine infectious anaemia virus
(EIAV) and visna-maedi virus, but distinguishing it from the oncogenic human retroviruses HTLV-1 and HTLV-II (Gonda et al., 1985). In the same year, the first complete nucleic acid sequence was obtained (Wain Hobson et al., 1985), and a closely related virus was isolated from a macaque with AIDS-like symptoms and named simian immunodeficiency virus, or SIV (Daniel et al., 1985). Other lentiviruses were discovered in primates whose geographical distribution matched that of the African HIV-1 epidemic. Over twenty SIVs have currently been identified, and each is followed by a short suffix to indicate the host from which they were isolated, e.g. SIV\textsubscript{MAC} isolated from macaques (Beer et al., 1999).

Research in central and western Africa revealed extensive anti-HIV-1 seroprevalence and many individuals with AIDS-like symptoms (Brun-Vezinet et al., 1984). In 1985, a human lentivirus with biological and morphological characteristics similar to HIV-1 was isolated from a Senegalese AIDS patient seronegative for HIV-1. Following further examination, this virus was adjudged sufficiently dissimilar to both HIV-1 and SIV\textsubscript{MAC} for it to be labelled LAV-II – now known as HIV-2 (Barin et al., 1985; Clavel et al., 1986).

At the time of writing (2002), it is estimated that over twenty million people have died from AIDS since the onset of the epidemic. The number of current infections is reckoned to be approximately 36.1 million (including cases of AIDS), with 5.4 million new infections in 1999 alone. Seventy percent of all HIV-infected individuals are in sub-Saharan Africa, nearly 18% in Asia, and nearly 5% in Latin America. It is thought that over 90% of these infected people are unaware of their infected status. Furthermore, HIV predominantly infects the economically important young adult demographic (25-44 years of age). Consequently, this pandemic is especially devastating to developing countries. The impact of HIV-1 in particular cannot be overstated as the vast majority of HIV infections worldwide are with this virus. Conversely, HIV-2 is largely restricted to countries epidemiologically linked with West Africa, e.g. Portugal, Brazil, and western India (Morison, 2001).


1.1. *Retroviridae*

HIV-1 belongs to the *Retroviridae* family of viruses – an extremely diverse and widespread group of infectious agents of vertebrates. Although the first retrovirus infecting humans (HTLV-1) was only reported in 1981, animal retroviruses were observed indirectly as long ago as 1904 when retrovirus-mediated leukaemia was transmitted between chickens. At New York’s Rockefeller Institute in 1911, Peyton Rous extracted cell-free material from a chicken with tumours. Using this, he was able to induce tumours in another chicken. The infectious agent responsible is now known to have been a retrovirus, and once isolated was named Rous Sarcoma Virus. The discovery and characterisation of many mainly oncogenic retroviruses followed these early steps, including the 1940s discovery of visna-maedi virus in Icelandic sheep, following which the name lentivirus (*lenti-* meaning slow) was coined to describe this subset of retroviruses with long incubation periods.

In 1966, Rous was awarded the Nobel Prize in Physiology or Medicine for his work on oncogenic viruses including retroviruses. Later that decade, Howard Temin observed that despite being RNA viruses, retrovirus replication could be prevented by inhibiting DNA synthesis. Together with Renato Dulbecco and David Baltimore, he shared the 1975 Nobel Prize in Physiology or Medicine for independently discovering reverse transcriptase enzymes and retroviral DNA intermediates.

Retrovirus virions are approximately 80-100nm in diameter, and take the form of spherical enveloped particles with surface glycoprotein projections. Beneath this envelope is situated the matrix, surrounding the capsid. This latter contains the usually conical core, where the genome-containing ribonucleoprotein complex is found. The envelope and the core of mature virions are visible under the electron microscope as electron-dense structures, whereas the matrix is relatively electron-transparent.

The single-stranded RNA genomes of retroviruses range from 7-11kb in length and conform to the largely conserved pattern shown in FIGURE 1-1. Like mRNA transcripts,
their 5' ends have 7-methyl guanine caps, and are poly-adenylated at their 3' ends. Three
major genes are common to all replication-competent retroviruses and are located centrally
in the order Gag, Pol, and Env, with frequent overlapping. These code for structural,
enzymatic, and envelope proteins respectively. Flanking the genes are non-coding regions
including a primer-binding site (PBS) and a poly-purine tract (PPT). These act as priming
sites during reverse transcription. Additionally, each non-coding region contains an
identical long terminal repeat sequence (LTR) essential to the regulation of viral gene
expression. Aside from these common attributes, four features of the retrovirus genome are
unique to the Retroviridae:

- Two genomes per virus particle make retroviruses truly diploid.
- The RNA genome is transcribed entirely by host cell machinery, with no contribution
  from virally-encoded polymerases.
- Specific cellular tRNA molecules are required for priming of reverse transcription.
- Retroviruses are unique amongst positive-sense single-stranded RNA viruses in that
  their genome does not act as mRNA immediately following infection of a host cell.
  Rather, virally-encoded reverse transcriptases generate DNA intermediates. After
  integration into host chromosomal DNA, these act as templates for gene expression.

Retroviruses used to be divided into three subfamilies based on their virion morphology
under electron microscopy, but phylogenetic data based on sequence analysis suggested
that this classification was inappropriate. Re-organisation of the family has led to the
current taxonomic structure of seven genera, as shown in TABLE 1-1. HIV-1, HIV-2, and
the SIVs belong to the lentivirus genus, described in further detail below (Coffin, 1996b).
In Introduction

FIGURE 1-1: The general organisation of retrovirus genomes. Protein coding regions are shaded orange. The yellow U5 and U3 sections contain sequences involved in the regulation of transcription. The PBS (primer binding site), the polypurine tract (PPT) and the R (repeat) regions comprise short sequences involved in reverse transcription.

TABLE 1-1: Retrovirus taxonomy.

<table>
<thead>
<tr>
<th>GENUS</th>
<th>SAMPLE MEMBERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammalian Type B</td>
<td>Mouse mammary tumour virus</td>
</tr>
<tr>
<td>Mammalian Type C</td>
<td>Murine leukaemia virus, feline leukaemia virus</td>
</tr>
<tr>
<td>Avian Type C</td>
<td>Avian leukosis virus, Rous sarcoma virus</td>
</tr>
<tr>
<td>Type D</td>
<td>Mason-Pfizer monkey virus, jaagsiekte virus</td>
</tr>
<tr>
<td>BLV-HTLV</td>
<td>HTLV-1, HTLV-2, bovine leukaemia virus</td>
</tr>
<tr>
<td>Spumaviruses</td>
<td>Human spuma virus, bovine syncytial virus</td>
</tr>
<tr>
<td>Lentiviruses</td>
<td>Human immunodeficiency virus type 1 (HIV-1), HIV-2</td>
</tr>
</tbody>
</table>

TABLE 1-2: Lentivirus taxonomy.

<table>
<thead>
<tr>
<th>SEROGROUP</th>
<th>SAMPLE MEMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bovine Lentivirus Group</td>
<td>Bovine immunodeficiency virus</td>
</tr>
<tr>
<td>Equine Lentivirus Group</td>
<td>Equine infectious anaemia virus (EIAV)</td>
</tr>
<tr>
<td>Feline Lentivirus Group</td>
<td>Feline immunodeficiency virus</td>
</tr>
<tr>
<td>Ovine / Caprine Lentivirus Group</td>
<td>Maedi-visna virus</td>
</tr>
<tr>
<td>Primate Lentivirus Group</td>
<td>HIV-1, HIV-2, simian immunodeficiency viruses</td>
</tr>
</tbody>
</table>
1.1.1. The Lentivirus Genus

Lentiviruses are exogenous, non-oncogenic retroviruses of vertebrates, and notwithstanding its relatively recent discovery, HIV-1 is the type species. Together with HIV-2, and the SIVs, HIV-1 belongs to the Primate Lentivirus Group, one of five subgroups of lentiviruses (TABLE 1-2). These are alternatively called serogroups, although lentiviruses are usually classified based on their hosts rather than their antigenic properties. Primate lentiviruses can also be characterised by their use of CD4 as a cell surface receptor and their lack of a dUTPase coding region (Coffin, 1996b).

After characterisation of early HIV-1 sequences derived from infected Europeans and North Americans, it was found that Central African isolates exhibited extensive diversity and could be grouped in several lineages each distinct from the European/American group. This suggests an African origin for HIV-1 from which the European/American group represented a later spread. HIV-1 has been divided into three major subgroups: O (found mainly in individuals linked to Cameroon), N, and M. FIGURE 1-2 demonstrates how subgroups M and N are more closely related to SIVCPZ(Pt), a virus of West African chimpanzees (Pan troglodytes troglodytes), than to the HIV-1 subgroup O lineage. However, the degree of sequence diversity within subgroup O is similar to that within subgroup M, implying they are of comparable evolutionary age. The recently discovered subgroup N seems to contain recombinant viruses with components of both HIV-1 subgroup M and SIVCPZ(Pt) sequences (Beer et al., 1999).

Subgroup M contains the majority of currently identified HIV-1 isolates and is further subdivided into eleven clades; A through K. Nucleotide sequence variation between clades can be as high as 30%, whereas 5-15% is typical within a clade. Retrospective phylogenetic analysis has placed a sample taken in the then Belgian Congo in 1959 close to the node at which clades B and D diverge (Zhu et al., 1998). This suggests that some of the clade diversity within HIV-1 subgroup M arose from a single human infected ‘founder’ individual and subsequent radiation rather than multiple zoonotic introductions (Yusim et
al., 2001). Moreover, 'mosaic' viruses have been observed whose Gag and Env sequences are derived from different clades. They are thought to have originated through in vivo recombination between viruses of different clades.

It is widely accepted that HIV-2 is a zoonosis arising from a virus found in sooty mangabeys (Cercocebus atys). Strong evidence comes from the phylogenetic tree shown in FIGURE 1-2, where SIV<sub>SM</sub> and HIV-2 are clustered. HIV-2 has also been subdivided with five distinct clades (A to E) recognized. However, clades C, D and E have only been described in a small number of individuals (Gao et al., 1994; McCutchan et al., 1996).

![Phylogenetic tree](image)

**KEY**

- **CPZ(Pt)** Chimpanzee (Pan troglodytes troglodytes – found in Gabon, Cameroon, Western Congo, etc.)
- **CPZ(Ps)** Chimpanzee (Pan troglodytes schweinfurthii – found in East African forests)
- **MND** Mandrill (Mandrillus sphinx)
- **AGM** African Green Monkey (Chlorocebus spp.)
- **SYK** Sykes' Monkey (Cercopithecus mitis albogularis)
- **SM** Sooty Mangabey (Cercocebus atys)

**FIGURE 1-2:** Phylogenetic tree (drawn to scale) based on Pol gene sequences showing the evolutionary relationships within the Primate Lentivirus Group, adapted from Hahn et al. (2000).
1.2. HIV-1 Infection

Human immunodeficiency virus type 1 (HIV-1) is almost universally acknowledged to be the aetiological agent of acquired immunodeficiency syndrome (AIDS), a disease characterised by progressive destruction of the immune system and depletion of CD4-positive T lymphocytes (CD4+ cells) in the peripheral blood system. AIDS leads to eventual death from infections and neoplasias which are either rarely seen at all, or rarely fatal in immunocompetent individuals. How HIV-1 causes CD4+ cell numbers to decline is unclear, but is probably a combination of a superantigen effect, apoptosis and direct viral cytopathic effects. Numerous attempts to prevent both the transmission and pathology of HIV-1 have only partially succeeded, and at the time of writing, no effective vaccine is available. However, chemotherapy can reduce virus replication and result in improved CD4+ cell counts, with newer therapies conferring longer-lasting effects.

Although the causative link between HIV-1 and AIDS has been disputed, the overwhelming circumstantial evidence reviewed in Luciw (1996) is briefly outlined here. Whilst these do not satisfy the adapted Koch's postulates developed by Evans (1982) — in particular, the lack of a suitable animal model for the full exploration of the natural history of HIV-1 pathogenesis — the levels of significance and the numbers involved in the epidemiological studies provide compelling corroboration of the link between virus and disease.

- All AIDS sufferers have detectable replicating virus both in blood and lymph nodes, and have antibodies to HIV-1 antigens in their plasma.
- The frequency of AIDS as a cause of death amongst haemophiliacs receiving contaminated blood products contaminated is enormously greater amongst those who subsequently became HIV-1-infected than those who did not.
- In vitro studies have demonstrated HIV-1 to be cytopathic to the macrophages and CD4+ T cells of the immune system.
Disease severity and progression to AIDS can be predicted by the correlates of viral burden, CD4\(^+\) T cell decline, and absolute CD4\(^+\) T cell number.

Successful reduction of HIV-1 replication through drug therapy temporally precedes a reduction in disease severity and an improvement in prognostic markers. When viruses subsequently acquire resistance and baseline replication levels are resumed, this pattern is reversed.

The demographic and geographical distributions of HIV-1 prevalence and incidence are strikingly similar to those for AIDS prevalence and incidence.

**1.2.1. Transmission**

HIV-1 is present in numerous body tissues and fluids, with genital secretions and blood being most important with regard to the routes of transmission. Casual contact does not appear to pose a transmission risk (Friedland and Klein, 1987). Consequently, HIV-1 is primarily transmitted via three routes – sexually, parenterally, and perinatally, leading to definable population subgroups (Berkelman and Curran, 1989).

Although the Western world first encountered HIV-1 in the homosexual population, heterosexual transmission in the developing world accounts for the majority of infections and this route is beginning to overtake homosexual transmission in the UK (Morison, 2001; CDSC, 2001). Transmission rates during sexual intercourse vary from approximately 0.01% to 0.8% depending upon the gender and role of the infected and uninfected individuals during intercourse (Mastro and Kitayaporn, 1998). Transmission risk is also influenced by three factors: viral burden in the infected host, integrity of the genital mucosa, and host genetics (Buchacz et al., 1998; Fleming and Wasserheit, 1999; Vernazza et al., 1999).

The second mode of transmission – parenteral – became associated with HIV-1 transmission early in the epidemic, with opportunistic infections indicative of the new
immunodeficiency syndrome diagnosed in transfusion recipients and injecting drug users (IDUs) (CDC, 1982; Des Jarlais et al., 1989). Antibody screening of blood donors has all but eliminated the former as a route of infection in the developed world, but IDUs have become the second largest source of new infections in the USA (Des Jarlais et al., 1989). Together with the bisexual population, the IDU population represents a bridge between the otherwise isolated heterosexual and homosexual groups, further facilitating the spread of HIV-1 (Rosenberg et al., 1992). Parenteral routes are the most efficient at establishing infection due to the large number of susceptible target cells in the immediate post-transmission environment, with rates from transfusion often approaching 95-100% (Coffin, 1996b).

Perinatal (or vertical) transmission of HIV-1 from mother to infant can be ante-, intra-, or post-partum. Virus crossing the placental barrier in utero is thought to account for relatively few perinatal transmission events. Nearly half of all perinatally acquired infections occur during delivery, through maternal blood either mixing with foetal blood, or contacting the mucosal surfaces of infants. A similar proportion of perinatal infections occur post-partum through breast-feeding. Colostrum often harbours high titres of HIV-1, and nipple abrasions provide a route for blood-borne infection (Luzuriaga and Sullivan, 1997).

Rates of vertical transmission vary widely, depending on geographical location, availability of drug therapy, quality of medical care, etc. In a European study, it has been estimated that between 14% and 40% of infants borne to HIV-1-infected mothers will become before 18 months, with approximately 14% of that rate being accounted for by post-partum transmission (Dunn et al., 1992). However, administration of antiretroviral therapy can reduce perinatal transmission rates by all routes by up to 70% (Bryson, 1996).

### 1.2.2. Aetiology

The first steps of infection remain unclear, and have tended to been inferred from experiments infecting rhesus macaques with SIV\textsubscript{MAC}. Transmission events not directly
inoculating the bloodstream are thought to cross the mucosa through intra-epithelial dendritic cells, Langerhans cells and CD4\(^+\) T lymphocytes, either by transcytosis or by infection (Pope, 1999). Once in the lamina propria, HIV-1 is also believed to infect mobile dendritic cells and macrophages, with local viral propagation. Infected dendritic cells are believed to be responsible for transporting the virus to lymph nodes, from where HIV-1 becomes distributed throughout the host (Granelli-Piperno et al., 1999).

For between two and four weeks post-transmission, unconstrained viral replication takes place and plasma viraemia levels rise dramatically (Daar et al., 1991). Up to half of individuals recently infected may present with an influenza-like illness (‘sero-conversion illness’), whose symptoms can include fever, arthralgia, rash, myalgia, night sweats, lymphadenopathy and diarrhoea lasting for between one and ten weeks (Tindall et al., 1988).

Following acute infection and the onset of specific humoral and cell-mediated immune responses, viraemia levels drop by up to 99.9%, and anti-HIV antibodies become detectable in the blood (Goudsmit et al., 1986). The infection enters a long period of clinical ‘latency’ during which the virus replicates without causing visible symptoms. This phase of chronic infection is characterised by continuous viral replication, and a gradual but unremitting depletion of CD4\(^+\) cells (Ho et al., 1995; Wei et al., 1995).

Exactly how HIV-1 causes the number of CD4\(^+\) T cells to decline is yet to be fully understood. The inverse relationship between plasma viraemia and CD4\(^+\) counts during early infection implies that HIV-1 directly lyses CD4\(^+\) T cells. Both the formation of multinucleated, giant cells (syncytia) and cell lysis occur \textit{in vitro}, but the ratio of uninfected to infected cells in the peripheral blood (~50 000:1) seems insufficient for this mechanism alone to drive progressive CD4\(^+\) T cell loss (Chun et al., 1997). Alternatively, a failure to generate new T cells due to HIV-1-mediated destruction of the thymus and bone marrow may be primarily responsible for this CD4\(^+\) T cell depletion. Supporting evidence for this view comes from decreasing T cell receptor repertoires and reduced naïve CD4\(^+\) T cell
counts during infection, and comparable half-lives of circulating CD4$^+$ T cells between infected and uninfected individuals (Hellerstein and McCune, 1997).

The chronic phase lasts on average between 5 and 10 years, although a minority of infected individuals called ‘rapid-progressors’ advance to AIDS within 3 years, and a similar minority of ‘long-term non-progressors’ with strong specific anti-HIV immune responses can remain asymptomatic for over 10 years (Haynes et al., 1996). It is unclear what influences the unpredictable duration of the asymptomatic phase, although both viral load and CD4$^+$ cell count immediately after the initial viraemia are predictive of disease progression (Connor et al., 1993).

Eventually clinical symptoms associated with AIDS begin to appear. They are accompanied by a phenotypic change in the virus from slowly-replicating non-syncytium inducing (slow-low) to rapidly-replicating syncytium-inducing (fast-high) phenotype (Asjo et al., 1986). A sharp decline in CD4$^+$ T cell count (below 200/μl defines AIDS) is matched by an equally rapid rise in viral load. Deterioration in CD4$^+$ cell count and elevation of plasma viraemia are accelerated: greater immune system destruction leads to more opportunistic infections, which in turn activate T cells and stimulate virus production. Strains with increasing pathogenicity are selected for, such that each generational cycle causes more damage than the last (Connor and Ho, 1994). The three stages (acute, asymptomatic, AIDS) are currently categorised A, B, and C respectively. Examples of the clinical symptoms associated with categories B and C are given below, in TABLE 1-3.
TABLE 1-3: Clinical conditions indicative of Category B and Category C HIV-1 infections.

<table>
<thead>
<tr>
<th>CATEGORY B</th>
<th>CATEGORY C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ano-rectal carcinoma</td>
<td>HIV-related wasting syndrome</td>
</tr>
<tr>
<td>Fever or diarrhoea</td>
<td>Burkitt’s Lymphoma</td>
</tr>
<tr>
<td>lasting over a month</td>
<td>Recurrent Salmonella septicaemia</td>
</tr>
<tr>
<td>HIV-related nephropathy</td>
<td>Toxoplasmosis of the brain</td>
</tr>
<tr>
<td>Pelvic inflammatory disease</td>
<td>Herpes simplex-related chronic ulcers, bronchitis, pneumonitis, or oesophagitis</td>
</tr>
</tbody>
</table>
1.3. The Life Cycle

The organisation of the integrated HIV-1 genome shown in FIGURE 1-3 conforms to the basic retroviral pattern shown in the previous chapter (FIGURE 1-1), but with additional elements, some of which are found in other lentiviruses. Aside from the Env gene and overlapping Gag and Pol domains, HIV-1 encodes four accessory proteins (Nef, Vif, Vpr, and Vpu) and two bi-cistronic regulatory proteins (Tat and Rev). These are outlined in TABLE 1-4.

Virions contain two positive-stranded RNA genomes of approximately 9.2kb, comprising only the two short repeat regions (R regions) and the sequences between them. Consequently, the mechanism of reverse transcription includes a facility for restoring the two intact LTR sequences (see section 1.3.4.2). The U3 region of the 5’ LTR domain comprises the HIV-1 transcriptional promoter, and contains a variable array of elements that bind cellular transcription factors including AP-1, Sp1, and NF-kB. Immediately upstream of the HIV-1 initiation site is a TATAA box, able to bind the transcription factors that help form the polymerase complex (Luciw, 1996). Because the genomic RNA is transcribed by cellular mechanisms, it becomes poly-adenylated and capped as per mRNA. As a further feature, HIV-1 has two polypurine tracts as the canonical 3’PPT is repeated centrally within the integrase domain of Pol.
FIGURE 1-3: Schematic of the HIV-1 genome. The upper, shaded figure represents the organisation of the integrated, double-stranded DNA genome, below which is an arrow showing the sequences encompassed by the smaller, single-stranded RNA genome found in virions. The colour coding is identical to that of FIGURE 1-1 (page 20), with the additional components of the regulatory proteins (green) and accessory proteins (blue).
TABLE 1-4: *Overview of the HIV-1 proteins.*

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein(s)</th>
<th>Size (kD)</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Matrix</strong></td>
<td>17</td>
<td>Virion structure, phosphorylated subset involved in nuclear import of viral nucleic acids</td>
<td></td>
</tr>
<tr>
<td><strong>Capsid</strong></td>
<td>24</td>
<td>Virion structure, possible role in virion uncoating, incorporating proteins into virions</td>
<td></td>
</tr>
<tr>
<td><strong>Gag</strong></td>
<td>p2gag / p1gag</td>
<td>2 &amp; 1</td>
<td>Virion structure, maintains genome architecture, widely used nucleic acid binding property</td>
</tr>
<tr>
<td><strong>Nucleocapsid</strong></td>
<td>7</td>
<td>Incorporating proteins into virions</td>
<td></td>
</tr>
<tr>
<td><strong>p6gag</strong></td>
<td>6</td>
<td>Incorporating proteins into virions</td>
<td></td>
</tr>
<tr>
<td><strong>Protease</strong></td>
<td>2 x 10</td>
<td>Cleavage of Gag and Gag-Pol precursor proteins</td>
<td></td>
</tr>
<tr>
<td><strong>Pol</strong></td>
<td><strong>Reverse Transcriptase</strong></td>
<td>51 &amp; 66</td>
<td>Reverse Transcription</td>
</tr>
<tr>
<td><strong>Integrase</strong></td>
<td>31</td>
<td>Integration of viral DNA into host chromosomes</td>
<td></td>
</tr>
<tr>
<td><strong>Env</strong></td>
<td>gp120 (SU)</td>
<td>120</td>
<td>Surface glycoprotein for CD4 receptor binding</td>
</tr>
<tr>
<td>gp41 (TM)</td>
<td>41</td>
<td>Transmembrane glycoprotein for envelope-cell membrane fusion</td>
<td></td>
</tr>
<tr>
<td><strong>Tat</strong></td>
<td>Tat</td>
<td>14</td>
<td>Transcriptional activator</td>
</tr>
<tr>
<td><strong>Rev</strong></td>
<td>Rev</td>
<td>19</td>
<td>Regulation of viral mRNA splicing and nuclear export, control of gene expression</td>
</tr>
<tr>
<td><strong>Nef</strong></td>
<td>Nef</td>
<td>27</td>
<td>Down-regulation of CD4 and MHC class I</td>
</tr>
<tr>
<td><strong>Vif</strong></td>
<td>Vif</td>
<td>23</td>
<td>Unclear</td>
</tr>
<tr>
<td><strong>Vpr</strong></td>
<td>Vpr</td>
<td>14</td>
<td>Transcriptional activation, nuclear import of viral nucleic acids, interference with host cell machinery</td>
</tr>
<tr>
<td><strong>Vpu</strong></td>
<td>Vpu</td>
<td>16</td>
<td>Down-regulation of CD4 and MHC class I</td>
</tr>
</tbody>
</table>
The HIV-1 life cycle illustrated in FIGURE 1-4 is typical of retroviruses. Beginning with the HIV-1 DNA genome integrated into host cell chromosomal DNA, viral proteins are expressed through host cell transcription and translation machinery, regulated in part by viral gene products. Together with two single-stranded positive-sense genomic RNA molecules, the major mature structural proteins derived from the HIV-1 Gag, Pol and Env genes become assembled at the cell membrane to form the virion framework. Either actively or by chance, other viral proteins and cellular molecules become incorporated into the virion structure. Mature virions subsequently ‘bud’ from the cell membrane.

Free virions are restricted in their target cell types by determinants in their surface glycoprotein molecules. Upon contact between the virion and a cell expressing the appropriate receptors, fusion between virion envelope and host cell membrane occurs. This is followed by penetration of the virion contents into the cell. Within the cytoplasm, genomic RNA is converted into double-stranded DNA by the virally encoded enzyme reverse transcriptase. By hijacking cellular pathways, this DNA molecule and other virion components enter the nucleus. Here, virally encoded integrase enzyme integrates the DNA genome into host chromosomal material (Coffin, 1996b; Luciw, 1996; Turner and Summers, 1999).

Integration may be followed by a period of latency during which the integrated genome is relatively inactive. The role of latency remains unclear. However, it is thought that the vast majority of in vivo viral production is from recently infected cells and not from reactivated latently infected cells (Wei et al., 1995).
1.3.1. Transcription and Translation

Responding to an as yet unclear signal and regulated in part by transcription factor binding sites in the U3 region of the 5' LTR, cellular RNA Polymerase II transcribes viral mRNA. Transcription initiates from the 5' end of the upstream R region of the integrated provirus and terminating after 9.2kb at the R-U5 junction in the 3' LTR. Early viral mRNAs are short and multiply spliced, and encode two small regulatory proteins Tat and Rev, and the accessory protein Nef. Tat binds a genomic RNA secondary structure at the transcription initiation site called the Tat-responsive element (TAR), and recruits both a cyclin-dependent kinase and a cyclin. Together, these phosphorylate RNA Polymerase II, enhancing processivity and consequently transcript length (Wei et al., 1998). Rev binds cis-acting sequences including the highly structured Rev-response element, and co-ordinates the splicing of viral mRNAs and hence gene expression. Through a nuclear-cytoplasmic shuttling mechanism, Rev transports these unspliced and partially spliced transcripts to the cytoplasm for translation. Multiply spliced transcripts continue to encode these early proteins, but from new, Rev-mediated species, three further accessory proteins Vpu, Vpr, and Vif are translated. A subset of full-length mRNAs is retained to act as genomes (Pollard and Malim, 1998).

Initiating at the first AUG codon of Gag and terminating at the first stop codon, conventional translation of unspliced RNA generates a 55kD precursor polyprotein called Pr55gag. However, the out-of-frame Pol gene lacks a start codon and overlaps the 3' end of Gag by approximately 40 nucleotides. For its translation, approximately 5% of Gag translations undergo a frame-shift event at a conserved ‘slippery’ site upstream of Pol (Jacks et al., 1988). The Pr55gag stop codon is suppressed, and Pr160gag-pol, a fusion protein containing both Gag and Pol, is translated. The N-terminal methionine residues of both Pr55gag and Pr160gag-pol are co-translationally removed, and the exposed glycine residues become myristoylated. This targets them to the cell membrane (Freed, 1998).

In the endoplasmic reticulum (ER), a 160kD precursor protein (gp160) is translated from singly spliced mRNA transcripts lacking the Gag-Pol region. Here, gp160
molecules associate as trimers and become extensively glycosylated. In the Golgi apparatus, gp160 is proteolytically processed into the two mature Env proteins – surface gp120 (SU) and transmembrane gp41 (TM). Trimeric, non-covalently associated TM-SU complexes are transported to the cell membrane (Wyatt and Sodroski, 1998). To reduce adventitious interactions between gp120 and CD4 (its ultimate ligand – see section 1.3.3), two mechanisms mediated by viral accessory proteins direct CD4 into cellular degradation pathways. Vpu down-regulates CD4 during its synthesis in the ER, whereas cytosolic Nef targets membrane-associated CD4. Additionally, Nef and Vpu both down-regulate cell surface expression of major histocompatibility class I molecules, possibly to protect infected cells against CTL activity (Piguet et al., 1999).

1.3.2. Virion Assembly and Release

The Gag proteins impart the basic virion structure, and although they alone can drive capsid assembly, retroviral genomic RNA is thought to act as an additional structural element (Muriaux et al., 2001). The N-terminal matrix protein components of Gag polyproteins associate and localise immediately beneath the host cell-derived virus envelope through interactions with both gp41 and cholesterol-rich regions of the cell membrane (Hunter, 1994). At the cell membrane, it is thought that the proximity of Pr160gag-pol protease domains at the cell membrane activates the enzyme through dimerisation (Lee et al., 1998). Functional protease orchestrates the ordered cleavage of both precursor polyproteins. Pr55gag releases the four mature Gag proteins matrix (p17), capsid (p24), nucleocapsid, and p6gag, together with two small peptides p1gag and p2gag. In addition, Pr160gag-pol cleavage releases the three Pol enzymes protease, integrase, and reverse transcriptase. However, due to the frame-shifting event, a protein called p6* is generated in place of p6gag and p1gag (Jacks et al., 1988; Wiegers et al., 1998). Capsid proteins form the conical capsid, which surrounds the ribonucleoprotein core. This contains two RNA genomes, linked at nucleotides 240-350 (the ‘Ψ region’ – also the encapsidation signal) by a ‘kissing loop’ structure and stabilised by nucleocapsid (Freed, 1998).
Recent research suggesting that co-precipitating cellular microvesicles contaminate analyses of virion composition has cast doubt upon the numerous claims of the incorporation of proteins into virions. The virally encoded accessory protein Vif is typical, and is now thought to be absent from virions (Dettenhofer and Yu, 1999). Vif is essential for HIV-1 infection of certain cell lines such as PBMCs but not for others. This may be due to a role in orchestrating the assembly of RNA and Gag proteins in virion cores (Ohagen and Gabuzda, 2000).

Defective Vpr genes have been found in viral genomes isolated from non-progressing infections, demonstrating it has a role in HIV-1 pathogenesis (Michael et al., 1995). It is almost certainly incorporated into virions and has multiple functions. These range from participating in the nuclear import of viral nucleic acids prior to integration, induction of post-entry cell-cycle arrest (a function shared by Vpr homologues of other lentiviruses), to interactions with transcription factors resulting in the activation of transcription from the HIV-1 LTR and cellular promoters (Bukrinsky and Adzhubei, 1999).

Other proteins believed to be incorporated into virions include the viral accessory protein Nef and cyclophilin A, a cellular heat shock protein. Incorporated Nef is cleaved by viral protease, and through an unclear mechanism has a positive downstream effect on infectivity (Marsh, 1999). Cyclophilin A appears essential for a post-entry step in the life cycle of HIV-1 (but not of other lentiviruses), possibly by ensuring that Gag proteins are correctly arranged during virion assembly (Braaten and Luban, 2001).

As virions are assembled, the cell membrane is induced to extrude and form the viral envelope in a process called budding. Membrane curvature is probably the result of spontaneously associating Gag precursors forming a symmetrical structure anchored to the membrane by the myristoylated matrix domains. A Pro-Thr/Ser-Ala-Pro motif within the p6gag protein appears critical for the 'pinching off' of the virion envelope from the cell membrane, but the mechanism by which this internal core protein influences the budding process remains unclear (Hunter, 1994; Huang et al., 1995).
1.3.3. Fusion and Entry

HIV-1 employs the cell surface marker CD4 as its primary receptor, but unusually for lentiviruses, it also needs a co-receptor for cell entry (Dalgleish et al., 1984; Miller, 1996). These belong to family of chemokine receptors, and their expression on different but often overlapping cell populations defines the host cell tropism of each virus. The two most commonly used co-receptors (CCR-5 and CXCR-4) have dissimilar host cell distributions, rendering most viruses either macrophage-tropic or T cell-tropic. Other chemokine receptors (e.g. CCR-3, CCR-2b) are less commonly employed as co-receptors by HIV-1 (Clapham and Weiss, 1997). Viruses isolated early in infection tend to be CCR-5-restricted, probably due to the CCR-5-expressing phenotype of target cells at the point of transmission. As disease progresses, there is often a switch in co-receptor usage to CXCR-4. This is the molecular basis of the change from the 'slow-low' phenotype to the fast-high phenotype mentioned in section 1.2.2, and is usually accompanied by accelerated disease progression and worsening clinical prognosis (Asjo et al., 1986; Scarlatti et al., 1997). Dually-tropic isolates able to use CCR-5 and CXCR-4 have been observed and may represent an intermediate phenotype during this switch (Doranz et al., 1996). Additionally, individuals with a rare 32 amino acid deletion within their CCR-5 receptors can be highly resistant to HIV-1 infection (Samson et al., 1996).

It is believed that fusion proceeds through a cascade of protein-protein interactions. It is well known that the binding of CD4 to gp120 induces a conformational change in gp120 (Sattentau and Moore, 1991). This is thought to facilitate interaction between the co-receptor-binding determinants of gp120 and the co-receptor itself, which exposes fusion domains of transmembrane viral glycoprotein gp41 trimers. Each gp41 molecule of the trimer contributes two α-helices to a six-helix barrel-shaped bundle – a common motif in the membrane proteins of enveloped viruses. Its insertion into the target membrane brings the two membranes together, facilitating their fusion (Berger et al., 1999).
1.3.4. Reverse Transcription

After virus entry, the virion core disassembles, releasing the RNA and virion proteins. A reverse transcription complex (RTC) is assembled in the cytoplasm. This contains the two RNA genomes, reverse transcriptase (RT), integrase, Vpr, and nucleocapsid, matrix and possibly capsid proteins. Its precise composition is uncertain as during maturation into an integration-competent pre-integration complex (PIC), various components are progressively shed (Fassati and Goff, 2001). As stated in section 1.1.1, retroviral RNA genomes are shorter than their integrated DNA genomes. Within the evolving framework of the RTC, the elegant mechanism of reverse transcription takes place, by which two LTRs are regenerated from the single non-contiguous copy present in the RNA. As this thesis deals extensively with the consequences of inhibitors targeting RT, this process is dealt with more thoroughly than the other life cycle stages.

1.3.4.1. The Structure of Reverse Transcriptase

A number of crystal structures of HIV-1 reverse transcriptase have been solved, and like several other DNA polymerases, it resembles a human hand. Five domains have been identified and labelled ‘fingers’, ‘thumb’, ‘palm’, ‘connector’, and RNAse H (Kohlstaedt et al., 1992). In its functional state, RT is heterodimeric, with subunits p66 and p51 named after their apparent size in kilodaltons. The p51 subunit comprises the N-terminal fragment of p66 after cleavage of a p66 molecule at a site upstream of the RNAse H domain. Interactions between leucine repeat motifs in the two ‘connector’ domains, and between the p51 ‘thumb’ and the p66 RNAse H domain stabilise the non-covalently associated heterodimer. Although the sequences of the amino termini of the p51 and p66 subunits are identical, the tertiary folding of p51 is dissimilar to that of p66, such that its active site is hidden. Consequently, all catalytic processes are performed by the p66 subunit, with p51 maintaining correct tertiary structure (Le Grice et al., 1991; Amacker and Hubscher, 1998).

The enzyme itself bears relatively little similarity to human cellular polymerases, although its ‘fingers’ and ‘palm’ domains share some homology at the amino acid level.
with the ribonucleoprotein telomerase complex found in all eukaryotes. Telomerases are responsible for maintaining repeat sequences at chromosome termini. Their over-activity is often considered responsible for extending the lifespan of tumour cells (Peng et al., 2001). The similarities between RT and telomerase have been considered sufficient for reverse transcriptase inhibitors such as AZT to be potential anti-cancer therapies (Murakami et al., 1999).

During reverse transcription, the sugar-phosphate backbone of the nucleic acid is accommodated in a cleft between the thumb and finger domains of p66 (Jacobo Molina et al., 1993). Within a flexible β-sheet at codons 183 to 186 within the p66 palm domain is found a YXDD amino acid motif (YMDD in HIV-1). This active site motif is common to all retroviral RT enzymes and similar to other viral polymerases such as that of hepatitis B virus. Together with several non-consecutive amino acids, this region also forms part of the dNTP binding pocket (Jacobo Molina et al., 1993; Ding et al., 1998).

The essential RNAse H activity of reverse transcriptase was mapped to the carboxy-terminal domain of p66, and crystallography revealed a general resemblance to the RNAse H of Escherichia coli (Davies et al., 1991). Two divalent cations interacting with four conserved acidic amino acids stabilise the active site conformation. Specific sequences and domains elsewhere within RT are crucial to RNAse H function, by maintaining correct alignment and proximity of the active site to its substrate (Smith et al., 1994a; Gao et al., 1998).
1.3.4.2. The Mechanism of Reverse Transcription

Reverse transcription can be divided into three stages – initiation, RNA-dependent DNA synthesis, and DNA-dependent DNA synthesis. The initiation step marks the assembly of the polymerase complex, the tRNA primer, and the genomic RNA template. In the second phase, RT catalyses the incorporation of nucleotides onto the genomic RNA template. The resultant RNA-DNA duplex is transient, as the RNase H component of RT degrades the genome template shortly after incorporation. The single-stranded cDNA copy acts as a template for second strand synthesis. The DNA-dependent DNA polymerase generates a complementary strand. Strand transfer events, described in more detail below, enable the regeneration of two LTRs from a single template copy. The mechanics of this process are detailed below, with FIGURE 1-5 providing an accompaniment to the text by charting the nucleic acid manoeuvres involved in reverse transcription.

**INITIATION**

The isoacceptor tRNA$^{\text{Lys3}}$ is incorporated into HIV-1 virions and specifically primes reverse transcription. Modified nucleotides within the tRNA$^{\text{Lys3}}$ anticodon loop specifically interact with a hairpin structure found in the U5 region of the genome. This facilitates complementary base-pairing between the 3'-terminal 18 nucleotides of tRNA$^{\text{Lys3}}$ and the HIV-1 primer-binding site (PBS) – a highly structured and highly conserved 18bp region located almost immediately downstream of the U5 hairpin (Marquet et al., 1995).

The nucleocapsid protein makes essential contributions to this duplex formation. It partially melts the tRNA and genomic RNA secondary structures and physically blocks RNase H cleavage of the genomic template (Druillennec et al., 1999a). Completion of this ‘initiation complex’ stimulates RT to incorporate five nucleotides complementary to the RNA template immediately downstream of the 3’ terminus of the tRNA$^{\text{Lys3}}$ primer (Lanchy et al., 1996a).
**RNA-Dependent DNA Synthesis**

After the initial incorporation of five nucleotides, a transition to quasi-processive elongation occurs, i.e. unlike the five initial incorporations, RT incorporates more than one nucleotide before dissociating from the primer-template heteroduplex. This transition is probably dependent upon interactions between RT, tRNA\(^{\text{Lys}}\), the upstream hairpin structure, and PBS sequences (Lanchy et al., 1998). Starting at the PBS, and continuing for approximately 180 nucleotides to the 5’ end of the template, reverse transcriptase synthesises ‘minus-strand strong-stop DNA’ (step 2 in FIGURE 1-5). Approximately 10-15 bases behind the point of nucleotide incorporation, the RNAse H domain completely degrades the RNA constituents of the newly synthesised RNA/DNA duplex (step 3 in FIGURE 1-5). However, the tRNA\(^{\text{Lys}}\)/PBS RNA/RNA homoduplex is resistant to this cleavage, allowing later copying of the PBS (Gotte et al., 1995).

Continuation of reverse transcription is marked by the annealing of the minus-strand strong-stop DNA to the complementary bases of the 3’ R region belonging either to the same template or to that of the co-packaged RNA genome molecule (Panganiban and Fiore, 1988). Homologous and heterologous acceptor genomes seem to be selected with equal frequency (van Wamel and Berkhout, 1998). This is the first ‘strand transfer’ event of reverse transcription, and is greatly assisted by three properties of nucleocapsid protein – strand-annealing, inhibition of self-priming by the minus-strand DNA, and effecting changes in RNAse H action (Peliska et al., 1994; Guo et al., 1997). Synthesis of minus-strand DNA restarts and extends across the genome to the PBS in the 5’ LTR. Again, the RNA template is degraded by the trailing RNAse H domain, with the exception of the two polypurine tracts (PPTs, step 4 in FIGURE 1-5).

**DNA-Dependent DNA Synthesis**

Co-incident with minus-strand synthesis, the retained PPT RNA sequences prime synthesis of two positive-sense DNA strands labelled the upstream and downstream segments (labelled U+ and D+ respectively, step 5 in FIGURE 1-5). D+ synthesis proceeds as far as the 3’PPT primer. Here, the strand displacement facility of RT separates the DNA/DNA duplex allowing elongation to the end of the U3-R-U5 LTR.
The displaced U+ positive-strand DNA acts as template for a further minus-strand synthesis extending to the 3’PPT, and so two LTRs are generated.

U+ synthesis halts at a methylated adenosine 19 bases inside the preserved tRNA^{Lys3} primer, thus copying the entire PBS (Ben Artzi et al., 1996). Eventually, RNAse H degrades the RNA at the two PPT sites and the tRNA primer is released. This allows the second strand transfer event to take place, facilitated by a dimeric integrase ‘bridge’ maintaining the homologous 5’ and 3’ termini in close proximity (Miller et al., 1997). The single-stranded regenerated U+ PBS (primer binding site, see Initiation above) anneals to the complementary minus-strand PBS (Panganiban and Fiore, 1988). Transcription is resumed and proceeds until the growing DNA strand reaches the first deoxy-nucleotide of the D+ segment. A second strand displacement event takes place, allowing the U+ strand to be extended by a further 99 nucleotides until the ‘central termination sequence’. Here, the minor groove of the DNA is compressed, ejecting the RT enzyme and terminating reverse transcription (Chameau et al., 1994). The three-strand ‘central DNA flap’ thus created is unique to lentiviral reverse transcription and was recently shown to be essential for the post-transcription nuclear import of PICs (Zennou et al., 2000).
FIGURE 1-5: Schematic diagram of the process of reverse transcription (after Gotte et al., 1999).

1. $\text{R} \quad \text{U5} \quad \text{PBS} \quad \text{Gag} \quad \text{Pol} \quad \text{Env} \quad \text{U3} \quad \text{R}$
   - Digested by RNAse H

2. 
   - RNAse H digestion retarded

3. 
   - Digested by RNAse H

4. 
   - RNAse H digestion retarded

5. $\text{D}^+ \quad \text{U}^+$
   - Digested by RNAse H

6. $\text{D}^+ \quad \text{U}^+$

7. $\text{U3} \quad \text{R} \quad \text{U5} \quad \text{PBS} \quad \text{Gag} \quad \text{Pol} \quad \text{Env} \quad 	ext{U3} \quad \text{R} \quad \text{U5}$
   - Elongation of DNA strand

Legend:
- **RNA**
- **DNA**
- **tRNA$_{\text{Lys}}^3$**
1.3.5. Integration and Latency

Most retroviruses require dissolution of the nuclear membrane during mitosis for the provirus to access the cellular genome. However, through nuclear localisation signals on the accessory protein Vpr and possibly a phosphorylated subset of p17 matrix protein, active transport of the pre-integration complex to the nucleus by cellular pathways allows HIV-1 and other lentiviruses to infect non-dividing and terminally differentiated cells such as macrophages (Whittaker and Helenius, 1998). Once in the nucleus, reverse-transcribed DNA becomes integrated into the host cell genome. Briefly, viral integrase recognises LTR sequences, and with little discernible target sequence preference, host chromosomal DNA is nicked, allowing ligation with the termini of the viral genome (Farnet and Bushman, 1996).

The numerous roles of the accessory protein Vpr are central to orchestration of the events between integration and expression of the viral genome. Two pathways appear most relevant. Firstly, Sp1, a cellular transcription factor whose binding sites are common in LTR sequences, and TFIIB, an essential component of the RNA polymerase II transcription complex, directly interact with Vpr and are thought to stimulate transcription. The second, more important pathway is intimately linked to the cell cycle arrest function of Vpr, although the precise details of this process are not certain. The two cellular proteins cyclin B1 and cdc2 associate to mediate transfer between the various stages of the cell cycle. Vpr can bind this complex, and in addition to inducing arrest in G2 phase, this interaction appears to influence transcription through p300 and the two transcription factors NF-κB and NF-IL-6. Binding sites for these two transcription factors are commonly found within LTR sequences (Bukrinsky and Adzhubei, 1999; Roux et al., 2000).
1.4. Antiretroviral Therapy

There are a number of stages in the HIV-1 life cycle described in the previous section at which therapeutic intervention is feasible. These include the early stages of transcription and translation, polyprotein precursor processing, virus maturation, budding, attachment (i.e. virion-cell interaction), fusion of the virion and cell membranes and capsid uncoating, reverse transcription, and finally integration. However, currently available clinically useful compounds target only two of these steps – precursor processing by viral protease, and reverse transcription. More encouragingly, several compounds effective against other components of the life cycle have recently been developed and many are already undergoing clinical trials.

Early compounds successful in inhibiting HIV-1 replication belonged solely to the nucleoside analogue class of reverse transcriptase (RT) inhibitor. Single-drug therapy successfully suppressed viral burdens and improved CD4\(^+\) cell counts. However, after prolonged treatment, many viruses became resistant to the drug and these benefits were reversed (Fischl et al., 1987; Larder et al., 1989). The introduction of dual nucleoside analogue therapies targeting RT delayed treatment failure by demanding viruses become resistant to both drugs, but cross-resistance, poor antiviral effect, and unpleasant side effects limited the benefits of this approach (Mayers, 1996).

The mid-nineties saw the advent of two new classes of drug – one targeting RT, one targeting protease. Combination therapy with compounds drawn from two or more classes of inhibitor heralded a major advance in antiretroviral therapies. Not only was clinical prognosis improved, the time to resistance was lengthened, functional immunity could be restored, and a wider range of salvage therapy options upon treatment failure became available (Vella and Palmisano, 2000). Unfortunately, the ultimate goal of eradicating virus from infected individuals still appears unattainable, as latent virus persists in sequestered reservoirs (Chun and Fauci, 1999; Finzi et al., 1999).

Interestingly, following the onset of highly active antiretroviral therapy (HAART), virus isolates taken from patients whose viruses previously exhibited CXCR-4 co-receptor
usage had predominantly CCR-5-utilising phenotypes. The rapidly replicating, more pathogenic T cell-tropic virus population may be disproportionately affected by antiretroviral therapy when compared to slower-growing M-tropic viruses (Philpott et al., 2001).

With the exception of foscarnet, all currently licensed reverse transcriptase inhibitors belong to one of three classes – nucleoside analogues, acyclic nucleotide analogues, and non-nucleoside RT inhibitors (NNRTIs). Nucleoside analogues competitively inhibit RT by mimicking the cellular precursors of the deoxy-ribonucleotide substrates of reverse transcriptase, and terminating strand extension. The acyclic nucleotide analogues have a similar mechanism of action to nucleoside analogues, but can be distinguished from the latter by partial phosphorylation and incomplete ring structures. In contrast, the newer NNRTIs act non-competitively by binding an allosteric site within RT, slowing the catalytic rate of the enzyme.

Falling into neither class, foscarnet (phosphonoformate, or PFA) is a pyrophosphate analogue, and inhibits pyrophosphorolysis during nucleic acid chain extension by lentivirus and herpesvirus DNA polymerases in vitro at concentrations well below those necessary for pyrophosphate feedback inhibition of cellular DNA polymerases. However, it rarely used to treat HIV-1 infected individuals as it is highly nephrotoxic and does not lower HIV-1 levels in vivo. However, inhibitory compounds derived from PFA are in development (Wagstaff and Bryson, 1994; Hammond et al., 2001).
1.4.1. Nucleoside Analogues

In 1987, the nucleoside analogue reverse transcriptase inhibitor 3'-azido-3'-deoxythymidine (AZT) became the first anti-HIV compound licensed for clinical use. After showing early promise, the emergence of resistance limited its effectiveness (Fischl et al., 1987; Larder et al., 1989). Later, four further nucleoside analogues with similar modes of action were approved (TABLE 1-5). They each resemble one of the four cellular nucleosides. Crucially, in each analogue, the hydroxyl group attached to carbon-3 (C^3) of the ribose moiety has been replaced by alternative groups. In the case of 3-TC, C^3 itself has been replaced by a sulphur atom. By lacking C^3 hydroxyl groups, nucleoside analogues block formation of the phosphodiester bond with the incoming nucleotide, prematurely terminating transcription (Arts and Wainberg, 1996).

Analogues of nucleosides are used because their nucleotides have poor cell penetration. Cellular kinases phosphorylate nucleosides to nucleotide triphosphates. For instance, thymidine kinase naturally converts thymidine to thymidine monophosphate (TMP), but also converts AZT to AZT monophosphate with similar efficiency. Phosphorylation of AZT monophosphate to AZT diphosphate by thymidylate kinase is relatively inefficient, as the enzyme has far higher affinity for thymidine monophosphate than AZT monophosphate. The final kinase, nucleotide diphosphate kinase, phosphorylates all nucleotide diphosphates (Lavie et al., 1997; Peter and Gambertoglio, 1998).

1.4.2. Acyclic Nucleotide Analogues

Abacavir and adefovir are newer analogues of deoxyadenine. They are termed acyclic nucleotide analogues as they are partially phosphorylated when administered, and the ring structures of their sugar moieties are incomplete. Both show similar or greater antiviral effect than AZT, although the oral bioavailability of adefovir is poor, and abacavir can provoke hypersensitivity and other side effects (Foster and Faulds, 1998; Noble and Goa, 1999). The efficacy of adefovir in clinical trials has been questioned, and is expected soon to be replaced by a derivative called tenofovir. Many new
nucleoside analogues are under development, including some that are active against RT enzymes resistant to other drugs of this class (De Clercq, 2001).

1.4.3. Non-Nucleoside Reverse Transcriptase Inhibitors

Three non-nucleoside reverse transcriptase inhibitors (NNRTIs) targeting specifically HIV-1 RT are currently licensed – nevirapine, delavirdine, and the newer efavirenz. They act by non-competitively binding a hydrophobic region of the p66 subunit of RT close to the tyrosines at codons 181 and 188 involved in active site structure (section 1.3.4.1). Upon binding, the orientations of these tyrosines are altered, forming a pocket. Crystallography has suggested this region forms part of a ‘hinge’ for the catalytic thumb domain, and NNRTIs may inhibit RT function by displacing amino acids in the primer grip region, loosening template-primer binding and destabilising the active site architecture. Binding of substrates to RT would not necessarily be affected, but their relative proximities may be altered, hindering the biochemical interactions needed for incorporation (Spence et al., 1995). Many new NNRTIs binding the hydrophobic pocket are currently being developed, with some showing high activity against NNRTI-resistant strains. These may inhibit RT through yet another mechanism (De Clercq, 1998; De Clercq, 2001).

TABLE 1-5: Table of nucleoside analogue compounds currently licensed for anti-HIV therapy.

<table>
<thead>
<tr>
<th>Trade Name</th>
<th>Common Names</th>
<th>Cellular analogue</th>
<th>Modification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrovir</td>
<td>AZT, Zidovudine</td>
<td>deoxythymidine</td>
<td>$C^3$ OH replaced by $N_3$</td>
</tr>
<tr>
<td>Videx</td>
<td>ddI, Didanosine</td>
<td>deoxyadenine</td>
<td>$C^3$ OH replaced by $H$</td>
</tr>
<tr>
<td>Hivid</td>
<td>ddC, Zalcitabine</td>
<td>deoxycytidine</td>
<td>$C^3$ OH replaced by $H$</td>
</tr>
<tr>
<td>Epivir</td>
<td>3-TC, Lamivudine</td>
<td>deoxycytidine</td>
<td>$C^3$ replaced by $S$</td>
</tr>
<tr>
<td>Zerit</td>
<td>d4T, Stavudine</td>
<td>deoxythymidine</td>
<td>$C^3$ OH group lost, $C^2$-$C^3$ bond doubled</td>
</tr>
</tbody>
</table>
1.4.4. Protease Inhibitors

Protease cleavage of polyprotein precursors is another enzymatic function coded for by the virus, and has become one of the two targets for clinically available anti-HIV drug therapy (the other being RT). Ease of recombinant protein production and purification has facilitated extensive research into the dimeric enzyme and its inhibition. Reference to other retroviral proteases together with biochemical and crystallographic analysis confirmed that HIV-1 protease comprises two identical 99 amino acid subunits, and like human renin, is an aspartyl protease (Navia et al., 1989; Fitzgerald and Springer, 1990). Its natural substrates are nine dissimilar and variable motifs at the sites of Pr55gag and Pr160gag-pol precursor cleavage. Failure to cleave these polyproteins leads to the release of immature, non-infectious virions (Kohl et al., 1988; Freed, 1998). Combinations of random screening and rational drug design led to the development of the five currently licensed protease inhibitors (PIs – saquinavir, ritonavir, indinavir, nelfinavir and amprenavir). These are ‘peptidomimetic’ in that they mimic the short peptide substrates of HIV-1 protease, binding the enzyme’s active site and competitively inhibiting the enzyme (Martin, 1992).

Significant side effects have been reported by patients on PI-containing regimens. Most notable are gastrointestinal problems such as nausea, diarrhoea and vomiting, and metabolic problems such as abnormal processing of lipids, dangerously elevated cholesterol levels, and hyperglycaemia. Future PI design must take these increasingly serious side effects of long-term PI use into account (Graham, 2000).

1.4.5. Other Antiretroviral Strategies

Unfortunately, of all potential drug targets, only protease and RT have yet proven susceptible to clinically useful compounds. The utility of potential compounds depends on many factors on top of their antiretroviral effects. Oral bioavailability, pharmacokinetics, toxicity, target specificity, activity against viruses to existing drugs, efficacy against diverse strains of HIV-1, compatibility with other therapies, ability to
cross the blood-brain barrier, and the cost of continuous treatment must all be taken into account.

Much early interest was in the interaction between the viral envelope glycoproteins and CD4. Soluble forms of CD4 (sCD4) were tested, but despite good \textit{in vitro} blocking activity, early doses proved ineffective \textit{in vivo} due to short half-lives, resistant clinical isolates and ultimately little effect on plasma viraemia. Elevated doses of sCD4 conjugated to immunoglobulin G molecules showed improved pharmacokinetics but little clinical benefit (Hadden, 1998).

The discovery of co-receptors led to renewed interest in virus entry as a target, and explained the mode of action of bicyclams – a class of compounds now known to inhibit HIV-1 replication by interfering with CXCR-4 usage. To date, AMD3100 is the most promising, but low oral bioavailability, toxicity, and the rapid emergence \textit{in vitro} of resistant virus continue to hamper its usefulness \textit{in vivo} (De Clercq, 2000).

A short peptide called T22 with anti-HIV properties was also found to act through CXCR-4 antagonism. Since its discovery in 1992, more potent and less cytotoxic derivatives with activity against bicyclam-resistant viruses have been developed. To date, viruses resistant to this class of inhibitor have not been reported, despite determinants of viral susceptibility having been located within the hypervariable V3 loop of gp120 – a region noted for its plasticity (LaBranche et al., 2001). Other developments towards fusion and entry inhibitors include an inhibitor of CCR-5-mediated entry with \textit{in vitro} anti-HIV activity, and 5-Helix, a promising synthetic fusion inhibitor mimicking a six-helix bundle found in gp41 essential for membrane fusion (Cammack, 2001).

The virally encoded enzyme integrase is essential to the viral life cycle and has no cellular homologue. Despite its catalytic domain having been crystallised, only weak inhibitors of integrase have so far been discovered. It is hoped that these will act as ‘leads’ in the development of more potent inhibitors (Pani and Marongiu, 2000). Interventions at other life-cycle stages are still in their infancy – licensed therapies for \textit{in vivo} use remain distant. Drug-based approaches include a peptide mimic of the
nucleocapsid protein (Druillennec et al., 1999b), and Vpr-binding compounds inhibiting nuclear import of viral nucleic acids (Dubrovsky et al., 1995). Others are employing gene therapy vectors to deliver specific DNA sequences to infected cells. The range of products that these could express includes antisense RNA designed to anneal to viral transcripts such as the TAR, decoy RNA to inhibit Rev, HIV-1-specific ribozymes, or mRNA to be translated into antibody fragments binding specific viral proteins (VandenDriessche et al., 1997).

In parallel to the design of novel therapeutic compounds, a vast amount of work is being done towards the development of both preventative and therapeutic HIV-1 vaccines. Immunogenic variability, sequestration, infection of many immune system cell types, inability to identify suitably protective immune mechanisms, and varied transmission routes all hinder advances in this field. The discovery of HIV-1 specific immunity in persistently exposed but uninfected individuals, and the successful application of new technologies to achieve protective immunity in primates have both been encouraging developments in this field. Many vaccine trials are currently being carried out, with systems as diverse as simple peptide stimulation through DNA vaccination, to bacterial and viral vector delivery systems (Hanke, 2001).

1.5. Resistance to Antiretroviral Compounds

A major cause of antiretroviral treatment failure is the evolution of drug-resistant HIV-1. Early therapies with single drugs swiftly resulted in the emergence of resistant viruses (Fischl et al., 1987; Larder et al., 1989). With the development of additional antiretroviral compounds, dual therapies became common, but this appeared to only delay resistance rather than preclude it (Mayers, 1996). The arrival of NNRTIs and protease inhibitors in the mid-1990s saw great improvements in treatment regimens, with the ability to target simultaneously two separate and essential virus proteins. This requires HIV-1 to become simultaneously resistant at two independent genes.
Consequently, the antiretroviral effect of treatment with one or more RTIs will inhibit development of resistance to a co-administered protease inhibitor and *vice versa*.

Phenotypic resistance is quantified *in vitro* by culturing virus in the presence of a serial dilution of a drug, and calculating the concentration required to inhibit virus growth by either 50% (the IC$_{50}$) or 90% (IC$_{90}$). Comparing IC$_{50}$ or IC$_{90}$ values of viruses varying only at a single codon allows the fold increase in resistance conferred by that mutation to be evaluated. Similarly, comparing levels of phenotypic resistance between pre-therapy and post-therapy isolates enables assessment of treatment efficacy. However, phenotypic testing is laborious and time-consuming, and genotyping is now commonly used, in which protease and RT genes are analysed by sequencing or hybridization assays. By comparing the mutational pattern to viruses whose phenotypic resistance profile has already been established, sample phenotypes can be inferred (Garcia Lerma and Heneine, 2001). The output of this latter technique comprises a range of susceptibilities, as the degree by which a given resistance mutation lowers susceptibility to a given drug is modulated by the genetic background of each individual virus (Rose et al., 1996; Kemp et al., 1998; Garcia Lerma et al., 2000; Precious et al., 2000).

Phenotypic resistance to RT or protease inhibitors can be achieved through single amino acid substitutions in the target gene, and for each currently licensed inhibitor, resistant isolates have been observed (Pillay et al., 2000). Indeed, isolates both phenotypically and genotypically resistant to multiple drugs from all classes have been found (Shafer et al., 1998; Palmer et al., 1999). Many mutations can confer resistance to a given drug, and interactions between resistance mutations both to a single drug and to multiple drugs can be variously additive, synergistic, and antagonistic, leading to further confusion. Most resistance mutations are selectively disadvantageous and are rarely seen in untreated viruses. However, improved detection methods have demonstrated that *in vivo* populations can contain low frequencies of genotypes that are resistant to drugs to which they have not been exposed (Najera et al., 1994, 1995).

A distinction has been made between ‘primary’ and ‘secondary’ resistance mutations. Primary mutations are the first to evolve under drug challenge, and are selected primarily for the reduction in susceptibility to the drug that they confer. The resistance
benefits that primary mutations are often accompanied by a reduction in enzyme functionality. The pattern by which secondary mutations evolve is frequently dependent upon the presence of particular primary mutations, and their contribution to drug resistance is not the only criterion for their selection. Their ability to restore functionality to an enzyme debilitated by primary mutations constitutes an equivalent selective pressure (Maeda et al., 1998; Nijhuis et al., 1999; Robinson et al., 2000).

Early observations revealed a relationship between the duration and dosage of AZT monotherapy with the frequency and degree of phenotypic resistance of virus isolates (Larder et al., 1989; Boucher et al., 1990). Sequence analysis revealed that the emergence of resistance was characterised by particular amino acid substitutions within the RT domain—D67N, K70R, T215Y/F, and K219Q). *In vitro* culture of virus in the presence of AZT led to the selection of drug resistance variants harbouring similar mutational patterns (Larder et al., 1991). Insertion of these mutations into an AZT-susceptible virus rendered viruses AZT-resistant, with the degree of resistance related to the number of mutations inserted (Larder and Kemp, 1989). Later, isolates exhibiting high levels of phenotypic AZT resistance with only partially resistant genotypes were found to possess a fifth mutation—M41L (Kellam et al., 1992). After long-term AZT therapy, a further substitution at RT codon 210 from leucine to a tryptophan is often detected. L210W is rarely detected in viruses lacking T215Y, as it is believed to reduce the destabilising effect of the 215Y mutation (Yahi et al., 2000). This is an example of a secondary mutation—L210W being dependent upon the presence of the primary mutation T215Y. The former not only contributes to resistance, but also improves RT function (and hence virus fitness) by partially restoring the defect incurred by the latter.

Evolution to an AZT-resistant genotype is frequently through a pattern of ordered mutation acquisitions (Boucher et al., 1992). Both the degree by which RT function is compromised by certain intermediately resistant genotypes and the mutational bias of RT itself are thought to be responsible for this pattern (Keulen et al., 1997a; Jeeninga et al., 2001).

Resistance to the other nucleoside analogues commonly arises through mutations other than those described above for AZT. T69D is frequently observed in ddC-resistant
isolates, but not in those resistant to ddI. This is the exception, as resistance to ddI and ddC is often acquired through the same mutations (K65R, L74V, and rarely, M184V), probably due to the structural similarities between the two drugs (St.Clair et al., 1991; Fitzgibbon et al., 1992; Zhang et al., 1994). M184V is the most common mutation seen in response to 3-TC therapy, increasing the IC₅₀ more than 100-fold. Due to the mutational bias of RT, it is often acquired in two stages, passing through an isoleucine intermediate (Boucher et al., 1993; Keulen et al., 1997a). Resistance to 3-TC can also be acquired through E44D and V118I, which together can confer between four and 50-fold resistance (Hertogs et al., 2000a). Isolates resistant to d4T exhibit almost no genotypic commonality. Mutations in RT codons are often observed during therapy, but only V75T and I50T occur with any frequency (Lin et al., 1999). Interestingly, recombinant viruses containing RT domains from d4T-resistant isolates often remain sensitive to the drug, suggesting mechanisms of resistance may lie elsewhere in the genome (Lin et al., 1994). Moreover, treatment with one nucleoside analogue can often select for mutations specifically associated with resistance to other drugs of this class, despite a lack of documented cross-resistance (Najera et al., 1994; Ross et al., 2001).

A small number of patients receiving dual nucleoside analogue therapy (AZT together with either ddI or ddC) develop a different pattern of mutations to those described above, with Q151M appearing first, followed by A62V, V75I, F77L, and F116Y. Together, these five mutations bestow high-level resistance to all nucleoside analogues (Shirasaka et al., 1995). The Q151M substitution confers only partial resistance to nucleoside analogues, and individually, the four secondarily acquired mutations have no effect upon resistance. However, together with Q151M, they both increase resistance and partially compensate for a reduced enzyme efficiency incurred by Q151M (Maeda et al., 1998). This is a classic example of the distinction between primary mutations (Q151M) and secondary mutations (the four subsequently acquired changes).

A second, even less common mechanism of resistance to multiple nucleoside analogues has also been seen in isolates from patients receiving AZT plus either ddI or ddC. The mutational pattern involves rearrangements in the β3-β4 loops of the finger domains. Most frequently, this takes the form of an insertion of between one and three amino acids (typically serines) in the region spanning codons 67 through 70 (Larder et al.,
1999). Four multinucleoside-resistant isolates exhibiting this type of mutational pattern illustrate the great capacity of reverse transcriptase to tolerate substantial mutation. One displayed five unusual substitutions in the amino acids surrounding a single insertion between codons 69 and 70. In another, a secondary mutation resulting in deletion of codon 67 compensated for the functionally deleterious T69G primary resistance-conferring mutation (Rakik et al., 1999; Imamichi et al., 2000). A recent retrospective study has identified two further RT domains with five and eleven amino acid insertions between codons 68 and 69 (Masquelier et al., 2001).

An observed infrequency of *in vivo* co-resistance to AZT and 3-TC supported earlier *in vitro* work in which M184V suppressed the effects of AZT resistance mutations (Tisdale et al., 1993; Larder et al., 1995). Recent work has shown how the RT mutation G333D/E can reverse this antagonistic effect (Kemp et al., 1998). Reverse transcriptases containing AZT-resistance mutations together with M184V may have reduced enzymatic function. This would inhibit their selection despite resistance advantage (Naeger et al., 2001). Similar research has revealed how in the absence of other AZT resistance mutations, viruses harbouring just M41L and K70R exhibit resistance to AZT at the cost of substantially reduced RT function, explaining the absence of this combination from AZT-treated individuals (Jeeninga et al., 2001).

Resistance to abacavir develops slowly, with K65R, L74V, Y115F, and M184V the mutations most frequently seen both *in vitro* and *in vivo*. L74V and M184V are the most frequent combination. The M184V mutation indicates cross-resistance with ddl, ddC, and 3-TC, but also possible incompatibility with AZT resistance (Harrigan et al., 2000). Conversely, M184V sensitises HIV-1 to adefovir, suggesting potential therapeutic synergy with 3-TC (Miller et al., 1999). In studies of adefovir resistance, K70E/R develops both *in vitro* and *in vivo*, although by no means frequently. To confound the issue, many other mutations have been observed *in vivo* (Mulato et al., 1998).

At the molecular level, the mutations described in the previous paragraphs are thought to confer resistance through a variety of mechanisms. Until recently, two mechanisms were invoked to describe how these mutations conferred resistance to nucleoside analogues. Firstly, codons 41, 184, 215 and 219 in the p66 subunit are located close to
the nucleotide binding pocket of RT. Substitution at these sites may directly affect nucleotide binding such that natural substrates are favoured over analogues (Tantillo et al., 1994; Gu et al., 1994). Secondly, codons 65 through 74, 215, and 219 of the p66 subunit contact the template-primer nucleic acid duplex. Their mutation is thought to alter the alignment of the duplex relative to RT at the polymerase active site. Although such conformational changes may reduce the overall rate of nucleotide incorporation, this could allow greater discrimination at the point of incorporation (Boyer et al., 1994a, 1994b; Tantillo et al., 1994). Thirdly, the recently discovered ability of HIV-1 RT to excise misincorporated bases provides a further mechanism by which RT can acquire resistance to nucleoside analogues (Canard et al., 1998). It appears that the AZT resistance mutations described above (at codons 41, 67, 70, 210 and 215) confer improved ability to excise misincorporated AZT (Meyer et al., 1999; Boyer et al., 2001).

Viruses resistant to the non-nucleoside reverse transcriptase inhibitors are easily obtained both in vitro and in vivo during monotherapy, although their appearance can be delayed under combination therapy regimes (Havlir and Lange, 1998). Most mutations conferring resistance to NNRTIs are found in two clusters near to the binding pocket in the folded enzyme – codons 98-108, and 179-190. TABLE 1-6 lists those found most commonly in NNRTI-resistant isolates. The mutations at positions 103, 181 and 188 are frequently observed in vivo and confer cross-resistance between NNRTIs. This contrasts with in vitro selection studies where drug-specific mutations are more commonly found (Maga et al., 1997). Interestingly, although no cross-resistance between NNRTI and nucleoside analogues would be expected, the NNRTI mutation Y181C partially reverses the effects of AZT resistance mutations, further enhancing the role of NNRTIs in combination therapy regimens (Byrnes et al., 1994). Finally, NNRTI resistance mutations (e.g. V108I, Y181C) are occasionally seen as natural polymorphisms in non-clade B viruses, in particular those belonging to clade O (Descamps et al., 1997; Cornelissen et al., 1997).

Numerous combinations of amino acid substitutions within protease accompany phenotypic resistance. Many mutations are inhibitor-specific; others are common to two or more protease inhibitor (PI) resistance patterns. Furthermore, such is the plasticity of
HIV-1 protease gene, resistance to a given PI can be acquired through several mutational pathways. For this reason, it is perhaps unsurprising that extensive cross-resistance between PIs is observed (Hertogs et al., 2000b). However, tipranavir, a recently developed non-peptidomimetic PI undergoing clinical trials, shows antiretroviral activity against resistant isolates (Larder et al., 2000). More recently, computational analysis of structural data is being used in the investigation of mechanisms of resistance to protease inhibitors and in the design of future inhibitors (Wang and Kollman, 2001).

Two crucial factors influence the development of drug resistance by HIV-1 – the genetic background of the virus, and the population genetics of the virus within the host. As touched upon above when discussing how genotypic information is used to infer phenotypic resistance levels, the genetic identity of protease and/or RT prior to treatment can modulate the mutational route by which HIV-1 acquires drug resistance. For example, in a recent study, the presence at baseline of non-consensus amino acids at codons 39 and 44 of RT was shown to be significant in predicting virological failure in patients receiving AZT and ddl, either with or without nevirapine (Precious et al., 2000). In protease, it has been shown in vitro that genetic background influences the degree by which different primary resistance mutations reduce enzymatic function. Consequently, which primary mutations will be selected under drug challenge will not be determined solely by their effects on resistance (Rose et al., 1996).

<table>
<thead>
<tr>
<th>NNRTI</th>
<th>Mutation(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nevirapine</td>
<td>A98G, L100I, K103N, V106A, V108I, Y181C, Y188C, G190A</td>
</tr>
<tr>
<td>Delavirdine</td>
<td>K103N/T, Y181C, P236L</td>
</tr>
<tr>
<td>Efavirenz</td>
<td>L100I, K101E/Q, K103N, V108I, Y188H/L, G190S/A/E, P225H</td>
</tr>
</tbody>
</table>

TABLE 1-6: Table of mutations in RT associated with lowered susceptibility to NNRTIs. Mutations in bold type are those most frequently found in vivo (Deeks, 2001).
1.6. Evolution of HIV-1

1.6.1. RNA Virus Population Dynamics

RNA viruses exhibit great genetic diversity through high mutation rates, recombination, rapid generation times, high progeny yields, genomic re-assortment and large population sizes (Domingo and Holland, 1997). Because of this genetic diversity, rather than comprising many copies of a single genotype, RNA virus populations takes the form of a dynamic distribution of related genotypes called a quasispecies. This concept was originally used to describe theoretical populations of self-replicating informational macromolecules (Eigen and Schuster, 1977), but is now widely applied to models of RNA virus evolution (Holland et al., 1992). Members of a quasispecies cluster in a theoretical multi-dimensional construct known as ‘sequence space’ in which all permutations of a given genome length are located.

A quasispecies comprises a subset \{S_k\} of points within sequence space, with each point \(S_k\) represented as a string of length \(N\) \(S_{ki}...S_{kn}\), with \(N\) being the length of the genome being described (approximately 9200 in the case of HIV-1). Each element of this string takes one of four values, corresponding to the four bases (A, C, G, or U). Under a given set of environmental conditions, each point in sequence space has an associated ‘relative fitness’. This is a dimensionless parameter representing the selection co-efficient and quantifies the relative capacity of each genotype to produce stable infectious progeny (Holland et al., 1991).

Key concepts regarding quasispecies evolution can be derived from the geometric evolution model proposed by Sir Ronald Fisher (Fisher, 1999). The basic principle is demonstrated below in FIGURE 1-6. The Cartesian axes \(x\) and \(y\) represent continuous phenotypic variables whose fitness optima are at the origin (‘O’). The fitness of each point in this space declines with its distance from the origin (the Hamming distance). Point ‘A’ lies on a contour of points of equal Hamming distance. The concentric circles centred upon point ‘A’ represent quasispecies distribution. A small quasispecies
population, with limited mutational reach can be represent by the smaller of the two concentric circles surrounding point ‘A’. Here, fitness gains are likely to be small, with progressive, step-wise evolution towards higher fitnesses. A larger population, or one with broad mutational reach, would be represented by a larger circle. As selection acts upon the quasispecies, rather than the dominant phenotype (point ‘A’), evolution to higher fitness is more likely to be through larger steps. This concept has been demonstrated experimentally using differently-sized populations of unfit bacteriophage Φ6. Larger populations tended to recover fitness by leaps, whereas fitnesses of smaller populations climbed only gradually (Burch and Chao, 1999).

FIGURE 1-6: A simplified two-axis representation of Fisher's geometric model of evolution.
Population size and mutation rate each have different influences upon evolution, despite both contributing to the diameter of the circle around point A. A higher proportion of mutations from point A are deleterious rather than beneficial, with the proportion increasing with the size of this circle. Therefore, the distribution of fitnesses describing the quasispecies around point A is normal, but with lower mean fitness than point A itself. The Muller’s ratchet effect, whereby deleterious mutations are gradually accumulated by asexually reproducing organisms, is accentuated in small populations, as they are vulnerable to stochastic evolution by genetic drift. Large populations, even with low mutational rates, evolve in a deterministic fashion as there are sufficient members with higher than average fitness to counteract the propensity of each individual to lose fitness.

Experimental demonstration of Muller’s ratchet comes from genetic bottleneck experiments. Very small subsets of RNA virus populations were used to found new populations. The magnitude of fitness loss experienced by a new population was inversely proportion to the size of the population transfer, i.e. smaller transfers led to more substantial fitness losses (Duarte et al., 1993; Escarmis et al., 1996; Elena et al., 1996). This has important implications regarding transmission events and in vivo compartmentalisation (Duarte et al., 1992).

High mutation rates increase the susceptibility of small populations to Muller’s ratchet due to a higher proportion of mutations leading to reduced fitness. Eventually, a point will be reached where replication is so inaccurate as to consistently lose fitness (Eigen and Biebricher, 1988). This ‘error catastrophe’ hypothesis has been demonstrated experimentally using chemical mutagens. When ribavirin, a compound used to treat hepatitis C virus infection, was applied to poliovirus cultures, the error rate of poliovirus replication was raised nearly 10-fold. A concomitant 99% loss of infectivity sent the populations into ‘error catastrophe’ and extinction (Crotty et al., 2001).
1.6.2. Sources of Genetic Variation in HIV-1

With the exception of genomic re-assortment, HIV-1 displays all the features that contribute to RNA virus variation (Duarte et al., 1994a; Domingo and Holland, 1997). Non-viral mechanisms may also increase the genetic diversity of HIV-1. Potentially, errors may be made during transcription of integrated HIV-1 genomes by RNA polymerase II. Measuring the accuracy of RNA polymerases is much more difficult than for DNA polymerases, and the fidelity of this enzyme is unknown. Error rates of similar RNA polymerases in *E. coli* and wheat-germ are estimated *in vitro* at $1 \times 10^{-4}$ per nucleotide. However, *in vitro* estimates of polymerase fidelities can be very different to those observed *in vivo* (Mansky and Temin, 1995). The extent by which transcription of the integrated genome introduces HIV-1 variation remains uncertain (Coffin, 1996b).

Recombination and high mutation rates during reverse transcription give rise to the majority of HIV-1 genetic variation (Coffin, 1995). Some recombinants arise through minus-strand strong stop DNA being accepted by the co-packaged heterologous genome during the first strand-transfer event (Panganiban and Fiore, 1988; Hu and Temin, 1990b). Errors during reverse transcription give rise to almost all the remaining variation. These are favoured by the processivity and fidelity of reverse transcriptase – two enzymatic parameters relating to the behaviour of polymerases. Processivity is quantified as the ratio between the rate of nucleotide incorporation ($k_{\text{pol}}$), and the rate of enzyme dissociation from its template ($k_{\text{off}}$), returning an average number of incorporations before enzyme-substrate dissociation ($k_{\text{D}}$) (Huber et al., 1989). When incorporating the first few nucleotides at initiation, RT behaves 'distributively', separating from the template after each nucleotide incorporation. However, once the elongation phase begins, RT becomes 'quasi-processive', such that it tends to remain attached to the template after each nucleotide incorporation. A truly processive polymerase would remain associated throughout, and complete replication of the entire template within a single cycle (Lanchy et al., 1996a, 1996b). RT processivity can fluctuate depending upon whether the template is RNA or DNA. Furthermore, regions of the genome with high levels of secondary structure are often associated with
'pausing', where \( k_{pol} \) values can be lowered by up to several orders of magnitude (Huber et al., 1989; Klasens et al., 1999). This quasi-processive property of RT makes inter-genomic recombination possible (Luo and Taylor, 1990; Stuhlmann and Berg, 1992). Recombination occurs with high frequency \textit{in vitro} (Kellam and Larder, 1995; Kuwata et al., 1997), and analysis of HIV-1 co-infections has provided strong evidence of \textit{in vivo} recombination (Zhu et al., 1995; Fultz et al., 1997).

The second enzymatic parameter – fidelity – is defined as the frequency by which a polymerase makes errors whilst copying a template strand. Due to the lack of a 3'-5' exonuclease ‘proof-reading’ ability, HIV-1 reverse transcriptases exhibit high error rates (Coffin, 1996b). The estimated error rate for HIV-1 RT is \( 3.4 \times 10^{-5} \) errors per nucleotide per round of replication (Mansky and Temin, 1995), and varies both between RNA and DNA templates and between sequences (Kerr and Anderson, 1997). Many studies have revealed how mutations within RT can increase fidelity, reduce fidelity, or alter the balance between types of error (Kunkel and Bebenek, 2000). Moreover, fidelity is not a property solely of RT – mutations in Vpr can significantly affect the RT error rate, although the mechanism by which this occurs remains unclear (Mansky, 1996).

Base substitutions comprise the majority of errors, with transitions more common than transversions (Mansky and Temin, 1995). Intracellular nucleotide pool imbalances influence the mutational bias of RT, causing purine transitions to be more common than pyrimidine transitions (Vartanian et al., 1997). This bias often causes ‘hypermutation’ – extensive G to A transitions throughout large stretches of a template (Vartanian et al., 1991). Insertion and deletion errors (frameshifts) are not uncommon, and frequently occur in homopolymeric sequences (Bebenek et al., 1993).

**1.6.3. The HIV-1 Quasispecies**

Numerous observations have demonstrated that HIV-1 populations display the evolutionary behaviours characteristic of quasispecies. Diverse, but related genotypes were found by cross-sectional sequence analysis of virus populations in infected individuals (Goodenow et al., 1989), and their continually evolving nature was
established shortly after (Rooke et al., 1991; Cichutek et al., 1992; Delwart et al., 1994). In vitro experimental populations of HIV-1 easily succumb to Muller’s ratchet (Yuste et al., 1999; Yuste et al., 2000). Genetic bottlenecks caused by antiretroviral therapy lead to stochastic evolution, reduced fitness and less diverse quasispecies (Nijhuis et al., 1998; Berkhout, 1999; Ibanez et al., 2000). Reducing the replication rate of HIV-1 through antiretroviral therapy, and consequently its population size, would constrain viral evolution by a deterministic model. However, under the quasispecies model, resistant viruses can swiftly emerge due to outgrowth of pre-existing resistant genotypes (Najera et al., 1994, 1995; Colgrove and Japour, 1999).

Mathematically predictable fitness landscapes such as the simple model described in section 1.6.1 do not exist in practice. Sewell Wright (1931) envisaged that under a given set of environmental conditions, sequence space would map onto a rugged landscape, where multiple local peaks of high fitness were connected to multiple troughs of low fitness by slopes of varying steepness. Stochastic evolution coupled with high mutation rates can allow even relatively small populations to ‘explore’ a wide array of fitness peaks, and possibly achieve distant peaks with superior fitness maxima than local peak (Red’ko, 1998).

HIV-1 populations are in a state of constant evolution, due to ever-changing selection pressures from the host immune system and therapeutic interventions (Wolinsky et al., 1996; Coffin, 1996a). This, in association with the long-lived subset of infected T cells allows the persistence of memory genomes. Memory genomes are elements within a quasispecies representing ancestral sequences adapted to previously encountered fitness topologies (Ruiz-Jarabo et al., 2000; Arias et al., 2001). Pre-treatment drug sensitive HIV-1 variants emerge after the cessation of drug therapy, despite an intervening period of substantial selective disadvantage (Lukashov et al., 2001).
1.6.4. Effects of Drug Resistance on the Evolution of HIV-1

Drug resistance can affect both the rate and the direction of HIV-1 evolution, either by altering the enzymatic parameters of RT governing genetic variation, or by modulating the selection pressures that govern HIV-1 population dynamics. As discussed in section 1.6.2, fidelity and processivity are the dominant parameters of RT involved in maintaining genetic variation. Altering fidelity can change either the rate at which errors are made or the types of error that are made, and changes to RT processivity will influence the frequency of recombination due to template switching.

Broad homology exists between the active sites of retroviral reverse transcriptases (RTs). However, possibly as an evolutionary mechanism, lentiviral RT enzymes have a lower fidelity than most (Coffin, 1992). Many mutations conferring resistance to nucleoside analogues result in RT being able to discriminate between the drug molecule and the natural nucleotide substrates (Martin et al., 1993; Tantillo et al., 1994; Shah et al., 2000). Consequently, one or more amino acid changes selected for by drug pressure may increase the ability of RT to recognise misincorporated nucleotides during reverse transcription, and thus bring about significant changes in fidelity. Several groups have reported altered RT fidelity due to resistance mutations (Rezende et al., 1998; Shah et al., 2000), but it has been the 3-TC-resistance mutation M184V that has stimulated most investigation. Originally, Wainberg et al. (1996) proposed that the acquisition of M184V inhibited HIV-1 evolution, based on the following observations:

i.) Patients harbouring 3TC-resistant virus had viral burdens consistently below pre-treatment levels, unlike viruses without codon 184 mutations.

ii.) In culture experiments, isolates from these patients failed subsequently to develop resistance to either d4T, AZT, nevirapine, delavirdine, or saquinavir.

iii.) Enzymatic studies suggested 184V-containing RT had 5- to 10-fold greater fidelity than 184M-containing RT.
Neutralizing antibody titres remained stable in 3TC-resistant patients, compared to sharp declines in untreated individuals, suggesting epitopes were not evolving in these patients.

Subsequent experiments initially supported this proposal, but further inquiry led to alternative explanations for the data. Firstly, in vitro experiments measuring the fidelity of purified RT on both non-HIV and HIV-2-derived RNA templates demonstrated that M184V both increased RT fidelity and altered the distribution of the types of polymerase errors (Oude Essink et al., 1997; Drosopoulos and Prasad, 1998). However, as Mansky and Temin demonstrated (1995), in vitro measures of fidelity are unlikely to reflect accurately the in vivo situation.

Secondly, M184V-containing viruses were found to develop resistance to AZT in vivo at a slower rate and through a larger number of mutations than for viruses lacking this mutation. However, this was believed to be due to mutual antagonism between M184V and the AZT resistance mutations (Nijhuis et al., 1997b; Kemp et al., 1998). Furthermore, Keulen et al. (1997b) were unable to replicate the delay in subsequent resistance acquisition in vitro, and it was asserted that any noticeable evolutionary effect would require vastly greater fidelity increases than those observed.

As with fidelity, it is not surprising that resistance mutations around the active site can affect RT processivity, and possibly viral evolution. For example, whilst K65R confers partial resistance to ddi, ddC, 3-TC, and the acyclic nucleotides, it has an additional effect of increasing processivity (Arion et al., 1996). Again, RT possessing multiple AZT mutations exhibits increased processivity, but also increased levels of misincorporated nucleotide removal (Caliendo et al., 1996; Boyer et al., 2001). Conversely, M184V reduces processivity, an effect magnified under limiting nucleotide concentrations (Back et al., 1996; Back and Berkhout, 1997), as does the adefovir-resistance mutation K70E (Miller et al., 1998). Whether these processivity changes affect the frequency of genomic recombination by template switching is unclear. Some RT mutations selected by drug therapy effect changes in RNAse H function, but again, whether they can influence recombination frequencies is unknown (Fan et al., 1996; Archer et al., 2000).
The majority of drug resistance mutations are largely absent from HIV-1 isolated from treatment-naïve individuals. This strongly suggests that they decrease fitness (i.e. confer replicative disadvantage), and are selected against in drug-free conditions. Two approaches to analyse the effects on fitness of resistance mutations have been employed. In the first, *in vivo* populations are monitored over time, and the rates at which mutant genotypes change in relation to wild-type sequences are measured. For example, the majority of viruses belonging to the post-transmission quasispecies of an infected but untreated individual harboured the AZT-resistance mutations M41L and T215Y. Over time, the tyrosine at codon 215 was gradually replaced by amino acids conferring sensitivity. The rate of replacement was used to infer the fitness differential between resistant and sensitive genotypes (Goudsmit et al., 1996). Conversely, the M41L mutation persisted and no loss of fitness could be ascribed to it (Goudsmit et al., 1997). Frost et al. used similar *in vivo* population dynamic methods to quantify the relative fitnesses of different amino acid identities at codon 184 of RT (Frost et al., 2000).

The second, most frequent means of fitness analysis, involves the generation of recombinant viruses containing the domain of interest in a constant (‘isogenic’) viral genetic background. Recombinant viruses are analysed either by experiments to assay single-round replication kinetics, or by pair-wise competition experiments. Fitness differences can be ascribed solely to differences between the inserted domains. Recombinant virus techniques have also been used to analyse many partially resistant genotypes intermediate between wild type and maximally resistant. The results help to explain the mutational patterns by which resistance is acquired (Back et al., 1996; Maeda et al., 1998; Jeeninga et al., 2001). Similarly, selective disadvantage has been shown to accompany many resistance mutations (Croteau et al., 1997; Gerondelis et al., 1999; Martinez Picado et al., 2000), including the M41L mutation described above (Harrigan et al., 1998; Keulen, 2000). However, these fitness data must be interpreted carefully. As with phenotypic drug resistance assays, the effects on fitness of specific mutations can be influenced by the genetic background of the virus – a possible explanation for post-therapy persistence of M41L (Dykes et al., 2001).
The initial fitness costs incurred by the selection of otherwise deleterious mutations can be gradually relieved by subsequent compensatory mutation. In addition to being prevalent in bacterial systems (Levin et al., 2000), many RNA viruses, including HIV-1, are amenable to compensatory evolution (Domingo and Holland, 1997). Examples can be drawn from almost all regions of the HIV-1 genome (Wang et al., 1996; Ono et al., 1997; Verhoef and Berkhout, 1999) including even apparently conserved sequences such as the TAR and the LTRs (Klaver and Berkhout, 1994; Liang et al., 2000).

In relation to drug resistance, compensatory mutation is readily observed. The genotypic changes that develop during the course of protease inhibitor therapy are commonly divided into primary and secondary mutations, based upon order of their appearance and their role in phenotypic resistance. Primary mutations arise first, and reduce the sensitivity of protease to a given drug. Subsequently acquired secondary mutations are dependent upon the existing profile of primary resistance mutations, and their effects are often not restricted to drug sensitivity. Their selection suggests they enhance virus fitness. Many restore functionality to enzymes compromised by primary resistance mutations and can be classed as compensatory mutations (Schock et al., 1996; Martinez Picado et al., 1999).

Changes in the Gag cleavage site substrates of HIV-1 protease have been documented in PI-treated isolates (Doyon et al., 1996). These can also be described as compensatory mutations as they offset reductions in substrate binding affinities caused by primary mutations within protease (Robinson et al., 2000; Cote et al., 2001). It now seems that non-cleavage site mutations within Gag are also selected for by PI treatment and play a role in resistance to PIs (Gatanaga et al., 2002).

Compensatory mutations are less easily observed in HIV-1 RT, although certain resistance mutations can be termed compensatory mutations. Examples include the four substitutions that follow the Q151M multinucleoside resistance mutation (Maeda et al., 1998), and the mutation at codon 333 that facilitates AZT and 3-TC co-resistance (Kemp et al., 1998). In the rare cases of drug resistant RT bearing multiple mutations along the template-primer binding cleft, it is likely that some are compensatory (Larder et al., 1999; Imamichi et al., 2000). Evidence of compensatory mutation suggesting that
RT has significant genetic flexibility comes from in vitro studies, where greater extremes of fitness than would be tolerated in vivo can be induced and observed (Olivares et al., 1999; Pelemans et al., 2001).

These observations reflect the complex topology of fitness landscapes. Periods of strong selective pressure can deform a landscape such that previously unfit variants can out-compete the dominant sequences of the pre-selection quasispecies. The landscape topology reverts upon removal of the selection pressure, but the new quasispecies may be distributed around a new fitness maximum. It is conceivable that this peak may be sufficiently distant from the original that only the distortion caused by the intervening period of selection enabled its attainment. Clearly, the fitness of the post-selection local maximum may be greater or lesser than that of the pre-selection peak. Evidence from bacteria demonstrates this ‘bridge’ concept. Streptomycin-resistant E. coli can acquire compensatory mutations such that reversion to a sensitive genotype at the resistance-conferring locus itself becomes disadvantageous (Schrag et al., 1997). Compensatory mutations within resistant protease can increase the fitness of post-treatment virus to a level greater than pre-treatment isolates demonstrated how a post-selection fitness peak could be higher than the pre-treatment peak (Nijhuis et al., 1999).

Transmission of drug-resistant virus was first reported in 1993, and has become increasingly common in recent years with the advent of widespread combination therapy (Erice et al., 1993; Yerly et al., 1999). This reduces the options for first-line treatment and necessitates the expense of widespread genotypic resistance testing. Furthermore, an increasing number of reports are emerging of viruses with potentially enhanced fitness and resistance profiles (de Ronde et al., 2001; Deeks, 2001). These raise public health issues regarding the HIV-1 epidemic. For example, as the risk of transmission increases with viral burden, viruses both drug resistant and fit may increase in prevalence. Additionally, controlling individual virus burdens will become more difficult, with consequent increases in transmission rates. Monitoring viral fitness should be considered as equally important as resistance analysis, although this may prove difficult in practice (Wainberg and Friedland, 1998).
1.7. Outline of the Thesis

This project was stimulated by observations made during the DELTA Trial, a randomised, double-blind clinical trial studying 3,207 subjects with low CD4\(^+\) cell counts (<350 mm\(^{-3}\)) taking one of three nucleoside analogue therapies – either AZT alone, AZT and ddI, or AZT and ddC (DELTA Coordinating Committee, 1996). A 240-strong treatment-naïve subset of these was selected for more extensive clinical and virologic analysis (Brun-Vezinet et al., 1997). The responses of the majority of subjects in this subset taking dual therapy conformed to one of two patterns – either sustained benefit was conferred, with lack of genotypic or phenotypic resistance to the non-AZT component of the therapy, or viral loads and CD4\(^+\) T cell counts returned to their baseline levels, with concomitant acquisition of resistance to both analogues. However, a minority of subjects exhibited a ‘delayed’ response, with the return of clinical parameters to baseline levels deferred by up to 48 weeks, without detectable genotypic or phenotypic resistance to the non-AZT component of their therapy.

It was postulated that genotypic changes in the reverse transcriptase (RT) domain secondary to the established resistance mutations had compensated for a reduced replicative capacity induced both by RT-compromising resistance mutations and by the continued effect of the non-AZT nucleoside analogue. To this end, it was decided to devise a system to study RT domains in vitro, such that phenotypic observations could be ascribed to each RT genotype. Of particular interest was comparing the relative fitness parameters of RT domains. This necessitated the comparative study of RT domains within a constant viral background. This section describes the development of recombinant virus techniques in which sample-derived reverse transcriptase (RT) domains could be inserted into a constant HIV-1 genetic background. Four strategies were successively attempted, with each being informed by the difficulties encountered during the previous ones (sections 2 to 5).

The second component of this thesis describes the development and application of methods used in the analysis of the relative fitnesses of recombinant viruses. It was decided that downstream pairwise competition experiments between viruses would be
monitored by a quantitative point mutation assay (qPMA) – a technique which measures the relative frequencies of each nucleotide at a specified locus within a known sequence (Kaye et al., 1992). The original qPMA method was improved upon by correcting several deficiencies inherent in the original design. Much of this development was performed by Dr. S. Kaye and Ms. J. Bennett in our laboratories, but it remained to be applied to the monitoring of virus cultures as used here.

During the course of the project, the original DELTA samples were replaced by a set of longitudinal samples obtained from a patient recruited into the AZT Resistance Study (ARS). Eleven HIV-positive subjects with Category C disease were recruited into this study, which investigated the effects of AZT monotherapy (500-1000mg per day) in patients with AIDS (Loveday et al., 1995).

Of particular interest was an individual (A3) whose response to therapy was unusual. At each of five timepoints throughout the study (days 0, 7, 77, 125, and 192), 20ml blood samples were taken, separated, and analysed for RNA viral load. A quantitative point mutation assay (Kaye et al., 1992) was used to analyse mutation frequencies at the known AZT resistance codons 41, 67, 70, 215, and 219 of RT. Interestingly, although AZT therapy was discontinued after 77 days, resistance mutations at codons 41 and 215 not only persisted until day 192, but also increased in prevalence (Loveday et al., 1995; Kaye, 1997).

Initially, a pairwise competition system was employed to confirm the results of previous work, where the relative fitness of HIV-1hxb2 was reduced by the introduction of the AZT resistance mutations M41L and T215Y into the RT domain (Harrigan et al., 1998; Keulen, 2000). Secondly, it was thought that pairwise competition experiments between recombinant viruses containing sequenced RT domains derived from each timepoint would reveal whether changes in the RT domain may have been responsible for fitness increases compensating for the potentially debilitating resistance mutations.
2. Design and Application of the pHIVEC Vector System

2.1. Introduction

The recombinant virus assay (RVA) devised by Kellam and Larder (1994) was adapted by workers at the University Hospital, Utrecht (Boucher et al., 1996). In this system, HIV-1 reverse transcriptase (RT) domains are amplified by nested RT-PCR from sample-derived viral RNA. The product of the second-round PCR step overlaps the termini of a linearised plasmid containing the full sequence of the HXB2 laboratory strain of HIV-1 but lacking the RT domain. The PCR product and linearised plasmid are co-transfected into SupT-1 cells, and through intracellular homologous recombination, the two molecules recombine to generate contiguous, full-length, and viable virus (FIGURE 2-1). This system is still widely used to analyse phenotypic drug resistance profiles of sample Pol domains (Nijhuis et al., 1997a; Hertogs et al., 1998; Garcia Lerma and Heneine, 2001).

However, RVA-based systems incorporate undefined amounts of sample sequences flanking the inserted domain (Boucher et al., 1996; Nijhuis et al., 1997a; Hertogs et al., 1998). These are likely to affect significantly the relative fitnesses of the recombinant viruses. Consequently, vector-based systems have been developed to avoid this issue. When inserts are cloned into vectors via restriction sites, the problem of undefined sample sequences being present in recombinant viruses is precluded. However, some existing vectors utilise pre-existing restriction sites in laboratory strains of HIV-1 for the transfer of Pol domains (Walter et al., 1999). Here, variable, but definable sample-derived sequences flanking the domains of interest still become incorporated into recombinant viruses. Protease sequences in particular are likely to affect virus fitness, and so these techniques present little improvement upon the RVA.
FIGURE 2-1: *The principle of homologous recombination. The shaded blocks represent sequences where the derivation (either from the vector or from the sample virus) is undefinable.*
Other vector systems attempt to incorporate only the RT domain of sample viruses. For example, a vector has been designed by Shi & Mellors (1997) in which unique XmaI and XbaI restriction sites were introduced into the RT domain of HIV-1_LAI. However, this vector permitted the cloning only of amino acids 16-490 of the 560 amino acid RT domain. A second cloning vector developed by Martinez-Picado et al. (Martinez-Picado et al., 1999) during the development of our vector improved upon Shi & Mellors' by incorporating an additional 35 C-terminal amino acids of RT into the background vector. These still fail to incorporate the extreme 3' end of RT, ignoring any contributions to RT fitness made by 3'-terminal sequences such as the RNAse H domain (Tisdale et al., 1991; Goobar-Larsson et al., 1993). This chapter describes a method devised to ensure that as much sample reverse transcriptase (RT) domain as possible was incorporated into each recombinant virus, without incorporating undefined flanking sequences.

Our vector was generated through modifications to pHXB2ΔRT – the RT-deletion clone used by the Utrecht group (Boucher et al., 1996, Nijhuis et al. 1997a). Two unique restriction sites were introduced flanking the RT deletion. The second-round PCR primers used to amplify RT domains were designed to introduce restriction sites allowing cross-ligation into this new vector. Our vector improves on all those described above by permitting the insertion of amino acids 7 through 548 of sample RT domains. The first part of this chapter describes the construction of our vector (pHIVEC) from pHXB2ΔRT, as the restriction sites in pHXB2ΔRT were unsuitable for cloning RT. The remainder of the chapter (2.3 to 2.5) describes our attempts to utilise this system for the generation of recombinant viruses. The system is outlined in schematic form in FIGURE 2-2.
FIGURE 2-2: Schematic outline of the pHIVEC vector system for the generation of recombinant HIV-1.
2.2. Construction of pHIVEC

The plasmid pHXB2(reverse)-RT, kindly donated by Wilco Keulen at University Hospital Utrecht, was originally derived from the pHIVARTTB(reverse)-stEI molecular clone of HIV-1 available from NIBSC, Potters Bar, UK (Kellam et al., 1992; Kellam and Larder, 1994). This parent vector was generated by cloning the entire proviral sequence of the HXB2 laboratory strain of HIV-1 into a commercially available vector. Bases 121 to 1566 of RT were deleted, and replaced by a BstEI restriction site. Boucher et al. (1996) modified pHIVARTTB(reverse)-stEI by replacing its partially-deleted RT domain with one lacking all 560 amino acids, with flanking Stai and XbaI sites at the 5' and 3' ends respectively. The presence of a second XbaI site within the human genomic sequences flanking the HIV-1 sequences precluded its use in an RT cloning site, and using the Stai site would introduce a glycine at the second codon of RT. The structure of pHXB2(reverse)-ART is shown in FIGURE 2-3.

An algorithm within the Vector NTI software package (InforMax Inc., MD) will search a specified nucleic acid sequence for potential silent restriction sites. This was applied to the sequence comprising the junction between the RT and integrase domains of HXB2, using a list of 18 restriction endonucleases that cut neither pHXB2(reverse)-ART nor HIV-1 RT, and cleave to give sticky ends. This way, a potential SacII site was discovered (FIGURE 2-4). However, when applied to the junction between HXB2 protease and RT, Vector NTI was unable to recommend any sites. In order to enable ligation at the 5' end of RT, non-identical restriction endonucleases producing complementary sticky ends upon cleavage were chosen to cleave the vector and the PCR-amplified RT domain. Two lists of candidate restriction endonucleases were assembled – those not cleaving pHXB2(reverse)-ART, and those not cleaving RT (102 candidates, not listed). Manual inspection of the protease-RT junction sequence revealed a design in which the vector would possess a BstBI site, and the RT domain a NarI site (FIGURE 2-5).
FIGURE 2-3: A schematic representation of plasmid pHXB2ART (~12.5kb), from which pHIVEC was derived.
FIGURE 2-4: The nucleotide sequence of the 3' terminus of HXB2 reverse transcriptase, showing the location of a potential silent SacII discovered using Vector NTI software.

FIGURE 2-5: The two-site method for ligation of PCR amplified RT inserts and the pHIVEC vector at the protease-RT junction.
Four oligonucleotides were designed specifically for the construction of pHIVEC. As they were not used elsewhere in this project, they are described below in TABLE 2-1 and FIGURE 2-6. The antisense primer Mut-aSense binds HXB2 at the 3' end of integrase, whereas the sense primer Sac-Site binds pHXB2ART across the pre-existing RT deletion and contains mismatches allowing the introduction of a SacII site. These two primers were used to amplify a 914bp product from pHXB2ART. Both this PCR product and pHXB2ART were digested using XmaI and MscI (FIGURE 2-7). The 363bp XmaI-MscI fragment of the PCR product was excised and purified using QIAquick Gel Extraction Kit.

The vector backbone molecule and the digested PCR product fragment were ligated together and transformed into SURE2 E. coli. Only a single colony was generated. This was shown to be positive for the insert by PCR screening with primers 706M and Mut-aSense (FIGURE 2-8). Subsequent SacII digestion of the PCR product verified successful mutagenesis (FIGURE 2-9). The colony was propagated in a 3ml culture and plasmid DNA extracted. This was subjected to sequential restriction digestion with Smal and SacII (2 hours at 25°C followed by 2 hours at 37°C) and fluorometric yield analysis (section A.12). Oligonucleotides AD-1 and AD-2 were mixed at an equimolar ratio, heated to 70°C for 2 minutes, and allowed to cross-anneal by cooling to 25°C at a rate of 1°C/min. T4 DNA ligase buffer (to a final ATP concentration of 1mM) and 1U T4 Polynucleotide kinase were added, and the reaction incubated at 37°C for 30 minutes.
TABLE 2-1: Oligonucleotides used specifically for the construction of pHIVEC.

<table>
<thead>
<tr>
<th>Mut-αSense</th>
<th>TAA TCC TCA TCC TGT CTA CTT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sac-Site</td>
<td>AAT TTT CCC GGG TCG AAA TTA GTC TCC GCG GGA ATC AGG</td>
</tr>
<tr>
<td>AD-1</td>
<td>ATT TCG AAT TTG CCC GGG GTC TCC GC</td>
</tr>
<tr>
<td>AD-2</td>
<td>GGA GAC CCC GGG CAA ATT CGA AAT</td>
</tr>
</tbody>
</table>

FIGURE 2-6: Schematic diagram showing how the oligonucleotides in TABLE 2-1 align with the Pol domain of pHXB2-RT. The restriction sites introduced by these primers are also shown.
FIGURE 2-7: Digests of pHXB2ΔART. Lane 1 – pHXB2ΔART (uncut). Lane 2 – pHXB2ΔART digested with MscI. Lane 3 – pHXB2ΔART digested with XmaI. Lane 4 – pHXB2ΔART digested with both MscI and XmaI. Lane 5 – PCR product from pHXB2ΔART using primers Sac-Site and Mut-aSense. Lane 6 – PCR product from Lane 5 digested with both MscI and XmaI.

FIGURE 2-8: PCR amplification using primers 706M and Mut-aSense. Lanes 1-3 respectively – pMCS79, pHXB2ΔART, and a colony obtained from ligation between the bands purified from FIGURE 2-7.

FIGURE 2-9: Lanes 1 and 2 – SacII digestion of the amplicons shown in Lanes 2 and 3 of FIGURE 2-8 respectively.
In 1X T4 DNA Ligase buffer containing 4U/reaction T4 DNA Ligase, annealed oligonucleotide adaptor and digested pHXB2ΔRT/SacII were mixed in a series of ratios from 1:1 to 10⁻⁶:1 in 10⁻⁵:1 steps. After overnight incubation at 12°C, these ligations were transformed into SURE2 E. coli, and yielded approximately 50 colonies each. Primers 707M and Mut-αSense were used to PCR screen 43 colonies selected from across all plates. Seventeen colonies generated products, which were subjected to further screening by BstBI digestion, of which four gave correct banding patterns (a 1,050bp PCR product cleaved by BstBI into two fragments of 907bp and 143bp – see FIGURE 2-10). Each of the four was cultured in 3ml volumes and plasmid DNA extracted. Automated DNA sequencing with the antisense RT primer 369L showed all to possess the correct sequence. One was selected to be pHVEC and propagated in 3ml culture from which a glycerol stock was prepared.

FIGURE 2-10: BstBI digestion of amplicons from 17 colonies obtained by transforming ligations between SacII-SmaI digested pHXB2ΔRT SacII and the products of the annealing of oligonucleotides AD-1 and AD-2. The expected amplicon of 1,050bp should cleave with BstBI to generate fragments of 907bp and 143bp. The arrows indicate the four products yielding correct amplification and digestion products. Lane 18 – amplicon from unmodified pHXB2ΔRT SacII subjected to BstBI digestion.
2.3. Methods

A second-round sense PCR primer ('Inner-1') was designed to introduce \( \text{NarI} \) sites at the 5' end of amplified RT domains. Two antisense primers were designed (Inner-2a and Inner-2b), both of which introduced \( \text{SacII} \) sites at the 3' end of amplified RT domains, but differing at a single (silent) nucleotide position within RT to enable future monitoring of paired recombinants by quantitative point mutation assay (TABLE 2-2).

Plasmid DNA was prepared from 50ml cultures seeded by 100\( \mu \)l of pHIVBC glycerol stock as described in section A.8.2 and digested with \( \text{SacII} \) and \( \text{BsrBI} \). 150\( \mu \)g of plasmid were digested in NEB buffer #4 containing 100\( \mu \)g/ml BSA and 150U of \( \text{SacII} \) for 24 hours at 37\(^\circ\)C. 150U of \( \text{BsrBI} \) were added to the reaction and the incubation temperature raised to 65\(^\circ\)C for a further 24 hours. The digest was purified by the phenol-chloroform method detailed in section A.8.3.

RNA was extracted from human plasma using the Boom protocol (section A.8.4). 5\( \mu \)l were entered into 20\( \mu \)l Superscript II reverse transcription reactions (section A.6), using 709M as the antisense primer. 2\( \mu \)l of cDNA were added to 20\( \mu \)l AmpliTaq PCRs. The primers used were 706M and 708M (sections A.2 and A.5.1), and the thermal cycling conditions were as follows: 35 cycles of 94\(^\circ\)C, 60s; 52\(^\circ\)C, 60s; 72\(^\circ\)C, 210s; followed by a single 7-minute extension reaction at 72\(^\circ\)C.

Firstly, anneal temperatures of each of the two second-round PCRs were optimised. First-round PCR product from a sample with a low RNA copy number was used as template in a second-round AmpliTaq PCR replicated across a range of anneal temperatures (from 42\(^\circ\)C to 56\(^\circ\)C in 2\(^\circ\)C steps, Stratagene Robocycler). The cycling parameters were as follows: 40 cycles of 94\(^\circ\)C, 60s, 42-56\(^\circ\)C, 60s, 72\(^\circ\)C, 150s, followed by a 7-minute extension reaction at 72\(^\circ\)C. The relative yield and purity of each product was evaluated by gel electrophoresis.

2\( \mu \)l of first round PCR products were added to 50\( \mu \)l second-round AmpliTaq PCR reactions containing primers Inner-1 and either Inner-2a or Inner-2b. Cycling conditions
were: 35 cycles of 94°C, 60s; 42°C (Inner-2b) or 48°C (Inner-2a – see page 87 and FIGURE 2-13, below), 60s; 72°C, 150s; followed by a single 7-minute extension at 72°C. Reactions were subjected to agarose gel electrophoresis and gel-purified as per section A.8.1. 20µl restriction endonuclease digests were performed on 15µl of each 30µl eluate, in 1X NEB #4, containing 100µg/ml BSA, and 5U each of NarI and SacII. They were incubated at 37°C for 2 hours followed by 20 minutes at 65°C.

Ligation reactions between digested vector and RT inserts were performed as described in section A.11, and subsequently screened using primers A588 and 470W (section A.2 – A588 lies within the Gag region of pHIVEC, 470W is an antisense primer in RT). Ligation were transformed into SURE2 E. coli, both onto plates and directly into 3ml cultures. Plasmid DNA was extracted from the 3ml cultures. DNA concentrations of purified plasmids were quantified by fluorometry (section A.12). Colonies and plasmid preparations were screened by PCR using primers 706M and 470W (TABLE A-1).

Transfection of putative full-length HIV-1 plasmids was as described in section A.15 with the sole difference of just 1µg of plasmid DNA being transfected rather than 2µg of each species as employed previously, due to transfection efficiencies of circular DNA molecules being higher than those obtained with linear DNA molecules. As in the previous chapter, cultures were monitored both by light microscopy and by p24 antigen assay.
TABLE 2-2: Second-round PCR primers unique to this section. Inner-2a and Inner-2b differed at a single nucleotide position (boxed).

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inner-1</td>
<td>AAT TTT CCC ATG GCG CCT ATT GAG</td>
</tr>
<tr>
<td>Inner-2a</td>
<td>CCT GAT TCC CGC GGA GAC TAA TTT ATC</td>
</tr>
<tr>
<td>Inner-2b</td>
<td>CCT GAT TCC CGC GGA GAC TAA TTT ATC</td>
</tr>
</tbody>
</table>

FIGURE 2-11: Alignment of the 'Inner' PCR primer set on HIV-1 Pol domains (upper set – protease-RT junction, lower set – 3' end of RT). Restriction sites introduced by the primers are highlighted in bold type.
2.4. Results

To verify that the two restriction enzymes used to digest pHIVEC were working efficiently, two digests were set up – one with $BstBI$ and $BsaBI$ ($BsaBI$ has a single site in Env), and the other with $SacII$ and $SpeI$ ($SpeI$ has a single site in Gag). The electrophoresis band pattern of each was as predicted (10,552bp and 3,371bp, and 12,852bp and 1,071bp respectively, FIGURE 2-12). BioMax analysis of the gel image indicated close to complete digestion.

![FIGURE 2-12: Double restriction digests of pHIVEC. Lane 1 – digestion with both BstBI and BsaBI. Lane 2 – digestion with both SacII and SpeI.]
Second-round PCRs using both second-round primer pairs were performed with a gradient of anneal temperatures in order to discover the optimal conditions for this step of the PCR (FIGURE 2-13). The pair Inner-1 and Inner-2b only amplified with a 42°C anneal temperature, whereas Inner-1 and Inner-2a amplified at temperatures up to 54°C, albeit inefficiently above 50°C. Additionally, in the first two lanes of the upper gel photo, extra bands can be observed whose molecular weight is below that of the desired band. Hence 48°C was chosen as the annealing temperature with the Inner-1/Inner-2a primer set, at which the balance of yield and purity was considered the best.

Using the first round products generated during the course of the experiments detailed in section 2, second-round PCRs were performed using primer set Inner-1 and Inner-2a or Inner-1 and Inner-2b (an example is shown in FIGURE 2-14). These PCRs gave mixed results, with some samples not amplifying and others giving variable yields of product. Those first round products which failed to amplify were screened using primers within RT located 420bp apart. These consistently gave positive results, indicating the presence of template in the failing PCRs (not shown).

Successful second-round PCRs were purified, digested with \textit{NarI} and \textit{SacII}, and further gel-purified as described. Yields were assayed relative to digested pHIVEC. Inserts were ligated into digested pHIVEC and the resultant reactions transformed as described in section A.10. FIGURE 2-15 shows a sample PCR screen of 12 ligations using primers A588 and 470W (1.1 kb expected product size). Between 5 and 50 colonies per transformation were obtained with screening revealing 0-40% ligation efficiency.

The presence of correctly-ligated DNA in the plasmid preparations was confirmed by PCR screening (primers 706M and 470W, FIGURE 2-16). However, FIGURE 2-17 shows a series of typical plasmid preparations, containing multiple DNA species of varying sizes, including that of the correct size, rather than just single species. Nevertheless, the proportion of correctly-sized DNA within each lane was assayed, and after fluorometric DNA quantification, approximate values for the concentration of correctly-sized DNA in each preparation were obtained.
FIGURE 2-13: Results of ‘gradient PCRs’ using primers Inner-1 and either Inner-2a (top picture) or Inner-2b (bottom picture). The temperature of the annealing steps ranged from 42°C to 56°C in 2°C increments.

FIGURE 2-14: Sample second-round PCRs of sample-derived first-round RT-PCR products using primers Inner-1 and Inner-2a (upper picture), and Inner-1 and Inner-2b (lower picture).
FIGURE 2-15: Lanes 1-12 – Sample PCR screen of ligations between pHIVEC and sample-derived RT inserts. Lane 13 – self-ligation (no insert). Lane 14 – negative control (pHlVEC). Lane 15 – positive PCR control (pHXB2). Lane 16 – negative PCR control (water). The band visible in lane 14 corresponds to plasmid template (pHIVEC) carried through the PCR.

FIGURE 2-16: Lanes 1-5 – Sample PCR screen of plasmids obtained from E. coli transformed with pHIVEC-RT insert ligations. Lane 6 – negative PCR control (water). Lane 7 – positive PCR control (pHXB2).
Although the plasmid preparations were shown to contain recombinant DNA with contiguous Gag and Pol regions, upon transfection, all but one failed to generate syncytia and were considered unproductive. Levels of p24 antigen in culture corresponded to the microscopy results. The successfully transfected plasmid contained an RT insert amplified from a modified pHXB2 molecular clone containing engineered AZT-mutations, and Inner-2a as the antisense primer. Cytopathic effect was visible by day 14 and was pronounced by day 19 when it was harvested. Antigen levels within the virus stocks gave a p24 signal of over 10ng/ml.

FIGURE 2-17: Five typical plasmid DNA preparations obtained after transforming ligations between pHIVEC and RT inserts amplified from RNA extracted from the A3 series samples. The molecular weight standards comprise linear DNA whereas plasmid preparations are predominantly composed of supercoiled DNA. Note how the band patterns are inconsistent.
2.5. Discussion

Problems were encountered throughout the development of this method, preventing the successful generation of recombinant viruses. One problem was the low efficiency of the second-round PCR despite low anneal temperatures and demonstrated presence of template in first round PCR products. Reference to FIGURE 2-11 shows the degree of mismatch between the primers and their putative binding sites. Inner-2a has four mismatches within a six-base sequence and Inner-2b five within seven, followed by 12-base ‘clamps’ at their 3’ ends. Inner-1 has four consecutive mismatched bases, with only a 9-base clamp at its 3’ end. The two restriction endonuclease sites introduced by the second-round primer pairs (SacII and NarI) have complementary G/C-rich recognition sequence. Such complementarity could facilitate cross-annealing between the two primers, and the palindromic repeats may give rise to secondary structures within each primer. These phenomena would have been favoured by the low anneal temperatures at which the primer sets were used, and may have contributed to the inefficiencies encountered in the second-round PCRs.

The method described in the next chapter uses redesigned second-round primers with longer clamping regions, at both their 5’ and 3’ ends. However, this compromises the method by specifying further amino acids of RT within the resultant recombinant virus. Inner-1 specified the first six codons of RT and both Inner-2 primers specified the final twelve codons. However, retrospective analysis of the sequence of Inner-1 and its putative binding site within HIV-1 challenges the suitability of Inner-1 as a PCR primer for amplification of RT domains from clade B viruses. Unfortunately this criticism also applies to the corresponding primers used in the following two chapters, and is discussed in section 4.5, page 130.

The efficiency of transfection may have been compromised both by low levels of full-length contiguous viral DNA, and possibly by having elements of the genome present on different DNA species within the transfected material (i.e. in trans rather than cis). It was shown that within the ligations, vector-derived protease and sample-derived RT sequences had ligated correctly. Furthermore, PCR analysis of the DNA obtained after
subsequent transformation of these ligations also demonstrated contiguous protease-RT sequences. However, as FIGURE 2-17 displays, these plasmid preparations contained multiple DNA species. Re-arrangement within *E. coli* of plasmid sequences containing retroviral sequences or direct repeats (e.g. HIV-1 LTRs) has been previously observed (Yamada et al., 1995; Rosin-Arbesfeld et al., 1998). Most *E. coli* used for the propagation of plasmid DNA have their recA DNA recombination systems deleted. However, even in the absence of recA, intra-LTR recombination can still occur. Furthermore, such re-arrangement is dependent upon the nature of the plasmid into which the HIV-1 sequences have been inserted, and is probably mediated through the R regions within the LTRs (Peden, 1992; Bi and Liu, 1996; Yamada et al., 1995; Rosin-Arbesfeld et al., 1998). This allows for the possibility that the contiguous *Pol* sequences were found within plasmids other than the species migrating at the correct size. It was also not certain that the correctly-sized bands represented RT-containing plasmids. The difference in size between pHIVEC and insert-positive vector was only 1.7kb, and differentiation between these species by electrophoresis was not possible, even using 0.7% agarose gels.

Finally, the number of colonies returned after ligation rarely exceeded 50. PCR screening of colonies revealed frequencies of colonies containing insert-positive plasmids in the region of zero to 40%, leading to between zero and 20 colonies per transformation which contained *Pol*-positive plasmid material of any nature. By inference, the *Pol*-positive plasmid DNA extracted from the cultures would also be derived from a similar number of positive ligation molecules. This was not considered adequate, as the final stock of recombinant virus would have harboured at best only 20 different sample-derived RT variants, and hence would not have been considered sufficiently representative of the *in vivo* quasispecies. In the next chapter, this shortcoming was addressed by improving PCR efficiency through longer primers and using a shorter vector.
3. Design and Application of a Shuttle Vector System (I)

3.1. Introduction

The poor performance of the assay described in the previous chapter made it necessary to re-design the cloning strategy. A new method was formulated, combining elements of the two approaches previously described. Firstly, each of the three primers used in the second-round PCR were extended at both ends to improve upon the poor second-round PCR performance. Secondly, the ‘protease-AΔRT-integrase’ sequence from pHIVEC was cloned into a short, commercially-available plasmid vector (pGEM-T Easy Vector). Second-round PCR products were ligated into this shuttle vector via the preserved pHIVEC cloning site.

Ligation products would have been shorter than those obtained using pHIVEC, allowing purification of insert-positive plasmids by gel electrophoresis. By using these separated plasmids as templates in a ‘final’ PCR whose products would span protease, RT and integrase, they could have been co-transfected with linearised deletion clone with homologous recombination used to generate recombinant viruses. Both primers used in this ‘final’ PCR were re-synthesised during the course of this experiment and lengthened.

This method was considered to have several advantages over those described previously. The homologous recombination component was superior to that employed in the RVA-based systems described in 2.1. This was due to the sequences flanking the RT domains being defined as HXB2. This removes the drawback of variable and undefined sample sequences flanking RT being incorporated into the recombinant virus. The homologous recombination system of mammalian cells has been extensively utilised for the manipulation of chromosomal DNA. Evidence suggests that the efficiency of recombination can be proportional to the degree of homology between the
two sequences to be recombined (Deng and Capecchi, 1992). As the overlapping sequences in this system showed 100% identity, recombination efficiency would not be thus inhibited. A schematic outline of the system is presented in FIGURE 3-1.

Being considerably shorter than pHIVEC, the shuttle vector enabled more efficient isolation of RT-containing plasmids. Given that smaller plasmids give higher transformation efficiencies than longer plasmids, this would allow greater representation of sample RT variants in the final recombinant stocks (Hanahan, 1983).
3.2. Construction of the Shuttle Vector

ApaI and EcoRII double restriction digests were performed, both on pHIVEC, and on a construct generated by a colleague, in which an unrelated sequence had been cloned into pGEM-T Easy Vector (FIGURE 3-2). Both the 2,962bp pGEM-T backbone and the 1,006bp Apal-EcoRI fragment of pHIVEC containing the RT cloning site were excised and gel-purified, although a co-migratory 1,095bp EcoRI-EcoRI fragment was likely to have been co-purified with this latter fragment. A 20μl ligation between these two fragments was performed at a 3:1 ratio of insert to vector, together with a self-ligation control. Overnight incubation at 12°C was followed by denaturation at 65°C for 20 minutes. 5U Neol were added to each ligation in order to digest any ligation products containing the 1,095bp fragment. Incubation for 2 hours at 37°C was followed by the addition of 5U BsaBI to linearise any remaining molecules of the original pGEM-T construct whose insert, uniquely amongst the other fragments involved, contained a BsaBI site. The reactions were incubated at 65°C for a further 2 hours.

Finally, the ligations were transformed into TOP10F’ E. coli. Thirty-nine colonies were selected from the insert-positive ligation transformation plate and screened by PCR for the presence of the insert using primers T7 and G5804 (TABLE A-1). Thirty-five were insert-positive. Six were selected and propagated in 3ml cultures from which their plasmid DNA was extracted. Each was subjected to ApaI and EcoRII double restriction digestion, of which five produced the correct banding pattern. One was selected to be the shuttle vector in the following experiments and named pGEM-TΔRT (FIGURE 3-3). Automated sequence analysis using primers T7 and aSp6 confirmed the correct arrangement of the protease-ΔRT-integrase insert.
FIGURE 3-2: Schematic of pGEM-T Easy Vector (from Promega Corporation, Madison, WI).

FIGURE 3-3: Schematic of pGEM-TART (3,965bp). Domains from pGEM-T are coloured blue, and the Pol domains from pHIVEC are coloured red.
3.3. Methods

The 'Inner' set of primers used in the previous chapter for second-round PCR amplification of RT domains were re-synthesised, with extensions at their 5' and 3' ends (TABLE 3-1).

<table>
<thead>
<tr>
<th>Specific primers used in this section.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IN-1</strong></td>
</tr>
<tr>
<td><strong>IN-2a</strong></td>
</tr>
<tr>
<td><strong>IN-2b</strong></td>
</tr>
</tbody>
</table>

FIGURE 3-4: Alignment of the 'IN' primer set on HIV-1 Pol domains (upper set – protease-RT junction, lower set – 3' end of RT). Restriction sites introduced by the primers are highlighted in bold type.
Large-scale preparation of doubly-digested pGEM-TΔRT was as described for pHIVEC (page 83, with the addition that after the BstBI step, Smal was added and the reaction incubated at 25°C overnight prior to phenol-chloroform purification. Amplification and digestion of RT domains was as per the previous section, but with primers IN-1, IN-2a and IN-2b (TABLE 3-1) instead of the 'Inner' primers. Before the ligation of inserts into digested pGEM-TΔRT, each component was quantified by BioMax gel image analysis (section A.13). 20µl ligations were set up at molar ratios of 3:1 between RT domains and pGEM-TΔRT, incubated overnight at 12°C, and transformed into TOP10F’ cells.

Plasmid DNA extracts from transformed bacterial cultures were separated by gel electrophoresis. Bands migrating at a rate consistent with insert-positive plasmid DNA were gel-purified and entered as templates in ‘final’ PCRs using primers 706M and 708M. These primers were later extended and labelled S-HR and AS-HR – see TABLE A-1. Ligations, transformed bacterial colonies and plasmid DNA extracts were all screened by PCR using one or both of two primer sets – T7-469W and 111W-αSp6. PCR products were gel-purified, and quantified by BioMax as above. These were co-transfected into SupT-1 cells together with linearised pHXB2ΔRT. Preparation of pHXB2ΔRT, transfection, and the subsequent maintenance of cultures was as described in section A.15.
3.4. Results and Analysis

Sequential digestion of pGEM-TΔRT by SacII, BstBI, and finally Smal, was monitored by sampling after each restriction endonuclease step and transforming this aliquot into TOP10F'. Over 10,000 colonies were yielded by an equivalent quantity of uncut plasmid. Each successive digestion reduced the counts to >1,000, 117, and 86 respectively.

When amplifying from plasmid templates (e.g. pHXB2) using the second-round PCR as a single round amplification, the efficiency was high, but as observed in the previous chapter, the amplification of second-round PCR products from sample-derived first round PCR product templates was inefficient and highly variable between samples (FIGURE 3-5). Furthermore, many samples failed to amplify with these extended primers (FIGURE 3-6), despite previous confirmation of template presence (page 86).

Ligations of digested pGEM-TΔRT and digested PCR products were performed as described. FIGURE 3-7 shows a typical PCR screen of ligations using both T7-469W and 111W-αSp6 primer pairs. With the exception of two ligations, one with a failed T7-469W amplification, and another with a failed 111W-αSp6 amplification, the ligations were PCR positive across both ligation junctions. FIGURE 3-8 shows a typical electrophoresis profile of DNA extracted from bacterial cultures into which ligations were directly transformed, with bands corresponding to self ligations and/or uncut pGEM-TΔRT, and insert-positive constructs both labelled. Other species were also present in these extracts. These are particularly visible in lanes 5 and 6. The agar plates corresponding to these transformations typically contained between 10 and 100 colonies (greater than that observed with pHIVEC in section 2). The percentage of colonies PCR positive for insert presence was similar to that observed for pHIVEC (generally between 10% and 40%, not shown); hence, the number of sample RT variants represented in the final preparation of each construct would have ranged from 1-40.
FIGURE 3-5: Sample gel showing the high yield variability amongst second-round PCR amplifications of sample-derived RT domains using the primer set IN-1 and IN-2a, compared to the efficiency of amplification from plasmid-derived template. Lane 1 – pHXB2 as template. Lanes 2-8 – sample-derived templates.

FIGURE 3-6: Sample gel showing how some sample-derived material failed to amplify using primers IN-1 and IN-2b. Lanes 1-5 – amplifications of sample-derived material. Lane 6 – Negative PCR control (water). Lane 7 – Positive PCR control (pHXB2).
FIGURE 3-7: PCR screening of ligations between pGEM-TART and RT inserts. Lanes 1-12 – using T7 and 469W as primers. Lanes 13-24 – using 111W and aSp6 as primers. In each row, the order of samples is as follows: eight ligations, self-ligation, negative PCR control (water) and positive PCR control (pHXB2).

FIGURE 3-8: Sample gel photo showing the plasmid DNA species extracted from bacteria transformed with ligations between pGEM-TART and RT inserts.
Bands thought to correspond to insert-positive plasmids were gel-purified and subjected to 'final' PCR amplification using primers 706M and 708M. As the sample gel shown in FIGURE 3-9 shows, these failed to generate product. In an attempt to address this problem, both 706M and 708M were re-synthesised and extended as S-HR and AS-HR respectively (TABLE A-1). However, with one exception (see below), amplification with these primers was also unsuccessful (FIGURE 3-10). As with the ligations, these purified plasmids were shown by PCR screening to contain contiguous protease-RT and RT-integrase sequences (data not shown). The smears seen on these two gels suggest mis-amplification, rather than absence of template.

To investigate the failure to amplify from purified plasmids, one such plasmid was transformed into *E. coli*. DNA was extracted and subjected to restriction mapping using *ApaI* in combination with *SacII*, *EcoRV*, or *EcoRI* (FIGURE 3-11). Both the *ApaI-SacII* and *ApaI-EcoRV* digests produced bands consistent with the predicted insert-positive pGEM-TART construct, and inconsistent with insert-negative pGEM-TART alone. However, the *ApaI* and *EcoRI* double digest failed to produce the correct band pattern of 2,959bp and 2,638bp. Instead, an array of multiple bands of varying intensity and size were observed. This was typical of 'star activity', a phenomenon observed when certain enzymes including *EcoRI* are employed in inappropriate buffer conditions. Under such circumstances, both the first and sixth position of the six-base recognition sequence of *EcoRI* lose their specificity, allowing cuts at non-specific sequences (Polisky et al., 1975).

Referring again to FIGURE 3-11, it can be seen that in addition to the strong bands in Lanes 1-3 corresponding to the correctly ligated pGEM-TART plus sample-derived RT domain, additional bands of higher molecular weight can be observed. The two uppermost bands have migrated at a rate consistent with concatemers of the lowermost band, but the band immediately above the lowermost band cannot be explained. FIGURE 3-12 shows a gel in which plasmid DNA extracted from ligation-transformed *E. coli* was separated overnight by electrophoresis to resolve potentially co-migrating doublet DNA bands. Here, a faint band migrating behind the putative insert-positive plasmid band can be seen. This band was indistinguishable from the 'correct' band after
shorter electrophoresis. It is probable that these second species were co-purified with insert-positive constructs.

FIGURE 3-9: PCR amplification of plasmid DNA migrating at a rate consistent with 5.6kb supercoils (as shown in FIGURE 3-8), using primers 706M and 708M. Lanes 1-4 – purified plasmid DNA as template. Lane 5 – Positive PCR control (pHXB2). Lane 6 – Negative PCR control (water).
FIGURE 3-10: PCR amplification of purified plasmid DNA migrating at a rate consistent with 5.6kb supercoils using primers S-HR and AS-HR. Lanes 1-5 – purified plasmid DNA as template. Lane 5 – Positive PCR control (pHXB2). Lane 6 – Negative PCR control (water).

FIGURE 3-11: Restriction endonuclease analysis of a plasmid obtained after transformation of a ligation between pGEM-TART and a plasmid template-derived RT insert. Lanes 1-3 – uncut supercoiled plasmid. Lane 4 – plasmid digested with ApaI and SacII. Lane 5 – plasmid digested with ApaI and EcoRV. Lane 6 – plasmid digested with ApaI and EcoRI. Note the multiple digestion in products the final lane.
FIGURE 3-12: Plasmid DNA extracts from bacteria transformed with pGEM-TART-RT insert ligation reactions and separated by overnight electrophoresis. RT inserts were obtained from plasmid templates (e.g. pHXB2). The banding patterns are discussed in the text.

FIGURE 3-13: Electrophoresis profiles of plasmid extracts obtained from bacteria transformed with pGEM-TΔRT-RT insert ligation reactions where the RT inserts were obtained from sample material, showing the presence of diverse DNA species smaller in size than pGEM-TΔRT.
Subsequent plasmid DNA extracts from *E. coli* transformed either with ligation reactions or with excised putative insert-positive DNA bands, when separated by electrophoresis, revealed arrays of bands inconsistent with any predicted ligation products (FIGURE 3-13). Excising and transforming putative insert-positive gel-purified plasmids typically resulted in DNA banding patterns similar to those seen in FIGURE 3-14, where correctly-sized bands could not be distinguished, let alone excised or amplified. In these constructs sample material was used as the sources of the RT domains, differing from those used previously where the RT domains were derived from HXB2-containing plasmids.

Using the protocol described in this chapter, only a single recombinant virus was generated, whose RT insert was derived from pHXB2 (pHXB2\(^8\)). No viruses with sample-derived RT domains were generated. Moreover, the template for the ‘final’ PCR was a colony rather than electrophoretically-separated insert-positive pGEM-TΔRT. FIGURE 3-15 illustrates the gel containing the ‘final’ PCRs containing products of the correct size. The product in Lane 4 with multiple bands failed to transfect due to an error during gel-purification leading to its loss – the strength of the correctly-sized band could not be reproduced. The band in Lane 5 was purified and transfected, together with that of the positive control (Lane 7). FIGURE 3-16 shows the development of p24 antigen signal over time with the two recombinant viruses. The similarity between the curves suggests that the silent mutations do not significantly affect replicative capacity, and confirms the efficacy of the transfection technique. In a final experiment confirming that the resultant virus was truly recombinant, rather than ancestral pHXB2 template, RNA from the recombinant virus was extracted, amplified using primers 111W and G5804, and subjected to SacII digestion (FIGURE 3-17).
FIGURE 3-14: Plasmid DNA extracts from bacteria 're-transformed' with gel-purified DNA from putative insert-positive pGEM-TART constructs seen in FIGURE 3-13.

FIGURE 3-15: Lanes 1-5: Amplifications from purified insert-positive pGEM-TART constructs using primers S-HR and AS-HR. Lane 6 – Negative PCR control (water), Lane 7 – Positive PCR control (pHXB2).
FIGURE 3-16: ELISA data (p24 antigen concentration) from the successful transfection of amplified Pol sequences from pGEM-T^HXB2-B and the transfection of Pol sequence amplified directly from pHXB2. The cut-off for p24 detection was 100pg/ml.

FIGURE 3-17: Lane 1 – RT-PCR product using recombinant virus pGEM-TΔRT^HXB2-B RNA as template and primers 111W and G5804. Lane 2 – SacII digestion of RT-PCR product in Lane 1. Lane 3 – identical amplification of recombinant virus generated using the positive PCR control product (HXB2).
3.5. Discussion

Unfortunately, the problems of poor PCR efficiency, plasmid re-arrangement in bacterial culture, and low efficiencies of ligation and transformation persisted through the development of this system. Although the second-round PCR primers were increased in length in an attempt to improve their efficiency, the pattern of non-amplification and variable yields amongst successful amplifications was similar to that observed in the previous chapter. As with Inner-1 in the previous chapter, subsequent analysis of clade B RT sequences called into question the suitability of IN-1 as a PCR primer. Again, as this also applies to the corresponding primer used in the following chapter, it is discussed in section 4.5, page 130.

Again, both PCR screening and the single successful transfection demonstrated the successful generation of plasmids containing inserts. Electrophoresis of plasmid DNA obtained after transformation only partially separated those containing insert from those lacking insert. Purification of the former at the expense of the latter may have co-purified a minority plasmid species. However, despite only HIV-1 Pol sequences rather than full-length HIV-1 genomes being retained in the vector, re-arrangement within transformed E. coli was still regularly observed, often producing plasmids smaller than pGEM-TΔRT itself. Purification of putative insert-positive plasmids and their subsequent re-transformation produced identical results. This was unexpected, as plasmid rearrangement within bacteria is commonly due to recombination between repeat sequences such as HIV-1 LTR sequences (Peden, 1992; Bi and Liu, 1996). In an attempt to overcome such problems, the TOP10F’ bacterial strain was used in this section as opposed to the SURE2 strain used in previous chapters. However, from the data presented here, TOP10F’ cells are equally poor at maintaining HIV-1 Pol-containing plasmids.

An alternative reason for the consistent failure of these bacteria to maintain these plasmids may be that the presence of non-HXB2 RT sequences inadvertently creates a reading frame for a protein toxic to E. coli. Selection of bacteria that have excised some
or all of this sequence from the plasmid may be the cause of the array of band patterns seen in FIGURE 3-12 and FIGURE 3-13.

From colony screens, it was found that the ligation efficiency was no different to that observed with pHIVEC in the previous chapter, but the absolute number of colonies was generally two-fold higher. This would have led to up to 40 sample RT variants per recombinant virus – higher than that obtained previously, but still considered inadequate. In the next chapter the second-round PCR was again re-designed, as was the shuttle vector, such that the number of deliberate mis-pairings was reduced to one per primer. Improved PCR performance would be expected to increase further the number of colonies and consequently the representation of sample RT variants.
4. Design and Application of a Shuttle Vector System (II)

4.1. Introduction

Each of the methods for the genesis of recombinant viruses described in the preceding sections suffered from the same problems – low efficiencies in the second-round PCR and in the digestion and ligation of these PCR products into plasmid vectors, intrabacterial re-arrangement of plasmids, and poor, if any, amplification during ‘final’ PCRs. As a result, the shuttle vector was re-designed with a new cloning site to improve both ligation and PCR efficiencies.

In the previous chapter, the cloning site of pHIVEC was itself cloned into a shuttle vector. However, when designing this cloning site, the list of possible restriction sites was constrained to those absent from the full-length pHXB2ART molecule. Here, the cloning site of pGEM-TART was re-designed using a broader range of potential restriction sites. The original BstBI and SacII restriction sites were replaced by MfeI and XbaI sites respectively.

In addition to retaining maximal amounts of sample-derived RT domains, this new vector (pRTshut) possesses two significant advantages over pGEM-TART. Firstly, ligation efficiency improves with longer sticky ends – MfeI and XbaI cleave to leave four-base sticky ends as opposed to the two-base overhangs left by each of BstBI, NaiI, and SacII. Secondly, only single base mismatches between the primers and target sequences are required for their introduction during second-round PCR amplification, contrasting favourably with the multiple mismatches with the ‘Inner’ and ‘IN’ primer sets (compare FIGURE 2-6 and 3-4 with FIGURE 4-7). Finally, the second-round antisense primer specifies less of the 3’ end of RT, allowing more sample RT sequence to be represented in each recombinant virus. As an additional measure, a third bacterial strain (STBL2, section A.3) was employed for its supposed suitability in maintaining
the integrity of Pol-containing plasmids. Aside from the identities of the restriction sites, this protocol is as per the previous shuttle vector system, illustrated in FIGURE 3-1, page 95

4.2. pRTshut

The set of restriction enzymes lacking sites in either pGEM-T△RT or HIV-1 RT is much greater than that used in the design of pHIVEC. This was due both to the shorter length of pGEM-T△RT relative to pHIVEC, and to the multiple cloning site of ancestral pGEM-T Easy Vector having been removed from pGEM-T△RT. Vector NTI software was used as described in section 2.2 to locate potential unique restriction sites which were silent with respect to the amino acid sequence of HIV-1 Pol (FIGURE 4-1). At the 5' end of RT, it was found that an MfeI site could be introduced across bases 11-16 (5 nucleotides downstream of the BstBI site). Similar mutagenesis scanning across the RT-integrase junction of Pol revealed three potential sites – a BamHI site spanning bases 1,664-1,669 of RT, a BseAI site across bases 1,667-1,672 of RT, and an XbaI site over bases 3-8 of integrase. To introduce either of the first two sites using second-round PCR primers would have necessitated an antisense primer specifying the nine 3'-terminal amino acids of RT. Moreover, introduction of the BamHI site into PCR products would have required two mismatched bases within the PCR primer, a strategy thought unlikely to improve on previous experience. Therefore, the XbaI site was selected as it required an antisense primer with just one deliberate mismatch, specifying only the three 3'-terminal amino acids of RT (FIGURE 4-2).
FIGURE 4-1: HXB2 sequences at the 5' and 3' ends of RT showing potential silent restriction sites for use in a new shuttle vector. Mismatches are highlighted in bold type.

FIGURE 4-2: Comparison of the nucleotide sequences and restriction maps of the RT-deletion regions of pGEM-ΔRT and pRTshut.
Two oligonucleotides were designed specifically for the construction of pRTshut. As they were not used elsewhere in this project, they are described below in TABLE 4-1 and FIGURE 4-3. Two PCRs were performed on pGEM-TΔRT to introduce the required restriction sites, one using the primer pair T7 and MUN-1, the other using INT-X and α-Sp6 (both T7 and α-Sp6 are detailed in TABLE A-1). To the former was added 5U of both SacII and ApaI, and to the latter 5U of both SacII and EcoRI (shortening the PCR products to 569bp and 438bp respectively). Reactions were incubated at 37°C for 2 hours. Three separate digests were performed on pGEM-TΔRT as detailed in TABLE 4-2.

FIGURE 4-3: Schematic diagram of the RT-deletion region of pGEM-TΔRT, showing the alignments of the mutagenic primers INT-X and MUN-1.

TABLE 4-1: Specific oligonucleotides used in the synthesis of pRTshut.

<table>
<thead>
<tr>
<th>INT-X</th>
<th>GGG TCT CCG CGG GAA TCA GGA AAG TAC TAT TTC TAG ATG GAA TAG ATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUN-1</td>
<td>TTT CCT GAT ACC CGC GGA GAC CCC GGG CAA TTG GGC TAA TGG GAA AAT TTA</td>
</tr>
</tbody>
</table>
All five digests were separated by agarose gel electrophoresis (FIGURE 4-4). The rearmost bands of the pGEM-TART digests and the stronger of the two bands in each of the digested PCR lanes were excised and purified (see boxed bands in FIGURE 4-4). Three cross-liguations were performed – one containing the Apal-SacII-digested backbone and PCR fragment, one containing the EcoRI-SacII-digested backbone and PCR fragment, and one containing both digested PCR fragments and the Apal-EcoRI-digested backbone. As before, these 20μl ligations had a 3:1 molar ratio of PCR fragment(s) to backbone, contained T4 DNA ligase in the appropriate buffer as supplied by NEB, and were incubated overnight at 12°C before being transformed into TOP10F’ E. coli (section A.10).

Colonies arising from each ligation were screened using S-HR and AS-HR (394bp product), followed by digestion with the enzyme appropriate for the mutagenic primer used to generate the insert, i.e. XbaI for those inserts derived from the INT-X-αSp6-primed PCR, and MfeI for those amplified by the T7-MUN-1-primer pair. The PCR product from one colony from the ‘INT-X-αSp6’ ligation successfully digested with XbaI, and amplicons from four colonies from the ‘T7-MUN-1’ ligation were successfully digested with MfeI. The single colony from the first plate was labelled pGEM-TART-XbaI; the four colonies from the second were labelled pGEM-TART-MfeI-A through D. The attempted ‘triple ligation’ containing the Apal-EcoRI-digested backbone and both digested PCR products produced colonies which amplified to give a 394bp product, but none of these products contained the desired restriction sites.

Plasmid DNA was extracted from cultures expanded from the five colonies detailed above by QIAGEN Plasmid Mini Kits (section A.8.2). Each extract was subjected to two separate restriction digests, one with SacII and ApaI, and the other with SacII and EcoRI. Each digest generated two bands, separated by gel electrophoresis. The smaller of the two bands were excised and purified from the SacII-EcoRI digest of pGEM-TART-XbaI and the ApaI-SacII digests of the four pGEM-TART-MfeI constructs (FIGURE 4-5). Conversely, the larger of the two bands were excised and purified from the ApaI-SacII digest of pGEM-TART-XbaI, and the SacII-EcoRI digests of the four pGEM-TART-MfeI constructs (FIGURE 4-6).
TABLE 4-2: The three restriction endonuclease combinations used to digest pGEM-TART, and the predicted band sizes thereof.

<table>
<thead>
<tr>
<th>Digest No.</th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enzymes</td>
<td>Apal – SacII</td>
<td>EcoRI – SacII</td>
<td>Apal - EcoRI</td>
</tr>
<tr>
<td>Expected band sizes (bp)</td>
<td>3,397</td>
<td>3,527</td>
<td>2,959</td>
</tr>
<tr>
<td></td>
<td>568</td>
<td>438</td>
<td>1,006</td>
</tr>
</tbody>
</table>

FIGURE 4-4: Restriction endonuclease digests of pGEM-TART and two mutagenic PCRs (see text). Lane 1 - pGEM-TART uncut (supercoiled plasmid). Lanes 2-4 – digests I-III detailed in TABLE 4-2. Lane 5 – T7-MUN-1-primed PCR digested with Apal and SacII. Lane 6 – INT-X-øSp6-primed PCR digested with SacII and EcoRI. The boxed bands were excised and purified.
FIGURE 4-5: Double restriction endonuclease digests on 'intermediate' plasmid constructs. Lane 1: pGEM-TART-Xbal digested with SacII and EcoRI, Lanes 2-5: pGEM-TART-MfeI A-D respectively, digested with ApaI and SacII. Boxed bands were excised and purified.

FIGURE 4-6: Double restriction endonuclease digests on 'intermediate' plasmid constructs. Uppermost bands from each lane were purified. Lane 1: pGEM-TART-Xbal digested with SacII and ApaI, Lanes 2-5: pGEM-TART-MfeI A through D, digested with EcoRI and SacII. Boxed bands were excised and purified.
Two sets of four reciprocal ligations were performed. All of the first set contained the ApaI-SaclII backbone of pGEM-TΔRT-XbaI and each contained one of the ApaI-SaclII fragments from pGEM-TΔRT-MfeI A through D. All of the second set contained the EcoRI-SaclII fragment from pGEM-TΔRT-XbaI and each contained one EcoRI-SaclII backbone from pGEM-TΔRT-MfeI A through D. Transformation into TOP10F' E. coli yielded between 1 and 45 colonies per ligation. Thirty-six colonies were screened by PCR using primers S-HR and AS-HR. All but two colonies successfully amplified using these primers. Eighteen amplicons were subjected to two separate restriction digests with MfeI and XbaI. All but one were shown to contain both XbaI and MfeI sites. Four were selected at random and cultured overnight as before.

Automated sequencing of plasmid DNA from the four candidate colonies using primers T7 and αSp6 (TABLE A-1) revealed each to be identical and to have the predicted nucleotide sequence. One was selected and labelled pRTshut.
4.3. Methods

New second-round PCR primers corresponding to IN-1, IN-2a and IN-2b were synthesised, enabling the introduction of the new restriction sites. These primers were designed with 20-base 5’ ‘tails’ both to improve second-round PCR efficiency, and to demonstrate successful digestion of PCR products,. The three replacement primers are detailed below, in TABLE 4-3 and FIGURE 4-7.

First round products amplified during the course of the experiments described in section 2 were used as templates for second-round PCRs. Experiments described in section 2.4 had previously demonstrated the presence of template DNA in these products. Prior to use, the optimal annealing temperatures of each primer pair (Mut-5 with either Mut-3T or Mut-3C) were assayed by ‘gradient PCRs’, as described in section 2. Products of second-round PCRs were gel-purified, and the eluates digested with 5U each of XbaI and MfeI at 37°C for 2 hours before being gel-purified.

Large-scale preparations of doubly-digested pRTshut were made as described earlier in section 2.3 for pHIVEC, but instead of sequential SacII and BstBI digestions, a simultaneous XbaI and MfeI digestion in NEB buffer #2 was performed overnight at 37°C.

Prior to ligation, both digested pRTshut vector and digested PCR products were quantified by relative yield analysis using BioMax software (section A.13). Inserts ligated with pRTshut vector were transformed into STBL2 E. coli (section A.3). Plasmid DNA extracted from resultant bacterial cultures was fractionated by gel electrophoresis.

Plasmid DNA bands migrating at a rate consistent with the presence of an insert were gel-purified. These were used both as templates in ‘final’ PCRs using primers S-HR and AS-HR, and as DNA for transformation again into STBL2 E. coli to increase yield. Plasmid DNA extracted from these latter cultures also acted as template in ‘final’ PCRs,
with the intention to co-transfect products with linearised deletion clone into SupT-1 cells as before.

PCR screening was performed at each stage using three primer pairs: S-HR and 469W to screen across the MfeI ligation site (500bp product), 111W and AS-HR to screen across the XbaI ligation site (335bp product), and S-HR and AS-HR to screen across the entire cloning site (393bp product). It must be noted that the extension time of these reactions was set at just 1 minute, so that signal would not be expected from insert-positive template with the S-HR-AS-HR-primed PCRs.

### TABLE 4-3: Primers synthesised to replace the 'IN' series.

<table>
<thead>
<tr>
<th>MUT-5</th>
<th>TTT AAT TTT CCC ATT AGC CCA ATT GAG AC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUT-3T</td>
<td>TTG GGC CTT ATC TAT TCC ATC TAG AAA TAG TAC TTT</td>
</tr>
<tr>
<td>MUT-3C</td>
<td>TTG GGC CTT ATC TAT TCC ATC TAG AAA CAG TAC TTT</td>
</tr>
</tbody>
</table>

**Protease**

<table>
<thead>
<tr>
<th>5'-TTTAAATTTT CCCATAGCCCACAATTGAGAC</th>
<th>Mut-5 (sense)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CACTTTAAATTTT CCCATAGCCCTATTGAGACTGTA</td>
<td>HXB2 (sense)</td>
</tr>
<tr>
<td>CT L N F P I S P I E T V</td>
<td></td>
</tr>
</tbody>
</table>

**RT**

<table>
<thead>
<tr>
<th>R K V L F L D G I D K A Q D</th>
<th>HXB2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGGAAAGTACTA TTTTTAGATGGAATAGATAAGGCCCAAGAT</td>
<td>HXB2 (sense)</td>
</tr>
<tr>
<td>TTTCATGAC AAGATCTACCTTTATCTATTCCGGT</td>
<td>Mut-3T/C (antisense)</td>
</tr>
</tbody>
</table>

**Integrate**

<table>
<thead>
<tr>
<th>R K V L F L D G I D K A Q D</th>
<th>HXB2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGGAAAGTACTA TTTTTAGATGGAATAGATAAGGCCCAAGAT</td>
<td>HXB2 (sense)</td>
</tr>
<tr>
<td>TTTCATGAC AAGATCTACCTTTATCTATTCCGGT</td>
<td>Mut-3T/C (antisense)</td>
</tr>
</tbody>
</table>

FIGURE 4-7: Alignment of the 'Mut' primer set on HIV-1 Pol domains (upper set – protease-RT junction, lower set – 3' end of RT). Restriction sites introduced by the primers are highlighted in bold type. The alternate bases at the third base of the final RT codon in the lower alignment were introduced for the purposes of future qPMA analysis of mixed virus populations.
4.4. Results and Analysis

FIGURE 4-8 shows the results of the gradient PCR experiments using primer pairs Mut-5-Mut-3T and Mut-5-Mut-3C. The optimal annealing temperature for both primer pairs was judged to be 47°C – amplification at higher temperatures led to extra product species with the first pair and lowered yields with the other. To ensure \textit{XbaI} and \textit{MfeI} were efficiently digesting the PCR products, control restriction digests were performed on Mut-5-Mut-3T-primed PCR product using pHXB2 as template, and separated on a 2% agarose gel (TABLE 4-4, and FIGURES 4-9 and 4-10). \textit{MfeI} efficiently cleaved the 5’ 122bp product of \textit{BsrGI} digestion, and \textit{XbaI} efficiently cleaved the 3’ 100bp product of \textit{NlaIV} digestion.

Amplification of first-round RT-PCR products from sample material was equally successful as amplification from plasmid templates. FIGURE 4-11 shows a post-purification yield analysis gel with six second-round PCR products from sample-derived material, two similar PCR products but using plasmid templates, and digested pRTshut.

Ligations were transformed into bacteria, giving colony counts ten-fold higher than previous transformations (typically 50-250), but with a typical frequency of insert positivity of just 10% (from PCR screening with primers T7 and 469W). FIGURE 4-12 shows the agarose gel visualising the transformed plasmid DNA extracts. Insert-positive pGEM-T\DeltaRT generated in the previous chapter was included as a molecular weight comparator. At 5,597bp, it would be identical in size to insert-positive pRTshut.
FIGURE 4-8: Gradient PCR using first round product amplified from a sample with low copy number RNA. Top row: Mut-5-Mut-3T, Bottom row: Mut-5-Mut-3C.

TABLE 4-4: Restriction endonuclease compositions of the six digest reactions.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>MfeI</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>XbaI</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>BsrGI</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>NlaIV</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

123
FIGURE 4-9: Schematic of the 1,717bp Mut-5-Mut-3T PCR product, showing the location of recognition sequences for restriction endonucleases XbaI, MfeI, BsrGI, and NlaIV.

FIGURE 4-10: Investigation into the activity of restriction endonucleases XbaI and MfeI on Mut-5 and Mut-3T/C-primed PCR products. Lanes 1-6 – reactions A-F respectively from TABLE 4-4.
Considering the distribution of species within ligations shown in FIGURE 4-12, it is perhaps unsurprising that the majority species after transformation was parental pRTshut, either from self-ligation, or as residual uncut plasmid. However, faint but visible bands (boxed in the figure) were observed migrating at a rate similar to the insert-positive pGEM-TART. Upon re-transformation and DNA extraction, a selection was digested with MfeI and XbaI (FIGURE 4-13). The insert-positive nature of these bands was confirmed, although one had failed to maintain the integrity of both restriction sites – suggesting it may have been derived from a single ligated molecule.

The constructs presented in FIGURE 4-12 and 4-13 contained inserts amplified from monotypic plasmid templates. Dramatically different results were observed when the RT inserts were derived from sample RNA. PCR screening of ligations indicated correct ligation at both restriction sites (FIGURE 4-14), and transformed bacteria produced similar numbers of colonies to previous experiments (typically 50-250). However, the electrophoretic profiles seen after transformation of plasmid preparations whose inserts derived from patient sample material (FIGURE 4-15) were reminiscent of those seen in previous chapters (FIGURE 2-17, page 90, 3-13, page 106, and 3-14, page 108), with multiple DNA species both larger than putative insert-positive pRTshut and smaller than insert-negative vector.
As they were of approximately the expected size, and absent from the self-ligation control lane, the row of bands from FIGURE 4-15 marked with an arrow were excised, purified, and re-transformed. Five electrophoretic profiles of subsequently extracted DNA are shown in FIGURE 4-16, alongside the purified band used in transformation, further reinforcing the suggestion of re-arrangement of plasmid sequences within *E. coli*. Due to the lack of ‘final’ PCR product, transfection was not attempted.

FIGURE 4-12: *Lane 1 – Insert-positive pGEM-TART plasmid from section 2. Lane 2 – undigested pRTshut. Lanes 3-6 – Plasmid DNA extracted from transformed pRTshut-RT insert ligations. Boxed bands represent the putative insert-positive constructs.*
FIGURE 4-13: Lanes 1-3 – Three DNA extracts from bacteria ‘re-transformed’ with bands corresponding to those boxed in FIGURE 4-12. Lanes 4-6. The same plasmids as in lanes 1-3 after digestion with XbaI and MfeI.

FIGURE 4-14: PCR screens of seven ligations using sample-derived RT domains. Top row: S-HR-469W-primed PCRs, Middle row: S-HR-AS-HR-primed PCRs, Bottom row: 111W-AS-HR-primed PCRs. In each row, lanes 1-7 – pRTshut-RT insert ligations, lane 8 – self-ligation, lane 9 – negative PCR control (water), lane 10 – positive PCR control (pHXB2).
FIGURE 4-15: Lanes 1-7 – electrophoretic profiles of seven DNA species extracted from bacterial cultures transformed with pRTshut ligated to sample-derived PCR-amplified RT domains. Lane 8 – DNA extracted from bacteria transformed with pRTshut self-ligation. The arrow indicates the row of bands that were excised.

FIGURE 4-17 shows the typical products of ‘final’ PCR amplifications with S-HR and AS-HR from purified putatively insert-positive pRTshut constructs. These were reminiscent of similar products observed in the previous chapter (FIGURE 3-9 and 3-10, page 104). The absence of specific product of the correct size was surprising considering the successful amplification of the unmodified pHXB2 plasmid control. Some distinction can be drawn between those in which the templates were plasmids in which the inserts were plasmid-derived (lanes 3 and 4) and those whose RT domains originated from sample material (other lanes). The presence of the strong band at approximately 100bp in both the negative and positive control lanes suggests it may be an artefact of the primer pair. Despite many experiments attempting to optimise the PCR conditions further, by adjusting the concentration of primers, DNA, magnesium cations, and/or deoxynucleotides, no improvement upon FIGURE 4-17 could be made. The products of the S-HR-AS-HR ligation screening PCRs shown in the middle row of FIGURE 4-14 also display a similar lack of discrete bands.
FIGURE 4-16: Lanes 1-5 – The bands excised from FIGURE 4-15 after purification. Lane 6 – insert-positive pGEM-TART from section 03 as a transformation control. Lanes 7-11 – plasmid DNA derived from transformation of the material in lanes 1-5 respectively.

FIGURE 4-17: 'Final' PCRs on purified insert-positive bands using primers S-HR and AS-HR. Lanes 1-9 – bands purified as per those shown in FIGURE 4-15 as templates. Lane 10 – negative PCR control (water). Lane 11 – positive PCR control (pHXB2). The templates for the PCRs shown in lanes 3 and 4 were putatively insert-positive plasmids containing RT domains derived from pHXB2-based plasmids (see text). The templates for the other PCRs were plasmids whose RT domains would have been sample-derived.
4.5. Discussion

With the exception of PCR performance, which was markedly improved with the ‘Mut’ primer set, similar problems to those experienced previously were observed, namely plasmid re-arrangement and inability to amplify transfectable material from purified plasmids containing RT inserts. Again, the quality of transformation was dependent upon the source of the RT insert ligated into the transformed vector. Constructs whose RT domains were derived from HXB2-based material gave clean separable bands after transformation, whereas with those whose RT domains derived from sample material, plasmid re-arrangement was invariably seen after transformation.

The use of a third bacterial strain (STBL2) failed to improve upon the results seen with SURE2 and TOP10F’ in previous chapters, suggesting that the fault does not lie with the inherent stability of the shortened plasmid sequences. Double purification of inserts (after both amplification and digestion) should have precluded any non-specific DNA sequences in the ligations. Furthermore, plasmid species smaller than the vector alone have been observed. The seemingly random distribution of the sizes of these plasmid species suggests non-specific excision is occurring during the propagation of ligated material. As suggested in the previous chapter, this may be due to the HIV-1 Pol sequences encoding a toxic protein product, although synthesis of recombinant Pol proteins in *E. coli* has become routine (Kim et al., 1996; Boretto et al., 2001). Furthermore, during propagation of pHXB2ΔRT, which contains almost all the sequences present in pRTshut including the two recombination-prone LTRs, plasmid re-arrangement has never been observed.

It was decided that in order to overcome the continued problems of plasmid-based manipulation of RT domains, a system would be devised in which the templates for the ‘final’ PCRs would comprise linear molecules of double stranded DNA in which HXB2 protease and integrase domains were ligated to sample RT domains. This plasmid-free system would preclude the need for bacterial propagation. For the design of this strategy, access to ‘Human retroviruses and AIDS’ (1999), a reference resource containing alignments of full-length HIV-1 nucleotide sequences. Examination of the
sense primers used for PCR amplification of clade B RT domains in the previous three chapters challenged their suitability. It was discovered that HXB2 is exceptional amongst clade B viruses in having a guanine at position 18 of RT as opposed to an adenine. This nucleotide position lies at the 3’-terminus of the Inner-1 primer-template duplex. Huang et al. (1992) showed how a mispair between a 3’-terminal adenine in the primer and a guanine in the template dramatically reduces the efficiency of Taq-mediated strand extension. In becoming IN-1, Inner-1 was lengthened by three bases, but with only a 3-base ‘clamp’ beyond the mispair, the efficiency of this PCR may not have been significantly improved. In Mut-5 from this section, this clamp was reduced to just two bases (FIGURE 4-11). Fortunately, research suggests the influence of single mismatch within a primer declines sharply away from the 3’-terminal position (Ayyadevara et al., 2000).

A relative lack of upstream mis-pairings as compared to Inner-1 and IN-1 may have improved the stability of the Mut-5-template duplex during amplification. In contrast, the binding sites for the Inner-2, IN-2, and Mut-3 primer pairs were similar and well conserved amongst clade B viruses. From sequencing data, it was found that all RT domains from DELTA samples and ARS samples possessed adenine at nucleotide 18. Hence, it was perhaps unsurprising that amplifications from HXB2-based templates were superior to those from patient-derived templates.
5. The Restriction-Endonuclease-Ligation System

5.1. Introduction

Whilst earlier methods attempted to clone RT fragments into plasmid vectors, intractable difficulties with plasmid propagation stimulated an fifth and final approach. The homologous recombination step for the generation of chimeric viruses was retained, but a new method used to obtain PCR products for co-transfection was applied. Instead of using bacterially-propagated HXB2-derived plasmid vectors containing sample RT domain inserts as templates for the ‘final’ PCR, products of simultaneous restriction digestion-ligation reactions (RELs) were used.

This ‘REL’ system is outlined in FIGURE 5-1. Briefly, HXB2 protease was PCR amplified such that its 3’ terminus possessed an *NgoMl* site, and integrase was amplified such that its 5’ terminus possessed an *MfEI* site. The second-round sense and antisense primers used to amplify the sample RT domain were designed in such a way that the PCR product overlapped the flanking HXB2 protease and integrase amplicons, with restriction sites possessing sticky ends complementary to those introduced into the protease and integrase amplicons (*BspEI* and *EcoRI* respectively). In a single reaction, all three molecules were simultaneously subjected to ligation by T4 DNA Ligase, and digestion with the four requisite enzymes. The complementary nature of the ligation junctions results in the destruction of the two restriction sites upon correct ligation, protecting the ligated molecule from further digestion. Conversely, ligation between two members of the same DNA species would preserve a restriction site, rendering the dimer vulnerable to further digestion. As a result, the reaction proceeds in an irreversible forward direction, with continuous digestion of undesired ligation products, and irreversible generation of the desired ligation products. As before, products from a ‘final’ PCR were co-transfected with digested pHXB2ΔRT, with intracellular homologous recombination generating recombinant viruses.
As with the previous systems, a means of distinguishing recombinants by quantitative point mutation assay was required. Again, RT domains were amplified twice with alternate sense primers differing at a single nucleotide position upstream of the EcoRI site. To ensure the presence of this site in recombinants, pHXB2ΔRT was digested with PshAI in addition to Smal. Primer 206V replaced AS-HR as the antisense primer in the 'final' PCRs to ensure sufficient overlap between vector and PCR product.

As well as having the advantages of being rapid, technically straightforward, and relatively inexpensive, the restriction-endonuclease-digestion (REL) protocol for producing recombinant viruses has been considerably more successful than previous methods.
Viral RNA extracted from sample

Sample Pol domain amplified by RT-PCR

Second-round PCR primers with HXB2 sequences introduce unique restriction sites to flank RT

BspEI

EcoRI

Single 'REL' reaction containing all four restriction endonucleases and T4 DNA Ligase incubated at 25°C for 4 hours, followed by 37°C for 4 hours.

Full-length molecular clone of HXB2

PR

RT

IN

Single-round PCR primers with HXB2 sequences introduce unique restriction sites in PR and IN

NgoMIV

MfeI

Reaction product PCR-amplified using primers located within PR and IN

Digested pHXB2ΔRT and final PCR product co-transfected into SupT-1 cell line. Homologous recombination generates chimæric virus

pHXB2ΔRT

PR

IN

HXB2 deletion clone digested with SmaI and PshAI

SmaI

PshAI

FIGURE 5-1: Schematic of the restriction-endonuclease-ligation (REL) method of recombinant virus generation.
5.2. Design

In the ‘pHIVEC’ method described in section 2, the termini of fragments to be ligated had been cleaved by different restriction endonucleases leaving complementary sticky ends. This technique was expanded to investigate the potential for similar ligation between PCR-amplified sample RT domains and HXB2 flanking sequences. Five interrelated concerns directed the design of the restriction-endonuclease-ligation (REL) system:

i.) Restriction endonuclease pairs for both ligation junctions.

ii.) The location of the primers to be used to generate the final PCR product.

iii.) The silent mutation to be used in a downstream point mutation assays.

iv.) Maximising representation of sample RT domains in the recombinant virus.

v.) Restriction endonuclease digestion of pHXB2ΔRT.

The design process was carried out as follows:

All nucleotide sequences that could encode the 15 3'-terminal amino acids of HXB2 protease were examined for 6-base sequences with central 4-base palindromes, of which six were found (FIGURE 5-2). TABLE 5-1 shows the analysis of these palindromes, with the possible final sequences, the two 6-base palindromes whose cleavage and cross-ligation would produce them, and the availability of suitable restriction enzymes that cleave to generate compatible 4-base overhangs. Palindromes 3, 5 and 6 were discarded due to lack of one or both appropriate restriction enzyme. 1 and 4 were rejected as the restriction endonuclease ‘pairs’ were identical in each case and correct ligation would fail to destroy their sites. Palindrome 2 became the sole option, and the resultant primer designs, with the cuts and cross-ligation are shown in FIGURE 5-3.
FIGURE 5-2: The 3'-terminal sequence of HXB2 protease showing potential 'silent' 6-base sequences with central 4-base palindromes.

TABLE 5-1: Analysis of the 4-base palindromes in FIGURE 5-2. Enzymes listed in parentheses leave overhangs upon cleavage rather than underhangs.
FIGURE 5-3: Nucleotide manipulations involved in restriction-endonuclease-ligation cross-ligation at the protease-RT junction.
The same process was followed to devise a site within integrase. The possible nucleotides of the first 15 amino acids of integrase were searched for palindromes (FIGURE 5-4). Analysis thereof (TABLE 5-2) isolated three potential pairs after rejecting those requiring either identical or unavailable restriction enzymes. The three potential enzyme pairs were XbaI-AvrII, MfeI-EcoRI, and BspHI-PciI. As the XbaI-AvrII pair would have been located just 2bp downstream of the end of RT, the 3’ terminus of the antisense RT-amplification primer to introduce the AvrII site would have needed to extend into RT. Such specifying of amino acids within RT would have violated point (4) in the above list. Primer design considerations informed the choice between the remaining two pairs. Introducing the PciI site would have necessitated a 43-mer primer, with more potential for inefficient priming of the PCR than the shorter primer needed to introduce the EcoRI site. Therefore, the MfeI-EcoRI site was chosen, (FIGURE 5-5).

In previous preparations of pHXB2ΔRT, Smal and XbaI were used to cleave the unique restriction sites flanking the deleted RT domain. However, the silent ‘PMA mutation’ had to be located inside these sites in order to ensure it was present in all recombinant viruses. Conversely, introduction of the PMA mutation could not have demanded primers that specified any of RT itself – the site had to be located outside the RT domain. The location of the Smal site precluded using the 5’ end of RT, and only two bases separate the end of the RT-deletion and the XbaI recognition sequence, neither of which can be changed without altering the phenylalanine codon. Moreover, for efficient amplification using the antisense primer, its 3’ end would have needed to extend into the RT domain, violating point (4). To overcome this difficulty, the unique PshAI site located 167bp inside the integrase domain was chosen to replace XbaI. Hence the antisense primer for the final PCR had to be located downstream of this point (206V was chosen), and introduction of the PMA mutation was through two antisense primers in the RT-amplification PCR.
FIGURE 5-4: The 5'-terminal sequence of HXB2 integrase showing potential ‘silent’ 6-base sequences with central 4-base palindromes. The nucleotide sequence shown begins with the first base of integrase.

TABLE 5-2: Analysis of the sequences shown in FIGURE 5-4. Enzymes listed in parentheses leave overhangs upon cleavage rather than underhangs.

<table>
<thead>
<tr>
<th>No.</th>
<th>Palindrome</th>
<th>Final Seq.</th>
<th>Sense Primer Seq.</th>
<th>Enzyme</th>
<th>Antisense Primer Seq.</th>
<th>Enzyme</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>YCTAGA</td>
<td>CCTAGA</td>
<td>TCTAGA</td>
<td>XbaI</td>
<td>CCTAGG</td>
<td>AvrII</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TCTAGA</td>
<td></td>
<td></td>
<td>TCTAGA</td>
<td>XbaI</td>
</tr>
<tr>
<td>2</td>
<td>CTCGAY</td>
<td>CTCGAC</td>
<td>GTCGAC</td>
<td>SalI</td>
<td>CTCGAG</td>
<td>XhoI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CTCGAT</td>
<td>ATCGAT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>GDATHG</td>
<td>GAATTG</td>
<td>CAATTG</td>
<td>MfeI</td>
<td>GAATTC</td>
<td>EcoRI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GGATCG</td>
<td>CGATCG</td>
<td></td>
<td>GGATCC</td>
<td>BamHI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GTATAG</td>
<td>CTATAG</td>
<td></td>
<td>GTATAC</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ATCGAY</td>
<td>ATCGAY</td>
<td>ATCGAT</td>
<td></td>
<td>ATCGAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATCGAY</td>
<td>GTCGAT</td>
<td>SalI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>ARGCYC</td>
<td>AAGCTC</td>
<td>GAGCTC</td>
<td>(SacI)</td>
<td>AAGCTT</td>
<td>HindIII</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AGGCCCG</td>
<td>GGGCCCG</td>
<td>PspOM I (Apal)</td>
<td>AGGCTCT</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>RCATGA</td>
<td>ACATGA</td>
<td>TCATGA</td>
<td>BspHI</td>
<td>ACATGT</td>
<td>PciI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GCATGA</td>
<td></td>
<td></td>
<td>GCATGC</td>
<td>(Sphi)</td>
</tr>
</tbody>
</table>
FIGURE 5-5: Nucleotide manipulations involved in restriction-endonuclease-ligation cross-ligation at the RT-integrase junction.
5.3. Materials and Methods

Five new primers were designed specifically for work in this section. Their sequences are detailed in TABLE 5-3, and their alignments with HXB2 Pol sequences in FIGURE 5-6.

Using an Env-deleted molecular clone of the HXB2 strain of HIV-1 as template, two AmpliTaq-mediated PCRs were performed – one using primers HPP3 and AS-MT5, and one using S-MT3 and 206V (TABLE 5-3 and A-1). Products were gel-purified as described in section A.8.1.

Rather than using first round PCR products generated during the experiments described in Section 1 as templates for second-round PCRs, single-tube Titan-mediated RT-PCRs were performed on the stored RNA extracts. Thermal cycling parameters were as follows: one round of 60min at 50°C, followed by 35 cycles of 1min at 92°C, 1min at 52°C, and 3min at 68°C. The reactions were completed by a ten-minute incubation at 68°C.

As with previous second-round PCRs, the optimum annealing temperature for the second-round PCR primers was established by ‘gradient PCR’ experiments. For each primer pair (S-MT5 and AS-MT3 or AS-MT3B) 8x20μl AmpliTaq PCRs were constructed containing Titan RT-PCR product amplified from a low RNA copy number sample. Each reaction was subjected to identical PCRs varying only in the temperature of the annealing step, which ranged from 42°C to 56°C at 2°C intervals: 95°C for 10min, followed by 35 cycles of 1min at 94°C, 1min at 42°C-56°C, 3min at 72°C, followed by a single round at 72°C for 7min.

1μl of first round RT-PCR product acted as template for each of two 25μl second-round AmpliTaq PCRs containing primers S-MT5 and either AS-MT3 or AS-MT3B. Thermal cycling parameters were as follows: 10min at 95°C, followed by 35 cycles of 1min at 94°C, 1min at 58°C (see pages 146 and 147, below), and 3min at 72°C. A seven-minute incubation at 72°C completed the reactions. 5μl of product were run down an agarose
geli to verify amplification before the remainder was purified as described in section A.8.1.

**TABLE 5-3: Specific primers used in this section.**

<table>
<thead>
<tr>
<th></th>
<th>ATA ATT GTC CGG AAT CTG TTG ACT CAG ATT GGT TGY ACT TTA AAT TT</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-MT5</td>
<td>CAA CAG ATG CCG GCC AAT TAT GTT GAC AGG TGT AGG</td>
</tr>
<tr>
<td>AS-MT5</td>
<td>TTT TTA GAT GCA ATT GAT AAG GCC CAA GAT GAA CA</td>
</tr>
<tr>
<td>S-MT3</td>
<td>TTC TTG GGC CTT ATG AAT TCC ATC YAA AAA</td>
</tr>
<tr>
<td>AS-MT3</td>
<td>TTC TTG GGC CTT ATG AAT TCC GTC YAA AAA</td>
</tr>
</tbody>
</table>

**FIGURE 5-6: Alignment on HIV-1 Pol domains (upper set – protease-RT junction, lower set – 3’ end of RT) of the primer set used in this section. Restriction sites introduced by the primers are highlighted in bold type.**
From correspondence with Paul Walsh at New England Biolabs, it was discovered that the half-life of T4 DNA ligase at 37°C is just 10 minutes, but this can be increased to 4 hours by reducing the temperature to 25°C. An investigation into the activities of each restriction enzyme at this temperature in a range of buffers was performed using unmethylated lambda phage DNA (NEB) as the substrate. The optimum NEB buffers for BspEI, MfeI, NgoMIV and EcoRI are #3, #4, #4, and U ('unique') respectively, but considering the potential for star activity of EcoRI (as seen in FIGURE 3-11, page 105), it was important to ensure optimal buffer conditions. Each enzyme was tested in its own optimum buffer, NEB buffer #4, and Stratagene Universal Buffer (SUB). Rather than the chloride ions found in NEB buffers, SUB contains acetate anions, which can lead to improved enzyme performance. Briefly, 20μl reactions containing 2μg of lambda DNA were subjected to digestion with 5U of enzyme in each specified buffer for a period of 2 hours at 25°C. Products were analysed by agarose gel electrophoresis.

Homodimers of the ‘flanker’ sequences possess two binding sites for one or other of the ‘final’ PCR primers. Although the 4-hour incubation at 37°C should significantly curtail their presence in completed REL reactions, PCR products corresponding to such templates were observed in ‘final’ PCRs. To investigate the effect of flanker:insert ratio upon the frequency of such templates in the completed REL reactions, four RELs were constructed. Each contained a constant absolute quantity of DNA, but with flanker:insert ratios of 1:3, 1:10, 1:30, and 1:100. These were performed in SUB supplemented with ATP (final concentration 1mM), with five units each of the four enzymes (NgoMIV, BspEI, EcoRI, and MfeI) and one unit of T4 DNA ligase. Incubation for 4 hours at 25°C was followed by 4 hours at 37°C and 20 minutes at 80°C (for the complete denaturation of BspEI and NgoMIV). The efficiency of the REL was assayed by PCR amplifying the products using primers S-HR and 206V (see below) and analysing the relative yield of each PCR product by gel electrophoresis.

The HXB2-derived flanker sequences and RT domains were quantified by Kodak-BioMax gel image analysis (A.13). Reactions were set up in Stratagene Universal Buffer supplemented with 10mM rATP containing both flanker sequences, RT domain DNA, five units of each of the four restriction endonucleases (NgoMIV, BspEI, EcoRI, and MfeI) and one unit of T4 DNA ligase. The relative molar concentrations of both
flanker sequences were equal, and were 3-fold lower than the molar concentration of the RT domain. Incubation was as follows: 4hrs at 25°C, 4hrs at 37°C, 20min at 80°C, cooled to 1°C until further use.

Each completed REL reaction was sequenced as per section A.14 using primers 469W and 111W, to ensure the REL junctions had been ligated correctly. Each was also used as a template in an Expand-mediated PCR using S-HR and 206V as primers. Furthermore, PCR amplification of dilution series of REL reactions was performed, to investigate sample RT variant representation.

Although amplicons from alternative templates were present in significant quantities in PCR products, the yields of 'final' PCRs were insufficient to recommend gel-purification of the correctly-sized band. Instead, a direct PCR purification as described in section A.8.1 was employed. It was hoped products from homodimer templates would not substantially interfere with homologous recombination. Co-transfection of purified PCR products with Smal-PshAI-linearised pHXB2ΔRT, and the subsequent maintenance of cultures were both as described in section A.15.
5.4. Results and Analysis

The primer pairs S-MT5 and either AS-MT3 or AS-MT3B were subjected to ‘gradient PCRs’, in which the temperature of the annealing step ranged from 42°C to 56°C in 2°C steps. However, other than an additional PCR product of low molecular weight, no loss of amplification was observed across this temperature range. The experiment was repeated using a range from 52°C to 66°C (FIGURE 5-7 shows the result for S-MT5 and AS-MT3B), within which amplification of the secondary band was abolished. Product of the predicted size (1,758bp) was observed at all temperatures except 66°C, although yield was substantially reduced over 60°C. This experiment informed the decision to use a temperature of 58°C during the annealing step of the second-round PCR. With the optimal annealing temperature of 58°C, second-round PCRs were performed on Titan-mediated first round products derived both from patient samples and from HXB2-derived plasmids. Amplification of the two flanker sequences using the primer sets HPP3/AS-MT5 and S-MT3/206V was also efficient and clean (FIGURE 5-8).

FIGURE 5-9 and 5-10 show the results of the buffer optimisation experiments in which unmethylated lambda phage DNA was digested by the four restriction endonucleases either singly or in combination. Upon each image are marked examples of possible star activity (seen with NgoMIV and EcoRI digestion in inappropriate buffers). BspEI digestion was consistent and accurate across the range of buffers (lanes 2-4, 6, and 7 of FIGURE 5-9). With only a single site in lambda DNA, NgoMIV digestion was detected in combination with ApaLI, where it cleaves a 16.2kb ApaLI-ApaLI fragment into a 14.4kb fragment and a 1,758bp fragment. As lanes 10, 12, and 14 show, digestion appears most specific in SUB, with unexpected bands seen in the digests using NEB buffers #3 and #4. EcoRI exhibited star activity both in its optimum buffer (NEB U) and in NEB buffer #4, and only digestion in SUB generated the expected banding pattern (lanes 2-4 of FIGURE 5-10). This result was also observed when MfeI was used in combination with EcoRI (lanes 10 and 11 in FIGURE 5-10). Like BspEI, MfeI digestion produced the predicted band pattern in each buffer used (lanes 6-8 in FIGURE 5-10). The superior nature of SUB over the NEB buffers is illustrated by the final digest reaction in which all four enzymes were employed. As FIGURE 5-10 shows, NEB
buffer #3 did not permit complete digestion of lambda DNA by all four enzymes (lane 13), and the NEB buffer #4 digest generated an unexpected band pattern from star activity (lane 14). In contrast, the digest in SUB produced a clean and correct banding pattern (lane 15).

FIGURE 5-7: Repeated 'gradient PCR' using primers S-MT5 and AS-MT3B with anneal temperatures ranging from 52°C to 66°C in 2°C intervals.

FIGURE 5-8: Agarose gel allowing relative yield analysis of five sample-derived RT domains (lanes 1-5), two plasmid-derived RT domains (lanes 6 & 7), and the two flancker sequences (lanes 8 and 9).
FIGURE 5-9: Enzyme optimisation by lambda digestion (I). Lanes 1, 5, 8, and 15 – molecular weight markers. Lanes 2-4 – BspEI in NEB #3, NEB #4, and SUB respectively. Lanes 6-7 – BspEI and NgoMIV in NEB #4 and SUB respectively. Lanes 9-14 – Paired ApaLI and ApaLI & NgoMIV digests in NEB #3, NEB #4, and SUB respectively.
FIGURE 5-10: Enzyme optimisation by lambda digestion (I). Lanes 1, 5, 9, 12, and 16 – molecular weight markers. Lanes 2-4 – EcoRI in buffers U, NEB #4, and SUB respectively. Lanes 6-8 – MfeI in buffers NEB #3, NEB #4, and SUB respectively. Lanes 10-11 – EcoRI & MfeI double digests in NEB #4 and SUB respectively. Lanes 13-15 – BspEI, NgoMIV, EcoRI, & MfeI multiple digests in buffers NEB #3, NEB #4, and SUB respectively.
FIGURE 5-11 shows the electrophoresis profile of 707M/206V-primed PCRs on RELs using a range of flanker to insert ratios. Raising the ratio of insert molecules to flanker molecules above 3:1 reduced the efficiency of PCR amplification of ‘correct’ templates, with signal abrogated above a 10:1 ratio.

A sample gel showing the products of S-HR/206V-primed ‘final’ PCRs using four optimised REL reactions as templates is presented in FIGURE 5-12. Together with the correctly-sized band corresponding to the expected PCR product, bands of shorter length were observed. All but one at 1.4kb were also found in the positive control lane to a variable degree, and may be due to suboptimal PCR conditions. The band unique to amplified REL reactions cannot be satisfactorily explained. Reasoning that products amplified from homodimers and other alternative templates would lack the binding sites within RT for 469W and 111W (TABLE A-1), these were used to sequence a selection of ‘final’ PCR products. All samples tested were found to have the correct sequence surrounding the REL junctions.

Co-transfection into SupT-1 cells of purified ‘final’ PCR products and pHXB2ART linearised by Smal-PshAI digestion was followed by maintenance of cultures as described previously. Monitoring of transfected cultures was by microscopy alone, as supplementary information from p24 ELISA testing did not assist in deciding when cultures should be harvested. Once the degree of cytopathic effect was judged to signify a sufficiently high virus titre, cultures were harvested and stored in liquid nitrogen prior to titration and use in subsequent competition experiments.
FIGURE 5-11: Experiment investigating the effect of flanked-insert ratio in REL reactions on the frequency of undesired 'final’ PCR products. Lanes 1-4 – flanked-insert ratios of 1:3, 1:10, 1:30, and 1:100. Lane 5 – negative PCR control (water). Lane 6 – positive PCR control (pHXB2).

FIGURE 5-12: An example of Expand-mediated S-HR/206V-primed PCRs on REL reaction products. Lanes 1-4 – REL reaction products. Lane 5 – negative water PCR control. Lane 6 – positive PCR control using pHXB2 as template.
5.5. Discussion

The REL protocol was successful in generating a panel of recombinant viruses differing only in their RT domains. In contrast to the four previous methods, the REL proved efficient, rapid, relatively straightforward, and inexpensive. Time-consuming and problematical propagation of plasmids in bacterial culture was avoided – a single-tube overnight step ligated second-round PCR products to flanking sequences, generating templates for ‘final’ PCRs. Again, unlike the corresponding steps in earlier strategies, the ‘final’ PCRs consistently generated transfectable products, albeit with superfluous co-amplified DNA species. Furthermore, unlike the strategies of chapters 2 to 4, the design of the second-round PCR primers led to the full 1,680 nucleotides of the sample RT being carried forward to the recombinant virus, with all remaining sequences in the recombinant viruses, from the first codon of integrase to the final codon of protease, being HXB2-derived. However, these operational constraints upon the location of the two primer pairs (S-RT5 and AS-RT3/AS-RT3B) precluded any latitude regarding optimal matching of primer sequences and target variability. As in the previous methods (chapters 2 to 4), the yields of the second round PCRs used to amplify sample RT domains remained variable (FIGURE 5-8). Nevertheless, unlike the previous methodologies, not one sample failed to amplify with the primer sets employed in the REL system.

However, it is clear that many components of the REL system as presented here could be improved. Firstly, although the second-round PCR was more efficient than its counterparts in previous sections, probably due to longer and better-designed primers with fewer mismatches (FIGURE 5-6), variation in yield was still observed (FIGURE 5-8). This may have been because the first round product templates were re-synthesised from stored RNA (>2 years at -70°C, with occasional freeze-thaw cycles). Coupled with the age (>7 years) and poor quality of the original plasma samples, it is quite feasible that RNA copy number had dropped considerably from the values obtained from the original tests. In contrast, amplification of the two HXB2-derived ‘flanker’ sequences was efficient, helped by a highly concentrated, monotypic plasmid template.
The REL reaction itself is most in need of further optimisation, in order to maximise the frequency of contiguous heterotrimeric *Pol* sequences at the expense of other species such as homodimers of flanker sequences. The buffer conditions have been optimised, but future investigation must address other parameters. The ratio between flanker and RT sequences influences the distribution of 'final' PCR products (FIGURE 5-11), although a further experiment is needed in which the total DNA quantity is maintained rather than that of just one component. Furthermore, although an improved proportion of the desired product appeared to be obtained at higher ratios of RT insert to flankers, extension of that ratio beyond 20:1 should be attempted.

Finally, there is the 'final' PCR used to amplify transfectable product from completed REL reactions. Unlike the relative failure in previous chapters of the S-HR/AS-HR primer pair, amplification using S-HR and 206V was largely successful, although products varied both in their yield and in the degree of amplification from inappropriate templates (e.g. residual homodimers of the flanking regions). Optimising the ratio of molecules entering the REL will confer some benefit, but attention to both the enzyme composition and the incubation schedule should further improve the frequency of full-length products. Data from NEB regarding the activity of each restriction endonuclease at 25°C have recently been obtained (TABLE 5-4). Consequently, in future RELs, the ratio of *BspEI*:*EcoRI*:*MfeI*:*NgoMI* should be set at 20:8:18:9. A correct ratio may well influence digestion efficiency during the final 37°C incubation. The initial 4 hour 25°C incubation was recommended by NEB, due to the half-life of T4 DNA Ligase decreasing as the incubation temperature climbs over 16°C, and represented a compromise between the conflicting temperature optima for digestion and ligation. As the critical step appears to be the digestion of homodimers, experimenting with cyclical incubations between 12°C and 25°C would not help. At 37°C, the half-life of the ligase is less than 10 minutes, so the presence of homodimers is due either to their inefficient cleavage or due to such products not being cleavable. This latter may be a product either of star activity (*EcoRI* and/or *NgoMIV*) or of too high a concentration of enzyme degrading cleaved sticky ends preventing their re-cleavage after ligation. Responding to the data from TABLE 5-4 should help minimise this possibility.
TABLE 5-4: Activities at 25°C of the enzymes used in the REL, relative to their activities at 37°C.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>BspEl</th>
<th>EcoRI</th>
<th>MfeI</th>
<th>NgoMIV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity at 25°C</td>
<td>30%</td>
<td>75%</td>
<td>33%</td>
<td>66%</td>
</tr>
</tbody>
</table>

Concurrent with further optimisation must be the important consideration of RT variant representation in recombinant viruses, as influenced by ‘bottlenecks’ at several stages of the experimental protocol. It is hoped that the technique can be developed such that over 200 sample RT domains become represented in each recombinant.

The extraction of RNA and the first round RT-PCR are probably the most critical steps, ensuring that the templates for the second round PCRs contain the requisite number of variants. Performing nested RT-PCRs on a 10-fold dilution series of RNA extracted from a plasma sample of known copy number would allow assessment of these steps’ efficiency. Increasing the volume of plasma entering the Boom protocol and doubling the volumes of RT-PCRs should improve representation.

In the experiments presented in this chapter, \(5 \times 10^6\) SupT-1 cells were transfected with 2\(\mu\)g of linearised deletion clone and variable amounts of ‘final’ PCR product. However, even after demonstration that the distribution of variants within the ‘final’ PCR products is satisfactorily representative of those in the original plasma sample, only a fraction of the RT domain molecules will recombine to generate virus. The electroporation step limits representation both by the high degree of target cell mortality upon electroporation, and by the independent probabilities of both molecules being cotransfected into each cell. It was estimated in the original protocol that approximately 8 000 recombinants are generated per transfection (Kellam and Larder, 1994). This exceeds the quota for adequate representation, but adaptations to the original system as outlined in (Boucher et al., 1996) may have reduced this number.

A final observation arising during these experiments is that the mutation to be monitored by the qPMA should have been introduced into REL products via the sense primer from the PCR used to generate the HXB2 integrase flanking sequence rather
than the antisense primer used to amplify sample RT domains. Therefore, the relatively
difficult amplification of RT domains would be performed just once per sample.
Additionally, the low amount of 'flanker' sequence DNA used in each REL would also
have led to just a single generation of each flanker variant. However, primarily because
the method evolved from previous ones that did not provide this choice, the mutation
was introduced into the antisense RT-amplification primer. As the two methods are
inter-related, this aspect of REL design is dealt with further in the discussion to the
qPMA methodology in the next chapter.
6. Development of a Quantitative Point Mutation Assay

6.1. Introduction

The point mutation assay (PMA) was originally designed by Kaye et al (1992) to monitor mutations in HIV-1 Pol domains conferring resistance to AZT. It is used to return the nucleotide identity at a specific locus within a specific nucleic acid sequence such as HIV-1 RT. A quantitative PMA (qPMA) was developed, and used for the analysis of mixed sequences, returning the proportions of each nucleotide at a specified locus. In this project, the qPMA was used to monitor the relative frequencies of two recombinant virus populations within mammalian cell cultures. As described in the previous chapter, each recombinant virus stock was engineered to contain either ‘A’ or ‘G’ at a specific (silent) site within integrase (the 'qPMA site'). The sequences surrounding the qPMA site were 100% homologous between recombinant viruses. This chapter details the qPMA protocol, and presents a series of experiments for its validation.

The principle of the qPMA is illustrated in FIGURE 6-1. Using one biotinylated and one non-biotinylated primer, PCR product is generated spanning the qPMA site. This is bound in duplicate to streptavidin-coated microwells, and the non-biotinylated strands stripped by sodium hydroxide treatment. An oligonucleotide probe whose 3' terminus extends to the nucleotide immediately preceding the qPMA site is annealed to the single-stranded DNA.

In the original protocol, the Klenow fragment of *E. coli* DNA Polymerase I was used to incorporate radio-labelled deoxy-nucleotides onto the 3’ end of the primer-template duplex. Output values were obtained by direct measurement of radioactivity in each nucleotide well. However, incorporation of radio-nucleotides was not restricted to the qPMA site – if one or more nucleotides immediately following the qPMA site were the
same as that at the qPMA site, extension with that deoxynucleotide could continue until a non-complementary template base was encountered. This necessitated a locus-dependent ‘correction co-efficient’ to the output counts of one or more of the nucleotides (Kaye, 1997).

Here, a chemiluminescent variation of the qPMA was employed, using fluorescein-conjugated dideoxy-nucleotides rather than radio-labelled deoxy-nucleotides. This improved upon the original qPMA method both by being safer and by precluding chain extension beyond the qPMA site. T7 sequenase was used to incorporate single dideoxy-nucleotides onto the DNA/DNA primer-template duplex (F12-ddATP in one of a pair of wells, F12-ddGTP in the other), and an anti-fluorescein alkaline phosphatase-conjugated antibody applied. The relative luminescence of enzyme substrate between each well indicates the relative frequencies of each nucleotide at the qPMA site in the original sample.
FIGURE 6-1: Schematic of the principle of the quantitative point mutation assay (qPMA).

1. Bind PCR product to streptavidin
2. Strip antisense strand with NaOH
3. Simultaneously anneal probe and extend by a single fluorescein-labelled ddNTP
4. Bind alkaline phosphatase-conjugated anti-fluorescein antibody
5. Apply Lumiphos substrate - enzyme causes substrate to luminesce
6. Use TopCount to measure luminescence

KEY

- **SAv**: Streptavidin
- **Bio**: Biotin
- **X**: qPMA site
- **ddY**: Labelled ddNTP
- **AP**: Enzyme-conjugated antibody
- **L**: Lumiphos
6.2. Materials and Methods

The primer sequences used (including the oligonucleotide probe) are detailed in TABLE 6-1 and their alignment within HIV-1 Pol is illustrated in FIGURE 6-2. The binding sites for both 987L and Pol-1-bio are well conserved among HIV-1 isolates. The binding sites for the two antisense primers and the probe are 100% homologous to their target sequences due to the method used to generate the recombinant viruses. TABLE 6-2 details the compositions of the buffers used in the quantitative point mutation assay.

TABLE 6-1: Primer sequences used in the final qPMA protocol.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>987L</td>
<td>GAA CTC CAT CCT GAT AAA TGG</td>
</tr>
<tr>
<td>AS-HR</td>
<td>GCT GAC ATT TAT CAC AGC TGG CTA CTA</td>
</tr>
<tr>
<td>Pol-1-bio</td>
<td>Biotin-CAG GAA AAT ATG CAA GAA TGA GG</td>
</tr>
<tr>
<td>G5804</td>
<td>TCA CTA GCC ATT GCT CTC CA</td>
</tr>
<tr>
<td>MT3-Pr</td>
<td>TTC TTG GGC CTT ATC AAT TCC</td>
</tr>
</tbody>
</table>

FIGURE 6-2: Location within Pol of the primers detailed in TABLE 6-1.
**TABLE 6-2: Compositions of the buffers used in the qPMA protocol.**

<table>
<thead>
<tr>
<th>Buffer A</th>
<th>PBS containing 10% Casein Blocking Buffer (Sigma, Poole, UK)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.5% Tween-20</td>
</tr>
<tr>
<td></td>
<td>0.1% sodium azide</td>
</tr>
<tr>
<td></td>
<td>Stored at 4°C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Buffer B</th>
<th>1.5M sodium chloride containing 10% Casein Blocking Buffer (Sigma, Poole, UK)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>150mM citric acid</td>
</tr>
<tr>
<td></td>
<td>0.5% Tween-20</td>
</tr>
<tr>
<td></td>
<td>0.1% sodium azide</td>
</tr>
<tr>
<td></td>
<td>Stored at 4°C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PMA Diluent</th>
<th>40mM Tris-Cl, pH 7.6 containing 20mM magnesium chloride 50mM sodium chloride</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stored at 4°C</td>
</tr>
</tbody>
</table>

| TTA            | 10mM Tris-Cl, pH 7.5 containing 0.05% Tween-20 0.1% sodium azide |

| Labelling Mix A| PMA Diluent containing 7nM MT-Pr (TABLE 6-1) 300μM dithiothreitol 100nM F^{12}-ddATP (NEN, Zaventem, Belgium) 16 Units/ml Sequenase (APB, Little Chalfont, UK) |

| Labelling Mix G| As per Labelling Mix A, but F^{12}-ddATP replaced by F^{12}-ddGTP |
RNA was extracted from cell-free supernatants using QIAamp Viral RNA Mini Kits, and subjected to 25μl Titan-mediated RT-PCRs using primers 987L and AS-HR (TABLE 6-1). Included in each batch of RNA test samples were two qPMA controls – RNA extracted from a virus containing ‘A’ at the qPMA site, and RNA from a virus containing ‘G’ at the qPMA site. Thermal cycling conditions were as follows: a cDNA synthesis step of 1 hour’s incubation at 55°C, followed by PCR cycling for 35 cycles of 1 minute at 92°C, 1 minute at 52°C, and 2 minutes at 68°C. Seven minutes’ incubation at 68°C completed the reactions. 2μl of each RT-PCR product were transferred to second-round 25μl AmpliTag-mediated PCRs using primers Pol-1-bio and G5804 (TABLE 6-1).

90μl of Buffer A (TABLE 6-2) were added to each well of a flat-bottomed, black 96-well microtitre plate. For each sample, 10μl of sample PCR product were added to a pair of wells. Finally, 10μl of each control PCR product was added to paired wells. The plates were briefly shaken, incubated at 45°C for 30 minutes, and washed with TTA. 100μl of 0.15M NaOH (freshly prepared from a 1.5M stock solution) were added to each well, and the plate incubated at room temperature for 5 minutes. The plate was washed with TTA and to one of each pair of wells 100μl Labelling Mix A were added (‘A wells’), and to the other 100μl Labelling Mix G (‘G wells’, TABLE 6-2). After 30 minutes’ incubation at 37°C, the plate was again washed with TTA. To each well were added 100μl of Buffer B (TABLE 6-2) containing alkaline phosphatase-conjugated anti-FITC antibody (NEN, Zaventem, Belgium) at a 1 in 1,000 dilution. Incubation for 30 minutes at 37°C was followed by eight further washes of the plate, with a 2-minute soak between the fourth and fifth washes. 100μl of Lumi-Phos (Lumigen Inc., MI) were added to each well. The plate was sealed before being incubated at room temperature in the dark for 1 hour.

Subsequent automated luminescence measurement was by a TopCount machine (Packard Bioscience Ltd, Pangbourne, UK) scanning each well for 6 seconds. Output was automatically multiplied by a factor of 10 to obtain counts per minute (photon emissions per minute), and analysed as follows:
(1) The count value of the ‘A well’ of the 100% G control was deducted from the counts of each ‘A well’. Counts for each ‘G well’ were reduced by the count value of the ‘G well’ of the 100% A control.

(2) Modified counts for each G well were multiplied by the modified count of the A well of the A control and divided by the modified count of the G well of the G control. A well counts were unmodified.

(3) Negative values were re-entered as zero.

(4) For each sample pair, the sum of the modified ‘A’ and modified ‘G’ counts were compared to the modified count of either control (identical after modification). If the total sample count was below 50% of the control count, then the result of that sample was rejected.

(5) For each sample pair, the proportion of ‘A’ viruses was calculated by dividing the modified ‘A’ count by the sum of the two count values (A and G), and expressed as a percentage.
6.3. Results and Analysis

To test the qPMA for reproducibility, a standard curve was constructed. RNA was extracted from two recombinant virus stocks - one with ‘A’ at the qPMA site, and one with ‘G’. Triplicate second-round 50μl PCRs were performed. Products were pooled into an ‘A’ batch, and a ‘G’ batch. Standards were prepared by mixing volumes from each batch to a range of ratios. Triplicate qPMAs were performed on these standards.

TABLE 6-3 shows the counts per minute (cpm) output from the TopCount, together with the intermediate values in the calculations used to assay the relative frequencies of each nucleotide at the qPMA site. The output values for the 0% and 100% input ratio standards were used to calculate the relative incorporation co-efficients of ‘A’ and ‘G’ in the assay. The output frequencies for intermediate input ratio standards were calculated using these co-efficients. FIGURE 6-3 and 6-4 present the data graphically, the first showing the three standard curve data sets, the second showing the mean calculated ‘A’ frequencies for each input ‘A’ frequency, together with the standard errors of these means. Regression analysis was performed on each data set not including the 0% and 100% controls, giving gradient co-efficients of 1.00, 0.99, and 1.02, with p-values of 1.7x10⁻⁷, 4.5x10⁻⁹, and 1.0x10⁻⁵ respectively. The corresponding statistics were 1.00 and 9.2x10⁻⁹ respectively for the regression equation of the mean values.

Duplicated second-round PCR and qPMA analysis was performed on products from five first-round RT-PCRs whose RNA templates were taken as part of the competition experiments described in the following two chapters. TABLE 6-4 shows the results and subsequent analysis. Although absolute counts varied between duplicates, their output frequencies showed high reproducibility. The two control PCRs were analysed by electrophoresis, by which differential amplification as a cause of the disparity between the outputs of the two controls was discounted (FIGURE 6-6).
TABLE 6-3: Data from three qPMA experiments using the same standard curve samples.

<table>
<thead>
<tr>
<th>% 'A' Input</th>
<th>Raw Cpm Data 'A'</th>
<th>Raw Cpm Data 'G'</th>
<th>- Background 'A'</th>
<th>- Background 'G'</th>
<th>Normalised 'A'</th>
<th>Normalised 'G'</th>
<th>% 'A' Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>983</td>
<td>21772</td>
<td>0</td>
<td>20743</td>
<td>0</td>
<td>163915</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>16120</td>
<td>15053</td>
<td>15137</td>
<td>14024</td>
<td>15137</td>
<td>110820</td>
<td>12.02</td>
</tr>
<tr>
<td>25</td>
<td>37163</td>
<td>15912</td>
<td>36180</td>
<td>14883</td>
<td>36180</td>
<td>117608</td>
<td>23.53</td>
</tr>
<tr>
<td>40</td>
<td>70989</td>
<td>14138</td>
<td>70006</td>
<td>13109</td>
<td>70006</td>
<td>103590</td>
<td>40.33</td>
</tr>
<tr>
<td>50</td>
<td>80632</td>
<td>13116</td>
<td>79649</td>
<td>12087</td>
<td>79649</td>
<td>95513.7</td>
<td>45.47</td>
</tr>
<tr>
<td>95</td>
<td>157060</td>
<td>1777</td>
<td>156077</td>
<td>748</td>
<td>156077</td>
<td>5910.83</td>
<td>96.35</td>
</tr>
<tr>
<td>100</td>
<td>164898</td>
<td>1029</td>
<td>163915</td>
<td>0</td>
<td>163915</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>% 'A' Input</td>
<td>Raw Cpm Data 'A'</td>
<td>Raw Cpm Data 'G'</td>
<td>- Background 'A'</td>
<td>- Background 'G'</td>
<td>Normalised 'A'</td>
<td>Normalised 'G'</td>
<td>% 'A' Output</td>
</tr>
<tr>
<td>0</td>
<td>983</td>
<td>19930</td>
<td>0</td>
<td>18889</td>
<td>0</td>
<td>118286</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>17453</td>
<td>18200</td>
<td>16470</td>
<td>17159</td>
<td>16470</td>
<td>107452</td>
<td>13.29</td>
</tr>
<tr>
<td>25</td>
<td>33622</td>
<td>16143</td>
<td>32639</td>
<td>15102</td>
<td>32639</td>
<td>94571.2</td>
<td>25.66</td>
</tr>
<tr>
<td>40</td>
<td>51836</td>
<td>12616</td>
<td>50853</td>
<td>11575</td>
<td>50853</td>
<td>72484.5</td>
<td>41.23</td>
</tr>
<tr>
<td>50</td>
<td>61612</td>
<td>10438</td>
<td>60629</td>
<td>9397</td>
<td>60629</td>
<td>58845.5</td>
<td>50.75</td>
</tr>
<tr>
<td>95</td>
<td>108984</td>
<td>1931</td>
<td>108001</td>
<td>890</td>
<td>108001</td>
<td>5573.33</td>
<td>95.09</td>
</tr>
<tr>
<td>100</td>
<td>119269</td>
<td>1041</td>
<td>118286</td>
<td>0</td>
<td>118286</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>% 'A' Input</td>
<td>Raw Cpm Data 'A'</td>
<td>Raw Cpm Data 'G'</td>
<td>- Background 'A'</td>
<td>- Background 'G'</td>
<td>Normalised 'A'</td>
<td>Normalised 'G'</td>
<td>% 'A' Output</td>
</tr>
<tr>
<td>0</td>
<td>1168</td>
<td>15199</td>
<td>0</td>
<td>14343</td>
<td>0</td>
<td>118596</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>10608</td>
<td>21975</td>
<td>9440</td>
<td>21119</td>
<td>9440</td>
<td>174624</td>
<td>5.13</td>
</tr>
<tr>
<td>25</td>
<td>41981</td>
<td>17471</td>
<td>40813</td>
<td>16615</td>
<td>40813</td>
<td>137382</td>
<td>22.90</td>
</tr>
<tr>
<td>40</td>
<td>68910</td>
<td>10810</td>
<td>67742</td>
<td>9954</td>
<td>67742</td>
<td>82305</td>
<td>45.15</td>
</tr>
<tr>
<td>50</td>
<td>93729</td>
<td>8188</td>
<td>92561</td>
<td>7332</td>
<td>92561</td>
<td>60625</td>
<td>60.42</td>
</tr>
<tr>
<td>95</td>
<td>127475</td>
<td>1787</td>
<td>126307</td>
<td>931</td>
<td>126307</td>
<td>7698</td>
<td>94.26</td>
</tr>
<tr>
<td>100</td>
<td>119764</td>
<td>856</td>
<td>118596</td>
<td>0</td>
<td>118596</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>
FIGURE 6-3: Graphical representation of the data from TABLE 6-3, showing the calculated output values for each of the three series of standard curve data.

FIGURE 6-4: Graphical representation of the data shown in TABLE 6-3, showing the mean output frequencies at each input percentage, with standard error bars. A unit gradient line (dashed) is also presented.
TABLE 6-4: TopCount data and analysis from experiments in which five RNA extracts were subjected to duplicated second-round PCRs and qPMAs.

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Raw Cpm Data</th>
<th>- Background</th>
<th>Normalised</th>
<th>% ‘A’ Output</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>‘A’</td>
<td>‘G’</td>
<td>‘A’</td>
<td>‘G’</td>
</tr>
<tr>
<td>‘A’ control</td>
<td>55184</td>
<td>877</td>
<td>54465</td>
<td>0</td>
</tr>
<tr>
<td>‘G’ control</td>
<td>719</td>
<td>17787</td>
<td>0</td>
<td>16910</td>
</tr>
<tr>
<td>1</td>
<td>48085</td>
<td>4055</td>
<td>47366</td>
<td>3178</td>
</tr>
<tr>
<td></td>
<td>47446</td>
<td>4420</td>
<td>46727</td>
<td>3543</td>
</tr>
<tr>
<td>2</td>
<td>11346</td>
<td>11298</td>
<td>10627</td>
<td>10421</td>
</tr>
<tr>
<td></td>
<td>20831</td>
<td>19881</td>
<td>20112</td>
<td>19004</td>
</tr>
<tr>
<td>3</td>
<td>27236</td>
<td>10031</td>
<td>26517</td>
<td>9154</td>
</tr>
<tr>
<td></td>
<td>33863</td>
<td>12340</td>
<td>33144</td>
<td>11463</td>
</tr>
<tr>
<td>4</td>
<td>52109</td>
<td>6913</td>
<td>51390</td>
<td>6036</td>
</tr>
<tr>
<td></td>
<td>48093</td>
<td>4928</td>
<td>47374</td>
<td>4051</td>
</tr>
<tr>
<td>5</td>
<td>67400</td>
<td>2836</td>
<td>66681</td>
<td>1959</td>
</tr>
<tr>
<td></td>
<td>89416</td>
<td>1880</td>
<td>88697</td>
<td>1003</td>
</tr>
</tbody>
</table>

FIGURE 6-5: Graph of the data in TABLE 6-4, with a line of equivalence (dashed).

FIGURE 6-6: Gel photo of the two controls from the experiment producing the data in TABLE 6-4. Lane 1: 100% ‘A’, Lane 2: 100% ‘G’.
6.4. Discussion

Accuracy and reproducibility were high when duplication of the second-round PCR was tested. However, it has not been shown that duplicating the entire protocol from the RNA extract would result in such reproducible data. As well as analysing a series of RNA extracts with duplication at each step, the efficiency of the nested RT-PCR could be improved to improve consistency between assays, and in particular to avoid potential sampling errors during small-scale competition experiments (see section 8, below). The sense primers need redesigning such that their binding sites are located much closer to the 3' end of RT, and both PCR steps should be optimised with regard to their annealing temperatures with a further investigation into varying the primer concentrations in the second-round PCR.

Although the dideoxy-nucleotides cause extension to terminate after a single incorporation, correction co-efficients are still required in order to allow for the differential incorporation efficiencies of each nucleotide. Referring to the 'A' and 'G' controls in the assays presented here (TABLE 6-3 and 6-4), the raw cpm data for \( \text{F}^{12} \text{ddATP} \) incorporation are approximately 3.5-8.5-fold higher than those for \( \text{F}^{12} \text{ddGTP} \), necessitating the presence of 100% 'A' and 100% 'G' controls in each assay. Absolute counts for each control varied in relation to the age both of the labelled dideoxy-nucleotides and of the Lumi-Phos substrate.

This disparity between the two nucleotides may be due to nucleotide-specific incorporation rates of T7 enzyme. However, after employing this technique to interrogate mutations at codon 215 of RT, a different co-efficient was needed (Julie Bennett, personal communication).

Another explanation is that upon becoming single stranded, the bound PCR product acquires a variable secondary structure dependent upon the nucleotide identity at the qPMA site. The presence of a 'C' at the qPMA site may distort the structure such that \( \text{F}^{12} \text{ddGTP} \) incorporation is inhibited, but when a 'T' is present at the site, the structure may be more amenable to \( \text{F}^{12} \text{ddATP} \) incorporation. In the chemiluminescent qPMA
presented here, the probe is both annealed to the single-stranded DNA template and extended by one base in a single step at 37°C. This constitutes a departure from the original protocol, in which the two steps were separated. Annealing was performed by heating the probe and template to 63°C and cooling slowly to room temperature, and followed by a 2 minute 37°C incorporation step (Kaye et al., 1992). Dissociation of these two steps in this assay may reduce the disparity between incorporation efficiencies.

Hence, increasing the amount of $^{12}$ddGTP in the labelling mix may not proportionately affect the relative counts for each incorporated nucleotide. Rather, as mentioned in the discussion to section 5 (page 151), the strategy of re-locating the qPMA site may be the best option.
7. Relative Fitness Analysis of the AZT Resistance Mutations M41L and T215Y

7.1. Introduction

Treatment of HIV-1 infection with AZT-containing therapy regimens often selects for viruses with mutations at codons 41 and 215 within their reverse transcriptase domains (M41L and T215Y/F). Together, these substitutions confer a 60-70-fold decrease in susceptibility to AZT (Larder and Kemp, 1989; Larder et al., 1991; Kellam et al., 1992). Rates of reversion at these codons in untreated individuals have suggested that these mutations lower the overall replicative capacity of HIV-1 (Goudsmit et al., 1997). However, reversion at these codons is often slow, indicating that these mutations result in only minor fitness debilitation (Albert et al., 1992; Goudsmit et al., 1997). In vitro studies also found that mutations M41L and 215Y incurred fitness loss (Harrigan et al., 1998; Keulen, 2000). Here, we attempted to replicate these findings using our techniques.

Using the REL method described in section 5, a panel of recombinant viruses varying solely in their RT domains was generated. All non-RT sequences were derived from the HXB2 laboratory strain of HIV-1, and were identical. RT domains were derived from HXB2 itself (rvHXB2), and from a molecular clone of HXB2 in which the two AZT mutations M41L and T215Y had been introduced by site-directed mutagenesis (rvHXB2-41/215). Furthermore, as described in the previous chapter, a qPMA site was engineered to enable relative frequency analysis of paired viruses in culture.

In this pair of experiments, the two RT domains were competed against each other and the change in their relative frequencies over time was monitored. Although a
reproducible trend was observed suggesting that HXB2 RT containing M41L and T215Y has lower relative fitness than HXB2 RT without these mutations, the results were not statistically significant. It is proposed that changes to the experimental protocol would enable more accurate and definitive assessment of the effect on fitness of these mutations.

7.2. Competition Protocol

As described in section 5, ‘final’ PCRs with primers S-HR and 206V were used to amplify from restriction-endonuclease-ligation reactions (RELs) in which templates comprised sample RT domains ligated to flanking HXB2-derived protease and integrase sequences. In the experiments described in this section, RT domains from two sources were used in the RELs – pHXB2 and pHXB2-41L/215Y. During the REL protocol two second-round PCRs were performed on each RT domain, one using AS-MT3 as the antisense primer, and one using AS-MT3B. Four RELs were performed – HXB2, HXB2^, HXB2-41L/215Y^, and HXB2-41L/215Y^.

Equal quantities (1µg) of two ‘final’ PCR products (either HXB2^ and HXB2-41L/215Y^ or HXB2-41L/215Y^ and HXB2^) were combined with 2µg of deletion clone DNA. Transfection of these mixtures into SupT-1 cells and the subsequent maintenance of cultures was as described in section A.15. 100µl cell-free supernatants were sampled from the cultures 3, 5, 7, and 10 days post-transfection, and entered into the first buffer of a QIAamp Viral RNA Mini Kit (section A.8.5). After 10 days, RNA extractions were completed, and qPMA analysis performed (section 6).
7.3. Results and Analysis

The qPMA data from the two direct competitions are shown in TABLE 7-1 and 7-2. Two timepoints only are given in the former as the RNA extracts from the intermediate timepoints failed to amplify by the nested RT-PCR. In the competition between rvHXB2\(^A\) and rvHXB2-41L/215Y\(^G\), the frequency of the former increased from 49.8% at day 3 to 52.5% by day 7 – a 0.013 \(\ln\) percentage point rise per day. Subjecting the data from the competition between rvHXB2-41L/215Y\(^A\) and rvHXB2\(^G\) to linear regression using time as the independent variable gave a gradient of -0.010 \(\ln\) % per day, but gave an \(F\)-value of 1.349 corresponding to a \(p\)-value of 0.365.

TABLE 7-1: qPMA data from a 10-day direct competition between rvHXB2\(^A\) and rvHXB2-41L/215Y\(^G\) showing the raw data, and the method of calculating virus frequency.

<table>
<thead>
<tr>
<th></th>
<th>Raw Cpm Data</th>
<th>- Background</th>
<th>Normalised</th>
<th>% Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>‘A’</td>
<td>‘G’</td>
<td>‘A’</td>
<td>‘G’</td>
</tr>
<tr>
<td>‘A’ control</td>
<td>44289</td>
<td>731</td>
<td>43461</td>
<td>0</td>
</tr>
<tr>
<td>‘G’ control</td>
<td>828</td>
<td>7809</td>
<td>0</td>
<td>7078</td>
</tr>
<tr>
<td>Day 3</td>
<td>33094</td>
<td>6038</td>
<td>32266</td>
<td>5307</td>
</tr>
<tr>
<td>Day 7</td>
<td>46468</td>
<td>7450</td>
<td>45460 6719</td>
<td>45460 41257</td>
</tr>
</tbody>
</table>
TABLE 7-2: qPMA data from a 10-day direct competition between HXB2-41L/215Y\textsuperscript{a} and HXB2\textsuperscript{G} showing the raw data, and the method of calculating virus frequency.

<table>
<thead>
<tr>
<th></th>
<th>Raw Cpm Data</th>
<th>- Background</th>
<th>Normalised</th>
<th>% Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>'A'</td>
<td>'G'</td>
<td>'A'</td>
<td>'G'</td>
</tr>
<tr>
<td>'A' control</td>
<td>44289</td>
<td>731</td>
<td>43,461</td>
<td>0</td>
</tr>
<tr>
<td>'G' control</td>
<td>828</td>
<td>7809</td>
<td>0</td>
<td>7078</td>
</tr>
<tr>
<td>Day 3</td>
<td>30430</td>
<td>5100</td>
<td>29602</td>
<td>4369</td>
</tr>
<tr>
<td>Day 5</td>
<td>32190</td>
<td>6190</td>
<td>31362</td>
<td>5459</td>
</tr>
<tr>
<td>Day 7</td>
<td>29194</td>
<td>5092</td>
<td>28366</td>
<td>4361</td>
</tr>
<tr>
<td>Day 10</td>
<td>23223</td>
<td>4715</td>
<td>22395</td>
<td>3984</td>
</tr>
</tbody>
</table>
7.4. Discussion

Two pairs of recombinant viruses were generated, varying only in their RT domains. The RT domains of one pair were derived from HXB2, whereas those of the other pair were derived from a molecular clone of HXB2 containing the two AZT resistance mutations M41L and T215Y. The two viruses within each pair differed only at a single nucleotide position, silent with respect to the amino acid sequence (the ‘qPMA site’). Using a quantitative point mutation assay (qPMA), the relative frequencies of each nucleotide at this position were used as markers for the relative frequencies of viruses in pairwise competition assays.

Two pairwise competition experiments were performed such that each virus from one of the two pairs was competed against that virus from the other pair possessing the alternative nucleotide at the qPMA site. Competitions were monitored by qPMA. After 10 days of culture, neither experiment had reached a position where one of the two viruses had successfully outgrown the other. Indeed, the gradients describing the changes in frequency of the viruses were not significantly different from zero.

The lack of evidence for fitness differences between mutant and wild-type virus was surprising as previous studies have shown M41L and T215Y to confer fitness losses to the HXB2 strain of HIV-1 (Harrigan et al., 1998; Keulen, 2000; Jeeninga et al., 2001). In vivo studies have recorded only slow reversion of M41L and T215Y upon withdrawal of AZT therapy suggesting the two mutations together confer only a low fitness deficit (Albert et al., 1992; Goudsmit et al., 1997; Bâtisse et al., 1998). However, in certain cases where additional resistance mutations in RT and/or protease have been present, wild-type sequences quickly dominated the in vivo population, suggesting rapid outgrowth of pre-treatment virus with significant fitness advantage over such multiple mutants (Devereux et al., 1999; Lukashov et al., 2001).

Clearly, the effects of codon changes at positions 41 and 215 on fitness are complex, and are probably dependent upon many additional factors including therapeutic history, genetic background, and the fitness parameters of the many possible intermediate
genotypes. Workers in Holland have performed experiments comparing in vitro fitnesses of HXB2 into which various combinations of AZT resistance mutations have been introduced by site-directed mutagenesis, showing the growth kinetics of the double mutant M41L–T215Y to be intermediate between wild-type virus and both single mutants (Keulen, 2000; Jeeninga et al., 2001). These experiments most closely resemble those performed here, but with the significant fitness differences being detected.

Several explanations can be invoked as to why these experiments failed to detect any fitness differences. The lack of replicates due to time constraints is clearly a problem in this study. As demonstrated in the following chapter (page 179), high variances of fitness parameters can be observed between replicate competition experiments. However, using monotypic plasmid templates for amplification of the RT domain should have minimised this risk.

The insignificant differentials observed in this assay may have reflected slow virus replication, such that after longer experimental duration, significant outgrowth may have been seen. This situation is considered unlikely as microscopy revealed the culture systems to contain high titres of virus (as measured by the degree of syncytium formation) as early as five days after transfection. Also, the duration of the competition experiments must be balanced by the potential for evolution – resistance mutations bestowing fitness deficit are likely to revert in culture in the absence of drug, and vice versa if drug pressure is maintained.

The possibility of evolution at codons 41 and/or 215 during the course of these experiments cannot be discounted. This is especially important in the light of recent evidence that single nucleotide substitutions at codon 215 can cause tyrosine to be replaced by amino acids conferring wild-type levels of RT fitness (Goudsmit et al., 1996, 1997; de Ronde et al., 2001; Garcia Lerma et al., 2001). Unfortunately, time constraints precluded investigation of the post-competition viruses, and this possibility could be neither confirmed nor denied.

Another possible source of error was recombination between the two virus species at sites between fitness determinants in RT and the qPMA site in integrase. Kellam and
Larder (1995) cultured a mixture of two virus stocks – identical save that one possessed the AZT resistance mutation M41L, and the other T215Y. Double mutants arising from recombination quickly appeared under AZT selection. Other researchers have demonstrated this effect using different resistance mutations (Moutouh et al., 1996; Yusa et al., 1997). However, Kellam and Larder (1995) observed no such changes in drug-free culture systems such as ours. Consequently, the fitness deficit incurred by M41L and/or T215Y in the absence of AZT was insufficient to select for double wild-type recombinants. This is surprising given that both mutations have been demonstrated to be deleterious when introduced into HIV-1_{HXB2} (Harrigan et al., 1998; Keulen, 2000), although the magnitude of the selective advantage of the double mutant under AZT pressure is probably much greater than the magnitude of its selective disadvantage in the absence of AZT. Hu and Temin (1990a) estimated the rate of recombination between co-packaged retroviral RNA genomes to be approximately 2% per kilobase per replication cycle. Considering the high virus titres obtained in these experiments, it is probable that such events unlinking the resistance mutations from the qPMA site occurred with some frequency. However, this cannot be demonstrated and is merely hypothetical. Using qPMA analysis to interrogate simultaneously the frequencies both of the 41L and 215Y mutants and of the qPMA site may shed light on this matter. Alternatively, a second qPMA site could be introduced at the 3' end of the protease domain, although cases of double recombination between the two sites would be undetectable. Using discordant frequencies at the two sites to calculate recombination frequency would be hampered by each individual culture system having a different frequency of doubly-infected host cells.
8. Relative Fitness Analysis of Patient RT Domains

8.1. Introduction

The ARS study of 11 HIV-1-infected individuals receiving AZT monotherapy described previously identified a subject ('A3') in which the virus displayed an unusual pattern of resistance mutations (Kaye, 1997). In samples taken during the 77 days of therapy, the frequency of known mutations within HIV-1 RT conferring AZT resistance adhered to a conventional pattern, with K70R appearing early, followed by low levels of T215Y (Boucher et al., 1992). However, once therapy was discontinued, rather than these codons reverting to their wild-type amino acids, both T215Y, and the previously undetected M41L mutation, increased in frequency. By day 192 (115 days post-therapy), the T215Y mutation was near fixation, with the frequency of M41L having climbed to 40% (FIGURE 8-1). This contrasts with repeated observations that viruses containing M41L and T215Y, either singly or together, have selective disadvantage compared to wild-type viruses and eventually disappear in the absence of AZT (Harrigan et al., 1998; Keulen, 2000).

In a previous study, AZT-resistance mutations persisted after the removal of AZT therapy, but in the context of continued therapy with other nucleoside analogues (Smith et al., 1994b). Low levels of cross-resistance between drugs of this class may account for this observation (Mayers et al., 1994). Even when reversion to wild type identities at AZT-resistance codons is observed, the rate of change is often much slower than for other resistant genotypes (Albert et al., 1992; Goudsmit et al., 1996, 1997; Lukashov et al., 2001).

Furthermore, when patient-derived AZT-resistant isolates are cultured without drug pressure, reversion at resistant loci is not always seen (Holguin et al., 1998). Upon removal of AZT, rather than simply returning to a pre-treatment threonine, 215-tyrosine
may evolve towards codons intermediate between tyrosine (TAY) and threonine (ACN), such as asparagine (AAY) and serine (TCY) (Goudsmit et al., 1996; de Ronde et al., 2001). In other instances, 215-tyrosine evolves away from threonine towards codons such as cysteine (TGY) and aspartic acid (GAY) (Goudsmit et al., 1997; de Ronde et al., 2001; Garcia Lerma et al., 2001). Many viruses displaying cysteine, aspartic acid or serine at codon 215 can replicate as efficiently as those with tyrosine at this position, although this is dependent upon the genetic background of the individual RT domain (Garcia Lerma et al., 2001).

Analysis of drug-resistant proteases has shown how discontinuation of drug pressure can result in continued evolution away from the pre-treatment sequence rather than reversion. Drug resistance mutations within protease had compromised enzyme function and consequently fitness. Restoration of fitness was through compensatory mutations elsewhere within protease, rather than reversion at resistance loci (Bacheler et al., 1998; Nijhuis et al., 1999).

It was speculated that within patient A3, additional genotypic markers were compensating for the debilitating effects of M41L and T215Y on RT function. As an initial step towards testing this speculation, the relative fitnesses of RT populations obtained from longitudinal samples taken from patient A3 were analysed. Rather than establishing mixed cultures at the point of transfection as was done in the previous chapter, individual recombinant virus stocks were generated and titrated prior to mixing. This allowed the level of infection to be controlled in an attempt to keep recombination frequencies to a minimum. Secondly, sequence analysis enabled the differences in relative fitnesses to be ascribed to genotypic changes other than those detected by earlier qPMA analysis.
FIGURE 8-1: Data from subject 'A3', showing RNA viral load pattern over time, together with PMA analysis of three AZT resistance-associated mutations (codons 41, 70 and 215) and duration of AZT therapy. Codons 67 and 219 were consistently wild-type (Kaye, 1997).
8.2. Methods

Plasma was separated from longitudinal blood samples obtained from subject A3. Total RNA was extracted by the Boom protocol (section A.8.4), and the 'REL' protocol used to generate five stocks of recombinant viruses containing the RT domain populations from these five A3 samples in a constant genetic backbone (section 5). These viruses were titrated (section A.17) and labelled rvA3-0^ through rvA3-192^, with the numerical suffix indicating the study day at which the blood sample was taken, and the 'A' indicating the nucleotide identity at the qPMA site (section 6). The RT population of each recombinant virus stock was subjected to automated sequence analysis according to the method described in section A.14.1.

Both one day prior to initiation, and on the day of initiation, a maintained culture of SupT-1 cells was harvested and resuspended in RPMI 10% FCS to a density of 0.1x10^6 ml^-1. Equivalent titres of two virus stocks were mixed (rvHXB2^, and one from rvA3-0^ through rvA3-192^). 40μl were removed and added to the first buffer of a QIAamp Viral RNA Mini Kit (section A.8.5), whilst the remainder was split, and added to duplicate wells of a 12-well Nunclon Delta Multidish containing 2ml of SupT-1 cell suspension. Cultures were mixed and incubated at 37°C in a humid 5% CO₂ atmosphere.

At various timepoints post-inoculation, 100μl of culture were removed and centrifuged at 6,000g for 30 seconds. 40μl of supernatant were removed and added to the first buffer of a QIAamp Viral RNA Mini Kit (section A.8.5). The sampled supernatant was replaced by 40μl of RPMI 10% FCS, and the 100μl sample mixed and returned to the dish. After a maximum of one week, the RNA extractions were completed and qPMA analysis performed as described in section 6.
8.3. Results and Analysis

To test the competition protocol, a competition between rvHXB2-41/215^A and rvHXB2-41/215^G was performed. The inoculum consisted of $7 \times 10^6\text{LD}_{50}$ (section A.17) of each virus. Supernatant samples were taken 0, 26, 50, 146 and 152 hours post-inoculation and analysed by qPMA. TABLE 8-1 shows the output from this experiment and the calculated frequencies of virus 'A' at each timepoint. FIGURE 8-2 displays these results graphically. The regression line has a positive gradient of $3.6 \times 10^{-4}\ln\%$ units per hour, but with a p-value of 0.17.

**TABLE 8-1: qPMA data from the reciprocal virus experiment.**

<table>
<thead>
<tr>
<th>Time</th>
<th>Raw Cpm Data 'A'</th>
<th>- Background 'A'</th>
<th>Normalised 'A'</th>
<th>% Frequency 'A'</th>
<th>ln (%A)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'A' control</td>
<td>11707 587</td>
<td>11052 0</td>
<td>11052 0</td>
<td>(100)</td>
<td>-</td>
</tr>
<tr>
<td>'G' control</td>
<td>655 3085</td>
<td>0 2498</td>
<td>0 11052</td>
<td>(0)</td>
<td>-</td>
</tr>
<tr>
<td>$t_0$</td>
<td>6190 1356</td>
<td>5535 769</td>
<td>5535 3402</td>
<td>61.9</td>
<td>4.126</td>
</tr>
<tr>
<td>$t_{26}$</td>
<td>7349 1465</td>
<td>6694 878</td>
<td>6694 3885</td>
<td>63.3</td>
<td>4.148</td>
</tr>
<tr>
<td>$t_{50}$</td>
<td>5252 1161</td>
<td>4597 574</td>
<td>4597 2540</td>
<td>64.4</td>
<td>4.165</td>
</tr>
<tr>
<td>$t_{146}$</td>
<td>8505 1604</td>
<td>7850 1017</td>
<td>7850 4500</td>
<td>63.6</td>
<td>4.153</td>
</tr>
<tr>
<td>$t_{152}$</td>
<td>5090 1060</td>
<td>4435 473</td>
<td>4435 2093</td>
<td>67.9</td>
<td>4.218</td>
</tr>
</tbody>
</table>
The data given below in TABLE 8-2 and 8-3 were taken from three competition experiments between rV3-0^A and rVHXB2^G. In two of the three experiments, the same inoculum was used to infect parallel SupT-1 cell cultures, whereas the third was an unduplicated experiment. As FIGURE 8-3 shows, over the course of each experiment there was an increase in the frequency of rV3-0^A. The percentage frequencies were transformed logarithmically (base e) and subjected to linear regression with time as the independent variable. The gradient of the increase was similar in the unduplicated experiment and one of the duplicated ones (0.2166 vs. 0.2104 ln % per day), but the other assay of the duplicated pair gave a more shallow gradient (0.0589 ln % per day).

Further competition experiments were performed, between individual recombinant ‘A’ viruses derived from A3 samples and rVHXB2^G. RNA extracts were taken from the competitions at two timepoints and analysed by qPMA. In each experiment, by dividing the loge of the change in percentage frequency of the A3-derived virus by the length of time between the two timepoints, ‘rates of change of virus frequency’ were calculated. For each A3 recombinant, the mean of the rates was taken, together with the standard error of the mean. TABLE 8-4 presents the results, and FIGURE 8-4 graphically...
demonstrates how the rates of change varied between the timepoints of the A3 samples from which the recombinant viruses were constructed. Virus rvA3-0A had a positive rate of change of frequency against rvHXB2G, in contrast to the recombinant virus derived from the day seven sample. With viruses derived from samples taken at subsequent timepoints (rvA3-7A through rvA3-192A), this rate gradually returned to a rate similar to that seen with rvA3-0A. Unlike the experiments described in section 7, competition experiments between recombinant viruses containing RT domains from clinical samples and the HXB2 reference strain were performed in replicate, enabling the inherent variability in the method to be recorded.

Application of two-tailed two-sample t-tests assuming unequal variances to each pairwise combination of gradient data sets presented in TABLE 8-4 generates the p values displayed in TABLE 8-5. The high variance amongst the rvA3-7A data set precludes determination of difference at 95% significance (unless one-tailed t-tests are performed instead). However, significant differences (p < 0.05) were found between the rvA3-0A data set and those for rvA3-77A and rvA3-125A. Significant differences (p < 0.01) were observed between the rvA3-192A data set and both the rvA3-77A and rvA3-125A data sets.

FIGURE 8-5 shows the correlation between HIV-1 plasma viral load and calculated relative RT fitnesses for the five samples taken from patient A3. Strong positive correlation was observed, with a correlation co-efficient (Pearson’s r) of 0.886, and p = 0.223.
TABLE 8-2: *qPMA* data and analysis from parallel secondary competition assays competing rvA3-0* against rvHXB2G.

<table>
<thead>
<tr>
<th>Raw Cpm Data</th>
<th>- Background</th>
<th>Normalised</th>
<th>% Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘A’</td>
<td>‘G’</td>
<td>‘A’</td>
<td>‘G’</td>
</tr>
<tr>
<td>‘A’ control</td>
<td>4639</td>
<td>55184</td>
<td>54465</td>
</tr>
<tr>
<td>‘G’ control</td>
<td>719</td>
<td>17787</td>
<td>0</td>
</tr>
<tr>
<td>Day 2</td>
<td>23188</td>
<td>4639</td>
<td>22469</td>
</tr>
<tr>
<td>Day 3</td>
<td>29765</td>
<td>5452</td>
<td>29046</td>
</tr>
<tr>
<td>Day 6</td>
<td>48085</td>
<td>4055</td>
<td>47366</td>
</tr>
<tr>
<td>Day 2</td>
<td>29638</td>
<td>18473</td>
<td>28919</td>
</tr>
<tr>
<td>Day 3</td>
<td>12861</td>
<td>6354</td>
<td>12142</td>
</tr>
<tr>
<td>Day 6</td>
<td>47446</td>
<td>4420</td>
<td>46727</td>
</tr>
</tbody>
</table>

TABLE 8-3: *qPMA* data and analysis from a secondary competition assay competing rvA3-0* against HXB2G.

<table>
<thead>
<tr>
<th>Raw Cpm Data</th>
<th>- Background</th>
<th>Normalised</th>
<th>% Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘A’</td>
<td>‘G’</td>
<td>‘A’</td>
<td>‘G’</td>
</tr>
<tr>
<td>‘A’ control</td>
<td>617</td>
<td>81932</td>
<td>81416</td>
</tr>
<tr>
<td>‘G’ control</td>
<td>5101</td>
<td>516</td>
<td>0</td>
</tr>
<tr>
<td>Day 2</td>
<td>11791</td>
<td>135887</td>
<td>135371</td>
</tr>
<tr>
<td>Day 3</td>
<td>7892</td>
<td>168421</td>
<td>167905</td>
</tr>
<tr>
<td>Day 6</td>
<td>1204</td>
<td>139073</td>
<td>138557</td>
</tr>
</tbody>
</table>
FIGURE 8-3: Graphical representation of the data presented in TABLE 8-2 and 8-3 with linear trend lines for each of the three series.
TABLE 8-4: Collected data from competition assays between A3 sample RT domains and HXB2 RT, showing the loge of the percentage change in frequency of the A3-derived virus per day over experimental periods of 5-7 days. Mean and standard error of the mean values for each of the five timepoint samples were calculated.

<table>
<thead>
<tr>
<th>vs. rvHXB2G</th>
<th>Ln % change per day</th>
<th>Mean</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>+0.1245</td>
<td>0.0383</td>
</tr>
<tr>
<td>rvA3-0^A</td>
<td>+0.2166</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0.0589</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0.0303</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0.2104</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0.1064</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rvA3-7^A</td>
<td>-0.1507</td>
<td>-0.3068</td>
<td>0.1335</td>
</tr>
<tr>
<td></td>
<td>-0.3686</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.6552</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.0530</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rvA3-77^A</td>
<td>-0.0321</td>
<td>-0.0317</td>
<td>0.0263</td>
</tr>
<tr>
<td></td>
<td>-0.0908</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0.0369</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.0408</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rvA3-125^A</td>
<td>+0.0782</td>
<td>-0.0206</td>
<td>0.0217</td>
</tr>
<tr>
<td></td>
<td>-0.0579</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.0178</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.0744</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.0301</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.0217</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rvA3-192^A</td>
<td>+0.1274</td>
<td>+0.1135</td>
<td>0.0328</td>
</tr>
<tr>
<td></td>
<td>+0.1037</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0.0909</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0.0400</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0.0551</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+0.2642</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 8-4: Graphical representation of the data presented in TABLE 8-4, with the day at which each sample was taken against the mean rate of change of virus frequency in competition with HXB2.
TABLE 8-5: p-values from two-tailed two-sample t tests assuming unequal variances between each permutation of paired data sets from those presented in TABLE 8-4. Values below 0.05 are highlighted in bold type.

<table>
<thead>
<tr>
<th></th>
<th>rvA3-0</th>
<th>rvA3-7</th>
<th>rvA3-77</th>
<th>rvA3-125</th>
<th>rvA3-192</th>
</tr>
</thead>
<tbody>
<tr>
<td>rvA3-0</td>
<td>0.0531</td>
<td>0.0120</td>
<td>0.0165</td>
<td>0.8332</td>
<td></td>
</tr>
<tr>
<td>rvA3-7</td>
<td></td>
<td>0.1365</td>
<td>0.1247</td>
<td>0.9913</td>
<td>0.0551</td>
</tr>
<tr>
<td>rvA3-77</td>
<td>0.0182</td>
<td></td>
<td>0.7538</td>
<td>0.0086</td>
<td></td>
</tr>
<tr>
<td>rvA3-125</td>
<td></td>
<td>0.7538</td>
<td></td>
<td>0.0078</td>
<td></td>
</tr>
</tbody>
</table>

FIGURE 8-5: Scatter plot of HIV-1 RNA viral load and relative RT fitness, from the A3 patient samples – $r = 0.886$, $p = 0.023$. 
Using the method described in section A.14.1, sample RT domains were sequenced using four sense primers and four antisense primers to ensure each nucleotide was scanned by a minimum of two sequencing reactions. TABLE 8-6 shows the summarised RT sequencing data from each of the five A3 samples. Only codons where the sample differed from the HXB2 reference sequence are shown. The majority of codons within A3 RT domains remained constant for the duration of the study, and constitute the genetic background of the virus, whereas others varied during the study period and can be classed as mutations. It must be stressed that the method of sequencing used here analysed RT populations en masse, such that detection of polymorphisms was restricted to those where the minority nucleotide was present at a frequency of at least 33%.

Pol sequences from 68 clade B viruses were downloaded from the Los Alamos Sequence Database¹, aligned by Sequencher software, and compared to the A3 sequences. All codon identities considered to constitute the genetic background of the series of A3 viruses occur frequently in these clade B viruses, with the exception of serine-376, isoleucine-469 and arginine-558, which were present in two, three, and one isolate respectively. Of all mutations, glutamine-324 has been seen in only a single isolate from the Netherlands (U34604), serine-554 in four, and valine-375 in five isolates.

The sub-population of RT domains within rvA3-192A in which a serine was present at codon 215 would have been overlooked by the previous qPMA analysis. The point mutation analysis presented in (Kaye, 1997) used an antisense probe to interrogate the central nucleotide of this codon. ‘C’ in the template was taken to indicate wild-type (threonine), with ‘A’ and ‘U’ taken to indicate tyrosine (TAC) and phenylalanine (TTC) mutants respectively. However, the serine discovered by sequencing (TCC) would have been scored as a wild-type threonine by this method.

¹ http://hiv-web.lanl.gov/
With reference to codon 41, the sequencing data correlated well with previously obtained qPMA data. In the rvA3-192^ virus, the peak heights of ‘A’ and ‘C’ at the first nucleotide of codon 41 were similar, indicating 50% methionine and 50% leucine. Data obtained by qPMA suggested 60% methionine and 40% leucine. Less well correlated were the data regarding codon 70. Sequencing analysis revealed an even mix of lysine and arginine, persisting from the day 77 sample until the day 192 sample. Analysis by qPMA gave the frequency of arginine-70 as declining from 40% to below 20% in the same series of samples (FIGURE 8-1).

Finally, FIGURE 8-6 presents the three data sets of virus load, relative RT fitness, and gross RT amino acid sequence in a single graph, to illustrate the changing relationships between the three parameters.
TABLE 8-6: Summary of sequencing data from the A3 samples. Only RT codons varying from the HXB2 reference sequence are presented. Polymorphic amino acids reflect relative peak heights of ~50%.

<table>
<thead>
<tr>
<th>RT-codon</th>
<th>41</th>
<th>69</th>
<th>70</th>
<th>122</th>
<th>196</th>
<th>211</th>
<th>214</th>
<th>215</th>
<th>245</th>
<th>272</th>
<th>277</th>
<th>324</th>
<th>326</th>
</tr>
</thead>
<tbody>
<tr>
<td>HXB2</td>
<td>M</td>
<td>T</td>
<td>K</td>
<td>E</td>
<td>G</td>
<td>R</td>
<td>L</td>
<td>T</td>
<td>V</td>
<td>P</td>
<td>R</td>
<td>D</td>
<td>I</td>
</tr>
<tr>
<td>A3-0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>K</td>
<td>E</td>
<td>K</td>
<td>F</td>
<td>-</td>
<td>M</td>
<td>A</td>
<td>K</td>
<td>-</td>
<td>V</td>
</tr>
<tr>
<td>A3-7</td>
<td>-</td>
<td>T/A</td>
<td>-</td>
<td>E</td>
<td>E</td>
<td>K</td>
<td>F</td>
<td>-</td>
<td>M</td>
<td>A</td>
<td>K</td>
<td>-</td>
<td>V</td>
</tr>
<tr>
<td>A3-77</td>
<td>-</td>
<td>-</td>
<td>K/R</td>
<td>E</td>
<td>E</td>
<td>K</td>
<td>F</td>
<td>-</td>
<td>M</td>
<td>A</td>
<td>K</td>
<td>-</td>
<td>V</td>
</tr>
<tr>
<td>A3-125</td>
<td>-</td>
<td>-</td>
<td>K/R</td>
<td>E</td>
<td>E</td>
<td>K</td>
<td>F</td>
<td>-</td>
<td>M</td>
<td>A</td>
<td>K</td>
<td>-</td>
<td>V</td>
</tr>
<tr>
<td>A3-192</td>
<td>M/L</td>
<td>T/A</td>
<td>K/R</td>
<td>E</td>
<td>E</td>
<td>K</td>
<td>F</td>
<td>Y/S</td>
<td>M</td>
<td>A</td>
<td>K</td>
<td>D/E</td>
<td>V</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RT-codon</th>
<th>375</th>
<th>376</th>
<th>390</th>
<th>400</th>
<th>460</th>
<th>461</th>
<th>468</th>
<th>469</th>
<th>483</th>
<th>519</th>
<th>540</th>
<th>554</th>
<th>558</th>
</tr>
</thead>
<tbody>
<tr>
<td>HXB2</td>
<td>I</td>
<td>T</td>
<td>K</td>
<td>T</td>
<td>N</td>
<td>R</td>
<td>T</td>
<td>L</td>
<td>Y</td>
<td>N</td>
<td>K</td>
<td>A</td>
<td>K</td>
</tr>
<tr>
<td>A3-0</td>
<td>-</td>
<td>S</td>
<td>R</td>
<td>A</td>
<td>D</td>
<td>K</td>
<td>S</td>
<td>I</td>
<td>H</td>
<td>S</td>
<td>-</td>
<td>-</td>
<td>K/R</td>
</tr>
<tr>
<td>A3-7</td>
<td>-</td>
<td>S</td>
<td>R</td>
<td>A</td>
<td>D</td>
<td>K</td>
<td>S</td>
<td>I</td>
<td>H</td>
<td>S</td>
<td>K/E</td>
<td>-</td>
<td>K/R</td>
</tr>
<tr>
<td>A3-77</td>
<td>I/V</td>
<td>S</td>
<td>R</td>
<td>A</td>
<td>D</td>
<td>K</td>
<td>S</td>
<td>I</td>
<td>H</td>
<td>S</td>
<td>K/E</td>
<td>-</td>
<td>K/R</td>
</tr>
<tr>
<td>A3-125</td>
<td>I/V</td>
<td>S</td>
<td>R</td>
<td>A</td>
<td>D</td>
<td>K</td>
<td>S</td>
<td>I</td>
<td>H</td>
<td>S</td>
<td>K/E</td>
<td>A/S</td>
<td>K/R</td>
</tr>
<tr>
<td>A3-192</td>
<td>I/V</td>
<td>S</td>
<td>R</td>
<td>A</td>
<td>D</td>
<td>K</td>
<td>S</td>
<td>I</td>
<td>H</td>
<td>S</td>
<td>K/E</td>
<td>A/S</td>
<td>K/R</td>
</tr>
</tbody>
</table>
FIGURE 8-6: The relationship between relative fitness of RT populations in patient A3 and RNA load over the 192 day study period, with sequence changes from the baseline sample highlighted (changes to codons associated with AZT resistance are emboldened).
8.4. Discussion

Recombinant HIV-1 stocks were generated in which reverse transcriptase (RT) populations derived from serial samples taken from a single subject participating in an AZT monotherapy trial were inserted into an isogenic HIV-1 background (HIV-1_{HXB2}). Multiple competition experiments between these and a reference stock (HIV-1_{HXB2} RT in the same background) were performed, and the relative fitnesses of each recombinant calculated. Sequence analysis of RT domains elucidated both the pre-treatment genetic background of the sample viruses and the genotypic changes that occurred within the population sequence during the course of the trial. Significant differences in relative fitness were observed between RT populations taken from the subject at different timepoints throughout the trial, as were several sequence changes. The isogenic genetic background of the recombinant viruses allows these differences to be ascribed solely to the differences in their reverse transcriptase domains.

As the sequencing was performed on a population level, a number of codons exhibited mixtures of amino acid identities. A lack of clonal analysis prevented both analysis of individual genotypes within the populations, and investigation into linkage between codon identities where two or more polymorphisms were detected in a single population sequence. Obtaining a different representation of the RT genotype population by examining many individual sequences per timepoint would have required the cloning and sequencing of a sufficiently large number of RT domains to be considered impractical in this project. Inevitably, this restricts the conclusions that can be drawn. Nevertheless, several hypotheses are proposed which can be addressed using our methods.

The relative fitnesses of sampled RT domains paralleled the plasma viral load, although mutations conferring high-level resistance to AZT were not observed until after therapy had been discontinued and HIV-1 RNA load had returned to the baseline level. Other workers have observed a similar pattern (Loveday et al., 1995), and one group has developed a
mathematical model of the viral dynamics under these conditions. In this model, the initial drop in HIV-1 RNA load as a result of initiating AZT therapy is accompanied by a rise in CD4⁺ T cell count. After the initial virus load nadir, this increased target cell population allows the incompletely-suppressed virus population to increase (de Jong et al., 1996). This rise is thus independent of resistance or fitness, although the evolution of both phenotypic parameters would be affected by the changing viral replication kinetics.

It is proposed that the changes in RT observed during the first seven days of AZT therapy were primarily due to genetic drift. At least two groups have shown how a reduction in the \textit{in vivo} HIV-1 population due to antiretroviral therapy reduces intra-sample nucleotide distances, constituting a genetic bottleneck event (Nijhuis et al., 1998; Ibanez et al., 2000). The prediction that genetic bottlenecks reduce the fitness of a population through the Muller's ratchet effect is a direct consequence of the mean fitness of a population being lower than the fitness of the mean population genotype (Fisher, 1999). This has been demonstrated with experimental RNA virus populations (Duarte et al., 1994b), and work with experimental HIV-1 populations has confirmed fitness loss upon genetic bottlenecking (Yuste et al., 1999, 2000). Additionally, the application of AZT would have distorted the fitness landscape such that the optimally fit genotype around which the pre-treatment quasispecies was clustered was itself rendered unfit.

As genetic drift is an essentially random process, mutations within the post-bottleneck virus populations will be distributed throughout the HIV-1 genome (Yuste et al., 2000). Fitness loss can be incurred through mutation at any locus, not simply the RT domain. Hence, although reverse transcriptase is the viral target of AZT, it will contribute only a fraction towards the overall fitness losses incurred by genetic drift. This may also explain the lack of significance of the fitness change between days zero and seven.

To ascertain whether the magnitude of the population reduction described in this thesis was sufficient to constitute a genetic bottleneck would require a thorough analysis of the quasispecies composition both before and after therapy. The scale of this task was beyond
the scope of this investigation. However, some insight can be gained from the gross sequence changes during the first seven days of therapy. T69A confers low-level resistance to AZT if introduced into HIV-1\textsubscript{NL4-3} (Winters and Merigan, 2001), which may have contributed to its survival at this timepoint. On the other hand, the two other sequence changes (K122E and K540K/E) are not thought to confer selective advantage. Codon 122 of RT appears to be naturally polymorphic in clade B viruses, with both lysine and glutamic acid equally represented (Human Retroviruses and AIDS, 1999; Shafer et al., 2000). This change, and that at codon 540 in the p66 RNAse H domain from lysine to glutamic acid in half of the virus population between days zero and seven, would appear to be a random event due to sampling effects during a bottleneck. Little work has been performed on specific mutations at the 3' end of the RT domain, and understanding of the role of mutations such as K540E is limited. Mutation immediately surrounding this codon has abrogated RNAse H activity and inhibited RT function (Tisdale et al., 1991), possibly due to its role in the formation of p66/p51 heterodimers (Restle et al., 1992).

The day zero RT population was consistently fitter than HIV-1\textsubscript{HXB2} RT, which in turn was consistently fitter than the day seven RT population. However, no significance could be attached to the detected differences in RT fitness between the first two timepoints. This is attributable to the high standard error of the measurements of the day seven sample, believed to be a result of the reduced population size in the day seven sample leading to sampling errors. The high variation of the relative fitness parameters would be explained by each value being obtained from a different population genotype. The efficiency of the nucleic acid extraction and amplification methodologies may not have been sufficient to ensure full and accurate population representation in each recombinant virus population. Unfortunately, time constraints precluded sequence analysis of each competed day seven recombinant virus stock, as this would have been useful in investigating this hypothesis.

The model of de Jong \textit{et al.} (1996) explains how the rising plasma virus load between days seven and 77 mimics host-parasite dynamics such that a suppressed HIV-1 population (the parasite) can recover in number through increased availability of CD4\textsuperscript{+} T cells (the host).
During this period, Fisher's evolutionary model (Fisher, 1999) would predict steady gains in fitness – increased replication leads to progressive increases in both population size and genetic variance, the substrate of natural selection. Again, this variance will be represented throughout the HIV-1 genome, but due to the selective pressures of AZT treatment, changes in reverse transcriptase are likely to account for a disproportionate fraction of fitness difference.

Once more, the large standard error of the measured day seven RT relative fitness parameters inhibits statistical comparison of RT fitnesses between days 7 and 77. The slightly increased resistance of the K70R mutation may have led to its selection at the expense of T69A (4-8-fold over 5-fold, (Larder and Kemp, 1989; Larder et al., 1991; Winters and Merigan, 2001)), but the crude tool of population sequencing failed to establish linkage between mutations within population subsets. If, at both timepoints, T69A and K70R had been linked to either the K540 or E540 subsets, the situation may have become clearer. Likewise, linkage between these mutations and I375V could not be established.

The K70R mutation may have been the critical genotypic change causing the significant difference in fitness between baseline day zero RT populations and those from the day 77 sample. Site-directed mutagenesis of HIV-1\textsubscript{HXB2} has demonstrated that K70R confers an approximate loss in fitness of between 3% and 8% in that genetic backbone under drug-free conditions (Harrigan et al., 1998; Keulen, 2000). However, the \emph{in vivo} fitness landscape is likely to have an inverted relationship, with resistant virus having increased RT fitness in the presence of AZT over baseline virus. Recent studies have found that K70R has less effect on AZT susceptibility that the 4-8-fold reduction originally suggested (Sharma et al., 1996; Keulen, 2000) suggesting its selective advantage in the presence of AZT may not be as profound as originally thought. Together, insufficient resistance to AZT and impaired RT function may not confer sufficient replicative advantage to R70 mutants compared to wild-type genotypes to result in their fixation. Again, the effects of other RT mutations in compensating for the K70R fitness deficit cannot be inferred from these data, as linkage
between polymorphic codons was not established. However, studies into the relative replicative capacities of the two codon 70 genotypes under varying levels of AZT pressure and their resistance profiles may shed light on this matter, particularly in the context of this RT background.

The persistence of K70R beyond the cessation of AZT therapy, and its presence in only a subset of the virus population suggests that the fitness deficit realised upon the removal of AZT does not confer sufficient selective disadvantage for its rapid extinction. Although little is known about the magnitude of non-lethal mutations in the p66 RNase H domain, it is thought unlikely that the A554S mutation could be responsible for stabilising the K70R mutant, although yet again, a lack of clonal analysis has prevented us from determining which mutations are linked. That the plasma viral load rose sharply between days 77 and 125 probably reflects the relaxation of AZT inhibition as the patient was taken off therapy at day 77.

Fisher’s model implies the narrower mutational reach of smaller populations is liable to exacerbate the existing predisposition towards ‘simple’ mutants. As the frequency of mutations tends to obey the Poisson distribution (Leitner and Albert, 1999), the distribution of mutants within a quasispecies should be weighted towards single mutants, with the frequency of members harbouring multiple mutants declining progressively with the number of mutations. HIV-1 reverse transcriptase is biased towards transitions (A→G or C→T), and the uneven intracellular distribution of the four nucleotides pushes this bias in favour of purine transitions, and G→A mutations in particular (Mansky and Temin, 1995; Vartanian et al., 1997; Keulen et al., 1997a). This tendency is reflected in the array of mutations encountered over the first 77 days of this study period (TABLE 8-7), where all of the six changes were A→G transitions. Contrast these with the transversion mutations (A/G→C/T) observed in the samples from day 125 and day 192 (TABLE 8-8). Of these, only A554S was observed in the day 125 sample – the remainder were undetected until day 192, when the plasma viral load had risen almost to the pre-treatment level.
Earlier, more sensitive qPMA data (FIGURE 8-1 and Loveday et al., 1995; Kaye, 1997) suggested that from 100% T215 at day 7, tyrosine increased in frequency in an essentially linear fashion, to approximately 18% by day 125. These data conflict with the proposition that the frequency of double transversion mutants (such as T215Y) in small virus populations such as those at days 7 and 77 is very low. ‘Blips’ of transient virus replication observed in patients undergoing HAART are associated with the development of resistance (Cohen Stuart et al., 2001). Temporary expansion of the virus population would increase quasispecies diversity, allowing the production of more complex mutants with favourable resistance or fitness phenotypes. It is feasible that between the timepoints for which data is available, such an event or events may have occurred in the plasma virus load of patient A3.

An alternative explanation is that the reading of the qPMA data as a T215Y mutation was based upon the central nucleotide of codon 215 being A. The presumption was that an adenine would be found only where the flanking nucleotides were consistent with tyrosine (TAC) (Kaye, 1997). This interpretation is perhaps unwarranted considering the recent reports of amino acids other than T, Y or F being observed at codon 215 (Goudsmit et al., 1996, 1997’ de Ronde et al., 2001). If the first nucleotide of codon 215 remained wild-type, an asparagine would be encoded (ACC—> AAC). This would require just a single transversion event, and would not have been detected by the earlier methods. Asparagine at codon 215 was previously observed as an intermediate evolutionary stage between tyrosine and aspartic acid (de Ronde et al., 2001). An inverse situation could be at work here, with asparagine representing an intermediate between threonine and tyrosine – stochastic evolution early in the study may have allowed the single mutant to emerge, prior to the appearance of the double transversion mutant when the population size had increased.

Although AZT-resistant isolates can retain their resistance mutations upon removal of therapy (Smith et al., 1994b), it remains to be explained why M41L/T215Y variants were selected in the absence of drug pressure. Furthermore, the earlier qPMA data suggests that the M41L mutation was present at a frequency half that of T215Y, suggesting that a subset
of genotypes harbouring 215Y has increased fitness in the absence of M41L. *In vitro* studies have calculated that in an HIV-1\textsubscript{HXB2} genetic background, M41L alone confers a fitness loss of between 13% and 20%, T215Y alone confers a 5%-15% deficit, and the double mutant has a relative fitness 7%-15% below wild-type (Harrigan et al., 1998; Keulen, 2000). It would seem that the genetic background of the A3 reverse transcriptase domains not only abrogated the fitness deficits commonly associated with these resistance mutations, but imbued a selective advantage to the double mutants. Although M41L and T215Y are considered primary AZT resistance mutations, they seem to be acting in a compensatory role, analogous to the evolution of protease under ritonavir therapy reported by Nijhuis et al. (1999). Here, the resistance-associated mutations K20R and I71V in protease restored viral fitness that had been diminished by earlier resistance mutations. However, unlike earlier observations (Molla et al., 1996; Lawrence et al., 1999), these secondary mutations did not increase phenotypic resistance to ritonavir.

The presence of a 215S (TCC) sub-population at day 192 could relate to previous observations where upon removal of drug pressure, AZT-resistant 215Y genotypes evolved serine rather than reverting to threonine (Goudsmit et al., 1997, 1998; Yerly et al. 1998; de Ronde et al., 2001; Garcia Lerma et al., 2001). Both *in vivo* and *in vitro* fitness calculations suggest that 215S variants are more fit than those with 215Y, and equal to those with 215T. However, the overall RT fitness of 215 mutants can be modulated by the genetic background of their RT domains (Garcia Lerma et al., 2001). This further confuses the picture. Mutations thought to confer decreased fitness emerged concomitantly with an increased relative fitness of the RT population. However, serine was observed at codon 215 – consistent either with post-treatment evolution, or with post-treatment reversion to a genotype intermediate between AZT-sensitive and AZT-resistant. Neither the order in which these mutations appeared, nor any linkage between them can be established from this data set, but it would be very interesting to interrogate further the day 192 sample.

In summary, the data collected chart the changes in relative RT fitness in a patient undergoing AZT monotherapy. Correlation was demonstrated between plasma virus load
and relative RT fitness. The predicted bottleneck-mediated loss of fitness immediately after the onset of therapy was observed, but also exposed the limitations of the system, as significance could not be confirmed. After an initial drop, fitness was gained in a step-wise fashion over 185 days, consistent with Fisher’s geometric evolutionary model. Fitness changes were paralleled by genotypic changes, but again, the population sequencing used here (due to time constraints) meant the interpretation was restricted to hypotheses.

A complex picture was revealed by the final sample, taken 115 days after the termination of therapy. RT fitness from this timepoint was significantly greater than from the preceding sample, concurrent with numerous genotypic changes. As with prior samples, the many codons were found to be polymorphic, indicating a mixed population. Particularly surprising was the presence of two AZT-resistance mutations, as they have been shown to reduce substantially RT fitness in other studies. It is likely that the genetic context in which these mutations were found inverted this fitness loss, enabling their selection in the absence of drug pressure.

The evolution of HIV-1 RT during drug therapy may involve a trade-off between fitness and resistance. During therapy, drug pressure is the predominant selective force on RT – upon cessation of therapy, selection in favour of improved RT function takes precedence. Repeated adjustments to the fitness landscape engendered by transient administration of therapy can impose restrictions on evolution – mutational pathways towards improved fitness can be modified by the genotypic changes induced by past selection. From the data obtained here, it would seem that the nature of the fitter genotypes detected at the end of the 192-day period were determined in part by intervening AZT selection.
TABLE 8-7: Expanded detail of the collected codon changes within RT observed in samples taken during the first 77 days of the study period.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Codon change</th>
<th>Nucleotide change</th>
</tr>
</thead>
<tbody>
<tr>
<td>T69A</td>
<td>ACT→GCT</td>
<td>A→G</td>
</tr>
<tr>
<td>K70R</td>
<td>AAA→AGA</td>
<td>A→G</td>
</tr>
<tr>
<td>K122E</td>
<td>AAA→GAA</td>
<td>A→G</td>
</tr>
<tr>
<td>I375V</td>
<td>ATA→GTA</td>
<td>A→G</td>
</tr>
<tr>
<td>K540E</td>
<td>AAA→GAA</td>
<td>A→G</td>
</tr>
</tbody>
</table>

TABLE 8-8: Expanded detail of the collected codon changes within RT observed in samples taken during the final 67 days of the study period.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Codon change</th>
<th>Nucleotide change</th>
</tr>
</thead>
<tbody>
<tr>
<td>M41L</td>
<td>ATG→CTG/TTG</td>
<td>A→C/T</td>
</tr>
<tr>
<td>T215Y</td>
<td>ACC→TAC</td>
<td>A→T &amp; C→A</td>
</tr>
<tr>
<td>T/Y215S*</td>
<td>ACC/TAC→TCC</td>
<td>A→T or A→C</td>
</tr>
<tr>
<td>D324E</td>
<td>GAC→GAA</td>
<td>C→A</td>
</tr>
<tr>
<td>A554S</td>
<td>GCT→TCT</td>
<td>G→T</td>
</tr>
</tbody>
</table>

* It is unclear whether the serines detected at codon 215 derive from wild-type threonines or from tyrosine intermediates.
9. General Discussion

Phenotypic and genotypic resistance to all licensed antiretroviral therapies have been observed in HIV-1 isolates (Pillay et al., 2000). However, in terms of replicative capacity (fitness), the cost to the virus of acquiring resistance is not well understood. *In vivo* studies have shown that mutations conferring resistance often revert to wild-type amino acids upon cessation of drug therapy (Land et al., 1991; Devereux et al., 1999; Devereux et al., 2001), and *in vitro*, introduction of resistance mutations into laboratory strains of HIV-1 demonstrably reduces their replicative capacities (Croteau et al., 1997; Harrigan et al., 1998; Keulen, 2000). These results support the theoretical concept that wild-type untreated viruses should occupy local fitness peaks on the sequence landscape.

However, in some cases, genotypic resistance mutations can persist long after discontinuation of the drug therapy by which they were selected (Albert et al., 1992; Smith et al., 1994b; Shafer et al., 1998), and some co-cultured resistant viruses have similar replication kinetics to wild-type strains (Schmit et al., 1996). Recombinant virus techniques similar to those used in this study have been used to compare the replicative capacities of recombinant viruses containing either wild-type or resistant protease domains, and the results suggest the evolutionary response of HIV-1 to protease inhibitor therapy has a two-stage pattern. Initially, primary mutations conferring resistance are selected, and once replication rate has recovered sufficiently, secondary mutations are selected which restore enzymatic functionality without necessarily increasing the level of resistance (Borman et al., 1996; Schock et al., 1996; Nijhuis et al., 1999).

Furthermore, the *in vivo* evolution of HIV-1 after stopping treatment can frequently diverge from simple reversion at mutant codons. These novel evolutionary pathways result in previously unreported codon identities at resistance-associated loci (Yerly et al., 1998; Garcia Lerma et al., 2001; de Ronde et al., 2001). A previous study (Loveday et al., 1995; Brun-Vezinet et al., 1997) identified a patient undergoing AZT monotherapy whose pattern
of development of resistance mutations was unusual (coded here as patient A3, see section 8, page 175-199). During the period of therapy, only a single mutation known to confer resistance to AZT was observed in the reverse transcriptase (RT) domain of circulating virions (K70R). Moreover, this mutation appeared only transiently and then only in a subpopulation of the plasma virus population. This early pattern has been fitted to a model of virus-host dynamics (de Jong et al., 1996). The subsequent evolution of variants harbouring RT mutations conferring high-level resistance to AZT was expected both from previous observations and from more recent characterisation of the intermediate genotypes (Boucher et al., 1992; Jeeninga et al., 2001). However, the selective advantage of these mutations was believed to be conditional upon continued AZT selection pressure (Harrigan et al., 1998; Keulen, 2000). In patient A3, outgrowth of AZT-resistant variants was observed only after AZT therapy had ceased. This thesis set out to address the question of how variants harbouring these deleterious mutations could emerge under conditions where they were thought to have selective disadvantage. Our hypothesis was that the emergence of these mutations was accompanied by increased fitness. This could indicate that rather than universally decreasing fitness, these mutations could increase fitness in an RT genotype-dependent manner.

An in vitro recombinant virus approach was taken in order to study RT phenotypes independently of other sample virus attributes. The work comprised two stages. Firstly, a system was developed whereby populations of sample RT domains could be inserted into an isogenic HIV-1 background. Secondly, pair-wise competition experiments between these viruses and a reference strain were performed. The isogenic virus background enabled differences in replication kinetics between viruses to be ascribed to the differences in RT domains.

In the first stage, persistent difficulties led to four separate methodologies being attempted before the establishment of a successful system. Propagation of plasmids in E. coli presented the most insuperable problems. Three different bacterial strains were employed, and although each was designed specifically for the maintenance of difficult plasmid
sequences, re-arrangement of plasmid sequences was consistently observed. One reason for these observations may have been the nature of the plasmids themselves. Peden (1992) reported how the source of a plasmid's origin of replication was critical to the stability of HIV-1-derived inserts, independently of the presence or absence of the bacterial recA system. In this thesis, plasmids were derived either from pHXB2ART, or from Promega's pGEM-T Easy Vector. Retroviral repeat sequences are particularly prone to instability in the pGEM series of plasmids. However, the effect was also observed in pHIVEC, the adapted pHXB2ART vector. Repeated propagation of pHXB2ART for co-transfection never resulted in plasmid re-arrangement, unlike RT-inserted pHIVEC (FIGURE 2-17). Furthermore, the inserts in the pGEM-T Easy Vector only comprised HIV-1 Pol domains. Unlike the twin LTRs, these are not considered substrates for intrabacterial plasmid recombination (Yamada et al., 1995; Rosin-Arbesfeld et al., 1998). The reasons for the re-arrangements remain unclear.

Another puzzling phenomenon was the failure to PCR-amplify from RT-inserted plasmids using the primer pair S-HR/AS-HR. Used in two of the methods, these primers were targeted at vector sequences and used to amplify across the RT inserts in order to generate transfectable material overlapping the termini of linearised pHXB2ART. The design of each system ensured that their binding sites were present with 100% homology, so mispriming should not have occurred. Outside of positive PCR control reactions, product was generated only once – from a colony shown by screening PCRs to contain vector with correctly ligated insert (FIGURE 3-15). Amplification from gel purified plasmids obtained by bulk transformation was otherwise unsuccessful (FIGURES 3-9, 3-10, and 4-17), despite the presence of the appropriate binding sites being demonstrated by screening PCRs.

PCR amplification of sample RT domains from RNA extracted from plasma presented yet another set of difficulties. In particular, poor primer design prevented optimal PCR performance. Until the HIV-1 sequence alignment database 'Human retroviruses and AIDS' (1999) became available, the diversity of potential target sequences was not
considered, leading to repeated mismatching between the 3'-terminal primer nucleotides and the template. One set of second round PCR primers in particular was not only mismatched at the 3'-terminal nucleotide position, but in such a way that strand extension efficiency was maximally abrogated (Huang et al., 1992). This was rectified in the final system by the design and application of appropriate primers.

Other considerations of primer design such as relative base composition, secondary structure, equal melting temperatures and repetitive bases were ignored, as the locations of the primers were dictated by the inflexibility of their prospective positions within HIV-1. Similar arguments led to consideration of codon usages being ignored. Multiple codons encode the same amino acids, and organisms can differ greatly in the relative frequencies by which each is employed within their genomes. The mutational bias of the HIV-1 replication machinery favours disproportionate use of adenine-rich codons (Berkhout and van Hemert, 1994). On several occasions in this project (see FIGURES 2-4, 2-5, 4-2, 5-2, or 5-5), the introduction of restriction sites by 'silent' mutagenesis necessitated an alternative codon at an amino acid locus. This is unlikely to have disrupted virus replication by forcing the use of a tRNA isoacceptor whose availability was limited, as all new codon identities are found elsewhere in the HXB2 genome.

In conclusion, the design of the REL system bypassed these obstacles by removing the need for bacterial propagation of plasmids, changing the primers used in the ‘final’ PCRs, and paying more regard to sequence variation when designing the primers used to amplify from sample sequences. Furthermore, the REL system represents an improvement upon others previously used to investigate fitness by incorporating the entire RT domains from sample virus populations into the recombinant viruses, including the entire RNAse H domains omitted by other methods (Shi and Mellors, 1997; Martinez-Picado et al., 1999). Many workers have established that resistance-associated mutations within the 5' terminus of the RT domain can not only affect the enzyme's polymerase function, but also impact upon the efficiency of RNAse H activity (Fan et al., 1996; Gerondelis et al., 1999; Archer et al., 2000). These effects are thought to occur through alterations in the juxtapositions of RT
domains and their relevant nucleic acid templates (Gao et al., 1998). It has also been shown that mutating amino acids in the RNAse H domain of HIV-1 can disrupt both the polymerase activity of RT (Tisdale et al., 1991) and proteolytic processing of the Gag-Pol precursor polyprotein (Mizrahi et al., 1994). Consequently, it was felt important to incorporate the entire sequence of the sampled RT domains into recombinant viruses, rather than looking only at codons surrounding the loci at which resistance mutations are found.

Additionally, no sequences flanking RT were incorporated, allowing all phenotypic differences between recombinants to be ascribed exclusively to differences in their RT domains. Systems based upon the recombinant virus assay of Kellam and Larder (1994) do not allow such conclusions, as varying and often undefined lengths of sample-derived protease and integrase sequences are incorporated into recombinants (Boucher et al., 1996; Nijhuis et al., 1997a; Hertogs et al., 1998; Walter et al., 1999).

Each system for generating recombinant viruses had to take into consideration how downstream pair-wise competition assays would be monitored. It was decided that a quantitative point mutation assay (qPMA), originally designed to measure nucleotide frequencies at resistance codons within RT, would be adapted to this end (Kaye et al., 1992). Each virus generated using the REL system possessed either an A or a G at a specified nucleotide position within their reverse transcriptase domains, although as stated in section 5.5, in retrospect it would have been preferable to place the variable locus in the protease or integrase flankers. This would both avoid the expense of amplifying RT domains twice, and possibly improve PCR efficiency by removing a mismatch in one of the second round antisense PCR primers (AS-MT3B, page 142). Further optimisation of the various steps in the REL protocol is also needed.

Modifications to the original qPMA design were required both to remove the radioactive component and to preclude erroneous measurements due to adjacent polynucleotide sequences. Inter-experimental variation in signal maxima was unexpectedly high, but after normalisation using controls containing monotypic nucleotide identities at the interrogated
locus, the system was shown to be robust and to measure nucleotide frequencies accurately. The sensitivity and resolution of this system enables more accurate monitoring than methods involving measurement of the relative peak heights from sequencing analysis.

The relative expense of the qPMA may favour an alternative system for the measurement of dimorphic nucleotides. Work subsequent to this thesis has successfully employed a non-isotopic adaptation of the densitometry method of Kidd et al. (2000). In this method, the qPMA site is substituted by a silent mutation that introduces or destroys a restriction site. Digestion of RT-PCR products amplified from RNA extracted from competition experiments generates mixtures of digested and undigested bands. The relative frequencies of each virus are represented by the relative intensities of each set of bands. These are easily measured using a digital camera and image analysis software. As with the qPMA, internal controls have been included, to enable normalisation in cases of partial digestion.

Several other options exist for the monitoring of mixed virus populations in competition experiments. Use of the Ligase Chain Reaction (LCR) cannot detect low frequencies of one of two variants (Kalin et al., 1992), but a gap-LCR can detect a minority variant up to a ratios of 1:10^5 (Abravaya et al., 1995). Recently, Lu and Kuritzkes (2001) have developed a sophisticated reporter gene assay. This latter system uses two vectors, one for each competitor. In each vector, the Nef gene has been replaced by a different enzyme, whose substrates undergo colour change upon catalysis. Quantifying the relative degree of colour change is used to monitor the relative amounts of each virus.

Degenerate heteroduplex tracking assay in which variation between probe and target leads to differential migration through a gel is inappropriate here, as only single nucleotide changes are present (Delwart et al., 1994). Reducing the complementarity of either the probe or the target may help, but maintaining as much identity within the invariant recombinant sequence as possible is preferable, in order to avoid vector-specific fitness contributions (Quinones Mateu et al., 2000; Resch et al., 2001). Line Probe technology is too crude, being no better than sequencing at detecting low frequencies in small populations.
General Discussion

(Descamps et al., 1998). Real-time PCR as described by Hance et al. (2000) may prove useful, although this technique is designed for the detection of mutant species at extremely low frequency within a mixture (as low as 0.1%). It is uncertain whether mixtures of two viruses whose frequencies differed by larger percentages would be accurately scored by this method, although the results from Hance et al. suggest this method may be as good, if not better, than the qPMA described here.

In the second phase of the project, competition experiments between viruses differing at this locus were monitored over time and the changing frequencies of each nucleotide were used to calculate relative fitness parameters. Two sets of experiments were performed with differing experimental conditions. In the first set, HIV-1\textsubscript{HXB2} RT containing the AZT resistance mutations M41L and T215Y was competed against unmutated HIV-1\textsubscript{HXB2}. Competitions were initiated at the point of transfection – approximately equal quantities of PCR products containing each RT domain were co-transfected with the isogenic HIV-1\textsubscript{HXB2} backbone into the competition culture. This contrasted with the protocol of the second set of competitions, where individual virus stocks were constructed by the REL system, titrated, and approximately equal amounts of each whole virus were competed.

The first competitions failed to detect a fitness differential between HIV-1\textsubscript{HXB2 (41L/215Y)} and HIV-1\textsubscript{HXB2} itself. Of the reasons suggested for this failure (insufficient virus growth, reversion, and/or recombination), recombination between the two RT genotypes was considered the most likely. It is thought that the frequency of recombination is related to the frequency of virions containing heterologous RNA molecules. In turn, this is believed to be proportional to the frequency of cells infected with multiple viruses. Consequently, cultures with reduced levels of infection should exhibit reduced rates of recombination. The level of infection was not well controlled in the first set of competitions, but the use of low multiplicities of infection and regular splitting of cultures in the second set may well have maintained sufficiently low titres of virus. Employing a tandem qPMA site upstream of RT would help to monitor levels of recombination and assist in the development of a protocol minimising such events.
In the second series of competition experiments, RT domains from a longitudinal series of samples taken from a patient undergoing transient AZT therapy were amplified and used to generate recombinant virus stocks. A series of replicated competitions between these viruses and one containing the RT domain of HIV-1HXB2 was performed. For each timepoint, the RT population was assigned a relative fitness parameter. Statistical analysis of these values attached significance to some fitness differences between timepoints, but wide variation amongst the fitness parameters obtained from the sample taken seven days into therapy precluded the observed changes from reaching acceptable levels of statistical significance.

By amplifying and transfecting RT populations from each timepoint, each recombinant virus stock contained a heterogeneous array of viruses. Sequencing the population of RT domains from each timepoint enabled detection only of gross genotypic changes. It was shown that fitness increased concurrently with the acquisition of M41L and T215Y by a significant proportion of the population (>33%). It was suggested that mutations that normally reduce RT functionality, in the context of a particular RT genotype, could actually restore functionality to a debilitated enzyme. However, many codons at a given timepoint exhibited mixed identities, and linkage between mutations could not be ascertained. Specific data regarding the precise genotypic context in which these two AZT resistance mutations were found remained elusive.

The methods designed and employed in this thesis seem well suited to addressing the questions raised by these data, although the putative recombination problem discussed above needs resolving. Furthermore, as Maree et al. (2000) recently explained, an additional variable in competition experiments is the overall rate of virus replication. Small inter-experimental differentials in overall virus growth kinetics can lead to significant miscalculations of fitness. Serial calculations of competitor frequency should account for the number of intervening cycles of replication in order to obtain meaningful co-efficients of selection. Replication rates in culture should be monitored by any assay that returns a quantitative measure of virus, such as p24 antigen ELISA, quantitative RT-PCR, or a
reverse transcriptase activity assay. After publication of the paper by Mareé, we intended to use a p24 antigen ELISA. However, our ELISA was unreliable; the time constraints precluded optimisation and due to the timepoint nature of the experiments, the opportunity to obtain the requisite samples was lost. However, measures to reduce virus levels in culture will necessitate a virus quantification system that has high sensitivity and high resolution at low titres. The variable RT domains within the recombinant virus competitors would invalidate use of reverse transcriptase assays, but a PCR-based quantification system such as the Enzyme-Linked Oligonucleotide Assay (Whitby and Garson, 1995) or a Real-Time quantitative PCR system may prove more useful, particularly as large regions of the competitor viruses’ genomes will be identical.

Even once these experimental procedures have been optimised, the question remains as to whether data generated by this system has relevance to the in vivo situation. The importance of Gag cleavage site mutations in resistance to protease inhibitors has established how compensatory mechanisms restoring fitness to resistant viruses may lie outside the drug target domains (Robinson et al., 2000). Introducing RT domains into an isogenic HIV-1 background ignores the contributions to RT function made by non-RT components of HIV-1. For example, recombinant viruses containing RT domains derived from d4T-resistant viruses often fail to exhibit reduced susceptibility to d4T (Lin et al., 1994). Furthermore, Peters et al. (2001) recently correlated nucleoside analogue resistance to the frequency of specific mutational patterns within the p6 domain of Gag. Moreover, mutants with these mutational patterns in p6 were shown to have increased replicative capacity when compared to viruses with wild-type sequences at these loci.

The question remains unresolved as to whether HIV-1HXB2 itself is an appropriate strain for the investigation of fitness. It has a premature stop codon rendering it effectively Nef-defective, a defective Vpu start codon and an abnormal frameshift sequence within Vpr. Close to the 3' terminus of the second Tat exon is another premature stop codon. Despite this, phenotypic drug resistance assays frequently utilise recombinant virus technology based upon the HIV-1HXB2 backbone, and the data so obtained has been clinically relevant.
(Kellam et al., 1992, 1994). It may also be postulated that HIV-1\textsubscript{HXB2} constitutes an appropriate genetic background only for the study of RT domains belonging to the same clade (B). Whether this can be expanded to include fitness may be contentious, particularly in light of the data presented here where mutations reducing fitness in the context of HIV-1\textsubscript{HXB2} RT (M41L and T215Y), actually increased fitness in a different RT genetic background (virus from patient A3). The behaviour of an RT domain can also vary according to the genetic background of the virus in which it is found (Peters et al., 2001). Hence, it may not be considered appropriate to use the HIV-1\textsubscript{HXB2} backbone when studying RT domains from non-clade B viruses. If this argument was extended further, it could lead to the conclusion that non-RT elements of the HIV-1\textsubscript{HXB2} genome could well distort the fitness of even clade B RT genotypes.

The necessarily artificial nature of the culture system in which viruses were competed requires the data acquired by these methods to be treated with great caution (Harrigan et al., 1998). For example, in culture systems, nucleotide concentrations may be unrepresentative of the uneven distributions that are found in vivo. This has importance, as some RT variants are sensitive to nucleotide concentrations (Back et al., 1996). Calculating fitness as a function of in vivo codon frequencies, virus load and time, whilst not as tightly controlled as the in vitro methodologies, has the great advantage of direct in vivo relevance (Goudsmit et al., 1996, 1997; Eastman et al., 1998; Frost et al., 2000). However, within the complex intra-host environment, numerous factors influence virus evolution, restricting conclusions to the context of the specific host from which the results were obtained. Recombinant virus technology inevitably operates at a distance from the in vivo situation, but can be usefully employed to explore hypotheses derived from in vivo data, away from the interfering and variable host factors that unpredictably distort in vivo data.

This thesis constitutes a preliminary series of experiments in the development of technologies that can be usefully applied to in vitro measurements of HIV-1 fitness. Although the series of experiments described here exposed defects in the system, these latter should be rectifiable without large-scale modifications, and notwithstanding the
caveats inherent within such an *in vitro* system, the methods presented here could be used to explore many aspects of HIV-1 evolution.

Firstly, the theory that a randomly selected subset of a quasispecies will have lower mean fitness than the fitness of the consensus quasispecies genotype could easily be tested. This has been shown for experimental populations of vesicular stomatitis virus (Duarte et al., 1994b), but not for *in vivo* HIV-1 populations. Although *in vivo* HIV-1 populations separated by therapy-mediated virus load drop have been shown to lose sequence diversity (Nijhuis et al., 1998; Ibanez et al., 2000), any changes in the relative fitnesses of drug target domains between the onset of therapy and viral load nadir has not been investigated. This is one explanation for the qualitative drop in fitness observed in our data from patient A3, where the relative fitness of the plasma virus RT population decreased after seven days of therapy (section 8.4). It was proposed that the onset of therapy and the consequent drop in viral load had constituted a genetic bottleneck, such that the virus population at the second timepoint was a random subset of those present at the onset of therapy. Further study into the behaviour of quasispecies during the early stages of therapy should involve a larger number of sample pairs and investigation on a clonal basis, in addition to population-based analysis.

Secondly, a sample taken at RNA load nadir would be presumed to be less fit, and could be allowed to evolve *in vitro*, in the presence or absence of continued drug pressure. Both the magnitude of fitness changes and their associated genotypic changes could be followed. The extent to which HIV-1 obeys the tenets of Fisher’s evolutionary model could be easily investigated (see pages 59 to 61).

The questions raised by the results of the analysis of RT domains from patient A3 could be expanded. Population sequence analysis means that amino acid ambiguity was reported only where the frequency of a minor population exceeded 33%. If individual RT domains were to be separated by cloning, it is possible that linkage between amino acids could be established. However, this system has its own weaknesses, due to the nature both of RT-
PCR amplification itself and of cloning diverse template populations. In a similar fashion to HIV-1 RT-mediated recombination, template switching by non-processive polymerases can generate recombinant genotypes absent from the original input template (Paabo and Wilson, 1988; Odelberg et al., 1995). This could undermine efforts to demonstrate linkage between amino acids at ambiguous loci. Differential amplification of similar template sequence due to variations in primer-binding sequences, length and denaturation temperature can lead to disproportionate representation of certain genotypes and as with the day seven sample from A3, sampling errors can occur if the frequency of target sequences is low (Walsh et al., 1992). Depending upon the subtlety of any fitness differences between sequences, these factors may profoundly affect experimental data.

Nevertheless, generating a matrix of genotypes containing all possible permutations of amino acid identities at the polymorphic codons and competing each against a reference strain could be used to establish a ‘fitness league table’ of genotypes. This may provide useful information as to how the resistance mutations M41L and T215Y came to be represented in post-therapy samples from patient A3.

Inspection of the sequences flanking the protease and Gag domains of HIV-1 should enable the development of a REL system for each combination of Gag, protease and RT (with the exception of sample Gag-RT and reference protease). These could be used to expand the range of patient samples that could be usefully investigated with our system, particularly in light of mutations in Gag affecting the resistance and fitness profiles of viruses subjected to protease and RT inhibitors (Robinson et al., 2000; Peters et al., 2001).

Reduced viral fitness incurred by resistance-associated mutations can be restored by secondary (compensatory) mutation to levels above the original wild-type. This phenomenon has been reported in relation to protease mutations (Nijhuis et al., 1999), Gag cleavage site mutations in response to protease inhibitor therapy (Robinson et al., 2000), and for the cluster of multinucleoside analogue resistance mutations in RT associated with Q151M (Maeda et al., 1998). However, the prevalence of compensatory mutation in
response to therapy is unknown. Rises in virus load in which the only apparent correlate is a change in Pol genotype could be investigated using the methods developed in this thesis, and mutations not previously associated with drug resistance could be demonstrated to improve Pol fitness. As the benefit of a given compensatory mutation may well be genotype-dependent, it could be difficult to establish links between specific genotypes and predicted responses. However, the data would reveal the prevalence of such mutational patterns, and the likelihood of non-resistance markers being implicated in the virological response to inhibitors.

In summary, this thesis demonstrates the possibility of viruses evolving high level drug resistance in conjunction with wild-type replication capacities. As Wainberg and Friedland (1998) noted, these viruses could pose a serious threat to public health. Due to wild-type levels of virus replication, the viral loads (and hence clinical prognoses) of individuals infected with such viruses be worse than in individuals infected with resistant, yet unfit viruses. By virtue of increased virus loads, transmission events involving these viruses would be more common than those involving resistant, yet unfit viruses, and to complete the cycle, an individual newly infected with such a virus would have a restricted range of treatment options, but a clinical prognosis similar to an infection with wild-type virus. Consequently, efforts to control the spread of the HIV-1 epidemic through therapeutic intervention would be significantly hampered, demonstrating the need for HIV-1 fitness to be better understood. It is hoped that the work in this thesis can provide a useful tool in further attempts to address these issues.
10. Bibliography


immunodeficiency viruses from high-risk, symptom-free individuals during therapy.
Lancet. 336, 585-590.


AZT-MP by drug-resistant human immunodeficiency virus reverse transcriptase. J. Virol.
75, 4832-4842.

Hughes, S.H. (1994b). Sensitivity of wild-type human immunodeficiency virus type 1
reverse transcriptase to dideoxynucleotides depends on template length; the sensitivity of
drug-resistant mutants does not. PNAS 91, 4882-4886.

Braaten, D. and Luban, J. (2001). Cyclophilin A regulates HIV-1 infectivity, as
demonstrated by gene targeting in human T cells. EMBO J. 20, 1300-1309.

Brun-Vezinet, F., Boucher, C.A., Loveday, C., Descamps, D., Fauveau, V., Izopet, J.,
Jeffries, D., Kaye, S., Krzyanowski, C., Nunn, A., Schuurman, R., Seigneurin, J.-M.,
phenotype, and resistance in a subset of drug-naive participants from the Delta trial. Lancet.
350, 983-990.

Brun-Vezinet, F., Rouzioux, C., Montagnier, L., Chamaret, S., Gruet, J., Barre-Sinoussi, F.,
Geroldi, D., Chermann, J.C., McCormick, J., Mitchell, S., Piot, P., Taelman, H.,
Mirlangu, K.B., Wobin, O., Mbendi, N., Mazebo, P., Kalambayi, K., Bridts, B., Desmyter, J.,
Feinsod, F.M., and Quinn, T.C. (1984). Prevalence of antibodies to lymphadenopathy-

interventions. AIDS. 10 Suppl 3, S33-42.


Bibilography


Bibliography


immunodeficiency virus type 1 RNA load suppression by zidovudine. PNAS 93, 5501-5506.


Bibliography


Bibliography


Ref Type: Abstract


BIBLIOGRAPHY


234


Bibliography


Keulen, W., Back, N.K., van Wijk, A., Boucher, C.A., and Berkhout, B. (1997a). Initial appearance of the 184Ile variant in lamivudine-treated patients is caused by the mutational...
bias of human immunodeficiency virus type 1 reverse transcriptase. J. Virol. 71, 3346-3350.


Liang, C., Rong, L., Russell, R.S., and Wainberg, M.A. (2000). Deletion mutagenesis downstream of the 5' long terminal repeat of human immunodeficiency virus type 1 is compensated for by point mutations in both the U5 region and gag gene. J. Virol. 74, 6251-6261.


BIBLIOGRAPHY


241


nucleoside analog reverse transcriptase inhibitors mediated by human immunodeficiency virus type 1 p6 protein. J. Virol. 75, 9644-9653.


shows a correlation between ex vivo human immunodeficiency virus type 1 fitness and disease progression. J. Virol. 74, 9222-9233.


Bibliography


Bibliography


Appendix A: Materials and General Methods

A.1 Plasma and Serum Samples

The DELTA trial monitored 3207 subjects with CD4⁺ cell counts below 350/mm³ taking either AZT, AZT and ddl, or AZT and ddC (DELTA Coordinating Committee, 1996). A 240-strong treatment-naïve subset of these were analysed more extensively (Brun-Vezinet et al., 1997).

The AZT-Resistance Study (ARS) comprised eleven HIV-positive subjects recruited into a study investigating the usefulness of AZT monotherapy (500-1000mg per day) in patients with category C disease (Loveday et al., 1995). Blood samples were taken at frequent intervals (every 4 days throughout the first month, and monthly thereafter), and promptly separated. Lymphocyte subsets were quantified by flow cytometry. Plasma RNA viral loads and p24 antigenemia levels were measured, and a quantitative point mutation assay (Kaye et al., 1992) applied to monitor the frequencies of AZT resistance mutations at codons 41, 67, 70, 215, and 219 of reverse transcriptase.
A.2 Primer and Oligonucleotide Sequences

Primers specific to each section are described therein. FIGURE A-1 and TABLE A-1 describe the locations within HXB2 and the specific nucleotide sequences respectively of primers common to all sections. TABLE A-2 displays the sequences and binding site locations of the primers used in the automated sequencing protocol described in section A.14.1 for discovering the nucleotide sequences of sample RT domains – FIGURE A-2 shows this schematically. All oligonucleotides were obtained from Oswel, Southampton, UK.

FIGURE A-1: Schematic of the 3' terminus of HXB2 Gag and the entire HXB2 Pol domain, showing the location of the binding sites for the primers listed in the upper part of TABLE A-1.
TABLE A-1: General primer sequences.

<table>
<thead>
<tr>
<th>Primers located within Gag and Pol</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A588</td>
<td>TTG CAG GGC CCC TAG GA</td>
</tr>
<tr>
<td>HPP3</td>
<td>GAG CCG ATA GAC AAG GAA CTG TAT</td>
</tr>
<tr>
<td>S-HR</td>
<td>ATA CAG GAG CAG ATG ATA CAG TAT TAG</td>
</tr>
<tr>
<td>706M</td>
<td>GAG CAG ATG ATA CAG TAT TAG</td>
</tr>
<tr>
<td>707M</td>
<td>CAA AGT AAG ACA GTA TGA TCA</td>
</tr>
<tr>
<td>469W</td>
<td>TAT TCC TAA TTG AAC TTC CC</td>
</tr>
<tr>
<td>470W</td>
<td>CCT ACA TAC AAA TCA TCC ATG</td>
</tr>
<tr>
<td>250L</td>
<td>TAG TCC CAT ATC CTG CTT TT</td>
</tr>
<tr>
<td>111W</td>
<td>GAA GTA AAC ATA GTA ACA G</td>
</tr>
<tr>
<td>708M</td>
<td>CAT TTA TCA CAG CTG GCT ACT A</td>
</tr>
<tr>
<td>AS-HR</td>
<td>GCT GAC ATT TAT CAC AGC TGG CTA CTA</td>
</tr>
<tr>
<td>709M</td>
<td>CAG TCT ACT TGT CCA TGC ATG GC</td>
</tr>
<tr>
<td>206V</td>
<td>CCT AAC CGT AGC ACC GGT G</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primers located within Env</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>618L</td>
<td>ATC CTC AGG AGG GGA CCC AG</td>
</tr>
<tr>
<td>628L</td>
<td>GTT CTA GAG ATT TAT TAC TCC AAC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primers located within pGEM-T sequences</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>T7</td>
<td>TAA TAC GAC TCA CTA TAG GGC GA</td>
</tr>
<tr>
<td>aSp6</td>
<td>ATT TAG GTG ACA CTA TAG ATT AC</td>
</tr>
</tbody>
</table>
TABLE A-2: Details of the primer sequences used for sequencing RT domains.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Position within RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-HR</td>
<td>ATA CAG GAG CAG ATG ATA CAG TAT TAG</td>
<td>RT -221-195 Sense</td>
</tr>
<tr>
<td>983L</td>
<td>TTG GGC CTG AAA ATC CAT AC</td>
<td>151-170 Sense</td>
</tr>
<tr>
<td>986L</td>
<td>CAT GGA TGA TTT GTA TGT AGG</td>
<td>551-571 Sense</td>
</tr>
<tr>
<td>DB-1</td>
<td>ACC ACA GAA AGC ATA GTA ATA T</td>
<td>1128-1149 Sense</td>
</tr>
<tr>
<td>469W</td>
<td>TAT TCC TAA TTG AAC TTC CC</td>
<td>284-265 Antisense</td>
</tr>
<tr>
<td>112V</td>
<td>GTC AGT CCA GCT GTC TTT TTC TGG</td>
<td>764-741 Antisense</td>
</tr>
<tr>
<td>MH6</td>
<td>CTC CCA CTC AGG AAT CCA GGT GGC</td>
<td>1247-1224 Antisense</td>
</tr>
<tr>
<td>G5804</td>
<td>TCA CTA GCC ATT GCT CTC CA</td>
<td>RT + 76-57 Antisense</td>
</tr>
</tbody>
</table>

FIGURE A-2: Location of the binding sites within HXB2 Pol of the sequencing primers detailed in TABLE A-2.
A.3 *Escherichia coli*

Three different strains of *E. coli* were used in the propagation of plasmid DNA – SURE2 (Stratagene, La Jolla, CA), TOP10F′ (Invitrogen, Carlsbad, CA), and STBL2 (also Invitrogen). All strains were obtained as competent cell aliquots, from which glycerol stocks were generated after overnight culture. Autoclave-sterilised 2xTY broth (16g bacto-tryptone, 10g yeast extract, and 5g sodium chloride, to 1 litre distilled water) was used as the liquid culture medium for *E. coli*. If selection of bacteria containing plasmids conferring ampicillin resistance was required, cooled post-autoclaved 2xTY broth was supplemented by the addition of 100μg/ml ampicillin. Unless stated otherwise, liquid cultures were grown overnight at 37°C in an orbital shaker set to ~225rpm. Solid-phase culture of bacterial colonies was on 2xTY agar plates, the composition of which was as per 2xTY broth, supplemented with 10g/L agarose added prior to autoclaving. If required, cooled autoclaved medium was supplemented by 100μg/ml ampicillin.

Each was selected for their ability to support the stable maintenance of foreign DNA sequences containing inverted repeats or secondary structures such as those often found in retroviral nucleic acids. SURE2 cells lack two DNA repair genes (*uvrC* and *umuC*). Further mutations in genes *recB* and *recJ* reduce the frequency of recombination between repeated sequences. TOP10F′ and STBL2 cells each have *recA* mutations conferring similar effects to the *recB* and *recJ* mutations of SURE2. STBL2 cells have additional genotypic markers specifically designed to maintain the integrity of transformed retroviral sequences (Trinh et al., 1994). Each strain has additional genomic mutations for improved recovery of plasmid DNA from transformed bacteria such as the *mcr, hsd, mrr, endAI*, and *recAI* systems. Other engineered defects render the bacteria auxotrophic such that they are not viable outside their specialised culture media. The full genotypes are detailed overleaf.
SURE2  e14\textsuperscript{-}(McrA\textsuperscript{-}) Δ(mcrCB-hsdSMR-mrr)\textsubscript{1}71 endA1 supE44 thi-1 gyrA96 relA1 lac recB recJ sbcC umuC::Tn5 (Kan\textsuperscript{r}) uvrC [F\textsuperscript{'} proAB lacI\textsuperscript{q}ZΔM15 Tn10 (Tet\textsuperscript{r})]

TOP10F\textsuperscript{'} mcrA Δ(mcrCB-hsdSMR-mrr) ) 80lacZΔM15 ΔlacX74 deoR recA1 araD139 Δ(ara-leu)7697 galU galK rpsL endA1 nupG [F\textsuperscript{'} lacI\textsuperscript{q} Tn10 (Tet\textsuperscript{r})]

STBL2  mcrA \textsuperscript{3} (mcrBC-hsdRMS-mrr) recA1 endA1 supE44 thi gyrA96 relA1 O \textsuperscript{3} (lac-proAB).

### A.4 Mammalian Cell Lines and Growth Media

Both mammalian cell lines were obtained as nitrogen stocks from the AIDS Reagent Project (Potters Bar, UK). SupT-1 cells were originally derived from a Non-Hodgkin’s T-cell lymphoma and express high levels of surface CD4 (Smith et al., 1984). MT-2 cells were established by co-culturing leukaemia cells harbouring human T-cell leukaemia-lymphoma virus subgroup 1 (HTLV-1) with cord blood lymphocytes from a human male infant (Miyoshi et al., 1981).

Both SupT-1 and MT-2 cells were maintained at cell densities between 2×10\textsuperscript{5} and 2×10\textsuperscript{6} per ml, in RPMI containing 10% FCS, supplemented with 1,000 U/L penicillin and 1mg/L streptomycin (referred to as ‘complete medium’ in the text). Incubation conditions were 37°C, 5.0% CO\textsubscript{2}, and a humid atmosphere.
A.5 Polymerase Chain Reaction

The polymerase chain reaction (PCR) (Saiki et al., 1985; Mullis et al., 1986) was widely used in this thesis for the specific amplification of HIV-1 sequences. As described below, amplification of viral genomic sequences from virion-derived RNA employed an initial reverse transcription step, generating a cDNA template for use in PCR itself.

Two PCR systems were employed in this thesis – AmpliTaq and Expand (both supplied by Roche, Indianapolis, IN). The enzymatic component of the former comprises solely Taq DNA polymerase, whereas the latter has an additional Pwo DNA polymerase component. Both polymerases are heat stable, but unlike Taq, Pwo possesses a proofreading 3’-5’ exonuclease property, and hence has higher fidelity than Taq. The Taq system was employed in screening PCRs due to it being considerably less expensive than the Expand system. However, for the amplification of sequences from viral nucleic acids where fidelity and yield were both important considerations, and in particular those with long products (>1.5kb), Expand was employed.

All PCR reaction mixes were set up in dedicated ‘clean areas’ free from environmental nucleic acid contamination, before being transferred to a different area for the addition of template and thermal cycling. To each reaction was added a 50μl mineral oil overlay to prevent evaporation of reactions during thermal cycling.

A.5.1 AmpliTaq-Mediated PCR

Unless otherwise stated, reactions comprised 1X manufacturer’s reaction buffer, 1.5mM MgCl₂ (from 25mM stock solution), 200μM each dNTP (from a stock solution containing each dNTP at a concentration of 10mM), 200nM each primer (from 60μM stocks), and 25 U/ml AmpliTaq polymerase. 20μl reactions were the norm unless otherwise stated in the text. Default thermal cycling conditions were as follows: 35 cycles of 94°C for 1 minute,
52°C for 1 minute, and 72°C for 1 minute, followed by a single cycle at 72°C for 7 minutes. Where conditions diverged from this default, they are detailed in the text.

**A.5.2 Expand-Mediated PCR**

Reactions were composed in two tubes to avoid pre-amplification degeneration of primers and template DNA by the exonuclease activity of *Pwo* polymerase. One tube contained 300nM each primer (from 60μM stocks), 200μM each dNTP (from a stock solution containing each dNTP at a concentration of 10mM), and 5mM dithiothreitol (DTT). The other comprised an equal volume to the first, and contained 52.5 U/ml Expand Enzyme mix in 2X supplied reaction buffer containing 2.25mM Mg\(^{2+}\) cations. Template was added to the first tube, followed by the addition of the contents of the second tube directly before adding mineral oil and the initiation of thermal cycling. Default cycling conditions were as per those for AmpliTaq-mediated PCR, but due to the decreased thermal stability and the lower optimal elongation temperature of *Pwo*, the 94°C steps were carried out at 92°C, and the 72°C steps at 68°C.

**A.6 Reverse Transcription**

Experiments performed early in the thesis employed separate reverse transcription (RT) and PCR steps, with cDNA synthesis from RNA carried out by the SuperScript II system (Invitrogen), followed by transfer of cDNA to a separate PCR reaction. Later experiments utilised the single-tube Titan RT-PCR system (Roche). Templates for both reactions were RNA extracted by the techniques described above in sections A.8.4 and A.8.5. Special care was taken during handling RNA to minimise the risk of degradation by environmental RNAses, specifically the use of aerosol barrier pipette tips, minimal exposure of RNA to air, frequent replacement of gloves, and use of a dedicated ‘clean’ area free of potential nucleic acid contaminants for the extraction of RNA.
A.6.1 The SuperScript II System

The separate cDNA synthesis reactions were prepared as follows. Firstly, primer was annealed to template. 200nM of primer was added to RNA template in a 10μl volume containing 1mM each dNTP, and the reaction incubated at 65°C for 5 minutes before supercooling the reactions on ice for 5 minutes. Added to the annealing reaction were 9μl of reaction mix such that the final concentrations of components were as follows: 1x SuperScript reaction buffer (from 10x stock solution), 10mM dithiothreitol, and 5mM magnesium chloride (all components supplied with the enzyme). After 2 minutes’ incubation at 42°C, 50 units of SuperScript II Enzyme were added, and the reaction continued for one hour at 42°C.

A.6.2 The Titan RT-PCR System

The single-tube Titan RT-PCRs were composed in two tubes, with the first containing RNA, 300nM each primer, 200μM each dNTP, and 5mM dithiothreitol (DTT). In the second part, an equal volume to the first, containing 20U/ml Titan Enzyme mix in 2X supplied reaction buffer containing 2.25 Mg²⁺ cations. Thermal cycling conditions were as described in the text.

A.7 Restriction Endonuclease Digestion

All enzymes for use in restriction endonuclease digests were supplied by New England Biolabs (NEB) and used in their supplied buffers. References to units when digests are described are to Weiss units (one unit is defined as that quantity of enzyme required to cleave 1μg of DNA substrate completely in 1 hour at the recommended incubation temperature in a 50μl volume). Incubations were for 2 hours at the temperature appropriate for each enzyme as listed in the NEB technical documentation. Reactions were 20μl in volume, and contained 500ng-1μg DNA and 5U of each enzyme (to maximise digestion in
cases of suboptimal conditions). Divergence from these conditions has been detailed in the text.

A.8 Nucleic Acid Purification

A.8.1 Purification of DNA from Agarose Gels, Restriction Digests and PCRs

For the purification of desired DNA species from agarose gels, QIAquick Gel Extraction Kits (Qiagen, Crawley, UK) were used according to the manufacturer’s instructions. For purifying DNA from PCRs, ligation reactions and restriction endonuclease digestions, QIAquick PCR Purification Kits (Qiagen) were used, also according to the manufacturer’s instructions. In the final step of each protocol, rather than a single 50μl elution, two consecutive 30μl elution steps with nuclease-free water were employed, to maximise nucleic acid recovery.

A.8.2 Preparation of Plasmid DNA from Bacterial Cultures

Plasmid-transformed bacteria were expanded by overnight culture in 2xTY broth containing ampicillin at 37°C in an orbital shaking incubator. Two volumes of culture were employed in this project – 3ml and 50ml.

DNA was extracted by alkaline precipitation followed by a spin column silica matrix-based bind-wash-elute system. For 3ml culture volumes, QIAGEN Plasmid Mini Kits were used, and for the larger size, QIAGEN Plasmid Midi Kits were used. Protocols were followed as per the manufacturer’s instructions. The volume of nuclease-free water used for the final re-suspension of the DNA pellet in the ‘Midi’ protocol varied according
to the quantity of DNA present as determined by visual examination of the DNA pellet. Typically, the DNA concentrations obtained by the ‘Midi’ method were approximately 1mg/ml.

A.8.3 DNA Purification by Phenol-Chloroform Extraction

DNA molecules greater than 10Kb in length were too large for the silica matrices in column-based extraction kits. Specifically, this method was used to purify the DNA of linearised RT-deleted molecular clones of HIV-1 from large-scale restriction digests. Employing sequential mixing and partitioning of organic and aqueous phases, this method eliminates organic molecules such as proteins and lipids from the aqueous phase. Finally, precipitation of the nucleic acid allows re-suspension of the nucleic acid in a suitable buffer, and to an appropriate concentration.

Firstly, an equal volume of phenol was added to the reaction to be purified, before vortexing for five minutes to ensure thorough mixing. Centrifugation at top speed in a microcentrifuge for 5 minutes (all further spins were of the same speed and length) resulted in a denser organic phase partitioned below a less dense aqueous phase. (The order of phases was similar after all further partitioning.) The aqueous phase was removed and transferred to a second equal volume of phenol, again mixed for five minutes. At this point, a ‘chaser’ was added to the residual phenol in the initial tube, comprising a volume of nuclease-free water equal to that of the remaining phenol. This latter was also thoroughly mixed by vortexing, and both the ‘sample’ and ‘chaser’ tubes were spun as before. Throughout this method, once the aqueous phase of the ‘sample’ tube had been transferred to the subsequent solvent tube, the aqueous phase of the ‘chaser’ was transferred to the left over organic phase, and the residual organic phase in the ‘chaser’ tube discarded. After the second phenol mix, three further mixes and spins were performed – one phenol-chloroform
(50% phenol, 50% chloroform), and two of chloroform alone, all volumes equal to the volumes of the transferred ‘sample’ aqueous phase.

After the final partitions (both ‘sample’, and then ‘chaser’), the aqueous phases were combined and transferred to a fresh tube. A volume of isopropanol was added equal to 70% of the aqueous volume. This was thoroughly mixed by vortexing and incubated at −20°C for one hour before being centrifuged at top speed in a microcentrifuge for 30 minutes. The supernatant was removed, and the pellet resuspended in 70% ethanol. A further 10-minute top speed centrifugation was followed by the complete removal of the ethanol, and the pellet was allowed to air-dry. Re-suspension was in Qiagen Elution Buffer (10mM Tris-Cl, pH 8.5), and the aliquoted at −20°C until required.

A.8.4 Purification of RNA from Plasma and Serum
– the Boom Protocol

RNA was extracted from patient plasma using the Boom extraction protocol (Boom et al., 1990). 100μl of plasma were added to 900μl of Lysis Buffer (TABLE A-3) containing 50μl of coarse silica. This was mixed by vortexing and shaken for 10 minutes at room temperature. Following centrifugation for 1 minute at 18000g, the supernatant was discarded and washed with 1ml Wash Buffer (TABLE A-3). The silica was pelleted (1’ centrifugation at 18,000g) and washed sequentially with another 1ml Wash Buffer, 2 x 1ml washes with 70% ethanol, and finally once with 1ml acetone. After discarding the acetone supernatant, the pellet was air dried in a hot block at 56°C for 10 minutes before being resuspended in 60μl nuclease-free water. A further incubation of 10 minutes at 56°C and a final vortex was followed by 2 minutes’ centrifugation at 18,000g, and the eluate transferred to a fresh tube. RNA extracts were stored at −20°C for up to one year. Extracts requiring longer storage periods were maintained at −70°C until use.
A.8.5 Extraction of RNA from Culture Supernatants

40μl of cell-free culture supernatant was entered into a QIAamp Viral RNA Mini Kit (Qiagen), and the protocol followed according to the manufacturer’s instructions, with the final step comprising a single elution using 60μl nuclease-free water. Storage of RNA extracts was as for those extracted by the Boom method above.

TABLE A-3: Buffer compositions for the Boom extraction protocol.

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wash Buffer</td>
<td>120g guanidinium thiocyanate to 100ml 0.1M Tris-Cl, pH 6.4. Stored at 4°C</td>
</tr>
<tr>
<td>Lysis Buffer</td>
<td>Wash Buffer, with additional 22ml 0.2M EDTA, pH 8.0, 2.6g Triton X-100. Stored at 4°C</td>
</tr>
</tbody>
</table>

A.9 Preparation of Competent Bacteria from Glycerol Stocks

100μl of a glycerol stock of E. coli were added to 3ml of 2xTY broth without ampicillin and incubated overnight at 37°C with vigorous shaking. 500μl of this culture was added to a 50ml volume of 2xTY broth in an Erlenmeyer flask and incubated for 3 hours at 37°C with vigorous shaking. After incubating the cells on ice for 20 minutes, cells were centrifuged at 4°C at 4,500 rpm (4,000g) for 5 minutes, the supernatant discarded and the cells resuspended in 20ml of ice-cold 0.1M calcium chloride. Following incubation on ice for 1 hour, the cells were spun again at 4°C at 4,500 rpm (4,000g) for 5 minutes, the supernatant discarded, and the cells resuspended in 4ml ice-cold 0.1M calcium chloride. The cells were further incubated on ice for at least one hour prior to use. Competent cells were stored for a maximum of 48 hours at 4°C before being discarded.
A.10 Transformation into *E. coli* of Plasmid DNA

The DNA to be transformed was placed in a 1.5ml flip-top Eppendorf tube and incubated on ice for 10 minutes before the addition of 200µl of competent bacterial cells (either freshly prepared as above, or obtained as competent from commercial suppliers). The mixture was left on ice for 40 minutes before being placed into a waterbath at 42°C for a period depending on the bacterial strain (the duration of the heat shock ranged from 30 to 90 seconds – manufacturers’ instructions were followed). The cells were replaced on ice for a further 10 minutes before being either plated out onto agar plates, added to a 3ml overnight culture, or both. The transformation control used was 10ng pUC18 supercoiled DNA.

A.11 Ligation

Unless otherwise stated, ligation reactions proceeded as follows:

DNA species (from either PCRs, restriction digests, etc.) were subjected to agarose gel electrophoresis and purified using QIAquick Gel Extraction Kits (Qiagen). 5µl aliquots of each species were subjected to a further round of agarose gel electrophoresis allowing the relative yields of each to be quantified using NIH Image analysis (section A.13, below). Ligation reactions were 20µl in volume, and consisted of 1X Ligase Buffer (NEB) containing 4U T4 DNA Ligase (NEB) and a 3:1 molar ratio of ‘insert’ DNA molecules to ‘vector’ DNA molecules. Reactions were incubated overnight at 12°C in a thermal cycler.
A.12 Hoescht Dye-Based Fluorometric DNA Quantification

This technique was used to quantify the nucleic acid components of transfections prior to their electroporation. 3μl sample DNA was mixed with 150μl assay diluent (100mM NaCl, 10mM Tris-Cl, pH7.5, 1mM EDTA) in duplicated black 96-well microtitre plates. Eight paired wells were reserved for the addition of pre-prepared calf thymus DNA standards. To each well was added 50μl of dye solution (a 0.5% solution of Hoescht dye 33258 in assay diluent). The plate was shaken briefly and incubated at room temperature in the dark for 5 minutes. Relative fluorescence values were assayed by a fluorometer, and relative DNA concentrations were calculated using ASCENT specialised software (Thermo Labsystems, Helsinki, Finland).

A.13 Kodak Gel-Based DNA Quantification

This method was used to ascertain relative concentrations of DNA in solution. After agarose gel electrophoresis, the gel was photographed under ultraviolet light (254nm), the image transferred to computer, and the relative band intensities were calculated by Kodak 1D Image Analysis software. Use of molecular weight markers of known concentration enabled measurement of absolute concentration. It must be noted that this technique was less accurate than the method described in section A.11, but had the advantage of being both rapid and more sensitive.
A.14 DNA Dideoxy-Sequencing

Sequencing of DNA regions was by the Sanger dideoxy sequencing method, although the protocol was largely automated. Briefly, sample DNA and primer were added to ‘Big-Dye Terminator’ reaction mix (Perkin Elmer). Reactions were subjected to cycle sequencing using a PE 2400 thermal cycler as per the manufacturer’s instructions. Reactions were purified by ethanol precipitation and loaded onto a gel matrix within an ABI 377 automated sequencer machine. This separated the cycle sequencing products according to size. By analysing the fluorescence spectra, the template nucleotide sequence from the primer-binding position for approximately 550bp in its 3’ direction could be elucidated. Ethanol precipitations, gel preparations, gel loading and operation of the ABI 377 were carried out within the department by Julie Bennett.

Sequencher software (Gene Codes Corporation, MI) was used to analyse the resultant sequences, and to generate ‘contigs’ – overlapping sequences derived from the same template but with different primers. This software enables direct visualisation of the chromatogram output of the ABI 377, allowing examination and modification of disputed nucleotide positions.

A.14.1 Sequencing of RT Domains

Templates for sequencing of sample RT domains were prepared by Expand-mediated PCR amplification of first-round Titan RT-PCR products using primers S-HR and AS-HR (TABLE A-1). Products were sequenced using the above method using the four sense primers and four antisense primers per domain described in TABLE A-2 and FIGURE A-2. Sequencher software was again used to align multiple sequences to obtain contiguous RT domain sequences.
A.15 Transfection

This method utilised the phenomenon of cell-membrane permeabilisation through the application of high voltage across mammalian cells to induce the internalisation of both circular and linearised molecules of foreign DNA. In this thesis it was used either to co-transfect sample RT sequences with RT-deleted molecular clones of HIV-1, or to transflect plasmid DNA containing full-length recombinant HIV-1.

To induce exponential growth of the SupT-1 mammalian cell culture, the cells were split to a density of 0.25x10^6 ml^-1 two days prior to transfection. A viable cell count was performed on the growing SupT-1 culture using a Trypan Blue exclusion assay before centrifugation at 1,200rpm (~500g) for 10 minutes. A volume of chilled complete RPMI medium containing an extra 10% FCS (i.e. final FCS concentration of 20%) was added to give a final cell density of 20 x 10^6 ml^-1. The cells were incubated on ice for 10 minutes.

The 2Hg of linearised p3 RT and 2Hg gel-purified PCR product were transferred to a 0.4cm electroporation cuvette (Invitrogen), and 250Î¿ of the SupT-1 suspension (5x10^6 cells) added. Using an Invitrogen Electroporator, the contents of the cuvette were electroporated by a 50-60ms shock at 330V and 1000Î¿F. The transfected cells were transferred to a 25cm^3 tissue culture flask containing 5mls of 1x10^5 SupT-1 cells/ml.

On day 3 post-transfection, 5ml of complete medium (10% final FCS concentration) was added to each culture. On day 4, 10 ml of complete medium were added and the culture transferred to a 75cm^3 tissue culture flask, and on day 5 post-transfection, a further 20ml of complete medium were added.

Regular microscopic examination of cultures for cytopathic effect was used to monitor the growth of virus in these cultures (the HIV backbone used in this technique derives from a tissue-culture-laboratory-adapted strain and forms syncytia in SupT-1 cells). Cultures judged to be highly infected (>90% target cells syncytial), were harvested by centrifugation
at 1,200rpm (~500g) for 10 minutes and 1ml aliquots of culture supernatants stored in liquid nitrogen. Non-productive cultures were discarded after 14 days.

**A.16 p24 Antigen ELISA**

**A.16.1 Preparation of Human Anti-p24 IgG**

From human plasma containing a high anti-p24 antibody titre (as assayed by routine serology in our diagnostic laboratory), the IgG fraction was purified. Firstly, a Protein-G column was equilibrated by washing with 30ml 0.1M Tris-Cl, pH 7.8, followed by loading 1ml of plasma diluted in 1ml 0.1M Tris-Cl, pH 7.8 onto the column. The column was further washed with 0.1M Tris-Cl, pH 7.8 to propel the sample through the column. After the sample had progressed through the column, as monitored by an attached spectroscope, the bound IgG was eluted into 0.35g Tris buffer by the application of 5mls 1M acetic acid. The eluate was dialysed overnight against PBS, followed by quantification by spectrophotometry at 280nm.

**A.16.2 ELISA**

Storage plates were prepared prior to sampling. 55μl of RPMI 10% FCS containing 0.2% sodium azide and 0.2% Triton-X100 were added to each well of a 96-well flat-bottomed Nunclon Delta MicroWell plate. Duplicate 55μl volumes of a 5-fold dilution series of high-titre virus stock was added to the first column – for the purposes of this experiment it was only necessary to calculate ratios between p24 titres rather than the absolute concentrations. Storage plates were stored in a moist environment at 4°C. 55μl cell-free culture supernatant samples were added to wells on storage plates and mixed thoroughly by pipetting.

Assay plates were prepared by coating each well of a 96-well Nunc Maxisorb plate with 50μl 50mM Tris-Cl, pH 9.2 with 2.5μg/ml human anti-p24 IgG, and incubating overnight at
To begin the ELISA, the assay plate was washed in PBS-Tween, and 50μl from each well of the storage plate were transferred to its corresponding well on the assay plate. The assay plate was incubated overnight at 4°C before being washed in PBS-Tween. 50μl per well TMT-A containing 0.1% of the anti-p24 monoclonal antibodies 38:96 and EF7 (AIDS Reagent Project, Potter’s Bar, UK) were added to each well and the plate incubated at 37°C for 1 hour. Further washing with PBS-Tween was followed by the addition of 50μl per well of TMT containing 0.05% horseradish peroxidase-conjugated goat anti-mouse antibody (Dako, Ely, UK). During a further incubation for 1 hour at 37°C, TMB solution was prepared. One tablet of freebase 3,3',5,5'-tetramethylbenzidine (TMB) was dissolved in 1ml dimethyl sulphoxide and added to 9ml phosphate-citrate buffer (51.4mM dibasic sodium phosphate, 48.6mM citric acid, pH 5.0). Immediately prior to use, 30μl of 3% v/v hydrogen peroxide was added.

After washing with PBS-Tween, 50μl per well of TMB solution were added to each well, and colour allowed to develop for approximately 20 minutes. The reaction was stopped by the addition of 20μl per well of 5N HCl. The plate was briefly shaken and optical densities at 450nm were measured by a plate-reader. To calculate the sample p24 concentrations by reference to the standards, a Visual Basic analysis program was devised.

A.17 Titration of Virus Stocks

In order to control accurately the levels of virus entering subsequent experiments, frozen stocks of culture supernatant containing recombinant virus were assayed for their infectivity. After thawing from liquid nitrogen storage, a five fold dilution series of the virus stock was made in complete medium, from undiluted supernatant to a 5⁻⁶ dilution and
a no-virus control. In microwell plates, MT-2 cells were infected with each dilution of virus in quadruplicate, and after five days, a substrate was applied whose colour changes upon cleavage by mitochondrial enzymes active only in living cells (Japour et al., 1993).

All plasticware and growth medium reagents were supplied by Life Technologies, Paisley, UK unless otherwise stated. One day prior to use, a maintained culture of MT-2 cells was diluted in complete medium to a final cell density of 1x10^6 ml^{-1}. On the day of titration, the cell density of the MT-2 culture was measured by a Trypan Blue exclusion method. Cells were centrifuged and resuspended in complete medium to a cell density of 0.4x10^6 ml^{-1}.

For each titration, 100μl of cell suspension were added to 4 columns of 8 wells (32 wells) of a 96 well flat-bottomed Nunclon Delta MicroWell Plate. On a separate Nunclon Delta MicroWell Plate, the virus dilution was constructed. To each of 8 wells in a column excepting the first, 180μl of RPMI 10% FCS were added. To the first well were added 235μl of virus stock. 45μl from the first well were transferred to the second well and thoroughly mixed by pipetting. 45μl from the second well were subsequently transferred to the third well and so forth until the seventh well. The eighth and final well was left blank as a negative control. 40μl of each dilution were transferred to each well in a row of four on the test plate. Plates were incubated at 37°C in a humid environment containing 5% CO₂.

After five days, 100μl of MTT solution (PBS containing 20mM 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2H-tetrazolium bromide filtered through a 0.22 micron filter) were added to each well, and the plate returned to the incubator for a further 2 hours. 50μl were carefully removed from each well avoiding removal of any blue formazan crystals at the base of the wells and replaced by 100μl of propan-2-ol containing 10% Triton X-100 and 0.4% hydrochloric acid. Plates were shaken for 10 minutes at 600rpm to dissolve the formazan, and the optical densities of each well at 610nm (OD_{610}) measured using a Titertek MkII Plate Reader. The mean OD_{610} values for each row of four wells were calculated, and the value for the virus-free control row was subtracted from each test row.
Each value was subsequently expressed as a percentage of the value for the first row, and the 50% endpoint dilution calculated by linear interpolation.

Values obtained with this method do not represent TCID$_{50}$s, as MTT cleavage occurs only in living cells. Hence, absence of colour change is proportional to the degree of cell death—values more accurately represent 50% lethal doses (‘LD$_{50}$s’). Correlation between TCID$_{50}$ values (established by microscopy) and LD$_{50}$ values was made.