



Universidade Nova de Lisboa
Instituto de Higiene e Medicina Tropical

**Caracterização de polimorfismos no gene *pol* associados a
resistência aos inibidores da protease utilizados em regimes
terapêuticos de prevenção da transmissão mãe-filho do
vírus da imunodeficiência humana tipo 1**

Cátia Sofia Sérgio Simões

DISSERTAÇÃO PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS
BIOMÉDICAS, ESPECIALIDADE EM BIOLOGIA MOLECULAR EM SAÚDE
TROPICAL E INTERNACIONAL

JULHO, 2016



Universidade Nova de Lisboa
Instituto de Higiene e Medicina Tropical

**Caracterização de polimorfismos no gene *pol* associados a
resistência aos inibidores da protease utilizados em regimes
terapêuticos de prevenção da transmissão mãe-filho do
vírus da imunodeficiência humana tipo 1**

Autor: Cátia Sofia Sérgio Simões

Orientador: Prof. Doutor João Piedade

Co-orientador: Inv.^a Doutora Elizabeth Pádua

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de mestre em Ciências Biomédicas, especialidade em Biologia Molecular em Saúde Tropical e Internacional

A presente dissertação foi redigida segundo o acordo ortográfico de 1945.

Elementos bibliográficos

Simões, C., Pádua, E., Mendes, A., Esteves, A., Parreira, R., Piedade, J. (2016). Genetic diversity and drug resistance profiles of human immunodeficiency virus type 1 (HIV-1) strains infecting pregnant women in the Greater Lisbon. *2nd Genetics Workshop*. Lisboa, Portugal, 30 de Maio.

Simões, C., Pádua, E., Mendes, A., Esteves, A., Parreira, R., Piedade, J. (2016). Genetic diversity and drug resistance profiles of human immunodeficiency virus type 1 (HIV-1) strains infecting pregnant women in the Greater Lisbon. *19th Annual Meeting of the European Society for Clinical Virology*. Lisboa, Portugal, 14-17 de Setembro.

Agradecimentos

A presente dissertação simboliza uma etapa muito importante a nível pessoal e académico, cuja realização só foi possível devido ao apoio e incentivo de todas as pessoas envolvidas. Quero deixar aqui os meus sinceros agradecimentos, por tornarem possível a sua concretização.

Um agradecimento especial ao Professor Doutor João Piedade, orientador desta tese, por toda a sua dedicação, apoio, paciência e constante disponibilidade. O seu entusiasmo pelo tema, partilha de conhecimentos e motivação tornaram este percurso gratificante. Agradeço ainda todas as críticas e sugestões essenciais para a elaboração do presente projecto.

À Investigadora Doutora Elizabeth Pádua pelo fornecimento das amostras biológicas e dos respectivos dados epidemiológicos. Muito obrigada pela sua disponibilidade e simpatia.

À Professora Doutora Aida Esteves e ao Professor Doutor Ricardo Parreira pelo apoio, simpatia e conhecimentos transmitidos ao longo do mestrado.

À Técnica Superior Ângela Mendes, por toda a sua simpatia, ajuda laboratorial e disponibilidade demonstrada.

Aos meus amigos que sempre me apoiaram e motivaram ao longo desta etapa. Um especial agradecimento à Nicole Lourenço, Ana Pacheco, Inês Guerra e Rita Pedrosa por toda a amizade, conselhos e momentos de convívio partilhados ao longo de todo o mestrado. À Mariana Nunes e Rute Marcelino por todos os momentos de alegria e boa disposição partilhados neste último ano.

Aos meus pais e à minha irmã pelo apoio imprescindível, confiança, motivação e compreensão demonstrados durante toda esta etapa.

A todos, Muito Obrigada!

Resumo

Segundo a UNAIDS (*Joint United Nations Programme on HIV/AIDS*), no final de 2014, o vírus da imunodeficiência humana (HIV) era responsável pela infecção de 36,9 milhões de pessoas, das quais, 2,6 milhões eram crianças com idade inferior a 15 anos. A transmissão vertical representa a principal causa de infecção em crianças e, embora o risco associado tenha diminuído drasticamente com a introdução da terapia anti-retroviral de alta eficácia, esta continua a verificar-se. Ainda que tenha sido alvo de progressos significativos, o acesso das mulheres grávidas à terapia ainda se mantém difícil em certas regiões do planeta. Em alguns casos, por outro lado, a presença de mutações virais associadas a resistência é responsável pela falha do regime profilático implementado e pode, conseqüentemente, levar à transmissão de vírus resistentes para o recém-nascido. Nos últimos anos, os inibidores da protease foram introduzidos nos regimes terapêuticos para prevenção da transmissão vertical, anteriormente constituídos somente por inibidores da transcriptase reversa.

O presente estudo incidiu sobre um grupo de 34 mulheres múltiplas infectadas com HIV-1, às quais foi efectuada a recolha de sangue periférico, num período máximo de dois dias após o parto, entre os anos de 1999 e 2008. Na sua grande maioria, as mulheres incluídas no estudo tinham cumprido regimes terapêuticos de prevenção da transmissão vertical do HIV-1 durante a gravidez. O DNA proviral das 70 amostras em estudo foi extraído a partir de células mononucleadas do sangue periférico, sendo efectuada a amplificação da região codificante da protease do gene *pol*, por *nested* PCR dupla. Após sequenciação nucleotídica, foi realizada a caracterização genotípica das estirpes virais por análise filogenética manual e a caracterização das mutações associadas a resistência, assim como, de outros polimorfismos genéticos, com recurso ao programa HIVdb, implementado através do *Genotypic Resistance Interpretation Algorithm* da *Stanford University HIV Drug Resistance Database*.

Este estudo revelou uma elevada diversidade genética do HIV-1 na população em causa, na qual foram identificados os subtipos G (47,8%), C (14,9%) e B (11,9%) como predominantes, destacando-se ainda uma elevada prevalência de formas recombinantes únicas (16,4%). Os subtipos não-B foram responsáveis pela infecção em todas as mulheres de origem africana, sendo o subtipo B apenas encontrado em mulheres de naturalidade portuguesa. As mulheres africanas foram as únicas infectadas com o subtipo C. Relativamente à pesquisa de mutações associadas a resistência aos inibidores da protease, foram detectadas duas mutações principais (D30N e M46I) e sete mutações acessórias (L10I, L10V, L33F, G48E, A71T, A71V e T74S), em 19 das seqüências estudadas. Destas, 16 eram de subtipos não-B, não sendo contudo demonstrada nenhuma associação estatisticamente significativa. Além disso, a maioria destas mutações foram detectadas em mulheres cujos regimes terapêuticos incluíam inibidores da protease, os quais podem ter levado à sua selecção. Dos restantes polimorfismos genéticos, não associados a resistência, a detecção foi superior em subtipos não-B, sendo alguns classificados como “assinaturas genéticas” de subtipo.

A identificação de mutações associadas a resistência aos inibidores da protease em mulheres cujos regimes terapêuticos incluem estes fármacos pode ter implicações na prevenção da transmissão vertical, sendo importante a sua monitorização.

Palavras-chave: HIV-1; mulheres grávidas; terapia anti-retroviral; resistência aos inibidores da protease; genotipagem.

Abstract

According to the UNAIDS (The Joint United Nations Programme on HIV/AIDS), the human immunodeficiency virus (HIV) infected 36.9 million people at the end of 2014, of whom 2.6 million were children under 15 years of age. Vertical transmission is the main cause of infection in children and while the associated risk has decreased dramatically with the introduction of highly active antiretroviral therapy (HAART), this transmission continues to occur. Although it has been the subject of significant progress in recent years, the access of pregnant women to therapy still remains difficult in certain regions of the world. Moreover, in some cases, the presence of viral mutations associated with resistance is responsible for the failure of the implemented prophylactic regimens and may consequently lead to transmission of resistant viruses to newborns. Not long ago, protease inhibitors were introduced in therapeutic regimens for prevention of vertical transmission, formerly including reverse transcriptase inhibitors only.

This study focused on a group of 34 multiparous women infected with HIV-1, from whom a sample of peripheral blood was collected within two days after delivery, between the years 1999 and 2008. The great majority of the women included in the study had followed therapeutic regimens for prevention of vertical transmission of HIV-1 during pregnancy. The proviral DNA of 70 samples analyzed was extracted and purified from peripheral blood mononuclear cells, being the amplification of the protease coding region in *pol* gene carried out by double nested PCR. After nucleotide sequencing, the genetic characterization of the viral strains by manual phylogenetic analysis was performed, along with the characterization of resistance-associated mutations, as well as other genetic polymorphisms, using the HIVdb program, implemented by the Genotypic Resistance Interpretation Algorithm from the Stanford University HIV Drug Resistance Database.

The study revealed a high genetic diversity of HIV-1 within this population, in which G (47.8%), C (14.9%), and B (11.9%) subtypes were predominant, with also a high prevalence of unique recombinant forms (16.4%). Non-B subtypes were responsible for the infection in all women of African origin, and the B subtype was found only in women of Portuguese nationality. Additionally, African women were the only infected with subtype C. Considering the mutations associated with resistance to protease inhibitors, two major mutations (D30N and M46I) and seven minor mutations (L10I, L10V, L33F, G48E, A71T, A71V, and T74S) were identified, in 19 of the sequences studied. Of these, 16 were classified as non-B subtypes, but no statistically significant association was found. Furthermore, most of these mutations were detected in women whose prophylactic regimens included protease inhibitors, which may have led to their selection. The remaining genetic polymorphisms, not associated with anti-retroviral resistance, were mainly detected in non-B subtypes, with some being classified as subtype "signatures".

The identification of HIV-1 mutations associated with resistance to protease inhibitors in women whose prophylactic regimens during pregnancy included this class of drugs may have implications for the prevention of vertical transmission of the virus, stressing the importance of mutation surveillance.

Keywords: HIV-1; pregnant women; antiretroviral therapy; resistance to protease inhibitors; genotyping.

Índice

Elementos bibliográficos	ii
Agradecimentos	iii
Resumo	iv
<i>Abstract</i>	vi
Índice	viii
Índice de Figuras.....	xii
Índice de Tabelas	xiv
Lista de Símbolos, Acrónimos e Abreviaturas	xv
1. Introdução.....	1
1.1. Vírus da imunodeficiência humana tipo 1	2
1.1.1. Origem do HIV-1 e história da pandemia.....	2
1.1.2. Epidemiologia	5
1.1.3. Organização estrutural e genómica do HIV-1	6
1.1.4. Ciclo replicativo.....	9
1.1.4.1. Adsorção e Entrada.....	10
1.1.4.2. Descapsidação, Transcrição Reversa, Importação Nuclear e Integração.....	11
1.1.4.3. Transcrição, Exportação Nuclear do RNA Viral e Síntese Proteica .	12
1.1.4.4. Morfogénese e Maturação	12
1.1.5. Diversidade genética do HIV-1	13
1.1.5.1. Mecanismos de diversidade genética	14
1.1.5.2. Diversidade genética e distribuição mundial.....	15
1.1.6. Principais vias de transmissão	16

1.1.6.1.	Transmissão vertical.....	16
1.1.7.	Gene <i>pol</i>	17
1.1.8.	Protease: estrutura e função	17
1.2.	Terapia anti-retroviral	18
1.2.1.	Regimes terapêuticos para prevenção da transmissão vertical	21
1.2.1.1.	Profilaxia pré-parto.....	22
1.2.1.2.	Profilaxia intraparto.....	22
1.2.1.3.	Profilaxia pós-parto	23
1.2.2.	Inibidores da protease	23
1.3.	Resistência associada aos anti-retrovirais	25
1.3.1.	Monitorização de mutações de resistência aos anti-retrovirais	26
1.3.2.	Mecanismos de resistência aos inibidores da protease	28
1.3.3.	Caracterização de mutações associadas à resistência aos inibidores da protease.....	29
1.4.	Objectivos	31
2.	Material e Métodos.....	32
2.1.	Desenho do estudo	33
2.2.	Caracterização da amostra populacional estudada.....	34
2.3.	Identificação e processamento das amostras biológicas	35
2.4.	Amplificação por <i>nested</i> PCR dupla da região codificante da protease no gene <i>pol</i>	36
2.4.1.	Seleccção dos <i>primers</i>	36
2.4.2.	Protocolo de amplificação	37
2.4.2.1.	Primeira e segunda reacções de PCR	38
2.4.2.2.	Terceira reacção de PCR	39
2.5.	Observação dos fragmentos de DNA em gel de agarose	40

2.6.	Purificação dos fragmentos de DNA.....	41
2.7.	Clonagem dos fragmentos de DNA	42
2.7.1.	Ligação dos fragmentos de DNA ao vector <i>pGEM[®]-T Easy</i>	43
2.7.2.	Preparação de células competentes de <i>Escherichia coli</i> e sua transformação por choque térmico	43
2.7.3.	Seleção de clones bacterianos e extracção do DNA plasmídico pelo método de lise alcalina	45
2.7.4.	Observação das moléculas de DNA plasmídico em gel de agarose	46
2.7.5.	Hidrólise enzimática do DNA plasmídico	46
2.7.6.	Extracção e purificação do DNA plasmídico para sequenciação nucleotídica.....	47
2.8.	Sequenciação nucleotídica de DNA.....	48
2.9.	Análise bioinformática das sequências nucleotídicas	48
2.9.1.	Edição de sequências e confirmação da sua identidade.....	48
2.9.2.	Genotipagem automática das sequências de HIV-1	49
2.9.3.	Análise filogenética	49
2.9.4.	Análise de sequências divergentes.....	50
2.10.	Análise bioinformática das sequências aminoacídicas.....	51
2.10.1.	Pesquisa de mutações associadas à diminuição da susceptibilidade aos anti-retrovirais inibidores da protease	51
2.10.2.	Análise da variabilidade dos domínios funcionais da protease	52
2.11.	Análise estatística	53
3.	Resultados.....	54
3.1.	Caracterização da amostra populacional estudada.....	55
3.2.	Amplificação da região codificante da protease por <i>nested PCR</i> dupla	57
3.3.	Clonagem dos produtos de PCR em <i>pGEM[®]-T Easy</i>	59

3.4.	Análise bioinformática das sequências nucleotídicas	60
3.4.1.	Genotipagem automática das sequências do HIV-1	61
3.4.2.	Análise filogenética	61
3.4.2.1.	Análise filogenética de sequências concatenadas.....	68
3.5.	Análise bioinformática das sequências aminoacídicas	71
3.5.1.	Identificação de mutações associadas a resistência aos inibidores da protease e outros polimorfismos genéticos.....	72
4.	Discussão e Conclusões.....	79
5.	Referências bibliográficas	101
6.	Anexos.....	122

Índice de Figuras

Figura 1. Área geográfica de distribuição natural de <i>Pan troglodytes troglodytes</i> e de <i>Cercocebus atys</i>	4
Figura 2. Distribuição relativa do número de indivíduos infectados por HIV, pelas diferentes regiões geográficas, no final de 2014.....	5
Figura 3. Representação esquemática da estrutura da partícula viral do HIV-1 imaturo e maduro	7
Figura 4. Representação esquemática do genoma de HIV-1	7
Figura 5. Representação do ciclo replicativo do HIV-1	10
Figura 6. Representação da estrutura tridimensional da protease do HIV-1, em conformação aberta e em conformação fechada.....	18
Figura 7. Representação genérica do modo de acção dos inibidores da protease	24
Figura 8. Representação esquemática das posições aminoacídicas na protease associadas a resistência aos inibidores da protease.	28
Figura 9. Representação do painel de mutações da região codificante da protease, compilado pela IAS-USA, associadas a resistência aos inibidores da protease licenciados pela FDA.....	30
Figura 10. Representação esquemática do protocolo de amplificação por <i>nested</i> PCR dupla utilizado na amplificação da região codificante da protease do HIV-1.....	38
Figura 11. Representação gráfica dos dados demográficos da amostra populacional em estudo, relativos à nacionalidade e idade (no nascimento do primeiro filho).....	55
Figura 12. Representação gráfica dos dados relativos à realização de terapia anti-retroviral pelas 34 mulheres infectadas pelo HIV-1 durante as 70 gravidezes consideradas.....	56
Figura 13. Representação gráfica dos dados relativos aos regimes aplicados nas 62 gestações sujeitas a terapia anti-retroviral	56

Figura 14. Representação gráfica dos resultados obtidos (em percentagem) após amplificação da região codificante da protease por <i>nested</i> PCR dupla.	57
Figura 15. Observação dos produtos de amplificação obtidos por <i>nested</i> PCR dupla, relativos à região codificante da protease, após electroforese em gel de agarose	58
Figura 16. A) Observação dos padrões de migração dos DNAs plasmídicos dos clones bacterianos recombinantes putativos seleccionados, após electroforese em gel de agarose. B) Análise dos produtos resultantes da hidrólise com a enzima de restrição EcoRI, após migração electroforética em gel de agarose	60
Figura 17. A) Relações filogenéticas entre 124 sequências nucleotídicas da região codificante da protease do HIV-1, com base num total de 321 nucleótidos. B) Análise de <i>bootscanning</i> , por <i>SimPlot</i> v.3.5.1, das sequências assinaladas com um rectângulo laranja.....	63
Figura 18. Distribuição dos génotipos prováveis das 67 sequências nucleotídicas do HIV-1 codificantes da protease segundo a nacionalidade das mulheres infectadas incluídas no estudo.....	68
Figura 19. A) Relações filogenéticas entre 99 sequências nucleotídicas concatenadas compreendendo as regiões codificantes da protease e da transcriptase reversa do gene <i>pol</i> do HIV-1, com base num alinhamento múltiplo com um total de 838 nucleótidos. B) Análise de <i>bootscanning</i> , por <i>SimPlot</i> v.3.5.1, da sequência PT_53B	70
Figura 20. Representação gráfica da prevalência das mutações associadas a resistência aos inibidores da protease (PIs) identificadas na região codificante da protease das sequências em estudo.....	73
Figura 21. Alinhamento múltiplo das 67 sequências aminoacídicas putativas da protease das estirpes do HIV-1 em estudo.....	75
Figura 22. Representação gráfica da sequência <i>weblogo</i> das 67 sequências de aminoácidos putativas da protease do HIV-1 em estudo.....	78

Índice de Tabelas

Tabela 1. Descrição das funções das proteínas do HIV-1 codificadas pelos genes estruturais.....	8
Tabela 2. Descrição das funções das proteínas do HIV-1 codificadas pelos genes reguladores.....	9
Tabela 3. Descrição das funções das proteínas do HIV-1 codificadas pelos genes acessórios.....	9
Tabela 4. Descrição do modo de acção das classes terapêuticas utilizadas na terapia anti-retroviral do HIV-1.....	19
Tabela 5. <i>Primers</i> utilizados nas reacções de PCR.....	37
Tabela 6. Condições de amplificação aplicadas na primeira PCR.....	39
Tabela 7. Condições de amplificação aplicadas na segunda PCR.....	39
Tabela 8. Condições de amplificação aplicadas na terceira PCR.....	40
Tabela 9. Caracterização genotípica das sequências em estudo.....	67
Tabela 10. “Assinaturas genéticas” identificadas nas sequências dos subtipos B, C e G ($p<0,01$), presentes na sequência aminoacídica putativa da protease das estirpes de HIV-1 em estudo.....	76

Lista de Símbolos, Acrónimos e Abreviaturas

A	Adenina, base orgânica constituinte dos nucleótidos
ABC	Abacavir
APOBEC	Do inglês, <i>apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like</i>
ATV	Atazanavir
AZT	Azidotimidina (Zidovudina)
BLAST	Do inglês, <i>Basic Local Alignment Search Tool</i>
C	Citosina, base orgânica constituinte dos nucleótidos
CA	Proteína da cápside do HIV-1
CCR5	Receptor de β -quimiocinas da classe CC (do inglês, <i>CC chemokine receptor 5</i>)
cDNA	Ácido desoxirribonucleico complementar (do inglês, <i>complementary deoxyribonucleic acid</i>)
CD4	Receptor celular (do inglês, <i>cluster of differentiation 4</i>)
CMSP	Células mononucleadas do sangue periférico
CRF	Forma recombinante circulante (do inglês, <i>circulating recombinant form</i>)
CXCR4	Receptor de β -quimiocinas da classe CXC (do inglês, <i>CXC chemokine receptor 4</i>)
DMSO	Dimetilsulfóxido
DNA	Ácido desoxirribonucleico (do inglês, <i>deoxyribonucleic acid</i>)
dNTP	Desoxirribonucleósidos 5'-trifosfatados
D.O.₆₀₀	Densidade óptica medida no comprimento de onda de 600 nm
DRV	Darunavir

EDTA	Ácido etilenodiamino tetracético
FDA	Entidade governamental dos EUA que regula a comercialização de alimentos e medicamentos (do inglês, <i>Food and Drug Administration</i>)
FPV	Fosamprenavir
FTC	Emtricitabina
g	Força gravitacional
G	Guanina, base orgânica constituinte dos nucleótidos
gp	Glicoproteína
HAART	Terapia anti-retroviral de alta eficácia (do inglês, <i>highly active antiretroviral therapy</i>)
HIV	Vírus da imunodeficiência humana (do inglês, <i>human immunodeficiency virus</i>)
IAS-USA	<i>International Antiviral Society-USA</i>
IC₅₀	Concentração mínima inibitória de 50% (neste trabalho, concentração de um fármaco necessária para inibir em 50% a replicação viral)
IDV	Indinavir
IN	Integrase do HIV-1
INSA	Instituto Nacional de Saúde Dr. Ricardo Jorge
IPTG	Isopropil-β-D-1-tiogalactopiranosídeo
IUPAC	Do inglês, <i>International Union of Pure and Applied Chemistry</i>
LB	Meio de lisogenia, do inglês <i>lysogeny broth</i> (vulgarmente designado de Luria-Bertani)
LPV	Lopinavir
LTR	Regiões repetitivas longas nas extremidades do genoma proviral do HIV (do inglês, <i>long terminal repeats</i>)
MA	Proteína da matriz do HIV-1

MEGA	Do inglês, <i>Molecular Evolutionary Genetics Analysis</i>
MHC	Complexo principal de histocompatibilidade (do inglês, <i>major histocompatibility complex</i>)
ml	Mililitro
mM	Milimolar
mRNA	Ácido ribonucleico mensageiro (do inglês, <i>messenger ribonucleic acid</i>)
NC	Proteína da nucleocápside do HIV-1
NFV	Nelfinavir
NNRTI	Inibidores não nucleosídicos da transcriptase reversa (do inglês, <i>non-nucleoside reverse transcriptase inhibitors</i>)
NRTI	Inibidores nucleosídicos da transcriptase reversa (do inglês, <i>nucleoside reverse transcriptase inhibitors</i>)
NVP	Nevirapina
pb	Par de bases
pbs	Sítio de ligação do <i>primer</i> (do inglês, <i>primer binding site</i>)
PCR	Reacção em cadeia da polimerase (do inglês, <i>polymerase chain reaction</i>)
PI	Inibidores da protease (do inglês, <i>protease inhibitors</i>)
PIC	Complexo de pré-integração (do inglês, <i>pre-integration complex</i>)
pmol	Picomole
PR	Protease do HIV-1
p/v	Peso por volume
RNA	Ácido ribonucleico (do inglês, <i>ribonucleic acid</i>)
RNase	Ribonuclease
rpm	Revoluções por minuto
RT	Transcriptase reversa do HIV-1 (do inglês, <i>reverse transcriptase</i>)

RTC	Complexo de transcrição reversa (do inglês, <i>reverse transcription complex</i>)
RT-PCR	Reacção em cadeia da polimerase antecedida de transcrição reversa (do inglês, <i>reverse transcription-polymerase chain reaction</i>)
RTV	Ritonavir
sida	Síndrome da imunodeficiência adquirida
SIV	Vírus da imunodeficiência símia (do inglês, <i>simian immunodeficiency virus</i>)
SQV	Saquinavir
ssRNA	Ácido ribonucleico de cadeia simples (do inglês, <i>single-stranded ribonucleic acid</i>)
SU	Glicoproteína de superfície do HIV-1
T	Timina, base orgânica constituinte dos nucleótidos
TAE	Tampão Tris-acetato-EDTA
TAR	Elemento de resposta à transactivação (por Tat) (do inglês, <i>transactivation response element</i>)
TDF	Tenofovir (disoproxil fumarato)
TM	Glicoproteína transmembranar do HIV-1
TPV	Tipranavir
tRNA	Ácido ribonucleico de transferência (do inglês, <i>transfer ribonucleic acid</i>)
TSS	Solução de transformação e armazenamento (do inglês, <i>transformation and storage solution</i>)
UNAIDS	Programa conjunto das Nações Unidas sobre HIV/sida (do inglês, <i>Joint United Nations Programme on HIV/AIDS</i>)
URF	Forma recombinante única (do inglês, <i>unique recombinant form</i>)

V	Volt
v/v	Volume por volume
X-Gal	5-bromo-4-cloro-3-indolil- β -D-galactopiranósido
°C	Grau Celsius
μg	Micrograma
μl	Microlitro
μM	Micromolar
3TC	Lamivudina

1. Introdução

1.1. Vírus da imunodeficiência humana tipo 1

O vírus da imunodeficiência humana tipo 1 (HIV-1, do inglês *human immunodeficiency virus type 1*) é responsável por uma das infecções com maior impacto na saúde pública mundial, a qual atingiu dimensões sem precedentes (Hemelaar, 2012). Desde o momento da sua descoberta até à data, tem sido alvo de um intenso e incessante estudo (Sharp & Hahn, 2011).

O HIV-1 pertence à família *Retroviridae*, à subfamília *Orthoretrovirinae*, ao género *Lentivirus* e ao grupo dos lentivírus dos primatas (Harrach et al., 2012). Este retrovírus infecta células do sistema imunitário, nomeadamente linfócitos T CD4⁺, causando a sua progressiva eliminação. Os linfócitos T CD4⁺ são essenciais no combate às infecções e a sua destruição conduz a uma supressão imunitária. Com a progressão da infecção pelo HIV, o sistema imunitário fica cada vez mais debilitado e surgem então as infecções oportunistas. Estas estão normalmente associadas ao estágio final de progressão da infecção, definido por uma contagem de linfócitos T CD4⁺ inferior a 200 células por µl, a que se dá o nome de síndrome da imunodeficiência adquirida (sida) (Kuritzkes & Koup, 2013; INSIGHT START Study Group, 2015).

A introdução da terapia anti-retroviral teve um enorme impacto no curso da infecção por HIV-1, melhorando significativamente a qualidade e a esperança média de vida dos indivíduos infectados. No entanto, a intolerância aos efeitos secundários e a emergência de estirpes virais resistentes representam alguns dos principais obstáculos no sucesso dos regimes terapêuticos actuais (Volberding & Deeks, 2010; European Centre for Disease Prevention and Control, 2012).

1.1.1. Origem do HIV-1 e história da pandemia

Os primeiros casos de sida foram descritos pela primeira vez em 1981 nos Estados Unidos da América, após um aumento súbito do número de casos de infecção por *Pneumocystis jirovecii* e de sarcoma de Kaposi, reportados em homens jovens homossexuais que apresentavam uma profunda deficiência imunitária, sem causa aparente (Centers for Disease Control and Prevention (CDC), 1981a, 1981b; Gottlieb et al., 1981; Centers for Disease Control and Prevention (CDC), 1982). A etiologia viral da sida foi descrita, em 1983, por Montagnier e colaboradores, no Instituto Pasteur de Paris, após o isolamento de um novo retrovírus a partir de gânglios linfáticos de um

indivíduo infectado (Barré-Sinoussi et al., 1983). A este novo retrovírus foi mais tarde atribuído o nome de vírus da imunodeficiência humana (HIV) (Coffin et al., 1986).

Em 1986, foi isolado um outro retrovírus (Clavel et al., 1986), que causava sida em pacientes na África Ocidental, morfológicamente semelhante ao primeiro, mas divergente a nível antigénico. Este novo vírus, designado de HIV-2, por contraposição ao primeiro vírus isolado, a partir de então designado HIV-1, era filogeneticamente distante do HIV-1 mas estreitamente relacionado com um vírus símio, genericamente denominado de vírus da imunodeficiência símia (SIV, do inglês *simian immunodeficiency virus*). Estes vírus, geralmente não patogénicos para os seus hospedeiros naturais, formam uma linhagem monofilética única com o HIV dentro do grupo dos lentivírus (Sharp & Hahn, 2011).

Em 1989, foram encontrados os SIV mais proximamente relacionados com o HIV-1 e o HIV-2 em primatas africanos (Hirsch et al., 1989; Peeters et al., 1989). O SIV_{smm}, encontrado em mangabeis fuliginosos (*Cercocebus atys*), apresenta uma elevada homologia genética e fenotípica com o HIV-2 (Gao et al., 1992). Pelo contrário, o HIV-1 está filogeneticamente mais relacionado com o SIV_{cpz} isolado dos chimpanzés da subespécie *Pan troglodytes troglodytes* (Gao et al., 1999). Além disso, há uma coincidência geográfica assinalável entre o habitat natural dos mangabeis fuliginosos e da subespécie de chimpanzés mencionada, com a endemicidade do HIV-2 na África Ocidental e do HIV-1 na África Centro-Ocidental (Gabão, República Democrática do Congo, Camarões) (Figura 1), respectivamente (Takebe et al., 2008; Sharp & Hahn, 2010, 2011).

Assim sendo, tornou-se evidente que o HIV emergiu na população humana através de transmissão de carácter zoonótico dos primatas não humanos infectados com SIV. Esta transmissão ocorreu, provavelmente, pelo contacto com sangue e/ou fluidos corporais de primatas infectados, no contexto da caça e abate de animais para consumo humano, assim como da sua captura e manutenção como animais de estimação (Hahn et al., 2000; Peeters et al., 2002). Esta teoria é sustentada por fortes evidências de natureza genética, filogenética, geográfica, epidemiológica e sociológica (Hahn et al., 2000).

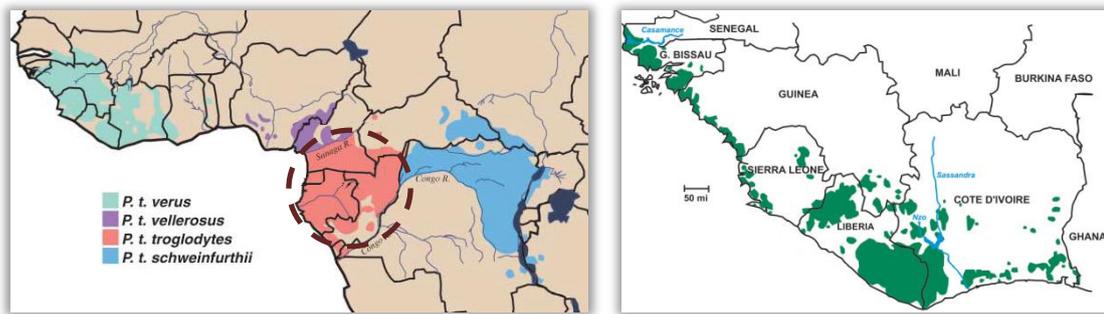


Figura 1. Área geográfica de distribuição natural de *Pan troglodytes troglodytes* (à esquerda) [adaptado de (Keele et al., 2006)] e de *Cercocebus atys* (à direita) [adaptado de (Santiago et al., 2005)].

Em suma, tanto o HIV-1 como o HIV-2 tiveram origem a partir de transmissão de carácter zoonótico, tendo ocorrido vários eventos de transmissão independentes, dos primatas não humanos para o Homem, resultando em várias linhagens de HIV. No HIV-1 ocorreram, pelo menos, quatro eventos de transmissão distintos, dando origem aos diversos grupos genéticos existentes (M, N, O e P), enquanto no HIV-2 aconteceram, no mínimo, oito transmissões independentes (grupos A-H) (Hemelaar, 2012).

Relativamente ao HIV-1, estudos filogenéticos indicam que os grupos M (do inglês, *main*) e N (do inglês, *non-M/non-O* ou *new*) são filogeneticamente mais relacionados com o SIV_{cpz} encontrado nos chimpanzés da subespécie *Pan troglodytes troglodytes*. Relativamente aos grupos O (do inglês, *outlier*) e P, estudos mais recentes demonstram que estes são mais proximamente relacionados com o SIV_{gor} encontrado em gorilas da subespécie *Gorilla gorilla gorilla*, no sul dos Camarões (Van Heuverswyn et al., 2006; Plantier et al., 2009; D'arc et al., 2015).

O grupo M foi o primeiro a ser descrito e é o responsável pela pandemia (Hemelaar, 2012). Nos últimos anos, foi proposto que a epidemia do HIV-1 teve origem na África Central, nomeadamente na República Democrática do Congo, devido, entre outros, à elevada diversidade genética aqui encontrada (Vidal et al., 2000; Kalish et al., 2004). Estudos mais recentes também o sugerem, indicando que o epicentro da epidemia foi, provavelmente, Kinshasa, capital da República Democrática do Congo, e que esta terá tido o seu início por volta de 1920 (Worobey et al., 2008; Faria et al., 2014). Nas décadas seguintes, o HIV-1 (grupo M) teve um período de expansão lento na África Central, associada à rede de transportes que ligava os principais centros populacionais a outras regiões de África subsariana. A partir de 1960 deu-se uma

propagação rápida do vírus, presumivelmente associada ao aumento da migração entre os países, à implementação de programas terapêuticos e de vacinação em larga escala, com utilização de agulhas não esterilizadas, e/ou a alterações significativas no comportamento sexual, com aumento da promiscuidade e da prevalência de infecções de transmissão sexual (Worobey et al., 2008; de Sousa et al., 2012; Pépin, 2013; Faria et al., 2014).

1.1.2. Epidemiologia

De acordo com o *Joint United Nations Programme on HIV/AIDS* (UNAIDS), no final de 2014, havia um total de 36,9 milhões [34,3 – 41,4 milhões] de pessoas infectadas com o HIV, das quais 25,8 milhões [24,0 – 28,7 milhões] viviam na África subsariana (Figura 2). Nesse ano, foram diagnosticados 2 milhões [1,9 – 2,2 milhões] de novos casos de infecção e reportada a morte de 1,2 milhões [0,98 – 1,6 milhões] de pessoas por doenças relacionadas com a sida (UNAIDS, 2015b).

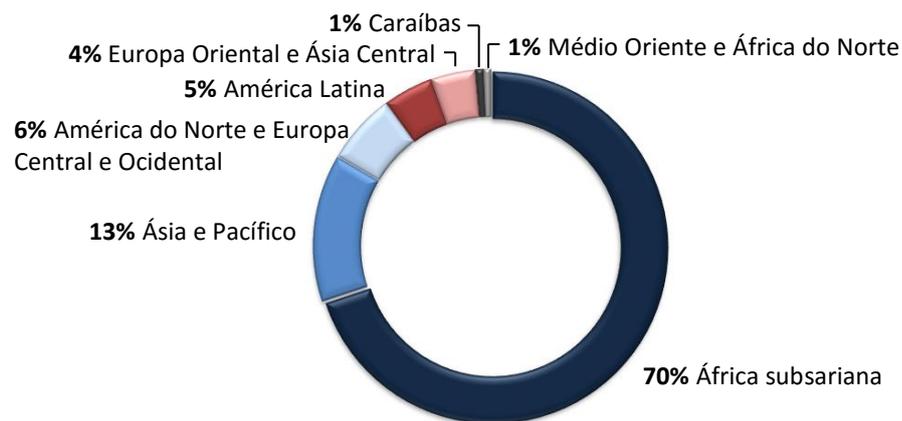


Figura 2. Distribuição relativa do número de indivíduos infectados por HIV, pelas diferentes regiões geográficas, no final de 2014 (UNAIDS, 2015b).

A UNAIDS estimou ainda que cerca de 17,4 milhões [16,1 – 20 milhões] de mulheres (com idade superior a 15 anos) estavam infectadas com o HIV, das quais 1,5 milhões foram mães, sem acesso a terapia anti-retroviral. A transmissão vertical representa a principal causa de infecção em crianças e, embora o número de novas infecções tenha sofrido um decréscimo de 58% desde 2000 (UNAIDS, 2015a), esta permanece elevada. Em 2014, 2,6 milhões [2,4 – 2,8 milhões] de crianças com idade inferior a 15 anos estavam infectadas com o HIV, dos quais 220000 casos correspondiam a novas infecções (UNAIDS & AIDSinfo, 2015; UNAIDS, 2015b).

Em Portugal, segundo o Instituto Nacional de Saúde Dr. Ricardo Jorge (INSA), até 30 de Junho de 2015, foram notificados 52694 casos cumulados de infecção por HIV, desde 1983. No ano de 2014, foram diagnosticados 920 novos casos de infecção, incluindo 167 novos casos de sida, comprovando uma percentagem de diagnósticos tardios muito elevada. Além disso, foram ainda registados 196 óbitos de indivíduos infectados pelo HIV (Departamento de Doenças Infecciosas do INSA & Programa Nacional para a Infecção VIH/SIDA, 2015).

Relativamente à infecção pelo HIV em crianças, com idade inferior a 15 anos, foram registados 526 casos desde 1983, sendo a transmissão vertical a principal via de transmissão (79,1% dos casos). O número destes casos tem vindo a diminuir ao longo dos últimos anos, fundamentalmente devido à implementação do rastreio da infecção na gravidez, no final dos anos 90 do século passado. Alguns dos casos registados nos últimos anos devem-se a diagnósticos realizados anos após o nascimento e/ou casos importados. No ano de 2014, foram diagnosticados seis novos casos de infecção pelo HIV-1 em crianças, todas com idades até aos sete anos de idade. A maioria das crianças residia na região de Lisboa (n=5) e a sua origem geográfica era Portugal (n=3) ou a África subsariana (n=3). Em todos os casos foi identificada a transmissão mãe-filho como via de transmissão do vírus (Departamento de Doenças Infecciosas do INSA & Programa Nacional para a Infecção VIH/SIDA, 2015).

Embora na última década o número de casos registados tenha uma tendência decrescente, quer nos adultos quer nas crianças (<15 anos), Portugal continua a ser o país da Europa Ocidental que apresenta uma taxa anual de novos diagnósticos de infecção pelo HIV mais elevada (Departamento de Doenças Infecciosas do INSA & Programa Nacional para a Infecção VIH/SIDA, 2015).

1.1.3. Organização estrutural e genómica do HIV-1

Os viriões apresentam uma morfologia aproximadamente esférica, com cerca de 100 nm de diâmetro, possuindo um invólucro lipídico, derivado da membrana citoplasmática das células hospedeiras, que contém as glicoproteínas de superfície (SU, do inglês *surface protein*) e transmembranar (TM, do inglês *transmembrane protein*). Rodeada pelo invólucro lipídico, encontra-se a matriz viral, constituída pela proteína da matriz (MA, do inglês *matrix protein*), que envolve a cápside viral, composta pela

proteína da cápside (CA, do inglês *capsid protein*), em forma de cone truncado. A cápside contém o genoma viral, associado às proteínas da nucleocápside (NC, do inglês *nucleocapsid protein*), as enzimas protease (PR, do inglês *protease*), transcriptase reversa (RT, do inglês *reverse transcriptase*) e integrase (IN, do inglês *integrase*). A partícula viral contém ainda as proteínas acessórias Vif, Nef e Vpr (Figura 3) (Turner & Summers, 1999; Sierra et al., 2005; Goff, 2013).

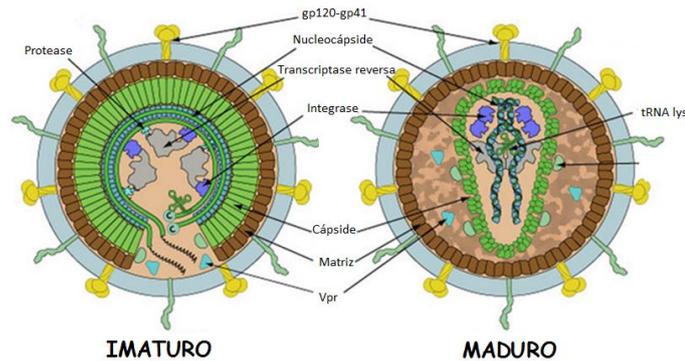


Figura 3. Representação esquemática da estrutura da partícula viral do HIV-1 imaturo (à esquerda) e maduro (à direita) [Adaptado de (ViralZone, 2015)].

O genoma viral é constituído por duas moléculas de RNA de cadeia simples (ssRNA) e polaridade positiva com cerca de 9,7 kb de comprimento. O genoma de RNA é obtido através da maquinaria enzimática da célula hospedeira e, por isso, contém características do mRNA celular, nomeadamente 5'-cap e cauda 3'-polyA (Sierra et al., 2005; Goff, 2013; Los Alamos National Laboratory, 2015b). O genoma do HIV-1 contém 9 genes que codificam 16 proteínas diferentes, como é representado na Figura 4.

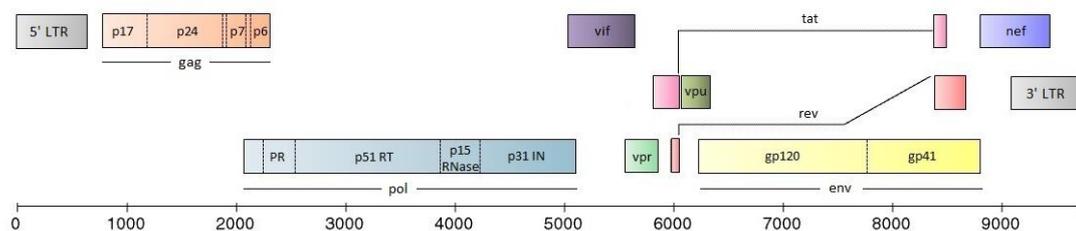


Figura 4. Representação esquemática do genoma de HIV-1 (escala em kb) [adaptado de (Los Alamos National Laboratory, 2015b)].

O genoma do HIV, na sua forma proviral, apresenta em ambas as extremidades regiões reguladoras da expressão génica viral, designados de LTR (do inglês, *long terminal repeats*). Estes são formados durante o processo de transcrição reversa, apresentando a estrutura U3-R-U5. A região LTR contém regiões reguladoras

importantes que promovem o início da transcrição do genoma proviral (Freed & Martin, 2013; Los Alamos National Laboratory, 2015b). Relativamente à região codificante do genoma, esta inclui os genes estruturais *gag*, *pol* e *env*, que codificam proteínas essenciais da partícula viral madura (*gag* e *env*), bem como as enzimas virais (*pol*) (Los Alamos National Laboratory, 2015b). O gene *gag* codifica a proteína precursora Pr55^{Gag}, que é processada nas proteínas p17 (MA), p24 (CA), p7 (NC) e p6. O gene *pol* codifica as enzimas virais protease (PR), transcriptase reversa (RT) e integrase (IN). Estas enzimas são sintetizadas como proteínas de fusão na poliproteína precursora Pr160^{Gag-Pol}, sendo depois processadas pela protease viral. O gene *env* codifica uma poliproteína precursora gp160, posteriormente processada por uma protease celular (furina) nas glicoproteínas virais gp120 (SU) e gp41 (TM). As funções das proteínas codificadas pelos genes estruturais *gag*, *pol* e *env* estão descritas na Tabela 1 (Turner & Summers, 1999; Freed, 2001; Freed & Martin, 2013; Los Alamos National Laboratory, 2015b).

Tabela 1. Descrição das funções das proteínas do HIV-1 codificadas pelos genes estruturais.

Genes estruturais		
<i>gag</i>	MA	Proteína estrutural da matriz. Implicada no processo de montagem do virião. Direciona as proteínas Gag e Gag-Pol para a membrana citoplasmática e promove a sua ligação.
	CA	Proteína estrutural da cápside. Envolvida no processo de montagem das partículas virais.
	NC	Proteína estrutural da nucleocápside. Responsável pelo reconhecimento e encapsidação do RNA genómico.
	p6	Envolvida na libertação dos viriões para o exterior da célula hospedeira.
<i>pol</i>	PR	Efectua a clivagem proteolítica das poliproteínas Gag e Gag-Pol, originando partículas virais infecciosas.
	RT	Responsável pela conversão do RNA genómico de cadeia simples em DNA complementar de cadeia dupla (transcrição reversa).
	IN	Realiza a inserção do DNA de dupla cadeia no DNA da célula hospedeira.
<i>env</i>	SU	Glicoproteína estrutural do invólucro. Responsável pela ligação do virião ao receptor celular CD4 e, posteriormente, aos co-receptores presentes na membrana da célula hospedeira.
	TM	Glicoproteína estrutural do invólucro. Responsável pela fusão entre o invólucro viral e a membrana celular.

Os outros genes, denominados genes reguladores (*tat* e *rev*) e acessórios (*vif*, *vpr*, *vpu*, *nef*), codificam proteínas que modulam o metabolismo da célula hospedeira e facilitam as diferentes etapas do ciclo replicativo do vírus (Turner & Summers, 1999; Freed, 2001; Freed & Martin, 2013; Los Alamos National Laboratory, 2015b) (Tabelas 2 e 3).

Tabela 2. Descrição das funções das proteínas do HIV-1 codificadas pelos genes reguladores.

Genes reguladores		
<i>tat</i>	Tat	Liga-se a uma estrutura <i>hairpin</i> dos mRNAs virais, o elemento TAR (do inglês, <i>transactivation response element</i>), activando o início da transcrição e o alongamento dos transcritos a partir da LTR viral, por indução da fosforilação da RNA polimerase II celular.
<i>rev</i>	Rev	Responsável pela exportação dos RNAs virais não processados e parcialmente processados para o exterior do núcleo.

Tabela 3. Descrição das funções das proteínas do HIV-1 codificadas pelos genes acessórios.

Genes acessórios		
<i>vif</i>	Vif	Potencia a infecciosidade viral. Actua sobre as proteínas celulares APOBEC que desaminam o DNA durante a transcrição reversa.
<i>vpr</i>	Vpr	Desempenha um papel importante no transporte do complexo de pré-integração (PIC) para o núcleo, em células que não se encontram em divisão celular; induz a permanência das células na fase G2 do ciclo celular e a sua apoptose.
<i>vpu</i>	Vpu	Presente apenas no HIV-1; induz a degradação do receptor CD4, potenciando a libertação das partículas virais a partir da membrana citoplasmática das células infectadas.
<i>nef</i>	Nef	Proteína multifuncional, essencial para a replicação do vírus <i>in vivo</i> ; modula negativamente a expressão do receptor CD4 e das moléculas MHC classe I, potenciando a infecciosidade viral.

1.1.4. Ciclo replicativo

O ciclo replicativo do HIV-1 (Figura 5) ocorre numa série de eventos que podem ser divididos, de um modo geral, em duas fases: precoce e tardia. A fase precoce tem início com o reconhecimento das células alvo pelo virião e envolve todos os processos até à integração do DNA genómico complementar no cromossoma da célula hospedeira.

Na fase tardia ocorre a expressão do DNA proviral, incluindo todos os processos até à formação de novas partículas virais infecciosas (Turner & Summers, 1999; Freed, 2001).

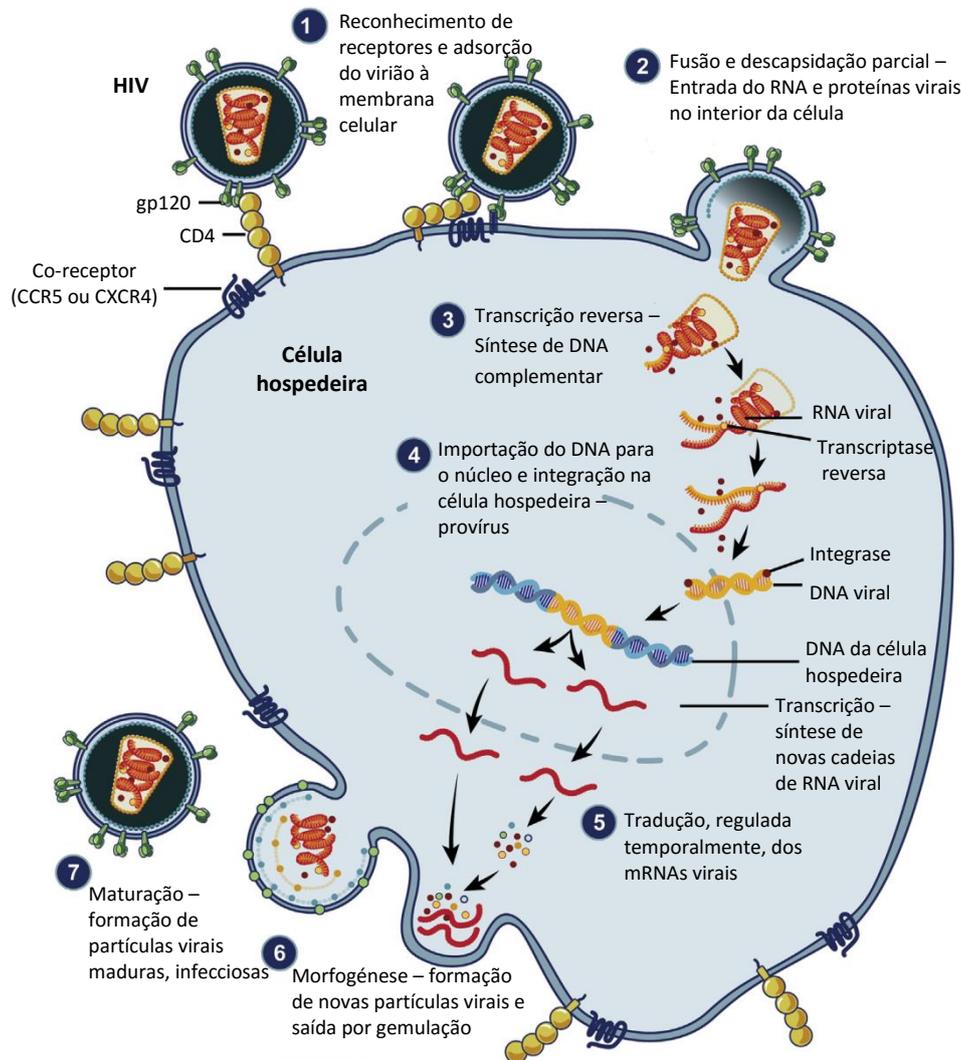


Figura 5. Representação do ciclo replicativo do HIV-1 [Adaptado de (Pau & George, 2014)].

1.1.4.1. Adsorção e Entrada

O ciclo replicativo do HIV-1 tem início após a adsorção do virião à membrana celular da célula susceptível. Este processo é mediado pela interação entre a glicoproteína viral gp120 e o receptor celular CD4. Embora a ligação do virião ao receptor CD4 seja essencial, a sua subsequente interação, via gp120, com um co-receptor celular é imprescindível para a ocorrência de infecção. Os principais co-receptores são os receptores celulares de quimiocinas CCR5 e CXCR4.

Após a ligação da partícula viral ao receptor CD4, dá-se uma alteração conformacional na gp120, aumentando a sua afinidade para o co-receptor, ao qual se liga. A ligação da gp120 ao co-receptor induz, por sua vez, alterações conformacionais na glicoproteína viral gp41, expondo o seu extremo N-terminal (também chamado de péptido de fusão). Estas alterações permitem a inserção do péptido de fusão na membrana da célula hospedeira, o que conduz finalmente à sua fusão com o invólucro viral e, conseqüentemente, à libertação da cápside viral no citoplasma da célula (Freed, 2001; Sierra et al., 2005; Freed & Martin, 2013).

1.1.4.2. Descapsidação, Transcrição Reversa, Importação Nuclear e Integração

Após a entrada da cápside no citoplasma da célula hospedeira, ocorre uma descapsidação parcial. Este processo é facilitado pela proteína celular ciclofilina A, a qual contribui também para a formação do complexo de transcrição reversa (RTC, do inglês *reverse transcription complex*). Este complexo é constituído pelo RNA genómico, e pelas proteínas virais RT, MA, CA, NC, IN e Vpr, a partir do qual é sintetizada uma molécula de DNA complementar de dupla cadeia (Freed, 2001; Freed & Martin, 2013).

De um modo sucinto, a transcrição reversa tem início após a ligação do *primer* tRNA^{Lys3} celular ao sítio pbs (do inglês, *primer binding site*), localizado na extremidade 5' do RNA genómico, a partir do qual a RT transcreve uma cadeia de DNA de polaridade negativa. Após a síntese desta cadeia de DNA, o motivo RNaseH da RT degrada a cadeia de RNA (que serviu de matriz), dando início à polimerização da segunda cadeia de DNA, de polaridade positiva, originando assim uma molécula de DNA de dupla cadeia. Este DNA complementar de dupla cadeia associado às proteínas virais MA, RT, IN e Vpr formam o complexo de pré-integração (PIC), que é transportado, em associação com os microtúbulos do citoesqueleto, até ao invólucro nuclear. A passagem do PIC através dos poros nucleares ocorre após uma descapsidação total, sendo dependente de Vpr. A posterior integração do DNA viral no genoma da célula hospedeira, dando origem ao provírus, é realizada pela IN viral e requer a acção de proteínas celulares (Freed, 2001; Freed & Martin, 2013).

1.1.4.3. Transcrição, Exportação Nuclear do RNA Viral e Síntese Proteica

Após a integração do DNA viral num cromossoma da célula hospedeira, o DNA é utilizado como matriz para a transcrição dos mRNAs virais que irão codificar as proteínas estruturais, reguladoras e acessórias. Dependendo do estado de activação da célula hospedeira, o provírus pode permanecer silenciado ou transcricionalmente activo, para a síntese eficiente de proteínas virais e de novas cópias de RNA genómico. A transcrição inicia-se pela ligação da RNA polimerase II celular à 5'-LTR viral, produzindo-se, ao longo do tempo, um elevado número de moléculas de RNA que podem sofrer vários níveis de *splicing* (nenhum, parcial ou múltiplo). Os RNAs que não sofrem *splicing* codificam as proteínas Pr55^{Gag} e Pr160^{Gag-Pol} ou são, mais tarde, utilizados como RNA genómico na génese de novos viriões. Os mRNAs maduros que resultam de um *splicing* parcial têm entre 4,3 a 5,5 kb de comprimento e são traduzidos para dar origem às proteínas Env, Vif, Vpr e Vpu. Os mRNAs maduros que resultam de um *splicing* múltiplo, com tamanho compreendido entre 1,7 e 2 kb, são os primeiros a ser produzidos, codificando as proteínas precoces Tat, Rev e Nef.

Após a transcrição e processamento dos diversos RNAs virais dá-se a sua exportação para o citoplasma (em alguns casos, dependente de Rev), onde é realizada a tradução dos vários mRNA virais, em momentos temporais bem definidos, produzindo-se assim todas as proteínas essenciais para o ciclo replicativo (Freed, 2001; Freed & Martin, 2013).

1.1.4.4. Morfogénese e Maturação

Na fase tardia do ciclo replicativo, o processo de montagem de novas partículas virais é então iniciado, no qual a poliproteína precursora Pr55^{Gag} tem um papel fundamental. O componente MA é responsável pela sua interacção com a membrana celular, enquanto o componente NC promove a encapsidação do genoma viral. Simultaneamente, dá-se o transporte das glicoproteínas gp120 e gp41 (resultantes da clivagem da gp160 por uma protease celular), via vesículas do sistema de Golgi, para a membrana celular. A gp41 associada à gp120 vai interagir com Pr55^{Gag}, permitindo assim a sua acumulação ao nível da membrana. Após o processo de montagem das partículas virais estar finalizado, dá-se a sua libertação por gemulação, processo no qual são requeridas as proteínas virais p6^{Gag} e Vpu, iniciando-se o processo de maturação

promovido pela protease viral. A protease vai clivar as poliproteínas precursoras Pr55^{Gag} e Pr160^{Gag-Pol}, conduzindo a uma série de rearranjos estruturais, dando origem a partículas virais maduras e, conseqüentemente, infecciosas (Freed, 2001; Freed & Martin, 2013).

1.1.5. Diversidade genética do HIV-1

Filogeneticamente, o HIV-1 é classificado em quatro grupos (M, N, O e P), como referido anteriormente, cujas distribuições geográficas variam consideravelmente. O grupo O encontra-se limitado a indivíduos com origem na África Central, essencialmente nos Camarões, Gabão e Guiné Equatorial. Por sua vez, os vírus dos grupos N e P foram apenas encontrados num pequeno número de indivíduos nos Camarões. O grupo M, pelo contrário, encontra-se distribuído por todo o mundo (Takebe et al., 2008; Tebit & Arts, 2011; Hemelaar, 2012). Durante o curso da epidemia do grupo M na população humana, as sequências nucleotídicas virais foram divergindo geneticamente, sendo necessário implementar um sistema de classificação (Hemelaar, 2012). Deste modo, o grupo M compreende nove subtipos (A, B, C, D, F, G, H, J e K) alguns dos quais, por sua vez, são ainda divididos em sub-subtipos. Estes últimos formam linhagens distintas dentro de um determinado subtipo, não tendo, no entanto, um grau de divergência genético suficientemente elevado para serem classificados como um novo subtipo (Robertson et al., 2000; Takebe et al., 2008; Hemelaar, 2012). O subtipo A está dividido em quatro sub-subtipos (A1, A2, A3 e A4), enquanto o F compreende dois sub-subtipos (F1 e F2) (Tebit & Arts, 2011). Por outro lado, os subtipos B e D deveriam corresponder a dois sub-subtipos, do mesmo subtipo genético, pois estão mais proximamente relacionados entre si do que com os restantes. No entanto, a nomenclatura originalmente proposta foi mantida, de modo a permitir a consistência dos artigos anteriormente publicados (Takebe et al., 2008).

O grupo M do HIV-1 inclui ainda numerosas formas genéticas recombinantes entre subtipos distintos. Os eventos de recombinação genética estão presentes nos mecanismos normais de replicação dos retrovírus, desempenhando um papel muito importante na geração de diversidade genética. Para que ocorra a formação de vírus recombinantes é necessário que duas estirpes virais distintas infectem a mesma célula, num evento simultâneo (co-infecção) ou sequencial (super-infecção) (Nájera et al.,

2002). Deste processo podem resultar novas partículas virais com genomas heterodiméricos. Tendo em conta que no processo de transcrição reversa a RT alterna frequentemente entre as duas moléculas de RNA do genoma viral, a presença de genomas heterodiméricos pode dar origem a genomas mosaico na progénie viral (Nájera et al., 2002; Takebe et al., 2008). Os vírus recombinantes, entre os diferentes subtipos do grupo M, podem ser designados por formas recombinantes circulantes, ou CRFs (do inglês, *circulating recombinant form*), ou por formas recombinantes únicas, ou URFs (do inglês, *unique recombinant form*). As CRFs correspondem a vírus com genomas mosaico que foram identificados em diversos indivíduos, desempenhando um papel importante na pandemia do HIV-1. Para que seja atribuída a categoria de CRF a uma nova forma de genoma mosaico têm de ser analisados três genomas virais completos, ou dois genomas completos juntamente com um terceiro genoma parcial que contenha diferentes segmentos genómicos, provenientes de indivíduos não relacionados do ponto de vista epidemiológico (Robertson et al., 2000; Nájera et al., 2002). As CRFs são designadas com um número de identificação (números consecutivos por ordem de descoberta), juntamente com os subtipos envolvidos (ex.: CRF02_AG). Quando na presença de mais de dois subtipos na origem do genoma mosaico atribui-se a designação de *cpx* (do inglês, *complex*) (Robertson et al., 2000; Nájera et al., 2002; Takebe et al., 2008). As URFs correspondem a vírus que apresentam um genoma recombinante, identificado em apenas um indivíduo ou num grupo de indivíduos relacionados a nível epidemiológico (Takebe et al., 2008). Até à data, estão descritas 75 CRFs do grupo M do HIV-1 (Los Alamos National Laboratory, 2015a).

1.1.5.1. Mecanismos de diversidade genética

O HIV-1 apresenta uma diversidade genética elevada, tal como foi descrito anteriormente. Esta característica permite ao vírus adaptar-se, contornando a pressão selectiva exercida quer pelo sistema imunitário do hospedeiro, quer pela terapia anti-retroviral (Santoro & Perno, 2013). Existem diversos factores que contribuem para a extraordinária heterogeneidade genética do HIV-1, nomeadamente:

- elevada taxa de mutação durante o processo de transcrição reversa (aproximadamente 3×10^{-5} mutações/nucleótido/ciclo replicativo), incluindo, a incapacidade da RT corrigir os nucleótidos incorrectamente incorporados na cadeia

nascente (ausência de actividade exonucleolítica 3'→5', vulgarmente designada por mecanismo de *proofreading*) (Cortez & Maldarelli, 2011; Hemelaar, 2012);

- elevada frequência de recombinação durante a transcrição reversa (6-7 transferências de cadeia/síntese proviral) (Cortez & Maldarelli, 2011; Freed & Martin, 2013);

- elevada taxa de replicação *in vivo*, originando um elevado número de novos viriões produzidos (10^9 viriões/dia), com tempos de semi-vida médios muito curtos (Freed & Martin, 2013; Santoro & Perno, 2013).

Estes mecanismos levam a que, num indivíduo infectado com HIV-1, ocorra a génese de uma população viral, geneticamente heterogénea, composta por micropopulações virais distintas, mas geneticamente relacionadas entre si, à qual se dá o nome de *quasiespecies* (Domingo et al., 2012; Hemelaar, 2012).

1.1.5.2. Diversidade genética e distribuição mundial

O HIV apresenta padrões epidemiológicos bastante heterogéneos por todo o mundo. Na África subsariana é onde se encontra uma maior diversidade genética. A maioria dos subtipos, sub-subtipos e muitas CRFs foram reportados na República Democrática do Congo e Camarões, sendo este facto consistente com a teoria de que o HIV-1 teve origem neste local, disseminando-se depois para todo o mundo (Takebe et al., 2008; Skar et al., 2011). À escala global, as formas genéticas mais prevalentes são os subtipos C, A, B, G e D e as CRF02_AG e CRF01_AE, estimando-se que cerca de metade dos indivíduos infectados o sejam pelo subtipo C (Takebe et al., 2008; Hemelaar et al., 2011; Wainberg & Brenner, 2012). A sua distribuição global não é, no entanto, de todo, homogénea, havendo alguns génotipos predominantes em determinadas regiões. O subtipo C é predominante na África Meridional e Oriental e Índia; o subtipo A na Europa Oriental, África Oriental e Ásia Central; o subtipo B na América do Norte, América Latina, Europa e Austrália; o subtipo G e a CRF02_AG são dominantes na África Ocidental e a CRF01_AE nos países do Sudeste Asiático (Hemelaar et al., 2011; Wainberg & Brenner, 2012).

Na Europa, segundo Abecasis et al. (2013), o subtipo B é predominante (66%), seguido dos subtipos A1, C, G e as CRF02_AG e CRF01_AE. Contudo, a sua distribuição geográfica não é uniforme entre os diversos países. A Polónia e a Eslovénia

representam os países com menor prevalência de subtipos não-B, contrariamente a Portugal, um dos países com maior percentagem destes subtipos. Em Portugal, predominam os subtipos B e G (Esteves et al., 2002; Abecasis et al., 2013; Carvalho et al., 2015), sendo os restantes genótipos mais frequentes os subtipos C, F1, A1 e as CRF14_BG e CRF02_AG (Esteves et al., 2003; Pereira et al., 2014; Carvalho et al., 2015).

1.1.6. Principais vias de transmissão

A maioria das infecções por HIV-1, em todo o mundo, ocorre por transmissão sexual. No entanto, existem outras vias de transmissão importantes, tais como, a transmissão vertical e a transmissão parentérica (transfusões de sangue contaminado, partilha de agulhas contaminadas, etc). O risco de transmissão varia com as diferentes vias de transmissão e depende de diversos factores, nomeadamente, carga viral, presença de co-infecções (tais como, outras doenças sexualmente transmissíveis, tuberculose, etc), factores comportamentais (uso de preservativo), entre outros. O tratamento anti-retroviral conduz a uma diminuição significativa do risco de transmissão (Skar et al., 2011; Kuritzkes & Koup, 2013).

1.1.6.1. Transmissão vertical

A transmissão vertical, ou transmissão mãe-filho, é a principal causa de infecção por HIV em crianças (Ahmad, 2011; da Silva et al., 2013). Na ausência de terapia anti-retroviral, o risco de transmissão pode variar entre os 20 e os 40%, sendo que esta pode ocorrer em três fases distintas (da Silva et al., 2013; Rosenvinge & Doerholt, 2013):

- durante a gravidez (*in utero*) – mais comum no final da gestação, podendo estar associada a um descolamento placentar;
- durante o parto – por contacto directo com sangue materno e/ou secreções genitais;
- durante o período de aleitamento materno.

A taxa de transmissão vertical pode ser reduzida para cerca de 2% se, durante a gravidez, for realizada uma terapia anti-retroviral eficaz e, por sua vez, uma diminuição significativa da carga viral (Cooper et al., 2002; Prieto et al., 2012). Outras medidas incluem a administração de anti-retrovirais durante o parto, a realização de cesariana

electiva, a não realização de amamentação natural e a administração de anti-retrovirais ao recém-nascido (Thorne & Newell, 2007).

1.1.7. Gene *pol*

O gene *pol*, com cerca de 3000 ribonucleótidos, na estirpe de referência HXB2, é uma das regiões mais conservadas do genoma do HIV-1, codificando as enzimas virais protease, transcriptase reversa e integrase, como referido anteriormente. Estas enzimas desempenham funções essenciais no ciclo replicativo do HIV, sendo alvos preferenciais da terapia anti-retroviral (Freed & Martin, 2013; Los Alamos National Laboratory, 2015b). A protease, tema central neste estudo, é descrita em seguida.

1.1.8. Protease: estrutura e função

A protease do HIV-1 pertence à família das proteases aspárticas, constituindo, mais especificamente, uma aspartil protease retroviral (ou retropepsina), e representa uma das enzimas proteolíticas mais estudadas nas últimas décadas (Louis et al., 2007; Konvalinka et al., 2015). A protease, na sua forma activa, apresenta-se como um homodímero simétrico, constituído por duas subunidades idênticas de 99 aminoácidos. Esta estrutura dimérica é estabilizada pela interacção entre os terminais N e C de cada monómero (posições aminoacídicas 1-4 e 96-99) e a sua dissociação leva a uma perda completa da capacidade catalítica da protease (Mager, 2001; Konvalinka et al., 2015). O centro activo da enzima é formado pela interacção entre as duas subunidades, sendo composto pela tríade catalítica Asp²⁵-Thr²⁶-Gly²⁷, sequência altamente conservada nas retropepsinas (Swanstrom & Wills, 1997; Turner & Summers, 1999; Mager, 2001). Os dois resíduos Asp²⁵ (um de cada cadeia) funcionam como resíduos catalíticos. Posicionadas acima do centro activo encontram-se duas regiões flexíveis, denominadas *flaps* (posições aminoacídicas 43 a 58, em cada cadeia), que se abrem e fecham sobre o centro activo (conformação aberta e fechada, respectivamente), permitindo a ligação e a estabilização do substrato ou inibidor (Freed & Martin, 2013; Konvalinka et al., 2015). Resumidamente, a nível estrutural existem três regiões fundamentais para a manutenção da actividade catalítica da protease, nomeadamente, o centro activo, os *flaps* e a zona de dimerização (Figura 6) (Weber & Agniswamy, 2009).

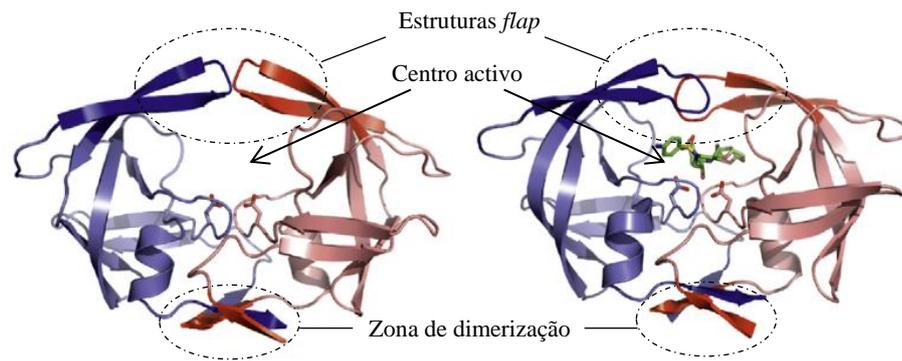


Figura 6. Representação da estrutura tridimensional da protease do HIV-1, em conformação aberta (à esquerda) e em conformação fechada (à direita) [Adaptado de (Konvalinka et al., 2015)].

A protease desempenha um papel fundamental no ciclo replicativo do HIV-1. É responsável pela clivagem proteolítica das poliproteínas Pr55^{Gag} e Pr160^{Gag-Pol}, dando origem às proteínas estruturais (MA, CA, NC e p6) e às enzimas virais (RT, PR e IN). Este processo, denominado maturação, leva a alterações conformacionais na partícula viral em formação e é indispensável para a formação de partículas virais infecciosas (Kohl et al., 1988). A inactivação da protease conduz à produção de partículas virais defectivas, tornando-a um excelente alvo terapêutico para o desenvolvimento de inibidores anti-retrovirais (Pokorná et al., 2009; Weber & Agniswamy, 2009).

1.2. Terapia anti-retroviral

O progresso da terapia anti-retroviral para o HIV constitui um dos maiores sucessos na história das doenças infecciosas (Looney et al., 2015). Inúmeros avanços no tratamento da infecção pelo HIV ocorreram desde a aprovação do primeiro fármaco anti-retroviral (zidovudina, AZT), em 1987, até aos dias de hoje, destacando-se a introdução da terapia anti-retroviral de alta eficácia (HAART, do inglês *highly active antiretroviral therapy*) em 1996 (Montaner et al., 2010; Looney et al., 2015). A HAART permitiu uma diminuição significativa da taxa de mortalidade e morbidade dos indivíduos infectados com o HIV, possibilitando que uma infecção invariavelmente fatal se tornasse numa infecção potencialmente crónica (Volberding & Deeks, 2010; Skar et al., 2011; Cotte et al., 2014).

Actualmente, a terapia anti-retroviral é simples, potente e geralmente bem tolerada, tendo como principal objectivo diminuir a replicação do vírus, não permitindo a sua evolução nem o desenvolvimento de resistências aos anti-retrovirais. Para isso, os níveis de RNA viral circulante (ou carga viral), devem manter-se abaixo dos níveis de detecção dos testes laboratoriais disponíveis para a sua quantificação. A terapia anti-retroviral pretende, também, concorrer para um aumento da contagem de linfócitos T CD4⁺, melhorando a função do sistema imunitário, diminuir a morbidade e o risco de transmissão e melhorar a qualidade e a esperança média de vida dos indivíduos infectados (Volberding & Deeks, 2010; US. Department of Health and Human Services et al., 2013; Pau & George, 2014; Looney et al., 2015). Os fármacos utilizados na terapia anti-retroviral têm como alvo diversas etapas do ciclo replicativo do HIV-1, sendo classificados de acordo com a denominação indicada na Tabela 4.

Tabela 4. Descrição do modo de acção das classes terapêuticas utilizadas na terapia anti-retroviral do HIV-1 (Pau & George, 2014).

Classes terapêuticas	
Inibidores de entrada	
Antagonista do co-receptor CCR5 (Maraviroc)	Inibidor de fusão (Enfuvirtide)
Liga-se ao co-receptor CCR5 presente na membrana celular da célula hospedeira, bloqueando a sua interacção com a gp120 viral.	Liga-se à gp41 do HIV, inibindo a fusão do invólucro viral com a membrana celular da célula hospedeira.
Inibidores da transcriptase reversa	
Análogos de nucleósidos/nucleótido - N(t)RTIs	Não análogos de nucleósidos - NNRTIs
Inibidores competitivos da RT, pois competem com os nucleótidos celulares durante a polimerização da nova cadeia de DNA complementar. A sua estrutura molecular impede que novos nucleótidos sejam adicionados à cadeia nascente, interrompendo assim a reacção de polimerização (Looney et al., 2015).	Inibidores não competitivos da RT. Ligam-se directamente a uma região hidrofóbica da transcriptase reversa, próxima do seu centro activo, induzindo alterações conformacionais na enzima que inibem a sua acção (Cortez & Maldarelli, 2011; Looney et al., 2015).

Inibidores da integrase

Inibem a formação de ligações covalentes entre as extremidades do DNA complementar viral e o DNA do hospedeiro, bloqueando assim a integração e a consequente formação do provírus (Looney et al., 2015).

Inibidores da protease

São inibidores competitivos que se ligam ao centro activo da protease, bloqueando a sua actividade enzimática. Deste modo, a protease é incapaz de processar proteoliticamente as poliproteínas precursoras virais, resultando em partículas virais imaturas, não infecciosas (Ali et al., 2010).

A HAART consiste na combinação de, pelo menos, três fármacos distintos de duas ou mais classes terapêuticas incluindo, preferencialmente, dois NRTIs (inibidores nucleosídicos da transcriptase reversa, do inglês *nucleoside reverse transcriptase inhibitors*), que apresentem um efeito sinérgico, juntamente com um fármaco de uma outra classe (Volberding & Deeks, 2010; US. Department of Health and Human Services et al., 2013; Pau & George, 2014). Actualmente, estão aprovados pela FDA (do inglês, *Food and Drug Administration*) 26 fármacos para o tratamento anti-retroviral do HIV-1 (AIDSinfo, 2015), a partir dos quais foram formuladas algumas combinações de dose “fixa”. Estas combinam três fármacos anti-retrovirais de uma ou mais classes terapêuticas, ou dois fármacos de uma mesma classe, num único comprimido, possibilitando a disponibilização de diversos regimes terapêuticos, cuja selecção deverá ser individualizada. Para tal, devem ser tidos em conta diversos factores, nomeadamente, efeitos adversos, interacções farmacológicas, presença de outras patologias, gravidez e nível potencial de adesão à terapêutica. Este último é fundamental na terapia anti-retroviral, pois uma toma descontinuada dos fármacos leva a uma incompleta supressão viral e, por sua vez, ao aparecimento de estirpes resistentes, limitando as opções futuras (Tsibris & Hirsch, 2010; Volberding & Deeks, 2010; Pau & George, 2014).

1.2.1. Regimes terapêuticos para prevenção da transmissão vertical

A prevenção da transmissão vertical (ou transmissão mãe-filho) tem evoluído bastante ao longo das últimas décadas, com a implementação de recomendações universais de diagnóstico e tratamento em mulheres grávidas infectadas pelo HIV-1. A maioria das recomendações terapêuticas para prevenção da transmissão vertical tem dois objectivos principais:

- o tratamento da mãe infectada pelo HIV-1;
- a profilaxia para diminuir o risco de transmissão perinatal de HIV-1 (European Centre for Disease Prevention and Control, 2012; World Health Organization, 2013; Panel on Treatment of HIV-infected pregnant women and prevention of perinatal transmission, 2014).

A terapia anti-retroviral é recomendada para todas as mulheres grávidas, independentemente da carga viral e da contagem de linfócitos T CD4⁺ (Panel on Treatment of HIV-infected pregnant women and prevention of perinatal transmission, 2014). Geralmente, o tratamento recomendado para prevenção da transmissão vertical é igual ao utilizado para as mulheres não grávidas. No entanto, o regime terapêutico seleccionado deve ser individualizado, tendo em conta a carga viral, a realização de terapia anti-retroviral anterior, possíveis efeitos adversos para a mãe e/ou para o feto e interacções medicamentosas (European Centre for Disease Prevention and Control, 2012; Panel on Treatment of HIV-infected pregnant women and prevention of perinatal transmission, 2014; Programa Nacional para a Infecção VIH/SIDA, 2015). Para além de ser realizada profilaxia durante a gravidez, esta também deve ser mantida durante o parto e no período de aleitamento materno (nos casos em que este não possa ser evitado) (World Health Organization, 2013; Panel on Treatment of HIV-infected pregnant women and prevention of perinatal transmission, 2014; Programa Nacional para a Infecção VIH/SIDA, 2015).

Em Portugal, o Programa Nacional para a Infecção VIH/sida publicou, em 2015, as *Recomendações Portuguesas para o tratamento da infecção por VIH-1 e VIH-2* (Programa Nacional para a Infecção VIH/SIDA, 2015). Estas englobam as medidas para a prevenção da transmissão mãe-filho, as quais são descritas em seguida.

1.2.1.1. Profilaxia pré-parto

O tratamento de uma mulher grávida, sem experiência prévia de terapia anti-retroviral, deve ser iniciado entre as 10 e as 12 semanas de gestação (excepto em casos de imunossupressão grave ou carga viral elevada, nos quais se deve iniciar mais cedo). Os regimes terapêuticos recomendados são constituídos por dois NRTIs e um PI (inibidor da protease, do inglês *protease inhibitor*) potenciado (com ritonavir). Para os NRTIs são aconselhadas uma das seguintes coformulações: AZT+3TC (zidovudina + lamivudina), TDF+FTC (tenofovir disoproxil fumarato + emtricitabina) ou ABC+3TC (abacavir + lamivudina). Relativamente aos PIs, são aconselhados: LPV/r (lopinavir), SQV/r (saquinavir) ou ATV/r (atazanavir).

Há, no entanto, casos particulares para os quais estas recomendações terapêuticas poderão ter que ser ajustadas, nomeadamente:

- Grávidas sob terapia anti-retroviral antes da concepção – deve manter-se o mesmo regime terapêutico (se eficaz);
- Grávidas sem terapia actual, mas com histórico de terapia anti-retroviral – deve ter-se em consideração os fármacos utilizados no regime terapêutico anterior e a possível presença de mutações genéticas associadas a resistência;
- Grávidas com diagnóstico tardio ou sem terapia anti-retroviral após as 28 semanas de gestação – deve iniciar-se, o mais rapidamente possível, a terapia anti-retroviral (em casos de carga viral elevada, ponderar a utilização de raltegravir) (Programa Nacional para a Infecção VIH/SIDA, 2015).

1.2.1.2. Profilaxia intraparto

Nas recomendações terapêuticas actuais, a administração (por via endovenosa) de AZT intraparto deixou de ser necessária em grávidas sob terapia anti-retroviral, com carga viral <1000 cópias de RNA/ml e boa adesão à terapêutica. Em todas as outras situações, como por exemplo, a presença de factores de risco (hemorragia por descolamento placentar, entre outros), a administração deve ser mantida. Nos casos em que as mulheres grávidas apresentem uma carga viral detectável (>1000 cópias de RNA/ml) deverá ponderar-se a toma de AZT juntamente com uma dose única de NVP (nevirapina) no início do trabalho de parto (Programa Nacional para a Infecção VIH/SIDA, 2015).

1.2.1.3. Profilaxia pós-parto

Em todos os recém-nascidos é efectuada profilaxia que, dependendo da situação clínica, pode ser realizada em regime de monoterapia (AZT) ou em regime combinado (AZT + 3TC + NVP). Geralmente a profilaxia é realizada durante quatro semanas.

Após o parto, recomenda-se fortemente que o aleitamento materno seja evitado. Assim sendo, a suspensão da terapia anti-retroviral pode ser considerada em mulheres que não tinham indicação para iniciar terapia antes da gravidez, segundo as recomendações no adulto (ou seja, com contagem de linfócitos T CD4⁺ >500/mm³ e sem qualquer sintoma). Esta decisão deve ser tomada durante a gravidez, tendo em conta diversos factores, nomeadamente, o estatuto serológico do parceiro, a adesão à terapia durante a gravidez, entre outros. Ao suspender a terapia anti-retroviral na mãe, a semi-vida dos fármacos deve ser tida em consideração, para evitar monoterapia funcional e, conseqüentemente, o risco de desenvolvimento de resistências (Programa Nacional para a Infecção VIH/SIDA, 2015).

1.2.2. Inibidores da protease

Tendo em conta o papel fundamental que a protease desempenha na maturação do HIV e, conseqüentemente, na formação de partículas virais infecciosas, o desenvolvimento de inibidores específicos da protease foi um sucesso na história da terapia anti-retroviral (Pokorná et al., 2009; Weber & Agniswamy, 2009; Wensing et al., 2010). Os PIs são inibidores competitivos, como já foi referido anteriormente, que se ligam com uma elevada afinidade ao centro activo da enzima (Weber & Agniswamy, 2009; Wensing et al., 2010). Todos são peptidomiméticos, à excepção do tipranavir (TPV), comportando-se como substratos naturais da protease (Wensing et al., 2010; Menéndez-Arias, 2013). No entanto, a sua estrutura química não permite que sejam clivados proteoliticamente e, por conseguinte, permanecem ligados à protease (conformação fechada), impedindo o processamento das poliproteínas virais (Figura 7) (Weber & Agniswamy, 2009; Ali et al., 2010; Wensing et al., 2010).

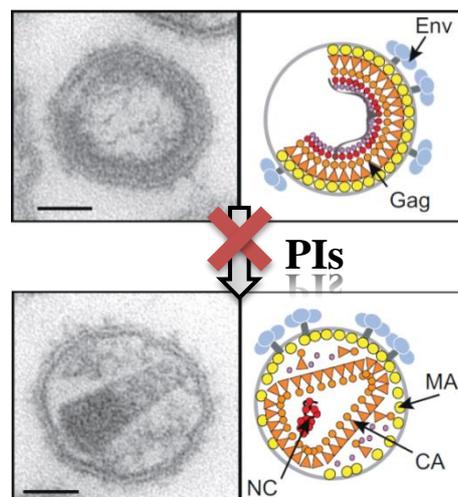


Figura 7. Representação genérica do modo de acção dos inibidores da protease. Demonstra o bloqueio da maturação do HIV-1, não ocorrendo alterações conformacionais na partícula viral, o que impede a aquisição de infecciosidade [Adaptado de (Konvalinka et al., 2015)].

Actualmente, são nove os inibidores da protease aprovados pela FDA para o tratamento anti-retroviral, nomeadamente: saquinavir (SQV), indinavir (IDV), ritonavir (RTV), nelfinavir (NFV), lopinavir (LPV), atazanavir (ATV), fosamprenavir (FPV), tipranavir (TPV) e darunavir (DRV) (AIDSinfo, 2015). Embora estes fármacos sejam bastante eficazes, todos eles apresentam o inconveniente de se ligarem a proteínas plasmáticas, levando a uma diminuição da sua concentração plasmática (Pokorná et al., 2009). Esta limitação pode ser superada pela administração concomitante com um potenciador farmacocinético ou *booster*. O RTV, acima mencionado, é um potente inibidor do citocromo P450 3A4, diminuindo o metabolismo dos restantes PIs e, desta forma, aumentando a sua concentração no plasma e nos tecidos. Assim, o RTV, embora originalmente tenha sido desenvolvido como um PI, nos regimes terapêuticos actuais, é apenas utilizado como potenciador farmacocinético, sendo co-administrado em concentrações sub-terapêuticas com outro PI, de modo a potenciar a sua actividade anti-retroviral (Zeldin & Petruschke, 2004; Pokorná et al., 2009; Pau & George, 2014). Os regimes terapêuticos que incluem um PI potenciado (*boosted PI*) apresentam níveis elevados de supressão viral, pelo que estes fármacos constituem hoje componentes essenciais da HAART. No entanto, estão, por vezes, associados a distúrbios metabólicos, destacando-se a dislipidemia, hiperglicemia, resistência à insulina e lipodistrofia. Além disso, podem causar doenças cardiovasculares, nefrotoxicidade,

hiperbilirrubinemia e aumento das enzimas hepáticas (Wensing et al., 2010; Pau & George, 2014).

A pesquisa de novos inibidores da protease tem sido constante e tem como objectivo potenciar a sua eficácia contra estirpes virais resistentes, bem como, melhorar as propriedades farmacocinéticas e reduzir os efeitos adversos. Nos últimos anos, têm sido estudados alguns compostos que se ligam a regiões fora do centro activo da enzima, tal como os *flaps* (interferindo com a funcionalidade da enzima) ou os terminais N e C de cada monómero (impossibilitando a dimerização da protease) (Pokorná et al., 2009).

1.3. Resistência associada aos anti-retrovirais

Apesar do sucesso da HAART, a emergência de resistências aos anti-retrovirais representa uma profunda limitação para a terapia actual (Cortez & Maldarelli, 2011; Menéndez-Arias, 2013). Diversos factores, quer sejam virais, farmacológicos, ou do hospedeiro, podem contribuir para o desenvolvimento de resistências. O HIV-1 apresenta uma elevada diversidade genética, tal como foi já referido, devido à sua elevada taxa de mutação e recombinação durante o processo de transcrição reversa, potenciada pela elevada taxa replicativa *in vivo*. Assim sendo, a população viral num indivíduo infectado consiste num conjunto complexo de estirpes virais, geneticamente aparentadas, mas heterogéneas (*quasispecies*), a partir do qual podem emergir estirpes resistentes aos anti-retrovirais (Cortez & Maldarelli, 2011; Menéndez-Arias, 2013; Santoro & Perno, 2013). Outros factores, como a fraca adesão à terapêutica por parte dos indivíduos infectados ou interacções medicamentosas que alterem os tempos de semi-vida dos anti-retrovirais, podem dar origem a uma incompleta supressão viral e, conseqüentemente, ao desenvolvimento de vírus resistentes. A selecção de estirpes virais resistentes depende, ainda, de factores como o *fitness* viral da estirpe com a nova mutação ou a chamada “barreira genética” (número mínimo de mutações necessárias para aquisição de resistência fenotípica), característica de cada classe de fármacos anti-retrovirais (Cortez & Maldarelli, 2011; Skar et al., 2011; Götte, 2012; Menéndez-Arias, 2013). Em suma, a presença de mutações que permitam a replicação de uma determinada estirpe viral, sob regime terapêutico, levam a que esta se torne rapidamente

dominante, resultando, conseqüentemente, na falha da terapia anti-retroviral. Há, no entanto, indivíduos que sem nunca terem sido submetidos a terapia anti-retroviral (indivíduos *naive*) apresentam vírus com mutações de resistência. Estes casos, associados à transmissão directa de vírus resistentes, limitam bastante as opções de tratamento e representam um grave problema de saúde pública.

O estudo de mutações associadas a resistência baseia-se, principalmente, no subtipo B, embora este represente actualmente apenas cerca de 10% das infecções de HIV-1 (Takebe et al., 2008; Hemelaar et al., 2011). Este facto é devido à predominância do subtipo B nos países onde a terapia anti-retroviral e os métodos de detecção de mutações de resistência foram primeiramente disponibilizados. Embora tudo indique que a terapia anti-retroviral seja eficaz contra todos os subtipos do grupo M do HIV-1, diversos estudos indicam que a divergência genética entre os subtipos pode resultar em diferentes padrões de resistência aos anti-retrovirais. Este facto pode conduzir a dificuldades na interpretação das mutações associadas a resistência, ou mesmo, a resultados duvidosos, de acordo com os diferentes algoritmos de interpretação utilizados. Assim sendo, é essencial compreender cada vez melhor o modo como a diversidade genética entre os diferentes subtipos pode afectar a resposta aos anti-retrovirais (Lessells et al., 2012; Wainberg & Brenner, 2012; Santoro & Perno, 2013).

1.3.1. Monitorização de mutações de resistência aos anti-retrovirais

Os testes de susceptibilidade aos anti-retrovirais são recomendados nos casos de infecção aguda, “falência virológica”, supressão viral incompleta e antes da implementação da terapêutica anti-retroviral. Estes devem ser também realizados em todas as mulheres grávidas, antes do início da terapêutica ou se a gravidez ocorrer durante o seu decurso, sempre que haja carga viral detectável (Panel on Antiretroviral Guidelines for Adults and Adolescents, 2015). Os testes de susceptibilidade aos anti-retrovirais podem ser divididos em dois grupos: genotípicos e fenotípicos. Os testes genotípicos permitem identificar as mutações responsáveis pela resistência aos anti-retrovirais presentes nos genes virais mais relevantes (genes *pol* e *env*). Estes testes baseiam-se na sequenciação nucleotídica do DNA complementar obtido por RT-PCR (do inglês, *reverse transcription-polymerase chain reaction*) a partir do genoma viral, no qual são identificadas e analisadas as mutações, por comparação com sequências

padrão. Existem diversas ferramentas disponíveis *online* que auxiliam na sua interpretação, nomeadamente, compilações com as mutações associadas a resistência que vão sendo descritas na literatura (disponível, por exemplo, através da *International Antiviral Society USA – IAS-USA*) e algoritmos informáticos, ou manuais, que permitem a elaboração de um relatório com as mutações presentes na sequência nucleotídica e estimam qual o grau de susceptibilidade para cada fármaco, de acordo com a sua classe terapêutica (por exemplo, *Stanford HIVdb Program*, HIV-GRADE, Geno2pheno, *ANRS*, *REGA*, entre outros) (Vercauteren & Vandamme, 2006; Cortez & Maldarelli, 2011; Panel on Antiretroviral Guidelines for Adults and Adolescents, 2015). Os testes fenotípicos avaliam a capacidade replicativa do vírus na presença de diferentes concentrações de fármacos anti-retrovirais. Tal como nos testes genotípicos, é realizada a amplificação das porções relevantes do genoma viral por RT-PCR, sendo os amplicões obtidos clonados num vector plasmídico de HIV-1, originando um clone molecular infeccioso que, no sistema celular adequado (linha de células permissivas), dá origem a partículas virais recombinantes infecciosas. Estas, bem como o vírus selvagem, de referência, são utilizadas para a infecção de culturas celulares expostas a diferentes concentrações do fármaco. De seguida, é calculada a concentração do fármaco necessária para inibir 50% da replicação viral (IC_{50}), em cada um dos casos. O rácio entre as IC_{50} do vírus recombinante e do vírus selvagem permite avaliar o nível de susceptibilidade ao fármaco apresentado pelo primeiro (Cortez & Maldarelli, 2011; Panel on Antiretroviral Guidelines for Adults and Adolescents, 2015).

Os testes genotípicos são os mais utilizados devido ao seu baixo custo e à rapidez de resposta, apresentando, também, uma maior sensibilidade para a detecção de resistência emergente. Por outro lado, os testes fenotípicos são mais vantajosos no caso de novos anti-retrovirais (não é necessário ter um conhecimento prévio das mutações associadas a resistência), assim como na avaliação de padrões de mutação mais complexos, dado que estimam directamente o efeito líquido conferido pela presença de mutações múltiplas (Vercauteren & Vandamme, 2006; Panel on Antiretroviral Guidelines for Adults and Adolescents, 2015). No entanto, ambos apresentam limitações relativamente à detecção de estirpes resistentes minoritárias presentes na *quasispecies* viral e, além disso, fazem uma avaliação individual para cada fármaco, enquanto a terapêutica actual se baseia, na maioria dos casos, na combinação de três ou

mais fármacos (Vercauteren & Vandamme, 2006; Panel on Antiretroviral Guidelines for Adults and Adolescents, 2015).

1.3.2. Mecanismos de resistência aos inibidores da protease

A fixação de certas mutações na região codificante da protease permite que a enzima se torne menos susceptível aos seus inibidores, continuando, no entanto, a reconhecer e a processar os seus substratos naturais (poliproteínas Gag e Gag-Pol) (King et al., 2004; Ali et al., 2010). Existem diversas mutações que conferem resistência aos PIs, as quais são classificadas em mutações principais e acessórias (Figura 8). As mutações principais (ou *major*) são, geralmente, as primeiras a ser seleccionadas na presença do fármaco e conferem um elevado nível de resistência. Porém, levam a uma perda da eficiência catalítica e, por sua vez, a uma diminuição do *fitness* viral. As mutações acessórias (ou *minor*) surgem, normalmente, mais tarde e, por si só, não conduzem a alterações fenotípicas significativas. Contudo, na presença de mutações *major*, melhoram a estabilidade e a actividade enzimática e, por conseguinte, aumentam a capacidade replicativa do vírus (Weber & Agniswamy, 2009; Menéndez-Arias, 2013; Wensing et al., 2015). Algumas destas mutações acessórias estão presentes como alterações polimórficas comuns em subtipos não-B do HIV-1.

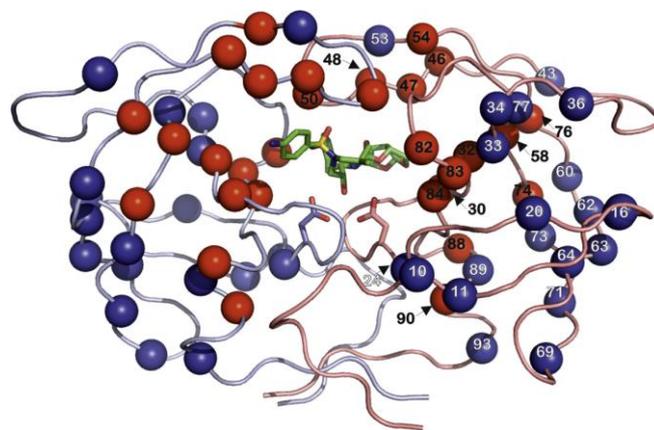


Figura 8. Representação esquemática das posições aminoacídicas na protease associadas a resistência aos inibidores da protease. As mutações principais e acessórias são apresentadas nas esferas vermelhas e azuis, respectivamente [Adaptado de (Konvalinka et al., 2015)].

Consoante o local na região codificante da protease onde as mutações ocorram, afectando, nomeadamente, o centro activo e/ou os *flaps*, diferentes mecanismos podem estar na origem de resistência aos PIs. Destes, destacam-se as mutações que alteram directamente a interacção entre a protease e o inibidor, levando a uma diminuição da eficiência do fármaco. Por outro lado, há mutações que conduzem a uma perda de estabilidade da estrutura molecular da enzima, podendo afectar a interacção com o inibidor (Weber & Agniswamy, 2009; Ali et al., 2010).

As mutações associadas a resistência aos PIs não estão, no entanto, somente delimitadas à região codificante da protease. Mutações nos locais de clivagem da poliproteína Gag, nomeadamente, NC/p1 e p1/p6, estão fortemente associadas a resistência a diversos PIs. Estas mutações induzem uma adaptação do substrato à enzima, melhorando a sua ligação e o seu processamento proteolítico. Inúmeros estudos descrevem que, em associação com mutações específicas na protease, originam um aumento da capacidade replicativa do HIV (Weber & Agniswamy, 2009; Cortez & Maldarelli, 2011; Fun et al., 2012; Konvalinka et al., 2015). Dados preliminares de um estudo recente sugerem que várias mutações que afectam a cauda citoplasmática de gp41 podem igualmente ser responsáveis pela reduzida susceptibilidade aos PIs em alguns indivíduos (Rabi et al., 2013).

1.3.3. Caracterização de mutações associadas à resistência aos inibidores da protease

O desenvolvimento de resistência aos PIs é causado pela acumulação de diversas mutações de resistência que podem ser classificadas como principais ou acessórias, como referido anteriormente. Dependendo da mutação em causa, as mutações principais podem causar resistência a um ou mais PIs. Estas afectam normalmente o centro activo da enzima, alterando a ligação do inibidor à protease (por exemplo, D30N, G48V, I50L/V, V82A/F/T, I84V) (Ali et al., 2010; Menéndez-Arias, 2013). Como se viu, as mutações principais levam a uma diminuição da capacidade catalítica da protease, a qual pode ser restabelecida com a selecção de mutações acessórias. Estas ocorrem em posições aminoacídicas situadas fora do centro activo da protease.

Periodicamente, a IAS-USA publica uma compilação comentada das mutações associadas a resistência aos anti-retrovirais, baseada numa busca sistematizada das

fontes bibliográficas e dos resultados disponíveis de estudos diversos na área. Os dados disponibilizados têm como objectivo ajudar os profissionais de saúde na identificação de mutações-chave associadas à resistência aos anti-retrovirais e, portanto, na tomada de decisões de carácter clínico. As mutações associadas a resistência aos PIs são apresentadas na Figura 9 (Wensing et al., 2015).

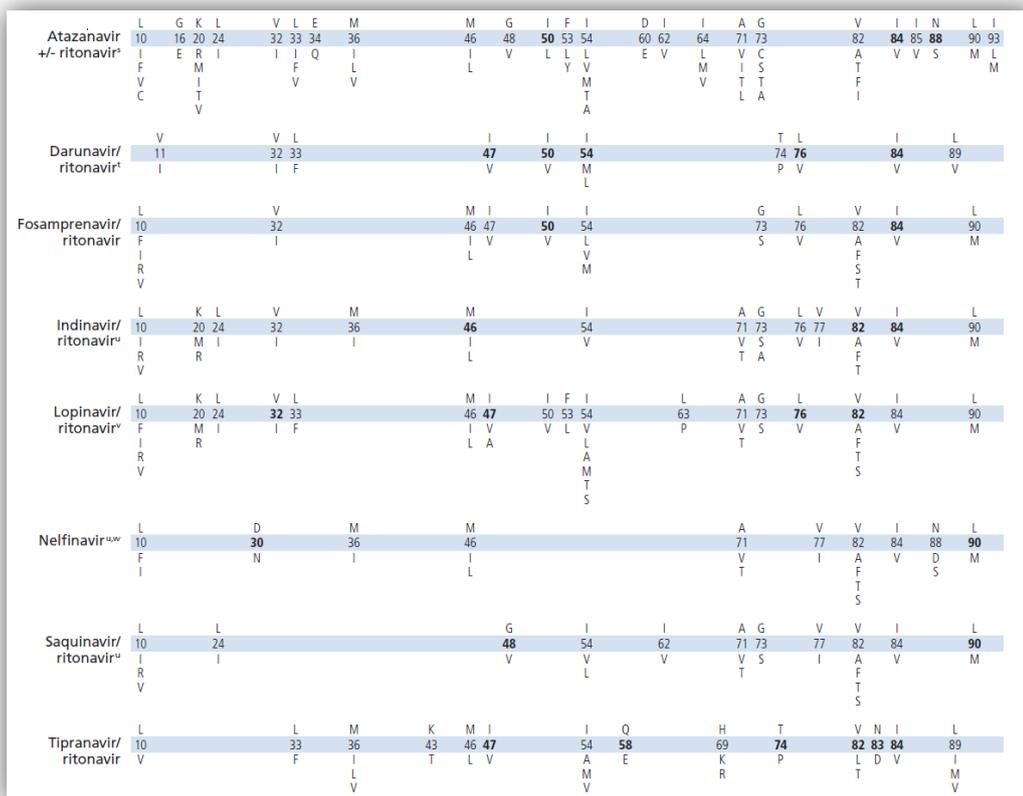


Figura 9. Representação do painel de mutações da região codificante da protease, compilado pela IAS-USA, associadas a resistência aos inibidores da protease licenciados pela FDA. São apresentadas as mutações principais (a negrito) e acessórias, com indicação da posição aminoacídica, tendo como referência o subtipo B de HIV-1. São indicados os aminoácidos presentes no vírus selvagem (letra acima da posição) e os aminoácidos presentes no vírus mutado (letra(s) abaixo da posição) [Adaptado de (Wensing et al., 2015)].

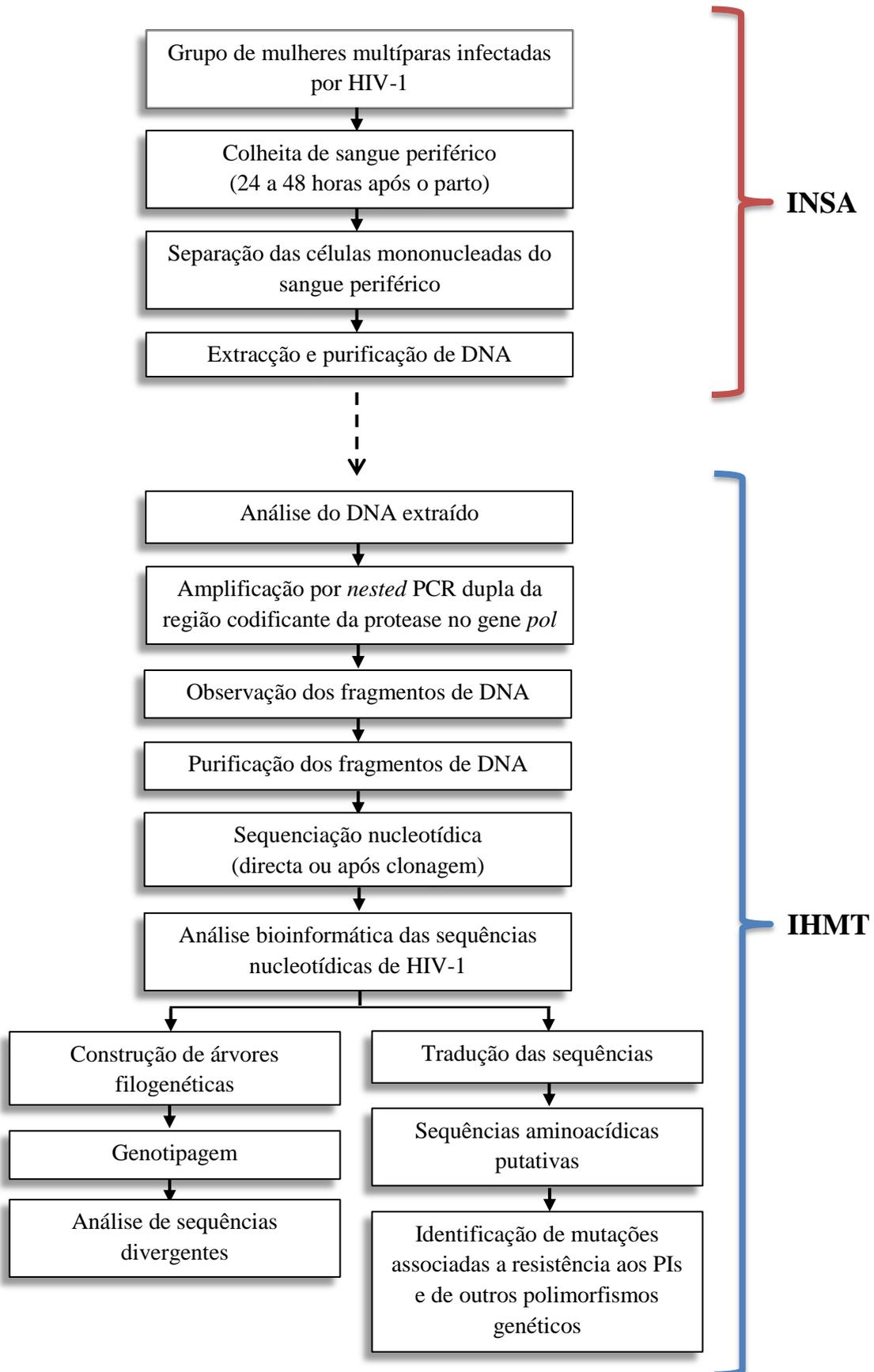
1.4. Objectivos

O presente estudo teve como objectivos principais:

- Caracterizar a diversidade genética da região codificante da protease numa amostra populacional de mulheres múltiparas infectadas pelo HIV-1;
- Identificar mutações de resistência associadas aos inibidores da protease;
- Analisar a evolução de sequências obtidas para cada caso, no contexto da eventual selecção das mutações de resistência, em função do cumprimento de um ou mais regimes terapêuticos durante a gravidez e dos respectivos fármacos;
- Investigar a presença de polimorfismos genéticos na região codificante da protease nas estirpes virais analisadas e estimar a sua prevalência;
- Correlacionar a presença dos polimorfismos genéticos encontrados com os subtipos genéticos descritos.

2. Material e Métodos

2.1. Desenho do estudo



2.2. Caracterização da amostra populacional estudada

O presente estudo incluiu 34 mulheres múltíparas infectadas com o HIV-1, às quais foi efectuada colheita de sangue periférico, por punção venosa, 24 a 48 horas após o nascimento de cada filho. Das mulheres em estudo, 32 foram mães de dois filhos, enquanto duas foram mães de três filhos, em períodos de tempo variáveis, obtendo-se assim um total de 70 amostras de sangue periférico maternas. Para cada amostragem foram colhidos 3 a 5 ml de sangue total (para tubos contendo anticoagulante EDTA) que foram enviados para o Laboratório Nacional de Referência HIV e Hepatites B e C do Instituto Nacional de Saúde Dr. Ricardo Jorge (INSA), entre 1999 e 2008, no âmbito de um protocolo para o diagnóstico precoce da transmissão vertical da infecção pelo HIV.

Após consulta em base de dados do laboratório pela respectiva responsável, foram fornecidos diversos dados relevantes, nomeadamente a idade e a nacionalidade das mulheres, o ano do parto, indicação da realização de terapia anti-retroviral e, caso aplicável, o regime terapêutico efectuado durante a gestação. Assim, as mulheres incluídas tinham uma idade compreendida entre os 17 e os 36 anos (no nascimento do primeiro filho) e a sua nacionalidade era, na maioria, portuguesa (n=22), sendo as restantes angolanas (n=8), moçambicanas (n=3) e de nacionalidade desconhecida (n=1). Relativamente à realização de terapia anti-retroviral, seis mulheres não efectuaram qualquer profilaxia durante a primeira gravidez, sendo que destas, duas também não efectuaram qualquer terapia durante a segunda gestação. A terapia anti-retroviral realizada durante a gravidez foi diferente entre as mulheres e, por vezes, de gestação para gestação. Os regimes terapêuticos administrados às diversas mulheres em estudo incluíram sempre dois NRTIs que serão, em diante, denominados como tratamento base. Este foi administrado isoladamente ou conjugado com um NNRTI (inibidor não nucleosídico da transcriptase reversa, do inglês *non-nucleoside reverse transcriptase inhibitor*) e/ou um PI. Dependendo da natureza e número dos fármacos utilizados, foram, portanto, realizados três regimes terapêuticos distintos, nomeadamente: tratamento base, tratamento base + NNRTI ou PI ou tratamento base + NNRTI + PI.

Por último, foram ainda recolhidos dados referentes à ocorrência de transmissão vertical do HIV-1, tendo-se esta verificado para três dos recém-nascidos.

O estudo aqui descrito, baseado em sobras de material biológico (DNA), sem qualquer outro préstimo, foi retrospectivo, anónimo e não relacionável.

2.3. Identificação e processamento das amostras biológicas

Neste estudo, a identificação das amostras foi efectuada tendo por base um número de código atribuído a cada mulher em estudo e os respectivos partos. Ou seja, após o prefixo PT (de Portugal), a cada amostra foi atribuído um número e a letra **A**, **B**, ou **C**, consoante a amostra de sangue periférico materno fosse correspondente ao nascimento do primeiro, segundo ou terceiro filho, respectivamente. Deste modo, as amostras incluídas neste estudo têm a seguinte identificação: PT_3A, PT_3B, PT_4A, PT_4B, PT_4C, PT_5A, PT_5B, PT_6A, PT_6B, PT_8A, PT_8B, PT_9A, PT_9B, PT_10A, PT_10B, PT_13A, PT_13B, PT_17A, PT_17B, PT_17C, PT_20A, PT_20B, PT_24A, PT_24B, PT_27A, PT_27B, PT_29A, PT_29B, PT_32A, PT_32B, PT_33A, PT_33B, PT_37A, PT_37B, PT_38A, PT_38B, PT_43A, PT_43B, PT_45A, PT_45B, PT_47A, PT_47B, PT_48A, PT_48B, PT_52A, PT_52B, PT_53A, PT_53B, PT_54A, PT_54B, PT_55A, PT_55B, PT_56A, PT_56B, PT_57A, PT_57B, PT_58A, PT_58B, PT_60A, PT_60B, PT_65A, PT_65B, PT_66A, PT_66B, PT_68A, PT_68B, PT_71A, PT_71B, PT_72A e PT_72B.

Após a chegada das amostras de sangue ao INSA foi feita a separação das células mononucleadas do sangue periférico (CMSP) por gradiente de densidade (World Health Organization, 2002), sendo estas mantidas a -80°C até à sua utilização. A partir das CMSP foi realizada a extracção do DNA total utilizando, para isso, o *kit QIAamp[®] DNA Blood* (QIAGEN, EUA) de acordo com o protocolo do fabricante (Espinheira, 2013). Para a elaboração do presente estudo, alíquotas das amostras de DNA purificado foram gentilmente cedidas ao Instituto de Higiene e Medicina Tropical (IHMT), tendo sido conservadas a -20°C até ao seu processamento.

Depois da sua recepção, procedeu-se à análise das amostras de DNA purificado, após migração por electroforese em gel de agarose (SeaKem[®] LE, Lonza, EUA), tendo como finalidade avaliar a sua qualidade, assim como, estimar a sua concentração aproximada. A electroforese decorreu em gel de agarose a 1% (p/v), em tampão TAE

(Tris-Acetato-EDTA) 0,5x (Anexo 1), com brometo de etídeo (0,5 µg/ml), sendo aplicado 1 µl de cada amostra juntamente com tampão de aplicação (Anexo 1).

2.4. Amplificação por *nested* PCR dupla da região codificante da protease no gene *pol*

2.4.1. Selecção dos *primers*

Para proceder à amplificação da região codificante da protease (PR), a partir do DNA proviral, foram utilizados *primers* adaptados (Videira e Castro, 2010) a partir do protocolo implementado por Steegen *et al.* (2006). Com o intuito de melhorar a eficiência da reacção de amplificação (sensibilidade e especificidade) foram efectuadas diversas experiências prévias de PCR, conjugando várias combinações entre os *primers* externos (Gag2, PR2, RT137 e RT3303) aplicados no protocolo referido, utilizando como matriz o DNA proviral extraído da linha celular 8E5/LAV (HIV-1) (Folks *et al.*, 1986; Gendelman *et al.*, 1987). De acordo com os resultados obtidos nas diversas reacções de PCR (dados não mostrados) estabeleceu-se que os *primers* a utilizar neste trabalho seriam, numa primeira reacção, Gag2 e RT137, e, numa segunda reacção, PR2 e RT137. A última reacção de PCR seria realizada com os *primers* internos PR3 e PR4, resultando na implementação de um protocolo de *nested* PCR dupla, diferente do original de Steegen *et al.* (2006). Na Tabela 5 estão descritas as sequências nucleotídicas dos *primers* e as suas respectivas posições, segundo a estirpe de referência HXB2 do subtipo B de HIV-1. Como pode ser observado, os *primers* PR2 e RT137 contêm posições degeneradas (Y= C ou T; R= G ou A). Os *primers* utilizados neste estudo foram sintetizados pela empresa STAB VIDA (Portugal), sendo utilizados na concentração de trabalho de 10 pmol/µl, após hidratação com água desionizada ultrapura, livre de nucleases, *Milli-Q*[®] (*Reagent Grade Ultrapure Systems*, Merck Millipore, EUA).

Tabela 5. *Primers* utilizados nas reacções de PCR [adaptados de (Steegeen et al., 2006)].

<i>Primers</i>	Posição nucleotídica em HXB2	Sequência 5' – 3'
Gag2	1404 - 1426	CAATGAGGAAGCTGCAGAATGGG
PR2	1822 - 1844	GAAATGATGACAGCATGYCAGGG
RT137	3301 - 3325	TTCTGTATRTCATTGACAGTCCAGC
PR3	2146 - 2166	CCAGAGCCAACAGCCCCACCA
PR4	2584 - 2605	GGGCCATCCATTCCTGGCTTTA

2.4.2. Protocolo de amplificação

Para a amplificação da região codificante da protease foram realizadas três PCRs sucessivas e, em cada uma delas, foram utilizadas diferentes combinações de *primers* e de condições de amplificação. Para definição destas últimas, foram considerados a temperatura de hibridação estimada dos *primers*, o tamanho dos fragmentos de DNA a amplificar e, naturalmente, as características intrínsecas da Taq DNA polimerase a utilizar. Na primeira PCR foi utilizado um par de *primers* externos (Gag2 e RT137), para amplificar fragmentos de DNA com um tamanho esperado de 1922 pb. Estes foram usados como matriz para uma segunda PCR (*hemi-nested*), na qual foi utilizado o par de *primers* PR2 e RT137, resultando na amplificação de fragmentos de 1504 pb. Por sua vez, estes serviram de matriz para uma terceira PCR (*nested*), através da qual foram amplificados fragmentos de 460 pb, com utilização dos *primers* internos PR3 e PR4 (Tabela 5 e Figura 10).

Considerados os objectivos do estudo, nas duas primeiras reacções de PCR foi utilizado o sistema *FastStart High Fidelity PCR* (Roche, Alemanha), principalmente devido às especificações da sua Taq DNA polimerase. Por um lado, esta é acompanhada por uma proteína com capacidade de *proofreading* e, por outro lado, só é activada após submissão a uma temperatura de 95°C durante 2 minutos (*hot start PCR*). Dado isto, evita-se a amplificação de produtos inespecíficos e a formação de dímeros de *primers*, com aumento da especificidade e da sensibilidade da reacção. A utilização deste sistema na primeira e segunda reacções de PCR prende-se ainda com o tamanho dos fragmentos amplificados (1922 pb e 1504 pb, respectivamente), que é significativamente maior do que na terceira reacção (460 pb), evitando-se a possível introdução cumulativa de nucleótidos incorrectos e consequente fixação de mutações de modo artificial.

As misturas de reacção foram preparadas em câmara de fluxo laminar (*Laminar Flow Cabinet-class 100*, Flow Laboratories, EUA), sendo utilizadas pontas de micropipeta com filtro (*MultiGuard™*, Sorenson BioScience, EUA). Além disso, foram efectuados controlos negativos (substituição do volume de DNA por água) em todas as reacções de amplificação, para confirmar a inexistência de contaminação. A adição de DNA das amostras a cada microtubo foi realizada em compartimento laboratorial distinto do local de preparação das misturas de reacção.

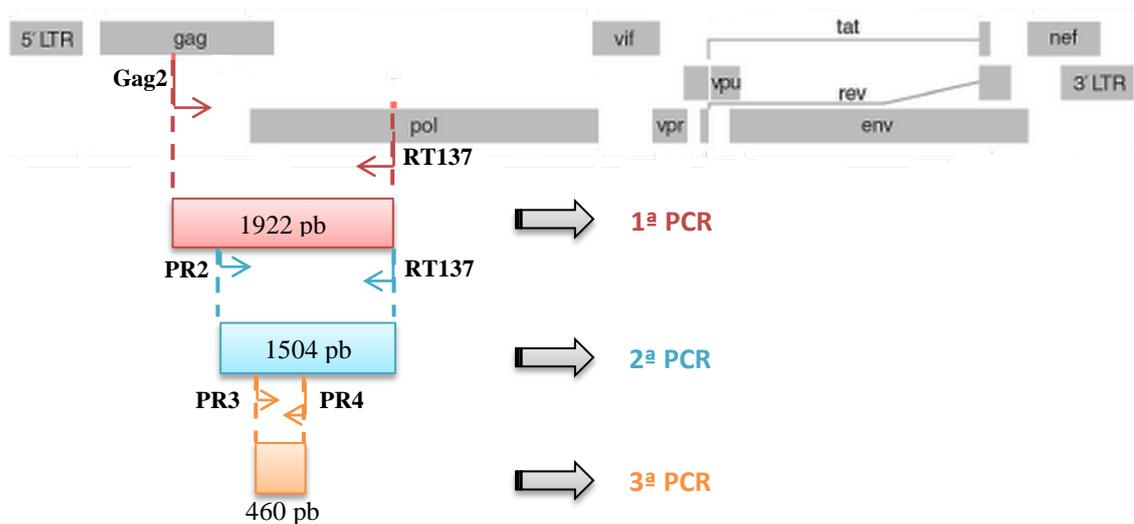


Figura 10. Representação esquemática do protocolo de amplificação por *nested* PCR dupla utilizado na amplificação da região codificante da protease do HIV-1 (fragmento de 460 pb) [figura adaptada de Los Alamos HIV sequence database (<http://www.hiv.lanl.gov>)].

2.4.2.1. Primeira e segunda reacções de PCR

Nas duas primeiras reacções de PCR foi utilizado o *kit FastStart High Fidelity PCR System, dNTPack* (Roche, Alemanha), como referido anteriormente. A mistura de reacção da primeira PCR foi preparada de acordo com as normas do fabricante e, por cada amostra, foram adicionados, por ordem: um volume de água *PCR Grade* (Roche, Alemanha) de modo a perfazer um volume final de reacção de 25 μ l; 2,5 μ l de tampão de reacção *FastStart High Fidelity* 10x com $MgCl_2$ (concentração final de 1,8 mM); 1 μ l de DMSO [concentração final de 4% (v/v)]; 0,5 μ l de uma mistura de dNTPs [*PCR Grade Nucleotide Mix* (Roche, Alemanha), para uma concentração final de 200 μ M cada]; 1 μ l de cada *primer* (concentração final de 0,4 μ M); 0,5/1 μ l de solução de DNA matriz (dependendo da sua qualidade e/ou concentração, anteriormente analisada, ver

2.3.); e, por fim, 0,25 µl da mistura de enzima Taq DNA polimerase modificada *FastStart High Fidelity Enzyme Blend* (concentração final de 2,5 U/reacção). Na segunda reacção de PCR, a mistura de reacção foi preparada exactamente do mesmo modo, mas sendo utilizado como matriz 1 µl do produto amplificado na reacção anterior.

As reacções de amplificação foram efectuadas no termociclador *PCR Express* (Thermo Hybaid, Reino Unido) e as condições de amplificação base da primeira e segunda PCR estão descritas nas Tabelas 6 e 7, respectivamente. De salientar que nas amostras inicialmente não amplificáveis com recurso a estas condições de amplificação, foram utilizados protocolos alternativos, com temperaturas de hibridação mais baixas (em diversas combinações).

Tabela 6. Condições de amplificação aplicadas na primeira PCR.

Etapas	Condições de amplificação			Número de ciclos
Desnaturação inicial	95°C – 2 min.			1
Amplificação	Desnaturação	Hibridação	Extensão	
	95°C – 30 seg.	45°C – 30 seg.	72°C – 2 min.	10
	95°C – 30 seg.	50°C – 30 seg.	72°C – 2,5 min.	10
	95°C – 30 seg.	55°C – 30 seg.	72°C – 3 min.	15
Extensão final	72°C – 7 min.			1
Manutenção	4°C - ∞			

Tabela 7. Condições de amplificação aplicadas na segunda PCR.

Etapas	Condições de amplificação			Número de ciclos
Desnaturação inicial	95°C – 2 min.			1
Amplificação	Desnaturação	Hibridação	Extensão	
	95°C – 30 seg.	45°C – 30 seg.	72°C – 1,5 min.	10
	95°C – 30 seg.	50°C – 30 seg.	72°C – 2 min.	10
	95°C – 30 seg.	55°C – 30 seg.	72°C – 2,5 min.	15
Extensão final	72°C – 7 min.			1
Manutenção	4°C - ∞			

2.4.2.2. Terceira reacção de PCR

Na última reacção de PCR foi aplicado o sistema *NZYTaq 2x Green Master Mix* (nzytech, Portugal), cuja solução já vem previamente preparada, contendo tampão de reacção, dNTPs e uma enzima Taq DNA polimerase recombinante modificada. A

mistura de reacção foi preparada de acordo com as especificações do fabricante, contendo, para cada amostra e um volume final de 25 μ l: 12,5 μ l de *NZYTaq 2x Green Master Mix* (concentração final: 50 mM de Tris-HCl, pH 9,0; 50 mM de NaCl; 2,5 mM de MgCl₂; 200 μ M de cada dNTP; 0,1 U/ μ l de Taq DNA polimerase modificada); 1 μ l de cada *primer* (concentração final de 0,4 μ M); 1 μ l do produto de amplificação da segunda PCR; e 9,5 μ l de água *PCR Grade* (Roche, Alemanha). Nesta reacção de amplificação foram utilizados os *primers* internos PR3 e PR4, como mencionado anteriormente, permitindo a amplificação de um fragmento de DNA de menores dimensões que contém a região codificante da protease.

Tal como nas reacções de amplificação anteriores, estas foram efectuadas no termociclador *PCR Express* (Thermo Hybaid, Reino Unido) e as condições de amplificação estão indicadas na Tabela 8.

Tabela 8. Condições de amplificação aplicadas na terceira PCR.

Etapas	Condições de amplificação			Número de ciclos
Desnaturação inicial	95°C – 2 min.			1
Amplificação	Desnaturação	Hibridação	Extensão	35
	95°C – 30 seg.	50°C – 30 seg.	72°C – 1 min.	
Extensão final	72°C – 7 min.			1
Manutenção	4°C - ∞			

2.5. Observação dos fragmentos de DNA em gel de agarose

Os fragmentos de DNA resultantes da *nested* PCR dupla foram observados após migração por electroforese em gel de agarose (SeaKem[®] LE, Lonza, EUA) a 1,2% (p/v), em tampão TAE 0,5x (Anexo 1), com brometo de etídeo (0,5 μ g/ml). Foram aplicados em gel 5 μ l de cada produto de amplificação e aplicada uma voltagem de 90 V. A estimativa do tamanho aproximado dos fragmentos de DNA amplificados foi realizada por comparação do respectivo padrão de migração com o do marcador de massa molecular *GeneRuler[™] 100 bp Plus DNA Ladder* (Fermentas, Lituânia). A visualização dos fragmentos de DNA foi efectuada com recurso ao equipamento *Gel-Doc XR* (Bio-Rad, EUA), que capta uma imagem do gel de agarose, resultante da sua exposição a radiação ultravioleta ($\lambda=254$ nm).

2.6. Purificação dos fragmentos de DNA

Após a observação dos fragmentos de DNA, os que apresentaram um tamanho estimado de 460 pb, foram seleccionados para posterior purificação. Contudo, o sistema utilizado foi distinto consoante os resultados obtidos. Deste modo, foram utilizados dois sistemas comerciais: o *DNA Clean & Concentrator*[™]-5 (Zymo Research, EUA) e o *Zymoclean*[™] *Gel DNA Recovery Kit* (Zymo Research, EUA).

O sistema *DNA Clean & Concentrator*[™]-5 (Zymo Research, EUA) foi aplicado em todos os produtos de PCR que apresentaram, após separação por electroforese em gel de agarose, um fragmento com um tamanho específico de 460 pb, com uma quantidade de DNA aceitável. Portanto, este sistema permitiu purificar os fragmentos de DNA directamente a partir dos produtos de PCR, tendo isto sido realizado de acordo com as normas do fabricante. Deste modo, foram adicionados 17 µl de cada produto de PCR a 85 µl de *DNA Binding Buffer* e, de seguida, após homogeneização, transferidos para uma coluna *Zymo-Spin*[™] previamente inserida num microtubo. Seguiu-se uma etapa de centrifugação a 16000 x g (Centrifuge 5415D, Eppendorf, Alemanha) durante 30 segundos, após a qual se descartou o eluato presente no microtubo. Posteriormente, foi efectuado um passo de lavagem, adicionando-se 200 µl de *DNA Wash Buffer* à coluna, seguido de uma centrifugação a 16000 x g durante 30 segundos (este passo foi realizado duas vezes). Nesta fase, a coluna foi transferida para um novo microtubo e, logo após, foram adicionados 16 µl de *DNA Elution Buffer* com um período de incubação, à temperatura ambiente, de 1 minuto. Por fim, foi feita a eluição por centrifugação a 16000 x g durante 30 segundos, obtendo-se assim o DNA purificado.

O sistema *Zymoclean*[™] *Gel DNA Recovery Kit* (Zymo Research, EUA) foi utilizado nos produtos de PCR que apresentaram diversos fragmentos de DNA resultantes de amplificação inespecífica, mas que continham o fragmento de DNA com o tamanho esperado (460 pb). Neste sistema, a purificação dos fragmentos de DNA foi efectuada a partir de gel de agarose a 1,2% (p/v), em tampão TAE 0,5x (Anexo 1), após migração electroforética, de acordo com as especificações do fabricante. O fragmento de DNA pretendido foi removido do gel de agarose, sob luz ultravioleta ($\lambda=254$ nm), com recurso a um escalpelo descartável, e colocado no interior de um microtubo. De seguida, adicionou-se *ADB Buffer* (num volume três vezes superior ao do pedaço de gel contendo o fragmento de DNA, cujo volume tinha sido anteriormente estimado por

pesagem), seguindo-se uma incubação a 55°C durante 5-10 minutos, para dissolução da agarose. Após liquefacção completa da agarose e homogeneização, foi realizada a sua transferência para uma coluna *Zymo-Spin*[™], previamente adaptada a um microtubo, prosseguindo o protocolo exactamente como descrito no parágrafo anterior.

Para confirmar se a purificação dos fragmentos de DNA, pelos sistemas acima descritos, foi bem sucedida procedeu-se à aplicação de 1 µl de cada produto de purificação, juntamente com tampão de aplicação (Anexo 1), em gel de agarose (SeaKem[®] LE, Lonza, EUA) a 1,2% (p/v) em tampão TAE 0,5x (Anexo 1), com brometo de etídeo (0,5 µg/ml).

2.7. Clonagem dos fragmentos de DNA

O método de clonagem molecular foi aplicado nos produtos de amplificação que, após sequenciação nucleotídica, apresentavam um cromatograma de difícil interpretação, incluindo a atribuição de bases degeneradas, com uma proporção igual ou superior a 1%, não permitindo a obtenção de uma sequência com qualidade desejável. Com base nestes critérios foram seleccionados os produtos de PCR e aplicado o método de clonagem recorrendo-se, para isso, ao sistema *pGEM*[®]-*T Easy Vector* (Promega, EUA). Este vector é comercializado sob a forma de uma molécula linearizada que contém resíduos de 3'-timidina em ambas as extremidades. Estes facilitam a ligação dos produtos de PCR (que contém resíduos extra de adenosina, adicionados pela Taq DNA polimerase utilizada) e, além disso, evitam a recircularização do vector. Este vector contém o gene *bla*, que codifica a enzima β-lactamase, a qual degrada a ampicilina possibilitando, assim, a selecção dos clones bacterianos transformantes (resistentes à ampicilina). O vector contém ainda os promotores de transcrição T7 e SP6, com orientações opostas, flanqueando o sítio de clonagem múltipla. Este foi inserido no gene que codifica o fragmento α da enzima β-galactosidase (*lacZ*), sendo a sua sequência interrompida quando ocorre a inserção de um fragmento de DNA exógeno, o que permite a selecção, em meio de cultura e na estirpe bacteriana apropriados, de clones bacterianos recombinantes (ausência de α-complementação). A consequência disto é visível em placas com meio de crescimento contendo X-Gal (5-bromo-4-cloro-3-indolil-β-D-galactopiranosídeo) e IPTG (isopropil-β-D-1-tiogalactopiranosídeo). O primeiro é

uma substância cromogénica, degradada por uma β -galactosidase funcional, ou seja, requerendo um α -péptido igualmente activo, pelo que clones não recombinantes são identificados pelo aparecimento de colónias de cor azul, na presença do IPTG. Ao invés, a ligação de fragmentos exógenos de DNA no sítio de clonagem múltipla origina, habitualmente, a inactivação do gene *lacZ*, impedindo que as bactérias produzam uma β -galactosidase funcional, não ocorrendo, portanto, a degradação do X-Gal, na presença do IPTG, o que dá origem a colónias de cor branca. Por último, este vector contém ainda locais de restrição duplos para três enzimas de restrição (BstZI, EcoRI e NotI) no sítio de clonagem múltipla, permitindo a utilização de qualquer uma das três para realizar a excisão do fragmento de DNA clonado.

2.7.1. Ligação dos fragmentos de DNA ao vector *pGEM[®]-T Easy*

Inicialmente, procedeu-se à ligação dos fragmentos de DNA ao vector *pGEM[®]-T Easy* (Promega, EUA). Para isso, foi preparada uma mistura de reacção constituída pelos seguintes componentes: 5 μ l de tampão de ligação 2x (Anexo 1); 0,75 μ l de solução de vector *pGEM[®]-T Easy* (Promega, EUA) (50 ng/ μ l); 2 μ l do produto de PCR; 1 μ l da enzima ligase de DNA do fago T4 (3 U/ μ l) e 1,25 μ l de água livre de nucleases *Milli-Q[®]* (Merck Millipore, EUA), perfazendo um volume total de 10 μ l. A mistura de ligação foi incubada a 4°C, em banho refrigerado (*Frigiterm*, P-Selecta, Espanha), durante a noite, com o intuito de aumentar a eficiência de ligação.

2.7.2. Preparação de células competentes de *Escherichia coli* e sua transformação por choque térmico

Após a ligação dos fragmentos de DNA ao vector plasmídico, procedeu-se à preparação de células bacterianas competentes para o passo de transformação. Para isso, foi utilizada a estirpe JM109 de *Escherichia coli*, mantida a -80°C em meio de criopreservação (Anexo 1). Esta estirpe, superprodutora do repressor Lac e susceptível à ampicilina, foi inoculada em 2 ml de meio LB (Anexo 1) e incubada a 37°C a 220 rpm, numa incubadora orbital (Infors HT, Suíça), durante a noite. Adicionalmente, foram efectuados dois controlos negativos para confirmação da esterilidade do meio de cultura e da pureza da estirpe bacteriana, mantidos sob as mesmas condições de incubação: um constituído apenas por meio LB e o outro consistindo numa inoculação da estirpe *E. coli*

JM109 em meio LB, mas ao qual fora adicionada ampicilina (concentração final de 100 µg/ml). Após o período de incubação, procedeu-se à transferência de 200 µl da suspensão celular saturada para 20 ml de meio LB (pré-aquecido a 37°C), presentes num nefalómetro. O crescimento bacteriano foi controlado por medição de densidade óptica (600 nm), até à obtenção de um valor compreendido entre 0,3 a 0,5, mantendo as condições de incubação anteriores. Depois, a cultura foi imediatamente centrifugada a 4000 rpm (Centrifuge 5810R, Eppendorf, Alemanha), durante 30 minutos, a 4°C, para sedimentação das células. De seguida, o sobrenadante foi rejeitado e o sedimento celular foi ressuscitado em 1 ml de meio TSS (do inglês, *transformation and storage solution*) (Anexo 1) refrigerado.

Após a adição do meio TSS, inicia-se então a fase de transformação, propriamente dita, das células bacterianas com o DNA exógeno. Assim sendo, adicionou-se 5 µl da mistura de ligação (descrita anteriormente, secção 2.7.1.) a 120 µl da suspensão celular em meio TSS. Foram ainda realizados dois controlos (positivo e negativo) processados exactamente do mesmo modo. No entanto, a mistura de ligação foi substituída por 1 µl de uma solução concentrada de DNA plasmídico purificado (*pBluescript II KS (+)*, Stratagene, EUA), no controlo positivo, e 5 µl de água livre de nucleases *Milli-Q*[®] (Merck Millipore, EUA), no controlo negativo. Após homogeneização suave das células, estas foram mantidas em gelo durante 60 minutos. Logo de seguida, foram sujeitas a um choque térmico, pela colocação num banho a 42°C durante 90 segundos e, sem demora, colocadas novamente em gelo, durante 5 minutos. Posteriormente, adicionou-se 900 µl de meio LB (pré-aquecido a 37°C) aos microtubos, seguindo-se uma incubação a 37°C, a 80 rpm, durante 1 hora. Findo o tempo de incubação, os microtubos foram centrifugados a 6000 rpm (Centrifuge 5415D, Eppendorf, Alemanha), durante 10 minutos, rejeitando-se, de seguida, 800 µl do sobrenadante. O sedimento foi ressuscitado no volume restante do mesmo meio (~200 µl) e distribuído por espalhamento, em condições de assepsia, em duas placas de Petri com meio LB sólido selectivo (15-20 ml/placa), contendo ampicilina (a 100 µg/ml), X-Gal (a 40 µg/ml) e IPTG (0,2 mM), as quais foram incubadas a 37°C, durante a noite.

2.7.3. Selecção de clones bacterianos e extracção do DNA plasmídico pelo método de lise alcalina

Esta fase teve início com a selecção de clones bacterianos transformantes (resistentes à ampicilina) e recombinantes putativos (colónias brancas na presença de X-Gal e IPTG), após o período de incubação, em meio LB sólido (ver secção anterior). As colónias brancas foram repicadas (em condições de assepsia) para tubos contendo 3 ml de meio LB líquido com ampicilina (concentração final de 100 µg/ml), seguindo-se uma incubação a 37°C, a 220 rpm (incubadora orbital Infors HT, Suíça), durante a noite.

De seguida, deu-se início à extracção do DNA plasmídico, propriamente dita, através do método de lise alcalina (Birnboim & Doly, 1979). Um volume de 2 ml da cultura bacteriana densa foi transferido para um microtubo e sujeito a uma centrifugação a 13500 rpm (Centrifuge 5415D, Eppendorf, Alemanha), durante 2 minutos. O sobrenadante foi rejeitado e o sedimento celular ressuspensionado em 100 µl de tampão TEG (Anexo 1) refrigerado. De seguida, foi adicionado o mesmo volume de uma solução de SDS alcalino (Anexo 1) e efectuadas inversões suaves dos microtubos, com o intuito de promover a lise celular. Posteriormente, adicionou-se 200 µl de acetato de potássio 3M, pH 4,8, frio, misturando novamente as soluções por inversão suave. Nesta fase ocorreu a separação entre o DNA plasmídico e o DNA cromossómico e os restos celulares. Os restos celulares foram removidos por centrifugação a 13500 rpm, durante 15 minutos, a 4°C (Centrifuge 3K18, Sigma, Alemanha). O sobrenadante, contendo o DNA plasmídico, foi cuidadosamente transferido para novos microtubos, aos quais foi adicionado 600 µl de isopropanol, com homogeneização por inversões sucessivas, seguindo-se uma nova centrifugação a 13500 rpm, durante 15 minutos, à temperatura ambiente (Centrifuge 5415D, Eppendorf, Alemanha), de modo a precipitar o DNA. O sobrenadante foi rejeitado e o sedimento lavado com 250 µl de etanol a 70% (v/v). Seguiu-se uma nova etapa de centrifugação, nas mesmas condições, mas durante 5 minutos, removendo-se o sobrenadante com auxílio de uma micropipeta. O sedimento de DNA plasmídico foi então seco sob vácuo (Concentrator 5301, Eppendorf, Alemanha), durante um mínimo de 5 minutos. Por último, foi efectuada a ressuspensão do sedimento em 30 µl de água livre de nucleases *Milli-Q*[®] (Merck Millipore, EUA) suplementada com RNaseA (a 50 µg/ml) seguida de uma incubação em banho-maria a 37°C durante 30 minutos. O DNA plasmídico foi mantido a -20°C.

2.7.4. Observação das moléculas de DNA plasmídico em gel de agarose

Para avaliação do sucesso das etapas anteriormente descritas, foram aplicados 2 µl da suspensão contendo o DNA plasmídico em gel de agarose (SeaKem[®] LE, Lonza, EUA) a 1% (p/v), em tampão TAE 0,5x (Anexo 1), corado com brometo de etídeo (0,5 µg/ml). Após a migração electroforética a 100 V, durante cerca de 1 hora, foi efectuada a observação dos padrões de migração das moléculas de DNA plasmídico. Destas, as que apresentavam uma migração retardada (sugerindo presença de DNA exógeno) foram seleccionadas para posterior hidrólise enzimática com a endonuclease de restrição EcoRI.

2.7.5. Hidrólise enzimática do DNA plasmídico

Este procedimento teve como objectivo confirmar a clonagem do fragmento de DNA pretendido no vector *pGEM[®]-T Easy* (Promega, EUA). Tal como referido anteriormente, este vector apresenta locais de restrição duplos para três enzimas de restrição (BstZI, EcoRI e NotI) no sítio de clonagem múltipla, flanqueando o local de inserção do DNA exógeno. Assim, a utilização de qualquer uma das três enzimas permite a hidrólise dos plasmídeos recombinantes, com excisão do fragmento de DNA clonado, comprovando o sucesso da experiência de clonagem. Para a hidrólise dos DNAs plasmídicos apresentando padrões de migração retardada (ver secção anterior) foram preparadas misturas de reacção compostas por: 1,5 µl de tampão EcoRI 10x (Anexo 1), 1 µl de enzima EcoRI (Fermentas, Lituânia) (10 U/µl), 4 µl de solução de DNA plasmídico e 8,5 µl de água livre de nucleases *Milli-Q[®]* (Merck Millipore, EUA), perfazendo um volume total de 15 µl. De seguida, procedeu-se à incubação dos microtubos a 37°C durante um mínimo de 2 horas. Finalmente, procedeu-se à análise dos produtos de digestão, em 10 µl da mistura de reacção, após separação por electroforese em gel de agarose (SeaKem[®] LE, Lonza, EUA) a 1,2% (p/v) em tampão TAE 0,5x (Anexo 1), corado com brometo de etídeo (0,5 µg/ml).

Caso no gel de agarose fossem observadas as bandas esperadas, ou seja, do fragmento de DNA clonado (460 pb) juntamente com o vector linearizado (cerca de 3000 pb), por comparação com o padrão de migração do marcador de massa molecular *GeneRuler[™] 100 bp Plus DNA Ladder* (Fermentas, Lituânia), prosseguia-se para nova extracção e purificação do DNA plasmídico dos clones bacterianos seleccionados, mas

com recurso a um *kit* comercial (ver secção seguinte), de modo a preparar o DNA para a sequenciação nucleotídica.

2.7.6. Extracção e purificação do DNA plasmídico para sequenciação nucleotídica

Para a extracção e purificação do DNA plasmídico para sequenciação nucleotídica, foi preparada uma cultura celular em 3,5 ml de meio LB líquido com ampicilina (a 100 µg/ml), a partir das culturas obtidas anteriormente por repicagem, após incubação a 37°C, a 220 rpm (incubadora orbital Infors HT, Suíça), durante a noite. No dia seguinte, o procedimento foi realizado recorrendo ao sistema *Wizard® Plus SV Minipreps DNA Purification System* (Promega, EUA), de acordo com as directivas do fabricante. Deste modo, procedeu-se à centrifugação de 3,5 ml da cultura bacteriana saturada a 16000 x g (Centrifuge 5415D, Eppendorf, Alemanha), durante 3 minutos, para sedimentação das células. O meio de cultura foi descartado, o sedimento celular ressuscitado em 250 µl da solução de ressuspensão celular e, de seguida, foi adicionado o mesmo volume de solução de lise. Os microtubos foram homogeneizados por inversão suave, de modo a facilitar a lise celular, e incubados à temperatura ambiente (entre 1 a 5 minutos), até a suspensão ficar translúcida. Logo após, foram adicionados 10 µl de uma solução de protease alcalina, seguindo-se uma homogeneização suave por inversão e uma incubação, durante 5 minutos, à temperatura ambiente. Para neutralizar o elevado pH do meio, foram adicionados 350 µl da solução de neutralização, com mistura por inversão, seguindo-se um passo de centrifugação a 16000 x g, durante 10 minutos. O lisado clarificado obtido (cerca de 860 µl), contendo o DNA plasmídico, foi transferido para uma coluna (*Spin Column*), adaptada a um tubo colector, sendo o conjunto centrifugado a 16000 x g, durante 1 minuto, e o eluato rejeitado. Procedeu-se então a dois passos consecutivos de lavagem. Um primeiro, no qual foram adicionados 750 µl de solução de lavagem e aplicada uma centrifugação a 16000 x g, durante 1 minuto, seguido de uma nova adição de solução de lavagem (250 µl) e um passo de centrifugação, à mesma velocidade, mas durante 2 minutos. Por último, à coluna, entretanto transferida para um novo microtubo, foram adicionados 100 µl de água livre de nucleases *Milli-Q®* (Merck Millipore, EUA), seguindo-se um passo

de centrifugação a 16000 x g, durante 1 minuto, para eluição final do DNA plasmídico. Este foi conservado a -20°C até utilização.

Para confirmar se a purificação foi bem sucedida, foram aplicados 5 µl de cada solução de DNA plasmídico em gel de agarose (SeaKem[®] LE, Lonza, EUA) a 1% (p/v) em tampão TAE 0,5x (Anexo 1), corado com brometo de etídeo (0,5 µg/ml), para migração electroforética e posterior visualização sob luz ultravioleta.

2.8. Sequenciação nucleotídica de DNA

A sequenciação nucleotídica foi realizada pela empresa STAB VIDA (Portugal), com recurso ao método de terminação de cadeia, originalmente descrito por Sanger et al. (1977), implementado num sistema automatizado. A sequenciação dos fragmentos de DNA obtidos por PCR foi efectuada em ambos os sentidos, sendo enviados 15 µl de cada amostra de DNA, juntamente com os *primers* PR3 e PR4 (50 pmoles/reacção), com uma quantidade mínima estimada de 400 ng de DNA por reacção de sequenciação. Por sua vez, para os fragmentos clonados procedeu-se ao envio de 25 µl de solução de DNA plasmídico purificado (para um mínimo de 1 µg de DNA por reacção de sequenciação), sendo utilizados os *primers* universais T7 e SP6 para a obtenção de sequências no sentido directo e reverso, respectivamente. Estes *primers* hibridam com sequências do vector *pGEM[®]-T Easy* (Promega, EUA) que flanqueiam o sítio de clonagem múltipla, permitindo a sequenciação do fragmento de DNA clonado.

2.9. Análise bioinformática das sequências nucleotídicas

2.9.1. Edição de sequências e confirmação da sua identidade

Tal como referido no subcapítulo anterior, a sequenciação nucleotídica foi realizada no sentido directo e reverso, obtendo-se assim duas sequências para cada molécula de DNA. Deste modo, os cromatogramas relativos a ambas as sequências foram analisados e manualmente editados, tendo como finalidade a construção de uma sequência consenso. Este procedimento foi efectuado com recurso ao programa *BioEdit Sequence Alignment Editor* (versão 7.0.9.0), disponível em <http://www.mbio.ncsu.edu/bioedit/bioedit.html>.

Após a obtenção de uma sequência final para todas moléculas de DNA, procedeu-se a uma pesquisa de regiões de semelhança local entre sequências, tendo como objectivo confirmar, de um modo rápido, a identidade das sequências obtidas, com recurso ao programa *BLAST* (do inglês, *Basic Local Alignment Search Tool*) (disponível em: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Este último efectua uma comparação entre a sequência submetida e as sequências presentes na base de dados, apresentando qual a percentagem de semelhança entre estas. Além disso, estima também a probabilidade da semelhança encontrada ser devida ao acaso (*E value*, compreendido entre 0-1).

2.9.2. Genotipagem automática das sequências de HIV-1

A determinação rápida automatizada dos genótipos virais foi realizada através do programa *REGA HIV-1 & 2 Automated Subtyping Tool* (versão 2.0) (disponível em: <http://bioafrica.net/reg-a-genotype/html/subtypinghiv.html>). Este programa considera um conjunto pré-definido de sequências de referência do grupo M de HIV-1 para a construção automatizada de uma árvore filogenética que inclui igualmente as sequências em estudo, o que permite a sua genotipagem. A classificação genotípica é, no entanto, apenas efectuada quando os valores de *bootstrap* associados ao agrupamento filogenético que contém a sequência em estudo são superiores a 70% (limite de significância). Esta ferramenta permite ainda detectar a presença de eventos de recombinação, por métodos de *bootscanning*. Contudo, esta análise é apenas realizada em sequências de tamanho superior a 800 pb.

2.9.3. Análise filogenética

A classificação genotípica das sequências nucleotídicas em estudo foi realizada com recurso à construção de uma árvore filogenética, através de métodos não automatizados. Este processo teve início com a construção de um alinhamento múltiplo, contendo as sequências em estudo, algumas sequências obtidas por pesquisa de regiões de semelhança local entre sequências (*BLAST*), sequências de referência do grupo M (subtipos, sub-subtipos e CRFs) e uma sequência do grupo N do HIV-1, como *outgroup*, disponíveis em *Los Alamos HIV sequence database* (<http://www.hiv.lanl.gov>). Este alinhamento foi realizado com recurso ao programa

ClustalW2 (disponível em: <http://www.ebi.ac.uk/Tools/msa/clustalw2/>) e posteriormente editado no programa *BioEdit Sequence Alignment Editor* (versão 7.0.9.0). A árvore filogenética foi construída utilizando o programa *MEGA* (do inglês, *Molecular Evolutionary Genetics Analysis*) versão 5 (Tamura et al., 2011), com recurso ao método de *Neighbor-Joining* (Saitou & Nei, 1987), a partir de matrizes de distâncias genéticas calculadas com base no método de dois parâmetros de Kimura (Kimura, 1980), assumindo-se uma frequência igual de bases, uma taxa de transversões igual à de transições e a opção *pairwise deletion*. A robustez das topologias sugeridas foi avaliada pelo método de *bootstrapping*, com base em 1000 amostragens aleatórias. Quando os valores de *bootstrap* foram iguais ou superiores a 70%, os agrupamentos formados foram considerados consistentes do ponto de vista filogenético.

Utilizando as condições acima descritas, foi também construída uma árvore filogenética a partir de sequências concatenadas que incluíam as regiões codificantes da protease (obtidas neste estudo) e da transcriptase reversa (sequências parciais), para as amostras em que ambas estavam disponíveis (n=53). Estas últimas foram obtidas no âmbito de uma dissertação de mestrado anterior (Espinheira, 2013), sendo gentilmente cedidas pela Inv.^a Doutora Elizabeth Pádua (INSA).

2.9.4. Análise de sequências divergentes

Para as sequências nucleotídicas cujos ramos de suporte na árvore filogenética construída evidenciavam posicionamento ambíguo, relativamente aos grupos monofiléticos formados, ou uma elevada divergência filogenética (ramos de suporte muito longos), procedeu-se a uma análise complementar, para possível identificação de fenómenos de recombinação ou de hipermutação. No primeiro caso, recorreu-se ao programa *SimPlot* versão 3.5.1 (disponível em: <http://sray.med.som.jhmi.edu/SCRsoftware/simplot/>), com o intuito de identificar possíveis eventos de recombinação, através do método de *bootscanning* (Lole et al., 1999). Nesta ferramenta bioinformática, foi aplicado o método de *Neighbor-Joining* (Saitou & Nei, 1987) para a inferência filogenética e as distâncias genéticas calculadas pelo método de dois parâmetros de Kimura (Kimura, 1980). A robustez da inferência filogenética foi estimada com base em 1000 amostragens aleatórias e considerada significativa quando $\geq 70\%$. Foram ainda pesquisados fenómenos de hipermutação

(Rose & Korber, 2000) por ocorrência de hiperadenilação (G→A), por acção das APOBEC (do inglês, *apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like*), com recurso ao programa *Hypermut 2.0* (disponível em: <http://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html>). Este último identifica as substituições nucleotídicas presentes nas sequências em estudo, comparativamente com uma sequência de referência, escolhida pelo operador. O relatório final contém os resultados obtidos para cada sequência, baseados no teste exacto de Fisher, e indica quais são estatisticamente significativos. Como tal, as sequências em estudo foram previamente agrupadas de acordo com o seu genótipo provável e, de seguida, alinhadas com uma sequência consenso de referência do genótipo respectivo, disponível na *Los Alamos HIV sequence database* (<http://www.hiv.lanl.gov>).

2.10. Análise bioinformática das sequências aminoacídicas

2.10.1. Pesquisa de mutações associadas à diminuição da susceptibilidade aos anti-retrovirais inibidores da protease

Com base nas sequências aminoacídicas putativas traduzidas a partir das sequências nucleotídicas em estudo foi efectuada a pesquisa de mutações associadas à diminuição da susceptibilidade aos anti-retrovirais inibidores da protease, assim como, da presença de outros polimorfismos genéticos. Para isso, recorreu-se ao programa *HIVdb*, implementado através do *Genotypic Resistance Interpretation Algorithm* da *Stanford University HIV Drug Resistance Database* (versão 7.0) (disponível em: <http://sierra2.stanford.edu/sierra/servlet/JSierra>). Os resultados obtidos através da implementação deste programa baseiam-se na comparação entre as sequências em estudo e uma sequência consenso de referência do subtipo B para a região codificante da protease de HIV-1. Deste modo, o programa identifica, para cada sequência introduzida, mutações principais (ou *major*), mutações acessórias (ou *minor*) e outras mutações não associadas a resistência aos anti-retrovirais (aqui denominadas “outros polimorfismos genéticos”). De acordo com a comparação efectuada, é atribuída uma pontuação individual a cada mutação, relativamente a cada fármaco considerado pelo algoritmo, reflectindo o grau de probabilidade, baseado em diversos tipos de estudos na

literatura, dessa mutação estar associada à aquisição de resistência fenotípica a esse mesmo fármaco. Por fim, é realizada a soma de todas as pontuações atribuídas para cada sequência e, consoante o valor final, é efectuada uma classificação do nível de resistência, em uma das seguintes categorias:

- Susceptibilidade – sem evidência de susceptibilidade reduzida comparativamente com a estirpe selvagem (pontuação entre 0 e 9);
- Resistência potencial de nível baixo – o vírus é susceptível ao fármaco; no entanto, pode indicar uma exposição prévia a uma determinada classe terapêutica anti-retroviral (pontuação entre 10 e 14);
- Resistência de nível baixo – o vírus pode apresentar susceptibilidade reduzida ao fármaco *in vitro* e/ou em indivíduos infectados, os quais podem apresentar uma resposta virológica incompleta ao tratamento anti-retroviral (pontuação entre 15 e 29);
- Resistência de nível intermédio – o vírus apresenta susceptibilidade reduzida ao fármaco, com impacto significativo na sua eficácia (pontuação entre 30 a 59);
- Resistência de nível elevado – o vírus apresenta níveis de susceptibilidade muito reduzidos, ou mesmo ausentes, em resposta ao tratamento anti-retroviral e, *in vitro*, vírus com genótipo similar apresentam os níveis mais elevados de resistência (pontuação superior a 60).

2.10.2. Análise da variabilidade dos domínios funcionais da protease

Por último, procedeu-se à análise dos polimorfismos encontrados nas sequências aminoacídicas, independentemente de estarem ou não associados a resistência aos anti-retrovirais. Esta análise teve como finalidade avaliar o grau de conservação dos diferentes domínios funcionais da protease. Para tal, foi construído um alinhamento múltiplo que permitiu comparar as sequências aminoacídicas em estudo com uma sequência aminoacídica consenso do subtipo B de HIV-1, disponível na *Los Alamos HIV sequence database* (<http://www.hiv.lanl.gov>), recorrendo ao programa *BioEdit Sequence Alignment Editor* (versão 7.0.9.0). Além disso, foi ainda construído um *weblogo*, com recurso à ferramenta bioinformática *AnalyzeAlign* (disponível em: http://www.hiv.lanl.gov/content/sequence/ANALYZEALIGN/analyze_align.html), a partir do alinhamento múltiplo das sequências em estudo. Como resultado final, este apresenta uma sequência aminoacídica, na qual, em cada posição, estão representados

os diversos aminoácidos identificados no conjunto de sequências, sendo que cada aminoácido é simbolicamente representado por uma letra, no código IUPAC (do inglês, *International Union of Pure and Applied Chemistry*), cujo tamanho é proporcional à probabilidade do mesmo estar presente naquela posição, quando considerado o conjunto de sequências estudado. Assim, esta ferramenta permite verificar qual o aminoácido mais frequente em cada posição, avaliar quais as posições mais conservadas, assim como inferir sobre o grau de conservação dos diferentes domínios funcionais da protease.

2.11. Análise estatística

Aos dados epidemiológicos e clínicos, de natureza categórica, que caracterizam basicamente a amostra populacional estudada, como nacionalidade e presença/ausência de terapia anti-retroviral e, em caso afirmativo, respectivo regime, foi aplicada uma análise de distribuição. Relativamente à idade (ao nascimento do primeiro filho), para além de uma análise de distribuição, calculou-se ainda a média e o desvio padrão. A associação de polimorfismos genéticos com os genótipos virais mais frequentes (B, C e G) em circulação na população estudada (“assinaturas”) foi estudada com recurso ao teste de χ^2 ou ao teste exacto de Fisher bilateral, aplicados à análise de tabelas de contingência 2x2 (disponíveis em <http://www.physics.csbsju.edu/stats/>), considerando um valor de significância de 1%.

3. *Resultados*

3.1. Caracterização da amostra populacional estudada

Este estudo incluiu 34 mulheres múltíparas, infectadas com HIV-1, às quais foi efectuada uma colheita de sangue periférico, após o nascimento de cada filho, como mencionado anteriormente (secção 2.2). A maioria das mulheres em estudo (64,7%) era de nacionalidade portuguesa, tendo as restantes origem em Angola (23,5%) e Moçambique (8,8%). À altura do nascimento do primeiro filho, tinham uma idade compreendida entre os 17 e os 36 anos (idade média de $26,9 \pm 5,09$ anos) (Figura 11; Anexo 2).

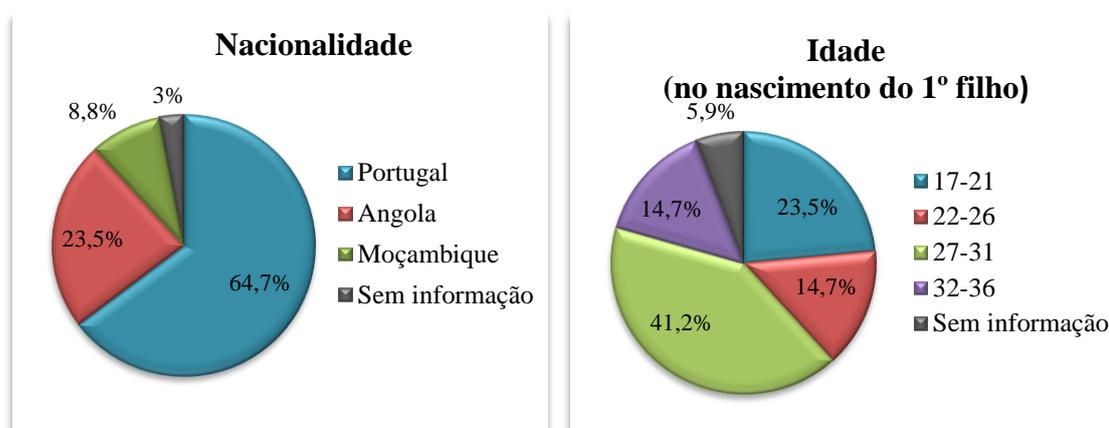


Figura 11. Representação gráfica dos dados demográficos da amostra populacional em estudo, relativos à nacionalidade e idade (no nascimento do primeiro filho).

Na amostra populacional em estudo, 32 mulheres foram mães de dois filhos e duas mulheres foram mães de três filhos, obtendo-se assim um total de 70 amostras de sangue periférico maternas. Foram recolhidos e analisados os dados referentes à realização de terapia anti-retroviral durante as gestações. Na primeira gravidez, seis mulheres não efectuaram qualquer terapia anti-retroviral e, destas últimas, duas também não realizaram durante a segunda gestação. Em suma, em 11,4% (8/70) das gestações não foi realizada terapia anti-retroviral (Figura 12; Anexo 2).

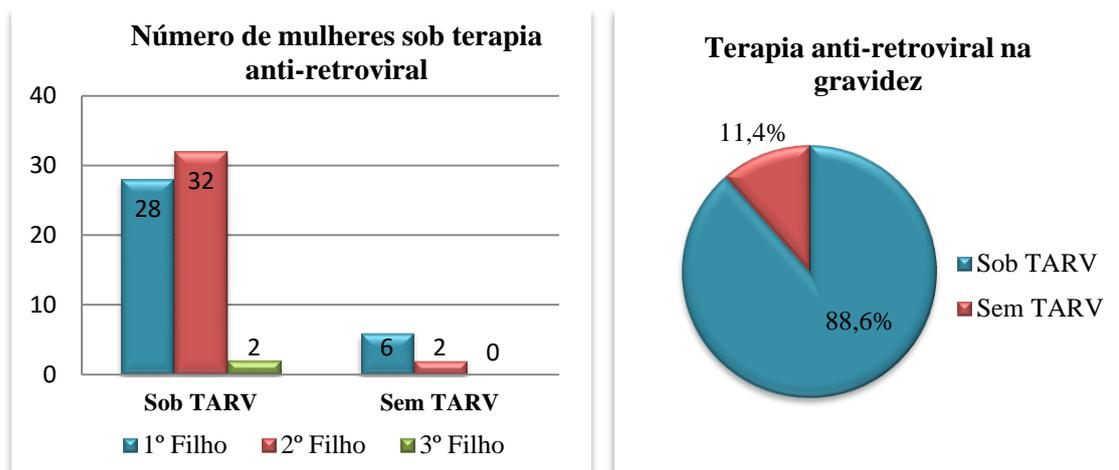


Figura 12. Representação gráfica dos dados relativos à realização de terapia anti-retroviral pelas 34 mulheres infectadas pelo HIV-1 durante as 70 gravidezes consideradas.

Relativamente às gestações nas quais foi efectuada terapia anti-retroviral (n=62), foram analisados quais os regimes terapêuticos aplicados e, por sua vez, quais incluíram inibidores da protease. Deste modo, em 90,3% (56/62) dos casos realizou-se HAART durante as gestações e verificou-se, ainda, que, no mínimo, em 37,1% (23/62) dos casos foi efectuada a toma de inibidores da protease (Figura 13; Anexo 2). Digno de realce que em cerca de um terço dos casos (32,3%), não se tinha informação concreta sobre o regime administrado, para além da base terapêutica com combinação de dois NRTIs.

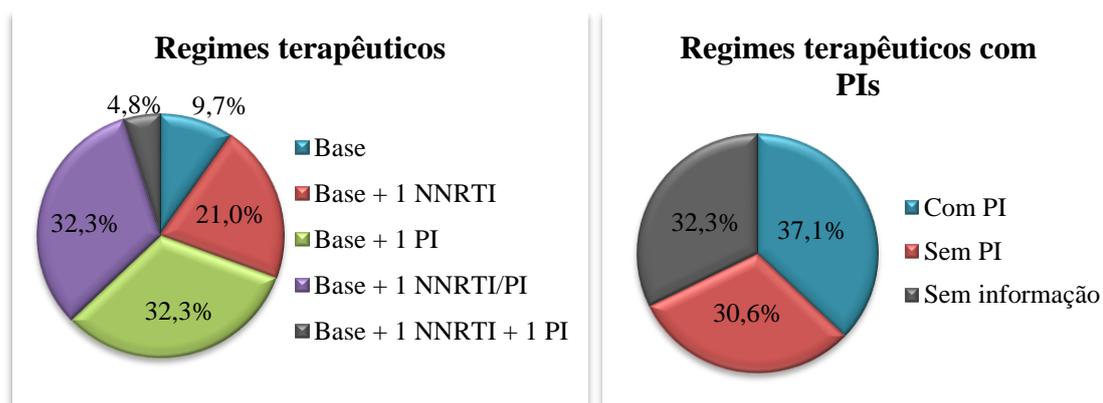


Figura 13. Representação gráfica dos dados relativos aos regimes aplicados nas 62 gestações sujeitas a terapia anti-retroviral.

Por fim, constatou-se a ocorrência de transmissão vertical do HIV-1 para três dos recém-nascidos (4,3%). Todos estes casos ocorreram em mulheres que não

efectuaram terapia anti-retroviral para prevenção da transmissão mãe-filho (PT_10A, PT_37A, PT_43A).

3.2. Amplificação da região codificante da protease por *nested* PCR dupla

Nas 70 amostras de sangue periférico maternas foi efectuada a extracção do DNA das células mononucleadas, contendo o DNA proviral, a partir do qual foi realizada a amplificação da região codificante da protease no gene *pol* do HIV-1. Esta amplificação foi realizada por *nested* PCR dupla, de acordo com o protocolo descrito no capítulo anterior (secção 2.4). Os *primers* internos, PR3 e PR4, utilizados na última reacção de PCR, permitem a amplificação de um fragmento de DNA com 460 pb (posições nucleotídicas baseadas na sequência de referência HXB2, indicadas anteriormente na Tabela 5). Após a execução do protocolo de amplificação base (ver secção 2.4.2.) ou de protocolos alternativos, com temperaturas de hibridação mais baixas, sempre que as amostras se revelavam inicialmente não amplificáveis, obteve-se um total de 67 amostras positivas (95,7%). Nas restantes amostras (n=3), apesar de vários protocolos de amplificação testados, não foi obtido o produto de amplificação pretendido (Figura 14).

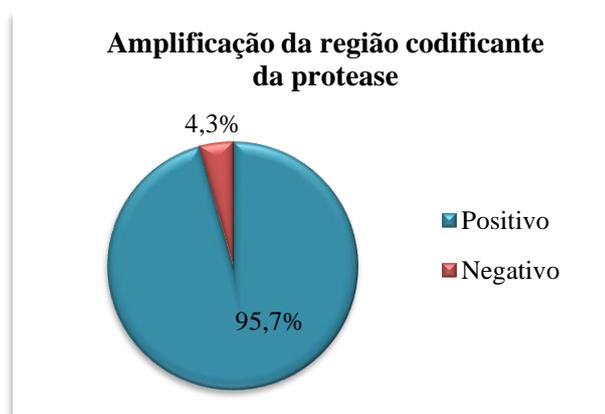


Figura 14. Representação gráfica dos resultados obtidos (em percentagem) após amplificação da região codificante da protease por *nested* PCR dupla.

A observação dos produtos de amplificação, obtidos por *nested* PCR dupla, foi realizada após migração por electroforese em gel de agarose (Figura 15). A análise dos diversos padrões de migração obtidos permitiu identificar amostras que apresentavam

produtos de amplificação específicos, com um tamanho aproximado de 460 pb (Figura 15, poços 9 e 10, por exemplo) e amostras que apresentavam um fragmento de DNA com o tamanho desejado, acompanhado, no entanto, de produtos inespecíficos (Figura 15, poços 13 e 14, por exemplo).

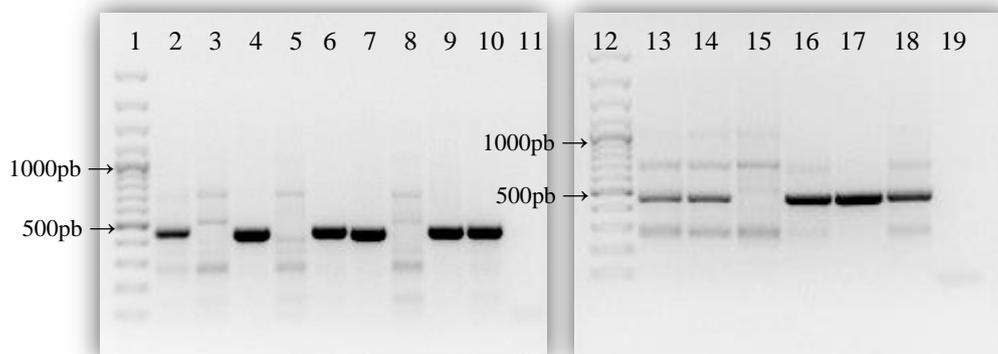


Figura 15. Observação dos produtos de amplificação obtidos por *nested* PCR dupla, relativos à região codificante da protease (≈ 460 pb), após electroforese em gel de agarose a 1,2% (p/v), em tampão TAE 0,5x, com brometo de etídeo (0,5 $\mu\text{g/ml}$). 1 e 12 – Marcador de massa molecular *GeneRuler™ 100 bp Plus DNA Ladder* (Fermentas, Lituânia); 2 a 10 e 13 a 18 – Amostras em estudo; 11 e 19 – Controlo negativo.

O procedimento seguinte divergia consoante os produtos de amplificação apresentassem maior ou menor especificidade, tal como anteriormente descrito (ver secção 2.6.). Relativamente às amostras (36/67, 53,7%) cujos produtos de amplificação apresentavam um fragmento específico, de tamanho esperado, com uma quantidade suficiente de DNA, foi realizada a sua purificação directamente a partir do produto de PCR. Por sua vez, nas amostras (26/67, 38,8%) que apresentavam fragmentos resultantes de amplificação inespecífica, para além do fragmento correspondente à região codificante da protease (460 pb), foi realizada a purificação deste último a partir do gel de agarose. Em alguns casos (5/67, 7,5%), e de acordo com os critérios descritos na secção 2.7., foi aplicado *a posteriori* o método de clonagem com recurso ao vector *pGEM®-T Easy* (Promega, EUA).

3.3. Clonagem dos produtos de PCR em *pGEM[®]-T Easy*

A clonagem molecular foi efectuada para os amplicões relativos às amostras PT_27A, PT_43A, PT_43B, PT_48B e PT_72A (n=5). O vector plasmídico utilizado, com a estirpe *E. coli* JM109 como hospedeiro, permite seleccionar os clones recombinantes putativos, após o processo de transformação. Esta selecção, baseada no fenómeno de α -complementação do gene da β -galactosidase, descrito anteriormente (secção 2.7), é realizada após o crescimento bacteriano em meio LB sólido selectivo, no qual as colónias podem apresentar coloração branca (clones recombinantes) ou azul (clones não recombinantes).

Assim, em cada experiência de clonagem molecular, as colónias brancas foram repicadas para meio LB líquido suplementado com ampicilina, para posterior extracção e purificação do DNA plasmídico através do método de lise alcalina. Findo o processo de extracção do DNA plasmídico dos diferentes clones seleccionados, procedeu-se à análise dos respectivos padrões de migração, após separação electroforética em gel de agarose. Nesta análise, é expectável que as moléculas de DNA plasmídico que contenham um fragmento de DNA exógeno apresentem um padrão de migração mais retardado. Na Figura 16 estão representados os DNAs plasmídicos referentes a três experiências independentes de clonagem, relativas a três amostras em estudo, e, tal como pode ser observado, os padrões de migração foram bastante semelhantes entre si (Figura 16.A), tendo sido seleccionados dois DNAs plasmídicos de cada experiência de clonagem (representados com o rectângulo vermelho), para posterior hidrólise com a enzima de restrição EcoRI. Os resultados dessa digestão do DNA foram visualizados após separação dos fragmentos obtidos por electroforese em gel de agarose (Figura 16.B). Nos perfis de digestão, foi possível observar um fragmento linearizado, de tamanho aproximado a 460 pb, correspondente ao DNA clonado, juntamente com o vector plasmídico linearizado (tamanho aproximado de 3000 pb). A presença destes dois fragmentos é indicativa de um processo de clonagem bem sucedido, com inserção no vector do fragmento de DNA amplificado. Digno de nota que todos os perfis obtidos (Figura 16.B) são indiciadores de reacções de digestão incompletas, devidas, muito provavelmente, a um excesso de DNA plasmídico.

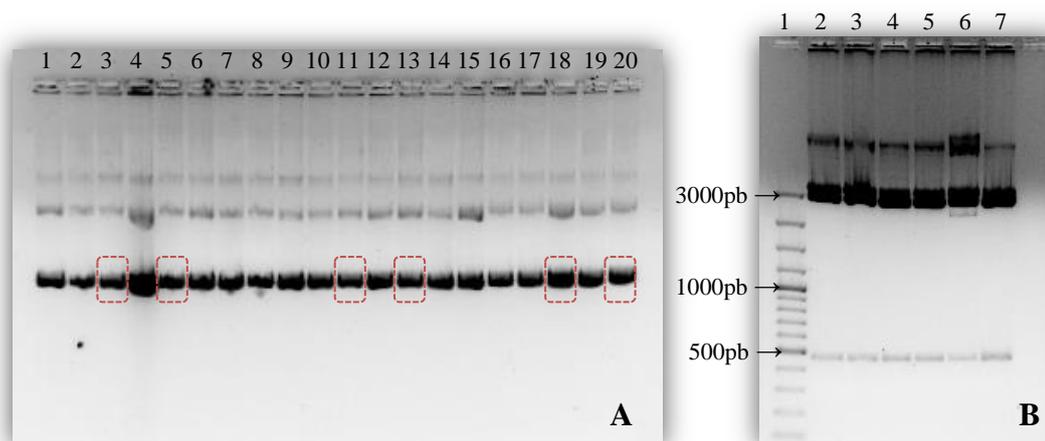


Figura 16. **A)** Observação dos padrões de migração dos DNAs plasmídicos dos clones bacterianos recombinantes putativos seleccionados, após electroforese em gel de agarose a 1,0% (p/v), em tampão TAE 0,5x, com brometo de etídeo (0,5 µg/ml). 1 a 6, 8 a 13 e 15 a 20 – DNAs plasmídicos resultantes de três experiências independentes de clonagem (os rectângulos de cor vermelha indicam possíveis recombinantes, seleccionados para posterior hidrólise enzimática); 7 e 14 – Controlo positivo (DNA plasmídico recombinante, contendo um fragmento de DNA exógeno com cerca de 460 pb). **B)** Análise dos produtos resultantes da hidrólise com a enzima de restrição EcoRI, após migração electroforética em gel de agarose a 1,2% (p/v), em tampão TAE 0,5x, com brometo de etídeo (0,5 µg/ml). 1 – Marcador de massa molecular *GeneRuler™ 100 bp Plus DNA Ladder* (Fermentas, Lituânia); 2 a 7 – Perfis de digestão dos DNAs plasmídicos assinalados em A.

Posteriormente, foram preparadas novas culturas celulares em meio LB líquido com ampicilina dos clones bacterianos com perfil de digestão do DNA plasmídico positivo (Figura 16.B). A partir destas culturas foi preparado novo DNA plasmídico, com recurso ao sistema *Wizard® Plus SV Minipreps DNA Purification System* (Promega, EUA), para posterior sequenciação nucleotídica.

3.4. Análise bioinformática das sequências nucleotídicas

Após a sequenciação nucleotídica dos 67 amplicões em estudo procedeu-se à análise dos respectivos cromatogramas. Para cada amostra foi obtida uma sequência no sentido directo e reverso, as quais foram comparadas e editadas de modo a obter uma sequência consenso final. Devido à existência de *quasiespecies*, em algumas amostras não foi possível definir um único nucleótido para determinadas posições do cromatograma, tendo sido atribuída, nestes casos, uma base degenerada, tendo em conta o código IUPAC. Após a obtenção de uma sequência nucleotídica consenso para cada amostra, procedeu-se à pesquisa de regiões de semelhança local entre sequências, com

recurso ao programa *BLAST*, com o intuito de confirmar, de um modo rápido, a sua identidade. Foram registadas quais as sequências da base de dados que apresentavam maior similaridade com as sequências em estudo, assim como o *E value* obtido. Deste modo, foram observadas percentagens de similaridade com sequências de HIV-1 iguais ou superiores a 91%, com *E value* igual ou muito próximo de zero ($\leq 1e^{-157}$) (Anexo 3).

3.4.1. Genotipagem automática das sequências do HIV-1

Antes de se proceder à construção da árvore filogenética para genotipagem, recorreu-se ao programa *REGA HIV-1 & 2 Automated Subtyping Tool* (versão 2.0), de modo a realizar-se uma genotipagem preliminar automatizada das sequências em estudo e, com isso, obter-se um conhecimento inicial da diversidade dos genótipos presentes. Verificou-se que os subtipos predominantes eram G (29/67, 43,3%), B (11/67, 16,4%) e C (10/67, 14,9%), sendo ainda detectado o sub-subtipo A1 (1/67, 1,5%). Além disso, de realçar que para 23,9% (16/67) das sequências não foi possível a sua classificação, sugerindo a presença de sequências muito divergentes ou recombinantes (Anexo 4). A confirmação dos genótipos atribuídos foi efectuada por análise filogenética não automatizada (ver secção seguinte).

3.4.2. Análise filogenética

A análise filogenética das sequências em estudo foi efectuada através da construção de uma árvore filogenética por métodos não automatizados (tal como descrito no subcapítulo 2.9.3.). Num processo iterativo, diversas sequências, incluídas no alinhamento múltiplo inicial, foram sendo excluídas, até à obtenção de uma árvore filogenética com a maior robustez topológica possível. Deste modo, a árvore filogenética final (Figura 17.A), construída a partir de um alinhamento múltiplo com um total de 321 nucleótidos, é constituída por 124 sequências nucleotídicas, incluindo, 67 sequências em estudo, 20 sequências obtidas da base de dados *GenBank/EMBL/DDBJ* por pesquisa de regiões de semelhança local entre sequências com recurso ao programa *BLAST*, 36 sequências de referência do grupo M do HIV-1 e uma sequência de referência do grupo N do HIV-1 (utilizada como grupo externo ou *outgroup*). Sucintamente, esta árvore filogenética foi construída tendo por base o método de *Neighbor-Joining* (Saitou & Nei, 1987), a partir de matrizes de distâncias genéticas calculadas pelo método de dois parâmetros de Kimura (Kimura, 1980), sendo

a robustez da sua topologia avaliada pelo método de *bootstrapping* (baseado em 1000 réplicas). Os agrupamentos gerados foram considerados consistentes, do ponto de vista filogenético, quando os valores de *bootstrap* obtidos foram iguais ou superiores a 70%.

Após a construção da árvore filogenética foi possível constatar que as 67 sequências nucleotídicas em estudo se encontram distribuídas por cinco agrupamentos distintos, dois dos quais (G/14_BG e C, na Figura 17.A) suportados por valores de *bootstrap* significativos (86% e 84%, respectivamente). No entanto, mesmo os agrupamentos não significativos do ponto de vista filogenético apresentam topologias, em geral, bem definidas. Assim, foi possível identificar de modo inequívoco os genótipos da maioria das sequências em estudo. Dado o seu posicionamento ambíguo na árvore filogenética ou a elevada divergência genética apresentada, relativamente ao nó ancestral, um grupo de 13 sequências, emparelhando maioritariamente junto de referências do subtipo B, foi seleccionado para análise de *bootscanning* (Figura 17.B), para pesquisa de eventuais fenómenos de recombinação intergenotípica.

Dando início a uma análise mais específica da árvore filogenética da Figura 17.A, observou-se que 32 das sequências em estudo (PT_3A, PT_3B, PT_6A, PT_6B, PT_10A, PT_10B, PT_13A, PT_13B, PT_17A, PT_17B, PT_17C, PT_24A, PT_24B, PT_29A, PT_29B, PT_32A, PT_32B, PT_33A, PT_33B, PT_43A, PT_43B, PT_52A, PT_52B, PT_54B, PT_57A, PT_57B, PT_60A, PT_60B, PT_65A, PT_65B, PT_66A, PT_66B) se encontravam agrupadas com sequências de referência do subtipo G e da CRF14_BG, com um valor de *bootstrap* significativo (86%). Convém realçar que a região codificante da protease na CRF14_BG é classificada no genótipo G (Anexo 6). Assim sendo, estas sequências foram definidas como pertencentes ao subtipo G (G/14_BG).

Para o grupo monofilético que inclui as sequências de referência da CRF02_AG não foi obtido um valor de *bootstrap* significativo que permita uma genotipagem muito consistente das sequências em estudo. Não obstante, para este grupo, a topologia da árvore aponta para uma evidente similaridade das sequências PT_5A, PT_5B, PT_72A e PT_72B (n=4) com esta forma recombinante. Por outro lado, as duas últimas emparelham de modo significativo com uma sequência CRF02_AG retirada do *GenBank*. A região nucleotídica em análise (posições 2167 a 2583, no genoma do HIV-1 HXB2) corresponde a uma região híbrida A/G do genoma da CRF02_AG (Anexo 6).

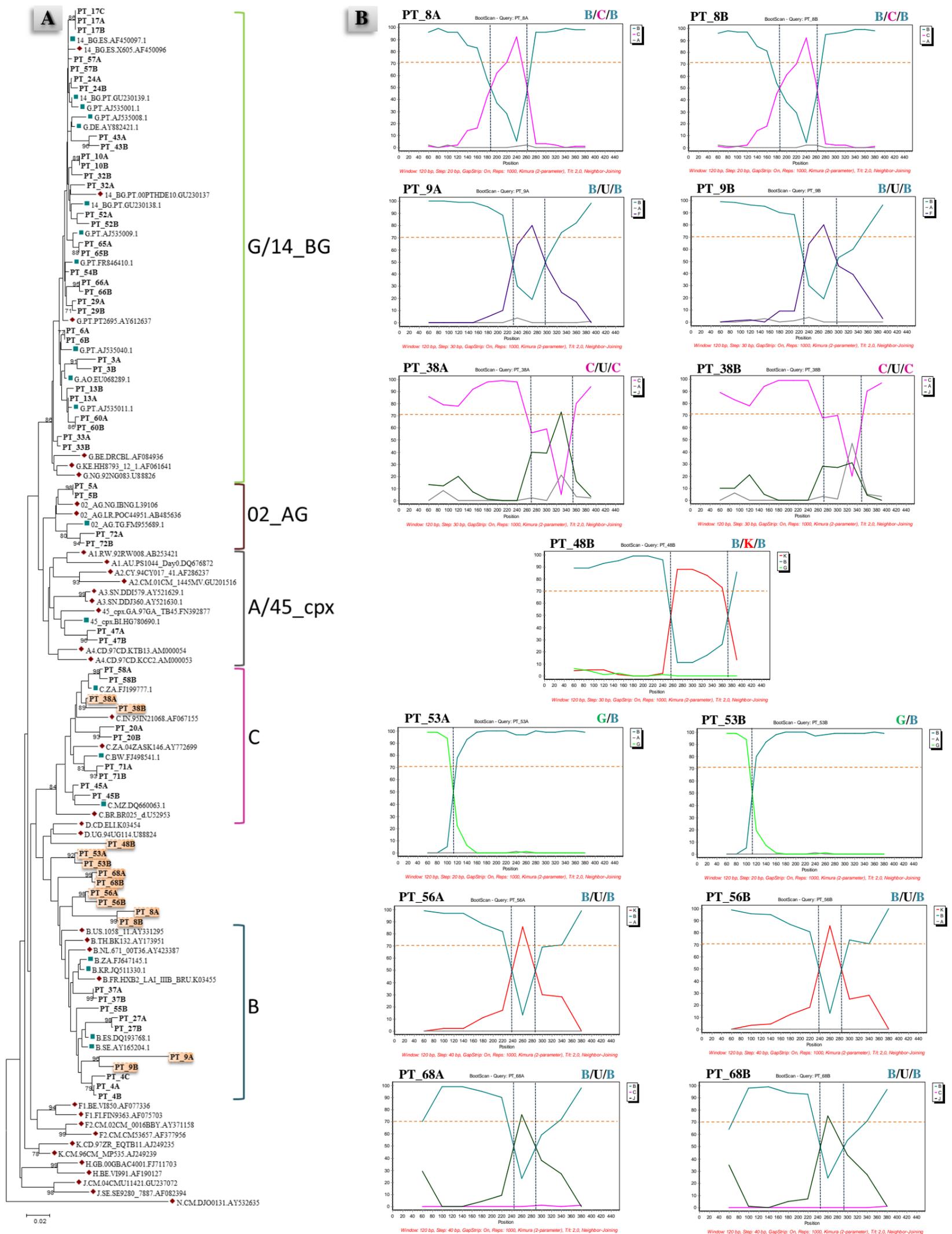


Figura 17. A) Relações filogenéticas entre 124 sequências nucleotídicas da região codificante da protease do HIV-1, com base num total de 321 nucleótidos. A árvore filogenética inclui as 67 sequências em estudo (representadas com o prefixo PT), 20 sequências obtidas do *GenBank* por pesquisa com o programa *BLAST* de regiões de semelhança local entre sequências (representadas com o símbolo ■), 36 sequências de referência do grupo M do HIV-1 (A1-A4, B, C, D, F1, F2, G, H, J, K, CRF02_AG, CRF14_BG e CRF45_cpx) e uma sequência de referência do grupo N (utilizada como *outgroup*) (representadas com o símbolo ◆). A designação das sequências de referência é iniciada com uma letra, correspondente ao genótipo, seguida do local de origem, nome original e número de acesso no *GenBank*. As origens geográficas das sequências de referência são: AO- Angola, AU- Austrália, BE- Bélgica, BI- Burundi, BR- Brasil, BW-Botsuana, CD- República Democrática do Congo, CM- Camarões, CY- Chipre, DE- Alemanha, ES- Espanha, FI- Finlândia, FR- França, GA- Gabão, GB- Reino Unido, IN- Índia, KE- Quênia, KR- Coreia do Sul, LR- Libéria, MZ- Moçambique, NG- Nigéria, NL- Holanda, PT- Portugal, RW- Ruanda, SE- Suécia, SN- Senegal, TG- Togo, TH- Tailândia, UG- Uganda, US- Estados Unidos da América, ZA- África do Sul. A inferência filogenética foi realizada com recurso ao programa *MEGA* v.5.0, tendo por base o modelo de *Neighbor-Joining* e o método de dois parâmetros de Kimura para o cálculo das distâncias genéticas (a escala indica 2% de distância genética). A robustez da inferência filogenética foi avaliada por *bootstrapping* de 1000 amostragens aleatórias, considerando-se significativos valores $\geq 70\%$. **B)** Análise de *bootscanning*, por *SimPlot* v.3.5.1, das sequências assinaladas com um retângulo laranja (modelos idênticos aos utilizados na construção da árvore), sendo a janela e o passo definidos individualmente (parâmetros indicados abaixo de cada imagem). As linhas horizontais tracejadas indicam o *cut-off* de 70%.

Relativamente ao agrupamento formado pelas sequências de referência do subtipo A (sub-subtipos A1 a A4) e da CRF45_cpx, este engloba duas das sequências em estudo (PT_47A e PT_47B). Ambas surgem agrupadas com uma sequência classificada como CRF45_cpx, mas com sinal filogenético baixo. Nesta CRF, o segmento do genoma viral em análise tem uma estrutura recombinante A1/U (“não genotipável”, do inglês *unknown*) (Anexo 6). Por isso, tendo em conta o grupo monofilético formado e a estrutura mosaico da CRF45_cpx pressupõe-se que estas sequências se encontrem mais próximas, do ponto de vista filogenético, do subtipo A (puro ou ancestral na base da CRF45_cpx).

Por sua vez, o grupo monofilético que corresponde ao subtipo C, incluindo 16 sequências, entre as quais 10 sequências em estudo, encontra-se suportado por um valor de *bootstrap* de 84%, permitindo uma genotipagem inequívoca das sequências PT_20A, PT_20B, PT_38A, PT_38B, PT_45A, PT_45B, PT_58A, PT_58B, PT_71A e PT_71B. No entanto, verificou-se que a sequência PT_38B é suportada por um ramo muito longo na árvore filogenética (Figura 17.A) e, por isso, foi seleccionada para pesquisa de eventos de recombinação por análise de *bootscanning* (Figura 17.B) e fenómenos de hipermutação (ver mais à frente). A sequência PT_38A foi também submetida a análise de recombinação, visto ter origem no mesmo indivíduo, relativamente à primeira gestação ocorrida quatro anos antes (Anexo 2).

Considerando as restantes sequências em estudo, dez (PT_4A, PT_4B, PT_4C, PT_9A, PT_9B, PT_27A, PT_27B, PT_37A, PT_37B e PT_55B) apresentam-se agrupadas juntamente com sequências de referência do subtipo B. Contudo, este agrupamento não está suportado por um valor de *bootstrap* consistente (<70%). Não obstante, a topologia da árvore evidencia que as presentes sequências sejam relacionadas filogeneticamente com o subtipo B, embora se apresentem algo divergentes (Figura 17.A). Assim sendo, foram incluídas no subtipo B. Neste agrupamento, constatou-se que a sequência PT_9A é suportada por um ramo muito longo, demonstrando divergência genética elevada. Dado isto, foi seleccionada para posterior análise de fenómenos de recombinação (Figura 17.B) e possíveis eventos de hipermutação. Neste caso, tal como no anteriormente descrito, a sequência PT_9B foi submetida ao mesmo tipo de análise, visto ter origem na mesma mulher, mas referente à segunda gestação, ocorrida três anos mais tarde (Anexo 2).

Por último, verificou-se que um conjunto de nove sequências (PT_8A, PT_8B, PT_48B, PT_53A, PT_53B, PT_56A, PT_56B, PT_68A e PT_68B) não se encontram incluídas em nenhum grupo monofilético consistente, do ponto de vista da inferência filogenética, e, como tal, foram assinaladas para pesquisa de fenómenos de recombinação e eventos de hipermutação. Em geral, estas sequências são suportadas por ramos precocemente divergentes, relativamente à topologia geral da árvore construída, demonstrando, maioritariamente, distâncias genéticas muito elevadas (Figura 17.A).

Os resultados obtidos através da análise filogenética não automatizada são, de uma forma geral, concordantes com os obtidos previamente recorrendo ao programa de genotipagem automática *REGA HIV-1 & 2 Automated Subtyping Tool* (versão 2.0) (Anexo 4). No entanto, foram observados resultados não inteiramente coincidentes entre as duas metodologias, particularmente no que se refere às sequências não genotipáveis pelo método automatizado. Algumas destas (PT_4C, PT_5A, PT_5B, PT_9B e PT_47B) embora genotipáveis pela análise da árvore filogenética obtida, encontram-se incluídas em agrupamentos menos significativos. As restantes (PT_10A, PT_13B, PT_52A, PT_72A e PT_72B) foram inequivocamente subtipadas após a construção da árvore filogenética. De sinal contrário, as sequências PT_56A, PT_56B e PT_68A foram subtipadas como B através do método automatizado, contudo a árvore filogenética construída não permitiu a sua caracterização genotípica, por não se encontrarem incluídas em grupos monofiléticos consistentes (Figura 17.A).

Pesquisa de genomas recombinantes

A identificação de sequências recombinantes intergenotípicas foi realizada com recurso ao programa *SimPlot* (versão 3.5.1), pelo método de *bootscanning*, como referido anteriormente (secção 2.9.4). Com base nos resultados obtidos, foram identificados indícios de recombinação em 11 das sequências estudadas, com graus variáveis de evidência (PT_8A, PT_8B, PT_9A, PT_9B, PT_48B, PT_53A, PT_53B, PT_56A, PT_56B, PT_68A e PT_68B) (Figura 17.B). Para algumas sequências (e.g. PT_48B, PT_53A, PT_53B), os gráficos de *bootscanning* obtidos demonstram claramente a presença de locais de recombinação, em associação com um nível elevado de significância. No entanto, outros casos existem (e.g. PT_9A, PT_9B, PT_68A, PT_68B), para os quais, embora haja indícios da ocorrência de recombinação, com

perda de sinal filogenético relativamente ao genótipo mais provável, o padrão obtido, e respectivo nível de significância, não permitem a inequívoca classificação das sequências em causa como recombinantes (pelo que devem ser consideradas como “recombinantes putativas”).

No que se refere a resultados mais específicos (Figura 17.B), as sequências PT_8A e PT_8B evidenciam fenómenos de recombinação entre os subtipos B e C, apresentando a estrutura mosaico B/C/B. Por sua vez, as sequências PT_9A, PT_9B, PT_56A, PT_56B, PT_68A e PT_68B apresentam em quase todo o seu comprimento uma maior similaridade com o subtipo B contendo, no entanto, uma região intermédia (de tamanho relativamente variável) para a qual o valor de significância obtido não foi significativo (<70%). Para estas regiões, onde não foi possível a genotipagem, foi atribuída a letra U, sendo a estrutura genómica das respectivas sequências designada como B/U/B. Claramente, não se pode igualmente descartar a hipótese destas seis sequências corresponderem, de facto, a formas divergentes do subtipo B, e não recombinantes B/U/B, pelo que devem ser consideradas “recombinantes putativas”. Relativamente à sequência PT_48B, foi identificada a estrutura mosaico B/K/B. Esta apresenta em ambas as extremidades uma homologia inequívoca com o subtipo B, contendo, no entanto, uma região, de, aproximadamente, 120 nucleótidos, que revela evidente similaridade com o subtipo K. Quanto às sequências PT_53A e PT_53B apresentam uma região inicial classificada como subtipo G, enquanto toda a restante sequência é subtipada como subtipo B, estabelecendo assim uma estrutura mosaico G/B. Por último, as sequências PT_38A e PT_38B apresentam uma elevada homologia com o subtipo C, à excepção de um segmento intermédio onde não foi possível efectuar a sua caracterização genotípica. Este facto deve-se à ausência de similaridade significativa com qualquer uma das sequências de referência utilizadas, por total ausência de qualquer sinal filogenético acima do nível de significância estabelecido (70%), sugerindo uma divergência genética elevada para esse segmento. Como tal, estas sequências foram classificadas como C divergentes. O grau de incerteza desta classificação é, no entanto, elevado. Por fim, procedeu-se à comparação dos padrões de recombinação obtidos com os das estruturas mosaico de todas as CRFs já descritas e depositadas na *Los Alamos HIV sequence database* (<http://www.hiv.lanl.gov>),

constatando-se uma total ausência de coincidência, pelo que estas sequências corresponderão, muito provavelmente, a formas recombinantes únicas (URFs).

Pesquisa de fenómenos de hipermutação

Para identificar a possível presença de fenómenos de hipermutação, por ocorrência de hiperadenilação (G→A), por acção das APOBEC, recorreu-se ao programa *Hypermut 2.0*, tal como descrito anteriormente (secção 2.9.4.). Após a análise dos resultados obtidos verificou-se que nas sequências PT_9A e PT_38B ocorreram eventos de hipermutação, estatisticamente significativos ($p < 0,05$). Este fenómeno responderá, ainda que parcialmente, pelo longo tamanho dos ramos que suportam ambas as sequências na árvore filogenética construída (Figura 17.A).

Em suma, os genótipos do HIV-1, para a região codificante da protease, identificados nesta amostra de mulheres múltiparas infectadas, encontram-se resumidos na Tabela 9.

Tabela 9. Caracterização genotípica das sequências em estudo.

Genótipos prováveis	Número de sequências e proporção	Número de mulheres e proporção
A/45_cpx*	2 (3,0%)	1 (2,9%)
B*	8 (11,9%)	4 (11,8%)
C	10 (14,9%)	5 (14,7%)
G/14_BG	32 (47,8%)	16 (47,1%)
02_AG*	4 (6,0%)	2 (5,9%)
URFs[§]	11 (16,4%)	6 (17,6%)
Total	67 (100%)	34 (100%)

* grupos monofiléticos com valores de *bootstrap* inferiores a 70%.

§ formas recombinantes únicas

Tendo em conta os resultados encontrados, os genótipos prováveis das 67 sequências nucleotídicas foram distribuídos segundo a nacionalidade das 34 mulheres infectadas incluídas no estudo (Figura 18; Anexo 2).

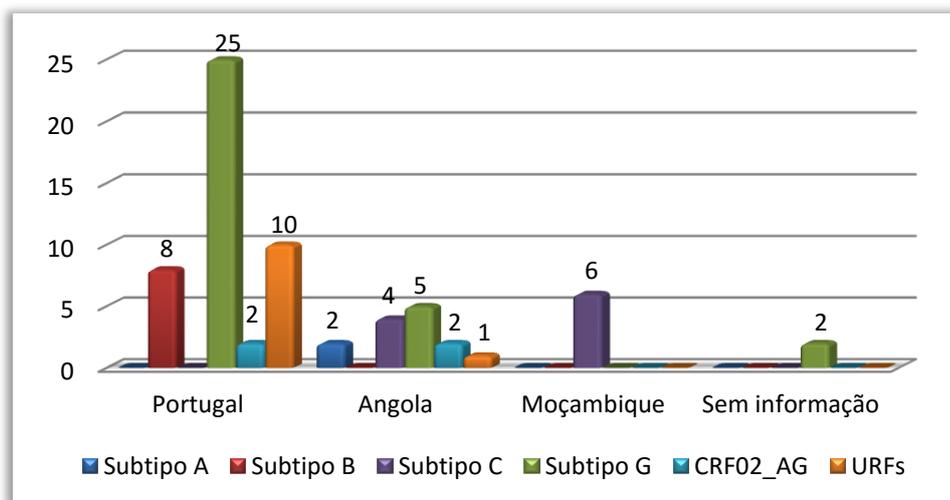


Figura 18. Distribuição dos genótipos prováveis das 67 sequências nucleotídicas do HIV-1 codificantes da protease segundo a nacionalidade das mulheres infectadas incluídas no estudo.

Assim sendo, e com base na Figura 18, verificou-se que nas mulheres infectadas com o HIV-1 com origem em Portugal há um claro predomínio (25/45, 55,6%) do genótipo G e/ou da forma recombinante sua derivada CRF14_BG, cuja região codificante da protease é do subtipo G (Anexo 6). Além disso, destacam-se, ainda, o subtipo B (8/45, 17,8%) e formas recombinantes únicas com origem no subtipo B (10/45, 22,2%). Por sua vez, nas mulheres angolanas, constatou-se a presença de uma elevada heterogeneidade genética, com uma distribuição das sequências pelas diversas formas genéticas encontradas, à excepção do subtipo B, ausente neste grupo. Pelo contrário, nas mulheres com origem em Moçambique, o único genótipo do HIV-1 identificado foi o subtipo C.

3.4.2.1. Análise filogenética de sequências concatenadas

Ainda no âmbito da análise filogenética foi, por fim, construída uma árvore filogenética baseada nas sequências nucleotídicas da região codificante da protease, obtidas neste estudo, juntamente com as sequências nucleotídicas da região codificante da transcriptase reversa, obtidas num estudo anterior (Espinheira, 2013), para amostras correspondentes. Esta árvore filogenética foi construída a partir de sequências artificiais, concatenadas, resultantes da simples justaposição das duas sequências individuais disponíveis, com as mesmas condições utilizadas para a análise filogenética das sequências da região da protease (Figura 17.A). Esta nova árvore compreende um

total de 99 sequências nucleotídicas, nomeadamente, 53 sequências concatenadas (região PR + RT), nove sequências obtidas por pesquisa de regiões de semelhança local entre sequências com recurso ao programa *BLAST*, 36 sequências de referência do grupo M do HIV-1 e uma sequência de referência do grupo N do HIV-1 (utilizada como *outgroup*). O facto de apenas ter sido possível a construção de sequências concatenadas para 53 das 67 sequências da protease estudadas anteriormente deve-se ao insucesso de amplificação por PCR da região codificante da transcriptase reversa das restantes, no estudo respectivo. Deste modo, estão ausentes as sequências PT_4A, PT_10A, PT_17A, PT_20A, PT_27B, PT_29A, PT_37A, PT_52B, PT_53A, PT_56B, PT_58A, PT_60A, PT_66B e PT_68A.

A genotipagem realizada a partir da árvore filogenética construída com as sequências concatenadas (838 nucleótidos) foi, de uma forma geral, concordante com a análise anterior para a região codificante da protease, ajudando, no entanto, a melhor definir eventuais padrões de recombinação ou a existência de sequências geneticamente divergentes. Pela observação da Figura 19.A, foi possível constatar que os valores de *bootstrap* obtidos, nos diversos grupos monofiléticos, em particular para os agrupamentos A, B e 02_AG, eram significativamente superiores aos observados na árvore filogenética anterior (Figura 17.A), na qual apenas alguns agrupamentos (G/14_BG e C) eram suportados por valores de *bootstrap* elevados. As principais diferenças de topologia foram observadas para as sequências PT_8A, PT_8B, PT_47A, PT_47B, PT_48B, PT_53B, PT_56A e PT_68B. Relativamente às sequências PT_47A e PT_47B, na árvore filogenética da região da protease, estas formaram um agrupamento com as sequências de referência do subtipo A e da CRF45_cpx, ainda que, sem valores de *bootstrap* significativos. Nesta árvore filogenética, ambas se encontram agrupadas somente com sequências de referência do subtipo A, com um valor de *bootstrap* de 86%, demonstrando maior semelhança filogenética com o sub-subtipo A3. Por outro lado, as sequências PT_8A, PT_8B, PT_56A e PT_68B, não genotipadas na árvore filogenética da região da protease, foram classificadas como pertencentes ao subtipo B na árvore concatenada. Este resultado poderá ser explicado pelo facto destas sequências apresentarem, em quase toda a sua extensão, uma maior similaridade com o subtipo B (dados não mostrados), com excepção de um pequeno segmento na região da protease, como demonstrado na análise de *bootscanning* realizada anteriormente (Figura

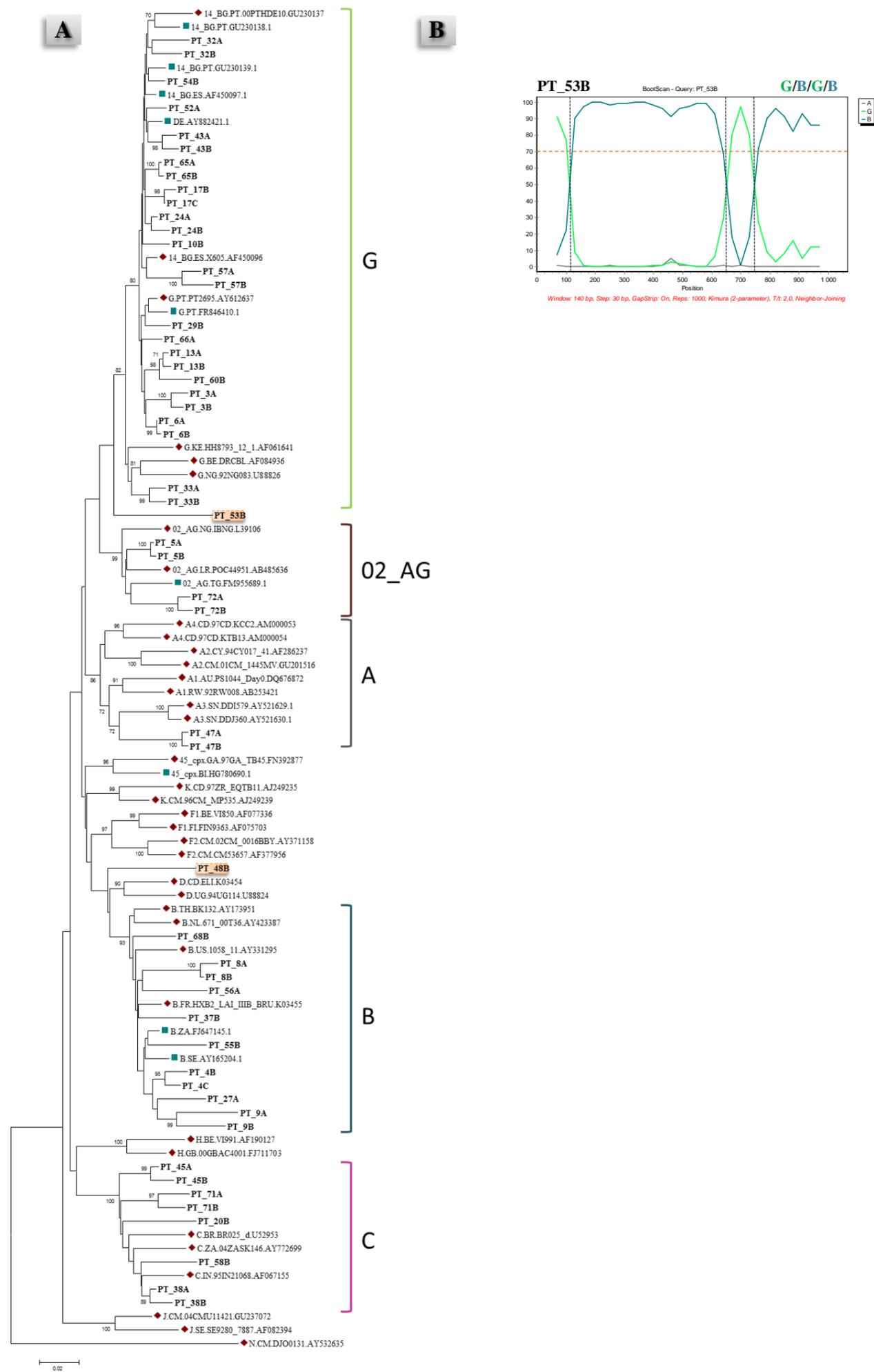


Figura 19. A) Relações filogenéticas entre 99 seqüências nucleotídicas concatenadas compreendendo as regiões codificantes da protease e da transcriptase reversa do gene *pol* do HIV-1, com base num alinhamento múltiplo com um total de 838 nucleótidos. A árvore filogenética inclui as 53 seqüências concatenadas em estudo (representadas com o prefixo PT); 9 seqüências obtidas do *GenBank* por pesquisa com o programa *BLAST* de regiões de semelhança local entre seqüências (representadas com o símbolo ■); 36 seqüências de referência do grupo M do HIV-1 (A1-A4, B, C, D, F1, F2, G, H, J, K, CRF02_AG, CRF14_BG e CRF45_cpx) e uma seqüência de referência do grupo N (utilizada como *outgroup*) (representadas com o símbolo ◆). A designação das seqüências de referência é iniciada com uma letra, correspondente ao respectivo genótipo, seguida do local de origem, nome original e número de acesso no *GenBank*. As origens geográficas das seqüências de referência são: AU- Austrália, BE- Bélgica, BI- Burundi, BR- Brasil, CD- República Democrática do Congo, CM- Camarões, CY- Chipre, DE- Alemanha, ES- Espanha, FI- Finlândia, FR- França, GA- Gabão, GB- Reino Unido, IN- Índia, KE- Quênia, LR- Libéria, NG- Nigéria, NL- Holanda, PT- Portugal, RW- Ruanda, SE- Suécia, SN- Senegal, TG- Togo, TH- Tailândia, UG- Uganda, US- Estados Unidos da América, ZA- África do Sul. A inferência filogenética foi realizada com recurso ao programa *MEGA* v.5.0, tendo por base o modelo de *Neighbor-Joining* e o método de dois parâmetros de Kimura para o cálculo das distâncias genéticas (a escala indica 2% de distância genética). A robustez da inferência filogenética foi avaliada por *bootstrapping* de 1000 amostragens aleatórias, considerando-se significativos valores $\geq 70\%$. **B)** Análise de *bootscanning*, por *SimPlot* v.3.5.1, da seqüência PT_53B (modelos idênticos aos utilizados na construção da árvore filogenética), sendo a janela e o passo os indicados abaixo da imagem. A linha horizontal tracejada indica o *cut-off* de 70%.

17.B). Por último, as sequências PT_48B e PT_53B apresentam, tal como na árvore filogenética para a região da protease, uma posição dúbia nesta árvore filogenética, reforçando a ideia de se tratar de sequências com elevada divergência genética. Por conseguinte, foram ambas submetidas a uma pesquisa de pontos de recombinação por análise de *bootscanning* (ver secção seguinte).

Pesquisa de genomas recombinantes

Tal como anteriormente, a pesquisa de eventos de recombinação foi realizada utilizando o programa *SimPlot*, pelo método de *bootscanning*. As sequências concatenadas PT_48B e PT_53B foram sujeitas a esta análise, das quais apenas para a segunda se obteve um resultado consistente. Na sequência PT_53B foi identificada a estrutura mosaico G/B/G/B (Figura 19.B), justificando assim o seu posicionamento ambíguo na árvore filogenética da Figura 19.A. Já relativamente à sequência PT_48B, não foi possível provar a existência de quaisquer pontos de recombinação, pelo que esta será considerada como uma sequência não genotipável, divergente do ponto de vista filogenético.

3.5. Análise bioinformática das sequências aminoacídicas

Para a análise das 67 sequências aminoacídicas traduzidas, no que se refere à presença de mutações associadas à diminuição da susceptibilidade aos anti-retrovirais inibidores da protease, assim como, de outros polimorfismos genéticos, recorreu-se ao programa *HIVdb*, implementado através do *Genotypic Resistance Interpretation Algorithm* da *Stanford University HIV Drug Resistance Database* (versão 7.0) (disponível em: <http://sierra2.stanford.edu/sierra/servlet/JSierra>), como descrito anteriormente (subcapítulo 2.10.1.). O referido programa emite um relatório interpretativo para cada uma das sequências nucleotídicas introduzidas, contendo a região codificante da protease, sendo os resultados obtidos baseados na comparação com uma sequência consenso do subtipo B. Deste modo, foi feito um levantamento das mutações associadas a resistência aos inibidores da protease (mutações principais e acessórias), assim como, de outros polimorfismos genéticos. As sequências aminoacídicas estudadas correspondem à totalidade da protease (99 aminoácidos). Com

base no mesmo algoritmo foram, ainda, recolhidos para cada sequência os perfis fenotípicos previstos para cada um dos inibidores da protease considerados. No relatório emitido, de acordo com a mutação identificada, foi apresentada uma pontuação individual para cada fármaco e, consoante o valor total obtido para cada sequência submetida, considerando todas as mutações encontradas, foi atribuída uma classificação do nível de resistência (susceptibilidade, potencial de nível baixo, nível baixo, intermédio ou elevado, como descrito no subcapítulo 2.10.1. Esta classificação foi atribuída para oito inibidores da protease (i.e., atazanavir, darunavir, fosamprenavir, indinavir, lopinavir, nelfinavir, saquinavir e tipranavir), pois o ritonavir é, actualmente, apenas utilizado como potenciador farmacocinético.

3.5.1. Identificação de mutações associadas a resistência aos inibidores da protease e outros polimorfismos genéticos

As mutações que conferem resistência aos inibidores da protease podem ser caracterizadas como mutações principais (ou *major*) ou mutações acessórias (ou *minor*), como referido anteriormente. Do conjunto de sequências em estudo, apenas na sequência PT_9A foram encontradas mutações principais, nomeadamente as mutações D30N e M46I (Figura 20). Estas mutações, tal como era expectável, têm um forte impacto na resposta à terapia anti-retroviral, quando esta inclua inibidores da protease. A sua presença confere ao vírus uma resistência de nível elevado ao nelfinavir e uma resistência potencial de nível baixo ao atazanavir, fosamprenavir, indinavir e lopinavir, respectivamente. Relativamente às mutações acessórias L10I/V, L33F, G48E, A71T/V e T74S, estas foram identificadas num total de 19 sequências e as respectivas percentagens estão representadas na Figura 20.

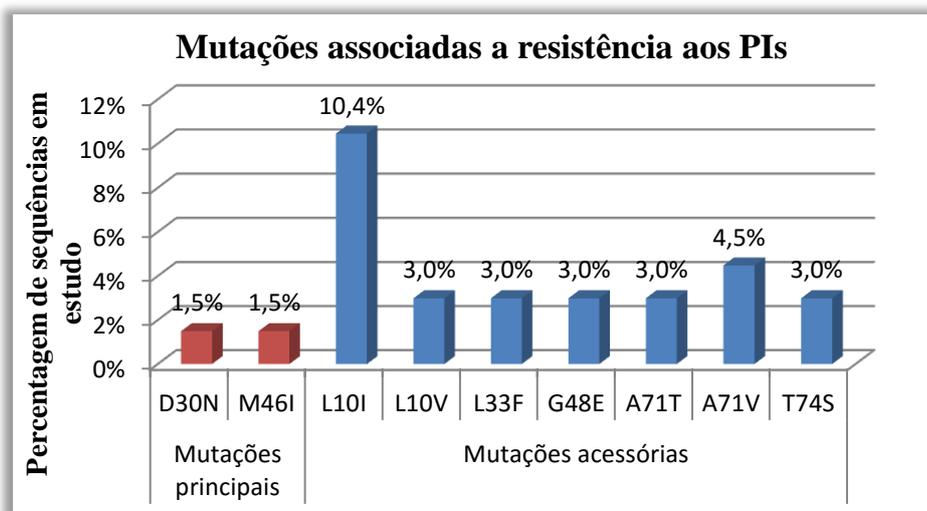


Figura 20. Representação gráfica da prevalência das mutações associadas a resistência aos inibidores da protease (PIs) identificadas na região codificante da protease das seqüências em estudo.

A mutação acessória L10I foi identificada em sete seqüências (PT_4C, PT_17A, PT_17B, PT_17C, PT_47A, PT_47B e PT_65B, nesta última, apenas numa subpopulação de vírus da *quasiespecies*), enquanto a mutação L10V apenas se verificou em duas seqüências (PT_8A e PT_8B). Por sua vez, a mutação L33F está presente nas seqüências PT_43A e PT_43B, enquanto a mutação G48E, uma mutação bastante invulgar, foi identificada nas seqüências PT_9A e PT_38B. As mutações A71T e A71V foram identificadas em duas (PT_27A, PT_27B) e três (PT_9A, PT_9B e PT_45B) seqüências, respectivamente. Por último, a mutação T74S foi encontrada nas seqüências PT_71A e PT_71B. Relativamente ao perfil fenotípico, a maioria destas mutações não estão associadas a uma perda de susceptibilidade aos inibidores da protease. No entanto, as mutações L33F e T74S podem afectar a resposta à terapia anti-retroviral. A primeira pode estar associada a uma resistência potencial de nível baixo ao fosamprenavir e ao tipranavir, embora o seu impacto seja potenciado quando na presença de outras mutações. Por sua vez, a mutação T74S está associada a uma resistência de nível baixo ao nelfinavir.

Ao analisar o genótipo das 19 seqüências nas quais foram detectadas as mutações acima descritas (duas principais e sete acessórias, num número total de 22 ocorrências), constatou-se que 16 correspondem a subtipos não-B. Como tal, foi realizado o teste exacto de Fisher bilateral, ou, quando mais apropriado, o teste de χ^2 , aplicados à análise de tabelas de contingência 2x2, com um nível de significância de

1%, para verificar da possível associação entre a presença destas mutações e os subtipos não-B. No entanto, não se verificou nenhuma associação estatisticamente significativa (Anexos 5.A, .B e .C).

Por último, foram pesquisados e analisados outros polimorfismos genéticos presentes nas 67 sequências da protease em estudo. Estes polimorfismos, segundo o algoritmo de interpretação de resistências genóticas da *Stanford HIVdb* utilizado, não estão associados a resistência aos inibidores da protease e não lhes foi atribuída qualquer pontuação quanto ao perfil fenotípico das estirpes nas quais são encontrados. Para além da utilização deste programa, que permite obter, de forma rápida, estes polimorfismos, foi, ainda, construído um alinhamento múltiplo incluindo as sequências em estudo e uma sequência consenso do subtipo B, utilizada como referência. Este alinhamento permite detectar os polimorfismos (quer sejam associados ou não a resistência aos inibidores da protease) presentes nas várias sequências quando comparadas com a sequência consenso (Figura 21).

Ao analisar os 99 aminoácidos constituintes da protease, nas 67 sequências aminoácídicas disponíveis, foram detectadas 53 posições não conservadas relativamente à sequência consenso do subtipo B (Figura 21), o que corresponde a um grau de conservação de 46,5% das posições. A maioria destes polimorfismos genéticos foram detectados em subtipos não-B, tendo sido realizado o teste exacto de Fisher bilateral, ou, se apropriado, o teste de χ^2 , aplicados à análise de tabelas de contingência 2x2, considerando um nível de significância de 1%, para verificar a possível associação entre os polimorfismos genéticos identificados para cada posição e os diferentes genótipos. Os resultados obtidos (Anexos 5.A, .B e .C) permitiram a identificação de vários polimorfismos associados aos subtipos B, C e G, com significância a nível estatístico ($p < 0,01$). Estes polimorfismos foram então designados como “assinaturas genéticas” desses mesmos subtipos, sendo identificados quatro para as sequências do subtipo B (M36L, R57K, L63P e V77I), cinco para o subtipo C (T12S, I15V, L19I, L19T e I93L) e, por último, dez no subtipo G (I13V, K14R, K20I, E35D, M36I, R41K, H69K, V82I, L89M e F99L) (Anexos 5 e Tabela 10). Digno de nota que a mutação I93L, para além da associação muito significativa encontrada com o subtipo C, foi também associada de um modo significativo ao subtipo B, mas, neste caso, com um valor limiar de probabilidade (Anexo 5.A), pelo que esta última associação não foi considerada.

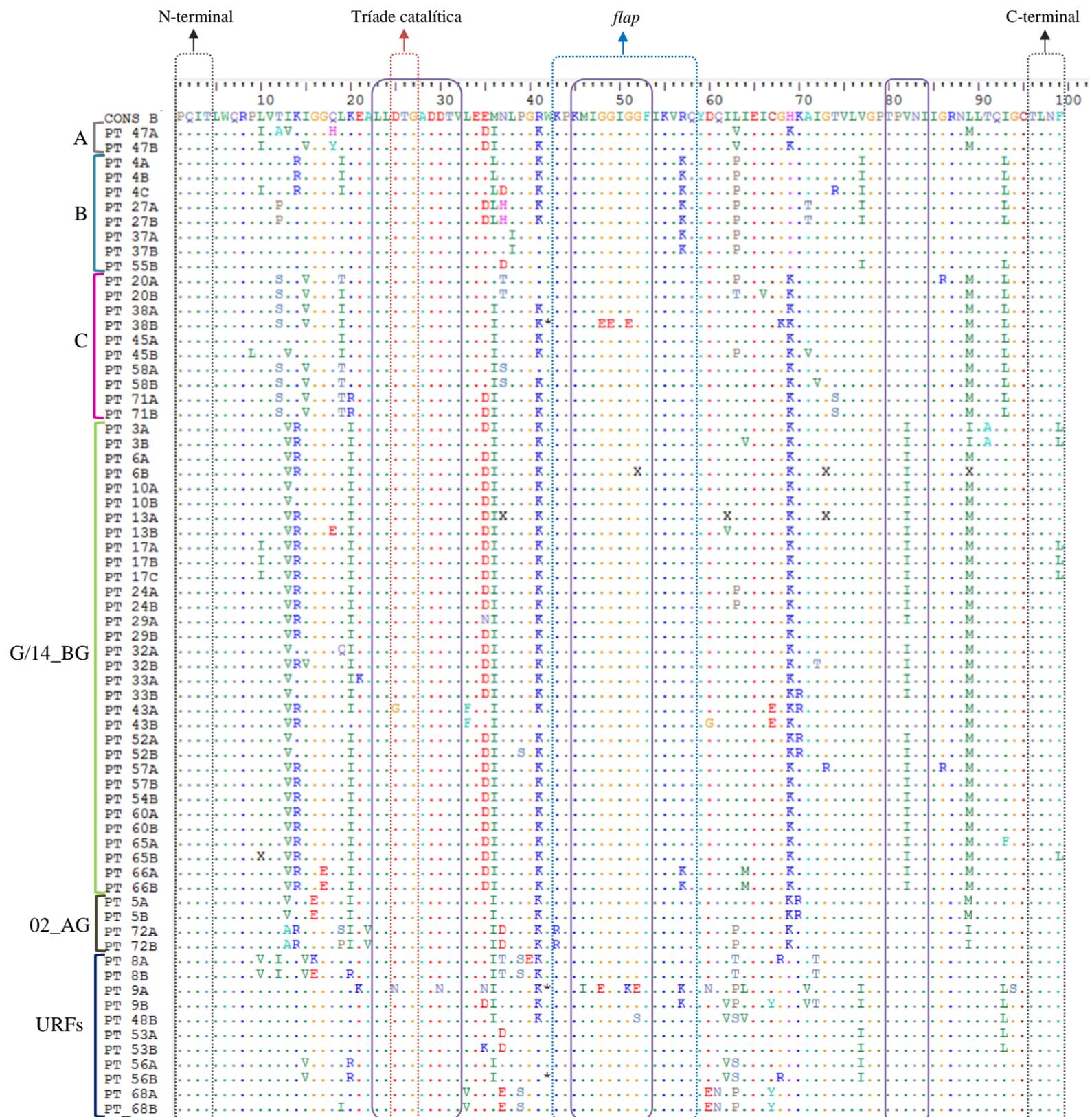


Figura 21. Alinhamento múltiplo das 67 seqüências aminoácídicas putativas da protease das estirpes do HIV-1 em estudo. As seqüências alinhadas foram comparadas com a seqüência consenso do subtipo B (denominada Cons_B), obtida da base de dados *Los Alamos HIV database* (www.hiv.lanl.gov). Os aminoácidos encontram-se representados pelo código internacional IUPAC de uma letra (ver Anexo 7). Os pontos indicam igualdade relativamente a Cons_B. A letra X indica a presença de uma base degenerada na seqüência nucleotídica. O símbolo * representa a presença de um códon de terminação precoce. Os domínios funcionais da protease estão delineados pelos retângulos. Os retângulos de cantos arredondados assinalam as posições aminoácídicas correspondentes ao centro activo da protease, nomeadamente, posições 23 a 32, 45 a 53 e 80 a 84. Os retângulos a tracejado azul e preto representam, respectivamente, a estrutura *flap* (posições 43 a 58) e os resíduos N e C terminais (1 a 4 e 96 a 99), envolvidos na dimerização.

Tabela 10. “Assinaturas genéticas” identificadas nas sequências dos subtipos B, C e G ($p < 0,01$), presentes na sequência aminoacídica putativa da protease das estirpes de HIV-1 em estudo. Os aminoácidos presentes em cada posição encontram-se representados pelo código internacional IUPAC de uma letra (Anexo 7).

“Assinaturas genéticas”			
Subtipos	Vírus selvagem*	Posição aminoacídica	Vírus mutado
B	M	36	L
	R	57	K
	L	63	P
	V	77	I
C	T	12	S
	I	15	V
	L	19	I
			T
I	93	L	
G	I	13	V
	K	14	R
	K	20	I
	E	35	D
	M	36	I
	R	41	K
	H	69	K
	V	82	I
	L	89	M
	F	99	L

* neste caso, corresponde a uma sequência consenso do subtipo B (cons_B), com origem na base de dados *Los Alamos HIV database* (www.hiv.lanl.gov).

Como é evidente pela observação do alinhamento múltiplo da Figura 21, verificou-se a presença frequente de polimorfismos genéticos nas sequências em estudo, quando comparadas com a sequência consenso do subtipo B. Foi igualmente analisada a distribuição destes polimorfismos ao longo da sequência aminoacídica da protease. As posições correspondentes aos domínios funcionais da enzima, nomeadamente, o centro activo (posições 23 a 32, 45 a 53 e 80 a 84), incluindo a tríade catalítica (D²⁵T²⁶G²⁷), assim como o *flap* (posições 43 a 58), são, em geral, bastante conservadas (Weber & Agniswamy, 2009; Ali et al., 2010; Konvalinka et al., 2015). O mesmo é válido para os resíduos, nas extremidades N e C do polipéptido (posições 1 a 4 e 96 a 99,

respectivamente), envolvidos na dimerização da forma activa da protease. Verificou-se assim que a grande maioria dos polimorfismos genéticos encontrados ocorre fora dos domínios funcionais da protease. Ainda assim, foram identificados alguns polimorfismos nestas regiões, nomeadamente, as “assinaturas genéticas” R57K, para o subtipo B, e V82I e F99L, para o subtipo G (Tabela 10). Foram ainda detectadas outras mutações, muito menos frequentes, designadamente, D25G (PT_43A), na tríade catalítica, K43R (PT_72A e PT_72B), na região do *flap*, e, por último, G52S (PT_6B, PT_48B), na região do *flap*, mas também a afectar o centro activo. É importante salientar que a mutação D25G é bastante invulgar, pois afecta a tríade catalítica da protease, sendo esta uma zona altamente conservada nas retropepsinas (Swanstrom & Wills, 1997; Turner & Summers, 1999; Mager, 2001). Também digno de nota, a presença de um maior número de polimorfismos nos domínios funcionais das sequências PT_9A e PT_38B, relativamente às restantes (D25N, D30N, M46I, G48E, G51K e G52E, para PT_9A, e G48E, G49E e G51E, para PT_38B). No entanto, deve ter-se em consideração que em ambas as sequências foram detectados fenómenos de hipermutação, estatisticamente significativos, por acção das APOBEC (Rose & Korber, 2000).

Por fim, é ainda importante referenciar que foram identificados codões de terminação precoces (ou codões *stop*) em três sequências em estudo (PT_9A, PT_38B e PT_56B), duas das quais com indícios de hipermutação (símbolo *, na Figura 21). Não foram observadas deleções nem inserções em qualquer uma das sequências, o que resultou na ausência de mutações de *frameshift*. Foi igualmente detectada a presença de *quasiespecies* em seis posições (L10, N37, G52, I62, G73 e L89), distribuídas por três sequências (PT_6B, PT_13A e PT_65B) (símbolo X, na Figura 21).

Com o intuito de se fornecer uma representação simplificada do grau de conservação relativo de cada posição aminoacídica no alinhamento múltiplo construído para a protease do HIV-1 (Figura 21), foi, por fim, construído um *weblogo*, com recurso à ferramenta bioinformática *AnalyzeAlign* (ver secção 2.10.2.). Esta ferramenta permite obter uma representação gráfica de uma sequência aminoacídica, artificial, construída a partir de um alinhamento múltiplo de sequências. Nesta sequência, cada aminoácido é simbolicamente representado através do código IUPAC de uma letra (Anexo 7), sendo o respectivo tamanho, em cada posição, proporcional à frequência na qual é encontrado

no conjunto de sequências em análise. Esta representação gráfica, permite, deste modo, visualizar a frequência relativa de cada aminoácido em cada posição de um modo muito imediato e, por sua vez, inferir quais as zonas mais conservadas da sequência. Por outro lado, ao utilizar um código de cores que representa o grau de hidrofiliidade/hidrofobicidade de cada aminoácido, permite igualmente verificar se as substituições encontradas são, ou não, de natureza conservativa. Ao observar a Figura 22, que apresenta o *weblogo* obtido a partir do alinhamento múltiplo das 67 sequências aminoacídicas putativas em estudo, é fácil verificar que os aminoácidos presentes nos domínios funcionais do centro activo (23 a 32, 45 a 53 e 80 a 84), incluindo a tríade catalítica (25 a 27), *flap* (43 a 58) e zona de dimerização da protease (1-4 e 96-99) permanecem, grosso modo, muito conservados (este facto é evidenciado pela presença de um aminoácido cuja probabilidade associada é igual ou próxima de 1,0). Por outro lado, nas posições mais polimórficas da sequência, as substituições evidenciadas são, na generalidade, de natureza conservativa.

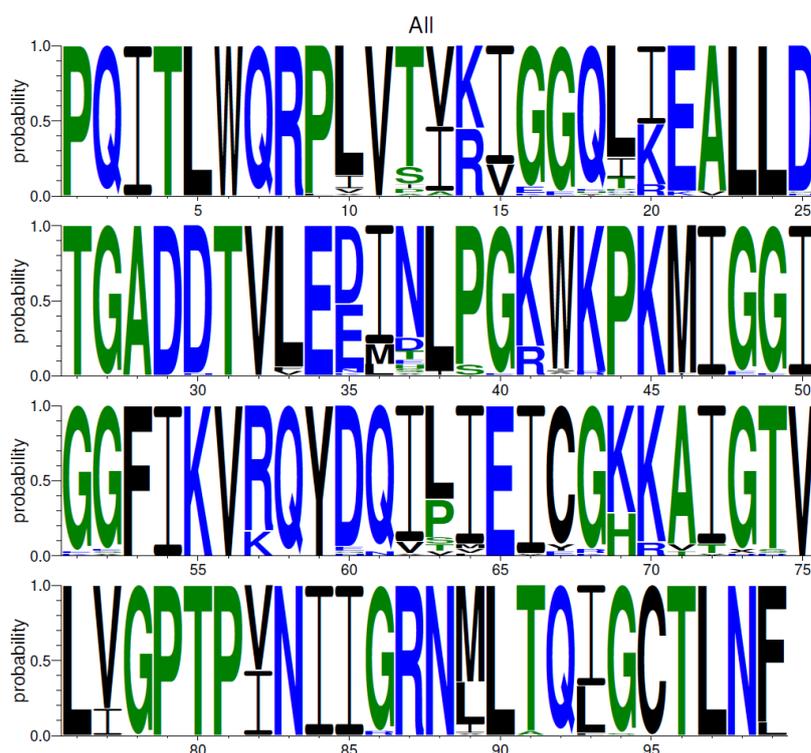


Figura 22. Representação gráfica da sequência *weblogo* das 67 sequências de aminoácidos putativas da protease do HIV-1 em estudo. A sequência é constituída por 99 posições, representando a probabilidade de cada aminoácido estar presente numa determinada posição. Os aminoácidos estão representados pelo código IUPAC de uma letra, com uma cor de acordo com o seu grau de hidrofobicidade, sendo atribuída a cor azul (hidrofílico), verde (neutro) ou preta (hidrofóbico).

4. Discussão e Conclusões

A pandemia do HIV-1 tem sido caracterizada por um aumento constante da diversidade genética viral, com aparecimento, ao longo do tempo, de diversos subtipos, sub-subtipos, CRFs e URFs, estas últimas descritas em número crescente. Esta extensa diversidade está relacionada com algumas características do próprio vírus, e respectivo ciclo replicativo, tais como, elevada taxa de replicação *in vivo*, com um tempo médio estimado de semi-vida dos viriões muito reduzido, elevada taxa de mutação durante o processo de transcrição reversa, potenciada pela ausência de *proofreading*, e elevada frequência de recombinação intermolecular (Cortez & Maldarelli, 2011; Santoro & Perno, 2013). Sobre este enorme potencial para a diversificação, actuam ainda as forças selectivas associadas à resposta imunitária do hospedeiro e, quando exista, à terapia anti-retroviral. Outros factores, como os padrões de migração da população humana, têm também contribuído para o aumento da diversidade genética nesta pandemia (Vermund & Leigh-Brown, 2012). Como tal, é assim essencial manter uma vigilância constante dos padrões epidemiológicos da infecção pelo HIV em todo o mundo. A compreensão da distribuição geográfica das diferentes formas genéticas do vírus, e respectiva taxa de evolução e disseminação no espaço e no tempo, revela-se fundamental para a definição de estratégias de prevenção, diagnóstico e tratamento.

As opções de tratamento para a infecção pelo HIV têm aumentado na última década, continuando limitadas a cinco classes terapêuticas (Skar et al., 2011; Looney et al., 2015). A terapia anti-retroviral de alta eficácia (HAART) contribuiu para uma diminuição acentuada da taxa de mortalidade e morbidade, tornando a infecção pelo HIV numa doença potencialmente crónica. Contudo, o aparecimento de estirpes virais resistentes à terapia anti-retroviral é uma das principais limitações dos regimes terapêuticos actuais (Menéndez-Arias, 2013; Santoro & Perno, 2013). A existência de mutações que conferem resistência aos anti-retrovirais em vírus que infectam mulheres grávidas pode ter um impacto significativo na eficácia da prevenção da transmissão mãe-filho, levando à transmissão de vírus resistentes para o feto/recém-nascido (Antunes et al., 2015). Nos últimos anos, observou-se uma evolução nas medidas profilácticas para a diminuição da transmissão vertical do HIV, adicionando-se os inibidores da protease a regimes terapêuticos que já incluíam inibidores da transcriptase reversa. Assim, pretendeu-se com este estudo caracterizar a diversidade genética da região codificante da protease no gene *pol* do HIV-1 e avaliar a presença de mutações

de resistência associadas aos inibidores da protease, assim como, de outros polimorfismos genéticos, em mulheres múltiparas infectadas pelo HIV-1. Este trabalho, baseado em remanescentes de material biológico (DNA extraído de células mononucleadas de sangue periférico), sem qualquer outra utilidade, consistiu num estudo retrospectivo, anónimo e não relacionável, no qual foram analisadas amostras colhidas após o parto, entre os anos de 1999 e 2008.

Deste modo, foi realizada a amplificação por *nested* PCR dupla da região codificante da protease a partir de 70 amostras de DNA proviral com origem em 34 mulheres infectadas pelo HIV-1, na sua grande maioria, mães de duas crianças. O processo de amplificação não foi bem sucedido em apenas três das amostras em estudo, podendo este facto estar associado à presença de sequências muito divergentes, impossibilitando a hibridação dos *primers*, ou a níveis de DNA proviral muito baixos, devido à terapia anti-retroviral implementada para prevenção da transmissão vertical. A degradação do DNA proviral foi igualmente considerada como uma das causas possíveis, dada a antiguidade das amostras. No entanto, após análise da qualidade do DNA, por electroforese em gel de agarose (secção 2.3.), este factor não demonstrou estar directamente relacionado com o insucesso no processo de amplificação (dados não mostrados).

O sucesso de amplificação do protocolo implementado foi quantificado em cerca de 96% (67/70). Todos os produtos de amplificação obtidos apresentaram um fragmento de DNA com cerca de 460 pb, com maior ou menor especificidade, sendo estes, após purificação, submetidos a sequenciação nucleotídica. Este procedimento foi realizado no sentido directo e reverso, reduzindo a probabilidade de erro que está associada à própria técnica e aumentando assim a fiabilidade dos resultados obtidos, factor particularmente relevante quando se pretende efectuar a pesquisa de mutações. Após a edição dos cromatogramas, a identidade das sequências nucleotídicas foi confirmada por pesquisa de regiões de semelhança local entre sequências com recurso ao programa *BLAST*, sendo, desde logo, registadas as sequências presentes na base de dados que apresentavam maior similaridade com as sequências de HIV-1 em estudo. De seguida, procedeu-se à genotipagem rápida destas sequências, de modo a obter um conhecimento prévio da diversidade genética presente na amostra populacional estudada. Para isso, recorreu-se ao programa *REGA HIV-1 & 2 Automated Subtyping Tool* (versão 2.0) que,

ao contrário de outros programas de genotipagem automática disponíveis *online*, apresenta os seus resultados baseados em análise filogenética, tendo assim em consideração a relação evolutiva entre as sequências, tornando a sua análise mais rigorosa (Abecasis et al., 2010). Deste modo, verificou-se que os genótipos mais frequentes eram os subtipos G, B e C (Anexo 4), mas com um número significativo de sequências “não genotipáveis”, sugerindo, desde logo, a presença de formas genéticas recombinantes ou muito divergentes. Como esta análise apresenta algumas limitações, a classificação genotípica definitiva das sequências em estudo foi realizada através da construção de árvores filogenéticas, aplicando métodos manuais, ou seja, não automatizados. Embora estes métodos sejam mais laboriosos e demorados, permitem seleccionar as sequências de referência utilizadas, bem como os modelos e algoritmos matemáticos aplicados, tornando os resultados obtidos mais fiáveis. A construção das árvores filogenéticas foi realizada com recurso ao programa *MEGA* (versão 5) (Tamura et al., 2011), no qual foi aplicado o método de *Neighbor-Joining* (Saitou & Nei, 1987) e as distâncias genéticas calculadas pelo modelo de dois parâmetros de Kimura (Kimura, 1980), sendo os agrupamentos formados considerados consistentes do ponto de vista filogenético quando suportados por valores de *bootstrap* com um valor mínimo de 70%.

Numa primeira abordagem à árvore filogenética da Figura 17.A, é, desde logo, possível constatar que para alguns agrupamentos formados não foram conseguidos valores de *bootstrap* superiores a 70%, o que limita a robustez da inferência filogenética realizada. Este facto será, presumivelmente, inerente ao tamanho algo reduzido da região analisada (321 nucleótidos), assim como à sua relativa conservação genética, no contexto do vírus estudado, traduzindo-se numa variabilidade mais reduzida e, por conseguinte, na obtenção de um sinal filogenético mais baixo. Tendo este facto em consideração, prosseguiu-se então para uma análise mais detalhada dos resultados obtidos.

O grupo monofilético G/14_BG (Figura 17.A), suportado por um valor de *bootstrap* significativo de 86%, inclui sequências de referência do subtipo G (ditas “puras”) e da CRF14_BG, sem subestruturação evidente, para além da geográfica, razão pela qual lhe foi atribuída aquela designação. Tendo em consideração a região do genoma estudada e a estrutura mosaico da CRF14_BG (Anexo 6), na qual a região codificante da protease é classificada no genótipo G, não é possível determinar com

precisão as relações filogenéticas entre as sequências em estudo e esta forma recombinante circulante. Para isso, seria necessário genotipar mais regiões do genoma viral, idealmente, obter sequências de genoma completo ou quase completo. Neste agrupamento monofilético verificou-se ainda a presença de um elevado número de sequências de referência com origem em Portugal e Espanha, indo ao encontro da história epidemiológica da CRF14_BG na Península Ibérica. Esta forma recombinante foi descrita, pela primeira vez, no ano de 2002, em Espanha (Galiza) (Delgado et al., 2002), pressupondo-se, no entanto, que terá tido origem em Portugal no início dos anos 90 do século passado, logo após a expansão da epidemia do HIV (Bártolo et al., 2011). Nessa altura, Portugal teria já uma elevada prevalência de subtipos não-B, em associação com população imigrante proveniente de África e, conseqüentemente, a co-circulação dos subtipos B e G terá estado na origem da CRF14_BG (Esteves et al., 2002, 2003). A sua disseminação para Espanha ocorreu, muito provavelmente, pela frequente mobilidade entre os dois países, aqui associada a indivíduos toxicodependentes por via endovenosa (Delgado et al., 2002; Esteves et al., 2003). Esta forma recombinante propagou-se rapidamente para outros países da Europa, aparentemente, sempre em associação com a prática da partilha de material de injeção associada à toxicodependência (Bártolo et al., 2011; Paraskevis et al., 2011, 2013, 2015; Niculescu et al., 2015). No entanto, nos últimos anos, a sua prevalência tem vindo a diminuir (Palma et al., 2007; Carvalho et al., 2015), sem que haja uma explicação clara para tal. Algumas causas, que podem estar associadas, são a sua elevada propensão para eventos de recombinação com outros subtipos/formas recombinantes (Niculescu et al., 2015; Fernández-García et al., 2016) ou, até mesmo, a sua elevada patogenicidade (Bártolo et al., 2011). Os indivíduos infectados progredem mais rapidamente para sida, o que está correlacionado com o facto de ser esta a forma genética do HIV-1 que apresenta um tropismo mais frequente para o co-receptor CXCR4.

Relativamente aos dois agrupamentos que incluem, respectivamente, sequências de referência da CRF02_AG e do genótipo A/CRF45_cpx (Figura 17.A), verificou-se que não foram obtidos valores de *bootstrap* estatisticamente significativos, não permitindo uma genotipagem robusta das sequências em estudo. Não obstante, para estes grupos, a topologia da árvore sugere fortemente que as sequências PT_5A, PT_5B, PT_72A e PT_72B, bem como as sequências PT_47A e PT_47B, sejam

filogeneticamente muito próximas da forma recombinante CRF02_AG e do subtipo A (A/CRF45_cpx), respectivamente. Relativamente às duas últimas, estas formam um subgrupo com sequências de referência da CRF45_cpx, intercalando a radiação A. Procedeu-se à análise dos pontos de recombinação da CRF45_cpx, o que permitiu constatar que, para a região codificante da protease, a estrutura mosaico desta forma recombinante é A/U (Anexo 6). Considerando o valor não significativo do *bootstrap* obtido, considerou-se, no entanto, como mais provável que ambas as sequências sejam filogeneticamente mais próximas do subtipo A “puro” e não propriamente daquela forma recombinante circulante complexa.

No que se refere às sequências genotipadas como subtipo C, com base no seu agrupamento com sequências de referência deste genótipo, suportado por um valor de *bootstrap* de 84% (Figura 17.A), foi possível constatar que correspondiam a mulheres cuja origem geográfica era Angola ou Moçambique (Figura 18 e Anexo 2). Este facto não é, de todo, inesperado, visto o subtipo C ser um dos genótipos predominantes em Angola (Bártolo et al., 2005; Castelbranco et al., 2010; Afonso et al., 2012), sendo quase exclusivo em Moçambique (Parreira et al., 2006; Abreu et al., 2008; Bila et al., 2015). Assim sendo, é possível que estas mulheres tenham adquirido a infecção pelo HIV-1 antes de emigrar para Portugal.

Finalmente, tendo em consideração as sequências em estudo incluídas no agrupamento com as sequências de referência do subtipo B, é importante salientar que este não apresenta um valor de *bootstrap* significativo (<70%) (Figura 17.A), não permitindo assim uma classificação genotípica muito consistente. No entanto, a topologia da árvore filogenética aponta para uma evidente similaridade das sequências em estudo com o subtipo B. Um dos factores que poderá estar na origem deste valor de *bootstrap* não significativo é a presença de estirpes virais geneticamente divergentes, relativamente às sequências de referência utilizadas. Num estudo anterior, foi demonstrada a presença de três subpopulações genéticas distintas do subtipo B em circulação em Portugal (Esteves et al., 2003), formando agrupamentos distintos e significativos na árvore filogenética construída. Duas dessas subpopulações são filogeneticamente relacionadas, respectivamente, com a CRF14_BG e com uma forma recombinante recentemente descrita, a CRF73_BG (Fernández-García et al., 2016), a qual inclui, para além das estirpes portuguesas, vírus circulantes em Espanha e também

uma forma recombinante BG descrita num toxicómano alemão, epidemiologicamente relacionado com Portugal (Harris et al., 2005). Assim sendo, no nosso caso, a presença de estirpes muito diversas do subtipo B pode representar uma das causas para a obtenção de um sinal filogenético baixo no agrupamento formado, devendo ainda considerar-se o tamanho relativamente reduzido (321 nucleótidos) do fragmento estudado. Na tentativa de clarificar este assunto, poderia construir-se uma nova árvore filogenética, englobando um maior número de sequências de referência do subtipo B, incluindo mais sequências de origem portuguesa, representativas de toda a diversidade circulante. Esta segunda árvore filogenética permitiria, muito provavelmente, aumentar o grau de certeza da inferência filogenética.

Na árvore filogenética da Figura 17.A, verificou-se também que estavam presentes sequências que não estavam incluídas em nenhum grupo monofilético consistente e/ou eram suportadas por ramos muito longos, sugerindo elevada divergência genética. Considerando isto, para estas sequências, foi realizada a pesquisa de eventos de recombinação, assim como, de fenómenos de hipermutação (Rose & Korber, 2000). Primeiramente, foi efectuada a pesquisa de possíveis eventos de recombinação, utilizando o programa *SimPlot* (v.3.5.1), cuja presença foi confirmada em onze das sequências em estudo, mas com graus variáveis de certeza. Como descrito anteriormente, a análise de *bootscanning* demonstrou a presença de sequências apresentando claros indícios de recombinação intergenotípica (PT_8A, PT_8B, PT_48B, PT_53A e PT_53B), mas também de sequências para as quais a atribuição do estatuto recombinante não foi feita sem reservas (PT_9A, PT_9B, PT_56A, PT_56B, PT_68A e PT_68B). As sequências PT_8A/PT_8B, PT_48B e PT_53A/PT_53B apresentam, respectivamente, as estruturas mosaico B/C/B, B/K/B e G/B (Figura 17.B). Os padrões identificados foram comparados com o genoma mosaico das CRFs já descritas (região da protease), não se verificando, no entanto, nenhuma analogia, pelo que estas sequências foram consideradas formas recombinantes únicas. Numa primeira abordagem, as sequências PT_8A e PT_8B, por exemplo, aparentavam ter uma estrutura mosaico semelhante à CRF64_BC (Hsi et al., 2014), contudo os pontos de recombinação não foram correspondentes (dados não mostrados). Por sua vez, nas sequências PT_9A/PT_9B, PT_56A/PT_56B e PT_68A/PT_68B foi descrita a estrutura B/U/B. Como se pode verificar na Figura 17.B, estas sequências apresentam elevada

semelhança genética com o subtipo B, na maior parte da sua extensão, possuindo, no entanto, uma região central onde não foi possível a atribuição de um genótipo. Tendo em conta este resultado, e o seu posicionamento na árvore filogenética, estas sequências foram consideradas como prováveis formas recombinantes únicas, não se podendo, no entanto, eliminar inteiramente a hipótese de corresponderem a formas muito divergentes do subtipo B. Ao analisar a estrutura mosaico das formas recombinantes únicas acima mencionadas, todas apresentam segmentos do subtipo B, o que explicará a sua posição relativa na árvore filogenética da Figura 17.A. Finalmente, as sequências PT_38A e PT_38B foram classificadas como C divergentes, dado o padrão de *bootscannig* encontrado (Figura 17.B). Ambas apresentam uma evidente homologia com o subtipo C, já expressa no respectivo posicionamento na árvore filogenética, excepto para uma pequena região intermédia, para a qual o sinal filogenético é perdido relativamente às sequências de referência utilizadas. Dado isto, e não havendo claros indícios de recombinação, estas sequências foram classificadas como formas geneticamente muito divergentes.

De seguida, procedeu-se à pesquisa de fenómenos de hipermutação, por acção das APOBEC (do inglês, *apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like*). A família APOBEC inclui as desaminases de citidina celulares APOBEC3, as quais apresentam uma potente actividade catalítica contra diversos retrovírus (Chiu & Greene, 2008). No Homem, descobriu-se que a APOBEC3G (hA3G) e a APOBEC3F (hA3F) têm um forte efeito inibidor na replicação do HIV-1, embora a actividade desta última seja um pouco mais fraca (Bishop et al., 2004; Henriot et al., 2009). Estas enzimas, expressas nos linfócitos T CD4⁺, macrófagos e monócitos, células alvo do HIV-1, inserem-se no interior dos viriões formados, actuando no ciclo replicativo seguinte. Actuam na cadeia de DNA de polaridade negativa, na fase de transcrição reversa, convertendo resíduos de citosina em uracilo. Assim, na formação da cadeia de DNA de polaridade positiva há a substituição de um resíduo de guanina por adenina, com introdução de uma mutação. Estas mutações podem levar à introdução de codões de terminação, tornando o vírus defectivo e, por conseguinte, não replicativo (Mangeat et al., 2004; Chiu & Greene, 2008; Henriot et al., 2009). Por norma, este fenómeno é inibido pela proteína Vif do HIV-1, que actua sobre hA3G e hA3F diminuindo a sua concentração celular e impedindo, assim, a sua incorporação na partícula viral. Alguns

dos mecanismos associados a esta supressão são a ubiquitinação das desaminases e, conseqüentemente, a sua degradação via proteassoma e/ou a ligação de Vif ao mRNA da hA3G inibindo a sua tradução (Yu et al., 2003; Mercenne et al., 2010). A pesquisa deste fenômeno foi realizada com recurso ao programa *Hypermot 2.0*, como referido no subcapítulo 2.9.4. Verificou-se a presença de fenômenos de hipermutação, estatisticamente significativos ($p < 0,05$), nas sequências PT_9A e PT_38B. A sua detecção justifica assim os ramos muito longos, que suportam ambas as sequências, na árvore filogenética (Figura 17.A), particularmente, no caso da sequência PT_9A. Por outro lado, este fenômeno poderá também concorrer para o padrão particular de *bootscanning* obtido para a sequência PT_38B, já discutido anteriormente. Por fim, a presença de hiperadenilação nestas sequências pressupõe que o DNA proviral amplificado, nestes dois casos, corresponda a formas virais defectivas, não infecciosas.

Adicionalmente, e a fim de ajudar a interpretar posicionamentos ambíguos de algumas sequências na árvore filogenética da Figura 17.A (construída a partir de um alinhamento múltiplo contendo 321 nucleótidos), e, assim, melhor definir eventuais padrões de recombinação ou a existência de sequências geneticamente divergentes, foi construída uma árvore filogenética a partir de sequências concatenadas, compreendendo as regiões codificantes da protease e da transcriptase reversa (parcial) (Figura 19.A). É importante salientar que, devido a um maior insucesso de amplificação para a região da transcriptase reversa, nesta árvore filogenética estão ausentes 14 amostras (referenciadas no subcapítulo 3.4.2.1.), para as quais se obteve apenas sequência para a protease. Assim, procedeu-se à construção da árvore filogenética contendo as sequências concatenadas ($n=53$), utilizando os mesmos algoritmos e modelos matemáticos aplicados para a árvore filogenética anterior. Após a sua obtenção foi, desde logo, possível observar que os grupos monofiléticos formados para os diferentes subtipos eram filogeneticamente mais consistentes, estando suportados por valores de *bootstrap* significativos ($\geq 70\%$). Este resultado estará relacionado com o maior tamanho das sequências concatenadas (838 nucleótidos), permitindo, assim, e *a priori*, a obtenção de uma inferência filogenética mais robusta. Na generalidade, os resultados obtidos vieram corroborar os anteriores, com algumas exceções. A posição relativa das sequências PT_8A, PT_8B, PT_56A e PT_68B, por exemplo, foi distinta entre as duas árvores filogenéticas. Na árvore filogenética concatenada (Figura 19.A), estas emparelham no

grupo monofilético que contém as sequências de referência do subtipo B, contrariamente ao observado na árvore relativa à região da protease. Esta discrepância deve-se ao facto destas sequências conterem maior homologia para o subtipo B na porção codificante da transcriptase reversa do que no segmento estudado anteriormente (protease). Por sua vez, e relativamente às sequências PT_47A e PT_47B, previamente agrupadas com sequências de referência da CRF45_cpx e do subtipo A, há também uma clarificação de posicionamento, uma vez que incluem agora um grupo monofilético significativo, constituído unicamente por sequências de referência do subtipo A (Figura 19.A). Portanto, estes resultados permitem confirmar a presença de uma maior homologia com o subtipo A para ambas as sequências, tal como tinha sido sugerido anteriormente, em particular com o sub-subtipo A3. Por fim, as sequências PT_48B e PT_53B apresentam uma posição ambígua em ambas as árvores filogenéticas construídas, sugerindo a presença de uma elevada divergência genética. Para ambas foi realizada uma análise de *bootscanning*, contudo, apenas na sequência PT_53B foram obtidos resultados consistentes. Para esta, a estrutura mosaico identificada foi G/B/G/B (Figura 19.B), vindo confirmar a sua classificação anterior como forma recombinante única. Para este caso, é importante ter em conta que as sequências concatenadas estudadas não são contínuas e, como tal, os pontos de recombinação na Figura 19.B não serão inteiramente representativos. Quanto à sequência PT_48B, é uma sequência muito interessante, do ponto de vista filogenético, aparentada com os genótipos B e D (de facto, sub-subtipos de um mesmo genótipo), mas com um posicionamento ambíguo, em ambas as árvores construídas. A total caracterização desta estirpe de HIV-1 passaria pela obtenção de um genoma completo ou quase completo. A sua origem geográfica em Angola, país com uma das epidemias de HIV-1 mais diversificadas, do ponto de vista genético (Abecasis et al., 2005; Bártolo et al., 2005, 2009; Castelbranco et al., 2010; Afonso et al., 2012), reforça o interesse na sua caracterização mais aprofundada.

Tendo em consideração os resultados obtidos, é possível constatar a presença de uma elevada diversidade genética do HIV-1 na amostra populacional estudada de mulheres múltiparas. De uma forma geral, os genótipos mais frequentes são os subtipos G (47,8%), C (14,9%) e B (11,9%), destacando-se ainda um elevado número de formas recombinantes únicas (16,4%). As restantes formas genéticas encontradas foram o subtipo A (3,0%) e a CRF02_AG (6,0%). Estes dados corroboram os resultados obtidos

em estudos anteriores, revelando uma elevada prevalência de subtipos não-B em Portugal (Esteves et al., 2002; Palma et al., 2007; Pereira et al., 2014; Carvalho et al., 2015). Na Europa, o subtipo B continua a ser predominante (Abecasis et al., 2013), embora, nos últimos anos, a prevalência de subtipos não-B tenha vindo a aumentar em diversos países, como, França (Brand et al., 2014), Itália (Ciccozzi et al., 2012), Suíça (von Wyl et al., 2011) e Suécia (Neogi et al., 2014), entre outros. Alguns estudos sugerem que este aumento se deve à presença de um elevado número de imigrantes, provenientes de países endémicos, cuja infecção pelo HIV ocorreu no país de origem (Holguín et al., 2008; von Wyl et al., 2011; Abecasis et al., 2013). Relativamente a Portugal, pressupõe-se que a presença de comunidades provenientes de África, América do Sul, Europa Ocidental e Europa do Leste (Departamento de Doenças Infecciosas do INSA & Programa Nacional para a Infecção VIH/SIDA, 2015) contribuem, muito provavelmente, para esta elevada heterogeneidade genética. Para além disso, uma elevada taxa de emigração no país, entre outros factores epidemiológicos e comportamentais, concorrerá igualmente para uma maior diversificação das formas genéticas virais circulantes. Além do subtipo B, relativamente frequente, Portugal, contrariamente aos restantes países europeus, tem também uma elevada prevalência do subtipo G, apresentando um padrão epidemiológico único (Esteves et al., 2002, 2003; Palma et al., 2007; Abecasis et al., 2013). No presente estudo, observou-se uma predominância do subtipo G (47,8%), assim como uma distribuição análoga entre os subtipos B e C na população estudada. No entanto, é importante realçar que 32,3% das mulheres incluídas no estudo eram de origem africana, sendo que a infecção pode ter ocorrido no país de origem, contribuindo para um aumento da prevalência de subtipos não-B. Um outro factor que poderá ter levado a um decréscimo na prevalência do subtipo B, comparativamente a outros estudos, é o facto da amostra populacional estudada ser constituída somente por mulheres. Num estudo realizado na Europa, verificou-se que a prevalência de subtipos não-B era superior em indivíduos heterossexuais, o que se reflecte numa maior proporção de mulheres infectadas com estas formas genéticas (Abecasis et al., 2013).

De seguida, procedeu-se à pesquisa de mutações associadas a resistência aos inibidores da protease, assim como, de outros polimorfismos genéticos, recorrendo ao programa *HIVdb*, implementado através do *Genotypic Resistance Interpretation*

Algorithm da *Stanford University HIV Drug Resistance Database* (versão 7.0). Este programa efectua uma comparação entre a sequência aminoacídica em estudo e uma sequência consenso do subtipo B (Stanford University, 2014). Antes de iniciar a análise dos resultados deve ter-se em consideração que, no presente estudo, as sequências foram obtidas a partir de DNA proviral, extraído de células mononucleadas do sangue periférico. O DNA proviral encontra-se integrado no genoma da célula hospedeira e, como tal, pode encontrar-se latente na célula, sem transcrição activa, não associado a vírus replicativos. Deste modo, o DNA proviral sequenciado pode não ser totalmente representativo da população viral presente no plasma. Esta é a razão pela qual o RNA viral, obtido a partir de plasma, é considerado o material biológico de eleição para a pesquisa de mutações associadas a resistência aos anti-retrovirais. Além disso, pressupõe-se que em caso de “falência virológica” numa fase inicial, a detecção de mutações associadas a resistência tenha maior sensibilidade quando realizada a partir do RNA viral (Turriziani et al., 2010). No entanto, a identificação de mutações a partir de DNA proviral pode apresentar algumas vantagens. Algumas estirpes resistentes podem permanecer latentes nos linfócitos T CD4⁺ durante longos períodos de tempo e, sob determinadas circunstâncias, reemergir e afectar a eficácia da terapia anti-retroviral (Turriziani et al., 2007, 2010; Kabamba-Mukadi et al., 2010). Aliás, num estudo anterior foi demonstrado que a identificação de mutações no DNA proviral, em indivíduos com supressão viral completa, esteve associada a uma falência da terapia anti-retroviral, nos dois anos subsequentes à sua identificação (Palmisano et al., 2009). Portanto, a utilização de DNA proviral pode ser útil na detecção de estirpes resistentes emergentes quando a carga viral é indetectável, assim como, em casos de interrupção ou alteração dos regimes terapêuticos (Turriziani et al., 2007; Kabamba-Mukadi et al., 2010; Lubke et al., 2015). Alguns autores afirmam, no entanto, que as mutações associadas à diminuição de susceptibilidade aos anti-retrovirais, encontradas, quer a partir de plasma, quer de células mononucleadas do sangue periférico são concordantes (Derache et al., 2015). Estes resultados foram, igualmente, demonstrados num estudo efectuado em mulheres sob terapia anti-retroviral para prevenção da transmissão mãe-filho (Soto-Ramirez et al., 2008). Porém, é importante salientar que quer seja utilizado DNA proviral ou RNA viral, a detecção de estirpes virais minoritárias (<20%) é bastante limitada (Palmer et al., 2005; Panel on Antiretroviral Guidelines for Adults and

Adolescents, 2015). Como se viu anteriormente, num indivíduo infectado, a população viral é constituída por diversas micropopulações genéticas, evolutivamente relacionadas, mas divergentes entre si, às quais, no seu conjunto, se dá o nome de *quasiespecies*. Deste modo, numa destas *quasiespecies* podem estar presentes, em níveis muito baixos (0,5-1%), formas virais resistentes aos anti-retrovirais, clinicamente relevantes, que podem emergir sob selecção do respectivo fármaco, levando, com o tempo, a uma falência terapêutica (Li et al., 2011; Fisher et al., 2012). Embora os testes genotípicos não detectem estas estirpes minoritárias, quando baseados em técnicas convencionais de sequenciação nucleotídica, são, de facto, os mais utilizados, devido à sua rapidez de resposta e baixo custo.

Após a análise dos resultados obtidos com recurso ao programa previamente mencionado, verificou-se a presença de duas mutações principais, nomeadamente D30N e M46I, ambas na sequência PT_9A. Nesta foram também detectadas duas mutações acessórias (G48E, A71V), descritas mais à frente. A mutação D30N está presente no centro activo da protease, afectando directamente a ligação do inibidor à enzima e, conseqüentemente, promovendo resistência aos inibidores da protease (Ali et al., 2010). Esta mutação está associada a uma resistência de nível elevado ao nelfinavir, traduzindo-se numa susceptibilidade muito reduzida, ou mesmo nula, do vírus a este fármaco (Stanford University, 2014). Por sua vez, M46I é uma mutação não polimórfica, seleccionada sob o efeito de inibidores da protease, que pode reduzir a susceptibilidade ao indinavir, nelfinavir, fosamprenavir, lopinavir e atazanavir (em especial, quando presente com outras mutações) (Stanford University, 2014). Em suma, com base no programa *HIVdb*, a presença simultânea destas mutações confere uma resistência de nível elevado ao nelfinavir e uma resistência potencial de nível baixo ao atazanavir, fosamprenavir, indinavir e lopinavir. Tendo em conta que a mulher em causa (PT_9), durante a gestação, esteve sob um regime terapêutico que incluiu o indinavir (Anexo 2), os resultados sugerem que, provavelmente, a eficácia da terapia anti-retroviral não teria sido afectada. No entanto, nesta sequência nucleotídica foi detectado o fenómeno de hipermutação, assim como a presença de um codão de terminação precoce, indicando que esta subpopulação viral seria defectiva, sem capacidade replicativa, não conseguindo infectar outras células. A mesma mulher teve um segundo filho, três anos mais tarde, sob o mesmo regime terapêutico aplicado na primeira

gestação (Anexo 2), não tendo sido detectadas mutações principais na sequência analisada (PT_9B). Neste caso é evidente a presença, num único indivíduo, de uma população viral temporalmente heterogénea.

Foram ainda detectadas sete mutações acessórias (L10I/V, L33F, G48E, A71T/V e T74S) em 19 das sequências estudadas. As mutações L10I/V, as mais frequentemente encontradas, presentes em nove sequências (13,4%), correspondem a um polimorfismo normalmente seleccionado após a toma de inibidores da protease. Estas mutações podem estar associadas a uma redução na susceptibilidade ao fármaco, ou potenciar a capacidade replicativa viral, quando presentes com outras mutações (Stanford University, 2014). Segundo os dados epidemiológicos recolhidos, quatro das mulheres (PT_4, PT_8, PT_17 e PT_47) estiveram, em, pelo menos, uma das gestações, sob terapêutica com inibidores da protease, o que poderá ter levado à selecção das mutações em causa. Apenas em uma das mulheres (PT_65) foi aplicado um esquema terapêutico que não incluía inibidores da protease. A mutação L33F, detectada nas sequências PT_43A e PT_43B (3,0%), estará associada a uma resistência potencial de nível baixo ao fosamprenavir e ao tipranavir (Stanford University, 2014). No entanto, esta mulher não realizou terapia anti-retroviral em nenhuma das gestações, ocorrendo mesmo a transmissão vertical do HIV-1 para o primeiro filho (correspondente à sequência PT_43A) (Anexo 2). Geralmente, a mutação L33F surge após selecção com os fármacos darunavir, fosamprenavir, lopinavir, nelfinavir e tipranavir, contudo, não há informação se esta mulher esteve sob terapia anti-retroviral anteriormente, podendo mesmo tratar-se de um caso de transmissão de resistência primária, ocorrida para indivíduo *naive*. Por sua vez, a mutação G48E, também identificada em duas sequências (3,0%), não está associada a uma alteração fenotípica, relativamente aos inibidores da protease, sendo considerada muito pouco comum, de acordo com o programa utilizado (Stanford University, 2014). No entanto, sabe-se que outras mutações presentes nesta posição, nomeadamente, G48V/M, estão associadas a uma perda da susceptibilidade aos inibidores da protease, pois afectam a sua ligação à enzima (Weber & Agniswamy, 2009; Stanford University, 2014). Curiosamente, a mutação G48E foi apenas detectada nas sequências PT_9A e PT_38B, nas quais foi identificado o fenómeno de hipermutação (G→A). Este estará na origem da alteração do codão GGA→GAA (presente em ambas as sequências), levando à substituição da glicina (aminoácido

presente na sequência consenso) pelo ácido glutâmico. Por sua vez, as mutações A71T/V, identificadas em cinco sequências (7,5%), são polimorfismos normalmente presentes em 2 a 3% de indivíduos *naive*. Contudo, a sua prevalência aumenta em indivíduos em terapêutica com inibidores da protease (Stanford University, 2014). Relativamente a quatro das sequências onde estas mutações foram detectadas (PT_9A, PT_9B, PT_27B, PT_45B), o regime terapêutico das respectivas mulheres incluía inibidores da protease, tendo provavelmente levado à sua selecção. No entanto, não se pode descartar a hipótese de se tratar de casos resultantes de transmissão primária de resistência, dada a presença destas mutações como polimorfismos naturais em indivíduos sem terapia. Finalmente, a mutação T74S foi identificada em 3,0% das sequências estudadas (PT_71A e PT_71B), constituindo um polimorfismo associado a resistência de nível baixo ao nelfinavir, podendo afectar o efeito anti-retroviral deste fármaco (Stanford University, 2014). Dos dados epidemiológicos referentes à mulher PT_71, apenas se obteve informação do regime terapêutico instituído durante a segunda gravidez (PT_71B). Nesta gestação, a mulher realizou a toma de lopinavir potenciado com ritonavir e, portanto, pressupõe-se que a presente mutação não tenha afectado a eficácia terapêutica do fármaco.

Em suma, no presente estudo foi detectada uma baixa prevalência de mutações principais, sendo estas identificadas em apenas uma sequência. Por outro lado, verificou-se uma prevalência superior de mutações acessórias, presentes em 19 das sequências estudadas, sendo que estes resultados estão de acordo com os dados anteriormente descritos para Portugal (Parreira et al., 2004; Palma et al., 2007). Estas mutações associadas à diminuição da susceptibilidade aos anti-retrovirais inibidores da protease foram detectadas maioritariamente em subtipos não-B (16/19), não sendo revelada, no entanto, qualquer associação estatisticamente significativa entre a presença das mutações descritas e os subtipos B/não-B. Estes resultados estão em concordância com os descritos, em estudos recentes, por outros autores (Barber et al., 2012; Grossman et al., 2014).

Adicionalmente, foram utilizadas outras ferramentas disponíveis *online* para pesquisa de mutações associadas a resistência, com o objectivo de avaliar, de uma forma genérica, o grau de divergência entre a previsão realizada pelos diversos algoritmos. Para isso, recorreu-se ao painel publicado anualmente pela *International*

Antiviral Society-USA (IAS-USA) e aos algoritmos bioinformáticos *REGA* v.9.1.0 (disponível em <https://rega.kuleuven.be/cev/avd/software/rega-algorithm>) e *ANRS* v.2015.09 (disponível em <http://www.medpocket.com/>). De acordo com a IAS-USA (Wensing et al., 2015), as mutações G48E e T74S, descritas anteriormente, não são classificadas como conferindo resistência aos inibidores da protease. Por outro lado, o mesmo painel inclui diversas mutações acessórias, não consideradas pelo algoritmo aplicado no presente projecto (*Stanford HIVdb*), nomeadamente: G16E, K20R, L33V, M36I/L, D60E, I62V, L63P, I64L/M/V, H69K, V77I, V82I, L89I/M e I93L (Figura 9). Foram ainda utilizados os algoritmos *ANRS* e *REGA* que, tal como *Stanford HIVdb* fornecem, para cada sequência, um perfil fenotípico provável de acordo com as mutações identificadas. Os perfis fenotípicos obtidos foram, de uma forma geral, concordantes entre os três algoritmos, com excepção das sequências PT_8B, PT_9A, PT_56A, PT_56B, PT_71A e PT_71B (dados não mostrados). A sequência PT_9A apresenta um perfil resistente ao nelfinavir, tal como no *Stanford HIVdb*. Contudo, contrariamente a este, é considerado resistente e resistente intermédio ao indinavir, segundo o *ANRS* e o *REGA*, respectivamente. Como se pode verificar, o grau de susceptibilidade ao indinavir não é concordante entre os diversos algoritmos aplicados. Além disso, o *ANRS* considera, ainda, que esta estirpe pode apresentar uma possível resistência ao lopinavir e atazanavir. De qualquer modo, esta sequência corresponderá a um vírus defectivo, não infeccioso, como já foi referido anteriormente. Por sua vez, na sequência PT_8B foi detectado um perfil fenotípico resistente ao saquinavir (L10V + I15V + K20R) e uma possível resistência ao atazanavir (L10V + G16E) segundo o *ANRS*. Este último diverge dos algoritmos *Stanford HIVdb* e *REGA*, para os quais a estirpe foi considerada susceptível. Relativamente às sequências PT_56A e PT_56B foram identificadas as mutações I15V, K20R e I62V, que, quando em simultâneo, conferem resistência ao saquinavir, de acordo com o *ANRS*. Tal como no caso anterior, esta interpretação difere da dos restantes algoritmos. Por último, nas sequências PT_71A e PT_71B, o perfil fenotípico apresentado também demonstrou ser diferente entre os algoritmos. Segundo o *Stanford HIVdb*, a presença da mutação T74S confere uma resistência de nível baixo ao nelfinavir, contrariamente ao *REGA* e ao *ANRS*. Por outro lado, este último considera que as mutações I15V e K20R podem estar associadas a uma possível resistência ao saquinavir. Por fim, o *ANRS*, contrariamente aos restantes

algoritmos, considerou que as estirpes referentes às sequências PT_8A, PT_13A, PT_13B, PT_17A, PT_17B, PT_17C, PT_32B, PT_47B e PT_65B podem apresentar uma possível resistência ao saquinavir (presença, em simultâneo, de duas das mutações L10I/V, I15V, K20I/R, I62V). Por sua vez, nas sequências PT_68A e PT_68B foi inferida uma possível resistência ao atazanavir (L33V + D60E). Tendo em conta o que foi referido, é possível concluir que a interpretação do significado fenotípico de mutações associadas a resistência não é totalmente concordante entre as diversas ferramentas disponíveis. Como foi demonstrado, nem todas aplicam a mesma classificação a uma mesma mutação, como, por exemplo, é o caso das mutações I15V, K20R e T74S. Além disso, a maioria destes sistemas utilizam como referência de comparação o subtipo B, continuando a verificar-se dificuldades na interpretação de mutações em subtipos não-B. Este facto deve-se à presença de polimorfismos naturais nestes subtipos, os quais são associados a resistência no subtipo B. Uma das alternativas possíveis seria a realização de testes fenotípicos, para cada uma das estirpes. Estes testes apresentam, entre outras, a vantagem de uma mais fácil interpretação, contudo são muito mais dispendiosos, demorados e laboriosos.

Por outro lado, foram ainda identificados os polimorfismos genéticos presentes nas 67 sequências em estudo. O seu papel, em termos de relevância clínica, é ainda controverso (Theys et al., 2013). Tendo em conta a variabilidade genética do HIV-1, sabe-se que diferentes subtipos contêm, no seu genoma, polimorfismos genéticos e/ou “assinaturas” (polimorfismos característicos de um determinado subtipo) que podem afectar as regiões codificantes das enzimas virais, sendo estas os principais alvos da terapia anti-retroviral (Santos & Soares, 2010; Theys et al., 2013). A relevância destes polimorfismos genéticos tem sido debatida ao longo dos últimos anos. Embora a maioria não afecte, significativamente, a resposta à terapia anti-retroviral, alguns estudos indicam que podem estar associados a um aumento na resistência aos anti-retrovirais, potenciar o *fitness* viral, na presença de mutações de resistência, ou, até mesmo, aumentar a susceptibilidade do vírus a um determinado fármaco (Bandaranayake et al., 2010; Santos & Soares, 2010; Santos et al., 2012). No presente estudo, foram identificadas 53/99 posições aminoacídicas em que, pelo menos, uma das sequências analisadas divergia da sequência consenso, sendo a maioria dos polimorfismos genéticos detectados em sequências de subtipos não-B. Contudo, a

utilização como referência de uma sequência consenso do subtipo B leva a que as divergências encontradas nas sequências pertencentes a este subtipo sejam naturalmente em número inferior. Alguns destes polimorfismos foram considerados como “assinaturas genéticas” relativas aos subtipos B, C e G ($p < 0,01$) (Anexo 5). A sua prevalência foi superior nos subtipos não-B verificando-se, no entanto, a presença de quatro “assinaturas” do subtipo B (M36L, R57K, L63P e V77I). Deve ter-se em consideração que estas sequências foram comparadas com uma sequência consenso do subtipo B, como referido anteriormente, e, ainda assim, revelaram a presença de “assinaturas”, comprovando a sua divergência genética relativamente à respectiva sequência de referência. Este facto corrobora a observação anterior do agrupamento formado por sequências do subtipo B, na árvore filogenética analisada (Figura 17.A), não ser suportado por um valor de *bootstrap* significativo. Por sua vez, no subtipo C verificou-se a presença de cinco “assinaturas”, nomeadamente, T12S, I15V, L19I, L19T e I93L. Estes dados estão genericamente em concordância com o que está descrito na literatura (Kantor et al., 2005; Abreu et al., 2008; Ali et al., 2010), excepto para as “assinaturas” L19I e L19T, não associadas, nos referidos estudos, ao presente subtipo. Por último, no subtipo G foi descrito um maior número de “assinaturas”, contabilizando-se um total de dez, designadamente, I13V, K14R, K20I, E35D, M36I, R41K, H69K, V82I, L89M e F99L. Estes dados corroboram os resultados obtidos por outros autores (Holguín et al., 2002; Parreira et al., 2004; Kantor et al., 2005; Ali et al., 2010), com excepção da “assinatura” F99L. Segundo o *Stanford HIVdb*, nenhuma das “assinaturas” descritas para os três subtipos está associada a qualquer alteração fenotípica relativamente à resposta anti-retroviral. No entanto, de acordo com o painel publicado pela IAS-USA, algumas são classificadas como mutações acessórias, nomeadamente, K20I, M36I/L, L63P, H69K, V77I, V82I, L89M e I93L (Wensing et al., 2015). Estas, embora não afectem substancialmente a eficácia dos fármacos, podem potenciar a capacidade replicativa do vírus, segundo o referido painel. O impacto *in vivo* destes polimorfismos genéticos, naturalmente presentes em determinados subtipos, permanece assim ainda por esclarecer. Enquanto alguns autores descrevem estes polimorfismos como potenciadores do *fitness* viral e/ou associados a resistência aos anti-retrovirais (Holguín et al., 2002, 2006; Lisovsky et al., 2010), outros não

corroboram estas ideias (Vergne et al., 2006; Abreu et al., 2008; Champenois et al., 2008).

Ao analisar a sequência traduzida da protease (99 aminoácidos), de todas as sequências nucleotídicas em estudo, verificou-se que os seus domínios funcionais se mantinham, de uma forma geral, preservados. Com base nas Figuras 21 e 22 é possível constatar que nestas regiões, importantes para a actividade da enzima, as substituições aminoacídicas detectadas são em número muito reduzido, contrariamente a outras regiões, nas quais se observou níveis de variabilidade mais elevados. Ainda assim, e excluindo desta análise as duas sequências (PT_9A e PT_38B) para as quais foi descrita hipermutação dependente da actividade das APOBEC, foram identificados alguns polimorfismos nos domínios funcionais da protease, nomeadamente, D25G, K43R, G52S, R57K, V82I e F99L. De acordo com o *Stanford HIVdb*, nenhum destes polimorfismos está associado a resistência aos inibidores da protease. Contudo, a mutação V82I, considerada como “assinatura genética” do subtipo G no presente estudo ($p=0,000$) é classificada, por alguns autores, como uma mutação acessória associada a resistência ao atazanavir, tal como já foi referido anteriormente (Santos & Soares, 2010; Wensing et al., 2015). De entre estes polimorfismos é ainda importante destacar o D25G, detectado na sequência PT_43A (Figura 21). Neste caso, o ácido aspártico (carregado negativamente) foi substituído pela glicina (sem carga) (GAC→GGC) conduzindo, muito provavelmente, a alterações conformacionais e funcionais na protease. Como se sabe, a posição aminoacídica 25 é um componente essencial da tríade catalítica da protease e, como tal, a substituição do ácido aspártico, presente nesta posição, leva a uma perda da actividade catalítica, tornando-a inactiva (Rosé et al., 1995; Sayer et al., 2008; Capel et al., 2014). Deste modo, a subpopulação viral, na qual esta mutação foi identificada, não teria uma protease funcional. Além disso, foram identificadas outras três subpopulações virais defectivas, designadamente, PT_9A, PT_38B e PT_56B, nas quais foram identificados codões de terminação precoce, originando proteínas truncadas e, por isso, não funcionais. Como tal, não ocorrerá a dimerização da protease, essencial para o processo de maturação das novas partículas da progénie viral, conduzindo, conseqüentemente, à formação de partículas virais defectivas e não infecciosas. Por sua vez, como se viu, nas sequências PT_9A e PT_38B foram identificados fenómenos de hiperadenilação, estatisticamente significativos

($p < 0,05$). Portanto, presume-se que nenhuma das quatro subpopulações virais discriminadas tenha capacidade replicativa, reflectindo a introdução de mutações letais resultantes da acção das APOBEC, ou as próprias características replicativas do HIV-1, que surge como uma *quasiespecies* constituída por diversas estirpes geneticamente distintas, das quais algumas serão defectivas. Resumindo, 6,0% (4/67) das sequências obtidas a partir do DNA proviral integrado nas células mononucleadas do sangue periférico, corresponderiam a formas defectivas do vírus. Este facto, representa uma das desvantagens de utilizar DNA proviral como matriz para a pesquisa de mutações, pois este pode não ser totalmente representativo da população viral presente no plasma, como referido anteriormente.

Foi igualmente avaliada a eficácia dos regimes terapêuticos utilizados para prevenção da transmissão vertical do HIV-1 na amostra populacional em estudo. Nas mulheres que efectuaram terapia anti-retroviral, independentemente do regime aplicado, constatou-se que esta foi efectiva, não se verificando nenhum caso de transmissão. No entanto, 11,4% (8/70) das amostras estudadas, colhidas de seis mulheres distintas, correspondiam a gestações nas quais não se realizou terapêutica anti-retroviral, reflectindo uma elevada proporção de mulheres nas quais não foi tomada qualquer medida preventiva. Destes oito casos, verificou-se a transmissão vertical do HIV-1 para três dos recém-nascidos (Anexo 2). Estes dados são análogos aos antes descritos por outros autores em Portugal. Num estudo anterior, verificou-se que cerca de 13% das mulheres não realizaram nenhum regime profiláctico anti-retroviral durante a gravidez, todas elas devido ao desconhecimento da sua situação clínica (Palma et al., 2007). Um estudo mais recente (Piñeiro et al., 2014), realizado num grupo de mulheres grávidas infectadas pelo HIV-1, entre 1999 e 2012, no Hospital de S. João (Porto), verificou que em 24% das mulheres a infecção foi diagnosticada já na gravidez. Neste mesmo estudo, observou-se a transmissão vertical do HIV-1 em três casos, num dos quais a mãe tinha iniciado a terapia anti-retroviral no segundo trimestre da gravidez (Piñeiro et al., 2014). No presente estudo, não há informação sobre as razões pelas quais as seis mulheres não realizaram terapia anti-retroviral, podendo, tal como nos estudos acima referenciados, dever-se ao desconhecimento do seu estado clínico. Segundo o INSA (Departamento de Doenças Infecciosas do INSA & Programa Nacional para a Infecção VIH/SIDA, 2015), desde 1983 até ao final de 2014, foram notificados, em Portugal, 426 casos de infecção

pelo HIV-1 devido a transmissão vertical. Após a introdução do rastreio na gravidez e da terapia anti-retroviral, o número de casos diagnosticados tem vindo a diminuir ao longo dos anos, reflectindo uma taxa de transmissão mãe-filho inferior a 2%, tal como em outros países europeus (Warszawski et al., 2008; Prieto et al., 2012; Townsend et al., 2014; Departamento de Doenças Infecciosas do INSA & Programa Nacional para a Infecção VIH/SIDA, 2015). Os casos de transmissão vertical que continuam a verificar-se podem ser devidos, entre outros, ao desconhecimento da infecção pelo HIV-1, a um início tardio da terapia anti-retroviral e a uma carga viral elevada no parto, mesmo sob terapia (indicativo de estirpes resistentes ou má adesão à terapêutica). A transmissão vertical representa a principal causa de infecção pelo HIV-1 em crianças com idade inferior a 15 anos, sendo a sua prevenção um dos principais objectivos da UNAIDS (UNAIDS, 2014). Considerando isto, estudos como o presente são importantes para a obtenção de mais dados epidemiológicos acerca deste grupo populacional. O conhecimento da diversidade genética do HIV-1 também é fundamental, pois os diferentes genótipos podem ter implicações na progressão da doença, na eficácia da terapia anti-retroviral, no desenvolvimento de resistências aos anti-retrovirais e, mesmo, na interpretação dos testes genotípicos para a detecção de estirpes resistentes (Santos & Soares, 2010; Santoro & Perno, 2013; Langs-Barlow & Paintsil, 2014). Por outro lado, a detecção atempada de mutações associadas à diminuição da susceptibilidade aos anti-retrovirais é essencial, pois durante a gravidez pode ocorrer a selecção de estirpes virais resistentes, quer por má adesão, quer por concentrações sub-terapêuticas do fármaco, entre outros factores. Em caso de transmissão vertical do vírus, se estas mutações forem transmitidas ao recém-nascido podem ter implicações graves na eficácia e nas opções futuras da terapia anti-retroviral (de Mulder et al., 2011; Rojas Sanchez & Holguín, 2014; Bure et al., 2015).

Em resumo, o presente estudo revelou uma vasta diversidade genética entre as sequências estudadas, indo ao encontro dos resultados anteriormente descritos para Portugal (Esteves et al., 2002; Palma et al., 2007; Pereira et al., 2014; Carvalho et al., 2015). No entanto, a prevalência dos subtipos B e G não foi semelhante, destacando-se uma elevada frequência do subtipo G (47,8%) comparativamente com o subtipo B (11,9%). Este estudo evidenciou ainda uma elevada prevalência de formas recombinantes únicas (16,4%). Relativamente às mutações associadas à diminuição da

susceptibilidade aos inibidores da protease, estas foram detectadas em 28,4% (19/67) das sequências estudadas. Destas, duas (D30N e M46I) foram classificadas como mutações principais, com implicações na resposta à terapia anti-retroviral. No entanto, ambas foram detectadas numa subpopulação viral sem capacidade replicativa aparente. Por sua vez, foram identificadas sete mutações acessórias (L10I/V, L33F, G48E, A71T/V e T74S) que, embora não afectem substancialmente a resposta à terapia, podem estar associadas a resistência quando na presença de outras mutações e/ou aumentar a capacidade replicativa do vírus. A maioria destas mutações (63,6%) foi detectada em mulheres cujo regime terapêutico incluía inibidores da protease e, portanto, presume-se que as mesmas tenham sido seleccionadas sob a pressão selectiva do fármaco. Além disso, foram também identificados diversos polimorfismos genéticos, não associados a resistência, de acordo com o algoritmo utilizado (Stanford University, 2014). Alguns destes polimorfismos, naturalmente presentes em subtipos não-B, são considerados como mutações acessórias no subtipo B e o seu impacto na resposta à terapia e/ou na capacidade replicativa do vírus permanece ainda por esclarecer.

Por fim, é importante salientar que o presente estudo apresenta algumas limitações, nomeadamente, o tamanho relativamente reduzido da amostra estudada, assim como, o período no qual a amostragem foi efectuada (até 2008). Como tal, os dados obtidos podem não reflectir com precisão a situação actual em Portugal. Além disso, o método de sequenciação aplicado não permite a detecção de estirpes virais minoritárias, as quais podem conter no seu genoma mutações que conferem resistência aos anti-retrovirais. Apesar destas limitações, os resultados obtidos reforçam a importância da monitorização da infecção pelo HIV-1 nas mulheres grávidas, tendo em conta a elevada diversidade genética do vírus e as implicações que esta pode ter no desenvolvimento de mutações de resistência aos anti-retrovirais, as quais podem afectar o sucesso da prevenção da transmissão vertical do vírus.

5. Referências bibliográficas

- Abecasis, A., Paraskevis, D., Epalanga, M., Fonseca, M., Burity, F., Bartolomeu, J., Carvalho, A.P., Gomes, P., Vandamme, A.M. & Camacho, R. (2005). HIV-1 genetic variants circulation in the North of Angola. *Infection, Genetics and Evolution*. 5 (3). pp. 231–237.
- Abecasis, A.B., Wang, Y., Libin, P., Imbrechts, S., de Oliveira, T., Camacho, R.J. & Vandamme, A.-M. (2010). Comparative performance of the REGA subtyping tool version 2 versus version 1. *Infection, Genetics and Evolution*. 10 (3). pp. 380–385.
- Abecasis, A.B., Wensing, A.M.J., Paraskevis, D., Vercauteren, J., Theys, K., Van de Vijver, D.A.M.C., Albert, J., Asjö, B., Balotta, C., Beshkov, D., Camacho, R.J., Clotet, B., De Gascun, C., Giskevicius, A., Grossman, Z., Hamouda, O., Horban, A., Kolupajeva, T., Korn, K., Kostrikis, L.G., Kücherer, C., Liitsola, K., Linka, M., Nielsen, C., Otelea, D., Paredes, R., Poljak, M., Puchhammer-Stöckl, E., Schmit, J.-C., Sönnernborg, A., Stanekova, D., Stanojevic, M., Struck, D., Boucher, C.A.B. & Vandamme, A.-M. (2013). HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics. *Retrovirology*. [Online]. 10 (7). Disponível em: <http://retrovirology.biomedcentral.com/articles/10.1186/1742-4690-10-7>.
- Abreu, C.M., Brindeiro, P.A., Martins, A.N., Arruda, M.B., Bule, E., Stakteas, S., Tanuri, A. & de Moraes Brindeiro, R. (2008). Genotypic and phenotypic characterization of human immunodeficiency virus type 1 isolates circulating in pregnant women from Mozambique. *Archives of Virology*. 153 (11). pp. 2013–2017.
- Afonso, J.M., Bello, G., Guimarães, M.L., Sojka, M. & Morgado, M.G. (2012). HIV-1 genetic diversity and transmitted drug resistance mutations among patients from the North, Central and South Regions of Angola. *PLoS ONE*. 7 (8). p. e42996.
- Ahmad, N. (2011). Molecular mechanisms of HIV-1 mother-to-child transmission and infection in neonatal target cells. *Life Sciences*. 88 (21-22). pp. 980–986.
- AIDSinfo (2015). *AIDSinfo Drug Database*. [Online]. Disponível em: <https://aidsinfo.nih.gov/drugs>. [Acedido: Junho 2015].
- Ali, A., Bandaranayake, R.M., Cai, Y., King, N.M., Kolli, M., Mittal, S., Murzycki, J.F., Nalam, M.N.L., Nalivaika, E.A., Özen, A., Prabu-Jeyabalan, M.M., Thayer, K. & Schiffer, C.A. (2010). Molecular basis for drug resistance in HIV-1 protease. *Viruses*. 2 (11). pp. 2509–2535.

- Antunes, F., Zindoga, P., Gomes, P., Augusto, O., Mahumane, I., Veloso, L., Valadas, E. & Camacho, R. (2015). Development of nevirapine resistance in children exposed to the prevention of mother-to-child HIV-1 transmission programme in Maputo, Mozambique. *PLoS ONE*. 10 (7). p. e0131994.
- Bandaranayake, R.M., Kolli, M., King, N.M., Nalivaika, E.A., Heroux, A., Kakizawa, J., Sugiura, W. & Schiffer, C.A. (2010). The effect of clade-specific sequence polymorphisms on HIV-1 protease activity and inhibitor resistance pathways. *Journal of Virology*. 84 (19). pp. 9995–10003.
- Barber, T.J., Harrison, L., Asboe, D., Williams, I., Kirk, S., Gilson, R., Bansi, L., Pillay, D., Dunn, D. & UK HIV Drug Resistance Database and UK Collaborative HIV Cohort (UK CHIC) Study Steering Committees (2012). Frequency and patterns of protease gene resistance mutations in HIV-infected patients treated with lopinavir/ritonavir as their first protease inhibitor. *Journal of Antimicrobial Chemotherapy*. 67 (4). pp. 995–1000.
- Barré-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W. & Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*. 220 (4599). pp. 868–871.
- Bártolo, I., Abecasis, A.B., Borrego, P., Barroso, H., McCutchan, F., Gomes, P., Camacho, R. & Taveira, N. (2011). Origin and epidemiological history of HIV-1 CRF14_BG. *PLoS ONE*. 6 (9). p. e24130.
- Bártolo, I., Epalanga, M., Bartolomeu, J., Fonseca, M., Mendes, A., Gama, A. & Taveira, N. (2005). High genetic diversity of human immunodeficiency virus type 1 in Angola. *AIDS Research and Human Retroviruses*. 21 (4). pp. 306–310.
- Bártolo, I., Rocha, C., Bartolomeu, J., Gama, A., Marcelino, R., Fonseca, M., Mendes, A., Epalanga, M., Silva, P.C. & Taveira, N. (2009). Highly divergent subtypes and new recombinant forms prevail in the HIV/AIDS epidemic in Angola: new insights into the origins of the AIDS pandemic. *Infection, Genetics and Evolution*. 9 (4). pp. 672–682.
- Bila, D., Boullosa, L., Vubil, A., Mabunda, N., Abreu, C., Ismael, N., Jani, I. & Tanuri, A. (2015). Trends in prevalence of HIV-1 drug resistance in a public clinic in Maputo, Mozambique. *PLoS ONE*. 10 (7). p. e0130580.
- Birnboim, H.C. & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*. 7 (6). pp. 1513–1523.

- Bishop, K.N., Holmes, R.K., Sheehy, A.M., Davidson, N.O., Cho, S.-J. & Malim, M.H. (2004). Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Current Biology*. 14 (15). pp. 1392–1396.
- Brand, D., Moreau, A., Cazein, F., Lot, F., Pillonel, J., Brunet, S., Thierry, D., Le Vu, S., Plantier, J.-C., Semaille, C. & Barin, F. (2014). Characteristics of patients recently infected with HIV-1 non-B subtypes in France: a nested study within the mandatory notification system for new HIV diagnoses. *Journal of Clinical Microbiology*. 52 (11). pp. 4010–4016.
- Bure, D., Makhdoomi, M.A., Lodha, R., Prakash, S.S., Kumar, R., Parray, H.A., Singh, R., Kabra, S.K. & Luthra, K. (2015). Mutations in the reverse transcriptase and protease genes of human immunodeficiency virus-1 from antiretroviral naïve and treated pediatric patients. *Viruses*. 7 (2). pp. 590–603.
- Capel, E., Parera, M. & Martinez, M.A. (2014). Epistasis as a determinant of the HIV-1 protease's robustness to mutation. *PLoS ONE*. 9 (12). p. e116301.
- Carr, J.K., Salminen, M.O., Albert, J., Sanders-buell, E., Gotte, D., Birx, D.L. & Mccutchan, F.E. (1998). Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G intersubtype recombinants. *Virology*. 247 (1). pp. 22–31.
- Carvalho, A., Costa, P., Triunfante, V., Branca, F., Rodrigues, F., Santos, C.L., Correia-Neves, M., Saraiva, M., Lecour, H., Castro, A.G., Pedrosa, J. & Osório, N.S. (2015). Analysis of a local HIV-1 epidemic in Portugal highlights established transmission of non-B and non-G subtypes. *Journal of Clinical Microbiology*. 53 (5). pp. 1506–1514.
- Castelbranco, E., da Silva Souza, E., Cavalcanti, A., Martins, A., de Alencar, L. & Tanuri, A. (2010). Frequency of primary resistance to antiretroviral drugs and genetic variability of HIV-1 among infected pregnant women recently diagnosed in Luanda-Angola. *AIDS Research and Human Retroviruses*. 26 (12). pp. 1313–1316.
- Centers for Disease Control and Prevention (CDC) (1981a). Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men- New York City and California. *MMWR Morbidity and Mortality Weekly Report*. 30 (25). pp. 305–308.
- Centers for Disease Control and Prevention (CDC) (1981b). Pneumocystis pneumonia- Los Angeles. *MMWR Morbidity and Mortality Weekly Report*. 30 (21). pp. 250–252.

- Centers for Disease Control and Prevention (CDC) (1982). Epidemiologic notes and reports persistent, generalized lymphadenopathy among homosexual males. *MMWR Morbidity and Mortality Weekly Report*. 31 (19). pp. 249–251.
- Champenois, K., Burban-Deuffic, S., Cotte, L., André, P., Choisy, P., Ajana, F., Bocket, L. & Yazdanpanah (2008). Natural polymorphisms in HIV-1 protease: impact on effectiveness of a first-line lopinavir-containing antiretroviral therapy regimen. *Journal of Medical Virology*. 80 (11). pp. 1871–1879.
- Chiu, Y.-L. & Greene, W.C. (2008). The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annual Review of Immunology*. 26. pp. 317–353.
- Ciccozzi, M., Santoro, M.M., Giovanetti, M., Andrissi, L., Bertoli, A. & Ciotti, M. (2012). HIV-1 non-B subtypes in Italy: a growing trend. *New Microbiologica*. 35 (4). pp. 377–386.
- Clavel, F., Guétard, D., Brun-Vézinet, F., Chamaret, S., Rey, M.A. & Santos-Ferreira, M.O. (1986). Isolation of new human retrovirus from west african patients with AIDS. *Science*. 233 (4761). pp. 343–346.
- Coffin, J., Haase, A., Levy, J.A., Montaigner, L., Oroszlan, S., Teich, N., Termin, H., Toyoshima, K., Varmus, H. & Vogt, P. (1986). What to call the AIDS virus? *Nature*. 321 (6065). p. 10.
- Cooper, E.R., Charutat, M., Mofenson, L., Hanson, I.C., Pitt, J., Diaz, C., Hayani, K., Handelsman, E., Smeriglio, V., Hoff, R., Blattner, W. & Women and Infants Transmission Study Group (2002). Combination antiretroviral strategies for the treatment of pregnant HIV-1- infected women and prevention of perinatal HIV-1 transmission. *Journal of Acquired Immune Deficiency Syndromes*. 29 (5). pp. 484–494.
- Cortez, K.J. & Maldarelli, F. (2011). Clinical management of HIV drug resistance. *Viruses*. 3 (4). pp. 347–378.
- Cotte, L., Bénet, T., Vanhems, P., Brochier, C., Perpoint, T., Ferry, T. & Chidiac, C. (2014). The effect of adherence to guidelines for initial antiretroviral therapy on 1-year outcomes: a French cohort study. *BMC Infectious Diseases*. 14. p. 596.

- D'arc, M., Ayouba, A., Esteban, A., Learn, G.H., Boué, V., Liegeois, F., Etienne, L., Tagg, N., Leendertz, F.H., Boesch, C., Madinda, N.F., Robbins, M.M., Gray, M., Cournil, A., Ooms, M., Letko, M., Simon, V.A., Sharp, P.M., Hahn, B.H., Delaporte, E., Mpoudi Ngole, E. & Peeters, M. (2015). Origin of the HIV-1 group O epidemic in western lowland gorillas. *Proceedings of the National Academy of Sciences USA*. 112 (11). pp. E1343–E1352.
- da Silva, R.C., Bedin, E., Mangano, A., Aulicino, P., Pontillo, A., Brandão, L., Guimarães, R., Arraes, L.C., Sen, L. & Crovella, S. (2013). HIV mother-to-child transmission: A complex genetic puzzle tackled by Brazil and Argentina research teams. *Infection, Genetics and Evolution*. 19. pp. 312–322.
- de Mulder, M., Yebra, G., Martín, L., Prieto, L., Mellado, M.J., Rojo, P., Muñoz-Fernández, M.Á., Jimenez de Ory, S., Ramos, J.T. & Holguín, Á. (2011). Drug resistance prevalence and HIV-1 variant characterization in the naive and pretreated HIV-1-infected paediatric population in Madrid, Spain. *Journal of Antimicrobial Chemotherapy*. 66 (10). pp. 2362–2371.
- de Sousa, J.D., Alvarez, C., Vandamme, A. & Muller, V. (2012). Enhanced heterosexual transmission hypothesis for the origin of pandemic HIV-1. *Viruses*. 4 (10). pp. 1950–1983.
- Delgado, E., Thomson, M.M., Villahermosa, M.L., Sierra, M., Ocampo, A., Miralles, C., Rodríguez-Pérez, R., Diz-Aren, J., Ojea-de Castro, R., Losada, E., Cuevas, L., Vázquez-de Parga, E., Carmona, R., Pérez Alvarez, L., Medrano, L., Cuevas, L., Taboada, J.A. & Nájera, R. (2002). Identification of a newly characterized HIV-1 BG intersubtype circulating recombinant form in Galicia, Spain, which exhibits a pseudotype-like virion structure. *Journal of Acquired Immune Deficiency Syndromes*. 29 (5). pp. 536–543.
- Departamento de Doenças Infecciosas do INSA & Programa Nacional para a Infecção VIH/SIDA (2015). *Infecção VIH/SIDA: a situação em Portugal a 31 de dezembro de 2014*. Instituto Nacional de Saúde Doutor Ricardo Jorge (INSA, IP) (ed.). [Online]. Lisboa. Disponível em: <http://repositorio.insa.pt/handle/10400.18/3205>.
- Derache, A., Shin, H.-S., Balamane, M., White, E., Israelski, D., Klausner, J.D., Freeman, A.H. & Katzenstein, D. (2015). HIV drug resistance mutations in proviral DNA from a community treatment program. *PLoS ONE*. 10 (1). p. e0117430.
- Domingo, E., Sheldon, J. & Perales, C. (2012). Viral quasispecies evolution. *Microbiology and Molecular Biology Reviews*. 76 (2). pp. 159–216.

- Espinheira, A. (2013). *Evolução da resistência in vivo aos inibidores da transcriptase reversa em mães multíparas infectadas pelo vírus da imunodeficiência humana tipo 1 que cumpriram regimes terapêuticos para prevenção da transmissão do vírus aos seus filhos*. Dissertação de Mestrado. Universidade de Lisboa.
- Esteves, A., Parreira, R., Piedade, J., Venenno, T., Franco, M., Germano de Sousa, J., Patrício, L., Brum, P., Costa, A. & Canas-Ferreira, W.F. (2003). Spreading of HIV-1 subtype G and envB/gagG recombinant strains among injecting drug users in Lisbon, Portugal. *AIDS Research and Human Retroviruses*. 19 (6). pp. 511–517.
- Esteves, A., Parreira, R., Venenno, T., Franco, M., Piedade, J., Germano de Sousa, J. & Canas-Ferreira, W.F. (2002). Molecular epidemiology of HIV type 1 infection in Portugal: high prevalence of non-B subtypes. *AIDS Research and Human Retroviruses*. 18 (5). pp. 313–325.
- European Centre for Disease Prevention and Control (2012). *Evaluating HIV treatment as prevention in the European context*. [Online]. Estocolmo. Disponível em: http://ecdc.europa.eu/en/publications/_layouts/forms/Publication_DispForm.aspx?ID=624&List=4f55ad51-4aed-4d32-b960-af70113dbb90.
- Faria, N.R., Rambaut, A., Suchard, M. A., Baele, G., Bedford, T., Ward, M.J., Tatem, A.J., Sousa, J.D., Arinaminpathy, N., Pépin, J., Posada, D., Peeters, M., Pybus, O.G. & Lemey, P. (2014). The early spread and epidemic ignition of HIV-1 in human populations. *Science*. 346 (6205). pp. 56–61.
- Fernández-García, A., Delgado, E., Cuevas, M.T., Vega, Y., Montero, V., Sánchez, M., Carrera, C., López-Álvarez, M.J., Miralles, C., Pérez-Castro, S., Cilla, G., Hinojosa, C., Pérez-Álvarez, L. & Thomson, M.M. (2016). Identification of an HIV-1 BG intersubtype recombinant form (CRF73_BG), partially related to CRF14_BG, which is circulating in Portugal and Spain. *PLoS ONE*. 11 (2). p. e0148549.
- Fisher, R., van Zyl, G.U., Travers, S.A.A., Kosakovsky Pond, S.L., Engelbrech, S., Murrell, B., Scheffler, K. & Smith, D. (2012). Deep sequencing reveals minor protease resistance mutations in patients failing a protease inhibitor regimen. *Journal of Virology*. 86 (11). pp. 6231–6237.
- Folks, T.M., Powell, D., Lightfoote, M., Koenig, S., Fauci, A.S., Benn, S., Rabson, A., Daugherty, D., Gendelman, H.E., Hoggan, M.D., Venkatesan, S. & Martin, M.A. (1986). Biological and biochemical characterization of a cloned Leu-3- cell surviving infection with the acquired immune deficiency syndrome retrovirus. *Journal of Experimental Medicine*. 164 (1). pp. 280–290.

- Freed, E.O. (2001). HIV-1 replication. *Somatic cell and molecular genetics*. 26 (1-6). pp. 13–33.
- Freed, E.O. & Martin, M.A. (2013). Human Immunodeficiency Viruses: replication. Em: D. M. Knipe & P. M. Howley (eds.). *Fields Virology*. Filadélfia: Lippincott Williams and Wilkins, pp. 1502–1560.
- Fun, A., Wensing, A.M.J., Verheyen, J. & Nijhuis, M. (2012). Human immunodeficiency virus gag and protease: partners in resistance. *Retrovirology*. [Online]. 9. p. 63. Disponível em: <http://www.retrovirology.com/content/9/1/63>.
- Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.B., Arthur, L.O., Peeters, M., Shaw, G.M., Sharp, P.M. & Hahn, B.H. (1999). Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature*. 397 (6718). pp. 436–441.
- Gao, F., Yue, L., White, A.T., Pappas, P.G., Barchue, J., Hanson, A.P., Greene, B.M., Sharp, P.M., Shaw, G.M. & Hahn, B.H. (1992). Human infection by genetically diverse SIVsm-related HIV-2 in west Africa. *Nature*. 358 (6386). pp. 495–499.
- Gendelman, H.E., Theodore, T.S., Willey, R., McCoy, J., Adachi, A., Mervis, R.J., Venkatesan, S. & Martin, M.A. (1987). Molecular characterization of a polymerase mutant human immunodeficiency virus. *Virology*. 160 (2). pp. 323–329.
- Goff, S.P. (2013). Retroviridae. Em: D. M. Knipe & P. M. Howley (eds.). *Fields Virology*. Filadélfia: Lippincott Williams and Wilkins, pp. 1424–1473.
- Götte, M. (2012). The distinct contributions of fitness and genetic barrier to the development of antiviral drug resistance. *Current Opinion in Virology*. 2 (5). pp. 644–650.
- Gottlieb, M.S., Schroff, R., Schanker, H.M., Weisman, J.D., Fan, P.T., Wolf, R.A. & Saxon, A. (1981). Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *The New England Journal of Medicine*. 305 (24). pp. 1425–1431.
- Grossman, Z., Schapiro, J.M., Levy, I., Elbirt, D., Chowers, M., Riesenber, K., Olstein-Pops, K., Shahar, E., Istomin, V., Asher, I., Gottessman, B.-S., Shemer, Y., Elinav, H., Hassoun, G., Rosenberg, S., Averbuch, D., Machleb-Guri, K., Kra-Oz, Z., Radian-Sade, S., Rudich, H., Ram, D., Maayan, S., Agmon-Levin, N. & Stoecker, Z. (2014). Comparable long-term efficacy of lopinavir/ritonavir and similar drug-resistance profiles in different HIV-1 subtypes. *PLoS ONE*. 9 (1). p. e86239.

- Hahn, B.H., Shaw, G.M., De Cock, K.M. & Sharp, P.M. (2000). AIDS as a zoonosis: scientific and public health implications. *Science*. 287 (5453). pp. 607–614.
- Harrach, B., Benkő, M., Both, G.W., Brown, M., Davison, A.J., Echavarria, M., Hess, M., Jones, M.S., Kajon, A., Lehmkuhl, H.D., Mautner, V., Mittal, S.K. & Wadell, G. (2012). Retroviridae. Em: A. M. Q. King, M. J. Adams, E. B. Carstens, & E. Lefkowitz (eds.). *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier, pp. 477–495.
- Harris, B., von Truchsess, I., Schatzl, H.M., Devare, S.G. & Hackett, J.J. (2005). Genomic characterization of a novel HIV type 1 B/G intersubtype recombinant strain from an injecting drug user in Germany. *AIDS Research and Human Retroviruses*. 21 (7). pp. 654–660.
- Hemelaar, J. (2012). The origin and diversity of the HIV-1 pandemic. *Trends in Molecular Medicine*. 18 (3). pp. 182–192.
- Hemelaar, J., Gouws, E., Ghys, P.D., Osmanov, S. & WHO-UNAIDS Network for HIV Isolation and Characterisation (2011). Global trends in molecular epidemiology of HIV-1 during 2000–2007. *AIDS*. 25 (5). pp. 679–689.
- Henriet, S., Mercenne, G., Bernacchi, S., Paillart, J.-C. & Marquet, R. (2009). Tumultuous relationship between the human immunodeficiency virus type 1 viral infectivity factor (Vif) and the human APOBEC-3G and APOBEC-3F restriction factors. *Microbiology and Molecular Biology Reviews*. 73 (2). pp. 211–232.
- Hirsch, V.M., Olmsted, R.A., Murphey-Corb, M., Purcell, R.H. & Johnson, P.R. (1989). An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature*. 339 (6223). pp. 389–392.
- Holguín, A., Álvarez, A. & Soriano, V. (2002). High prevalence of HIV-1 subtype G and natural polymorphisms at the protease gene among HIV-infected immigrants in Madrid. *AIDS*. 16 (8). pp. 1163–1170.
- Holguín, A., de Mulder, M., Yebra, G., López, M. & Soriano, V. (2008). Increase of non-B subtypes and recombinants among newly diagnosed HIV-1 native Spaniards and immigrants in Spain. *Current HIV Research*. 6 (4). pp. 327–334.
- Holguín, A., Suñe, C., Hamy, F., Soriano, V. & Klimkait, T. (2006). Natural polymorphisms in the protease gene modulate the replicative capacity of non-B HIV-1 variants in the absence of drug pressure. *Journal of Clinical Virology*. 36 (4). pp. 264–271.

- Hsi, J., Wei, H., Xing, H., Feng, Y., He, X., Liao, L., Jia, M., Wang, N., Ning, C. & Shao, Y. (2014). Genome sequence of a novel HIV-1 circulating recombinant form (CRF64_BC) identified from Yunnan, China. *AIDS Research and Human Retroviruses*. 30 (4). pp. 389–393.
- INSIGHT START Study Group (2015). Initiation of antiretroviral therapy in early asymptomatic HIV infection. *New England Journal of Medicine*. 373 (9). pp. 795–807.
- Kabamba-Mukadi, B., Duquenne, A., Henrivaux, P., Musuamba, F., Ruelle, J., Yombi, J.-C., Bodéus, M., Vandercam, B. & Goubau, P. (2010). HIV-1 proviral resistance mutations: usefulness in clinical practice. *HIV Medicine*. 11 (8). pp. 483–492.
- Kalish, M.L., Robbins, K.E., Pieniazek, D., Schaefer, A., Nzilambi, N., Quinn, T.C., St. Louis, M.E., Youngpairoj, A.S., Phillips, J., Jaffe, H.W. & Folks, T.M. (2004). Recombinant viruses and early global HIV-1 epidemic. *Emerging infectious diseases*. 10 (7). pp. 1227–1234.
- Kantor, R., Katzenstein, D.A., Efron, B., Carvalho, A.P., Wynhoven, B., Cane, P., Clarke, J., Sirivichayakul, S., Soares, M.A., Snoeck, J., Pillay, C., Rudich, H., Rodrigues, R., Holguin, A., Ariyoshi, K., Bouzas, M.B., Cahn, P., Sugiura, W., Soriano, V., Brigido, L.F., Grossman, Z., Morris, L., Vandamme, A.-M., Tanuri, A., Phanuphak, P., Weber, J.N., Pillay, D., Harrigan, P.R., Camacho, R., Schapiro, J.M. & Shafer, R.W. (2005). Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase genotype: results of a global collaboration. *PLoS Medicine*. 2 (4). p. e112.
- Keele, B.F., Heuverswyn, F. V., Li, Y., Bailes, E., Takehisa, J., Santiago, M.L., Bibollet-ruche, F., Chen, Y., Wain, L. V., Liegeois, F., Loul, S., Ngole, E.M., Bienvenue, Y., Delaporte, E., Brookfield, J.F.Y., Sharp, P.M., Shaw, G.M., Peeters, M. & Hahn, B.H. (2006). Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science*. 313 (5786). pp. 523–526.
- Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*. 16 (2). pp. 111–120.
- King, N.M., Prabu-Jeyabalan, M., Nalivaika, E.A. & Schiffer, C.A. (2004). Combating susceptibility to drug resistance: lessons from HIV-1 protease. *Chemistry & Biology*. 11 (10). pp. 1333–1338.

- Kohl, N.E., Emini, E.A., Schleif, W.A., Davis, L.J., Heimbach, J.C., Dixon, R.A.F., Scolnick, E.M. & Sigal, I.S. (1988). Active human immunodeficiency virus protease is required for viral infectivity. *Proceedings of the National Academy of Sciences USA*. 85 (13). pp. 4686–4690.
- Konvalinka, J., Kräusslich, H.-G. & Müller, B. (2015). Retroviral proteases and their roles in virion maturation. *Virology*. 479-480. pp. 403–417.
- Kuritzkes, D.R. & Koup, R.A. (2013). HIV-1: pathogenesis, clinical manifestations, and treatment. Em: D. M. Knipe & P. M. Howley (eds.). *Fields Virology*. Filadélfia: Lippincott Williams and Wilkins, pp. 1561–1583.
- Langs-Barlow, A. & Paintsil, E. (2014). Impact of human immunodeficiency virus type-1 sequence diversity on antiretroviral therapy outcomes. *Viruses*. 6 (10). pp. 3855–3872.
- Lessells, R.J., Katzenstein, D.K. & de Oliveira, T. (2012). Are subtype differences important in HIV drug resistance? *Current Opinion in Virology*. 2 (5). pp. 636–643.
- Li, J.Z., Paredes, R., Ribaldo, H.J., Svarovskaia, E.S., Metzner, K.J., Kozal, M.J., Hullsiek, K.H., Balduin, M., Jakobsen, M.R., Geretti, A.M., Thiebaut, R., Ostergaard, L., Masquelier, B., Jonhson, J.A., Miller, M.D. & Kuritzkes, D.R. (2011). Low-frequency HIV-1 drug resistance mutations and risk of NNRTI-based antiretroviral treatment failure: a systematic review and pooled analysis. *JAMA*. 305 (13). pp. 1327–1335.
- Lisovsky, I., Schader, S.M., Martinez-Cajas, J.-L., Oliveira, M., Moisi, D. & Wainberg, M.A. (2010). HIV-1 protease codon 36 polymorphisms and differential development of resistance to nelfinavir, lopinavir, and atazanavir in different HIV-1 subtypes. *Antimicrobial Agents and Chemotherapy*. 54 (7). pp. 2878–2885.
- Lole, K.S., Bollinger, R.C., Paranjape, R.S., Gadkari, D., Kulkarni, S.S., Novak, N.G., Ingersoll, R., Sheppard, H.W. & Ray, S.C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *Journal of Virology*. 73 (1). pp. 152–160.
- Looney, D., Ma, A. & Johns, S. (2015). HIV therapy - the state of art. *Current Topics in Microbiology and Immunology*. 389. pp. 1–29.

- Los Alamos National Laboratory (2015a). *HIV sequence database: HIV circulating recombinant forms (CRFs)*. [Online]. Disponível em: <http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>. [Acedido: Junho 2016].
- Los Alamos National Laboratory (2015b). *HIV sequence database: HIV-1 gene map*. [Online]. Disponível em: <http://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html>. [Acedido: Maio 2015].
- Louis, J.M., Ishima, R., Torchia, D.A. & Weber, I.T. (2007). HIV-1 protease: structure, dynamics, and inhibition. *Advances in Pharmacology*. 55. pp. 261–298.
- Lubke, N., Di Cristanziano, V., Sierra, S., Knops, E., Schulter, E., Jensen, B., Oette, M., Lengauer, T. & Kaiser, R. (2015). Proviral DNA as a target for HIV-1 resistance analysis. *Intervirology*. 58 (3). pp. 184–189.
- Mager, P.P. (2001). The active site of HIV-1 protease. *Medicinal Research Reviews*. 21 (4). pp. 348–353.
- Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L. & Trono, D. (2004). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature*. 424 (6944). pp. 99–103.
- Menéndez-Arias, L. (2013). Molecular basis of human immunodeficiency virus type 1 drug resistance: Overview and recent developments. *Antiviral Research*. 98 (1). pp. 93–120.
- Mercenne, G., Bernacchi, S., Richer, D., Bec, G., Henriot, S., Paillart, J.-C. & Marquet, R. (2010). HIV-1 Vif binds to APOBEC3G mRNA and inhibits its translation. *Nucleic Acids Research*. 38 (2). pp. 633–646.
- Montaner, J.S.G., Lima, V.D., Barrios, R., Yip, B., Wood, E., Kerr, T., Shannon, K., Harrigan, P.R., Hogg, R.S., Daly, P. & Kendall, P. (2010). Expanded HAART coverage is associated with decreased population-level HIV-1-RNA and annual new HIV diagnoses in British Columbia, Canada. *Lancet*. 376 (9740). pp. 532–539.
- Nájera, R., Delgado, E., Pérez-Alvarez, L. & Thomson, M.M. (2002). Genetic recombination and its role in the development of HIV-1 pandemic. *AIDS*. 16 (Suppl 4). pp. S3–S16.

- Neogi, U., Häggblom, A., Santacatterina, M., Bratt, G., Gisslén, M., Albert, J. & Sonnerborg, A. (2014). Temporal trends in the Swedish HIV-1 epidemic: increase in non-B subtypes and recombinant forms over three decades. *PLoS ONE*. 9 (6). p. e99390.
- Niama, F.R., Vidal, N., Bazepeo, S.E., Mpoudi, E., Toure-Kane, C., Parra, H.J., Delaporte, E. & Peeters, M. (2009). CRF45_AKU, a circulating recombinant from Central Africa, is probably the common ancestor of HIV type 1 MAL and HIV type 1 NOGIL. *AIDS Research and Human Retroviruses*. 25 (12). pp. 1345–1353.
- Niculescu, I., Paraschiv, S., Paraskevis, D., Abagiu, A., Batan, I., Banica, L. & Otelea, D. (2015). Recent HIV-1 outbreak among intravenous drug users in Romania: evidence for cocirculation of CRF14_BG and subtype F1 strains. *AIDS Research and Human Retroviruses*. 31 (5). pp. 488–495.
- Palma, A.C., Araújo, F., Duque, V., Borges, F., Paixão, M.T. & Camacho, R. (2007). Molecular epidemiology and prevalence of drug resistance-associated mutations in newly diagnosed HIV-1 patients in Portugal. *Infection, Genetics and Evolution*. 7 (3). pp. 391–398.
- Palmer, S., Kearney, M., Maldarelli, F., Halvas, E.K., Bixby, C.J., Bazmi, H., Rock, D., Falloon, J., Davey, R.T.J., Dewar, R.L., Metcalf, J.A., Hammer, S., Mellors, J.W. & Coffin, J.M. (2005). Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *Journal of Clinical Microbiology*. 43 (1). pp. 406–413.
- Palmisano, L., Galluzzo, C.M. & Giuliano, M. (2009). The importance of testing genotypic resistance in proviral DNA of patients fully responding to highly active antiretroviral therapy. *Journal of Acquired Immune Deficiency Syndromes*. 51 (2). pp. 233–234.
- Panel on Antiretroviral Guidelines for Adults and Adolescents (2015). *Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents*. [Online]. Disponível em: <https://aidsinfo.nih.gov/contentfiles/adultandadolescentgl003400.pdf>. [Acedido: Julho 2015].
- Panel on Treatment of HIV-infected pregnant women and prevention of perinatal transmission (2014). *Recommendations for use of antiretroviral drugs in pregnant HIV-1-infected women for maternal health and interventions to reduce perinatal HIV transmission in the United States*. [Online]. Disponível em: <https://aidsinfo.nih.gov/contentfiles/lvguidelines/perinatalgl.pdf>. [Acedido: Julho 2015].

- Paraskevis, D., Nikolopoulos, G., Fotiou, A., Tsiara, C., Paraskeva, D., Sypsa, V., Lazanas, M., Gargalianos, P., Psychogiou, M., Skoutelis, A., Wiessing, L., Friedman, S.R., Des Jarlais, D.C., Terzidou, M., Kremastinou, J., Malliori, M. & Hatzakis, A. (2013). Economic recession and emergence of an HIV-1 outbreak among drug injectors in Athens metropolitan area: a longitudinal study. *PLoS ONE*. 8 (11). pp. 1–10.
- Paraskevis, D., Nikolopoulos, G., Tsiara, C., Paraskeva, D., Antoniadou, A., Lazanas, M., Gargalianos, P., Psychogiou, M., Malliori, M., Kremastinou, J. & Hatzakis, A. (2011). HIV-1 outbreak among injecting drug users in Greece, 2011: a preliminary report. *Eurosurveillance*. 16 (36).
- Paraskevis, D., Paraschiv, S., Sypsa, V., Nikolopoulos, G., Tsiara, C., Magiorkinis, G., Psychogiou, M., Flampouris, A., Mardarescu, M., Niculescu, I., Batan, I., Malliori, M., Otelea, D. & Hatzakis, A. (2015). Enhanced HIV-1 surveillance using molecular epidemiology to study and monitor HIV-1 outbreaks among intravenous drug users (IDUs) in Athens and Bucharest. *Infection, Genetics and Evolution*. 35. pp. 109–121.
- Parreira, R., Piedade, J., Domingues, A., Lobão, D., Santos, M., Venenno, T., Baptista, J.L., Mussa, S.A., Barreto, A.T., Baptista, A.J. & Esteves, A. (2006). Genetic characterization of human immunodeficiency virus type 1 from Beira, Mozambique. *Microbes and Infection*. 8 (9-10). pp. 2442–2451.
- Parreira, R., Santos, M., Piedade, J. & Esteves, A. (2004). Natural polymorphism of HIV-1 subtype G protease and cleavage sites. *AIDS*. 18 (9). pp. 1345–1346.
- Pau, A.K. & George, J.M. (2014). Antiretroviral therapy: current drugs. *Infectious Disease Clinics of North America*. 28 (3). pp. 371–402.
- Peeters, M., Courgnaud, V., Abela, B., Auzel, P., Pourrut, X., Bibollet-Ruche, F., Loul, S., Liegeois, F., Butel, C., Koulagna, D., Mpoudi-Ngole, E., Shaw, G.M., Hahn, B.H. & Delaporte, E. (2002). Risk to human health from a plethora of simian immunodeficiency viruses in primate bushmeat. *Emerging Infectious Diseases*. 8 (5). pp. 451–457.
- Peeters, M., Honoré, C., Huet, T., Bedjabaga, L., Ossari, S., Bussi, P., Cooper, R.W. & Delaporte, E. (1989). Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon. *AIDS*. 3 (10). pp. 625–630.
- Pépin, J. (2013). The origins of AIDS: from patient zero to ground zero. *Journal of Epidemiology & Community Health*. 67 (6). pp. 473–475.

- Pereira, N.R., Duro, R., Piñero, C., Malheiro, L., Soares, J., Serrão, R. & Sarmiento, A. (2014). Characterization of HIV-1 subtypes in a Portuguese cohort. *Journal of the International AIDS Society*. 17 (4 Suppl 3). p. 19683.
- Piñeiro, C., Fernandes, S., Figueiredo, C., Santos, A.S., Moucho, M., Serrão, R., Montenegro, N. & Sarmiento, A. (2014). Prevention of mother-to-child transmission: experience of a Portuguese centre. *Journal of the International AIDS Society*. 17 (4 Suppl 3). p. 19705.
- Plantier, J.C., Leoz, M., Dickerson, J.E., De Oliveira, F., Cordonnier, F., Lemée, V., Damond, F., Robertson, D.L. & Simon, F. (2009). A new human immunodeficiency virus derived from gorillas. *Nature Medicine*. 15 (8). pp. 871–882.
- Pokorná, J., Machala, L., Řezáčová, P. & Konvalinka, J. (2009). Current and novel inhibitors of HIV protease. *Viruses*. 1 (3). pp. 1209–1239.
- Prieto, L.M., González-Tomé, M.I., Muñoz, E., Fernández-Ibieta, M., Soto, B., Del Rosal, T., Cuadrado, I., Navarro, M.L., Roa, M.A., Beceiro, J., de José, M.I., Santos, M.J., Lora, D., Ramos, J.T. & Madrid Cohort of HIV-infected mother-infant pairs (2012). Low rates of mother-to-child transmission of HIV-1 and risk factors for infection in Spain: 2000-2007. *The Pediatric Infectious Disease Journal*. 31 (10). pp. 1053–1058.
- Programa Nacional para a Infecção VIH/SIDA (2015). *Recomendações portuguesas para o tratamento da infecção VIH-1 e VIH-2*. [Online]. Disponível em: <https://www.dgs.pt/documentos-e-publicacoes/recomendacoes-portuguesas-para-o-tratamento-da-infecao-por-vih-1-e-vih-2-versao-10-2015.aspx>.
- Rabi, S.A., Laird, G.M., Durand, C.M., Laskey, S., Shan, L., Bailey, J.R., Chioma, S., Moore, R.D. & Siliciano, R.F. (2013). Multi-step inhibition explains HIV-1 protease inhibitor pharmacodynamics and resistance. *The Journal of Clinical Investigation*. 123 (9). pp. 3848–3860.
- Robertson, D.L., Anderson, J.P., Bradac, J.A., Carr, J.K., Foley, B., Funkhouser, R.K., Gao, F., Hahn, B.H., Kalish, M.L., Kuiken, C., Learn, G.H., Leitner, T., McCutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P.M., Wolinsky, S. & Korber, B. (2000). HIV-1 nomenclature proposal. *Science*. 288 (5463). pp. 55–56.
- Rojas Sanchez, P. & Holguín, A. (2014). Drug resistance in the HIV-1-infected paediatric population worldwide: a systematic review. *Journal of Antimicrobial Chemotherapy*. 69 (8). pp. 2032–2042.

- Rosé, J.R., Babé, L.M. & Craik, C.S. (1995). Defining the level of human immunodeficiency virus type 1 (HIV-1) protease activity required for HIV-1 particle maturation and infectivity. *Journal of Virology*. 69 (5). pp. 2751–2758.
- Rose, P.P. & Korber, B.T. (2000). Detecting hypermutations in viral sequences with an emphasis on G --> A hypermutation. *Bioinformatics*. 16 (4). pp. 400–401.
- Roseninge, M.M. & Doerholt, K. (2013). HIV in mothers and children. *Medicine*. 41 (8). pp. 461–465.
- Saitou, N. & Nei, M. (1987). The Neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*. 4 (4). pp. 406–425.
- Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences USA*. 74 (12). pp. 5463–5467.
- Santiago, M.L., Range, F., Keele, B.F., Li, Y., Bailes, E., Bibollet-Ruche, F., Fruteau, C., Noe, R., Peeters, M., Brookfield, J.F.Y., Shaw, G.M., Sharp, P.M. & Hahn, B.H. (2005). Simian immunodeficiency virus infection in free-ranging Sooty Mangabeys (*Cercocebus atys atys*) from the Tai Forest, Côte d’Ivoire: implications for the origin of epidemic human immunodeficiency virus type 2. *Journal of Virology*. 79 (19). pp. 12515–12527.
- Santoro, M.M. & Perno, C.F. (2013). HIV-1 genetic variability and clinical implications. *ISRN Microbiology*. pp. 1–20.
- Santos, A.F. & Soares, M.A. (2010). HIV genetic diversity and drug resistance. *Viruses*. 2 (2). pp. 503–531.
- Santos, A.F.A., Tebit, D.M., Lalonde, M.S., Abecasis, A.B., Ratcliff, A., Camacho, R.J., Diaz, R.S., Herchenroder, O., Soares, M.A. & Arts, E.J. (2012). Effect of natural polymorphisms in the HIV-1 CRF02_AG protease on protease inhibitor hypersusceptibility. *Antimicrobial Agents and Chemotherapy*. 56 (5). pp. 2719–2725.
- Sayer, J.M., Liu, F., Ishima, R., Weber, I.T. & Louis, J.M. (2008). Effect of the active site D25N mutation on the structure, stability, and ligand binding of the mature HIV-1 protease. *Journal of Biological Chemistry*. 283 (19). pp. 13459–13470.
- Sharp, P.M. & Hahn, B.H. (2010). The evolution of HIV-1 and the origin of AIDS. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 365 (1552). pp. 2487–2494.

- Sharp, P.M. & Hahn, B.H. (2011). Origins of HIV and the AIDS pandemic. *Cold Spring Harbor Perspectives in Medicine*. 1 (1). p. a006841.
- Sierra, S., Kupfer, B. & Kaiser, R. (2005). Basics of the virology of HIV-1 and its replication. *Journal of Clinical Virology*. 34 (4). pp. 233–244.
- Skar, H., Hedskog, C. & Albert, J. (2011). HIV-1 evolution in relation to molecular epidemiology and antiretroviral resistance. *Annals of the New York Academy of Sciences*. 1230. pp. 108–118.
- Soto-Ramirez, L.E., Rodriguez-Diaz, R., Durán, A.S., Losso, M.H., Salomón, H., Gómez-Carrillo, M., Pampuro, S., Harris, D.R., Duarte, G., De Souza, R.S. & Read, J.S. (2008). Antiretroviral resistance among HIV type 1-infected women first exposed to antiretrovirals during pregnancy: plasma versus PBMCs. *AIDS Research and Human Retroviruses*. 24 (6). pp. 797–804.
- Stanford University (2014). *HIVdb Program - Genotypic Resistance Interpretation Algorithm*. [Online]. Disponível em: <http://sierra2.stanford.edu/sierra/servlet/JSierra>. [Acedido: Agosto 2015].
- Steege, K., Demecheleer, E., De Cabooter, N., Nges, D., Temmerman, M., Ndumbe, P., Mandaliya, K., Plum, J. & Verhofstede, C. (2006). A sensitive in-house RT-PCR genotyping system for combined detection of plasma HIV-1 and assessment of drug resistance. *Journal of Virological Methods*. 133 (2). pp. 137–145.
- Swanstrom, R. & Wills, J.W. (1997). Synthesis, assembly, and processing of viral proteins. Em: J. M. Coffin, S. H. Hughes, & H. E. Varmus (eds.). *Retroviruses*. [Online]. Cold Spring Harbor Laboratory Press, pp. 263–312. Disponível em: <http://www.ncbi.nlm.nih.gov/books/NBK19456/>.
- Takebe, Y., Uenishi, R. & Li, X. (2008). Global molecular epidemiology of HIV: understanding the genesis of AIDS pandemic. *Advances in pharmacology*. 56 (07). pp. 1–25.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*. 28 (10). pp. 2731–2739.
- Tebit, D.M. & Arts, E.J. (2011). Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *The Lancet Infectious Diseases*. 11 (1). pp. 45–56.

- Theys, K., Abecasis, A.B. & Vandamme, A.-M. (2013). HIV-1 drug resistance: where do polymorphisms fit in? *Future Microbiology*. 8 (3). pp. 303–306.
- Thorne, C. & Newell, M.-L. (2007). HIV. *Seminars in Fetal and Neonatal Medicine*. 12 (3). pp. 174–181.
- Townsend, C.L., Burne, L., Cortina-Borja, M., Thorne, C., de Ruiter, A., Lyall, H., Taylor, G.P., Peckam, C.S. & Tookey, P.A. (2014). Earlier initiation of ART and further decline in mother-to-child HIV transmission rates, 2000-2011. *AIDS*. 28 (7). pp. 1049–1057.
- Tsibris, A.M.N. & Hirsch, M.S. (2010). Antiretroviral therapy in the clinic. *Journal of Virology*. 84 (11). pp. 5458–5464.
- Turner, B.G. & Summers, M.F. (1999). Structural biology of HIV. *Journal of Molecular Biology*. 285 (1). pp. 1–32.
- Turriziani, O., Andreoni, M. & Antonelli, G. (2010). Resistant viral variants in cellular reservoirs of human immunodeficiency virus infection. *Clinical Microbiology and Infection*. 16 (10). pp. 1518–1524.
- Turriziani, O., Bucci, M., Stano, A., Scagnolari, C., Bellomi, F., Fimiani, C., Mezzaroma, I., D’Ettorre, G., Brogi, A., Vullo, V. & Antonelli, G. (2007). Genotypic resistance of archived and circulating viral strains in the blood of treated HIV-infected individuals. *Journal of Acquired Immune Deficiency Syndromes*. 44 (5). pp. 518–524.
- UNAIDS (2014). *The Gap Report*. [Online]. Disponível em: http://www.unaids.org/en/resources/documents/2014/20140716_UNAIDS_gap_report.
- UNAIDS (2015a). *AIDS by the numbers 2015*. [Online]. Disponível em: http://www.unaids.org/en/resources/documents/2015/AIDS_by_the_numbers_2015.
- UNAIDS (2015b). *Fact Sheet 2015 - Global Statistics*. [Online]. Disponível em: http://www.unaids.org/sites/default/files/media_asset/20150901_FactSheet_2015_en.pdf.
- UNAIDS & AIDInfo (2015). *Indicators*. [Online]. Disponível em: <http://aidsinfo.unaids.org/#>. [Acedido: Agosto 2015].

- US. Department of Health and Human Services, Food and Drug Administration & Center for Drug Evaluation and Research (2013). *Guidance for industry: human immunodeficiency virus-1 infection: developing antiretroviral drugs for treatment*. [Online]. Clinical Antimicrobial. Disponível em: <http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm355128.pdf>.
- Van Heuverswyn, F., Li, Y., Nell, C., Bailes, E., Keele, B.F., Liu, W., Loul, S., Butel, C., Liegeois, F., Bienvenue, Y., Ngolle, E.M., Sharp, P.M., Shaw, G.M., Delaporte, E., Hahn, B.H. & Peeters, M. (2006). Human immunodeficiency viruses: SIV infection in wild gorillas. *Nature*. 444 (7116). p. 164.
- Vercauteren, J. & Vandamme, A.-M. (2006). Algorithms for the interpretation of HIV-1 genotypic drug resistance information. *Antiviral Research*. 71 (2-3). pp. 335–342.
- Vergne, L., Stuyver, L., Van Houtte, M., Butel, C., Delaporte, E. & Peeters, M. (2006). Natural polymorphism in protease and reverse transcriptase genes and in vitro antiretroviral drug susceptibilities of non-B HIV-1 strains from treatment-naive patients. *Journal of Clinical Virology*. 36 (1). pp. 43–49.
- Vermund, S.H. & Leigh-Brown, A.J. (2012). The HIV epidemic: high-income countries. *Cold Spring Harbor Perspectives in Medicine*. 2 (5). p. a007195.
- Vidal, N., Peeters, M., Mulanga-Kabeya, C., Nzilambi, N., Robertson, D., Ilunga, W., Sema, H., Tshimanga, K., Bongo, B. & Delaporte, E. (2000). Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. *Journal of Virology*. 74 (22). pp. 10498–10507.
- Videira e Castro, S. (2010). *Diversidade genética e resistência aos anti-retrovirais inibidores enzimáticos de vírus da imunodeficiência humana tipo 1 (HIV-1) de uma população de toxicod dependentes por via endovenosa da grande Lisboa*. Dissertação de Mestrado. Universidade Nova de Lisboa.
- ViralZone (2015). *Human immunodeficiency virus 1*. [Online]. Disponível em: http://viralzone.expasy.org/all_by_species/7.html. [Acedido: Abril 2015].
- Volberding, P.A. & Deeks, S.G. (2010). Antiretroviral therapy and management of HIV infection. *Lancet*. 376 (9734). pp. 49–62.

- von Wyl, V., Kouyos, R.D., Yerly, S., Boni, J., Shah, C., Burgisser, P., Klimkait, T., Weber, R., Hirschel, B., Cavassini, M., Staehelin, C., Battegay, M., Vernazza, P.L., Bernasconi, E., Ledergerber, B., Bonhoeffer, S., Gunthard, H.F. & Swiss HIV Cohort Study (2011). The role of migration and domestic transmission in the spread of HIV-1 non-B subtypes in Switzerland. *Journal of Infectious Diseases*. 204 (7). pp. 1095–1103.
- Wainberg, M.A. & Brenner, B.G. (2012). The impact of HIV genetic polymorphisms and subtype differences on the occurrence of resistance to antiretroviral drugs. *Molecular Biology International*. 2012. pp. 1–10.
- Warszawski, J., Tubiana, R., Le Chanadec, J., Blanche, S., Teglas, J.P., Dolfus, C., Faye, A., Burgard, M., Rouzioux, C., Mandelbrot, L. & ANRS French Perinatal Cohort (2008). Mother-to-child HIV transmission despite antiretroviral therapy in the ANRS French Perinatal Cohort. *AIDS*. 22 (2). pp. 289–299.
- Weber, I.T. & Agniswamy, J. (2009). HIV-1 protease: structural perspectives on drug resistance. *Viruses*. 1 (3). pp. 1110–1136.
- Wensing, A.M., Calvez, V., Gunthard, H.F., Johnson, V.A., Paredes, R., Pillay, D., Shafer, R.W. & Richman, D.D. (2015). 2015 Update of the drug resistance mutations in HIV-1. *Topics in Antiviral Medicine*. [Online]. 23 (4). Disponível em: <https://www.iasusa.org/sites/default/files/tam/23-4-132.pdf>.
- Wensing, A.M.J., van Maarseveen, N.M. & Nijhuis, M. (2010). Fifteen years of HIV protease inhibitors: raising the barrier to resistance. *Antiviral Research*. 85 (1). pp. 59–74.
- World Health Organization (2002). *WHO-UNAIDS Guidelines for standard HIV isolation and characterization procedures*. Suíça.
- World Health Organization (2013). *Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach*. [Online]. Suíça. Disponível em: http://apps.who.int/iris/bitstream/10665/85321/1/9789241505727_eng.pdf.
- Worobey, M., Gemmel, M., Teuwen, D.E., Haselkorn, T., Bunce, M., Muyembe, J., Kabongo, J.M., Kalengayi, M., Marck, E. Van, Gilbert, M.T.P. & Wolinsky, S.M. (2008). Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature*. 455 (7213). pp. 661–664.
- Yu, X., Yu, Y., Liu, B., Luo, K., Kong, W., Mao, P. & Yu, X.F. (2003). Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science*. 302 (5647). pp. 1056–1060.

Zeldin, R.K. & Petruschke, R.A. (2004). Pharmacological and therapeutic properties of ritonavir-boosted protease inhibitor therapy in HIV-infected patients. *Journal of Antimicrobial Chemotherapy*. 53 (1). pp. 4–9.

6. *Anexos*

Anexo 1

Meios e soluções utilizados	Composição
Solução de agarose 1,2% (p/v) (50 ml)	<ul style="list-style-type: none"> ▪ 0,6 g de agarose ▪ 50 ml de tampão TAE 0,5x
Solução de agarose 1,0% (p/v) (50 ml)	<ul style="list-style-type: none"> ▪ 0,5 g de agarose ▪ 50 ml de tampão TAE 0,5x
Tampão Tris-Acetato-EDTA (TAE) 50x	<ul style="list-style-type: none"> ▪ 242 g Tris-base ▪ 57,1 ml ácido acético glacial ▪ 100 ml 0,5 M EDTA, pH 8,0 ▪ Adicionar água para 1000 ml
Tampão de aplicação	<ul style="list-style-type: none"> ▪ 0,25% (p/v) azul de bromofenol ▪ 40% (p/v) sacarose
Tampão de ligação da T4 DNA ligase 2x (Promega, EUA)	<ul style="list-style-type: none"> ▪ 60 mM Tris-HCl (pH 7,8) ▪ 20 mM MgCl₂ ▪ 20 mM DTT ▪ 2 mM ATP ▪ 10% (v/v) de PEG
Meio LB	<ul style="list-style-type: none"> ▪ 10 g Bacto-triptona ▪ 5 g extrato de levedura ▪ 10 g NaCl ▪ 15 g agar* ▪ Adicionar água para 1000 ml e autoclavar * Não adicionar para meio líquido
Meio de criopreservação	<ul style="list-style-type: none"> ▪ 10% (v/v) DMSO ▪ 10% (v/v) glicerol
Tampão Tris-EDTA-glucose (TEG)	<ul style="list-style-type: none"> ▪ 50 mM glucose ▪ 10 mM EDTA ▪ 25 mM Tris-HCl, pH 8,0
Solução de lise (SDS alcalino)	<ul style="list-style-type: none"> ▪ 0,2 N NaOH ▪ 1,5% (v/v) SDS
Meio TSS	<ul style="list-style-type: none"> ▪ LB 1x ▪ 10% (v/v) PEG 6000 ▪ 5% (v/v) DMSO

	<ul style="list-style-type: none">▪ 50 mM MgSO₄ (pH 6,5)
Tampão EcoRI 10x (Fermentas, Lituânia)	<ul style="list-style-type: none">▪ 50 mM Tris-HCl (pH 7,5 a 37°C)▪ 10 mM MgCl₂▪ 100 mM NaCl▪ 0,02% (v/v) Triton X-100▪ 0,1 mg/ml BSA

Anexo 2

Dados epidemiológicos relativos às 34 mulheres incluídas no estudo e respectivas gestações. A tabela apresenta a identificação (ID) das amostras (n=70) recolhidas após o parto, assim como, idade das mulheres (à altura do parto), nacionalidade, ano do parto, terapia anti-retroviral (TARV) realizada, com indicação específica para os inibidores da protease (PIs), e ocorrência de transmissão vertical do HIV-1 para o recém-nascido [ND = não determinado].

ID	Idade (no parto)	Nacionalidade	Ano do parto	TARV	Regime terapêutico	PIs	Transmissão vertical
PT_3A	34	Portugal	2003	Não	-	-	Não
PT_3B	36		2005	Sim	2NRTIs + PI	Indinavir	Não
PT_4A	27	Portugal	2001	Sim	2NRTIs + NNRTI/PI	-	Não
PT_4B	28		2002	Sim	2NRTIs + NNRTI/PI	-	Não
PT_4C	31		2005	Sim	2NRTIs + PI	Lopinavir/Ritonavir	Não
PT_5A	21	Portugal	2002	Sim	2NRTIs + NNRTI/PI	-	Não
PT_5B	26		2007	Sim	2NRTIs + PI	Indinavir	Não
PT_6A	28	Angola	2001	Sim	2NRTIs + NNRTI/PI	-	Não
PT_6B	31		2004	Sim	2NRTIs + NNRTI	-	Não
PT_8A	28	Portugal	2002	Sim	2NRTIs + NNRTI/PI	-	Não
PT_8B	29		2003	Sim	2NRTIs + PI	Indinavir	Não
PT_9A	30	Portugal	2004	Sim	2NRTIs + PI	Indinavir	Não
PT_9B	33		2007	Sim	2NRTIs + PI	Indinavir	Não
PT_10A	20	Portugal	2004	Não	-	-	Sim
PT_10B	24		2008	Sim	2NRTIs + NNRTI/PI	-	Não

ID	Idade (no parto)	Nacionalidade	Ano do parto	TARV	Regime terapêutico	PIs	Transmissão vertical
PT_13A	21	Portugal	2002	Não	-	-	Não
PT_13B	24		2005	Sim	2NRTIs + NNRTI	-	Não
PT_17A	19	Portugal	2004	Sim	2NRTIs + PI	Nelfinavir	Não
PT_17B	20		2005	Sim	2NRTIs + PI	Nelfinavir	Não
PT_17C	21		2006	Sim	2NRTIs + PI	Saquinavir/Ritonavir	Não
PT_20A	31	Moçambique	2005	Sim	2NRTIs + NNRTI/PI	-	Não
PT_20B	34		2008	Sim	2NRTIs + NNRTI + PI	-	Não
PT_24A	ND	Portugal	1999	Sim	2NRTIs + NNRTI/PI	-	Não
PT_24B	ND		2003	Sim	2NRTIs + NNRTI	-	Não
PT_27A	31	Portugal	1999	Sim	2NRTIs + NNRTI/PI	-	Não
PT_27B	37		2005	Sim	2NRTIs + PI	Nelfinavir	Não
PT_29A	26	Portugal	2000	Sim	2NRTIs + NNRTI/PI	-	Não
PT_29B	30		2004	Sim	2NRTIs + PI	Nelfinavir	Não
PT_32A	26	Portugal	2000	Sim	2NRTIs	-	Não
PT_32B	32		2006	Sim	2NRTIs + NNRTI + PI	-	Não
PT_33A	27	Angola	2000	Sim	2NRTIs	-	Não
PT_33B	33		2006	Sim	2NRTIs + NNRTI + PI	-	Não
PT_37A	26	Portugal	2001	Não	-	-	Sim
PT_37B	30		2005	Sim	2NRTIs + PI	Nelfinavir	Não
PT_38A	29	Moçambique	2001	Sim	2NRTIs + NNRTI	-	Não

ID	Idade (no parto)	Nacionalidade	Ano do parto	TARV	Regime terapêutico	PIs	Transmissão vertical
PT_38B	33		2005	Sim	2NRTIs + NNRTI	-	Não
PT_43A	19	Portugal	2001	Não	-	-	Sim
PT_43B	24		2006	Não	-	-	Não
PT_45A	33	Moçambique	2001	Sim	2NRTIs	-	Não
PT_45B	38		2006	Sim	2NRTIs + PI	Saquinavir/Ritonavir	Não
PT_47A	26	Angola	2002	Sim	2NRTIs + NNRTI/PI	-	Não
PT_47B	32		2008	Sim	2NRTIs + PI	Lopinavir/Ritonavir	Não
PT_48A	28	Angola	2002	Sim	2NRTIs + NNRTI/PI	-	Não
PT_48B	34		2008	Sim	2NRTIs + NNRTI	-	Não
PT_52A	32	Portugal	2002	Sim	2NRTIs + NNRTI/PI	-	Não
PT_52B	36		2006	Sim	2NRTIs + PI	Nelfinavir	Não
PT_53A	17	Portugal	2002	Sim	2NRTIs	-	Não
PT_53B	21		2006	Sim	2NRTIs + NNRTI	-	Não
PT_54A	30	Angola	2002	Sim	2NRTIs + NNRTI/PI	-	Não
PT_54B	34		2006	Sim	2NRTIs + PI	Saquinavir/Ritonavir	Não
PT_55A	34	Portugal	2002	Sim	2NRTIs	-	Não
PT_55B	38		2006	Sim	2NRTIs + NNRTI	-	Não
PT_56A	19	Portugal	2002	Sim	2NRTIs + NNRTI/PI	-	Não
PT_56B	24		2007	Sim	2NRTIs + NNRTI/PI	-	Não
PT_57A	27	Portugal	2002	Sim	2NRTIs + NNRTI/PI	-	Não

ID	Idade (no parto)	Nacionalidade	Ano do parto	TARV	Regime terapêutico	PIs	Transmissão vertical
PT_57B	32		2007	Sim	2NRTIs + PI	Nelfinavir	Não
PT_58A	25	Angola	2002	Sim	2NRTIs + NNRTI/PI	-	Não
PT_58B	30		2007	Sim	2NRTIs + NNRTI	-	Não
PT_60A	ND	ND	2003	Não	-	-	Não
PT_60B	ND		2008	Não	-	-	Não
PT_65A	29	Portugal	2003	Sim	2NRTIs	-	Não
PT_65B	33		2007	Sim	2NRTIs + NNRTI	-	Não
PT_66A	31	Portugal	2003	Sim	2NRTIs + NNRTI	-	Não
PT_66B	35		2007	Sim	2NRTIs + PI	Lopinavir/Ritonavir	Não
PT_68A	20	Portugal	2004	Sim	2NRTIs + PI	Nelfinavir	Não
PT_68B	23		2007	Sim	2NRTIs + NNRTI/PI	-	Não
PT_71A	36	Angola	2004	Sim	2NRTIs + NNRTI/PI	-	Não
PT_71B	39		2007	Sim	2NRTIs + PI	Lopinavir/Ritonavir	Não
PT_72A	31	Angola	2004	Sim	2NRTIs + NNRTI	-	Não
PT_72B	35		2008	Sim	2NRTIs + NNRTI	-	Não

Anexo 3

Resultados obtidos na pesquisa de regiões de semelhança local entre sequências com recurso ao programa *BLAST* (do inglês, *Basic Local Alignment Search Tool*). Na tabela estão indicados os números de acesso, na base de dados *GenBank/EMBL/DDBJ*, das sequências nucleotídicas que apresentaram uma percentagem de semelhança mais elevada com as sequências em estudo (ID), assim como o *E value* obtido para a respectiva comparação, incluindo, sempre que possível, sequências de origem portuguesa.

ID	Pesquisa de semelhança genética (Nº acesso)	Identidade (%)	<i>E value</i>
PT_3A	EU068289.1	94	1e-177
	AJ535040.1	94	3e-174
PT_3B	EU068289.1	95	0.0
	AJ535040.1	94	3e-174
PT_4A	FJ647145.1	95	0.0
PT_4B	FJ647145.1	96	0.0
PT_4C	JN214919.1	94	3e-178
PT_5A	AJ535004.1	99	0.0
PT_5B	AJ535004.1	99	0.0
PT_6A	AJ535040.1	98	0.0
PT_6B	AJ535040.1	98	0.0
PT_8A	KJ140266.1	93	1e-167
PT_8B	KJ140266.1	93	7e-170
PT_9A	HM469975.1	91	1e-157
PT_9B	AY165204.1	93	1e-172
PT_10A	AJ535023.1	96	0.0
PT_10B	AJ535023.1	96	0.0
PT_13A	AJ534996.1	99	0.0
PT_13B	AJ534996.1	97	0.0
PT_17A	AJ535005.1	99	2e-175
PT_17B	AJ535005.1	99	2e-174
PT_17C	AJ535005.1	99	9e-174
PT_20A	FJ498326.1	96	0.0
PT_20B	FJ498326.1	96	0.0
PT_24A	AF450097.1	98	0.0
	GU230139.1	98	0.0
PT_24B	GU230139.1	98	0.0
	AF450097.1	98	0.0
PT_27A	DQ193768.1	96	0.0
PT_27B	AY713409.1	95	0.0
PT_29A	AJ534991.1	98	0.0
PT_29B	AJ534991.1	97	0.0
PT_32A	AF423755.1	97	0.0
	AJ535023.1	97	0.0
PT_32B	AJ535023.1	97	0.0
PT_33A	EF108467.1	96	0.0

ID	Pesquisa de semelhança genética (N° acesso)	Identidade (%)	<i>E value</i>
	GU230138.1	95	2e-179
PT_33B	EF119528.1	96	0.0
	GU230138.1	94	1e-177
PT_37A	JQ511330.1	95	0.0
PT_37B	JQ511330.1	95	0.0
PT_38A	KM050460.1	98	0.0
PT_38B	FJ199583.1	96	0.0
PT_43A	AY882421.1	97	0.0
	AJ535008.1	97	0.0
PT_43B	AY882421.1	96	0.0
	AJ535031.1	96	0.0
PT_45A	GQ433815.1	98	0.0
PT_45B	DQ660063.1	96	0.0
PT_47A	HG780690.1	95	0.0
PT_47B	HG780690.1	94	7e-180
PT_48B	AY102509.1	93	6e-166
PT_52A	AF450097.1	97	0.0
	AJ535001.1	97	0.0
PT_52B	AJ535001.1	98	0.0
PT_53A	AF349382.1	93	4e-172
PT_53B	AF349382.1	93	4e-172
PT_54B	FR846410.1	99	0.0
PT_55B	JF689859.1	96	0.0
PT_56A	AY102490.1	97	0.0
PT_56B	AY102490.1	96	0.0
PT_57A	AF423755.1	98	0.0
	GU230139.1	98	0.0
PT_57B	AF423755.1	99	0.0
	GU230139.1	99	0.0
PT_58A	FJ199777.1	97	0.0
PT_58B	FJ199777.1	97	0.0
PT_60A	AJ535011.1	99	0.0
PT_60B	AJ535011.1	98	0.0
PT_65A	AJ535009.1	97	0.0
PT_65B	AJ535009.1	97	0.0
PT_66A	AY586548.2	97	0.0
	AJ535023.1	96	0.0
PT_66B	AY586548.2	98	0.0
	AJ535023.1	96	0.0
PT_68A	AY388691.1	96	0.0
PT_68B	AY388820.1	95	0.0
PT_71A	KM049923.1	95	0.0
PT_71B	FJ498541.1	96	0.0
PT_72A	FM955689.1	96	0.0
PT_72B	FN599725.1	96	0.0

Anexo 4

Comparação dos resultados de genotipagem obtidos para as amostras estudadas (n=67) com recurso ao programa *REGA HIV-1 & 2 Automated Subtyping Tool* (v.2.0) e através da árvore filogenética construída com base em métodos não automatizados. ID – identificação das sequências nucleotídicas; NA – não atribuído; * – valor de *bootstrap* inferior a 70%; § – não inclusão em grupo monofilético consistente.

ID	<i>REGA subtyping tool 2.0</i>	Árvore filogenética
PT_3A	G	G
PT_3B	G	G
PT_4A	B	B*
PT_4B	B	B*
PT_4C	NA	B*
PT_5A	NA	02_AG*
PT_5B	NA	02_AG*
PT_6A	G	G
PT_6B	G	G
PT_8A	NA	NA [§]
PT_8B	NA	NA [§]
PT_9A	B	B*
PT_9B	NA	B*
PT_10A	NA	G
PT_10B	G	G
PT_13A	G	G
PT_13B	NA	G
PT_17A	G	G
PT_17B	G	G
PT_17C	G	G
PT_20A	C	C
PT_20B	C	C
PT_24A	G	G
PT_24B	G	G
PT_27A	B	B*
PT_27B	B	B*
PT_29A	G	G
PT_29B	G	G
PT_32A	G	G
PT_32B	G	G
PT_33A	G	G
PT_33B	G	G
PT_37A	B	B*
PT_37B	B	B*
PT_38A	C	C
PT_38B	C	C
PT_43A	G	G

ID	REGA subtyping tool 2.0	Árvore filogenética
PT_43B	G	G
PT_45A	C	C
PT_45B	C	C
PT_47A	A1	A/45_cpx*
PT_47B	NA	A/45_cpx*
PT_48B	NA	NA [§]
PT_52A	NA	G
PT_52B	G	G
PT_53A	NA	NA [§]
PT_53B	NA	NA [§]
PT_54B	G	G
PT_55B	B	B*
PT_56A	B	NA [§]
PT_56B	B	NA [§]
PT_57A	G	G
PT_57B	G	G
PT_58A	C	C
PT_58B	C	C
PT_60A	G	G
PT_60B	G	G
PT_65A	G	G
PT_65B	G	G
PT_66A	G	G
PT_66B	G	G
PT_68A	B	NA [§]
PT_68B	NA	NA [§]
PT_71A	C	C
PT_71B	C	C
PT_72A	NA	02_AG
PT_72B	NA	02_AG

Anexo 5.A

Listagem dos polimorfismos genéticos identificados nas 67 sequências em estudo, associados ou não a resistência aos inibidores da protease. A tabela apresenta a comparação entre sequências do subtipo B e não-B. As mutações assinaladas a vermelho e verde estão associadas ao subtipo B e não-B, respectivamente, com base na aplicação do teste de Fisher bilateral ($p < 0,01$).

Mutações	Subtipo B (n=8)	Subtipos não-B (n=59)	<i>p</i>
P9L	0	1	1,000
L10I	1	6	1,000
L10V	0	2	1,000
T12A	0	1	1,000
T12I	0	2	1,000
T12P	2	0	0,013
T12S	0	8	0,582
I13A	0	2	1,000
I13V	0	36	0,001
K14R	3	27	0,722
I15V	0	14	0,189
G16E	0	3	1,000
G16K	0	1	1,000
G17E	0	2	1,000
Q18E	0	1	1,000
Q18H	0	1	1,000
Q18Y	0	1	1,000
L19I	3	6	0,068
L19P	0	1	1,000
L19Q	0	1	1,000
L19S	0	1	1,000
L19T	0	5	1,000
K20I	0	35	0,002
K20R	0	5	1,000
E21K	0	2	1,000
A22V	0	2	1,000
D25G	0	1	1,000
D25N	0	1	1,000
D30N	0	1	1,000
L33F	0	2	1,000
L33V	0	2	1,000
E35D	2	33	0,139
E35K	0	1	1,000
E35N	0	2	1,000
M36I	0	53	4,6E-07
M36L	5	0	5,8E-06
N37D	2	5	0,193
N37E	0	2	1,000
N37H	2	0	0,013
N37S	0	2	1,000
N37T	0	4	1,000
L38I	2	0	0,013

Mutações	Subtipo B (n=8)	Subtipos não-B (n=59)	p
P39S	0	5	1,000
G40E	0	1	1,000
R41K	5	49	0,179
W42*	0	3	1,000
K43R	0	2	1,000
M46I	0	1	1,000
G48E	0	2	1,000
G49E	0	1	1,000
G51E	0	1	1,000
G51K	0	1	1,000
G52E	0	1	1,000
G52S	0	2	1,000
R57K	7	4	2,9E-06
D60E	0	2	1,000
D60G	0	1	1,000
D60N	0	1	1,000
Q61N	0	2	1,000
I62V	0	6	1,000
L63P	7	10	1,5E-04
L63S	0	3	1,000
L63T	0	3	1,000
L63V	0	2	1,000
I64L	0	1	1,000
I64M	0	2	1,000
I64V	0	2	1,000
I66V	0	1	1,000
C67E	0	2	1,000
C67Y	0	3	1,000
G68K	0	1	1,000
G68R	0	2	1,000
H69K	0	48	1,2E-05
K70R	0	6	1,000
A71T	2	0	0,013
A71V	0	3	1,000
I72T	0	4	1,000
I72V	0	1	1,000
G73R	0	3	1,000
T74R	1	0	0,119
T74S	0	2	1,000
V77I	6	6	2,2E-04
V82I	0	29	0,008
G86R	0	2	1,000
L89I	0	5	1,000
L89M	0	44	7,5E-05
T91A	0	2	1,000
I93F	0	1	1,000
I93L	6	15	0,009
G94S	0	1	1,000
F99L	0	6	1,000

Anexo 5.B

Listagem dos polimorfismos genéticos identificados nas 67 sequências em estudo, associados ou não a resistência aos inibidores da protease. A tabela apresenta a comparação entre sequências do subtipo C e não-C. As mutações assinaladas a vermelho e verde estão associadas ao subtipo C e não-C, respectivamente, com base na aplicação do teste de Fisher bilateral ($p < 0,01$).

Mutações	Subtipo C (n=10)	Subtipos não-C (n=57)	<i>p</i>
P9L	1	0	0,149
L10I	0	7	0,583
L10V	0	2	1,000
T12A	0	1	1,000
T12I	0	2	1,000
T12P	0	2	1,000
T12S	8	0	6,9E-09
I13A	0	2	1,000
I13V	1	35	0,004
K14R	0	30	0,002
I15V	8	6	1,7E-05
G16E	0	3	1,000
G16K	0	1	1,000
G17E	0	2	1,000
Q18E	0	1	1,000
Q18H	0	1	1,000
Q18Y	0	1	1,000
L19I	5	4	0,002
L19P	0	1	1,000
L19Q	0	1	1,000
L19S	0	1	1,000
L19T	5	0	2,6E-05
K20I	0	35	2,6E-04
K20R	2	3	0,157
E21K	0	2	1,000
A22V	0	2	1,000
D25G	0	1	1,000
D25N	0	1	1,000
D30N	0	1	1,000
L33F	0	2	1,000
L33V	0	2	1,000
E35D	2	33	0,039
E35K	0	1	1,000
E35N	0	2	1,000
M36I	8	45	1,000
M36L	0	5	1,000
N37D	0	7	0,583
N37E	0	2	1,000
N37H	0	2	1,000
N37S	2	0	0,020
N37T	2	2	0,103
L38I	0	2	1,000

Mutações	Subtipo C (n=10)	Subtipos não-C (n=57)	<i>p</i>
P39S	0	5	1,000
G40E	0	1	1,000
R41K	7	47	0,394
W42*	1	2	0,389
K43R	0	2	1,000
M46I	0	1	1,000
G48E	1	1	0,278
G49E	1	0	0,149
G51E	1	0	0,149
G51K	0	1	1,000
G52E	0	1	1,000
G52S	0	2	1,000
R57K	0	11	0,195
D60E	0	2	1,000
D60G	0	1	1,000
D60N	0	1	1,000
Q61N	0	2	1,000
I62V	0	6	0,580
L63P	2	15	1,000
L63S	0	3	1,000
L63T	1	2	0,389
L63V	0	2	1,000
I64L	0	1	1,000
I64M	0	2	1,000
I64V	0	2	1,000
I66V	1	0	0,149
C67E	0	2	1,000
C67Y	0	3	1,000
G68K	1	0	0,149
G68R	0	2	1,000
H69K	10	38	0,052
K70R	0	6	0,580
A71T	0	2	1,000
A71V	1	2	0,389
I72T	0	4	1,000
I72V	1	0	0,149
G73R	0	3	1,000
T74R	0	1	1,000
T74S	2	0	0,020
V77I	0	12	0,188
V82I	0	29	0,004
G86R	1	1	0,278
L89I	0	5	1,000
L89M	10	34	0,012
T91A	0	2	1,000
I93F	0	1	1,000
I93L	10	11	1,4E-06
G94S	0	1	1,000
F99L	0	6	0,580

Anexo 5.C

Listagem dos polimorfismos genéticos identificados nas 67 sequências em estudo, associados ou não a resistência aos inibidores da protease. A tabela apresenta a comparação entre sequências do subtipo G e não-G. As mutações assinaladas a vermelho e verde estão associadas ao subtipo G e não-G, respectivamente, com base na aplicação do teste de Fisher bilateral ou, se apropriado, o teste de χ^2 ($p < 0,01$).

Mutações	Subtipo G (n=32)	Subtipos não-G (n=35)	<i>p</i>
P9L	0	1	1,000
L10I	4	3	0,701
L10V	0	2	0,493
T12A	0	1	1,000
T12I	0	2	0,493
T12P	0	2	0,493
T12S	0	8	0,005
I13A	0	2	0,493
I13V	32	4	0,000*
K14R	25	5	0,000*
I15V	1	13	0,001*
G16E	0	3	0,240
G16K	0	1	1,000
G17E	2	0	0,224
Q18E	1	0	0,478
Q18H	0	1	1,000
Q18Y	0	1	1,000
L19I	0	9	0,002
L19P	0	1	1,000
L19Q	1	0	0,478
L19S	0	1	1,000
L19T	0	5	0,054
K20I	31	4	0,000*
K20R	0	5	0,054
E21K	1	1	1,000
A22V	0	2	0,493
D25G	1	0	0,478
D25N	0	1	1,000
D30N	0	1	1,000
L33F	2	0	0,224
L33V	0	2	0,493
E35D	28	7	0,000*
E35K	0	1	1,000
E35N	1	1	1,000
M36I	32	21	0,000*
M36L	0	5	0,054
N37D	1	6	0,108
N37E	0	2	0,493
N37H	0	2	0,493
N37S	0	2	0,493
N37T	0	4	0,115
L38I	0	2	0,493

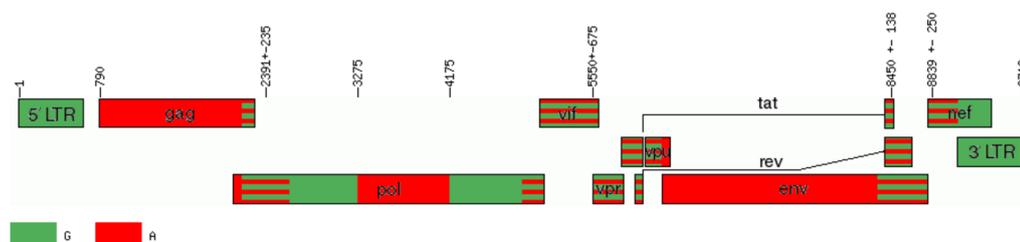
Mutações	Subtipo G (n=32)	Subtipos não-G (n=35)	<i>p</i>
P39S	1	4	0,358
G40E	0	1	1,000
R41K	31	23	0,001*
W42*	0	3	0,240
K43R	0	2	0,493
M46I	0	1	1,000
G48E	0	2	0,493
G49E	0	1	1,000
G51E	0	1	1,000
G51K	0	1	1,000
G52E	0	1	1,000
G52S	1	1	1,000
R57K	2	9	0,032*
D60E	0	2	0,493
D60G	1	0	0,478
D60N	0	1	1,000
Q61N	0	2	0,493
I62V	2	4	0,675
L63P	2	15	0,001*
L63S	0	3	0,240
L63T	0	3	0,240
L63V	0	2	0,493
I64L	0	1	1,000
I64M	2	0	0,224
I64V	1	1	1,000
I66V	0	1	1,000
C67E	2	0	0,224
C67Y	0	3	0,240
G68K	0	1	1,000
G68R	0	2	0,493
H69K	32	16	0,000*
K70R	4	2	0,414
A71T	0	2	0,493
A71V	0	3	0,240
I72T	1	3	0,615
I72V	0	1	1,000
G73R	3	0	0,104
T74R	0	1	1,000
T74S	0	2	0,493
V77I	0	12	0,000*
V82I	29	0	0,000*
G86R	1	1	1,000
L89I	3	2	0,664
L89M	30	14	0,000*
T91A	2	0	0,224
I93F	1	0	0,478
I93L	0	21	0,000*
G94S	0	1	1,000
F99L	6	0	0,009

* aplicado o teste de χ^2

Anexo 6

Representação do genoma mosaico das formas recombinantes circulantes CRF02_AG, CRF14_BG e CRF45_cpx (Los Alamos National Laboratory, 2015a).

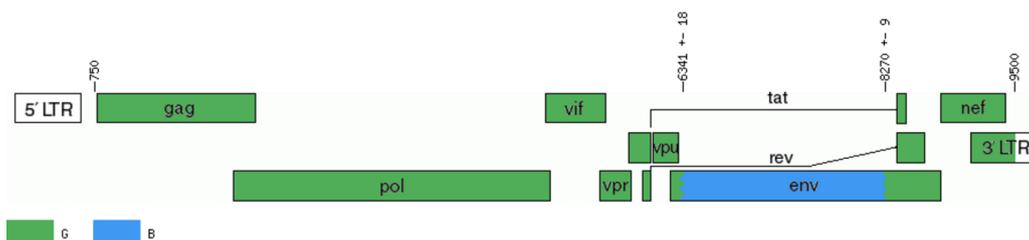
▪ CRF02_AG



Pontos de recombinação (Carr et al., 1998):

1 a 789	G
790 a 2155	A
2156 a 2625	A/G*
2626 a 3274	G
3275 a 4174	A
4175 a 4874	G
4875 a 6224	A/G
6225 a 8311	A
8312 a 9089	A/G
9090 a 9719	G

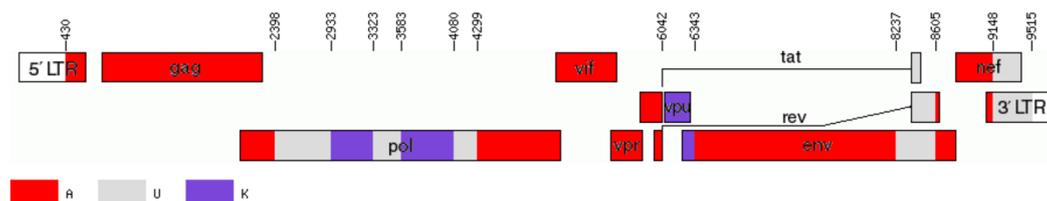
▪ CRF14_BG



Pontos de recombinação (Delgado et al., 2002):

750 a 6322	G*
6323 a 6359	B/G
6360 a 8260	B
8261 a 8279	B/G
8280 a 9500	G

▪ CRF45_cpx



Pontos de recombinação (Niama et al., 2009):

430 a 2397	A (tipo CRF09)*
2398 a 2932	U*
2933 a 3322	K (tipo CRF09)
3323 a 3582	U (tipo K)
3583 a 4079	K
4080 a 4298	U (tipo CRF04)
4299 a 6041	A
6042 a 6342	K
6343 a 8236	A
8237 a 8604	U (tipo CRF09)
8605 a 9147	A
9148 a 9515	U

* Segmentos nos quais está contida a região codificante da protease no gene *pol* (posições nucleotídicas 2253 a 2550, em HXB2).

Anexo 7

Nomenclatura dos aminoácidos, utilizando o código IUPAC de uma e três letras.

Nomenclatura dos aminoácidos		
Ácido aspártico	Asp	D
Ácido glutâmico	Glu	E
Alanina	Ala	A
Arginina	Arg	R
Asparagina	Asn	N
Cisteína	Cys	C
Fenilalanina	Phe	F
Glicina	Gly	G
Glutamina	Gln	Q
Histidina	His	H
Isoleucina	Ile	I
Leucina	Leu	L
Lisina	Lys	K
Metionina	Met	M
Prolina	Pro	P
Serina	Ser	S
Tirosina	Tyr	Y
Treonina	Thr	T
Triptofano	Trp	W
Valina	Val	V