



INSTITUTO DE HIGIENE E
MEDICINA TROPICAL
DESDE 1902

Universidade Nova de Lisboa
Instituto de Higiene e Medicina Tropical

Análise genética e funcional dos genes *pol* e *env* e
caracterização da resposta em anticorpos
neutralizantes em adultos infetados por VIH-1 de
Moçambique

Paloma Catarina Barradas Gonçalves

**DISSERTAÇÃO PARA A OBTENÇÃO DO GRAU DE
MESTRE EM CIÊNCIAS BIOMÉDICAS**

JULHO 2021



INSTITUTO DE HIGIENE E
MEDICINA TROPICAL
DESDE 1902

Universidade Nova de Lisboa
Instituto de Higiene e Medicina Tropical

Análise genética e funcional dos genes *pol* e *env* e
caracterização da resposta em anticorpos neutralizantes
em adultos infetados por VIH-1 de Moçambique

Autor: Paloma Catarina Barradas Gonçalves

Orientador: Prof. Doutor Nuno Taveira

Coorientador: Prof. Doutor João Piedade

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do
grau de Mestre em Ciências Biomédicas

“Don’t be afraid of hard work. Nothing worthwhile comes easily. Don’t let others discourage you or tell you that you can’t do it. In my day I was told women didn’t go into chemistry. I saw no reason why we couldn’t.”

Gertrude B. Elion

Agradecimentos

Esta jornada foi marcada por um ano que pareceu uma volta na montanha russa, dada a adrenalina, a ansiedade, o receio, a euforia e a **felicidade** com que foi vivida. Acreditar e terem acreditado em mim foi a força para chegar até aqui.

Agradeço desde já ao projeto RIA2016MC-1615 que financiou o meu trabalho (Título: *Neonatal HIV early infant diagnosis (EID) versus standard of care EID – Impact on inFant hEalth (LIFE): a feasibility study of point-of care testing at birth vs at 6 weeks of age, on the uptake of ART and infant prophylaxis, and on rates of infant survival, morbidity and retention in care*), cujo sponsor é European & Developing Countries Clinical Trials Partnership (EDCTP), UE.

Gostaria ainda de agradecer a todas as pessoas que me acompanharam e contribuíram para que este trabalho fosse possível.

Prof. Doutor Nuno Taveira, agradeço a oportunidade de realizar a minha dissertação numa das minhas áreas de eleição e por acreditar nas minhas capacidades. Obrigada pela sua orientação, pelo apoio, pelas ideias e por ser sempre tão otimista. Obrigada pela oportunidade de evoluir enquanto profissional e por me mostrar que a persistência e o **esforço** compensam sempre.

Rute Marcelino, não tenho palavras para agradecer todo o seu apoio, atenção e preocupação. Obrigada pela sua prontidão tão genuína em ajudar e em dar sempre o **melhor de si**. Tive muita sorte em tê-la encontrado no meu percurso. Enriqueceu o meu caminho a nível profissional, mas também pessoal. Mais uma vez obrigada por tudo.

José Marcelino, um profissional fantástico e uma pessoa igualmente brilhante. Obrigada por ter orientado os meus primeiros passos no mundo da ciência e por me ter continuado a guiar até mais uma meta. É um **privilegio** trabalhar e aprender consigo.

Aos meus colegas de laboratório, Inês Bártolo, Inês Moranguinho, Susana Tracana, Marta Calado, Rita Elias e Rita Ferreira pela amizade, pelo apoio, por todos os conselhos e pelos momentos de **convívio** que tornaram o meu percurso mais feliz. Agradeço também a todas as pessoas do departamento de imunologia e microbiologia da Faculdade de Farmácia que me proporcionaram as melhores condições de trabalho.

Aos meus pais, Matilde e Paulo, por todo o investimento na minha educação e formação profissional e pessoal. Há pessoas que acordam a nossa vontade de nos tornarmos sempre alguém melhor e sem dúvida vocês são essas pessoas para mim. Sem vocês não poderia ter chegado aqui enquanto profissional e agradeço imenso por isso, mas agradeço ainda mais por me terem ensinado o respeito pelo outro, a bondade, a generosidade e um coração que sabe ser grato todos os dias pelo **amor** que sempre me deram.

Aos meus avós, Margarida e Inácio, obrigada. Para vocês faltam-me as palavras, porque não encontro forma exprimir o que sinto e o quanto vos sou grata por tudo. Vocês representam a minha infância, o meu percurso, a minha vida até aqui... representam tudo o que sou hoje. Avós, vocês são canção mais bonita, a história de encantar que termina sempre com “felizes para sempre”, são o sorriso de cada regresso, o abraço reconfortante, a certeza de que nunca estarei sozinha e que serei para sempre aquela **criança feliz**. Obrigada por tudo.

Aos meus amigos, em especial, Marlene, Sara, Raquel, Tiago, Paulo, João, André, Bárbara e Maria por estarem comigo ao longo de todos estes anos. Esta é só mais uma das maratonas que já terminámos juntos e em que acabamos a **admirar-nos** mais e a torcer para que a vida nos reserve coisas fantásticas. Obrigada amigos!

Ao Tiago, obrigada por teres sido o meu companheiro da volta na montanha russa e teres vivido comigo todos altos e baixos, mesmo sem comprar bilhete. Foste imprescindível para me conseguir superar a mim própria e lutar pelos meus sonhos e ambições. Consegui porque estavas **sempre** lá para dizer que era capaz mesmo quando eu não acreditava que era. Um apoio incondicional. Estes anos só me mostraram que serás sempre o meu companheiro das voltas na montanha russa da vida. Obrigada.

Resumo

No início dos anos 80, o VIH-1 foi introduzido em Moçambique e desde então houve um aumento exponencial do número de casos de infeção, atingindo em 2020 cerca de 2 200 000 pessoas infetadas, o que equivale a uma prevalência de 12,9% (11,9%-14,4%). Neste trabalho, caracterizou-se a diversidade genética das estirpes de VIH-1 que circulam em Maputo e a frequência e distribuição de polimorfismos genéticos nas regiões codificantes das enzimas transcriptase reversa (RT), protease (PR) e integrase (IN) associados ou não a resistência aos antirretrovirais em uso clínico. Analisou-se pela primeira vez a resposta de anticorpos neutralizantes e não neutralizantes de indivíduos moçambicanos infetados pelo VIH-1. Foram obtidas sequências da região C2V3C3 (N=58) do gene *env* e das regiões codificantes da RT (N=26) PR (N=34) e IN (N=22) do gene *pol*. A análise filogenética mostrou a circulação maioritária de vírus do subtipo C de fenótipo R5 (86%, N=50). Observou-se a introdução de outros subtipos, tais como A, B, D e CRF02_AG e outros recombinantes. A análise das sequências aminoacídicas da RT permitiu identificar seis mutações associadas a resistência aos NRTI (K65R, D67N, K70R, M184V, T215V e K219E) e quatro mutações principais (K103N, V106A, Y181C e G190A) e quatro mutações acessórias (A98G, V108I, E138A e F227L) associadas à resistência aos NNRTI. Estes resultados sugerem que 27% (7/26) dos indivíduos se encontram infetados com vírus potencialmente resistentes aos regimes terapêuticos de primeira linha definidos em Moçambique e 12% (3/26) aos regimes terapêuticos de segunda linha. Para a classe dos PI, foram identificadas quatro mutações acessórias (I47R, Q58E, G73S e L10F) e para a classe dos INSTI apenas uma mutação acessória (G163R). Em todas as sequências identificou-se um número considerável de polimorfismos genéticos, o que é consistente com a grande diversidade genética do subtipo C. Avaliou-se a resposta dos anticorpos de ligação para polipéptidos recombinantes C2V3C3 dos subtipos B, C, G, H e CRF02_AG. A maioria dos doentes (77%) reagiu contra o polipéptido CRF02_AG, mas houve também uma forte reatividade (>40%) contra os polipéptidos G, C e H, confirmando a forte antigenicidade da região C2V3C3 do invólucro do VIH-1. A maioria dos indivíduos possuíam anticorpos capazes de neutralizar o isolado VIH-1 NL4.3 (subtipo B), seguido dos pseudovírus TRO11 (subtipo B), BJOX2000 (CRF07_BC) e CE1176 (subtipo C). Observou-se uma correlação entre o espectro e a potência de neutralização e ainda uma associação inversa entre o número de locais de N-glicosilação e a potência de neutralização. Os indivíduos com atividade

neutralizante tinham uma carga viral 1,4 vezes mais baixa ($P= 0,0391$) que os indivíduos sem anticorpos neutralizantes, sugerindo que estes anticorpos contribuem para o controlo da carga viral nesta população. Em conclusão, os nossos resultados mostram que a epidemia VIH-1 em Moçambique está a evoluir rapidamente para uma maior complexidade genética e epidemiológica, acompanhada pelo aumento de mutações de resistência que tornam ineficientes os regimes terapêuticos de primeira e segunda linha em utilização em Moçambique. A introdução dos INSTI no programa nacional de tratamento poderia contribuir para o controlo do VIH no país. Estudos adicionais são necessários para definir o perfil de neutralização em indivíduos infetados por VIH-1 em Moçambique e para identificar os principais determinantes de neutralização.

Palavras-chave: Infecção VIH-1; Moçambique; Genotipagem; Mutações de resistência e Polimorfismos; Tropicismo viral; Anticorpos de ligação; Anticorpos neutralizantes.

Abstract

In the early 1980s HIV-1 was introduced in Mozambique and since then there has been an exponential increase in the number of cases of infection reaching around 2.200.000 infected people in 2020, equivalent to a prevalence of 12.9% (11.9%-14.4%). In this work we characterized the genetic diversity of HIV-1 strains circulating in Maputo as well as the frequency and distribution of genetic polymorphisms in the coding regions of the reverse transcriptase (RT), protease (PR) and integrase (IN) enzymes commonly associated with resistance to antiretroviral drugs in clinical use. In addition, the neutralizing and non-neutralizing antibody response of HIV-1 infected Mozambican patients was analyzed for the first time. Sequences of C2V3C3 (N=58), RT (N=26) PR (N=34) and IN (N=22) genes were obtained and phylogenetic analysis showed that the majority of circulating viruses belong to subtype C and have with the R5 phenotype (86%, N=50). Other subtypes, such as A, B, CRF02_AG and other recombinants were also detected. The analysis of RT amino acid sequences allowed the identification of six mutations associated with NRTI resistance (K65R, D67N, K70R, M184V, T215V and K219E), and four primary mutations (K103N, V106A, Y181C and G190A) secondary mutations (A98G, V108I, E138A and F227L) associated with resistance to the NNRTIs. These results show that 27% (7/26) of the individuals are infected with viruses potentially resistant to first line therapeutic regimens in use in Mozambique and 12% (3/26) to the second line therapeutic regimen. For the IP class, four secondary mutations were identified (I47R, Q58E, G73S and L10F) while for the INSTI class only one secondary mutation was detected (G163R). A considerable number of genetic polymorphisms was identified in all sequences, which is consistent with the great genetic diversity of subtype C. Binding antibody reactivity to C2V3C3 recombinant polypeptides of subtypes B, C, G, H and CRF02_AG was evaluated. Most patients (77%) reacted against CRF02_AG polypeptide, and there were also many patients (>40%) reacting with G, C and H polypeptides, confirming the C2V3C3 region as a dominant antigenic region in the HIV-1 envelope. Most individuals had antibodies able to neutralize the HIV-1 NL4.3 isolate (subtype B), followed by the pseudovirus TRO11 (subtype B), BJOX2000 (CRF07_BC) and CE1176 (subtype C). There was a correlation between the neutralization breadth and potency and an inverse association between the number of N-glycosylation sites and neutralization potency. Individuals with neutralizing activity had a viral load 1.4 times lower ($P = 0.0391$) than individuals without neutralizing antibodies, suggesting that these

antibodies contribute to the control of viral load in this population. In conclusion, our results show that the HIV-1 epidemic in Mozambique is rapidly evolving in terms of genetic and epidemiological complexity, and that there is an increase in the prevalence of resistance mutations that render Mozambique's first and second line therapeutic regimes inefficient. The introduction of INSTI in the national treatment programme could contribute to HIV control in the country. Additional studies are needed to define the neutralization profile in HIV-1 infected individuals in Mozambique and to identify the main determinants of neutralization.

Keywords: HIV-1 infection; Mozambique; Genotyping; Resistance mutations and Polymorphisms; Viral tropism; Binding antibodies; Neutralizing antibodies.

Lista de acrónimos e abreviaturas

ABC	Abacavir
ADCC	Do inglês, <i>Antibody-dependent cell-mediated cytotoxicity</i>
ADCVI	Do inglês, <i>Antibody-dependent cell-mediated virus inhibition</i>
ADN	Ácido desoxirribonucleico
APOBEC	Do inglês, <i>Apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like</i>
ARN	Ácido ribonucleico
ARV	Antirretroviral
ATV	Atazanavir
AZT	Azidotimidina (Zidovudina)
BLAST	Do inglês, <i>Basic Local Alignment Search Tool</i>
BIC	Bictegravir
bNAb	Anticorpo amplamente neutralizante
CA	Proteína da cápside do VIH-1
CCD	Domínio do centro ativo
CCR5	Recetor de β -quimiocinas da classe CC
CD4	Do inglês, <i>cluster of differentiation 4</i>
CD4bs	Do inglês, <i>CD4 binding site</i>
CDC	Do inglês, <i>Center for disease control and prevention</i>
CDR	Regiões determinantes de complementaridade
CRF	Formas recombinantes circulantes
CTD	Domínio C-terminal
CTL	Linfócitos T citotóxicos
CXCR4	Recetor de β -quimiocinas da classe CXC
D4T	Estavudina
ddI	Didanosina
dNTPs	Desoxirribonucleósidos 5'-trifosfatados
DOR	Doravirina
DRM	Do inglês, <i>drug resistance mutation</i>
DRV	Darunavir
DTG	Dolutegravir
EFV	Efavirenz
EI	Inibidores de Entrada
ETR	Etravirina
EUA	Estados Unidos da América

EVG	Elvitegravir
Fab	Fragmento de ligação ao antígeno
Fc	Fragmento cristalizável
FCR	Recetores Fc
FDA	Do inglês, <i>Food and Drug Administration</i>
FDC	Células dendríticas foliculares
FPV	Fosamprenavir
FTC	Emtricitabina
FTR	Fostemsavir
GC	Centro germinativo
HAART	Do inglês, <i>highly active antiretroviral therapy</i>
HLA	Do inglês, <i>human leukocyte antigen</i>
HR1	Do inglês, <i>heptad repeat region 1</i>
HR2	Do inglês, <i>heptad repeat region 2</i>
HSC	Células Estaminais hematopoiéticas
IDV	Indinavir
Ig	Imunoglobulina
IN	Integrase do VIH-1
INSTI	Do inglês, <i>integrase strand-transfer inhibitors</i>
IUPAC	Do inglês, <i>International Union of Pure and Applied Chemistry</i>
LAV	Do inglês, <i>Lymphadenopathy-associated Virus</i>
LLPC	Do inglês, long-lived plasma cells
LPV	Lopinavir
LTR	Repetição terminal longa
MA	Proteína da matriz do VIH-1
MDR	Do inglês, <i>multidrug resistance</i>
MEGA	Do inglês, <i>Molecular Evolutionary Genetics Analysis</i>
MHC	Complexo principal de histocompatibilidade
MPER	Região proximal externa da membrana
MRV	Maraviroc
NAb	Anticorpo neutralizante
NC	Proteína da nucleocápside do VIH-1

NFV	Nelfinavir
NHP	Primatas não humanos
NNRTI	Do inglês, <i>non-nucleoside reverse transcriptase inhibitors</i>
NRTI	Do inglês, <i>nucleoside reverse transcriptase inhibitors</i>
NTD	Domínio N-terminal
NVP	Neviparina
OMS	Organização Mundial da Saúde
PAMP	Do inglês, <i>Pathogen-associated molecular patterns</i>
PCR	Reação em cadeia de polimerase
PI	Do inglês, <i>protease inhibitors</i>
PIC	Complexo de pré-integração
PR	Protease do VIH-1
PRR	Recetores de reconhecimento de padrões
RAL	Raltegravir
RNase H	Ribonuclease H
RPV	Rilpivirina
RT	Transcriptase Reversa do VIH-1
SIDA	Síndrome da Imunodeficiência Adquirida
SLPC	Do inglês, <i>short-lived plasma cells</i>
SQV	Saquinavir
SU	Glicoproteína de superfície do VIH-1
TAM	Do inglês, <i>thymidine analogue mutations</i>
TARV	Terapia antirretroviral
TDF	Tenofovir
Tfh	Células T auxiliares foliculares
TM	Glicoproteína transmembranar do HIV-1
TPV	Tipranavir
UNAIDS	Do inglês, <i>Joint United Nations Programme on HIV/AIDS</i>
URF	Formas recombinantes únicas
VIH-1	Vírus da imunodeficiência humana tipo 1
VIH-2	Vírus da imunodeficiência humana tipo 2
VIS	Vírus da imunodeficiência símia
3TC	Lamivudina

Índice

<i>Agradecimentos</i>	<i>I</i>
<i>Resumo</i>	<i>III</i>
<i>Abstract</i>	<i>V</i>
<i>Lista de acrónimos e abreviaturas</i>	<i>VII</i>
<i>1. Introdução</i>	<i>1</i>
1.1 Descoberta do VIH	2
1.2 Origem biológica do VIH	2
1.3 Diversidade genética do VIH-1	3
1.3.1 Classificação do VIH-1	3
1.3.2 Dados estatísticos da pandemia	5
1.4 Genoma e estrutura do VIH-1	7
1.4.1 Gene <i>pol</i>	9
.....	11
1.4.2 Gene <i>env</i>	14
1.5 Replicação Viral	16
1.5.1 Tropismo Celular e Recetores.....	16
1.5.2 Ciclo Replicativo	17
1.6 História natural da infeção por VIH-1	19
1.6.1 Vias de transmissão	19
1.6.2 Patofisiologia e marcadores de infeção	19
1.7.1 Imunidade Inata	21
1.7.2 Imunidade Adaptativa.....	22
1.8 Terapia antirretroviral	30
1.8.1 Classes de antirretrovirais	30
1.8.1.1 Inibidores de entrada (EI)	31
1.8.1.3 Inibidores da RT não análogos a nucleósidos (NNRTI).....	32
1.8.1.4 Inibidores da integrase (INSTI)	32
1.8.1.5 Inibidores da protease (PI)	32
1.8.2 Regime terapêutico	33
1.8.4 Mutações de resistência em cada classe de antirretrovirais	35
1.8.5 Polimorfismos genéticos naturais	42
1.10 Epidemiologia da infeção VIH/SIDA em Moçambique.....	42
1.10.1 Dados estatísticos.....	42
1.10.2 Subtipos circulantes no país.....	44
1.10.3 Terapia antirretroviral	44
<i>2. Materiais e Métodos</i>	<i>48</i>
2.1 Caracterização da amostra populacional estudada	49
2.2 Extração de ARN viral a partir de plasma	51
2.3 Amplificação por <i>nested</i> PCR da região C2V3C3 do gene <i>env</i> e da região codificante das enzimas RT, PR e IN do gene <i>pol</i>	51
2.3.1 Seleção dos <i>primers</i>	51

2.3.2 Protocolo de amplificação da região codificante C2V3C3 do gene <i>env</i>	52
2.3.3 Protocolo de amplificação da região codificante das enzimas RT, PR, IN do gene <i>pol</i>	54
2.4 Observação dos fragmentos de ADN em gel de agarose.....	56
2.5 Purificação dos produtos de PCR	56
2.6 Sequenciação de Sanger.....	57
2.7 Ensaio imunoenzimático (ELISA, Enzyme Linked Immunosorbent Assay) para testar a diversidade antigénica dos doentes moçambicanos	57
2.8 Purificação e quantificação de anticorpos IgG	58
2.9 Ensaio de Neutralização da infeção viral	59
2.10 Análise bioinformática das sequências nucleotídicas.....	60
2.10.1 Edição de sequências e homologia genética.....	60
2.10.2 Subtipagem por análise filogenética.....	60
2.10.3 Análise de hipermutação e de recombinação.....	61
2.11 Análise bioinformática das sequências aminoacídicas	61
2.11.1 Análise de resistências aos antirretrovirais.....	61
2.11.2 Variabilidade das sequências aminoacídicas obtidas.....	62
2.11.4 Análise de Locais de N-glicosilação.....	63
2.11.5 Predição do tropismo viral.....	63
2.12 Análise estatística	64
3. Resultados.....	65
3.1 Análise filogenética das sequências nucleotídicas.....	66
3.1.1 Análise filogenética das sequências nucleotídicas C2V3C3	66
3.1.2 Análise filogenética das sequências nucleotídicas RT	69
3.1.3 Análise Filogenética das sequências nucleotídicas PR.....	71
3.3 Distâncias genéticas.....	76
3.3 Análise bioinformática das sequências aminoacídicas da transcriptase reversa, protease e integrase.....	78
3.3.1 Identificação de mutações associadas a resistência aos inibidores da transcriptase reversa e outros polimorfismos genéticos	78
3.3.3 Identificação de mutações associadas a resistência aos inibidores da protease e outros polimorfismos genéticos	84
3.4 Análise bioinformática das sequências aminoacídicas da região C2V3C3	87
3.4.1 Polimorfismos genéticos da <i>loop</i> V3	89
3.5 Diversidade antigénica nos doentes moçambicanos.....	93
3.6 Análise da atividade neutralizante dos anticorpos IgG.....	95
4. Discussão e conclusões.....	99
5. Bibliografia.....	108
6. Anexos.....	130

Índice de Figuras

Figura 1 :	Árvore filogenética de VIS e VIH-1.....	4
Figura 2 :	Distribuição global dos principais subtipos, CRFs e URFs do VIH-1.....	6
Figura 3 :	Mapa genómico do ADN proviral e esquema representativo de uma partícula de VIH-1.....	9
Figura 4 :	Mapa genómico do gene <i>pol</i> do VIH-1, com base na estirpe de referência HIXB2.....	10
Figura 5 :	Domínios funcionais e estruturais da enzima integrase do VIH-1.....	11
Figura 6 :	Estrutura do heterodímero p66/p51 da transcriptase reversa do VIH-1.....	12
Figura 7 :	Estrutura da protease do VIH-1.....	13
Figura 8 :	Ilustração esquemática da glicoproteína SU de VIH-1 em HXB2.....	15
Figura 9 :	Representação da organização estrutural do monómero gp120 e do trímero gp120-gp41.....	15
Figura 10:	Resumo do ciclo replicativo do VIH.....	18
Figura 11:	Decurso geral de uma infeção por VIH-1.....	20
Figura 12:	Estrutura do anticorpo.....	24
Figura 13:	Ativação e diferenciação de células B em células B de memória e células plasmáticas.....	26
Figura 14:	Representação esquemática dos mecanismos de ação dos anticorpos neutralizantes e não neutralizantes.....	27
Figura 15:	Mapeamento dos locais de atuação do bNAbs nas glicoproteínas do invólucro do VIH-1.....	29
Figura 16:	Desenho base do estudo	47
Figura 17:	Árvore Filogenética de sequências C2V3C3 de Moçambique.....	68
Figura 18:	Mapa genómico da CRF07_BC.....	69
Figura 19:	Árvore filogenética de sequências codificantes da RT de Moçambique.....	70
Figura 20:	Árvore Filogenética de sequências codificantes da PR de Moçambique.....	72
Figura 21:	Mapa genómico da CRF02_AG.....	73
Figura 22:	Árvore Filogenética de sequências codificantes da IN de Moçambique.....	74
Figura 23:	Representação gráfica da prevalência dos 56 polimorfismos genéticos identificados na RT.....	79
Figura 24:	WebLogo do alinhamento múltiplo das sequências aminoácidas MZ da enzima RT.....	81
Figura 25:	Representação gráfica da prevalência dos 45 polimorfismos genéticos identificados na PR.....	82
Figura 26:	WebLogo do alinhamento múltiplo das sequências aminoácidas MZ da enzima PR.....	84

Figura 27:	Representação gráfica da prevalência dos 63 polimorfismos genéticos identificados na IN.....	85
Figura 28:	WebLogo do alinhamento múltiplo das sequências aminoacídicas MZ da enzima IN.....	87
Figura 29:	WebLogo do alinhamento múltiplo das sequências aminoacídicas MZ da <i>loop</i> V3.....	88
Figura 30:	Alinhamento múltiplo das 57 sequências aminoacídicas da região V3 dos isolados de VIH-1 de Moçambique em estudo.....	90
Figura 31:	Representação gráfica da prevalência dos 44 polimorfismos genéticos identificados na região V3.....	91
Figura 32:	Representação gráfica da carga total de aminoácidos da região V3 em função do tropismo viral.....	92
Figura 33:	Gráfico de entropia da região V3 de isolados virais de acordo com o tropismo.....	92
Figura 34:	Frequência com que anticorpos IgG anti-C2V3C3 reagem contra determinado número de polipéptidos.....	93
Figura 35:	Frequência (%) com que anticorpos IgG anti-C2V3C3 contra os polipéptidos recombinantes.....	94
Figura 36:	Reatividade média dos anticorpos de ligação (OD/Cut-off) contra os diferentes polipéptidos recombinantes.....	95
Figura 37:	Número de isolados virais que possuem atividade neutralizante contra o painel de pseudovírus usado.....	96
Figura 38:	Correlação entre a potência e o espectro da atividade neutralizante....	96
Figura 39:	Atividade neutralizante percentual em função do número de locais de N-glicosilação da região C2V3C3.....	97
Figura 40:	Carga viral (cópias ARN/ml) dos indivíduos em função de possuírem ou não atividade neutralizante e terapia.....	98

Índice de tabelas

Tabela 1:	Regimes terapêuticos de primeira e segunda linha.....	33
Tabela 2 :	Mutações de resistência mais comuns aos NRTI.....	36
Tabela 3 :	Mutações de resistência mais comuns aos NNRTI.....	38
Tabela 4 :	Mutações de resistência mais comuns aos INSTI.....	39
Tabela 5 :	Mutações de resistência mais comuns aos PI.....	41
Tabela 6 :	Distribuição regional da prevalência de HIV-1 em Moçambique.....	43
Tabela 7 :	Registo dos regimes terapêuticos de Moçambique.....	45
Tabela 8 :	Registo de mutações de resistência.....	46
Tabela 9 :	Caracterização epidemiológica e clínica dos pacientes de Moçambique incluídos no estudo.....	49
Tabela 10:	Primers utilizados nas reações de amplificação e sequenciação das regiões codificantes das enzimas RT, PR e IN do gene <i>pol</i> e da região C2V3C3 do gene <i>env</i>	52
Tabela 11:	Condições de amplificação aplicadas na primeira reação de PCR para amplificação da região C2V3C3 do gene <i>env</i>	53
Tabela 12:	Condições de amplificação aplicadas na segunda reação de PCR para amplificação da região C2V3C3 do gene <i>env</i>	54
Tabela 13:	Condições de amplificação aplicadas na primeira reação de PCR para amplificação da região codificante das enzimas RT, PR e IN que constituem o gene <i>pol</i>	55
Tabela 14:	Condições de amplificação aplicadas na segunda reação de PCR para amplificação da região codificante das enzimas RT, PR e IN que constituem o gene <i>pol</i>	56
Tabela 15:	Pseudovírus VIH-1 utilizados nos ensaios de neutralização.....	59
Tabela 16:	Subtipos genéticos em circulação em Moçambique.....	66
Tabela 17:	Distâncias genéticas entre os isolados virais de cada região do genoma em estudo.....	76
Tabela 18:	Matrizes de distância genética entre os subtipos das diferentes regiões do genoma viral.....	77
Tabela 19:	Distâncias genéticas médias (%) dos isolados virais em cada subtipo nas diferentes regiões do genoma viral.....	77
Tabela 20:	Perfis de resistência dos isolados virais de Moçambique em estudo.....	79
Tabela 21:	Polimorfismos genéticos associados estatisticamente ($p<0,05$) aos subtipos C e D.....	80
Tabela 22:	Mutações associadas aos PI e os respetivos perfis de resistência dos isolados virais de Moçambique em estudo.....	83
Tabela 23:	Polimorfismos genéticos associados estatisticamente ($p<0,05$) aos subtipos A e C.....	85
Tabela 24:	Polimorfismos genéticos associados estatisticamente ($p<0,05$) aos subtipos C e D.....	86

1. Introdução

1.1 Descoberta do VIH

O vírus da imunodeficiência humana (VIH), como agente etiológico da sida remonta a Junho de 1981 nos Estados Unidos da América com a descrição de casos clínicos de pneumonia por *Pneumocystis carinii* em jovens homossexuais do sexo masculino [1-5]. Entre 1982 e 1983, a sida foi identificada no Haiti, África e Europa [6-8], em grupos populacionais específicos, nomeadamente, utilizadores de drogas injetáveis, hemofílicos, praticantes de relações sexuais desprotegidas e bebés infetados por transmissão vertical [9-11]. O principal agente etiológico da sida, o VIH-1, foi isolado em 1983 por Luc Montaigner e Françoise Barré-Sinoussi do Instituto Pasteur de Paris, a partir de um gânglio linfático de um doente com linfadenopatias generalizadas [12-14]. Este novo isolado correspondia a um retrovírus, o qual foi designado de LAV (do inglês *lymphadenopathy-associated virus*) [12]. A sida tornou-se pandémica muito rapidamente e nos últimos meses de 1986, oitenta e cinco países distribuídos por África, América, Ásia, Europa e Oceânia, já haviam relatado casos de indivíduos com infeção por VIH-1 e sida à Organização Mundial da Saúde [15]. Em 1985, um novo retrovírus foi isolado e caracterizado em indivíduos com sida da Guiné-Bissau e Cabo Verde, com características morfológicas, estruturais e imunológicas semelhantes ao então designado VIH, mas com algumas diferenças a nível genético e antigénico. Este novo vírus foi classificado como vírus da imunodeficiência humana tipo 2 (VIH-2) [16,17].

1.2 Origem biológica do VIH

O vírus da imunodeficiência humana (VIH) teve origem em múltiplos eventos de transmissão de carácter zoonótico do vírus da imunodeficiência simia (VIS) de primatas não humanos para humanos [18]. Esta transmissão aconteceu no continente africano, onde são encontradas cerca de 45 espécies de primatas infetados naturalmente por diversos tipos de vírus da imunodeficiência simia [19]. A elevada prevalência de VIS nos seus hospedeiros naturais sem causar doença é indicativo de um longo período de circulação desses vírus nas respetivas espécies, o que reforça a sua ancestralidade em relação ao VIH [19,20]. De acordo com considerações filogenéticas e geográficas, concluiu-se que uma subespécie do chimpanzé comum (*Pan troglodytes troglodytes*) e

de gorila (*Gorilla gorilla gorilla*), infectados com VIScpzPtt e VISgor, respetivamente, terão sido o reservatório natural para a transmissão zoonótica de VIH-1 [19-21]. Em relação ao VIH-2, a sua origem biológica está intimamente relacionada com estirpes de VISmm encontradas em macacos *mangabeys* (*Cercocebus atys*) existentes na África Ocidental, entre o Senegal e Costa do Marfim, regiões caracteristicamente endémicas [22].

1.3 Diversidade genética do VIH-1

1.3.1 Classificação do VIH-1

Os estudos filogenéticos, epidemiológicos e moleculares permitiram concluir que a infeção humana por VIH-1 teve origem na transmissão cruzada de VIScpzPtt e VISgor em pelo menos quatro ocasiões independentes correspondentes aos quatro grupos de VIH-1 existentes: M (*Major*), P, O (*Outlier*) e N (*non-M, non-O*) [19,20]. Os grupos M e N descendem de VIScpzPtt transmitido dos chimpanzés para os humanos no sul dos Camarões, enquanto o grupo P descende de VISgor existente em gorilas-do-ocidente. O grupo O, está mais relacionado filogeneticamente com VISgor do que com VIScpzPtt, no entanto não é claro se o precursor imediato do vírus humano infetou chimpanzés ou gorilas (Figura 1) [18,19]. Os vírus do grupo M divergiram progressivamente na população humana, no início do século XX, aproximadamente em 1908 (1884-1924), e atualmente são responsáveis por cerca de 33 milhões de infeções a nível mundial [19]. Este grupo é composto por dez subtipos diferentes designados A, B, C, D, F, G, H, J, K e L [23-25]. Os grupos de VIH-1 divergem entre si ao nível da sequência de aminoácidos cerca de 30% em Gag e 47% em Env [23]. Os subtipos A e F do grupo M dividem-se ainda em sub-subtipos A1-A6 e F1-F2, respetivamente [19,24]. Os diferentes subtipos distinguem-se por divergência nos seus aminoácidos em cerca de 15-30% em Env, 15-20% em Gag e 10% em Pol [24]. Dentro de um sub-subtipo pode verificar-se divergência, ao nível da sequência de aminoácidos, na ordem de 8–17% em Env [24,25]. Tendo em conta as considerações anteriores, atualmente os subtipos B e D seriam considerados sub-subtipos de um único subtipo, no entanto, por razões históricas é difícil alterar as designações [20].

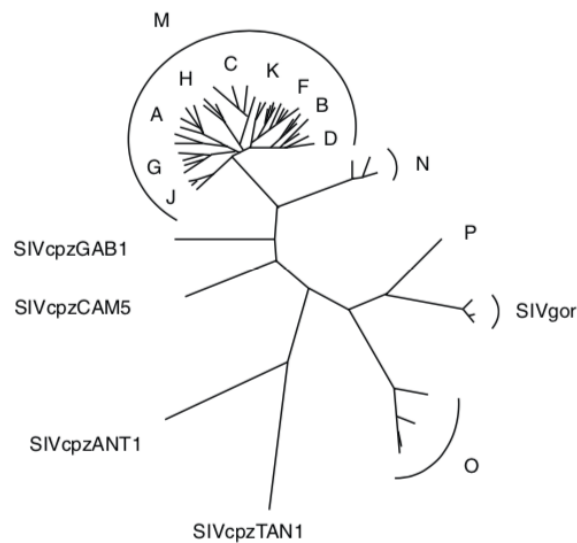


Figura 1 – Árvore filogenética de VIS e VIH-1. Árvore filogenética baseada em sequências do gene *pol* de referência de VIS e VIH-1. Encontram-se representados os grupos de VIH-1 (M, N, O e P), bem como os subtipos (A-D, F-H, J e K) do grupo M. Para o VIS dos chimpanzés foram incluídas sequências de SIVcpz. Foram incluídas também sequências representativas de SIVgor. Adaptado de [18].

O grupo M inclui ainda cerca de 102 formas recombinantes circulantes (CRF) e inúmeras formas recombinantes únicas (URF) que advêm de fenómenos de recombinação que ocorrem quando uma célula está infectada por dois ou mais vírus de subtipos diferentes [20,26]. As CRF correspondem a vírus com genomas mosaico, idênticos entre si e que foram identificados em pelo menos três indivíduos não relacionados epidemiologicamente; por outro lado, as URF correspondem a vírus com genoma mosaico que foram identificados num único indivíduo, ou num conjunto de indivíduos epidemiologicamente relacionados. As CRF constituídas por três ou mais subtipos são designadas de CRF complexas (CRF_cpx) [18]. Os grupos N e P só existem nos Camarões, onde foram descobertos em 1998 e 2009, respetivamente. Ambos possuem uma menor diversidade genética intragrupo, o que sugere uma introdução mais recente na população humana [18,20,24]. Relativamente ao grupo O, estima-se que tenha iniciado a sua dispersão paralelamente ao grupo M, no entanto, este grupo infeta apenas alguns milhares de pessoas, com maior incidência na África Ocidental, nos Camarões, Gabão e Guiné Equatorial [18,20].

1.3.2 Dados estatísticos da pandemia

Segundo os últimos dados globais de 2020 da *Joint United Programme on HIV/AIDS* (UNAIDS) existem cerca de 37,6 milhões pessoas infetadas pelo VIH dos quais 19,2 milhões são mulheres e 1,7 milhões são crianças com menos de 15 anos [27]. As regiões do globo que mais contribuem para estes resultados estatísticos são a África Oriental e Austral, que registam 20,6 milhões do total de infeções [27]. Relativamente ao número total de novas infeções verificou-se uma diminuição global desde 2010, sendo mais acentuada em crianças, nas quais se registou, em 2020, cerca de 160 000 novos casos num total de 1,5 milhões. Ainda no mesmo ano, registou-se 690 000 mortes (95 000 em crianças) por sida [27].

1.3.2.1 Distribuição global dos subtipos do grupo M, CRF e URF do VIH-1

A pandemia apresenta-se como uma rede epidemiológica complexa, bastante heterogénea e dinâmica, com diversos subtipos nas diferentes regiões geográficas, afetando grupos populacionais diferentes [28,29]. A dispersão do VIH-1, documentada em torno de Kinshasa, atual República Democrática do Congo, é sustentada pela maior diversidade genética de estirpes do vírus encontrada nessa região e nos países vizinhos, como Camarões, Angola, República Centro-Africana, Gabão e Guiné Equatorial (Figura 2) [28].

Globalmente, predomina o subtipo C, que representa cerca de 50% de todas as infeções por VIH-1 no mundo, seguido do subtipo B (12%), subtipo A (10%) e das formas recombinantes CRF02_AG (8%) e CRF01_AE (5%) [28]. Os subtipos F, H, J e K são responsáveis, em conjunto, por 1% das infeções, enquanto os subtipos G e D representam 5% e 3%, respetivamente. As restantes CRF representam cerca de 4% das infeções que, em conjunto com as URF (6%), elevam a proporção da infeção global por VIH-1 atribuível a recombinantes para os 23% (Figura 2) [28]. Nos países da África Austral, o subtipo C é o mais prevalente. Na África Oriental, o subtipo A é responsável por cerca de 53% das infeções, embora o subtipo D (17%) e o subtipo C (15%) circulem significativamente, tal como algumas CRF e URF (13%). A ocidente do continente africano, a CRF02_AG está a tornar-se predominante em vários países (46%), registando-

se uma influência significativa de estirpes virais do subtipo G e A (27%) e algumas URF (16%) [28-30].

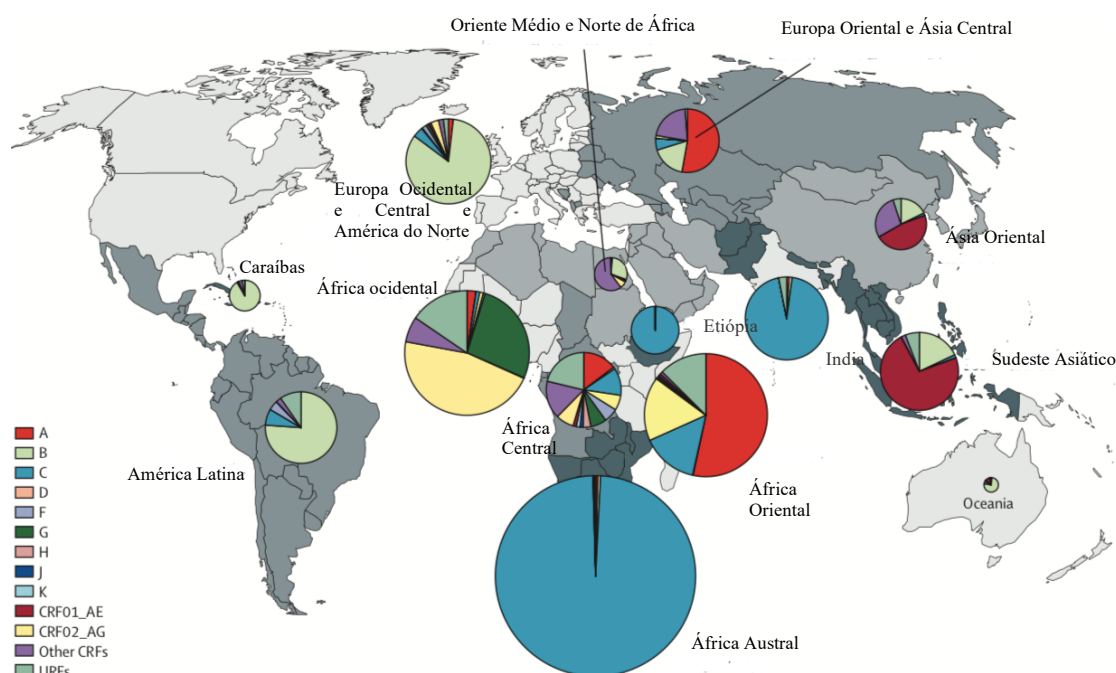


Figura 2 - Distribuição global dos principais subtipos, CRF e URF do VIH-1. Todos os países foram agrupados em 14 regiões. Os países que formam uma região possuem a mesma cor no planisfério. A área de cada gráfico corresponde ao número de pessoas infectadas pelo VIH-1. A cor de cada secção dos gráficos representa os subtipos, CRF ou URF circulantes na região. Adaptado de [28].

Na Europa, em particular na Europa Ocidental e Central, o subtipo B tem sido o mais prevalente (75%), bem como na América do Norte e Sul e na Oceânia. Ao longo do tempo, principalmente através da imigração, o subtipo F e a CRF02_AG foram também introduzidos nestas regiões [31-34]. Relativamente à Europa Oriental e Ásia Central, mais de metade das infeções são originadas por vírus do subtipo A e as restantes por vírus do subtipo B e algumas CRF [28,35]. No Sudeste Asiático têm surgido inúmeras CRF, nomeadamente CRF01_AE, CRF07_BC e CRF08_BC que são responsáveis por mais de 70% das infeções [28,29,36]. No Médio Oriente e no Norte de África verifica-se um aumento considerável do número de CRF, responsáveis por 60% do número total de infeções, acompanhadas de uma diminuição nas infeções causadas por vírus do subtipo B [28,37].

1.3.3 Mecanismos responsáveis pela diversidade genética do VIH-1

O VIH apresenta uma extensa variabilidade genética intra e interindividual, que resulta de uma combinação de vários fatores que permitem que o vírus se diversifique em inúmeras formas genéticas, adote mecanismos para subverter a resposta do sistema imunitário e desenvolva variantes resistentes aos antirretrovirais [38]. A evolução rápida do vírus resulta da progressiva acumulação de mutações durante o processo de transcrição reversa, devido à ausência de um mecanismo de *proofreading*, e através de mecanismos de recombinação que ocorrem em células co-infetadas com duas ou mais estirpes virais genotipicamente diferentes [38,39]. A frequência de recombinação é cerca de $8,3 \times 10^{-4}$ recombinações por ciclo de replicação [40] enquanto a acumulação de mutações ocorre numa frequência média de $2,4 \times 10^{-5}$ por nucleótido por ciclo de replicação e $6,4 \times 10^{-3}$ por nucleótido por ano [38,40]. A marcada heterogeneidade genética do VIH é ainda ampliada devido ao elevado número de células infetadas num indivíduo sem terapia antirretroviral ($\cong 10^7$ - 10^8), pela elevada taxa reprodutiva do vírus *in vivo* (cada célula infetada produz entre 10^2 e 10^4 viriões) e pelo número de vírus que se podem produzir diariamente nestes indivíduos (cerca de 10^9) [40]. Existe ainda a atividade de fatores mutagénicos do hospedeiro, como a APOBEC3G, que contribuem para a elevada taxa mutagénica [38]. A manifestação genotípica dos fenómenos apresentados traduz-se na constante manutenção de uma extensa variabilidade de sequências nucleotídicas do genoma viral, sendo que, por vezes, é possível encontrar num único indivíduo uma complexa população de variantes víricas heterogéneas, relacionadas geneticamente, mas com propriedades antigénicas distintas designadas por *quasispecies* [41].

1.4 Genoma e estrutura do VIH-1

O vírus da imunodeficiência humana (VIH) pertence à família *Retroviridae*, subfamília *Orthoretrovirinae* e ao género *Lentivirus* [42]. O genoma do VIH-1 é constituído por duas moléculas idênticas de ARN de cadeia simples e polaridade positiva, com uma extensão nucleotídica de cerca de 10 kb [42-44]. Por sua vez, o genoma do ADN proviral, consiste numa molécula de cadeia dupla resultante da retrotranscrição do ARN viral aquando da infeção da célula hospedeira [42]. A cadeia de ADN é flanqueada nas suas extremidades por sequências LTR (do inglês, *long terminal repeats*) – a região

5' LTR contém importantes elementos de controlo transcricional enquanto a região 3' LTR é responsável pela poliadenilação dos ARNs virais [44].

A organização do genoma é complexa e compreende nove genes que codificam as principais proteínas do virião. Estes genes estão divididos em três grupos: estruturais (*gag*, *pol*, *env*), reguladores (*tat*, *rev*) e acessórios (*vif*, *vpr*, *vpu*, *nef*) e encontram-se representados na figura 3A [42-44]. As proteínas estruturais e enzimáticas estão envolvidas no processo replicativo do vírus, sendo que as proteínas reguladoras e acessórias estão geralmente associadas a processos transcricionais e pós-transcricionais da expressão génica e também a mecanismos de evasão à resposta imunológica do hospedeiro [45,46]. O gene *gag* codifica para uma poliproteína precursora, Pr55^{Gag}, que é processada pela protease viral (PR), gerando as proteínas da matriz (MA, do inglês *matrix protein*), da cápside (CA, do inglês *capsid protein*), da nucleocápside (NC, do inglês *nucleocapside protein*) e a proteína p6. O gene *pol* codifica uma proteína precursora, Pr160^{gagpol}, cujo processamento proteolítico, pela PR, permite obter as enzimas virais: protease (PR), integrase (IN) e transcriptase reversa (RT). Por sua vez, o gene *env* codifica uma poliproteína precursora (gp160) que é posteriormente cindida por enzimas celulares no complexo de Golgi, originando as glicoproteínas gp120 (SU, do inglês *surface protein*) e gp41 (TM, do inglês *transmembrane protein*) [43-45]. Os genes reguladores, denominados *tat* e *rev*, codificam as proteínas reguladoras Tat e Rev (p19), essenciais na replicação viral [43-45]. Os genes acessórios *vif*, *vpr*, *vpu* e *nef* codificam as proteínas acessórias Vif, Vpr, Vpu e Nef, respetivamente, e possuem funções de regulação tardia, sendo importantes na infecciosidade viral e na eficiência da replicação do VIH [43-45].

As partículas virais maduras apresentam uma morfologia esférica com aproximadamente 100 a 120 nm de diâmetro e possuem um invólucro externo de natureza fosfolipídica, com origem na membrana plasmática das células infetadas (Figura 3B) [47]. Esta bicamada fosfolipídica é constituída pelas glicoproteínas de superfície gp120 e transmembranares gp41 e outras proteínas constituintes da membrana celular (e.g. HLA-I e HLA-II) [43-45]. Este invólucro lipídico envolve uma matriz viral, constituída pela proteína da matriz p17 (MA), em cujo interior se encontra a cápside viral, em forma de cone truncado, composta pela proteína p24 (CA) [42-45]. No interior da

cápside viral encontra-se o complexo ribonucleoproteico, constituído pela proteína da nucleocápside p7 (NC) associada ao genoma viral, proteínas virais de natureza enzimática (transcriptase reversa, integrase e protease), algumas proteínas acessórias e outras proteínas de origem celular (Figura 3B) [42-45].

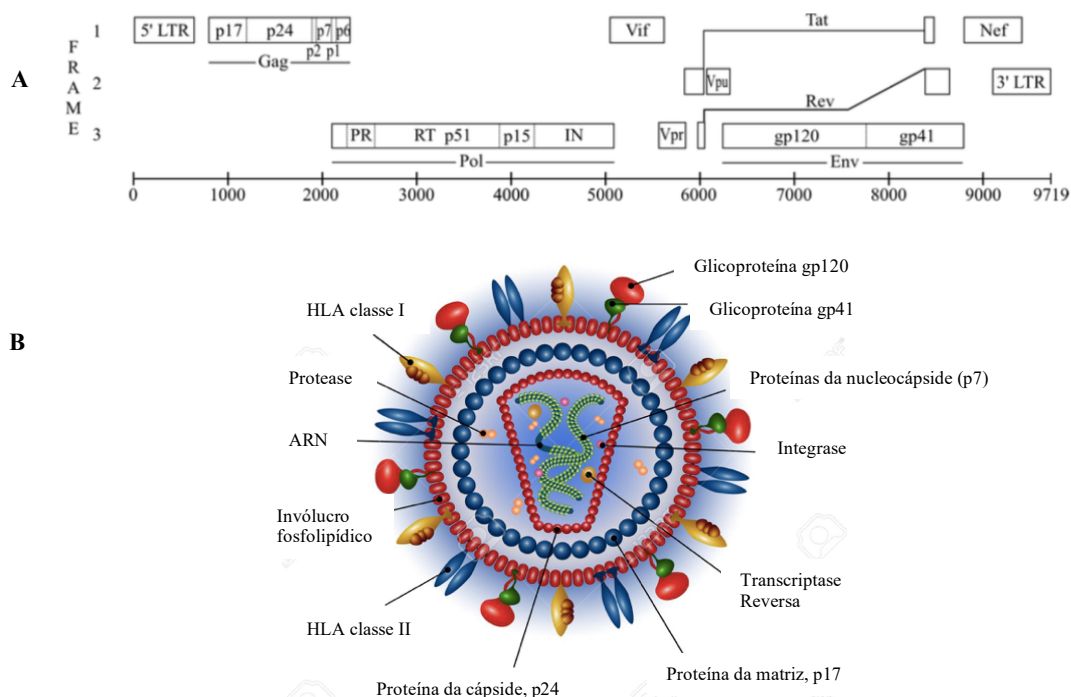


Figura 3 – Mapa genómico do ADN proviral e esquema representativo de uma partícula de VIH-1. A) Localizações relativas dos genes *gag*, *pol*, *env*, *vif*, *vpr*, *vpu*, *nef*, *tat* e *rev* e indicação das LTR em 5' e 3'. **B)** Representação esquemática da morfologia do virião do VIH-1. O invólucro lípido envolve a matriz viral em cujo interior se encontra a cápside viral que encerra o genoma viral, para além das enzimas virais (RT, IN e PR). Adaptado de [46].

1.4.1 Gene *pol*

O gene *pol* codifica as proteínas enzimáticas do VIH-1 (RT, PR e IN) e possui cerca de 3012 nucleótidos, segundo a estirpe de referência HXB2 (Figura 4). Esta região é uma das mais bem conservadas do genoma do vírus, sendo por isso, os seus produtos alvos preferenciais da terapia antirretroviral [44,45,48]. Este gene encontra-se parcialmente sobreposto ao gene *gag* numa grelha de leitura diferente. Devido à existência de um codão *stop* no final da sequência codificante do gene *gag*, é efetuada uma mudança na fase de leitura durante a tradução, para que o gene *pol* seja expresso, o que resulta numa poliproteína de fusão Pr160^{gagpol} [44]. O processamento proteolítico da poliproteína

ocorre durante o processo de maturação do virião, pela PR, resultando na separação das proteínas estruturais e enzimáticas [42,44].

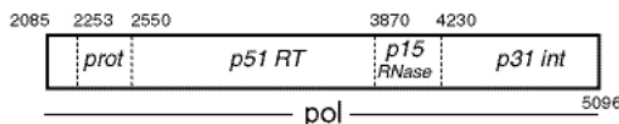


Figura 4 – Mapa genômico do gene *pol* do VIH-1, com base na estirpe de referência HXB2.

A região genômica codifica as enzimas protease, transcriptase reversa e integrase. Estas enzimas são produzidas como uma poliproteína precursora Gag-Pol, que é processada pela protease viral. Adaptado de [49].

1.4.1.1 Protease: Estrutura e função

A protease do VIH-1 pertence à família das proteases aspárticas e intervém na conversão proteolítica das poliproteínas precursoras Pr55^{gag} e Pr160^{gagpol} nas proteínas virais enzimáticas e estruturais. A sua atividade enzimática é imprescindível no processo de maturação e obtenção de partículas virais infecciosas [50]. A protease apresenta-se como um homodímero simétrico, constituído por duas subunidades idênticas de 99 aminoácidos [50]. A sua estrutura dimérica é estabilizada pela interação entre a região N-terminal de um monómero e a região C-terminal do segundo, nas posições aminoacídicas 1-4 e 96-99, sendo que a sua dissociação leva à perda da atividade catalítica (Figura 5) [51]. O centro ativo da enzima localiza-se na interface dos dois monómeros sendo composto pela tríade catalítica D25-T26-G27, altamente conservada [50,51]. Na sua forma ativa, a protease possui uma estrutura flexível em ambos os monómeros, designada de *flap* (resíduos de aminoácidos 43 a 58) que alteram a sua conformação (aberto e fechado) de modo a permitir a ligação de substratos ou inibidores ao centro ativo da enzima [50,51]. Existem outros resíduos de aminoácidos conservados no centro ativo (D25, G27, A28, D29 e D30), na região flexível *flap* (G48, G49 e I47) e na região C-terminal (I84 e P81), que são críticos para a ligação do substrato ao centro ativo da enzima. Estas regiões estão identificadas como alvos terapêuticos de inibidores da protease devido à sua importância para o correto funcionamento da enzima [50-52].

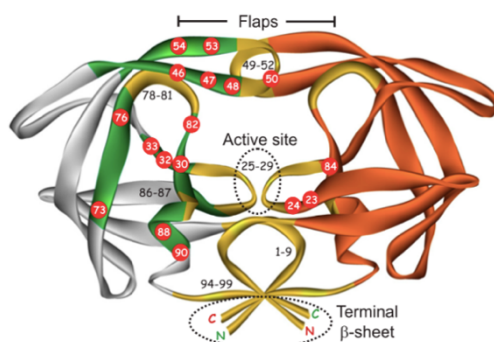


Figura 5 – Estrutura da protease do VIH-1. As regiões altamente conservadas estão representadas a amarelo, as regiões de maior variabilidade entre os subtipos do grupo M encontram-se a cinza. As regiões alvo das principais mutações de resistência (DRM) estão a verde. Os círculos vermelhos indicam as posições das principais DRM. Adaptado de [52].

1.4.1.2 Transcriptase Reversa: Estrutura e função

A transcriptase reversa do VIH-1 é uma enzima multifuncional responsável por três atividades essenciais para a replicação viral: 1) ADN polimerase ARN-dependente, responsável pela conversão do genoma de ARN de cadeia simples do vírus, em ADN; 2) RNase H, que hidrolisa a molécula de ARN genómico presente nos híbridos ARN/ADN formados durante o processo de transcrição reversa; e 3) ADN polimerase ADN-dependente, que permite a formação da cadeia de ADN complementar [53]. Esta enzima apresenta-se como um heterodímero assimétrico constituído por uma subunidade com 66 kDa (p66) e outra com 51 kDa (p51) (Figura 6) [53-55].

A subunidade p66 apresenta dois domínios catalíticos com funções distintas: domínio polimerase, que permite a síntese do ADN proviral e o domínio RNase H (15 kDa) [53,54]. A subunidade p51 deriva da excisão proteolítica da extremidade C-terminal da p66 correspondente à ribonuclease H (RNase H) levada a cabo pela PR [53,54]. As duas subunidades possuem conformações estruturais distintas entre si, sendo que a subunidade p51 possui apenas um papel estrutural que é essencial para a estabilidade e atividade da enzima [53,54]. O domínio polimerase da subunidade p66 é constituído por quatro subdomínios que são descritos com analogia à anatomia da mão humana: *fingers* (resíduos 1-85 e 118-155), *palm* (resíduos 86-117 e 156-236) e *thumb* (resíduos 237-318)

que intervêm na polimerização do ADN viral e o subdomínio *connection* (resíduos 319-426) que permite a ligação ao domínio RNase H (Figura 6) [53-55].

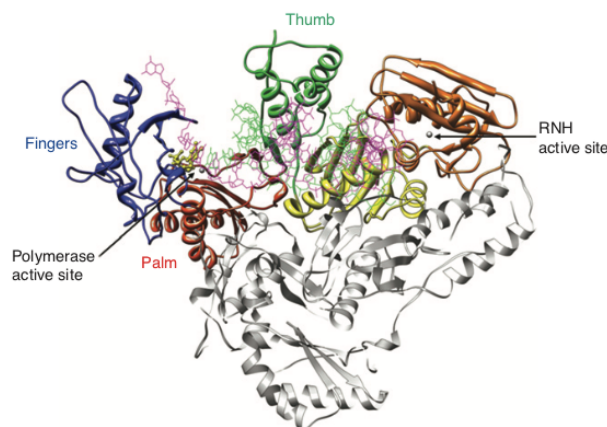


Figura 6 – Estrutura do heterodímero p66/p51 da transcriptase reversa do VIH-1. A subunidade p66 está representada a cores enquanto a subunidade p51 está em cinza. Os vários subdomínios da p66 encontram-se representados: *fingers* (azul), *palm* (vermelho), *thumb* (verde), *connection* (amarelo) e RNase H (laranja). A cadeia de ácido nucleico encontra-se a magenta. Adaptado de [55].

O centro ativo do domínio polimerase é composto por uma tríade catalítica conservada (D110, D185 e D186) no subdomínio *palm*. Esta tríade catalítica forma complexos de coordenação com catiões divalentes (Mg^{2+} , utilizado *in vivo*), essenciais à atividade da enzima [55]. Os resíduos D185 e D186 fazem parte do motivo YXDD, altamente conservado nos retrovírus (no caso do VIH-1, X corresponde a metionina) [55]. Existem outras regiões conservadas, tal como o motivo estrutural *DNA primer grip* (E224-H235) que é essencial para o correto posicionamento da extremidade 3'-OH do primer $tRNA_{Lys3}$ no centro ativo da enzima para o ataque nucleofílico ao dNTP a incorporar na cadeia nascente. O domínio RNase H está compreendido entre os aminoácidos 427 e 560 da subunidade p66 e possui duas regiões conservadas: o centro ativo, composto por quatro aminoácidos – D442, E478, D498 e D549; e o *RNase H primer grip*, composto por uma extensa rede de aminoácidos - T473, I505, K476, Q475, Y501, Q448, N474 e R500 – que interagem com as cadeias de ADN e ARN próximas do centro ativo [55].

1.4.1.3 Integrase: Estrutura e função

A integrase é uma proteína com uma massa molecular de 32 kDa, constituída por 288 aminoácidos e pertence à família das nucleotidiltransferases [56]. A enzima catalisa a integração do ADN complementar no genoma da célula hospedeira e possui três domínios funcionais: N-terminal (NTD), centro catalítico (CCD) e C-terminal (CTD), como se encontra representado na figura 7 [47,56-58].

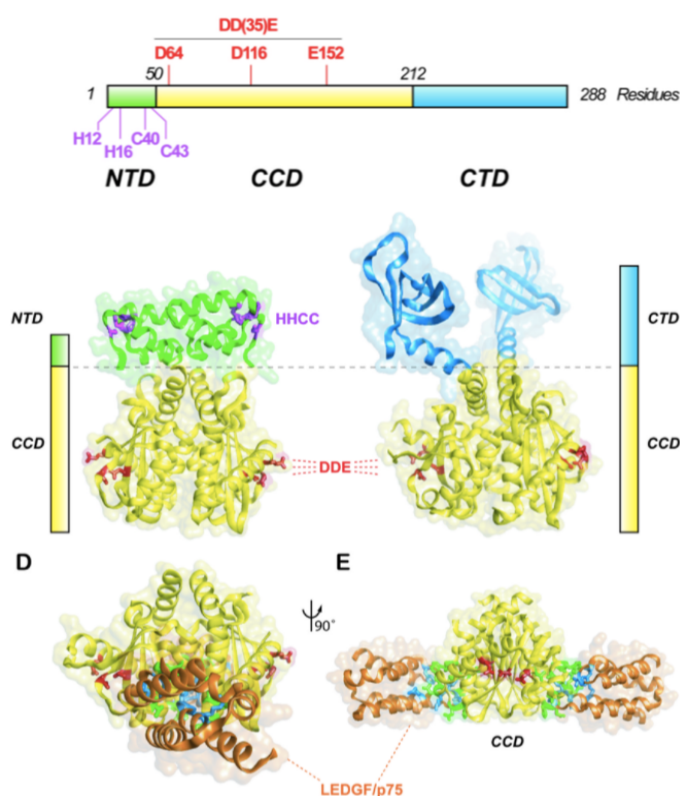


Figura 7 – Domínios funcionais e estruturais da enzima integrase do VIH-1. O domínio NTD tem quatro resíduos de aminoácidos essenciais e conservados (H12, H16, C40 e C43) que coordenam um átomo de zinco. O domínio CCD contém a tríade catalítica DDE *motif* (D64, D116, E152) coordenando um, ou possivelmente dois metais divalentes. O domínio CTD interage com os dois domínios anteriores promovendo a ligação ao ADN. Adaptado de [57].

O domínio N-terminal (NTD) (1-49 resíduos de aminoácidos) contém quatro aminoácidos essenciais e conservados, duas histidinas (H12 e H16) e duas cisteínas (C40 e C43) que coordenam um átomo de zinco [56-58]. A formação do complexo de coordenação entre o átomo de zinco e o motivo $H^{12}H^{16}C^{40}C^{43}$ contribui para a

oligomerização da integrase e para o aumento da atividade catalítica da mesma [56-58]. O domínio catalítico (CCD) abrange os resíduos de aminoácidos 50-212 e contém uma tríade catalítica ácida altamente conservada (DDE *motif*) formada por dois aminoácidos de ácidos aspárticos (D64, D116) e um de ácido glutâmico (E152). Estes aminoácidos são considerados essenciais para a IN, uma vez que a sua alteração culmina na perda da atividade catalítica [56-58]. Os resíduos D64 e D116 formam um complexo de coordenação com iões divalentes (manganésio ou magnésio) [56,58]. Outros resíduos importantes neste domínio estão incluídos no motivo $K^{186}R^{187}K^{188}$, que participa na multimerização da enzima e possui um sinal de localização nuclear, que permite a importação do ADN complementar para o núcleo da célula infetada [59]. A região central da integrase é essencial para a ligação a cofatores celulares de integração (*e.g.* LEDGF/p75) que intervêm na ligação da integrase à cromatina, permitindo a translocação do ADN complementar do citoplasma para o núcleo [57,58]. Por último, o domínio carboxiterminal (CTD), compreendido entre os aminoácidos 213-288, corresponde à região menos conservada da integrase [49]. Este domínio tem um papel fundamental na formação da estrutura quaternária da enzima e é ainda responsável pela ligação do ADN viral ao ADN da célula hospedeira [47,56-59].

1.4.2 Gene *env*

O gene *env* do VIH-1 possui cerca de 2570 nucleótidos e codifica uma glicoproteína viral de 160 kDa (gp160) [48,60]. A glicoproteína gp160, após tradução, é clivada por proteases celulares originando as glicoproteínas de superfície gp120 (SU) e transmembranar gp41 (TM). Estas, por sua vez, permanecem covalentemente ligadas no invólucro viral formando um heterodímero, que se dispõe em trímeros, onde a gp120 permite a interação com os recetores celulares e a gp41 media a fusão do vírus com a membrana da célula hospedeira [60]. A glicoproteína de superfície gp120 é constituída por cinco regiões conservadas (C1-C5) e cinco regiões variáveis (V1-V5), como representado na figura 8 [47,61]. Na sua conformação nativa, gp120 possui um domínio interno e um domínio externo. O domínio externo (252-482, em VIH-1 HXB2) é altamente glicosilado e está envolvido na ligação ao recetor CD4 e aos co-recetores CCR5/CXCR4 [48,60,62]. Este domínio serve também como principal alvo da resposta imunológica do hospedeiro, pois contém a maior parte dos determinantes antígenicos da

gp120, incluindo epítomos neutralizantes, que são alvo dos anticorpos neutralizantes [62]. O domínio interno é hidrofóbico e essencial para a ligação entre a gp120 e a gp41[47].

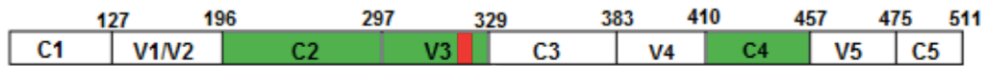


Figura 8 – Ilustração esquemática da glicoproteína SU do VIH-1 HXB2. A glicoproteína de superfície inclui cinco regiões conservadas (C1-C5) e cinco regiões variáveis (V1-V5). Adaptado de [61].

As regiões variáveis V1-V4 tendem a formar *loops* delimitados por ligações dissulfureto e encontram-se expostas à superfície do invólucro viral (Figura 9) [61]. As regiões V1-V2 são consideradas as mais variáveis em termos de sequência e tamanho e apresentam uma extensa glicosilação. A exclusão das regiões V1/V2 torna o vírus altamente suscetível à neutralização, uma vez que o seu padrão de glicosilação permite a oclusão de domínios neutralizantes constituintes da região V3 [60,63]. As próprias *loops* V1/V2 constituem um alvo importante de ação de anticorpos neutralizantes e não neutralizantes. A sensibilidade das estirpes virais aos anticorpos neutralizantes está também relacionada com as dimensões das *loops* V1/V2. Geralmente, os vírus que possuem regiões V1/V2 mais longas são menos suscetíveis à neutralização [63].

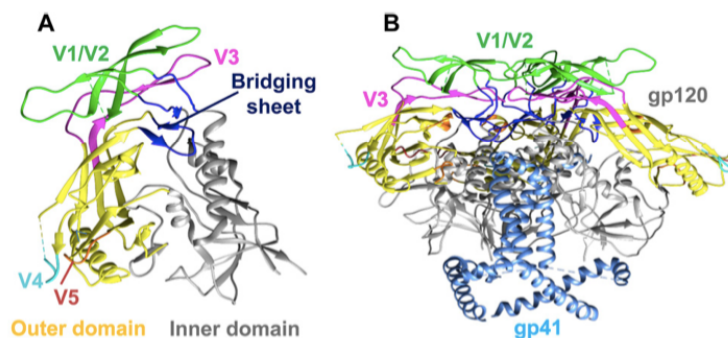


Figura 9 – Representação estrutural do monómero gp120 e do heterodímero gp120-gp41. A) A gp120 encontra-se dividida no seu domínio interno (cinza), externo (amarelo) e *bridging sheet* (azul escuro). As *loops* correspondentes às regiões variáveis V1-V5 estão representadas a verde, magenta, azul claro e vermelho, respetivamente. B) O heterodímero é composto pela gp120, exposta à superfície (colorida como em A) e a gp41 representada a azul. Adaptado de [62].

A região V3 da gp120 é geralmente constituída por 35 aminoácidos, podendo variar entre 34-36. [64,65]. A sua sequência aminoacídica é muito variável nos subtipos do grupo M, à exceção dos vírus do subtipo C onde a sequência de V3 é geralmente mais conservada, comparativamente com outros subtipos [66]. Para além da sua função no início do ciclo replicativo (ligação a um co-recetor de quimiocina), o domínio V3 foi identificado como o principal domínio neutralizante do invólucro viral, sendo alvo de vários anticorpos monoclonais e policlonais anti-V3 [64-66].

1.5 Replicação Viral

1.5.1 Tropismo Celular e Recetores

O VIH revela um tropismo particular para células do sistema imunitário que expressem o recetor CD4 e os co-recetores CCR5 e/ou CXCR4, nomeadamente linfócitos T CD4⁺ auxiliares, células de linhagem monocitária e células dendríticas [67]. A entrada do vírus, na maioria das células, é iniciada pela ligação da gp120 ao recetor CD4 (*cluster* de diferenciação 4), desencadeando alterações conformacionais que permitem a exposição da região V3, responsável pela ligação a co-recetores essenciais para uma infeção eficiente da célula hospedeira [67,68]. Os co-recetores fundamentais para a entrada do vírus nas células são CCR5 (recetor de β -quimiocinas) e CXCR4 (recetor de α -quimiocinas), possuem cerca de 352 aminoácidos, apresentam-se na membrana das células e têm sete domínios transmembranares [67,69]. O co-recetor CCR5 é expresso em monócitos, macrófagos, células *natural killer* (NK), células T de memória, células T ativadas e células dendríticas imaturas, incluindo as células de Langerhans da mucosa genital [69,70]. O co-recetor CXCR4 é expresso numa maior classe de células, como a maior parte das células hematopoiéticas, incluindo neutrófilos, monócitos, macrófagos, linfócitos T primários (*naive*), linfócitos B e células dendríticas. Este recetor é ainda expresso em células endoteliais vasculares, neurónios e na glia [67,68,70]. As estirpes do VIH-1 que utilizam CCR5 (estirpes R5) predominam durante a fase assintomática e em cerca de 60% dos indivíduos nos estágios terminais da infeção. As estirpes que utilizam o co-recetor CXCR4 (estirpes X4) surgem em estágios mais tardios da infeção e estão associados à diminuição do número de linfócitos T CD4⁺ e progressão da doença. Existem estirpes que apresentam tropismo duplo, sendo por isso designadas de R5/X4. As estirpes

X4 são em geral mais suscetíveis à ação de anticorpos neutralizantes do que as estirpes R5 [67-70]. O tropismo diferencial do VIH-1 pelos co-recetores é, em geral, determinado pela sequência em aminoácidos, carga e dimensão da região V3 da gp120 [67,71,72]. A acumulação de mutações nessa região promove a alteração da carga global da ansa V3, podendo também levar à substituição de aminoácidos específicos nas posições 11 e 25 que estão diretamente envolvidos na interação da gp120 com os co-recetores [67,73]. Quando estes aminoácidos possuem carga positiva, os vírus usam, em geral, CXCR4 para entrar nas células [67,73]. Este facto pode ser explicado pela atração eletrostática entre a *loop* V3 e a superfície extracelular do co-recetor CXCR4, uma vez que este possui mais cargas negativas que CCR5 [67,73]. A alteração do padrão de N-glicosilação da *loop* V3 também pode contribuir para a ligação a um co-recetor em particular [74]. Atualmente são usados algoritmos informáticos que permitem prever com elevada precisão o tipo de co-recetor que é utilizado por determinada estirpe de VIH-1 [67-74].

1.5.2 Ciclo Replicativo

O ciclo replicativo do VIH-1 consiste numa série de eventos sequenciais com vista à produção de novos viriões capazes de infetar outras células alvo [75]. Genericamente, o ciclo replicativo pode ser dividido numa fase precoce e numa fase tardia (Figura 10). A fase precoce tem início com a ligação do virião à superfície celular e termina com a integração do ADN proviral no genoma da célula hospedeira, enquanto a fase tardia engloba a transcrição dos genes virais até à maturação dos viriões [75,76]. A entrada do VIH-1 em células suscetíveis inicia-se com a ligação da gp120 ao recetor celular CD4, sofrendo um conjunto de alterações conformacionais que expõem a região V3 permitindo a sua interação com o CCR5 ou CXCR4. Após esta interação, dá-se a exposição do péptido de fusão da gp41 que culmina com a subsequente fusão das membranas viral e celular [75,77]. As partículas virais sofrem uma descapsidação parcial permitindo a libertação da cápside viral no citoplasma [75]. Subsequentemente, inicia-se a transcrição reversa do genoma viral, a partir do qual é sintetizada uma molécula de ADN complementar de cadeia dupla [76-79]. A nova cadeia de ADN, associada a várias proteínas virais (MA, NC, Vpr, RT e IN), formam o complexo de pré-integração (PIC), transportado até ao núcleo, através de uma rede de microtúbulos citoplasmáticos [76,79].

No núcleo, através da ação da integrase viral, ocorre a integração do ADN viral no genoma da célula hospedeira [48,56,76]. Uma vez integrado, o ADN proviral serve de matriz para a síntese do genoma viral e de ARNm que codificam as proteínas virais, recorrendo à maquinaria enzimática celular [76,80].

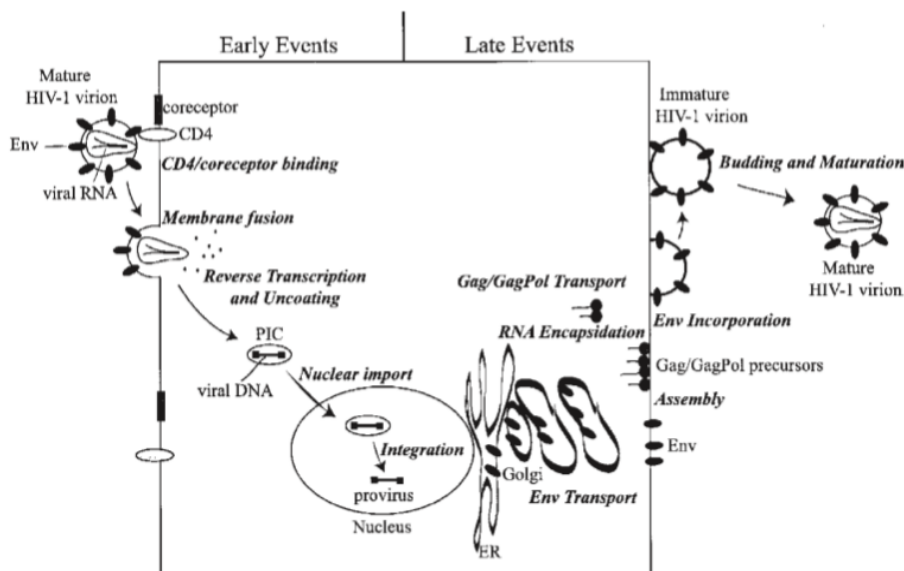


Figura 10 – Resumo do ciclo replicativo do VIH. A fase precoce tem início com a ligação do vírus à superfície celular, seguido da sua entrada, síntese do DNA proviral e integração do mesmo no genoma da célula hospedeira. A fase tardia engloba a transcrição dos genes do vírus, o processamento dos ARNm, síntese de proteínas virais, montagem e gemulação de novas partículas virais imaturas, que por fim sofrem o processo de maturação. Adaptado de [76].

Depois de transcritos, os ARNm sofrem algumas transformações, como a poliadenilação na extremidade 3' ou o processamento por *splicing* (nenhum, parcial, total). Os ARN que não sofrem *splicing* codificam as proteínas Pr55^{Gag} e Pr160^{Gag-Pol} ou são utilizados como ARN genómico na génese de novos viriões [76,78]. Os ARNm parcialmente processados dão origem a Env, Vif, Vpr e Vpu. Os ARNm maduros que resultam de um *splicing* total são os primeiros a ser produzidos e codificam as proteínas precoces Tat, Rev e Nef [78-80]. Os ARNm completamente processados são transportados para o citoplasma como qualquer ARN celular. No entanto, o transporte de ARNm não processados ou parcialmente processados são assegurados pela proteína Rev, precocemente expressa [79,81]. A síntese das glicoproteínas do envólucro ocorre ao nível dos polirribossomas associados à membrana do retículo endoplasmático, sob a forma do precursor poliproteico gp160 [66,78]. As restantes proteínas são traduzidas em

polirribossomas livres no citoplasma da célula [42,75]. Após a síntese as glicoproteínas virais são transportadas até junto da membrana plasmática, através de vesículas do complexo de Golgi [76,82]. O processo de montagem de novos viriões ocorre no citoplasma da célula, junto à membrana plasmática (onde já encontram as proteínas do invólucro) num processo que envolve cópias de ARN do genoma viral e as poliproteínas precursoras Pr55^{Gag} e Pr160^{Gag-Pol}. O domínio MA direciona a proteína precursora para a membrana plasmática, enquanto a NC permite a encapsidação do genoma viral [75,83]. Ocorre a gemulação das partículas virais para o exterior, ainda de forma imatura e não infecciosa, coordenada principalmente pelo domínio p6 da Pr55^{Gag} e pela proteína Vpu [75,82]. No final do ciclo replicativo do VIH-1, ocorre a maturação das partículas virais, que pode ou não coincidir com a sua saída da célula, na qual ocorre a clivagem das poliproteínas precursoras Pr55^{Gag} e Pr160^{Gag-Pol}, mediada pela PR, dando origem a partículas virais maduras e infecciosas [84].

1.6 História natural da infeção por VIH-1

1.6.1 Vias de transmissão

A infeção por VIH pode ser contraída através de uma exposição parentérica (inoculação percutânea e transfusão de sangue contaminado), após contacto sexual sem proteção, através da mucosa rectal e genital e por transmissão vertical de mãe para filho, durante a gestação, no parto ou através da amamentação [88]. O risco associado a estas diferentes vias de exposição difere, mas independentemente da via de transmissão, regista-se um aparecimento uniforme, cronológico e ordenado dos marcadores virais de infeção [88].

1.6.2 Patofisiologia e marcadores de infeção

1.6.2.1 Infeção primária e síndrome da infeção aguda

O primeiro estágio da infeção pelo VIH-1, denominado de fase aguda, ocorre aproximadamente, três a seis semanas após o evento de transmissão. A maior parte dos indivíduos revelam algumas manifestações clínicas semelhantes à gripe, febre, linfadenopatia, cefaleias, artralgia, náuseas e vômitos, diarreia, perda de peso e erupções

cutâneas [85]. Estas manifestações estão associadas a uma intensa replicação viral, associada ao aumento exponencial da carga viral, acompanhada por uma diminuição transitória do número de células T CD4⁺ no sangue periférico, devido à interação direta do vírus com as mesmas, e através da resposta celular do sistema imunológico do hospedeiro, que promove a apoptose das células infectadas (Figura 11) [86-88].

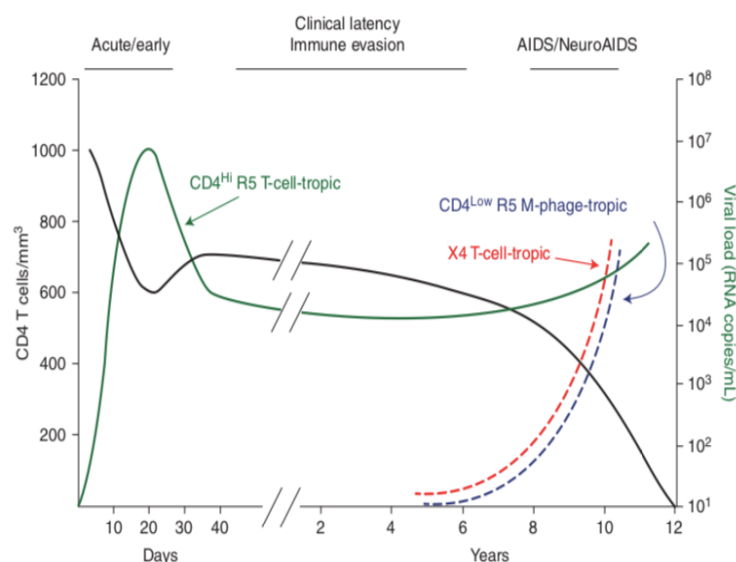


Figura 11 – Decurso geral de uma infecção por VIH-1. Cronograma dos diferentes estágios de infecção e respetivas variações nos níveis de carga viral (verde) e do número de células T CD4⁺ (linha preta). Geralmente, o vírus utiliza o co-recetor CCR5 no início da infecção, mas em estágios mais tardios pode evoluir por forma a utilizar o co-recetor CXCR4 (vermelho) e pode adquirir a capacidade de infetar células com baixos níveis de CD4 (CD4^{Low} R5 M-phage-tropic, roxo). Adaptado de: [88]

Durante e após a infecção aguda, o vírus encontra-se, sobretudo, em células T da região paracortical dos nódulos linfáticos, que se encontram produtivamente infetadas [86,87]. Neste estágio, observa-se também uma depleção marcada das células T CD8⁺ e T CD4⁺ de memória, em particular das que expressam o co-recetor CCR5 nos tecidos linfoides secundários [86,87]. Nas semanas subsequentes, verifica-se uma diminuição gradual da carga viral plasmática e um aumento do número de células T CD4⁺, até ser atingido um equilíbrio entre a replicação viral e a resposta imunitária do hospedeiro, denominado *viral set point*, sendo determinante na forma de evolução da doença (Figura 11) [86-88]. Geralmente, três a doze semanas após a infecção, inicia-se a seroconversão com deteção de anticorpos [89].

1.6.2.2 Infecção Crónica: Período de latência clínica

O segundo estágio da infecção pelo VIH-1 é considerado uma fase de latência clínica ou assintomática e pode durar entre oito a dez anos [85,88]. É caracterizado por baixos, mas persistentes, níveis de replicação viral e uma estimulação contínua do sistema imunológico do hospedeiro [85,88,90]. O vírus induz a ativação crónica do sistema imunológico, que reverte em seu benefício, permitindo a manutenção da infecção de novas células e da replicação viral. A contínua ativação celular leva à expansão clonal de linfócitos e induz anergia e aumento da suscetibilidade à apoptose, promovendo a falência do sistema imunológico [90].

1.6.2.3 SIDA

A maioria dos indivíduos infetados pelo VIH-1, na ausência de tratamento, tornam-se suscetíveis à ocorrência ou reativação de infeções oportunistas (*e.g.* pneumonia, tuberculose, candidíase), bem como o desenvolvimento de neoplasias, na sua maioria relacionadas com a reativação de microrganismos oportunistas, como certos linfomas e carcinomas (*e.g.* vírus de Epstein-Barr, papilomavírus e herpesvírus 8) [91]. Estas manifestações clínicas definem um estágio de doença avançada associado à depleção progressiva de células T CD4⁺ que se designa por sida [85,91]. De acordo com a definição redigida pelo Centro de Controlo e Prevenção de Doenças (CDC), o indivíduo entra neste último estágio quando a contagem de células T CD4⁺ se encontra abaixo de 200 células/mm³ de sangue associada a infeções oportunistas [92]. A maioria dos doentes com sida morrem num período médio de 3 anos se não forem tratados com antirretrovirais [85,91].

1.7 Resposta imunitária contra o VIH-1

1.7.1 Imunidade Inata

A resposta imunológica à infecção pelo VIH-1 tem início com a deteção do vírus e a sinalização imediata da atuação do sistema inato. A deteção do vírus pelo organismo é possível pelo reconhecimento de padrões moleculares associado ao patógeno (PAMP), sendo realizado por recetores de reconhecimento de patógenos (PRR) expressos por

células do sistema imunológico inato [93,94]. O reconhecimento do vírus permite a ativação de células responsáveis pela libertação de IFN- γ e TNF- α , bem como citocinas e quimiocinas pró-inflamatórias que recrutam e ativam novas células do sistema inato para os locais de infecção com o objetivo de inibir a replicação viral [93,94].

1.7.2 Imunidade Adaptativa

1.7.2.1 Resposta Celular

As respostas celulares contra o VIH-1 são mediadas por vários tipos de linfócitos T, nomeados de acordo com a sua função, tais como: células T auxiliares ($CD4^+$), células T citotóxicas ($CD8^+$) e células T foliculares (Tfh) [95]. Após a infecção, as células T *naive*, que circulam nos órgãos linfoides, reconhecem epítomos específicos que se encontram na superfície das células apresentadoras de antígenos, no contexto da apresentação dependente de HLA classe II, induzindo a sua expansão proliferativa e diferenciação em células T efectoras [95-97]. As células T $CD4^+$ promovem a libertação de um conjunto de interleucinas e citocinas que estimulam autocrinamente a proliferação linfocitária e a amplificação da resposta de diferentes componentes do sistema imunitário, como o desenvolvimento de linfócitos T $CD8^+$ (CTL) [96]. Estas células estimulam ainda, a diferenciação dos linfócitos B em células plasmáticas e o recrutamento de neutrófilos e macrófagos para os locais de infecção [96]. Geralmente, a população de células T efectoras ativadas leva à produção de um conjunto de células T de memória responsáveis pela resposta rápida a uma reexposição ao patógeno [96,97]. As células auxiliares foliculares (Tfh), localizadas nos órgãos linfoides secundários, intervêm na estimulação das células B a produzir anticorpos e diferenciação em células B de memória [96,98]. Deste modo, as células T $CD4^+$ intervêm na ativação e expansão das células T $CD8^+$, contribuem para a maturação das células B e para o desenvolvimento de memória imunitária [95-98]. No decurso da infecção até à SIDA, as alterações qualitativas dos linfócitos T $CD4^+$ traduzem-se na diminuição da sua capacidade proliferativa e alteração da produção de citocinas, que em conjunto comprometem a resposta imune celular e humoral [85-88]. A depleção linfocitária é causada por eventos de apoptose promovidos por uma variedade de proteínas do VIH, pela ativação crónica do sistema imunológico e pelo estado hiperinflamatório característico da infecção [99].

No entanto, a maior parte das respostas celulares contra o VIH-1 são mediadas por linfócitos T CD8⁺ citotóxicos no contexto da apresentação antigénica pelo complexo de histocompatibilidade (HLA) classe I, induzindo a lise direta das células infetadas [95-97]. Os CTL também podem promover a libertação de fatores supressores solúveis, como citocinas - TNF- α e IFN- γ , responsáveis pela inibição da replicação viral – e quimiocinas (MIP-1 α , MIP-1 β e RANTES) que competem com o vírus para o uso do co-recetor CCR5 e bloqueiam, desta forma, a entrada do mesmo na célula [95-97]. A expansão dos CTL ocorre no início da infeção aguda, podendo atingir concentrações muito elevadas no sangue periférico (até 1% das células mononucleadas) e parece ser crucial, para o declínio posterior da viremia plasmática, interferindo com o *set point*, e, portanto, com a taxa de progressão da doença [95]. Ao longo da fase aguda da infeção verifica-se um declínio rápido das células T CD8⁺, muitas vezes explicado pela ação das proteínas virais Tat, Nef e Vpu do VIH-1 que têm a capacidade de promover a diminuição da expressão de moléculas HLA-I, da superfície das células, impedindo o reconhecimento do antígeno pelos linfócitos TCD8⁺ [97]. Outro fator que contribui para o declínio do número de células T CD8⁺ é seleção inicial de mutações que interferem no reconhecimento dos antígenos virais pelas CTL, permitindo o escape viral. Para além disso, a ativação crónica e a perda de função das células T auxiliares, com a progressiva depleção de linfócitos T CD4⁺, pode culminar numa anergia funcional dos CTL, bem como um aumento da suscetibilidade à apoptose, contribuindo para a perda de atividade citolítica com a progressão da doença [97].

1.7.2.2 Resposta Humoral

A resposta humoral é caracterizada pela produção de anticorpos (ou imunoglobulinas) neutralizantes e não neutralizantes por plasmócitos – células derivadas dos linfócitos B [100].

1.7.2.2.1 Classes de anticorpos

Os anticorpos são proteínas glicosiladas presentes na superfície de células B (imunoglobulinas de superfície) que servem como recetores de antígenos (BCR, do inglês, *B cell receptor*), ou são secretados no espaço extracelular, onde se ligam aos seus antígenos alvo [100]. Os anticorpos são constituídos por quatro cadeias polipeptídicas –

duas cadeias pesadas e duas cadeias leves – que se encontram ligadas por pontes dissulfeto (Figura 12) [100].

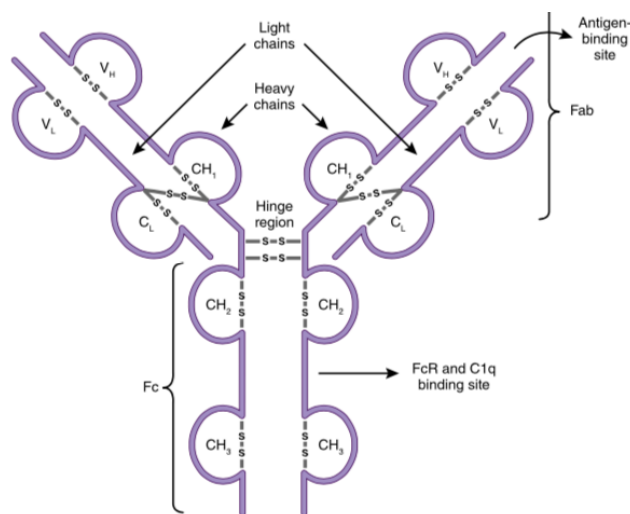


Figura 12 – Estrutura do anticorpo. Os anticorpos são compostos por duas cadeias pesadas (VH e CH) e duas cadeias leves (VL e CL). O fragmento de ligação ao antígeno, Fab, é composto por um domínio variável de cada cadeia pesada e leve (VH e VL). Os domínios constantes CH2 e CH3 da cadeia pesada constituem o fragmento Fc, que medeia as funções efectoras da Ig. Adaptado de: [100]

A estrutura das imunoglobulinas apresenta, a nível funcional, duas regiões principais: o fragmento cristalino (Fc) e o fragmento de ligação ao antígeno (Fab). A região Fab é responsável pela ligação ao determinante antígeno ou epítipo e é composta por um domínio constante e outro variável de cada uma das cadeias leve e pesada do anticorpo. Em cada domínio variável existem três regiões, denominadas regiões determinantes de complementaridade (do inglês, *complementarity determining region*, CDR), que entram em contato direto com o antígeno [100,101]. O fragmento cristalizável (Fc) é composto pelos domínios constantes das cadeias pesadas (CH2 e CH3) e é responsável pelas funções efectoras do anticorpo através da ligação aos recetores Fc ou a proteínas do sistema complemento [100]. Com base em sequências de aminoácidos únicas na região constante da extremidade C-terminal das cadeias pesadas, os anticorpos são classificados como IgM, IgD, IgG, IgA e IgE. Cada classe possui propriedades funcionais e estruturais específicas [102]. A classe de anticorpos IgG é a mais abundante no plasma e encontra-se subdividida em quatro subclasses: IgG1, IgG2, IgG3 e IgG4 [102]. Os indivíduos infectados pelo VIH-1, durante o estágio agudo e crónico da doença, produzem

principalmente anticorpos IgG1 e IgA (encontrado principalmente nas mucosas), em que títulos mais altos dessas subclasses de anticorpos estão associados a um melhor controle da replicação viral [103,104].

1.7.2.2.2 Produção de anticorpos – Células B

A produção de células B ocorre pela primeira vez na fase embrionária e após o nascimento a sua produção ocorre na medula óssea a partir de células estaminais hematopoiéticas (HSC) [100,105]. O desenvolvimento e maturação destas células ocorre segundo um processo altamente regulado, onde se dá o rearranjo dos genes de imunoglobulina nas cadeias pesadas e leves e eventos de seleção negativa e positiva que selecionam as células B imaturas que possuem um recetor BCR funcional e não autoreativo [106]. Após estes processos, as células B *naive* maduras, que expressam IgM e IgD à sua superfície, migram para os tecidos linfoides e para a corrente sanguínea [107]. A interação inicial entre as células T CD4⁺ e os linfócitos B ocorre na extremidade dos folículos linfoides após apresentação antigénica pelas células dendríticas, promovendo a primeira linha de defesa humoral contra o vírus (Figura 13, passo 1) [100,108]. Esta interação resulta na produção de níveis baixos de anticorpos com baixa afinidade para o antígeno circulante. Os plasmócitos gerados nesta interação são geralmente de vida curta (SLPC, do inglês *short-lived plasma cells*) e produzem anticorpos por poucas semanas. São geradas poucas células B de memória (Figura 13, passo 2) [100,108,109].

Algumas células T CD4⁺ auxiliares ativadas pela apresentação antigénica por parte das células dendríticas migram para o interior do folículo, diferenciando-se em células T foliculares (TFh) (Figura 13, passo 3 e 4) [100,110]. As células B migram também para o interior do folículo formando uma estrutura organizada chamada de centro germinativo, que se encontra dividido numa zona clara e numa zona escura (Figura 13, passo 5) [100,108]. Na zona escura do centro germinativo, as células B expandem e diversificam os seus recetores por meio de hipermutação somática extensiva das regiões variáveis do gene Ig e alteração da classe da cadeia pesada para IgG, IgA ou IgE, promovendo a alteração das funções efectoras do anticorpo [100,106,108]. Na zona clara do centro germinativo as células dendríticas foliculares apresentam o antígeno às células B do centro germinativo (Figura 13, passo 6). Estas internalizam e processam o antígeno em vesículas para apresentação às células TFh (Figura 13, passo 7) [100,108].

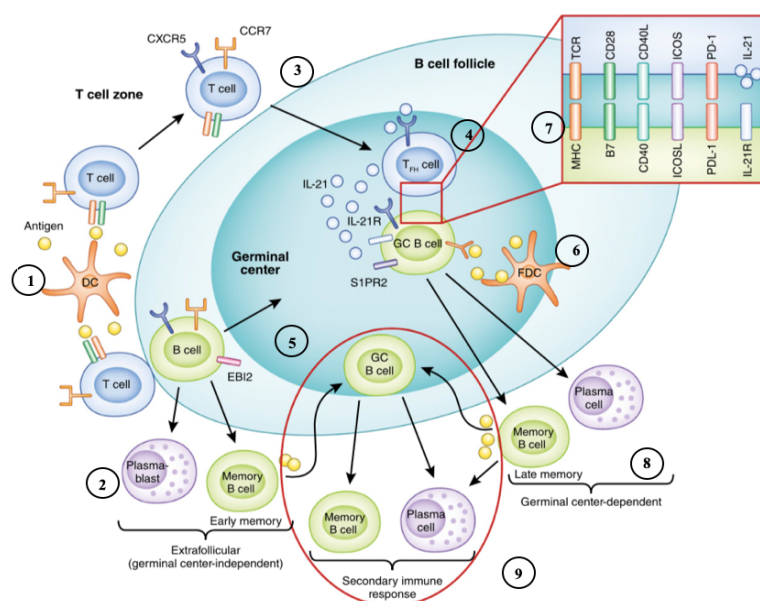


Figura 13 – Ativação e diferenciação de células B em células B de memória e células plasmáticas. Apresentação antigénica inicial (passo 1 e 2). A migração de algumas células T $CD4^+$ e células B para o interior do folículo permite a formação de um centro germinativo, ocorrendo a expansão clonal e diversificação de células B (passos 3 a 8). Nova exposição ao antígeno promove a ativação de células B de memória e/ou diferenciação das mesmas em células plasmáticas de vida longa ou migração para o folículo para formação de novo centro germinativo (passo 9). Adaptado de: [100]

Estas reações no centro germinativo promovem a produção de células B de memória e células plasmáticas de vida longa (LLPC, do inglês *long-lived plasma cells*) que migram para medula óssea onde produzem anticorpos por períodos prolongados, para uma proteção persistente (Figura 13, passo 8) [100,108,109]. No caso de nova exposição ao mesmo antígeno, as células B de memória são ativadas e podem diferenciar-se em células plasmáticas de vida longa para a produção de novos anticorpos ou iniciarem uma nova expansão clonal através da migração para o folículo, formando um novo centro germinativo (Figura 13, passo 9) [100,108].

1.7.2.3 Anticorpos neutralizantes e não neutralizantes produzidos na infeção por VIH-1

A resposta inicial das células B à infeção pelo VIH-1 ocorre cerca de uma semana após viremia detetável e é dirigida, inicialmente, contra o antígeno p24 e depois contra epítopos variáveis e não neutralizantes do invólucro viral [111]. No invólucro, esta

resposta é direcionada, numa primeira instância, contra a gp41 e cerca de 28 dias após carga viral detetável contra a gp120 [107,111]. Estes anticorpos não neutralizantes atuam através da sua ligação a células infetadas promovendo o recrutamento de células efectoras, que por sua vez induzem a citólise ou a apoptose das primeiras [89]. Estes mecanismos são conhecidos como citotoxicidade dependente de anticorpos (ADCC, do inglês *antibody-dependent cell-mediated cytotoxicity*) e inibição viral mediada por células dependentes de anticorpos (ADCVI, do inglês *antibody-dependent cell-mediated vírus inhibition*). O primeiro compreende a formação de um complexo entre a região Fab do anticorpo com a proteína de superfície do invólucro viral e a ligação da porção Fc aos recetores FcγRs em células efectoras (e.g. células NK, macrófagos, células dendríticas e neutrófilos) (Figura 14) [89]. ADCVI trata-se da inativação de vírus livres e/ou associados a células, por morte celular, mas também inibição do vírus por citocinas e outros fatores secundários (Figura 14) [89,112].

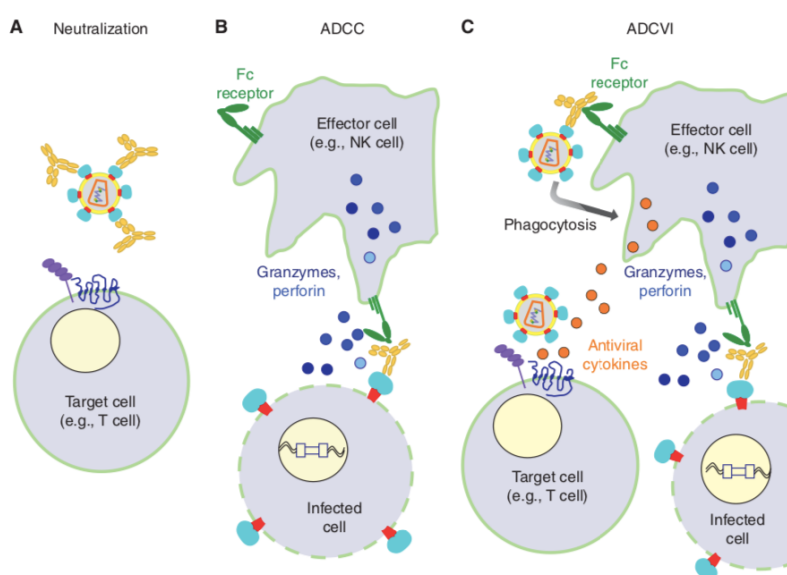


Figura 14 – Representação esquemática dos mecanismos de ação dos anticorpos neutralizantes e não neutralizantes. A) Os anticorpos neutralizantes ligam-se às glicoproteínas do invólucro viral e impedem a entrada do vírus na célula. **B)** Em ADCC a formação de um complexo entre o anticorpo, a proteína viral de superfície e as células efectoras promove a lise da célula infetada. **C)** Em ADCVI ocorre inativação de vírus por morte celular e/ou por libertação de citocinas e outros fatores secundários. Adaptado de: [89].

Os primeiros anticorpos neutralizantes (NAbs), direcionados para regiões específicas das glicoproteínas do invólucro viral (gp120 e gp41), surgem cerca de dois a três meses após a infecção e atuam através da ligação a vírus livres, impedindo que o virião infete novas células (Figura 14) [89,111]. No entanto, operam apenas sob a estirpe infetante (vírus autólogo), sendo incapazes de neutralizar vírus mais divergentes (vírus heterólogos) [106,112,113]. Contudo, os NAbs autólogos exercem uma pressão seletiva sobre o vírus que rapidamente adquire mutações que possibilitam o escape à neutralização [106]. O escape viral é em grande parte resultado da variabilidade genética de Env e das suas características estruturais que limitam ou bloqueiam o acesso dos anticorpos aos epítomos neutralizantes conservados [106]. A título de exemplo, tais características incluem a alteração do padrão de glicanos para ocultar epítomos ou a exposição momentânea de alguns desses epítomos aos anticorpos (*e.g.* conformação de pré-fusão da gp41). Para além dos fatores mencionados, a ligação dos anticorpos a epítomos não funcionais (*e.g.* precursores não clivados e monómeros de gp120) e o tamanho das imunoglobulinas que impedem o acesso a alguns epítomos (*e.g.* CD4i), contribuem também para o escape viral [106]. No decurso da infecção, aproximadamente 20-30% dos indivíduos infetados pelo VIH-1 desenvolvem respostas de anticorpos neutralizantes contra vírus heterólogos e aproximadamente 1% dos indivíduos desenvolvem uma potente e ampla resposta de anticorpos neutralizantes (bNAbs) [70,114]. Os bNAbs são anticorpos raros capazes de neutralizar *in vitro* várias estirpes de VIH-1 de diversos subtipos e geralmente surgem dois a quatro anos após seroconversão (estádio crónico da infecção), estando ausentes em indivíduos recentemente infetados [106,114]. Estes anticorpos possuem em geral características que tornam difíceis a sua indução, tais como: auto e poli-reatividade, uma longa região determinante de complementaridade (CDR3) e uma extensa hipermutação somática [115]. Estas características sugerem que a infecção persistente e a exposição repetida aos antígenos virais são necessárias para a produção de bNAbs [106,116]. O desenvolvimento de novas técnicas de clonagem de anticorpos, em complementaridade com métodos de seleção baseados na ligação a antígenos, permitiu mapear e definir estruturalmente as regiões do invólucro viral que são alvo dos bNAbs, sendo elas: local de ligação ao recetor CD4 (CD4bs), epítomos associados aos N-glicanos nas *loops* V1/V2 e V3, local de interface gp120/gp41, péptido de fusão e região externa proximal de membrana da gp41 (MPER, do inglês *membrane proximal external region*) (Figura 15) [106,115].

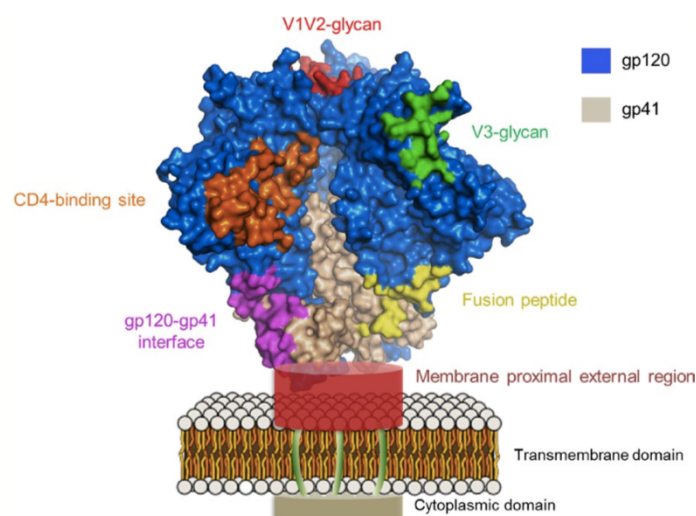


Figura 15 – Mapeamento dos locais de ligação de bNABs nas glicoproteínas do invólucro do VIH-1. As regiões alvo dos bNABs são: epítomos associados aos N-glicanos das *loops* V1/V2 (vermelho) e V3 (verde), péptido de fusão (amarelo), região externa proximal de membrana da gp41 (castanho), local de interface gp120/gp41 (roxo) e o local de ligação ao recetor CD4 (CD4bs) (laranja). Adaptado de: [106]

Os bNABs de primeira geração foram isolados de amostras de indivíduos infetados com VIH-1, no início dos anos noventa do século passado e embora neutralizassem diversas estirpes de VIH-1 *in vitro*, a sua potência e amplitude eram insuficientes, uma vez que suprimiam apenas moderadamente a viremia [106,116,117]. A aplicação de novas técnicas de clonagem de anticorpos e o desenvolvimento de ensaios de neutralização de alto rendimento permitiram o isolamento e a caracterização de uma nova geração de bNABs com maior potência e amplitude para profilaxia imunológica e terapia da infeção por VIH-1 [117]. Em ensaios de imunoterapia, a administração de bNABs em animais cronicamente infetados resultou numa diminuição do ARN viral plasmático para níveis indetetáveis, uma redução do ADN proviral na mucosa gastrointestinal e nos nódulos linfáticos e ainda um aumento da resposta imunitária celular [117,118]. A indução de uma resposta imunitária potente, capaz de induzir bNABs, tornou-se palco de atuação de muitos virologistas, com a tentativa de desenvolver uma vacina para combater a pandemia [113]. Desde 1987, foram desenvolvidos mais de trinta candidatos a vacina contra o VIH-1, em que alguns produziram uma resposta protetora em modelos primatas não humanos e uma pequena parte avançou para ensaios clínicos em humanos [119-121]. Estes estudos utilizaram diferentes tipos de estratégias para induzir uma resposta

imunológica, nomeadamente, proteínas de envólucro viral gp120 recombinante, péptidos multiméricos com base na região V3 da proteína gp120, vetores víricos, e ainda, combinações de vetores e subunidades do envólucro viral, em diversos esquemas de imunização primária e de reforço [119-122]. Após décadas de pesquisa, uma vacina eficaz permanece por descobrir. Alguns dos principais obstáculos encontrados são: 1) diversidade abrangente e evolutiva do VIH-1 como consequência da sua alta taxa de mutação; 2) destruição precoce das células T CD4⁺; 3) difícil identificação de imunogénios capazes de induzir uma resposta celular abrangente e duradoura; 4) dificuldade em definir estruturas e estratégias de imunização capazes de produzir anticorpos neutralizantes; 5) estabelecimento precoce de reservatórios virais latentes [119-122].

1.8 Terapia antirretroviral

Na inexistência de uma vacina eficaz contra o VIH e apesar das abordagens de saúde pública no controlo da pandemia (e.g promoção do uso de preservativos e programas de prevenção entre os toxicodependentes), os antirretrovirais são considerados a única opção efetiva no combate à infeção por VIH [123]. Em 1987, o primeiro fármaco antirretroviral, chamado Zidovudina (AZT), foi aprovado pela *Food and Drug Administration* (FDA) nos EUA. Desde então, foram desenvolvidos e aprovados cerca de 31 antirretrovirais adicionais descritos na tabela A1 em anexo [123,124].

1.8.1 Classes de antirretrovirais

Os antirretrovirais, até agora desenvolvidos, enquadram-se em cinco classes dependendo do local da sua atuação no ciclo replicativo do VIH-1: 1) inibidores de entrada (antagonistas de co-recetores, inibidores de fusão e inibidor de ligação); 2) inibidores da RT análogos a nucleósidos/nucleótidos (NRTI, do inglês *nucleoside/nucleotide reverse transcriptase inhibitors*); 3) inibidores da RT não análogos a nucleósidos (NNRTI, do inglês *nonnucleoside reverse transcriptase inhibitors*); 4) inibidores da IN (INSTI, do inglês *integrase strand transfer inhibitors*) e 5) inibidores da PR (PI, do inglês *protease inhibitors*) [124-126].

1.8.1.1 Inibidores de entrada (EI)

Os EI impedem a entrada do vírus na célula hospedeira, intervindo nos principais eventos da fase precoce do ciclo replicativo do VIH: ligação ao recetor CD4, ligação ao co-recetor CCR5 e fusão do invólucro viral com a membrana citoplasmática da célula. Atualmente, existem três antirretrovirais desta classe em uso clínico, um inibidor de fusão (enfuvirtide), um antagonista do CCR5 (maraviroc) e um inibidor de ligação (fostemsavir) [127,128]. O enfuvirtide é um polipeptídeo de 36 aminoácidos de sequência idêntica à região HR2 da glicoproteína gp41 do VIH-1. Este composto interage com a região HR1 na gp41, exposta temporariamente durante a fusão do VIH com as células alvo, impedindo a associação de HR1 com HR2, o que impossibilita que a glicoproteína gp41 assuma a sua conformação de fusão e consequentemente a entrada do vírus na célula [127,129]. Relativamente ao maraviroc, este interage com o co-recetor CCR5, utilizado pelas estirpes R5, impedindo a ligação da gp120 viral ao mesmo [130]. Por último, o fostemsavir é produzido sob a forma de pró-fármaco, sendo necessárias reações de hidrólise por enzimas celulares para o converter na sua forma ativa (temsavir) [131]. Este fármaco atua ligando-se diretamente à gp120 do invólucro viral inibindo a interação da mesma com o recetor CD4 e por isso impedindo a entrada do vírus na célula [131].

1.8.1.2 Inibidores da RT análogos a nucleósidos/nucleótidos (NRTI)

Os NRTI atuam no centro ativo da transcriptase reversa como inibidores competitivos, uma vez que substituem os dNTPs celulares durante a síntese da cadeia de ADN proviral, induzindo a sua terminação precoce. A estrutura molecular desta classe de inibidores evidencia a ausência do resíduo 3'-OH (grupo hidroxilo) no carbono C3, o que impede a formação da ligação fosfodiéster necessária à continuação da reação de polimerização [124-127]. Estes antirretrovirais são produzidos sob a forma de pró-fármacos, sendo necessárias reações de fosforilação (adição de três grupos fosfato) por enzimas celulares para os converter na sua forma ativa. Nesta classe de inibidores encontra-se inserido um análogo de nucleótido (NtRTI). O NtRTI possui um grupo fosfato, sendo necessária a adição de apenas dois grupos fosfato para ser convertido na sua forma ativa. Esta mesma categorização, apesar da diferença mencionada, deve-se à estrutura molecular e

mecanismo de ação semelhante à dos NRTI [125]. Atualmente, estão aprovados oito fármacos desta classe, discriminados na tabela A1 em anexo.

1.8.1.3 Inibidores da RT não análogos a nucleósidos (NNRTI)

Os NNRTI são moléculas de natureza hidrofóbica, não necessitam ser metabolizados a nível celular para se tornarem ativos e possuem diferentes estruturas moleculares entre si [125,128]. A atividade dos antirretrovirais inseridos nesta classe é também direcionada para a RT, atuando como inibidores não competitivos específicos [124]. Estes ligam-se a um centro alostérico, localizado no domínio *palm* da subunidade p66, perto do centro ativo da RT, promovendo alterações conformacionais da enzima, impedindo o correto funcionamento da mesma [125-127]. Os seis fármacos aprovados e inseridos nesta classe de antirretrovirais encontram-se discriminados na tabela A1 em anexo.

1.8.1.4 Inibidores da integrase (INSTI)

Os INSTI interagem com o cofator da enzima integrase, um complexo de coordenação de iões divalentes (Mg^{2+}) alterando a sua atividade. Para além disso, os compostos desta classe de inibidores impedem a formação de ligações covalentes entre as extremidades do cADN viral e o ADN da célula hospedeira, bloqueando a integração e a consequente formação do provírus [124,126,127]. Atualmente, estão aprovados cinco fármacos desta classe, discriminados na tabela A1 em anexo.

1.8.1.5 Inibidores da protease (PI)

Os PI atuam com base no princípio da mimetização peptídica, ou seja, assemelham-se aos péptidos virais que servem de substrato à protease viral [126]. Estes inibidores são considerados competitivos, uma vez que se ligam ao centro ativo da enzima no lugar nas poliproteínas precursoras virais Pr55^{Gag} e Pr160^{Gag-Pol}, impedindo a etapa de maturação do ciclo replicativo viral, resultando na produção de partículas virais imaturas e não infecciosas [126,128]. Foram aprovados dez fármacos desta classe, discriminados na tabela A1 em anexo.

1.8.2 Regime terapêutico

Inicialmente, os antirretrovirais eram administrados em regimes de monoterapia, no entanto, em 1996, surgiu o conceito de terapia antirretroviral de alta eficácia (HAART, do inglês, *Highly Active Antiretroviral Therapy*) [127]. A HAART consiste geralmente, na administração de uma combinação de três antirretrovirais, com pelo menos dois alvos terapêuticos distintos [127, 132]. Desde 2018, os regimes de primeira linha incluem um inibidor de integrase (DTG) em combinação com dois inibidores de transcriptase reversa análogos a nucleósidos/nucleótido [TDF e 3TC (ou FTC)]. Os restantes regimes encontram-se representados na tabela 1.

Tabela 1 – Regimes terapêuticos de primeira e segunda linha. Regimes de primeira e segunda linha e segunda linha alternativa, atualizados no último relatório da OMS (Julho 2019). TDF – tenofovir; 3TC – lamivudina; FTC – emtricitabina; DTG – dolutegravir; EFV – efavirenz; NVP – nevirapina; AZT – zidovudina; ATV/r – atazanavir/ritonavir; LPV/r – lopinavir/ritonavir; DRV/r – darunavir/ritonavir. Adaptado de: [132].

População	Regime de primeira-linha	Regime de segunda-linha	Regime de segunda-linha alternativa
Adultos e adolescentes	TDF + 3TC (ou FTC) + DTG	AZT + 3TC + ATV/r (LPV/r)	AZT + 3TC + DRV/r
	TDF + 3TC (ou FTC) + EFV (ou NVP)	AZT + 3TC + DTG	AZT + 3TC + ATV/r (ou LPV/r ou DRV/r)
	AZT + 3TC + EFV (ou NVP)	TDF + 3TC (ou FTC) + DTG	TDF + 3TC (ou FTC) + ATV/r (ou LPV/r ou DRV/r)

No entanto, em alguns países em vias de desenvolvimento, como Moçambique, os regimes de primeira linha ainda são baseados em antirretrovirais da classe dos NRTI e NNRTI [132]. A HAART oferece uma melhor qualidade de vida aos indivíduos infetados pelo VIH, pois na maioria dos casos, promove a diminuição da replicação viral e, por consequência, a recuperação parcial do sistema imunológico culminando com a interrupção da progressão da doença [127]. Atualmente, verifica-se um prognóstico otimista para os indivíduos infetados pelo VIH-1 e uma diminuição significativa das taxas de mortalidade associadas, graças à implementação da HAART [133].

1.8.3 Resistências associadas as antirretrovirais

Apesar dos benefícios da terapia, esta não é uniformemente eficaz para todos os indivíduos e o desenvolvimento de resistências aos antirretrovirais é uma consequência quase inevitável [126]. Apesar da HAART diminuir a probabilidade do vírus adquirir resistências, uma vez que atua em diferentes alvos fundamentais do ciclo replicativo viral, existem diversos fatores, quer sejam virais (*e.g.* elevada diversidade genética e altas taxas de mutabilidade), farmacológicos (*e.g.* absorção variável do composto no organismo e a interação com outros medicamentos) ou associados ao próprio indivíduo infectado (*e.g.* falta de adesão aos regimes terapêuticos), que contribuem fortemente para o desenvolvimento de vírus resistentes [127]. Atualmente, estão descritas mutações de resistência para todos os medicamentos em uso clínico para o combate à infecção pelo VIH-1 [134]. As mutações de resistência podem ser classificadas como principais e acessórias. As mutações principais geralmente, ocorrem em locais estruturalmente essenciais e conservados (*e.g.* próximo do local catalítico de uma enzima) e conferem resistência direta a um ou mais antirretrovirais [135]. Por vezes, uma única mutação deste tipo pode conferir resistência cruzada a todos os ARV dentro de uma determinada classe de medicamentos, ou seja, o vírus é definido como resistente a medicamentos aos quais nunca foi exposto, em resultado de mutações que foram selecionadas pelo uso de outro medicamento [136]. As mutações acessórias emergem durante a contínua pressão seletiva dos ARV e geralmente estão associadas a mutações principais. Estas têm a capacidade de reconstituir o *fitness* replicativo do vírus, afetado em maior ou menor grau pelas mutações primárias [135,136]. É de referir que uma mutação considerada principal para um determinado fármaco pode ser considerada secundária para outro [134]. Relativamente aos polimorfismos naturais, consistem em variantes genéticas, sem impacto significativo na replicação viral, e ocorrem, com frequência, em indivíduos não submetidos a qualquer terapêutica [137]. A presença de mutações que permitam a replicação de determinadas estirpes virais sob regime terapêutico, levam a que se tornem rapidamente dominantes, resultando na falha da terapia antirretroviral. Por esta razão, é extremamente importante a monitorização de resistências, que pode ser realizada com recurso a testes de suscetibilidade fenotípicos - análise da capacidade replicativa do vírus na presença de diferentes concentrações do fármaco, com recurso a culturas celulares - e genotípicos, que consistem na identificação de perfis genéticos de resistência, baseados na análise de

sequências nucleotídicas dos genes virais *pol* e *env*, cujos produtos proteicos constituem alvos terapêuticos dos fármacos [138].

1.8.4 Mutações de resistência em cada classe de antirretrovirais

1.8.4.1 Mutações de resistência aos inibidores de entrada

O inibidor de fusão, enfurvirtida, possui uma barreira genética relativamente baixa e a sua capacidade de controlo da carga viral é de curta duração. As mutações que conferem resistência a este antirretroviral ocorrem entre os codões 32 e 39 da região codificante do domínio HR1 da gp41 [69,139]. Relativamente ao maraviroc, a resistência pode ocorrer através do mecanismo habitual de seleção de mutações na ansa V3 do gene *env*, que permitem a ligação do vírus ao co-recetor CCR5, mesmo na presença do fármaco. No entanto, a ineficácia do maraviroc está muitas vezes, associada à seleção do co-recetor CXCR4 como preferencial à entrada do vírus na célula [139]. A resistência ao recente fármaco aprovado – fostemsavir – está muitas vezes associada a mutações em regiões específicas da gp120 (*e.g* L116P, A204D, S375T, M426L, M434I, S475I e V506M) que diminuem a afinidade entre esta e o fármaco permitindo a ligação da glicoproteína do invólucro ao recetor CD4 [140].

1.8.4.2 Mutações de resistência aos NRTI

A resistência aos NRTI está associada a mutações na região codificante da transcriptase reversa, as quais estão associadas a dois mecanismos distintos: diminuição da afinidade do fármaco para a RT e pirofosforólise [141]. No primeiro mecanismo, as mutações existentes conferem à RT uma maior seletividade que permite a distinção entre dNTPs celulares e os seus análogos, impedindo a incorporação dos NRTI na cadeia nascente de ADN complementar [141]. Por sua vez, a pirofosforólise resulta da aquisição de mutações que conferem uma alteração da atividade fosforolítica da RT, que permite a excisão de NRTI incorporados na cadeia nascente a partir da extremidade 3', revertendo a terminação da mesma e permitindo a continuação da sua polimerização [136,141]. Os registos das mutações mais comuns aos NRTI encontram-se representadas na tabela 2 [125,134,142].

Tabela 2 –Mutações de resistência mais comuns aos NRTI. A vermelho encontram-se assinaladas as mutações associadas aos níveis mais baixos de suscetibilidade. As mutações a negrito diminuem a suscetibilidade aos NRTI. As mutações em texto simples contribuem para a redução da suscetibilidade em combinação com outras mutações de resistência. Os aminoácidos apresentados em cada posição estão representados pelo código internacional IUPAC. Adaptado de [142].

Mutações de Resistência a NRTI													
	Mutações discriminatórias					TAMs						Mutações MDR	
Fármaco	M 184	K 65	K 70	L 74	Y 115	M 41	D 67	K 70	L 210	T 215	K 219	T 69	Q 151
3TC	VI	R										Ins	M
FTC	VI	R										Ins	M
ABC	VI	R	E	VI	F	L			W	FY		Ins	M
ddI	VI	R	E	VI		L			W	FY		Ins	M
TDF	***	R	E		F	L		R	W	FY		Ins	M
D4T	***	R	E			L	N	R	W	FY	QE	Ins	M
AZT	***	***		***		L	N	R	W	FY	QE	Ins	M

Ins: Mutação de inserção. *** Reversão da suscetibilidade ao fármaco.

A resistência aos NRTI pode ocorrer através de uma única mutação ou por acumulação sucessiva de mutações. As TAMs (do inglês, *thymidine analogue mutations*) são mutações não polimórficas que se acumulam de forma sequencial, são seleccionadas principalmente pelos análogos de timidina AZT e d4T e são detetáveis, com maior frequência, nos codões 41,210 e 215 e, com menor frequência nos codões 67, 70 e 219 [125,142]. Relativamente às mutações discriminatórias (tabela 2), uma das mutações de resistência mais comuns é M184V/I e localiza-se no motivo altamente conservado YMDD da transcriptase reversa [138]. Geralmente, esta mutação é a primeira a aparecer sob seleção do ABC no decurso da acumulação de mutações que incluem L74V, K65R e Y115F [138,142]. Por outro lado, a presença desta mutação pode restaurar, temporariamente, a suscetibilidade ao AZT, d4T ou TDF, num vírus previamente resistente a esses fármacos [142]. Outro exemplo de interação entre mutações, é a reversão quase total da resistência ao TDF que se verifica quando ocorre uma mutação no codão 65 e se associa à mutação no codão 184 [125,142]. L74V/I ocorre comumente em indivíduos que recebem tratamento com ABC ou ddI. L74I parece menos eficaz na

redução da suscetibilidade aos fármacos indicados e não interfere no aumento da suscetibilidade a AZT, ao contrário de L74V [125,142]. No entanto, a combinação de L74V/I com pelo menos uma das mutações M184V, K65R e Y115F, sendo a primeira o conjunto mais comum, além de reduzir a suscetibilidade a ABC e ddI, pode resultar na resistência cruzada aos fármacos análogos de timidina. A mutação K65R ocorre no subdomínio *fingers* da RT e é selecionada por TDF, ABC e ddI, conferindo um alto nível de resistência aos mesmos [125]. K65R promove a redução de suscetibilidade a 3TC, FTC e d4T e em contraste aumenta a suscetibilidade a AZT, exceto quando se encontra em combinação com Q151M [125,142]. K65R raramente ocorre em combinação com TAMs [142]. A mutação Q151M foi identificada em indivíduos sujeitos a terapia com ddI em combinação com AZT ou d4T [138]. Ocorre, geralmente, em combinação com duas ou mais mutações acessórias (*e.g.* A62V, V75I, F77L e F116Y), conferindo resistência a todos os NRTI [125,142,143]. As inserções que ocorrem entre os códons 66 e 70 são por convenção atribuídas à posição 69. Estas geralmente compreendem a inserção de dois aminoácidos (habitualmente duas serinas) e ocorrem em combinação com múltiplas TAMs, conferindo assim, uma resistência de nível intermédio a 3TC e FTC e resistência de alto nível aos restantes NRTI. Estas mutações são, usualmente, descritas após regimes terapêuticos prolongados [138].

1.8.4.3 Mutações de resistência aos NNRTI

A resistência a esta classe de inibidores resulta da aquisição de mutações no seu local de ligação à RT ou em resíduos adjacentes, promovendo uma redução da afinidade da ligação dos mesmos à enzima [144]. Os NNRTI possuem uma baixa barreira genética, pelo que a fixação de uma ou duas mutações podem resultar em elevados níveis de resistência com um impacto mínimo no *fitness* replicativo [144]. Assim, observam-se fenómenos frequentes de resistência cruzada nesta classe de inibidores, uma vez que a maioria das mutações reduzem substancialmente a suscetibilidade a múltiplos NNRTI [144]. Os registos das mutações mais comuns associadas aos NNRTI encontram-se representados na tabela 3 [142]. As mutações registadas nesta classe de inibidores ocorrem, maioritariamente, no subdomínio *palm* da subunidade p66 da RT. A análise de genótipos virais resistentes aos NNRTI mostram que K103N, Y181C e G190A são as mutações mais prevalentes [142,144].

Tabela 3 – Mutações de resistência mais comuns aos NNRTI. A vermelho encontram-se assinaladas as mutações associadas aos níveis mais baixos de suscetibilidade. As mutações a negrito diminuem a suscetibilidade aos NNRTI. As mutações em texto simples contribuem para a redução da suscetibilidade em combinação com outras mutações de resistência. Os aminoácidos apresentados em cada posição estão representados pelo código internacional IUPAC. Adaptado de [142].

Mutações de Resistência a NNRTI									
Fármaco	Mutações principais					Mutações acessórias			
	Y 181	K 103	V 106	Y 188	G 190	E 138	L 100	M 230	K 101
DOR	CIV		AMI	LHC	SE		I	L	EP
EFV	CIV	NS	AM	LCH	ASE		I	L	EP
ETR	CIV			L	ASE	AGKQ	I	L	EP
NVP	CIV	NS	AM	LCH	ASE		I	L	EP
RPV	CIV			L	ASE	AGKQ	I	L	EP

A mutação K103N é selecionada em indivíduos que recebem tratamento com NVP e EFV, reduzindo a suscetibilidade a estes fármacos [128]. Adicionalmente, surgem mutações como V108I e P225H que potenciam a resistência conferida por K103N [128]. A elevada prevalência de K103N pode ser explicada pelo facto da NVP e o EFV serem frequentemente usados na terapia de primeira linha e pela mutação em si não possuir incompatibilidades relevantes com outras mutações de resistência a inibidores da RT, não alterando significativamente o *fitness* replicativo [123]. Relativamente a Y181C/I/V, esta é selecionada por NVP, ETR e RPV. Esta mutação causa uma elevada redução da suscetibilidade a NVP e intermédia a ETR e RPV. Y181C reduz a suscetibilidade a DOR e EFV apenas associada a outras mutações e suprime a mutação T215Y associada à resistência a AZT [123,126,136,142]. O tratamento combinado de AZT com NVP favorece a emergência de mutações que causam resistência elevada à NVP nos codões 103,188 e 190. G190A/S são mutações não polimórficas selecionadas por NVP e EFV. Relativamente a G190S confere elevada resistência a ambos os fármacos. Esta mutação ocorre mais comumente em vírus do subtipo A [123,126,136,142]. Y188L é uma mutação não polimórfica de dois pares de base, selecionada por NVP, EFV e DOR, conferindo um alto nível de resistência a todos os NNRTI, à exceção de ETR. Y188C é menos comum que a anterior, sendo selecionada por NVP e EFV conferindo um elevado nível de

resistência aos mesmos. V106A é selecionada por NVP e DOR e por si só causa reduções intermediárias na suscetibilidade a DOR, mas quando associada a outras mutações, como F227C, confere elevados níveis de resistência a este fármaco, a NVP e EFV. V106M é selecionada, principalmente, por EFV, NVP e DOR, e é particularmente comum em vírus do subtipo C [123,126,136,142]. Relativamente às mutações acessórias, L100I ocorre muitas vezes em combinação com K103N promovendo um elevado nível de resistência ao EFV e NVP e um nível intermédio de resistência a RPV e ETR. Esta mutação acessória causa hipersuscetibilidade a AZT. K101E, K101P e M230L são selecionadas por todos os NNRTI e geralmente ocorrem em combinação com outras mutações [123,126,136,142].

1.8.4.4 Mutações de resistência aos INSTI

Os INSTI bloqueiam a integração do ADN complementar no ADN da célula infetada e representam a última classe de antirretrovirais a ser aprovada para o tratamento de indivíduos infetados pelo VIH-1 [145]. A resistência a esta classe de antirretrovirais surge a partir da alteração do alvo enzimático, seja por alteração conformacional no centro ativo da enzima ou por perturbações na coordenação dos iões metálicos, resultando numa afinidade reduzida entre o inibidor e a integrase [127]. As mutações mais comuns aos inibidores da integrase encontram-se representadas na tabela 4 [142].

Tabela 4 –Mutações de resistência mais comuns aos INSTI. A vermelho encontram-se assinaladas as mutações associadas aos níveis mais baixos de suscetibilidade. As mutações a negrito diminuem a suscetibilidade aos INSTI. As mutações em texto simples contribuem para a redução da suscetibilidade em combinação com outras mutações de resistência. Os aminoácidos apresentados em cada posição estão representados pelo código internacional IUPAC. Adaptado de [142].

Mutações de Resistência a INSTI										
Fármaco	T 66	E 92	G 118	E 138	G 140	Y 143	S 147	Q 148	N 155	R 263
BIC	K	Q	R	KAT	SAC	-	-	HRK	H	K
DTG	K	Q	R	KAT	SAC	-	-	HRK	H	K
EVG	AIK	Q	R	KAT	SAC	-	G	HRK	H	K
RAL	AIK	Q	R	KAT	SAC	RCH	-	HRK	H	K

O RAL e EVG apresentam uma baixa barreira genética, ou seja, a resistência a estes fármacos é mais provável, podendo ser conferida apenas com uma mutação [145,146]. As mutações na região codificante da integrase mais comuns e responsáveis pela resistência ao RAL, são definidas como mutações principais - Y143R/C/H, N155H e Q148H/R/K - e geralmente encontram-se associadas a outras mutações acessórias [145,146]. O RAL e EVG têm em comum as principais mutações de resistência e por isso verifica-se resistência cruzada entre os fármacos, sendo que, no entanto, mutações em T66 e E92 são predominantemente selecionadas pelo EVG [145,146]. As mutações de resistência para estes fármacos desaparecem quando se interrompe o tratamento, mas caso haja uma nova introdução do fármaco, estas voltam a surgir devido à presença na *quasispecies* de estirpes virais minoritárias resistentes [147]. Esta observação mostra que as mutações principais, selecionadas por RAL e EVG, provocam um impacto significativo no *fitness* viral e por isso são frequentemente acompanhadas por uma ou mais mutações acessórias específicas [147]. Em contraste, DTG e o BIC possuem um perfil de resistência marcadamente distinto, acompanhado de uma barreira genética mais elevada [146,148]. A combinação da mutação R263K com G118R, T66I e/ou N155H aumenta a resistência a estes dois fármacos, no entanto, não interferem com a capacidade replicativa do vírus, devido à reconstituição do *fitness* viral conferido por estas [146]. O DTG e o BIC são eficazes mesmo na presença de mutações que conferem resistência ao RAL e ao EVG, como N155H, Y143R/C/H, T66A/I/K e E92Q, exceto na presença da mutação Q148H/R/K, em que estes não atuam de forma tão eficaz, principalmente quando essa mutação se encontra associada a mutações acessórias [148].

1.8.4.5 Mutações de resistência aos PI

A resistência aos PI está associada a mutações que provêm a diminuição do volume das cadeias laterais dos aminoácidos constituintes do centro ativo da enzima. Esta alteração promove a perda de interações não-covalentes entre a PR mutante e o inibidor, resultando num decréscimo da afinidade de ligação entre estes, permitindo a reativação funcional da protease [125]. Existe uma complexidade acrescida na interpretação dos testes genotípicos de suscetibilidade aos fármacos para a região do genoma codificante da PR, pois esta apresenta-se como altamente polimórfica. Algumas mutações podem existir como polimorfismos naturais em estirpes virais de indivíduos nunca submetidos a

terapia antirretroviral [149,150]. O registo das mutações mais comuns aos inibidores de protease encontram-se representadas na tabela 5 [142].

Tabela 5 –Mutações de resistência mais comuns aos PI. A vermelho encontram-se assinaladas as mutações associadas aos níveis mais baixos de suscetibilidade. As mutações a negrito diminuem a suscetibilidade aos PI. E as mutações em texto simples contribuem para a redução da suscetibilidade em combinação com outras mutações de resistência. Os aminoácidos apresentados em cada posição estão representados pelo código internacional IUPAC. Adaptado de [142].

Mutações de Resistência a PI													
Fármaco	Mutações principais						Mutações acessórias						
	D 30	G 48	I 50	V 82	I 84	L 90	V 32	L 33	M 46	I 47	I 54	L 76	N 88
ATV	-	VM	L	ATFS	V	M	I	F	IL	V	VTALM	-	S
DRV	-	-	V	F	V	-	I	F	-	VA	LM	V	-
FPV	-	-	V	ATSF	V	M	I	F	IL	VA	VTALM	V	-
IDV	-	-	-	ATFS	V	M	I	-	IL	V	VTALM	V	S
LPV	-	VM	V	ATFS	V	M	I	F	IL	VA	VTALM	V	-
NFV	N	VM	-	ATFS	V	M	-	F	IL	V	VTALM	-	DS
SQV	-	VM	-	AT	V	M	-	-	-	-	VTALM	-	S
TPV	-	-	-	TL	V	-	I	F	IL	VA	VAM	-	-

O desenvolvimento de um elevado grau de resistência ao IDV e LPV requer a acumulação de sucessivas mutações. Geralmente, a primeira mutação detetável ocorre no codão 82, seguida da seleção de três mutações adicionais nos codões 46, 54 e 84 que diminuem, significativamente, a eficácia destes fármacos e conferem resistência cruzada a praticamente todos os PI [149,150]. A resistência ao SQV está associada, muitas vezes, às mutações G48V e L90M, sendo esta última a primeira a surgir com maior frequência, reduzindo a suscetibilidade a todos os PI, exceto ao TPV e DRV. A estas mutações podem associar-se múltiplas mutações acessórias que conferem resistência cruzada a outros fármacos desta classe [150,151]. A resistência ao LPV resulta do aparecimento de uma mutação no codão 50 – substituição de uma isoleucina por uma valina. Uma mutação no mesmo codão, mas com um aminoácido diferente – leucina em vez de valina – é característica da resistência ao ATV e confere um aumento da suscetibilidade aos demais PI [142]. Apesar das mutações ocorrerem no mesmo codão, não conferem resistência cruzada entre estes dois ARVs. No entanto a mutação seleccionada pelo LPV confere

resistência cruzada significativa ao DRV. A mutação I50V também reduz a suscetibilidade a FPV e DRV, mas aumenta a suscetibilidade a TPV [142,150]. A resistência ao NFV está associada à substituição de um ácido aspártico por uma arginina, no codão 30 da região codificante da protease. Quando esta mutação ocorre em combinação com outras mutações de resistência aos PI, está associada a uma suscetibilidade reduzida a toda esta classe [149,150].

1.8.5 Polimorfismos genéticos naturais

Os estudos que incidem na análise das resistências aos ARV, atualmente disponíveis, têm sido centrados em indivíduos infetados por isolados virais do subtipo B. No entanto, a análise dos restantes subtipos é absolutamente legítima e necessária, uma vez que se verifica um potencial para diferenciais genéticos entre subtipos para produzir padrões de mutação diferenciados que conferem resistência aos antirretrovirais [152,153]. Os polimorfismos genéticos naturais são mutações que ocorrem em vírus que nunca foram expostos à pressão seletiva dos fármacos. No entanto, alguns destes polimorfismos podem contribuir para a resistência aos antirretrovirais quando o indivíduo é sujeito à terapêutica. Deste modo, diferenças entre subtipos virais a nível de polimorfismos genéticos naturais podem culminar no desenvolvimento de padrões de resistência únicos e diferentes níveis de eficácia dos fármacos [152,154].

1.10 Epidemiologia da infeção VIH/SIDA em Moçambique

1.10.1 Dados estatísticos

A África Subsariana é a região com mais casos de infeção por VIH-1, onde cerca de 25,3 milhões de pessoas são portadoras do vírus, representando 67% de todas as infeções no mundo [155]. Moçambique localiza-se nesta região, com uma população estimada de 29,5 milhões e uma esperança média de vida de 59 anos [155,156]. Estima-se que no início dos anos 80 do século XX [157] o VIH-1 tenha sido introduzido em Moçambique e desde então houve um aumento exponencial do número de novos casos de infeção, registando-se, em 2019, cerca de 2 200 000 pessoas infetadas [156,158]. As mulheres representam cerca de 60% do total de infeções no país (1 300 000) e as crianças (<15 anos) cerca de 7% (150 000) [156,158]. A proximidade de Moçambique com países que

apresentam elevadas taxas de prevalência de infeção em adultos (Tanzânia 4,8%, Malawi 8,9%, Zâmbia 11,5%, Zimbábue 12,8%, Essuatini 27,0% e África do Sul 19,0%) e os fatores históricos e sociológicos (*e.g.* guerra civil, migração, cultura, condições económicas precárias) contribuem para uma disseminação massiva e sustentada do vírus no país [159-161]. Em 2019 foram registadas 130 000 novas infeções e 51 000 mortes [156,158]. No estudo de Nuttor et al. (2020), estimaram que a prevalência da infeção pelo VIH-1 em Moçambique é cerca de 12,6% (11,9%-14,4%) em adultos (15-49 anos), variando entre 3,2% e 28,5% em diferentes províncias [156,158,162]. As prevalências mais elevadas registam-se na região centro e sul de Moçambique, com taxas de prevalência de 12% e 19% respetivamente, em comparação com a região norte que possui uma taxa de 9% [161-163]. As taxas de prevalência encontram-se discriminadas na tabela 6.

Tabela 6 – Distribuição regional da prevalência de HIV-1 em Moçambique. Os dados registados mostram a prevalência de HIV-1 no país em função do género (feminino e masculino). Adaptado de: [162].

Províncias	Prevalências (%)	
	Feminino	Masculino
Niassa	9,80	4,40
Cabo Delgado	15,7	10,6
Nampula	5,10	6,20
Zambézia	16,2	12,7
Tete	6,10	3,20
Manica	15,2	11,2
Sofala	18,4	12,7
Inhambane	16,1	8,00
Gaza	28,5	18,8
Maputo	28,0	16,7
Cidade de Maputo	21,7	12,1

Os dados mostram também que Gaza e Maputo registam a maior prevalência de VIH-1 entre as onze províncias de Moçambique e que as mulheres são as mais afetadas, exceto na província de Nampula (Tabela 6) [162]. Com a exceção de Gaza, Maputo e Cidade de Maputo, as restantes províncias registaram um maior número de casos em áreas urbanas

face às áreas rurais [162]. Os novos casos ocorrem predominantemente na faixa etária dos 35-39 anos nas províncias de Tete, Manica, Cidade de Maputo e Maputo. Por sua vez em Gaza, Zambézia e Cabo Delgado a faixa etária com maior incidência da infecção é de 30-34 anos e em Nampula e Sofala entre os 40-49 anos [162].

1.10.2 Subtipos circulantes no país

Na epidemia de VIH-1 em Moçambique, os vírus do subtipo C são predominantes [159, 162-166]. No entanto, foi demonstrado que existem em circulação vírus de outros subtipos como A, D, G e algumas formas recombinantes (B/D, C/A1, C/G, B/C) [162, 159, 164-168]. O padrão epidémico descrito até hoje revela uma predominância acentuada no subtipo C na região sul do país, que se torna menos evidente, sem deixar de ser predominante, quando nos deslocamos para norte [167]. A presença crescente de subtipos não-C no norte do país pode ser explicada pela influência dos países da África Oriental, como Quênia, Ruanda, Tanzânia e Uganda, nos quais circulam ativamente vírus dos subtipos não-C (em particular A, D e algumas CRF) [164, 167]. Por outro lado, a região sul do país sofre uma maior influência e tráfego migratório da África do Sul, onde o subtipo C é altamente predominante [161, 164, 167].

1.10.3 Terapia antirretroviral

Em 2004, os regimes de terapia antirretroviral altamente ativa (HAART) foram instituídos de forma gratuita no país a um número crescente de indivíduos verificando-se, em consequência, uma diminuição significativa da taxa de mortalidade e morbilidade associada à infecção pelo VIH-1, para além de melhorias consideráveis no prognóstico dos indivíduos infetados [157, 163, 168]. No entanto, apesar da meta 90-90-90 [169], definida até 2020, em Moçambique, apenas 77% das pessoas infetadas conhecem a sua condição, das quais apenas 60% (1 213 000) estão em tratamento, sendo que 65% do número total são mulheres [161, 165]. Além disso, a taxa estimada de indivíduos em terapêutica com supressão viral é de apenas 45%. Relativamente às crianças, apenas 63% (95 080) têm acesso à terapêutica [156, 158]. Apesar do acesso gratuito a estes esquemas de tratamento padronizados e simplificados, ainda existem muitas dificuldades em fazer chegar os antirretrovirais às comunidades mais isoladas devido à dificuldade de deslocação dos

indivíduos infectados aos pontos de recolha ou até mesmo dos distribuidores a esses mesmos locais [170-172]. Este obstáculo promove muitas vezes a compra dos antirretrovirais no mercado negro ou a interrupção da terapêutica por longos períodos de tempo. Além disso, em países com recursos limitados, como Moçambique, há falta de infraestruturas adequadas para uma monitorização personalizada de cada indivíduo [168,170]. Os fatores mencionados, acoplados às características virais promovem uma constante alteração genética do vírus circulante associada a novas mutações que podem levar ao rápido aparecimento e disseminação de estirpes virais resistentes que limitarão as opções terapêuticas disponíveis atualmente [143,163,166,168]. Além disso, a transmissão direta de estirpes virais resistentes aos antirretrovirais em uso clínico revela-se um grande problema de saúde pública, embora, em Moçambique, as taxas de resistência transmitida aos antirretrovirais (TDR, do inglês *transmitted drug resistance mutations*) permaneçam abaixo dos 5% [168]. A terapia de primeira linha aplicada em Moçambique consiste em dois NRTI e um NNRTI e a de segunda linha, num PI (potenciado por ritonavir) e dois NRTI, como representado na tabela 7 [156,163,168]. A monitorização do tratamento é realizada usando essencialmente parâmetros imunológicos (contagem de linfócitos TCD4⁺) e critérios clínicos [168].

Tabela 7 – Registo dos regimes terapêuticos de Moçambique. Terapêutica de primeira e segunda linha usados no país.

População	Regime de primeira-linha	Regime de segunda-linha
Adultos e adolescentes	AZT + 3TC + NVP	TDF + 3TC + ATV/r
	TDF + 3TC + EFV	AZT+3TC+ATV/r
	TDF + FTC + EFV	TDF+3TC+LPV/r
		AZT +3TC+LPV/r
		TDF+FTC+ LPV/r

1.10.3.1. Mutações de resistência

Com referência às opções de tratamento de primeira linha, o número de mutações descritas é maior e mais incidente nas classes de antirretrovirais NNRTI e NRTI, como representado na tabela 8 [168,173]. Uma vez que os PI são usados em regime de segunda linha, os poucos estudos realizados indicam que não existe um número significativo de mutações principais que conferem resistência a esta classe de antirretrovirais. No entanto,

foram registadas algumas mutações acessórias, já descritas como polimorfismos comuns em estirpes virais do subtipo C (Tabela 8), que reduzem a suscetibilidade a alguns fármacos desta classe [168,174,175].

Tabela 8 – Registo de mutações de resistência. As mutações representadas são as mais comumente descritas em estudos anteriores para NRTI, NNRTI, PI e INSTI, em Moçambique.

	Mutações principais	Mutações acessórias
NRTI	M184V, T215F/Y/N, D67N, K219E/R, T69A/D/N, M41L, L210W, K70R	A62V, V75I
NNRTI	K103N, Y181C, V106A/M, G190A, K101E, V179A/D	A98G, K101Q, V108I, E138A, M230L
PI		K20R, M36I/T/L, D60E, L63A/P, I64M, V77I, I85M, L89M, I93L
IN		L74M, T97A, E157Q

Em 2018, ponderou-se a inclusão de INSTI, nomeadamente, DTG e RAL, nos regimes de tratamento, uma vez que se trata de uma classe de antirretrovirais mais recente, permitindo abordar alguns casos clínicos associados a vírus extensivamente resistentes às classes de antirretrovirais anteriormente inseridas nos regimes terapêuticos [173]. Estudos mostram que não foram detetadas resistências primárias a DTG e RAL em Moçambique, no entanto foram detetadas algumas mutações polimórficas acessórias entre estirpes virais do subtipo C (Tabela 8) [176]. A ausência de mutações de resistência principais a INSTI, em isolados de Moçambique, pode sugerir uma boa perspetiva para a introdução desta classe de antirretrovirais no país [176]. Em suma, a implementação em larga escala da HAART em Moçambique teve início há 17 anos e é de extrema importância investir em estudos longitudinais que analisem a diversidade genética e a evolução molecular do VIH-1, bem como estudos que incluam análises filogenéticas e a identificação de polimorfismos de resistência aos antirretrovirais em uso clínico, para um controlo eficaz da epidemia no país [157,166,177].

1.11 Objetivos e plano de trabalho

Os principais objetivos deste trabalho foram: 1) Caracterizar os subtipos de VIH-1 em circulação em Moçambique em 2014; 2) Caracterizar os vírus em circulação em Moçambique no que concerne ao uso dos co-recetores CCR5 e/ou CXCR4; 3) Caracterizar a frequência e distribuição de mutações de resistência aos antirretrovirais; e 4) Iniciar a caracterização das respostas em anticorpos de ligação e neutralização em Moçambique. O desenho simplificado do estudo e objetivos definidos para o trabalho desenvolvido encontra-se representado na figura 16.

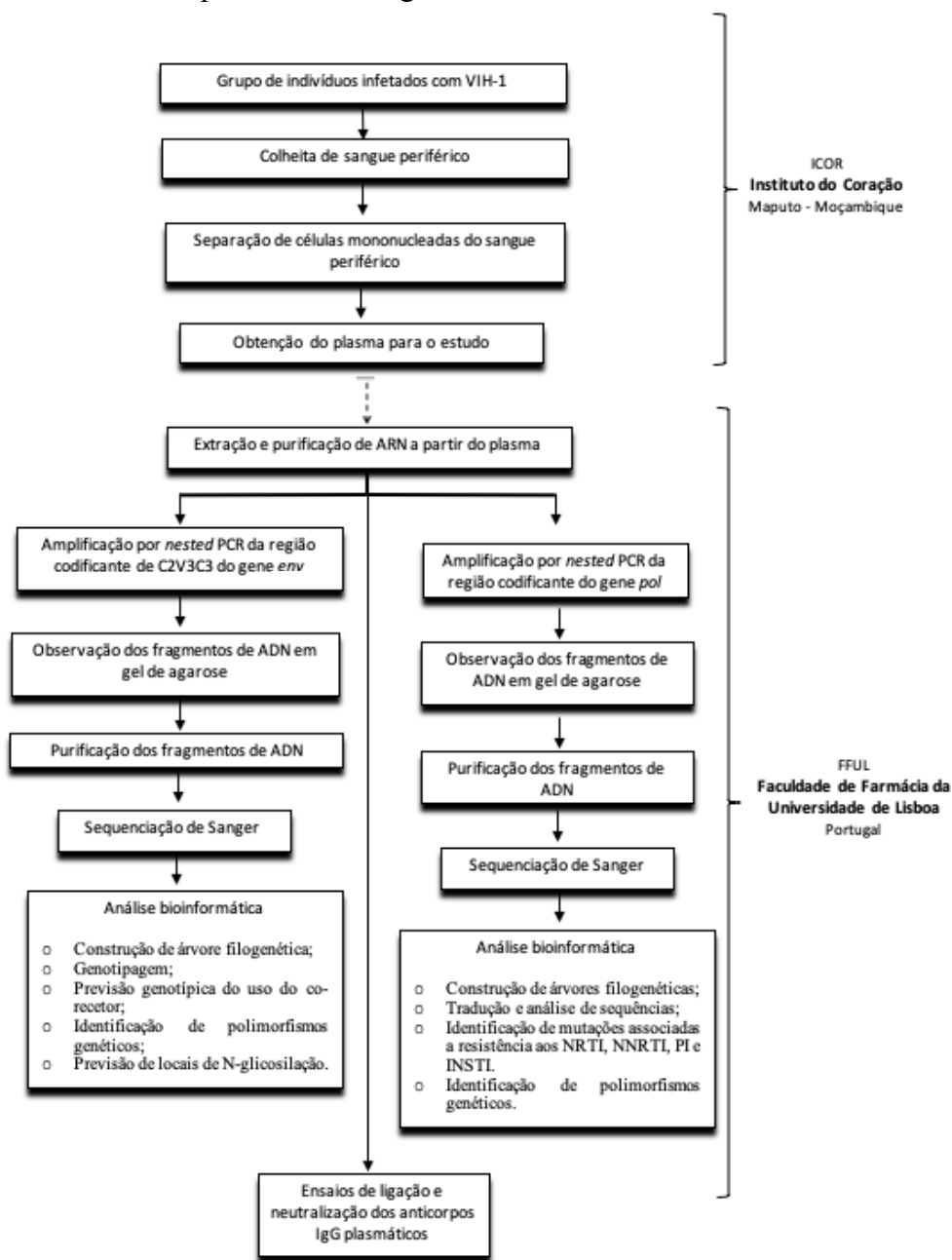


Figura 16 – Desenho base do estudo e objetivos definidos para o trabalho desenvolvido.

2. Materiais e Métodos

2.1 Caracterização da amostra populacional estudada

O presente estudo inclui 58 indivíduos seropositivos para VIH-1, aos quais foi efetuada, em 2014, uma colheita de sangue periférico para fins de diagnóstico ou controlo da infeção por VIH-1, a cargo do Instituto do Coração (ICOR) em Maputo, Moçambique. Após a consulta em base de dados do laboratório do ICOR, pela respetiva responsável e com o consentimento dos pacientes, foram fornecidos alguns dados relevantes, nomeadamente: ano de diagnóstico da infeção por VIH-1, idade, sexo, terapia antirretroviral aplicada, número de linfócitos T CD4⁺ e carga viral (Tabela 9). O estudo inclui 24 mulheres e 35 homens com idades entre os 4 e 77 anos, cujo ano de diagnóstico de infeção varia entre 2004-2014. A identificação das amostras foi realizada tendo por base um código numérico aleatório atribuído a cada paciente. O plasma obtido do sangue total foi armazenado a -80°C (duas alíquotas de 1000 µl e o restante em tubos de stock).

Tabela 9 - Caracterização epidemiológica e clínica dos pacientes do ICOR em Moçambique incluídos no estudo.

Amostra	Sexo	Idade	Ano de diagnóstico	Linfócitos T CD4 ⁺ (nº células/mm ³)	Carga viral (cópias ARN viral/ml)	Regime terapêutico
17	F	33	-----	394	-----	-----
19	F	53	-----	465	-----	-----
22	F	37	-----	433	-----	-----
24	M	48	-----	38	-----	-----
25	M	32	2013	346	682	-----
28	M	29	-----	199	-----	-----
33	M	57	-----	130	-----	-----
36	M	46	2007	43	838	AZT+3TC+NVP
38	F	37	2008	83	3566	Duovir (AZT+3TC)
44	M	48	-----	30	5490	Duovir (AZT+3TC)
45	F	34	-----	114	-----	-----
48	M	48	2014	350	-----	-----
55	M	41	-----	11	-----	-----
57	M	54	-----	92	-----	-----
62	F	20	2012	502	-----	Naive
65	M	39	-----	266	58505	Atripla (EFV+FTC+TDF)
67	F	52	2012	188	-----	Naive
70	M	76	-----	289	-----	-----
71	F	37	-----	241	-----	-----
72	F	33	2008	411	-----	Darunavir
73	F	27	2008	156	98470	Naive

74	F	58	-----	87	243984	Atripla (EFV+FTC+TDF)
77	M	77	-----	75	5040	-----
84	F	38	-----	124	143381	Atripla (EFV+FTC+TDF)
93	M	49	2012	387	-----	Duovir (AZT+3TC)
94	F	49	2009	260	1290	Naïve
95	M	47	2014	345	7929	Naïve
103	F	39	-----	487	-----	-----
105	M	45	2011	499	-----	Efavirenze
109	F	43	2004	783	370	Naïve
111	M	42	-----	68	93	-----
113	M	51	2005	851	1230	Duovir (AZT+3TC)
114	M	34	-----	102	143381	Naïve
123	F	37	-----	295	-----	-----
127	F	45	-----	184	18782	-----
130	F	36	-----	132	-----	-----
134	M	40	-----	244	-----	-----
135	F	16	2006	33	15665	Atripla (EFV+FTC+TDF)
137	M	43	-----	195	97	EVF+3TC
139	M	49	2006	343	13044	Darunavir
142	F	11	-----	670	3816	D4T + 3TC+ NVP
146	F	57	-----	167	-----	-----
150	M	37	-----	68	128970	Atripla (EFV+FTC+TDF)
151	M	69	-----	101		-----
153	F	38	-----	75	677271	Não faz
154	M	43	-----	193	-----	-----
157	M	36	-----	9	-----	-----
158	M	26	2013	344	-----	-----
160	M	44	-----	19	-----	Duovir (AZT+3TC)
163	M	46	-----	314	240	-----
170	F	14	-----	290	-----	-----
178	M	36	-----	90	-----	-----
184	M	44	-----	179	19140	-----
186	M	38	-----	167	194975	Atripla (EFV+FTC+TDF)
187	M	30	-----	360	4036	-----
188	M	47	-----	34	192735	-----
192	M	30	-----	299	91403	Atripla (EFV+FTC+TDF)
197	M	4	2008	262	611259	Atripla (EFV+FTC+TDF)

----- Dados não facultados.

2.2 Extração de ARN viral a partir de plasma

O processamento das amostras de plasma iniciou-se com a extração do ARN viral com o *QIAamp® Viral RNA Mini KIT (QIAGEN)*. Utilizou-se 1mL de plasma e procedeu-se à sua concentração através de centrifugação a 23000xg durante 60 minutos a 4°C [178]. O sobrenadante foi recolhido e armazenado a -80°C para posterior utilização em ensaios de ligação e neutralização de anticorpos IgG plasmáticos. O *pellet* foi dissolvido em 560µl de tampão de lise (AVL+*Carrier RNA*) sob condições desnaturantes, inativando possíveis RNases em solução. Adicionou-se 560µl de etanol absoluto para precipitar o ARN e transferiu-se a solução para a coluna de extração QIamp [179]. O ARN viral adsorve à membrana de sílica da coluna após duas breves etapas de centrifugação a 6000xg, ambas de 1 minuto. Posteriormente, adicionou-se 500µl de um tampão de lavagem (AW1) seguido de uma centrifugação de 6000xg 1 minuto [179]. Por forma a terminar a etapa de lavagens é adicionado um novo tampão de lavagem (AW2) para a remoção completa de contaminantes durante uma centrifugação de 6000xg 3 minutos. É feita uma nova centrifugação de 6000xg 1 minuto para remover o excesso de tampão [179]. Por fim, adicionou-se 60µl de tampão de eluição AVE, composto por água e azida de sódio (0,04%), que impede o crescimento microbiano, seguido de 2 minutos de incubação à TA e de uma etapa de centrifugação de 6000xg durante 1 minuto [179]. Fez-se alíquotas de 20µl do ARN viral eluído e armazenadas a -80°C. O risco de perda de ARN está associado ao número de descongelamentos a que é sujeito, não devendo por isso, ser descongelado mais de duas vezes [178].

2.3 Amplificação por *nested PCR* da região C2V3C3 do gene *env* e da região codificante das enzimas RT, PR e IN do gene *pol*

2.3.1 Seleção dos *primers*

Os *primers* selecionados para a amplificação da região C2V3C3 do gene *env* foram adaptados de estudos anteriores com recurso ao protocolo implementado por Bártolo et al. (2005) [180,181]. Relativamente aos *primers* para a amplificação do gene *pol* foram desenhados especificamente para este estudo, de acordo com os subtipos circulantes em Moçambique. Na Tabela 10 estão descritas as sequências nucleotídicas dos *primers* e a sua respetiva localização, segundo a estirpe de referência HXB2 de VIH-1 [48]. Estes

possuem posições degeneradas (Y= C ou T; R= G ou A) e foram sintetizados pela empresa *Invitrogen Custom Primers, by Thermo Fisher Scientific*. Os *primers* liofilizados foram hidratados com água desionizada para uma concentração de trabalho de 10pmol/μl.

Tabela 10 – Primers utilizados nas reações de amplificação e sequenciação das regiões codificantes das enzimas RT, PR e IN do gene *pol* e da região C2V3C3 do gene *env*.

<i>Primers</i>	<i>Posição^a</i>	<i>Região</i>	<i>Sequência</i>
JA167_II ^b	6856-6880	V2 loop end	AYCCAATTCCYATACATTATTGTGC
JA170 ^b	7412-7388	V4 <i>start</i>	GTGATGTATTTCARTAGAAAAATTC
JA168F ^{c,d}	6959-6983	V2 <i>loop end</i>	ACAATGYACACATGGAATTARGCCA
JA169B ^{c,d}	7350-7326	V4 <i>start</i>	AGAAAAATTCYCCTCYACAATTAAA
PR_1F ^b	2017-2037	<i>Gag p2 start</i>	AAAAGGGCTGTTGGAAATGT
PR_2R ^b	2759-2782	<i>Pol protease end</i>	CTGAAATCTACTAATTTYCTCCA
PR_3F ^{c,d}	2109-2130	<i>Gag p1 end</i>	ACAAGGGRAGGCCAGGRAATT
PR_4R ^{c,d}	2584-2606	<i>Pol p66 end p51 RT start</i>	GGGCCATCCATTCTGGCTTTA
RT_1F ^b	2387-2409	<i>Gag p6 end</i>	ATGATAGGAGGRATTGGAGGTT
RT_2R ^b	3693-3715	<i>Pol p15 RNase H start</i>	TCTTTCCCCATATTACTATGCT
RT_3F ^{c,d}	2485-2506	<i>Pol protease end</i>	CCTACACCTGTCAACATAATT
RT_4R ^{c,d}	3554-3578	<i>Pol p15 RNase H start</i>	CTTGRTAAATTTGATATGTCCATT
IN_1F ^b	4024-4046	<i>Pol p15 RNase H start</i>	AGTAAAYATAGTAACAGACTCA
IN_2R ^b	5193-5214	P31 integrase end	TGGGATGTGTACTTCTGAACT
IN_3F ^{c,d}	4171-4193	Pol p31 Integrase start	TTGGAGGAAATGAACAAGTAGA
IN_4R ^{c,d}	5076-5096	P31 integrase end	TCTAATCTTCATCCTGTCTA

^a HIV-1 HXB2, ^b Primers externos para PCR, ^c Primers internos para PCR, ^d Primer de sequenciação

2.3.2 Protocolo de amplificação da região codificante C2V3C3 do gene *env*

A região C2V3C3 foi amplificada seguindo um protocolo de amplificação por PCR de duas etapas – *Nested PCR*. As condições de amplificação foram baseadas na temperatura de hibridação dos primers (temperatura de *annealing*), no tamanho dos fragmentos de ADN a amplificar e nas características das enzimas utilizadas em cada uma das reações de PCR. As reações de polimerização enzimática em cadeia foram efetuadas no termociclador *C1000 Touch Thermal Cycler* (Bio-Rad, Portugal).

2.3.2.1 One-Step RT-PCR

A primeira reação de PCR é mediada pelo par de primers externos (JA167_III e JA170) para dar origem a um fragmento de ADN de 556pb. Utilizou-se o kit *MyTaqTM One-Step RT-PCR* (Bioline, Londres) [182]. A enzima *MyTaq HS DNA Polymerase* é ativada à temperatura de 95°C (*hot start PCR*), prevenindo a amplificação inicial de produtos inespecíficos e a formação de dímeros de *primers*, garantindo um aumento da especificidade da reação de amplificação [182]. Para além disso, suporta temperaturas mais elevadas, importante para superar as estruturas secundárias para um acesso completo à região alvo [186]. Foram usados 10µl ($\cong$ 300ng) de ARN total extraído e uma mistura de reação de 25µl constituída por: 2x MyTaq One Step Mix, 12,5µl; *Primer* JA167_II (10µM), 0,75µl; *Primer* JA170 (10µM), 0,75µl; RiboSafe RNase Inhibitor, 0,5µl; *MyTaq HS DNA Polymerase*, 0,25µl e dH₂O até perfazer o volume total. As condições de amplificação encontram-se descritas na Tabela 11.

Tabela 11 – Condições de amplificação aplicadas na primeira reação de PCR para amplificação da região C2V3C3 do gene *env*.

Etapas	Condições de amplificação	Número de ciclos
Transcrição Reversa	45°C – 20 minutos	1
Ativação da Polimerase	95°C – 1 minuto	1
Desnaturação	95°C – 30 segundos	45
<i>Annealing</i>	58°C – 35 segundos	
Extensão	72°C – 50 segundos	
Manutenção	4°C - ∞	

2.3.2.2 Segunda reação de PCR

Para a realização da segunda reação de PCR utilizou-se os primers internos JA168F e JA169B (Tabela 10), para dar origem a um fragmento de ADN final de 391pb. Foi utilizada a *NZYTaq II DNA polymerase* (Nzytech, Portugal), que consiste numa polimerase de ADN otimizada para reações de PCR padrão, produzindo bons rendimentos na amplificação de ADN em tempos de execução mais curtos [183]. A mistura de reação foi preparada para um volume final de 50µl, constituída por: 10x Reaction buffer, 5µl; *Primer* JA168F (10µM), 1,5µl; *Primer* JA169B (10µM), 1,5µl; dNTPs mix (10mM), 1µl; MgCl₂ (50mM), 3µl; NZYTaq III (5U/µl), 0,5µl e dH₂O até perfazer o volume total.

Adicionou-se 5µl do produto de amplificação do PCR anterior. As condições de amplificação encontram-se descritas na Tabela 12.

Tabela 12 – Condições de amplificação aplicadas na segunda reação de PCR para amplificação da região C2V3C3 do gene *env*.

Etapas	Condições de amplificação	Número de ciclos
Desnaturação inicial	95°C – 3 minutos	1
Desnaturação	94°C – 30 segundos	40
<i>Annealing</i>	63°C – 35 segundos	
Extensão	72°C – 40 segundos	
Extensão final	72°C – 5 minuto	1
Manutenção	4°C - ∞	

2.3.3 Protocolo de amplificação da região codificante das enzimas RT, PR, IN do gene *pol*

O protocolo de amplificação das regiões codificantes do gene *pol* correspondentes à transcriptase reversa, protease e integrase, também segue a amplificação por PCR em duas etapas. A amplificação de cada uma das enzimas é realizada por reações de PCR independentes. As condições de amplificação são baseadas nos parâmetros mencionados na secção anterior. As reações de PCR foram efetuadas no termociclador *C1000 Touch Thermal Cycler* (Bio-Rad, Portugal).

2.3.3.1 One-Step RT-PCR

A primeira reação de PCR, para cada uma das enzimas, é mediada pelo par de *primers* externos representados e organizados na tabela 13, bem como o tamanho dos fragmentos originados após as reações de retrotranscrição e amplificação. Utilizou-se o kit *MyTaqTM One-Step RT-PCR* (Bioline, London), já descrito anteriormente [182]. As condições de amplificação, bem como os volumes de preparação da mistura de reação com um volume total de 25µl, encontram-se descritas na tabela 13.

Tabela 13 – Condições de amplificação aplicadas na primeira reação de PCR para amplificação da região codificante das enzimas RT, PR e IN que constituem o gene *pol*.

<i>One Step RT-PCR</i>			
Par de primers	Tamanho do fragmento	Mistura de reação	Condições de amplificação
RT1_F RT2_R	1328pb	⇒ 2x MyTaq One-Step Mix – 12,5µl ⇒ Primer 1 (10µM) – 1,25µl ⇒ Primer 2 (10µM) - 1,25µl	50°C, 40min - 1 ciclo 95°C, 1min - 1 ciclo 95°C, 20seg; 57°C, 45seg; 72°C 1,30min – 10 ciclos 95°C, 20seg; 57°C, 45seg; 72°C 1,30min com incrementos de 3seg – 35 ciclos
PR1_F PR2_R	766pb	⇒ RiboSafe RNase Inhibitor – 1,0µl ⇒ Enzima – 0,25µl ⇒ dH ₂ O – 1,75µl ⇒ ARN - 7µl (≈300ng)	50°C, 40min - 1 ciclo 95°C, 1min - 1 ciclo 95°C, 15seg; 52°C, 45seg; 72°C 1,10min – 10 ciclos 95°C, 15seg; 52°C, 45seg; 72°C 1,10min com incrementos de 3seg – 35 ciclos
IN1_F IN2_R	1187pb		45°C, 30min - 1 ciclo 95°C, 1min - 1 ciclo 95°C, 20seg; 50°C, 45seg; 72°C 1,30min – 10 ciclos 95°C, 20seg; 50°C, 45seg; 72°C 1,30min com incrementos de 3seg – 35 ciclos

2.3.3.2 Segunda reação de PCR

Para a realização da segunda reação de PCR utilizou-se os primers representados e organizados na tabela 14. Para a preparação das misturas de reação utilizou-se o kit QIAGEN LongRange PCR. O *LongRange PCR Enzyme Mix* é constituído por uma mistura de polimerases de ADN termoestáveis recombinantes que permitem um aumento da processabilidade, garantindo uma alta taxa de extensão e uma maior fidelidade na amplificação de fragmentos mais longos [184]. O tamanho final dos fragmentos amplificados, as condições de amplificação, bem como os volumes de preparação da mistura de reação com um volume total de 50µl, encontram-se descritas na tabela 14.

Tabela 14– Condições de amplificação aplicadas na segunda reação de PCR para amplificação da região codificante das enzimas RT, PR e IN que constituem o gene *pol*.

Nested PCR			
Par de primers	Tamanho do fragmento	Mistura de reação	Condições de amplificação
RT3_F	1093pb	⇒ LongRange PCR buffer, 10X – 5µl ⇒ Primer 1 – 2,5µl	93°C, 3min - 1 ciclo
RT4_R			93°C, 30seg; 57°C, 45seg; 68°C 1,20min – 40 ciclos
PR3_F	596pb	⇒ Primer 2 – 2,5µl ⇒ dNTP mix (10mM) – 2,5µl ⇒ Enzima – 0,4µl	93°C, 3min - 1 ciclo
PR4_R			93°C, 30seg; 62°C, 45seg; 68°C 1min – 40 ciclos
IN3_F	925pb	⇒ dH ₂ O – 34,1µl ⇒ cDNA - 3µl	93°C, 3min - 1 ciclo
IN4_R			93°C, 30seg; 53°C, 45seg; 68°C 1,20min – 40 ciclos

2.4 Observação dos fragmentos de ADN em gel de agarose

Os fragmentos de ADN obtidos foram observados após migração por eletroforese em gel de agarose (SeaKem® LE, EUA) a 1% (p/v), em tampão SB 1X (Tabela A2 em anexo), com *Greensafe Premium* (0,05µl/ml) (Nzytech, Portugal). O volume de cada produto de amplificação aplicado no gel de agarose foi de 5µl, a uma voltagem de 90V durante, aproximadamente, 1 hora. Por forma a verificar a presença do fragmento de ADN pretendido, pelo padrão de migração criado em função dos pesos moleculares, foi usado o marcador *NzyDNA Ladder III* (Nzytech, Portugal). A visualização das bandas obtidas foi efetuada com recurso a um transiluminador (*Cleaver Scientific*, UK), com exposição a radiação ultravioleta ($\lambda=254\text{nm}$).

2.5 Purificação dos produtos de PCR

O sistema utilizado para proceder à purificação dos produtos obtidos foi o NucleoSpin® Gel and PCR Clean-Up (*Macherey-Nagel*, Alemanha) que permite purificar os fragmentos de ADN diretamente a partir dos produtos de PCR [185]. O processo inicia-se com a adição de 10µl dH₂O estéril a 45µl do produto de amplificação aos quais se junta o dobro do volume de partida, neste caso 110µl de tampão de ligação NTI, permitindo

uma melhor adsorção do ADN à coluna de purificação (*NucleoSpin® Gel and PCR Cleanup Column*). Procedeu-se à centrifugação da amostra na coluna durante 30 segundos a 11000xg. De seguida adiciona-se 700µl tampão de lavagem NT3 para a remoção de contaminantes e centrifugou-se 30 segundos a 11000xg. Repete-se a etapa anterior com apenas 250µl tampão de lavagem NT3. Ao protocolo do fabricante adicionou-se uma nova etapa de centrifugação de 3 minutos a 11000xg para remoção completa do tampão NT3. Por fim, o ADN puro é eluído com 30µl de tampão NE (5mM Tris/HCl, pH 8,5) após 2 minutos de incubação à temperatura ambiente, seguidos de 3 minutos de centrifugação a 11000xg.

2.6 Sequenciação de Sanger

A sequenciação nucleotídica dos fragmentos de ADN purificados foi realizada pela empresa STAB VIDA (Portugal) pelo sistema You Tube It, com recurso ao método de terminação de cadeia, descrito por Sanger et al. (1977) [186]. A sequenciação foi efetuada nos sentidos *forward* e *reverse*, com os primers internos descritos para cada reação de PCR em 2.4.1. Enviou-se 10µl do produto de amplificação purificado, com uma concentração maior ou igual a 20ng/µl, aos quais foram adicionados 3µl de primer por reação.

2.7 Ensaio imunoenzimático (ELISA, Enzyme Linked Immunosorbent Assay) para testar a diversidade antigénica dos doentes moçambicanos

Para testar a antigenicidade de polipéptidos recombinantes da região C2V3C3 da gp120 (SU) dos subtipos B, C, G, H e CRF02_AG do VIH-1, previamente expressos em *E.coli*, utilizou-se um ensaio imunoenzimático ELISA descrito previamente [187-189]. Os ensaios ELISA foram realizados em microplacas de poliestireno (*MaxiSorp – Thermo Fisher Scientific*, US) de 96 poços. Estas foram revestidas com 100µl, por poço, com os péptidos a testar, numa concentração de 0,5µg/ml em tampão de bicarbonato de sódio e incubou-se durante 24h a 4°C. Após três lavagens com PBS (137mM NaCl, 2,7mM KCl, 100mM Na₂HPO₄, pH 7,4), contendo 0,05% de Tween-20 (PBS-T), adicionou-se 200µl de bloqueante (gelatina a 1% em tampão PBS) em cada poço e incubou-se durante 1 hora à temperatura ambiente. Procedeu-se à lavagem da microplaca (3 vezes) e adicionou-se 100µl das diferentes amostras de plasmas (diluídas a 1:100 em PBS-T com 1% de gelatina) em duplicado, seguido de um período de incubação de 2 horas e 30 minutos.

Uma mistura de plasmas de indivíduos saudáveis seronegativos para VIH-1) serviu como controlo do ensaio, utilizado para critérios de validação do teste. Lavou-se a microplaca (5 vezes) e adicionou-se 100µl de anticorpo de cabra anti-IgG humano conjugado com a enzima peroxidase (*Jakson Immuno Research*, UK) e incubou-se durante 1h. Após 5 lavagens com tampão PBS-T, adicionou-se 50µl de substrato da enzima (tetrametilbenzidina) durante 10 minutos. Por fim, adicionou-se 50µl de ácido sulfúrico (1N) para parar a reação. Na presença de anticorpos de ligação específicos para os antígenos testados, observa-se a cor azul. A leitura das absorvâncias foi realizada a um comprimento de onda de 450 nm no luminómetro (*Microplate Reader Biotek*, EUA). Os resultados do ensaio são expressos quantitativamente como proporções entre a densidade ótica (DO) da amostra clínica e o cut-off (DO média da amostra clínica negativa mais duas vezes o valor de absorvância correspondente ao desvio padrão). Considera-se que existe interação entre os anticorpos de ligação e os polipéptidos em estudo quando $DO/cut-off > 1$ [188,189].

2.8 Purificação e quantificação de anticorpos IgG

O processo de purificação de anticorpos IgG, a partir do plasma dos indivíduos infetados por VIH-1 em estudo, foi efetuado com recurso ao kit Protein A and G Mag Sepharose™ Xtra (*GE Healthcare*, EUA) [190]. Este sistema promove uma purificação por afinidade com base em esferas magnéticas constituídas pelas proteínas A e G, provenientes de bactérias, as quais interagem de forma específica com a região Fc dos anticorpos presentes na amostra biológica [191]. O processo inicia-se com 100µl da solução que contém as esferas magnéticas, num tubo de 1,5ml e a sua lavagem com 500µl de tampão PBS [190,192]. Com o magneto, remove-se o tampão e adiciona-se às esferas magnéticas 150µl de plasma, seguido de um período de incubação de 60 minutos, sob agitação orbital. Segue-se duas etapas de lavagem com 500µl de tampão PBS, homogeneizando a solução com o vórtex entre lavagens. O tampão é removido com auxílio do magneto. Adiciona-se 100µl de tampão de eluição (100mM glicina-HCl, pH 2,8) ao tubo de 1,5ml, por forma a quebrar a interação entre as IgG e as proteínas A e G [190,192]. O sobrenadante contendo as IgG é recolhido para um novo tubo de 1,5ml, contendo 10µl de Tris-HCl (1M, pH 9,0), que auxilia na preservação da atividade dos

anticorpos a um pH de 7,0-7,4 [190,192]. A concentração das IgG purificadas foi determinada utilizando o reagente de Bradford [193].

2.9 Ensaios de Neutralização da infecção viral

A atividade neutralizante dos anticorpos IgG purificados foi testada contra três pseudovírus do painel global de estirpes de referência definida por Montefiori et al (2014) [194] e um vírus *tier 1*- pNL4.3 (Tabela 15) [195]. Para os ensaios foi usada a linha celular TZM-bl com gene repórter luciferase [196].

Tabela 15 – Pseudovírus VIH-1 utilizados nos ensaios de neutralização.

Pseudovírus	Genótipo	Origem	<i>Tier</i>	Ano
pTRO11 env	B	Itália	2	1995
pBJOX2000 env	CRF07_BC	China	2	2007
pCE1176 env	C	Malawi	2	2004
pNL4.3 env	B	-----	1	1991

Os ensaios de neutralização foram efetuados como descrito anteriormente [187,188,192]. Resumidamente, 10000 células em 100µl de meio de crescimento completo (DMEM suplementado com 10% de soro bovino fetal, 1% L-glutamina e 0,5% penicilina/estreptomicina) foram adicionadas a placas de cultura de fundo plano de 96 poços (Nunc) por um período de incubação de 24h a 37°C. Subsequentemente, 100µl de cada vírus (correspondendo a 200 TCID₅₀) foi incubado durante 1h a 37°C, com 75µg/ml de anticorpos purificados (diluição de 1:100), num volume total de 250µl de meio de crescimento completo, contendo DEAE-Dextran (19,6µg/ml). Esta mistura foi adicionada às células por um período de incubação de 48h a 37°C. Para cada amostra, o ensaio foi realizado em duplicado, com um controlo de células (DMEM com células) e um controlo de vírus (DMEM com células e vírus) para efeitos de validação do ensaio. Removeu-se o meio de cultura de cada poço e analisou-se a atividade da luciferase num luminómetro (TECAN), a um comprimento de onda de 613nm, usando o kit Pierce™ Firefly Luc One-Step Glow Assay (*Thermo Scientific*, US) [195,196]. A atividade neutralizante percentual foi calculada através: $[1-(CVA-CC)/(CV-CC)] \times 100$, sendo CVA a média de RLUs (unidade de luz relativa) dos poços contendo células, vírus e anticorpos; CC e CV a média de RLUs correspondentes aos poços contendo o controlo de células e o controlo de

infecção, respetivamente. Os resultados são considerados válidos se a média de RULs obtida em CV for cerca de 10 vezes superior à média de RLUs obtida em CC.

2.10 Análise bioinformática das sequências nucleotídicas

2.10.1 Edição de sequências e homologia genética

A sequenciação nucleotídica foi realizada no sentido direto e reverso, obtendo-se duas sequências por cada isolado viral. Deste modo, os cromatogramas relativos a ambas as sequências foram analisados e manualmente editados no programa BioEdit Sequence Alignment Editor (versão 7.0.9.0), tendo como objetivo a construção de uma sequência consenso para cada isolado viral. Isto permite definir de forma mais precisa e acertada que nucleótido ocupa efetivamente cada posição. Procedeu-se à pesquisa de homologia genética entre as sequências nucleotídicas dos isolados de cada doente e sequências nucleotídicas de outros doentes e de isolados de referência existentes nas bases de dados, através do programa on-line BLAST (*Basic Local Alignment Search Tool*) (disponível em <https://blast.ncbi.nlm.nih.gov/Blast.cgi>), com o objetivo de confirmar a identidade das sequências obtidas.

2.10.2 Subtipagem por análise filogenética

As sequências obtidas para cada uma das regiões codificantes em estudo (C2V3C3, RT, IN e PR) foram alinhadas, através do programa *on-line* MAFFT version 7 (*Multiple alignment program for amino acid or nucleotide sequences*) (disponível em <https://mafft.cbrc.jp/alignment/server/>) [197], com sequências de referência de VIH-1 do grupo M (subtipos, sub-subtipos e as CRF mais comuns) obtidas a partir da base de dados *Los Alamos HIV Database* (disponível em <https://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html>). Os alinhamentos múltiplos obtidos foram editados no programa BioEdit Sequence Alignment Editor (versão 7.0.9.0).

O modelo de substituição nucleotídica mais adequado às sequências em estudo foi determinado utilizando o programa ModelTest e as árvores filogenéticas foram construídas com recurso ao método *Maximum Likelihood* [198] no programa IQ-Tree V.1.6.12 (disponível em <http://www.iqtree.org>). Os modelos escolhidos foram GTR+F+R6 para as sequências da região C2V3C3 e GTR+F+R5, TVM+F+R4 e

TVM+F+R3 para as sequências das regiões codificantes da RT, PR e IN, respetivamente. Foi usado o método de *bootstrapping*, com base em 1000 amostragens aleatórias, para garantir a robustez da inferência filogenética e das topologias sugeridas. Considerou-se significativo obtenção de valores de bootstrap iguais ou superiores a 70% [199]. As árvores filogenéticas foram visualizadas e editadas no programa FigTree (V1.4.4) (disponível em <http://tree.bio.ed.ac.uk/software/figtree/>). As distâncias genéticas entre as sequências dos diferentes genes em estudo foram calculadas segundo os modelos aplicados na construção das árvores filogenéticas com o método *pairwise distance*, implementado no programa MEGA V7 (disponível em <https://www.megasoftware.net>) [200].

2.10.3 Análise de hipermutação e de recombinação

As sequências nucleotídicas, cujo posicionamento na árvore filogenética se revelou ambíguo relativamente aos grupos monofiléticos formados, ou que apresentaram uma elevada divergência filogenética foram analisadas quanto à identificação de potenciais recombinantes através do método *bootscanning* implementado no programa Simplot 3.5.1. Os isolados virais com genoma mosaico, também foram analisados pelo método de *bootscanning* a partir de sequências concatenadas, resultantes da justaposição das sequências nucleotídicas das regiões codificantes C2V3C3, RT, PR e IN no programa ChromasPro (V2.1.10). As formas recombinantes identificadas foram desenhadas através da ferramenta on-line Recombinant HIV-1 Drawing Tool (disponível em https://www.hiv.lanl.gov/content/sequence/DRAW_CRF/recom_mapper.html). Todas as sequências em estudo foram analisadas quanto à presença de fenómenos de hipermutação por ocorrência de hiperadenilação (G→A), por ação das APOBEC, através da ferramenta on-line Hypermut 2.0 (disponível em https://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html?sample_%20input=1) [201].

2.11 Análise bioinformática das sequências aminoacídicas

2.11.1 Análise de resistências aos antirretrovirais

As sequências aminoacídicas correspondentes às enzimas virais RT, PR e IN foram analisadas quanto à existência de mutações associadas à diminuição da suscetibilidade aos antirretrovirais, bem como à existência de polimorfismos genéticos. Esta análise foi

realizada no programa on-line HIVdb baseado no algoritmo interpretação de resistência genotípica da *Stanford University HIV Drug Resistance Database* (versão 8.9) (disponível em <https://hivdb.stanford.edu/hivdb/by-sequences/>). O relatório obtido para cada sequência evidencia as mutações principais, mutações acessórias e outras mutações não associadas a resistência. Para cada mutação é atribuído um valor gerado pelo algoritmo, relativo a cada fármaco, que reflete a probabilidade dessa mutação estar associada à aquisição de resistência fenotípica a esse mesmo fármaco. Após esta classificação, os dados são convertidos qualitativamente em cinco categorias: suscetível (0-9), resistência potencial de nível baixo (10-14), resistência de nível baixo (15-29), resistência de nível intermédio (30-59) e resistência de nível elevado (>60). A resistência potencial de nível baixo definida pelo algoritmo do programa indica uma exposição prévia a determinado antirretroviral da respetiva classe terapêutica, embora o vírus seja suscetível ao fármaco. A resistência de nível baixo indica uma diminuição da suscetibilidade ao fármaco e por isso em caso de tratamento com o mesmo, o indivíduo pode apresentar uma resposta virológica incompleta. No nível de resistência intermédia, os níveis de suscetibilidade ao fármaco são significativamente reduzidos e no nível de resistência elevada, o vírus apresenta suscetibilidade muito reduzida ou inexistente ao antirretroviral em questão.

2.11.2 Variabilidade das sequências aminoacídicas obtidas

As sequências nucleotídicas em estudo foram traduzidas, segundo o código genético universal, no programa on-line EMBOSS Transeq (disponível em https://www.ebi.ac.uk/Tools/st/emboss_transeq/). A construção do alinhamento múltiplo contendo as sequências em estudo foi realizada no programa MAFFT e submetidas na ferramenta bioinformática *Analyse Align* (disponível em https://www.hiv.lanl.gov/content/sequence/ANALYZEALIGN/analyze_align.html). Esta ferramenta gera WebLogos que permitem analisar as sequências aminoacídicas quanto ao seu grau de conservação de forma simplificada e imediata. O *output* gerado apresenta-se como um gráfico, onde cada coluna do alinhamento se encontra representada por letras, que correspondem aos aminoácidos das sequências em estudo, segundo o código IUPAC (do inglês, *International Union of Pure and Applied Chemistry*) [202]. O tamanho de cada letra é proporcional à frequência com que determinado aminoácido é observado naquela posição. As letras de cada coluna são ordenadas da mais frequente

para a menos frequente, de modo a que a sequência consenso esteja representada no topo da coluna. O padrão de cores selecionado representa o grau de hidrofiliidade ou hidrofobicidade de cada aminoácido, permitindo avaliar a natureza conservativa ou não dos polimorfismos encontrados [202]. A variabilidade das sequências também foi analisada com recurso à ferramenta on-line *Entropy* (disponível em https://www.hiv.lanl.gov/content/sequence/ENTROPY/entropy_one.html). Esta ferramenta permite analisar quantitativamente a variabilidade das sequências aminoacídicas em todas as posições de um determinado alinhamento múltiplo, permitindo restringir e definir regiões com interesse específico. O *output* gerado consiste numa tabela, onde se encontra o valor de entropia atribuído a cada posição aminoacídica – valor mais baixo corresponde a uma menor variabilidade e vice-versa.

2.11.4 Análise de Locais de N-glicosilação

As sequências aminoacídicas da região C2V3C3 foram analisadas quanto a potenciais locais de N-glicosilação (motivo NXS/T, em que X é qualquer aminoácido diferente de prolina). Esta análise foi realizada na ferramenta on-line *N-GlycoSite*, disponível em <https://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html>.

2.11.5 Predição do tropismo viral

A análise de sequências aminoacídicas da *loop* V3, com o objetivo de determinar o uso preferencial dos co-recetores CCR5 e/ou CXCR4, foi realizada nos programas on-line geno2pheno_[coreceptor] (disponível em <https://coreceptor.geno2pheno.org>), PhenoSeq (disponível em <http://tools.burnet.edu.au/phenoseq/>) e Web PSSM (disponível em <https://indra.mullins.microbiol.washington.edu/webpssm/>).

Geno2pheno_[coreceptor] é uma ferramenta bioinformática, desenvolvida pelo Instituto de informática Max Planck, que para além da genotipagem automática de sequências V3, permite determinar a possibilidade do uso do co-recetor CCR5 por comparação com a sequência de referência HXB2 [203]. Além disso, o programa permite a análise de parâmetros clínicos adicionais (*e.g.* carga viral, número de células T CD4⁺ e células T CD8⁺) que correlaciona com marcadores clínicos de indivíduos *naive*, inseridos no sistema [204]. O nível de significância usado foi de 10%, de acordo com as recomendações do *European Consensus Group* para a gestão dos testes de tropismo de

VIH-1 [205]. PhenoSeq, desenvolvido pelo Instituto Burnet, funciona como o programa descrito anteriormente, por comparação das sequências em estudo com a sequência de referência HXB2. Esta ferramenta bioinformática é específica para os subtipos A, B, C e D do VIH-1 e as CRF01_AE e CRF02_AG [206]. Por fim, o programa PSSM analisa as sequências em estudo gerando matrizes com valores de probabilidade para a determinação do tropismo, com base na acumulação de mutações na região V3 que levam à alteração da sua carga global bem como à substituição de aminoácidos específicos nas posições 11 e 25. Estas matrizes são comparadas com as de sequências com fenótipo conhecido gerando o resultado mais provável. A desvantagem mais evidente deste programa reside no facto da análise ser limitada a isolados virais do subtipo B e C [207].

2.12 Análise estatística

A análise estatística foi realizada no GraphPad Prim versão 5.1 para Windows (GraphPad Software), com um nível de significância de 5%. Testes não paramétricos foram usados para comparar medianas entre duas variáveis, nomeadamente o teste de Mann-Whitney. Para testes paramétricos entre duas variáveis foi usado o teste t de Student. Para comparar 3 ou mais variáveis foi usado o teste one-way ANOVA. O teste exato de Fisher bilateral e o teste Chi-Quadrado (χ^2), aplicados à análise de tabelas de contingência 2x2, foram usados para avaliar estatisticamente diferenças entre frequências obtidas nos grupos em estudo. Para estudar a evolução de duas variáveis foram realizadas regressões lineares e coeficientes de correlação de Spearman.

3. Resultados

3.1 Análise filogenética das sequências nucleotídicas

Os resultados obtidos, relativos à subtipagem através da análise filogenética encontram-se resumidos na tabela 16. Todas as sequências consideradas para a construção de árvores filogenéticas foram analisadas quanto à ocorrência de fenómenos de hipermutação. Nenhuma sequência obteve valores estaticamente significativos ($p < 0,05$) após análise pelo teste exato de Fisher (incluído no algoritmo do programa Hypermut 2.0) para considerar a ocorrência de hipermutações.

Tabela 16 – Subtipos genéticos em circulação em Moçambique.

Genótipo					Genótipo				
Amostra	C2V3C3	PR	RT	IN	Amostra	C2V3C3	PR	RT	IN
17	C	----	C	----	109	C	----	----	----
19	B	----	----	----	111	B	C	C	----
22	C	----	----	----	113	C	C	----	----
24	C	C	C	C	114	C	C	----	C
25	C	C	C	----	123	C	----	----	----
28	C	C	C	C	127	C	----	C	C
33	C	----	----	C	130	C	C	----	----
36	C	----	----	----	134	C	----	----	----
38	C	C	----	----	135	C	C	C	C
44	C	C	----	----	137	C	----	----	----
45	C	C	C	C	139	C	A	B	C
48	B	C	C	C	142	C	----	----	----
55	C	C	C	C	146	C	C	----	----
57	C	C	C	C	150	C	C	C	C
62	C	----	----	----	151	C	C	----	----
65	C	----	----	----	153	C	C	C	D
67	B	----	----	----	154	C	C	C	C
70	C	C	----	----	157	C	C	C	C
71	C	----	----	----	158	C	----	----	----
72	C	C	----	----	160	C	C	C	C
73	C	C	C	----	163	C	----	----	----
74	C	----	----	----	170	C	A	D	C
77	C	CRF02_AG	D	C	178	C	A	C	----
84	C	C	----	----	184	C	C	C	C
93	C	----	----	----	186	C	B	C	D
94	C	----	----	----	187	C	C	D	----
95	C	----	----	----	188	C	C	C	C
103	C	----	----	----	192	C	C	C	C
105	C	----	----	----	197	A1	A	----	----

---- Não foi possível amplificar.

3.1.1 Análise filogenética das sequências nucleotídicas C2V3C3

A análise filogenética das 58 sequências nucleotídicas da região C2V3C3 mostra que a maioria das sequências pertence ao subtipo C (91%, N=53), havendo ainda algumas sequências dos subtipos B (7%, N=4) e A1 (2%, N=1) (Tabela 16 e Figura 17). As sequências MZ_160, MZ_94 e MZ_158 formam um grupo monofilético com a sequência de referência C_BE_HQ912711 de origem belga. Este grupo encontra-se distante dos restantes isolados do subtipo C sugerindo que estes isolados são mais relacionados com o subtipo C proveniente da Bélgica do que com os restantes isolados de subtipo C de origem asiática ou africana.

As sequências MZ_109, MZ_103, MZ_134 e MZ_137 aparentam estar muito mais próximas entre si, do ponto de vista filogenético, do que de qualquer sequência de referência incluída na árvore. Este fato coloca em consideração a possibilidade de contaminação entre amostras aquando da manipulação laboratorial, ou a hipótese destes indivíduos estarem epidemiologicamente relacionados, através da transmissão numa pequena comunidade ou num grupo com um comportamento de risco semelhante que permita a transmissão do vírus entre si. No entanto, os dados facultados dos pacientes não permitem confirmar estas possibilidades.

As 20 sequências designadas por MZ_44, MZ_24, MZ_154, MZ_74, MZ_127, MZ_178, MZ_170, MZ_105, MZ_123, MZ_146, MZ_142, MZ_186, MZ_151, MZ_84, MZ_25, MZ_188, MZ_157, MZ_28, MZ_62 e MZ_139 apresentam um posicionamento ambíguo na árvore filogenética e uma elevada diversidade genética relativamente ao nó ancestral, não obstante a topologia da árvore remeter para uma evidente homologia das sequências em estudo com sequências de referência do subtipo C. Estas sequências foram por isso, analisados por *bootscanning* para pesquisa de eventuais fenómenos de recombinação intergenotípica, os quais não se verificaram (gráficos de *bootscanning* não apresentados).

Relativamente ao grupo monofilético formado pelas sequências de referência representativas da CRF07_BC e pelas sequências MZ_36, MZ_135 e MZ_153, a inferência filogenética consistente (bootstrap 72%) permite, em primeira instância, classificar as sequências em estudo como sendo CRF07_BC. No entanto os gráficos de

bootscanning revelam homologia genética com o subtipo C ao longo do fragmento em análise, e por isso a ausência de pontos de recombinação (gráfico de *bootscanning* não apresentado).



Figura 17 – Árvore Filogenética de sequências C2V3C3 de Moçambique. Relações filogenéticas entre 173 sequências nucleotídicas da região codificante C2V3C3, com base num total de 392 nucleótidos. A árvore filogenética inclui 58 sequências em estudo (representadas com o prefixo MZ), 100 sequências obtidas no GenBank por pesquisa de homologia genética com o programa BLAST e 15 sequências de referência do grupo M do VIH-1 (F, G, H, J, K, CRF05_DF, CRF10_CD, CRF41_CD, CRF12_BF). A designação das sequências de referência é iniciada com a letra correspondente ao genótipo, seguida do local de origem e o número de acesso no GenBank. As origens geográficas das sequências de referências são: BD – Bangladesh, BE – Bélgica, CM – Camarões, CN – China, CY – Chipre, ET – Etiópia, IN – Índia, IT – Itália, KE – Quênia, MW – Malawi, PG – Papua Nova Guiné, PT – Portugal, RDC – República Democrática do Congo, SE – Suécia, TW – Taiwan, TZ – Tanzânia, UK – Reino Unido, ZA – África do Sul, ZM – Zâmbia, ZW – Zimbabwe.

Este resultado permite concluir que os segmentos do genoma viral em análise são semelhantes ao subtipo C da CRF07_BC, o que justifica a posição destas CRF junto de outras sequências de referência que pertencem ao subtipo C. Assim, as sequências MZ serão subtipo C com grande homologia com a região C destas CRF07_BC (Figura 18). As restantes sequências incluídas na árvore filogenética foram classificadas inequivocamente como subtipo C.

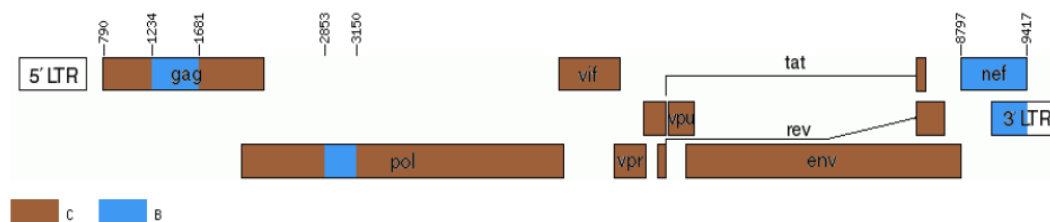


Figura 18– Mapa genómico da CRF07_BC. As regiões azuis representam o subtipo B e as castanhas o subtipo C.

3.1.2 Análise filogenética das sequências nucleotídicas RT

A análise revelou que 85% dos isolados virais eram do subtipo C, 12% do subtipo D e 3% do subtipo B (Figura 19). As sequências associadas a nós terminais de ramos muito longos, que representam possíveis sequências divergentes, foram analisadas quanto à presença de sequências recombinantes o que não se verificou em nenhum dos casos.



Figura 19 – Árvore filogenética de sequências codificantes da RT de Moçambique. Relações filogenéticas entre 145 sequências nucleotídicas da região codificante da RT, com base num total de 1021 nucleótidos. A árvore filogenética inclui 26 sequências em estudo (representadas com o prefixo MZ), 102 sequências obtidas no GenBank por pesquisa de homologia genética com o programa BLAST e 17 sequências de referência do grupo M do VIH-1 (A, D, F, G, H, J, K, CRF07_BC e CRF08_BC). A designação das sequências de referência é iniciada com a letra correspondente ao genótipo, seguida do local de origem e o número de acesso no GenBank. As origens geográficas das sequências de referências são: AT – Áustria, BW – Botswana, CA – Canadá, ES – Espanha, IN – Índia, ME – Montenegro, MW – Malawi, MZ – Moçambique, SE – Suécia, SN – Senegal, UG – Uganda, USA – Estados Unidos da América, UK – Reino Unido, ZA – África do Sul, ZM – Zâmbia, ZW – Zimbabwe.

3.1.3 Análise Filogenética das sequências nucleotídicas PR

Relativamente à análise da árvore filogenética das sequências nucleotídicas da protease, conclui-se que 85% das sequências pertenciam a isolados virais do subtipo C, 9% do subtipo A, 3% do subtipo B e 3% CRF02_AG (Tabela 16 e Figura 20).

As sequências MZ_113, MZ_188, MZ_38, MZ_157, MZ_70, MZ_151, MZ_160, MZ_135, MZ_48 e MZ_25 foram analisadas quanto à existência de possíveis eventos de recombinação por estarem associadas a nós terminais de ramos muito longos, bem como as sequências MZ_28 e MZ_57 por apresentarem segregações precoces do grupo que contém as demais sequências de referência. No entanto não foram detetados eventos de recombinação nestas sequências (gráficos de *bootscanning* não apresentados).

A ausência de eventos de recombinação também se verificou no gráfico de *bootscanning* da sequência MZ_77, classificada como CRF02_AG. O gráfico revelou uma homologia inequívoca com o subtipo G. A análise da CRF02_AG (Figura 21) mostra que a região do genoma correspondente à protease (2253-2549 nucleótidos) é maioritariamente do subtipo G, justificando, possivelmente, o resultado obtido através do programa Simplot. Os grupos formados pelas sequências de referência do subtipo A e pelas sequências de referência CRF02_AG, incluindo a sequência MZ_77, encontram-se relacionados do ponto vista filogenético com um valor significativo de *bootstrap* de 71%. Desta forma, as considerações apresentadas suportam a hipótese de MZ_77 se tratar efetivamente de uma CRF02_AG, apesar de não ser evidente somente pela análise do gráfico de *bootscanning*.

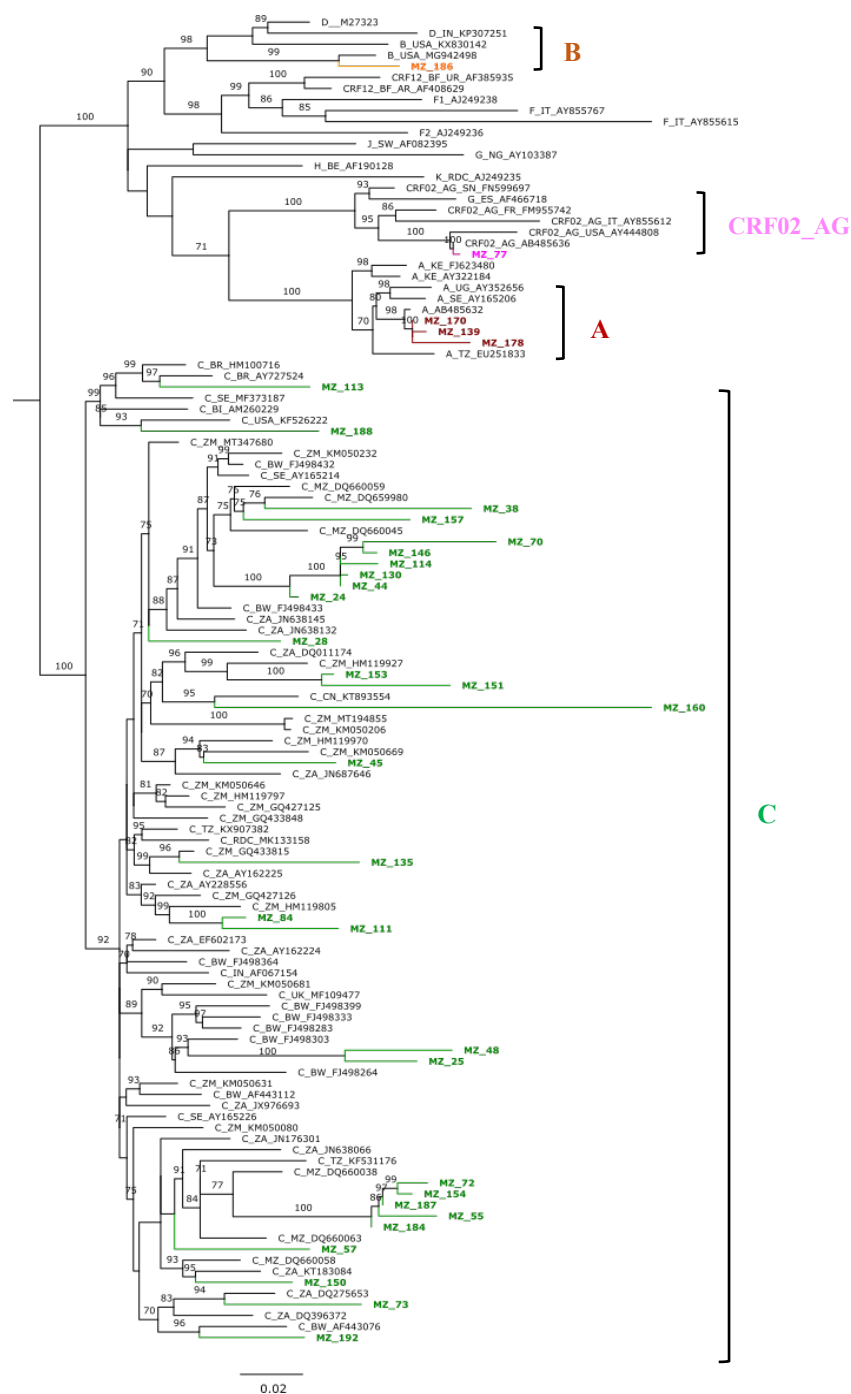


Figura 20 – Árvore Filogenética de sequências codificantes da PR de Moçambique. Relações filogenéticas entre 120 sequências nucleotídicas da região codificante da PR, com base num total de 460 nucleótidos. A árvore filogenética inclui 34 sequências em estudo (representadas com o prefixo MZ), 76 sequências obtidas no GenBank por pesquisa de homologia genética com o programa BLAST e 10 sequências de referência do grupo M do VIH-1 (F, G, H, J, K e CRF12_BF). A designação das sequências de referência é iniciada com a letra correspondente ao genótipo, seguida do local de origem e o número de acesso no GenBank. As origens geográficas das sequências de referências são: AT – Áustria, BW – Botswana, CA – Canadá, ES – Espanha, IN – Índia, ME – Montenegro, MW – Malawi, MZ – Moçambique, SE – Suécia, SN – Senegal, UG – Uganda, USA – Estados Unidos da América, UK – Reino Unido, ZA – África do Sul, ZM – Zâmbia, ZW – Zimbábwe.

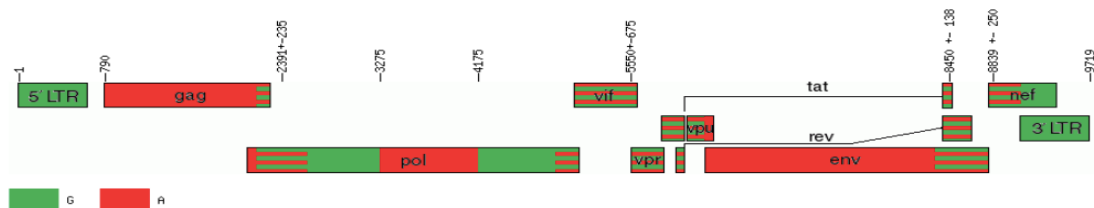


Figura 21 – Mapa genômico da CRF02_AG. As regiões vermelhas representam o subtipo A e as verdes o subtipo G. Adaptado de [26].

3.1.4 Análise Filogenética das sequências nucleotídicas IN

A árvore filogenética das 22 sequências nucleotídicas da região codificante da enzima integrase encontra-se representada na figura 22, onde se verifica que 91% das sequências estão agrupadas com sequências de referência do subtipo C e os restantes 9% com sequências de referência do subtipo D.

A sequência MZ_184 forma um grupo monofilético com sequências de referência do subtipo C de Moçambique, no entanto, encontra-se inserido no grupo que diverge para dar origem à maioria das estirpes do subtipo não-C. A topologia observada pode sugerir que MZ_184 está relacionado do ponto de vista filogenético com um ou mais subtipos deste grupo ou que existem alterações significativas ao nível das sequências nucleotídicas de referência do subtipo C provenientes de outras regiões em comparação com as duas sequências de referência do subtipo C de Moçambique. Por forma a confirmar a classificação do isolado viral em estudo como subtipo C, procedeu-se à análise de possíveis eventos de recombinação, os quais não se verificaram.

As sequências MZ_157, MZ_170, MZ_45, MZ_192 associadas a nós terminais de ramos muito longos, representando possíveis sequências divergentes, foram analisadas pelo método *bootscanning*, tal como as sequências MZ_24, MZ_55, MZ_28 e MZ_135 por apresentarem segregações precoces do grupo que contém as demais sequências de referência. Todavia, nenhuma das sequências analisadas revelou a existência de pontos de recombinação (gráficos de *bootscanning* não apresentados).

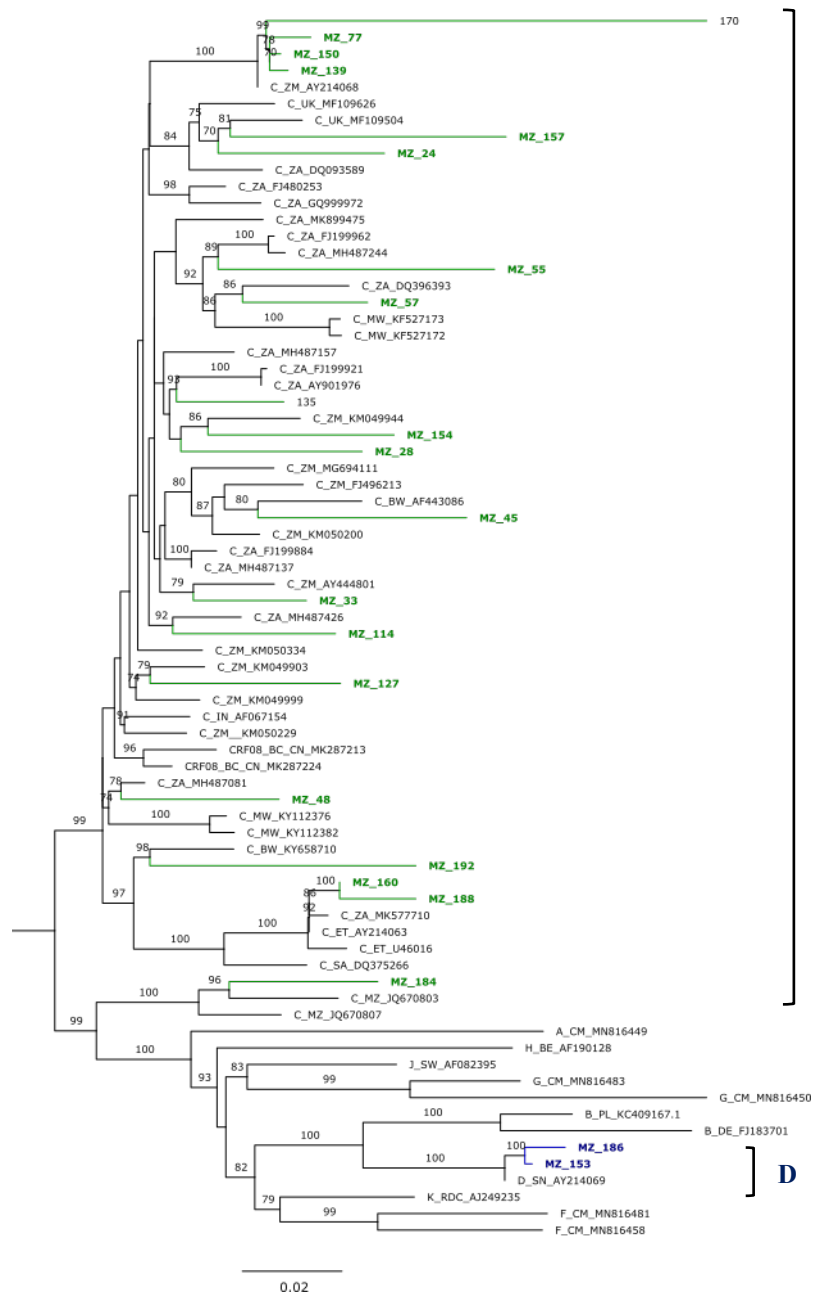


Figura 22 – Árvore Filogenética de sequências codificantes da IN de Moçambique. Relações filogenéticas entre 74 sequências nucleotídicas da região codificante da IN, com base num total de 892 nucleótidos. A árvore filogenética inclui 22 sequências em estudo (representadas com o prefixo MZ), 42 sequências obtidas no GenBank por pesquisa de homologia genética com o programa BLAST e 10 sequências de referência do grupo M do VIH-1 (A, B, F, G, H, J, K). A designação das sequências de referência é iniciada com a letra correspondente ao genótipo, seguida do local de origem e o número de acesso no GenBank. As origens geográficas das sequências de referências são: BW – Botswana, CM – Camarões, CN – China, DE – Alemanha, ET – Etiópia, MW – Malawi, MZ – Moçambique, PL – Polónia, SA – Arábia Saudita, SN – Senegal, UK – Reino Unido, ZA – África do Sul, ZM – Zâmbia.

3.2 Análise das sequências MZ com genomas mosaico

Após a análise das quatro regiões do genoma viral em estudo, conclui-se que as sequências MZ_48, MZ_111, MZ_139, MZ_153, MZ_77, MZ_170, MZ_178, MZ_186 e MZ_187 possuem genomas recombinantes (Tabela 16). A sequência MZ_48 apresenta fenómenos de recombinação entre os subtipos B e C, com a estrutura mosaico C/C/C/B (PR, RT, IN e C2V3C3, respetivamente) (Figura A1, Anexo 3). Em MZ_48 é evidente a homologia genética com o subtipo C ao longo do gene *pol* e posteriormente com o subtipo B em C2V3C3. Relativamente a MZ_111, apresenta fenómenos de recombinação entre os subtipos B e C, com uma estrutura em mosaico semelhante às sequências anteriormente apresentadas, no entanto, não foi possível amplificar a região codificante da enzima integrase, pelo que, não pode ser confirmada a semelhança total entre estas três sequências (Figura A1, Anexo 3).

A sequência MZ_77 apresenta uma estrutura mosaico CRF02_AG/D/C/C (Figura A2, Anexo 3). Na parte inicial do gráfico de bootscanning, correspondente à região da enzima protease é observável um ponto de recombinação entre os subtipos A e G, no entanto abaixo do nível de significância de 70%. Todavia dada a justificação apresentada na secção 3.1.3, o isolado viral é considerado recombinante nesta região do genoma. A secção do gráfico correspondente à região C2V3C3 (1800-2100 nucleótidos) apresenta-se por completo abaixo do valor de significância estabelecido de 70%. Uma possível justificação para este facto, tendo em conta a árvore filogenética da secção 3.1.1, seria baseada no tamanho do ramo que representa a sequência bem como a segregação precoce do grupo que contém as demais sequências de referência do subtipo C, que levaria a considerar MZ_77 como divergente desse mesmo subtipo. No entanto, é observável um possível ponto de recombinação com o subtipo G, que se revela incompatível com os dados obtidos da análise da árvore filogenética, uma vez que esta sequência seria invariavelmente classificada como subtipo C devido à formação de um agrupamento com sequências de referência desse subtipo com um valor de *bootstrap* de 89%. Além disso, a topologia da árvore mostra uma distância filogenética significativa das sequências de referência do subtipo G. Dada a inexistência de sequências de referência relativas a uma possível forma recombinante entre os subtipos C e G não é possível estabelecer uma comparação direta entre sequências.

Tal como MZ_77, as sequências MZ_153 e MZ_170 apresentam na região C2V3C3 um possível ponto de recombinação entre o subtipo C e G. No entanto, a dimensão do fragmento C2V3C3 (392 nucleótidos) não permite atingir um nível de significância acima dos 70% por forma a confirmar esta estrutura mosaico. Para uma genotipagem inequívoca destas sequências C2V3C3 seria necessário sequenciar o gene *env* por completo. Assim, as sequências MZ_153 e MZ_170 apresentam uma estrutura mosaico C/C/D/C e A/D/C/C, respetivamente (Figura A3, Anexo 3). As sequências MZ_139 e MZ_186 apresentam estruturas em mosaico A/B/C/C e B/C/D/C (PR, RT, IN, C2V3C3), respetivamente (Figura A4, Anexo 3). MZ_187 não apresenta dados sobre a região codificante da enzima integrase que não foi possível amplificar, no entanto apresenta a estrutura mosaico C/D/C (PR, RT, C2V3C3) (Figura A4, Anexo 3). Os dados são concordantes com os obtidos aquando da análise filogenética. Em relação a MZ_178, esta revela uma estrutura mosaico A/C/C (PR, RT e C2V3C3). A região correspondente à região C2V3C3 (1000-1300 nucleótidos) no gráfico de bootscanning (Figura A5, Anexo 3) encontra-se totalmente abaixo do valor de significância de 70%. Apesar deste facto, em comparação com a análise filogenética e inequívoca classificação do fragmento como subtipo C, considera-se esta genotipagem correta.

3.3 Distâncias genéticas

A distância genética média foi calculada entre todos os isolados virais de cada região do genoma em estudo e encontra-se resumida na tabela 17. Os valores médios de distância genética obtidos foram de 23,4% para C2V3C3, 11,9% para PR, 8,87% para RT e 7,81% para IN.

Tabela 17 – Distâncias genéticas entre os isolados virais de cada região do genoma em estudo.

Região	Distância Genética (média %)	S.E	Máximo (%)	Mínimo (%)	Número de isolados
C2V3C3	23,4	0,0140	38,7	0,660	58
RT	8,87	0,0051	16,4	0,440	26
PR	11,9	0,0085	22,3	0,230	34
IN	7,81	0,0052	17,1	0,570	22

As distâncias genéticas entre subtipos para as regiões do genoma viral em análise também foram calculadas (Tabela 18). Verificou-se que o gene *env* é o mais variável, com valores calculados de 25,9% a 30,6% entre os isolados virais dos subtipos A, B e C. As regiões

codificantes das enzimas constituintes do gene *pol* variam entre 8,6% e 16,8%. A distância genética mais elevada para a RT é encontrada entre os subtipos C e D e para a PR entre o subtipo C e CRF02_AG.

Tabela 18 – Matrizes de distância genética entre os subtipos das diferentes regiões do genoma viral.

Matrizes de Distâncias Genéticas entre subtipos											
C2V3C3			RT			PR				IN	
C	B	A	C	B	D	C	B	A	CRF02	C	D
C			C			C				C	
B	30,2%		B	11,9%		B	14,7%			D	11,7%
A	30,6%	25,9%	D	13,1%	8,60%	A	13,7%	12,0%			
						CRF02	16,8%	15,4%	11,9%		

As distâncias genéticas médias entre os isolados virais do mesmo subtipo também foram analisadas para as quatro regiões do genoma em estudo e encontram-se representadas na tabela 19. Os dados revelam uma maior variabilidade genética entre as sequências do subtipo C em comparação com os restantes subtipos independentemente da região em análise.

Tabela 19 – Distâncias genéticas médias (%) dos isolados virais em cada subtipo nas diferentes regiões do genoma viral.

Distância genética [média% (SE)]				
Subtipo	C2V3C3	RT	PR	IN
A	-----	-----	1,14(0,0039)	-----
B	2,90(0,0076)	-----	-----	-----
C	22,5(0,0140)	7,53(0,0050)	11,1(0,0113)	7,26(0,0052)
D	-----	0,50(0,0020)	-----	0,8(0,0027)

----- Não foram identificados os subtipos na região do genoma em análise.

Os vinte isolados virais destacados na secção 3.1.1 (MZ_44, MZ_24, MZ_154, MZ_74, MZ_127, MZ_178, MZ_170, MZ_105, MZ_123, MZ_146, MZ_142, MZ_186, MZ_151, MZ_84, MZ_25, MZ_188, MZ_157, MZ_28, MZ_62 e MZ_139) que revelaram um posicionamento ambíguo das sequências MZ na árvore filogenética, remetendo para uma maior relação entre si do ponto de vista filogenético do que com as sequências de referência do subtipo C, podiam sugerir a existência de um sub-subtipo do

subtipo C, uma vez que não foram detetados eventos de recombinação. Dada esta hipótese, calculou-se a distância genética entre as sequências MZ destacadas e as restantes sequências correspondentes a isolados do subtipo C. O valor da distância genética obtido foi de $19,7 \pm 0,0266\%$, remetendo para a possibilidade de as sequências em estudo poderem representar efetivamente um sub-subtipo C [208]. Por forma a confirmar esta observação seria necessário obter as sequências genómicas integrais dos isolados, confirmar que não possuem um genoma mosaico e que estes indivíduos não estão epidemiologicamente relacionados [209].

3.3 Análise bioinformática das sequências aminoacídicas da transcriptase reversa, protease e integrase

As sequências em estudo foram analisadas quanto à presença de mutações associadas à diminuição da suscetibilidade aos antirretrovirais em uso clínico, bem como à existência de polimorfismos genéticos.

3.3.1 Identificação de mutações associadas a resistência aos inibidores da transcriptase reversa e outros polimorfismos genéticos

Nas vinte e seis sequências em estudo da região RT foram encontradas as seguintes mutações principais associadas à diminuição da suscetibilidade aos NRTI: K65R (4%), D67N (4%), K70R (4%), M184V (8%), T215V (4%) e K219E (4%). Quanto aos NNRTI foram encontradas as seguintes mutações: A98G (4%), K101E (4%), K103N (12%), K103S (4%), V106A (4%), V106M (4%), V108I (4%), E138A (8%), Y181C (8%), G190A (8%) e F227L (4%). Estas mutações foram encontradas em oito isolados virais (14%) do subtipo C, cujos perfis de resistência se encontram representados na tabela 20. Considerando os oito isolados virais onde foram detetadas as mutações referidas e a análise dos dados representados na tabela 20, constatou-se que 15% (N=4) revelam mutações de resistência aos NRTI e 31% (N=8) possuem pelo menos uma mutação de resistência aos NNRTI. Desta forma, estes indivíduos em estudo, encontram-se infetados com vírus potencialmente resistentes ao tratamento com a maioria dos fármacos de ambas as classes de antirretrovirais abordadas. Por último, foram analisados outros polimorfismos genéticos presentes nas 26 sequências da transcriptase reversa. Estas, segundo o algoritmo de interpretação de resistências genotípicas de Stanford HIVdb, não

estão associadas a resistência aos inibidores da RT e encontram-se representados na figura 23.

Tabela 20 – Perfis de resistência dos isolados virais de Moçambique em estudo.

Isolado	Mutações		Perfil de resistência			
	NRTI	NNRTI	P. Baixo	Baixo	Intermédio	Alto
MZ_24		E138A	ETR	RPV		
MZ_45		K103N				EFV, NVP
MZ_55		K103N, G190A	ETR	RPV		EFV, NVP
MZ_57	D67N	V106A, F227L	ABC, DDI, TDF	AZT, D4T,		DOR, EFV, NVP
MZ_127	K65R	Y181C	DOR		ABC, FTC, 3TC, EFV, ETR, RPV	D4T, DDI, TDF, NVP
MZ_135		K103N, V106M, E138A	ETR	RPV	DOR	EFV, NVP
MZ_153	M184V	K103N	DDI	ABC		FTC, 3TC, EFV, NVP
MZ_188	K70R, M184V, T215V, K219E	V108I, Y181C, G190A, A98G	TDF		ABC, AZT, D4T, DDI	FTC, 3TC, DOR, EFV, ETR, NVP, RPV

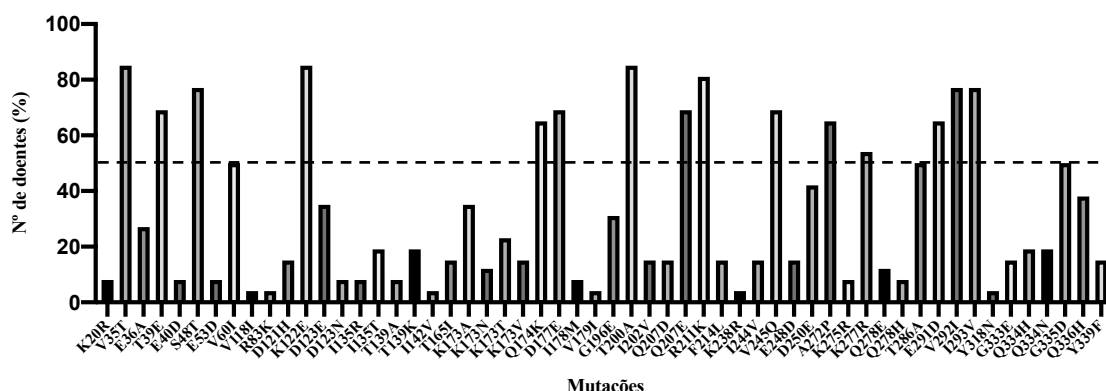


Figura 23 – Representação gráfica da prevalência dos 56 polimorfismos genéticos identificados na RT. Os polimorfismos representados não estão associados a resistência aos inibidores da RT (NRTI e NNRTI). A linha a tracejado assinala o cut-off de 50%.

Os 56 polimorfismos foram encontrados em isolados virais do subtipo B, C e D, revelando maior incidência sob isolados do subtipo C. As mutações encontradas com maior frequência (>50%) são: V35T (85%), T39E (69%), S48T (77%), V60I (50%), K122E (85%), Q147K (65%), D177E (69%), T200A (85%), Q207E (69%), R211K (81%), V245Q (69%), A272P (65%), K277R (54%), T286A (50%), E291D (65%), V292I (77%), I293V (77%) e G335D (50%). As frequências dos polimorfismos genéticos mais encontrados nos isolados em estudo, são coincidentes com os descritos na base de dados

de Stanford (Tabela A2, Anexo 4) [210]. Em relação aos polimorfismos menos frequentes (<50%), e estabelecendo a comparação com a base de dados, as mutações E36A e T286A são encontradas com maior frequência, em isolados do subtipo C, do que as registradas neste estudo enquanto que para as mutações T139K, G196E e G333E ocorre o inverso [214]. A incidência dos restantes polimorfismos é coincidente com os dados já descritos (Tabela A2, Anexo 4) [210]. Os isolados do subtipo D revelam diferenças mais significativas em relação aos dados de Stanford. Os polimorfismos V35T, V60I, K122E, D177E, Q207E, D250E, T286A e I293V possuem frequências semelhantes às descritas em estudos anteriores (Tabela A2, Anexo 4). Todavia, os polimorfismos D121H, I135T, K173N, Q174K e T200A possuem valores de incidência mais elevados nos isolados virais em estudo quando comparados com os valores de referência, ao contrário do que se verifica com as mutações G335E, A272P e R211K, que apesar de descritos frequentemente para isolados do subtipo D, não foram registados nos isolados virais em estudo (Tabela A2, Anexo 4) [210]. Ainda foram encontradas as mutações G336H e Y339F com uma frequência de 41% e 14% respetivamente, em isolados do subtipo C e a mutação Y339F, com uma frequência de 33%, para o subtipo D, no entanto estas não se encontram registradas nas bases de dados para estes subtipos [210]. Alguns polimorfismos foram mais frequentemente encontrados nos subtipos C e D do que nos outros subtipos (Tabela 21).

Tabela 21– Polimorfismos genéticos associados estatisticamente ($p < 0,05$) aos subtipos C e D.

Subtipos	Vírus Selvagem	Posição aminoacídica	Vírus mutado	Valor p
C	T	39	E	0,0047
	S	48	T	0,0010
	R	211	K	0,0144
	K	245	Q	0,0047
	A	272	P	0,0084
	K	277	R	0,0331
	E	291	D	0,0084
D	D	121	H	0,0004
	I	135	T	0,0038
	K	173	N	0,0004
	I	224	V	0,0015
	D	250	E	0,0462

3.3.1.2 Análise da variabilidade da transcriptase reversa

A variabilidade das sequências aminoacídicas da RT foi analisada ao longo das secções anteriores, no entanto, com o intuito de fornecer uma representação simplificada do grau de conservação relativo de cada posição aminoacídica no alinhamento múltiplo, foi construído um WebLogo com recurso à ferramenta bioinformática AnalyzeAlign. Na representação gráfica da figura 24, cada aminoácido representado pelo código IUPAC, possui um tamanho específico em cada posição, proporcional à frequência na qual é encontrado no conjunto de sequências que formam o alinhamento, mostrando de modo imediato e simplificado quais as regiões mais conservadas.

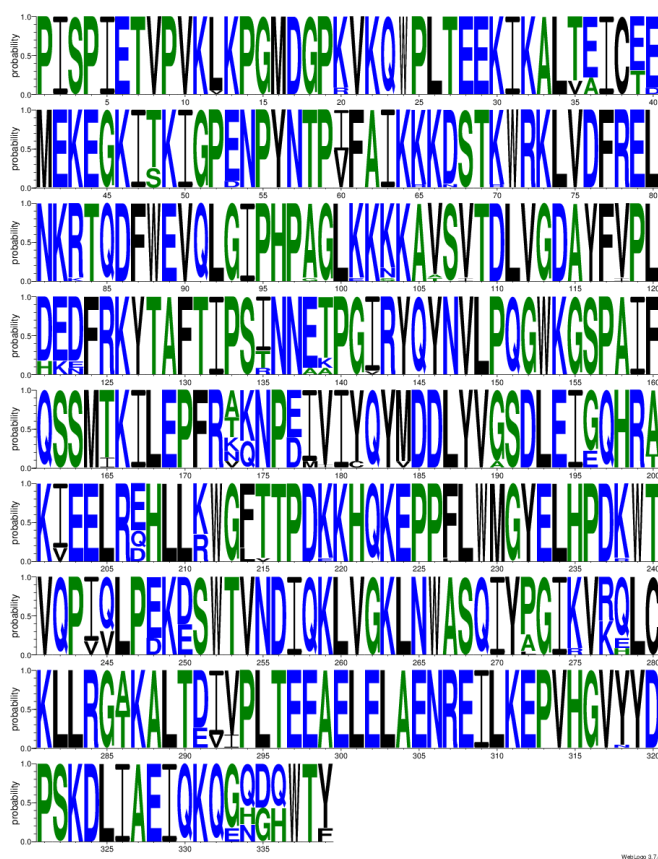


Figura 24— WebLogo do alinhamento múltiplo das sequências aminoacídicas MZ da enzima RT. Representação gráfica da sequência aminoacídica weblogo constituída a partir do alinhamento múltiplo das sequências MZ para a RT. Os aminoácidos apresentados em cada posição estão representados pelo código internacional IUPAC. As cores representam o grau de hidrofobicidade: azul – hidrofílico, verde – neutro, preto – hidrofóbico.

Os subdomínios *fingers* (resíduos 1-85 e 118-155), *palm* (resíduos 86-117 e 156-236), *thumb* (resíduos 237-318) e *connection* (resíduos 319-426) fazem parte do domínio

polimerase da subunidade p66 da transcriptase reversa (descrição pormenorizada na secção 1.4.1.2) [53-55]. No subdomínio *palm* encontra-se o centro ativo da enzima, composto por uma tríade catalítica, geralmente, conservada (D110, D185 e D186). É possível confirmar pela análise do WebLogo da figura 24, a conservação da tríade catalítica, evidenciada pela presença da letra D, representativa do ácido aspártico, cuja probabilidade associada é igual a 1. Os resíduos D185 e D186 fazem parte do motivo YXDD, altamente conservado nos retrovírus (no caso do VIH-1, X corresponde a metionina) [53-55], no entanto verifica-se a existência da mutação primária M184V. Verificou-se que a maioria dos polimorfismos genéticos encontrados ocorrem no subdomínio *fingers*, apesar de se encontrarem distribuídos de forma significativa pelos restantes subdomínios. Por fim, ao analisar o código de cores que representa o grau de hidrofilicidade/hidrofobicidade de cada aminoácido, permite concluir que a maioria dos polimorfismos encontrados são de natureza conservativa.

3.3.2 Identificação de mutações associadas a resistência aos inibidores da integrase e outros polimorfismos genéticos

Nas vinte e duas sequências aminoacídicas da integrase analisadas foi encontrada apenas uma mutação acessória G163R num isolado do subtipo C. Esta mutação por si só tem pouco efeito sobre a suscetibilidade aos INSTI, no entanto, promove uma resistência de nível baixo ao EVG e RAL, devido à baixa barreira genética destes antirretrovirais. Foram encontrados 63 polimorfismos genéticos, não associados a resistência aos INSTI, em isolados virais do subtipo C e D (figura 27).

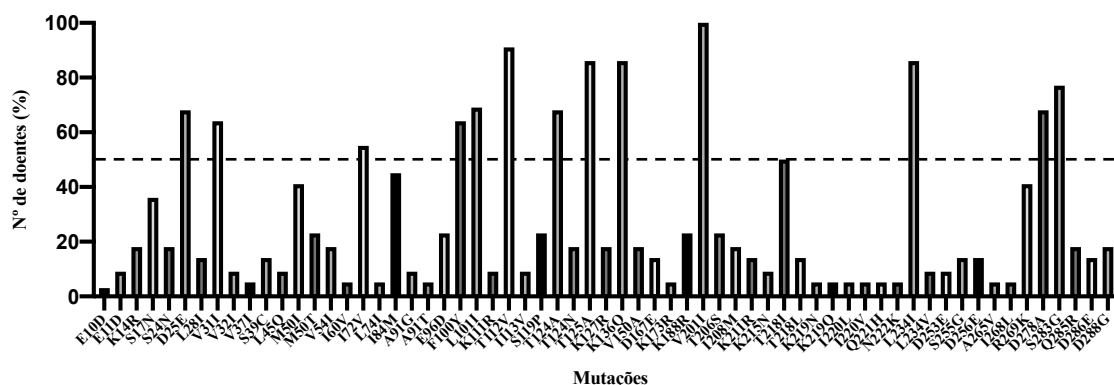


Figura 27 – Representação gráfica da prevalência dos 63 polimorfismos genéticos identificados na IN. Os polimorfismos representados não estão associados a resistência aos INSTI. A linha a tracejado assinala o cut-off de 50%.

Os polimorfismos mais frequentes (>50%) são: D25E (68%), V31I (64%), I72V (55%), F100Y (64%), L101I (69%), T112V (91%), T124A (68%), T125A (86%), K136Q (86%), V201I (100%), T218I (50%), L234I (86%), D278A (68%) e S283G (77%). As frequências dos polimorfismos genéticos são coincidentes com os descritos na base de dados de Stanford (Tabela A7, Anexo 8) [210]. A análise dos polimorfismos menos frequentes nos isolados virais do subtipo C permitiu identificar os polimorfismos Q285R (20%), D286E (15%) e D288G (15%) que não se encontram registados nas bases de dados para os subtipos em estudo. A mutação D288G também foi encontrada num isolado viral do subtipo D (Tabela A7, Anexo 8) [210]. Alguns polimorfismos foram mais frequentemente encontrados nos subtipos C e D do que nos outros subtipos (Tabela 22).

Tabela 22 – Polimorfismos genéticos associados estatisticamente ($p < 0,05$) aos subtipos C e D.

Subtipos	Vírus Selvagem	Posição aminoacídica	Vírus mutado	Valor p
C	L	101	I	0,0260
	T	125	A	0,0130
	K	136	Q	0,0130
	S	283	G	0,0433
D	V	37	I	0,0043
	I	113	V	0,0043
	K	211	R	0,0130
	D	256	E	0,0130

3.3.3.1 Análise da variabilidade das sequências MZ para a integrase

A variabilidade das sequências aminoacídicas da IN foi analisada como descrito para as restantes enzimas nas secções anteriores, a partir da construção de um WebLogo representado na figura 28. A enzima possui três domínios funcionais: N-terminal (NTD, resíduos 1-49), o domínio do centro catalítico (CCD, 50-212 resíduos) e o domínio C-terminal (CTD, 213-288 resíduos) [47, 56-59]. O domínio NTD possui quatro aminoácidos essenciais e conservados, duas histidinas (H12 e H16) e duas cisteínas (C40 e C43), que coordenam um átomo de zinco, e a análise do weblogo permite confirmar a conservação das posições mencionadas nos isolados em estudo [47, 56-59]. Por sua vez, o domínio CCD contém uma tríade catalítica ácida altamente conservada (*DDE motif*) formada por dois aminoácidos de ácidos aspárticos (D64, D116) e um de ácido glutâmico (E152), o que também se verifica pela análise do WebLogo [47,56-59]. Ainda neste

domínio, o motivo K¹⁸⁶R¹⁸⁷K¹⁸⁸, envolvido na importação do ADN proviral para o núcleo da célula infetada, apresenta o polimorfismo K188R em um quarto dos isolados virais em estudo. Por fim, o domínio CTD, apesar de fundamental para a conservação da estrutura quaternária da enzima e pela ligação do ADN viral ao ADN da célula hospedeira, corresponde à região mais variável da integrase no estudo, associada a um maior número de polimorfismos. Na maioria dos estudos anteriores, este domínio também está caracterizado por uma maior variabilidade [47, 211-213]. Pela distribuição de cores que representam a hidrofobicidade ou hidrofiliicidade dos aminoácidos é possível concluir que a maior parte das substituições são de natureza conservativa.

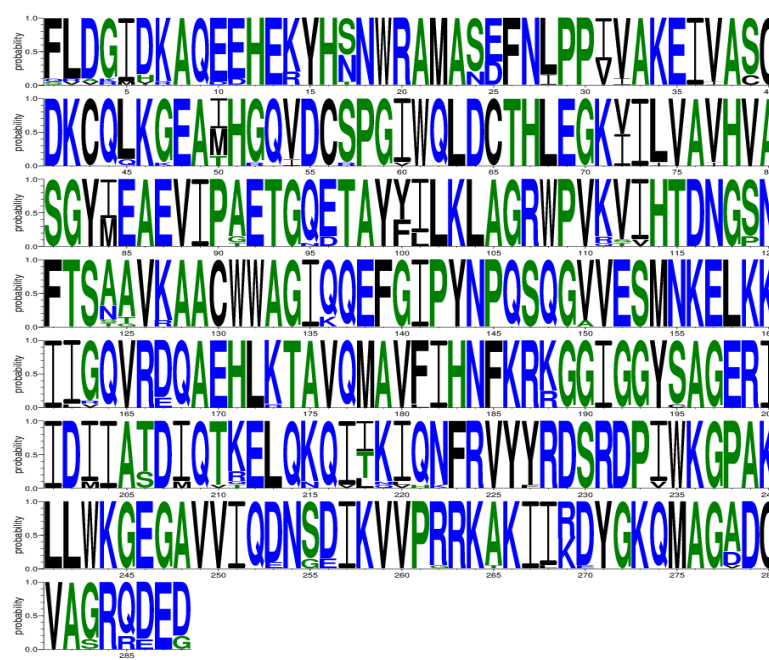


Figura 28–WebLogo do alinhamento múltiplo das sequências aminoacídicas MZ da enzima IN. Representação gráfica da sequência aminoacídica constituída a partir do alinhamento múltiplo das sequências MZ da enzima IN. Os aminoácidos apresentados em cada posição estão representados pelo código internacional IUPAC. As cores representam o grau de hidrofobicidade: azul – hidrofílico, verde – neutro, preta – hidrofóbico.

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

A análise das 34 sequências aminoacídicas da região PR permitiu identificar quatro mutações secundárias associadas à diminuição da suscetibilidade aos inibidores de protease, sendo estas: L10F (3%), I47R (6%), Q58E (3%) e G73S (3%). Estas mutações

foram encontradas em três isolados virais do subtipo A e C, cujos perfis de resistência se encontram representados na tabela 23.

Tabela 23 – Mutações associadas aos PI e os respectivos perfis de resistência dos isolados virais de Moçambique em estudo.

Isolado	Mutações	Perfil de resistência			
		P. Baixo	Baixo	Intermédio	Alto
MZ_55	I47R, Q58E, G73S	ATV, FPV, LPV	IDV, NFV, SQV, TPV		
MZ_113	L10F	DRV, IDV, LPV	FPV, NFV		
MZ_178	I47R				

O perfil de resistência mostra que estes indivíduos encontram-se infetados com vírus potencialmente resistentes ao tratamento com fármacos desta classe. É importante referir que a mutação I47R, encontrada em dois isolados virais, é altamente incomum nesta posição pelo que é desconhecido o seu impacto na suscetibilidade aos inibidores de protease. Outros polimorfismos genéticos, não associadas a resistência aos PI, encontrados nas 34 sequências aminoacídicas estão representados no gráfico da figura 25.

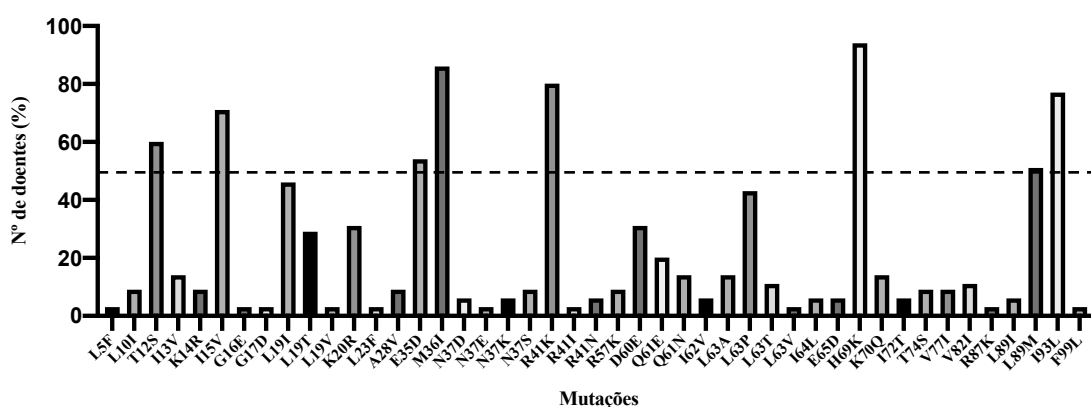


Figura 25 – Representação gráfica da prevalência dos 45 polimorfismos genéticos identificados na PR. Os polimorfismos representados não estão associados a resistência aos inibidores da PR. A linha a tracejado assinala o cut-off de 50%.

Os 45 polimorfismos foram encontrados em isolados virais do subtipo A, B, C e na CRF02_AG e segundo o gráfico, as mutações encontradas com maior frequência (>50%) são: T12S (60%), I15V (71%), E35D (54%), M36I (86%), R41K (80%), H69K (94%), L89M (51%) e I93L (77%). As frequências dos polimorfismos encontradas nos isolados

em estudo, são coincidentes com os descritos na base de dados de Stanford (Tabela A5, Anexo 6) [210]. No entanto, é importante destacar a baixa incidência da mutação L10I nos isolados virais em estudo, uma vez que esta mutação se revela transversal e altamente frequente (>99%) nos subtipos A, B e C e na CRF02_AG em análise. Esta mutação foi encontrada apenas nas sequências aminoacídicas correspondentes à região da PR do subtipo A [210]. O polimorfismo A28V não está descrito para os subtipos B, C e para a CRF02_AG, sendo relatado com uma frequência de apenas 0,1% para o subtipo A [210]. Contudo, foi encontrado em todas as sequências do subtipo A em estudo para esta região do genoma (Tabela A5, Anexo 6). Alguns polimorfismos foram mais frequentemente encontrados nos subtipos A e C do que nos outros subtipos (Tabela 24).

Tabela 24 – Polimorfismos genéticos associados estatisticamente ($p < 0,05$) aos subtipos A e C.

Subtipos	Vírus Selvagem	Posição aminoacídica	Vírus mutado	Valor p
A	L	10	I	0,0002
	I	13	V	0,0017
	A	28	V	0,0002
C	T	12	S	0,0046
	I	15	V	0,0005
	L	19	I	0,0465
	R	41	K	0,0362

4.3.2.1 Análise da variabilidade das sequências MZ para a protease

A análise da variabilidade das sequências aminoacídicas da enzima protease foi realizada a partir do WebLogo representado na figura 26. A protease possui uma estrutura dimérica estabilizada pela interação da região N-terminal de um monómero e a região C-terminal do segundo, nas posições aminoacídicas 1-4 e 96-99, respetivamente [50,51]. A análise do weblogo permite atestar a conservação de ambas as regiões nos isolados em estudo. O centro ativo da enzima localiza-se na interface dos dois monómeros (resíduos 23-32, 45-53 e 80-84) e é composto pela tríade catalítica D25 T26 G27, altamente conservada, como evidencia a figura 26 [50,51]. No entanto, no centro ativo, nomeadamente nas posições 23, 28, 47 e 82 foram detetados os polimorfismos L23F, A28V, I47R e V82I. Na sua forma ativa, a protease possui uma estrutura flexível em ambos os monómeros, designada de *flap* (resíduos 43-58) que permite a ligação aos

substratos e é geralmente conservada [50,51]. Pela análise da representação gráfica WebLogo é detetável a presença no polimorfismo R57K nessa região. Pela distribuição de cores que representam a hidrofobicidade ou hidrofiliicidade dos aminoácidos é possível concluir que à exceção de algumas posições polimórficas da sequência (L19, A28, N37, I47, L63 e I72), as substituições são, na generalidade, de natureza conservativa.

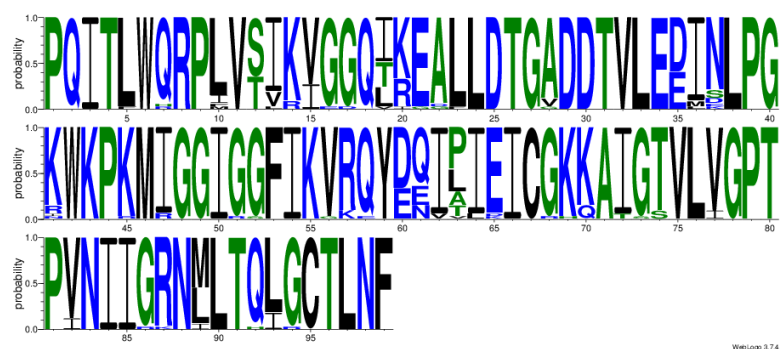


Figura 26– WebLogo do alinhamento múltiplo das sequências aminoacídicas MZ da enzima PR. Representação gráfica da sequência aminoacídica constituída a partir do alinhamento múltiplo das sequências MZ da enzima PR. Os aminoácidos apresentados em cada posição estão representados pelo código internacional IUPAC. As cores representam o grau de hidrofobicidade: azul – hidrofílico, verde – neutro, preta – hidrofóbico.

3.4 Análise bioinformática das sequências aminoacídicas da região C2V3C3

O alinhamento múltiplo das 58 sequências aminoacídicas da região C2V3C3 (Figura A6, Anexo 10) foi convertido num WebLogo (figura 29), por forma a analisar de forma imediata e sintetizada a variabilidade das mesmas, uma vez que esta se trata da região mais heterogénea do genoma do VIH-1. No Weblogo a região C2 encontra-se entre as posições 1-30, V3 entre 31-67 e por fim C3 de 68-120. Uma vez que o padrão de glicosilação do gene *env* do VIH-1 pode influenciar o escape à neutralização e o tropismo viral analisou-se os potenciais locais de N-glicosilação nas sequências em estudo [214,215]. Os locais de N-glicosilação que estão associados, geralmente, a um motivo NXS/T (X é qualquer aminoácido diferente de prolina) [194]. Na região C2 existem 3 locais de N-glicosilação putativos, segundo a literatura, nas posições destacadas a laranja na figura 29 (posições 11, 24 e 30) que não se apresentam conservadas. A análise das sequências através da ferramenta on-line *N-GlycoSite*, evidenciou a variabilidade destas regiões que se traduzem numa perda de locais de glicosilação de 14% nas posições 11-13

(277-279, no HIV-1 HXB2), 41% nas posições 24-26 (290-292, segundo numeração HXB2) e 79% nas posições 30-33 (295-297, no HXB2).

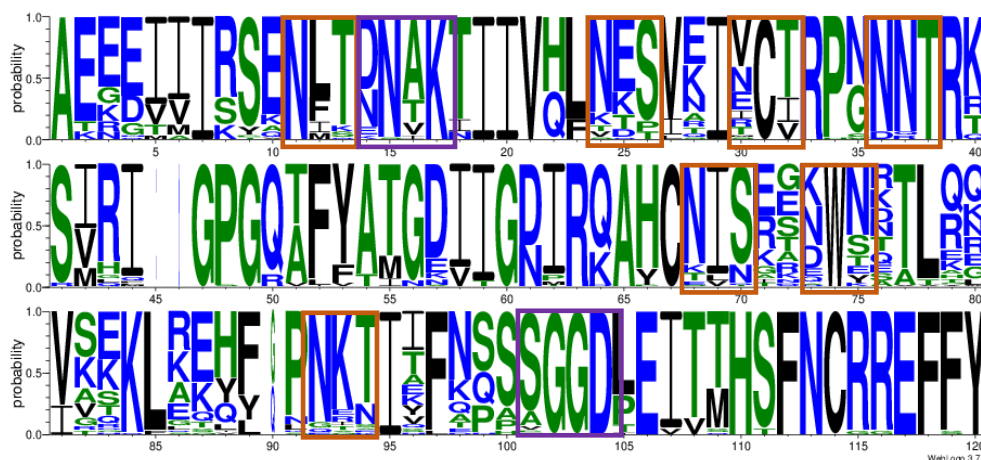


Figura 29– WebLogo do alinhamento múltiplo das sequências aminoacídicas MZ da *loop* V3. Representação gráfica da sequência aminoacídica constituída a partir do alinhamento múltiplo das sequências MZ da região C2V3C3. Os aminoácidos apresentados em cada posição estão representados pelo código internacional IUPAC. As cores representam o grau de hidrofobicidade: azul – hidrofílico, verde – neutro, preta – hidrofóbico. As posições destacadas por um retângulo laranja representam locais putativos de N-glicosilação e a roxo locais de ligação ao recetor CD4.

A região V3 apresenta nas posições 36-38 (301-303, no HXB2) outra região de N-glicosilação que se encontra conservada em 95% dos isolados virais em estudo. Este local é de extrema importância pois oculta grande parte dos epítomos CD4bs por forma a promover o escape à neutralização [216]. Por fim a região C3 possui três regiões descritas na literatura como locais de N-glicosilação nas posições 68-70 (332-334, segundo numeração HXB2), 73-75 (337-339, segundo numeração HXB2) e 92-94 (356-358, segundo numeração HXB2), nas quais se registou, respetivamente, em 19%, 74% e 24% dos isolados virais a perda de glicanos. No entanto, verificou-se que existe em algumas sequências, a compensação dos locais de N-glicosilação perdidos (Figura A6, Anexo 10). Os isolados que evidenciam a perda dos locais de N-glicosilação na posição 68 (332, segundo a numeração HXB2), adquirem mutações adicionais que permitem substituir esse local por outro na posição 70 (334, segundo a numeração HXB2). O mesmo se verifica para a posição 73 e 92 (356, segundo numeração HXB2), em que a perda de locais de N-glicosilação dessas regiões é compensada a jusante, em alguns isolados virais, nas posições 75 (339, segundo numeração HXB2) e 80 (344, segundo numeração HXB2)

e para a segunda região de perda, as mutações que permitem adquirir novos locais de glicosilação ocorre em 94 (358, segundo numeração HXB2) e 98 (362, segundo numeração HXB2). Em resumo, o número de locais de N-glicosilação variou significativamente entre V3 e C3 (teste *t* de Student, $p=0,0444$), sendo que os locais mais conservados estavam localizados em V3. Em cada um dos 58 isolados virais, o número de locais de N-glicosilação varia entre 2 e 8. A comparação da distribuição destes locais entre os genótipos encontrados (A, B e C), revela que esta é genericamente aleatória, no entanto, o número médio de locais de N-glicosilação, revelou ser maior em isolados virais do subtipo B (teste *t* de Student, $p = 0,0083$).

3.4.1 Polimorfismos genéticos da *loop* V3

A região V3 da gp120 é delimitada por duas cisteínas e é constituída por 34 a 35 aminoácidos e as suas sequências aminoacídicas caracterizam-se por serem altamente variáveis (Figura 29) [64,65]. Esta região para além de ser identificada como o principal domínio neutralizante do invólucro viral, possui um papel essencial na determinação do tropismo viral [66]. Por forma a analisar as sequências aminoacídicas da região em estudo construiu-se o alinhamento múltiplo representado na figura 30. A *loop* V3, não obstante, a sua proeminente variabilidade, possui uma região conservada designada de motivo coroa– GPGR/Q – considerado o epítipo mais antigénico de toda a V3. As sequências MZ_19, MZ_48, MZ_67 e MZ_111, do subtipo B, apresentam o motivo GPGR enquanto que as restantes sequências do subtipo C, incluindo uma sequência do subtipo A1, apresentam o motivo GPGQ (Figura 30). Estes dados são concordantes com os dados da literatura [217,218].

O gráfico da figura 31 representa a frequência (%) das 44 mutações encontradas nas sequências aminoacídicas da região V3. As mutações encontradas com maior frequência (>50%) no conjunto dos 58 isolados virais são: R306S (97%), R315Q (91%), V318Y (81%), T319A (90%), I320T (76%), K322T (74%), N325D (67%) e M326I (93%). Foram encontradas quatro mutações com associação estatística aos subtipos mencionados – R315Q ($p<0,0001$) e A316T ($p=0,0238$) para o subtipo C e R308H ($p<0,0001$) e K305R ($p=0,0002$) para o subtipo B.

		N301					
		▲				▲	
HXB2_ :	CTRPNNNTRKRIRIQFGPGRAFVTIGK-IGNMRQAHC						
				Carga total V3	Posição 11/25	Co-recetor	
MZ_17 :	.I.....S...--...Q..YAT.DI..DI.....			4	SD	CCR5	
MZ_19 :	RS.H.--...V.YAT.DI..DI.....			5	SD	CCR5	
MZ_22 :	.I.....S...--...Q..YAT.DI..DI.....			4	SD	CCR5	
MZ_24 :	TSV...--...QT.YAT.DI..DI.....			3	SD	CCR5	
MZ_25 :	G.....SV...--...Q..FATNEV..DI.....			4	SE	CCR5	
MZ_28 :	...H.....SV...--...QTIFAT.DI..DI.K...			6	SD	CCR5	
MZ_33 :	.V...DR..TS.L.--...QT.YAM.DI...I.K...			4	SD	CCR5	
MZ_36 :	G.....S...--...QT.YAT.AI..DI.....			5	SA	CCR5	
MZ_38 :	.I.....SV...--...QT.YAT.EV..DI.K...			5	SE	CCR5	
MZ_44 :	TSV...--...QT.YAT.DI..DI.....			3	SD	CCR5	
MZ_45 :	.I.....S...--...Q..YAT.DI..DI.....			4	SD	CCR5	
MZ_48 :	RS.H.--...YAT.DI..DI.....			5	SD	CCR5	
MZ_55 :	S..F...--...Q..YAT.DI..DI.....			4	SD	CCR5	
MZ_57 :	.V.....S...--...Q..YAT.DI..DI.....			4	SD	CCR5	
MZ_62 :	S.....SV...--...QTLFAT.NI..DI.....			5	SN	CXCR4	
MZ_65 :	SV...--...Q..YAT.DI..DI.R...			4	SD	CCR5	
MZ_67 :	RS.H.--...YAT.DI..DI.....			5	SD	CCR5	
MZ_70 :	.V...DR..TS.L.--...QT.YAM.DI...I.K...			4	SD	CXCR4	
MZ_71 :	S...--...Q..YAT.DI..DI.....			4	SD	CCR5	
MZ_72 :	...G....TS...--...QT.FAT.EV..DI.....			3	SE	CCR5	
MZ_73 :	...G....RS...--...QT.F.TN.I...I.....			8	SK	CCXR4	
MZ_74 :	TS...--...Q..Y.T.DIT.DI...Y.			2	SD	CCR5	
MZ_77 :	.V.....SV...--...QT.YAT.DI..DI.....			4	SD	CCR5	
MZ_84 :	G.....SV...--...Q..FATNEV..DI.....			4	SE	CCR5	
MZ_93 :	.I.....S...--...Q..YAT.DI..DI.....			4	SD	CCR5	
MZ_94 :	QS.H.--...Q..LY.T-NI...I.K.Y.			5	SN	CXCR4	
MZ_95 :	.I..S.....S...--...QT.YAT.EV...I.....			5	SE	CCR5	
MZ_103:	.V.....--...Q..YAT.DI..DI...Y.			4	RD	CXCR4	
MZ_105:	.I..G....SM...--...QT.YAM.DI...I.K...			6	SD	CXCR4	
MZ_109:	...GDQ...SMQQ--...QT.YAM.DI...I.K...			4	SD	CCR5	
MZ_111:	RS.H.--...YAT.DI..DI.....			5	SD	CCR5	
MZ_113:	.I..DQ...SMQQ--...QT.YAM.DI...I.K...			4	SD	CCR5	
MZ_114:	G....RSV...--...Q..YAT.DI..DI.....			4	SD	CCR5	
MZ_123:	.I..G....SM...--...QT.YAM.DI...I.K...			6	SD	CCR5	
MZ_127:	G....QS...--...QT.YAT.DI..DI.....			3	SD	CCR5	
MZ_130:	G.....SV...--...QV.YAT.DIT.DI.....			4	SD	CCR5	
MZ_134:	.V.....S...--...Q..YAT.DI..DI...Y.			3	SD	CCR5	
MZ_135:	G.....S...--...Q..YAT.EV..DP.K...			5	SE	CCR5	
MZ_137:	.V.....S...--...Q..YAT.DI..DI...Y.			3	SD	CCR5	
MZ_139:	.I..G....RSV...--...QT.FAT.DI..DI.....			4	SD	CCR5	
MZ_142:	G....SM...--...QT.YAM.DI...I.K...			6	SD	CCR5	
MZ_146:	.I..G....SM...--...QT.YAM.DI...I.K...			6	SD	CCR5	
MZ_150:	S...--...Q..YATNEV...I.K...			7	SE	CXCR4	
MZ_151:	G.....SV...--...QT.YTM.DI...I.....			5	SD	CCR5	
MZ_153:	G.....S...--...Q..YAT.EV..DP.K...			5	SE	CCR5	
MZ_154:	SM...--...QT.YAT.DI..DI...Y.			3	SD	CCR5	
MZ_157:	G....TS...--...QT.FAT.DI...I.....			4	SD	CCR5	
MZ_158:	.V.....RS..L--...Q..YA..DI..DI.....			4	SD	CCR5	
MZ_160:	QS.H.--...Q..LY.T-NI...I.K.Y.			5	SN	CXCR4	
MZ_163:	SV...--...Q..YAT.DI..DI.....			4	SD	CCR5	
MZ_170:	.I..G....SM...--...QT.YAM.DI...I.K...			6	SD	CCR5	
MZ_178:	G....TSVG...--...QT.YAT.DV..DI...Y.			1	SD	CCR5	
MZ_184:	S....SMG...--...QT.YAT.DI..DI.....			3	SD	CCR5	
MZ_186:	SV...--...QT.YAM.DI...I.....			5	SD	CCR5	
MZ_187:	G.....SV...--...Q..YAT.DI...P.K...			6	SD	CCR5	
MZ_188:	G....TS...--...QT.FAT.DI...I.....			4	SD	CCR5	
MZ_192:	--...QT.Y.T.DI..DI.....			4	SD	CCR5	
MZ_197:	SV...--...Q..YA..DVT.DI.....			4	SD	CCR5	

Figura 30– Alinhamento múltiplo das 58 sequências aminoacídicas da região V3 dos isolados de VIH-1 de Moçambique em estudo. As sequências alinhadas foram comparadas com a sequência aminoacídica de HXB2, obtida na base de dados Los Alamos HIV database (www.hiv.lanl.gov). Os aminoácidos apresentados em cada posição estão representados pelo código internacional IUPAC. Os pontos indicam semelhança relativamente à sequência HXB2. As posições aminoacídicas em destaque na região V3 representam: os aminoácidos das posições 11 e 25 (relevantes na identificação do tropismo viral), o local de inserção comum a quase todas as sequências MZ (laranja) e motivo coroa conservado GPGR/Q (verde). Os triângulos denotam resíduos de cisteína conservados. O potencial local de N-glicosilação está representado em cinzento (N301).

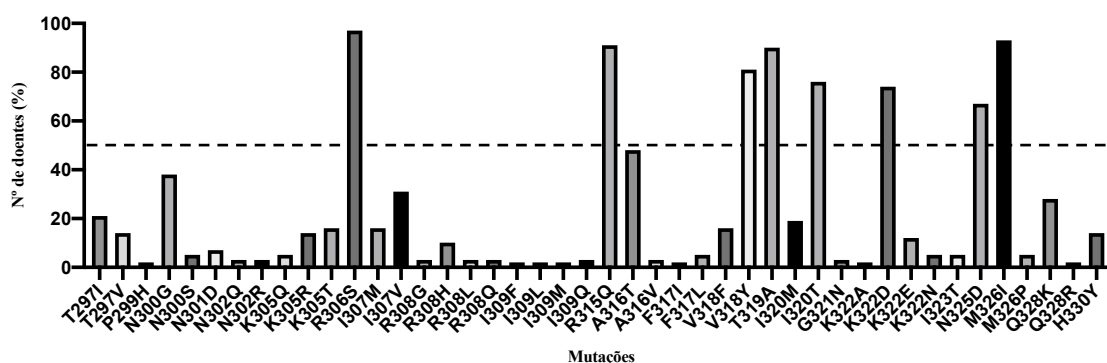


Figura 31 – Representação gráfica da prevalência dos 44 polimorfismos genéticos identificados na região V3. A linha a tracejado assinala o cut-off de 50%.

3.4.1.1 Tropismo Viral

O tropismo diferencial do VIH-1 pelos co-recetores foi determinado através de algoritmos informáticos que estabelecem a comparação entre as sequências de aminoácidos dos isolados virais em estudo com a sequência aminoacídica de HXB2, tendo em conta a substituição de aminoácidos específicos nas posições 11 e 25 que estão diretamente envolvidos na interação da gp120 com os co-recetores, a carga e dimensão da região V3 [67,73] (Figura 30). A análise das 58 sequências da região V3 permitiu identificar 8 isolados virais (14%) que utilizam CXCR4 como co-recetor aquando da entrada na célula hospedeira, sendo que os restantes utilizam o co-recetor CCR5. Os isolados virais identificados pelo uso de CXCR4 pertencem ao subtipo C. Independentemente do tropismo viral, todas as sequências são constituídas por 35 aminoácidos. As mutações T319A ($p=0,0293$), K322D ($p=0,0068$) e N325D ($p=0,0011$) estão associados aos isolados virais que utilizam o co-recetor CCR5 e as mutações K305Q ($p=0,0472$), F317L ($p=0,0018$), K322N ($p=0,0018$) e Q328K ($p=0,0401$) aos isolados

que utilizam CXCR4. Os vírus que usam CCR5 possuem uma carga total na região V3 significativamente inferior aos vírus que usam CXCR4 (mediana para vírus R5= 4 vs 6 para vírus X4, $P = 0,0127$) (Figura 32). Estes resultados são consistentes com o descrito na literatura [217-222].

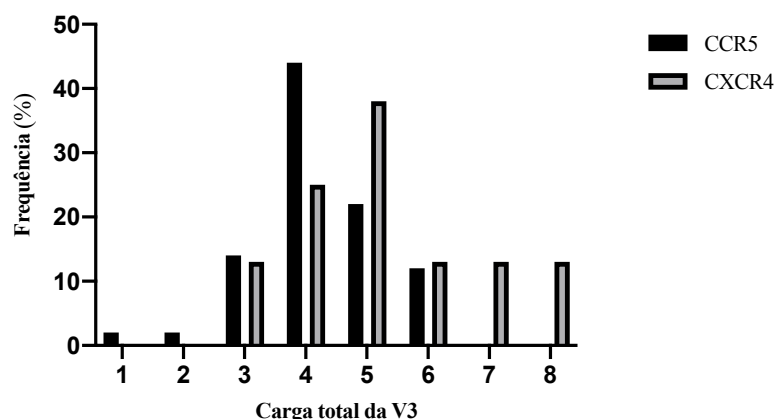


Figura 32 – Representação gráfica da carga total de aminoácidos da região V3 em função do tropismo viral. O gráfico permite fazer uma comparação entre a carga de aminoácidos da *loop* V3 dos isolados de VIH-1 que utilizam o co-recetor CCR5 ou CXCR4 para entrar nas células hospedeiras.

Por fim, foi comparada a entropia das sequências aminoacídicas da região C2V3C3 dos isolados virais que utilizam CCR5 e CXCR4 (Figura 33). O gráfico de entropia permite analisar a variabilidade aminoacídica de locais específicos do alinhamento múltiplo construído com as sequências em estudo. Um valor de entropia mais elevado está associado a uma maior variabilidade da região em análise [223,224].

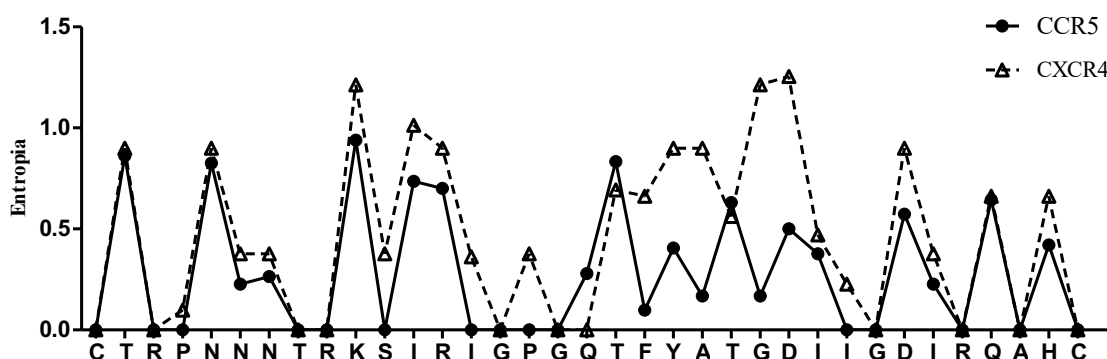


Figura 33 – Gráfico de entropia da região V3 de isolados virais de acordo com o tropismo. Variabilidade em termos de aminoácidos, em cada posição na *loop* V3, de acordo com o uso do co-recetor CCR5 ou CXCR4.

As variantes CXCR4 em estudo apresentam uma maior variabilidade ao longo da região V3 em comparação com os isolados virais que utilizam CCR5 (teste *t* de Student, $p = 0,0406$). As posições aminoacídicas (segundo a numeração HXB2) em que se observa uma maior variabilidade entre os isolados com diferentes tropismos são 305-307, 317-322 e 325 e 328.

3.5 Diversidade antigénica nos doentes moçambicanos

Esta secção tem como objetivo caraterizar a diversidade antigénica nos indivíduos moçambicanos com base na resposta em anticorpos contra polipéptidos da região C2V3C3 de diferentes subtipos de VIH-1 (B, C, G, H, CRF02_AG) [187]. A maioria dos 58 plasmas testados possuía anticorpos IgG que se ligaram a pelo menos um dos polipéptidos recombinantes em teste ELISA (Figura 34). Apenas onze amostras não apresentaram reatividade contra nenhum dos polipéptidos recombinantes. Das oito amostras que reagiram contra apenas um polipéptido, seis reagiram contra o polipéptido AG e duas contra o polipéptido C. As amostras que não reagiram contra nenhum polipéptido, ou que reagiram apenas contra um, são de indivíduos infetados com subtipo C. Relativamente às doze amostras que reagiram contra os cinco polipéptidos, 42% (N=5) são de doentes infetados com vírus recombinantes, 50% (N= 6) com subtipo C e 8% (N=1) com subtipo A. Os nove doentes com vírus recombinantes identificados no estudo possuem anticorpos de ligação contra três ou mais péptidos - 56% reagem contra todos os polipéptidos, 33% contra três e 11% contra quatro.

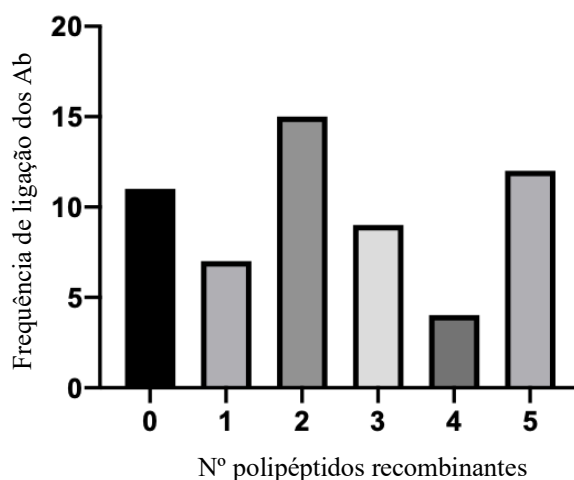


Figura 34 – Frequência com que anticorpos IgG anti-C2V3C3 presentes no plasma dos doentes em estudo reagem contra determinado número de polipéptidos recombinantes.

Cerca de 77% dos doentes possuía anticorpos de ligação contra o polipéptido recombinante AG, seguido de G (47%), C (45%), H (43%) e B (31%) (Figura 35). Existem diferenças estatisticamente significativas relativamente ao número de isolados virais que reagem contra os polipéptidos AG ($p<0,0001$) e B ($p=0,0027$). A baixa ligação ao polipéptido B é consistente com o reduzido número de doentes infetados por este subtipo viral e com as diferenças significativas existentes entre o subtipo B e os restantes subtipos ao nível da região em causa [225] (Figura A7, Anexo 13). Pelo contrário, a elevada antigenicidade do péptido AG foi algo surpreendente e poderá estar relacionada com a conservação das regiões CD4bs que são semelhantes à sequência consenso dos isolados virais MZ bem como dos motivos de N-glicosilação (Figura A7, Anexo 13).

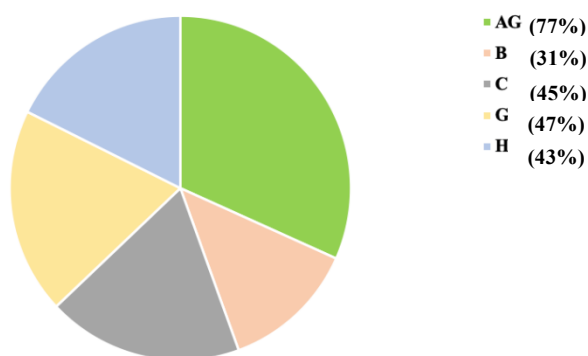


Figura 35 – Frequência (%) com que anticorpos IgG anti-C2V3C3 presentes no plasma dos doentes em estudo reagem contra os polipéptidos recombinantes AG (verde), B (laranja), C (cinzento), G (amarelo) e H (azul).

Quanto aos níveis de reatividade, estes foram maiores para o polipéptido G (mediana 2,09) seguido do péptido C (mediana 1,89). Contudo só se observou diferenças significativas entre os diferentes polipéptidos AG e G (Figura 36). Os níveis mais elevados de reatividade, encontrados contra o péptido G em relação ao péptido AG, podem ser explicados com base na semelhança do motivo – GPGQ – presente em quase todos os isolados virais em estudo e nos polipéptidos recombinantes G e C e ausente nos restantes, incluindo o polipéptido AG (Figura A7, Anexo 13).

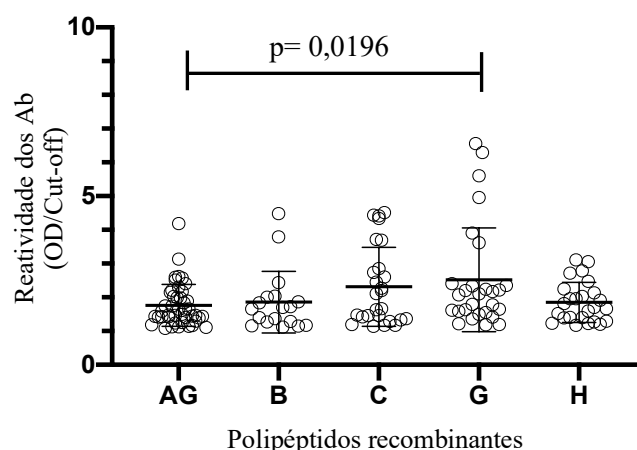


Figura 36 – Reatividade média dos anticorpos de ligação (OD/Cut-off) das 58 amostras de plasma contra os diferentes polipéptidos recombinantes. O teste One-Way ANOVA foi usado para comparar os valores medianos.

3.6 Análise da atividade neutralizante dos anticorpos IgG

No total de indivíduos em estudo, apenas 32% (N=19) possuíam anticorpos neutralizantes contra pelo menos um dos vírus/pseudovírus selecionados para os ensaios - pNL4.3 (subtipo B), TRO11(subtipo B), CE1176 (subtipo C) e BJOX2000 (CRF07_BC). Os indivíduos que apresentavam uma neutralização acima dos 20% estavam infectados por isolados dos subtipos C (93%) e B (7%) na região C2V3C3 e todos apresentavam tropismo para o co-recetor CCR5. A figura 37 representa o número de isolados virais que possuem atividade neutralizante contra o painel de pseudovírus usado. O pseudovírus CE1176 foi o menos frequentemente neutralizado ($p=0,0345$). Dado este facto, estabeleceu-se a comparação entre região C2V3C3 de uma sequência consenso MZ e as sequências dos pseudovírus do painel usado, todavia não foram encontradas nesta região alterações que permitam justificar o menor número de amostras capazes de neutralizar CE1176 (Figura A8, Anexo 14).

Consistente com isto, o vírus mais facilmente neutralizável foi NL4.3 (média de neutralização=40,7%), seguido de TRO11 (37,3%), BJOX200 (35,9%) e por fim CE1176 (32,8%). Como seria de esperar, o espectro de neutralização, ou seja, número de vírus neutralizados por cada amostra, correlacionou-se com a potência de neutralização (% de neutralização à diluição testada) ($F=4,837$; $p=0,0466$) (Figura 38).

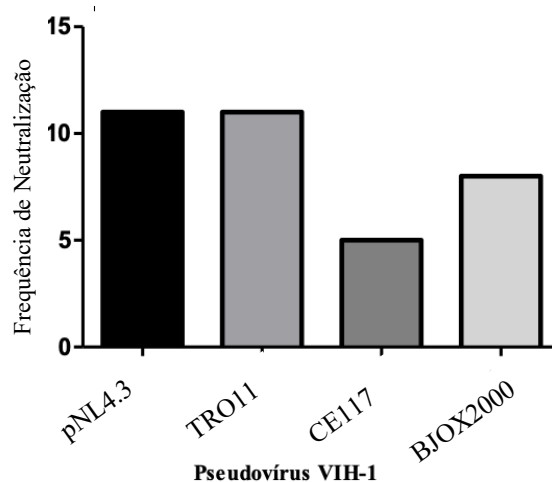


Figura 37 – Número de isolados virais que possuem atividade neutralizante contra o painel de pseudovírus usado. O teste exato de Fisher bilateral foi usado para comparar os valores de frequência. Apenas CE1176 (subtipo C) apresentou diferenças estatisticamente significativas ($p=0,0345$) entre a frequência com que era neutralizado comparativamente com os restantes pseudovírus – pNL4.3 (subtipo B), TRO11 (subtipo B) e BJOX2000 (CRF07_BC).

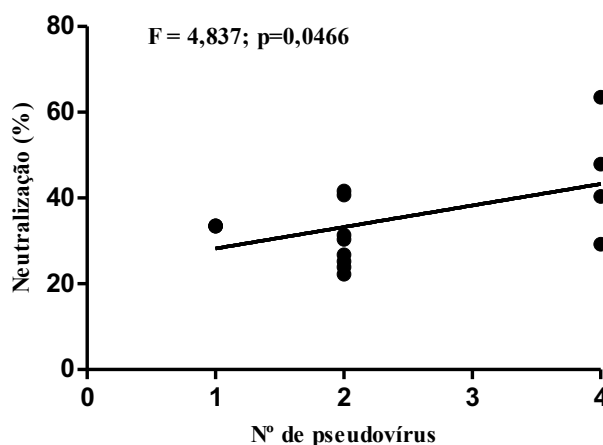


Figura 38 – Correlação entre a potência e o espectro da atividade neutralizante.

Com base na evidência de que o padrão de glicosilação da região C2V3C3 pode influenciar a atividade neutralizante [74,189,226] o número de locais de N-glicosilação da região C2V3C3 de cada amostra foi analisado em função da mesma. Observou-se uma associação inversa entre o número de locais de N-glicosilação na região C2V3C3 dos isolados virais dos indivíduos moçambicanos e a atividade neutralizante ($F=4,198$; $p=0,0485$) (Figura 39). Consistente com isto, os isolados virais com quatro locais de N-glicosilação em C2V3C3 possuem uma média de atividade neutralizante de 64,4%, que é significativamente diferente em relação às médias dos restantes isolados com números

de locais de N-glicosilação superiores (One-Way ANOVA, $p=0,0012$). Estes resultados sugerem que um menor número de locais de N-glicosilação está associado a uma maior atividade neutralizante.

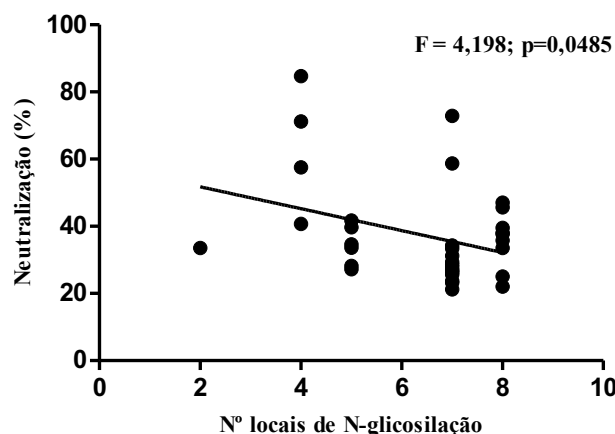


Figura 39 – Atividade neutralizante percentual em função do número de locais de N-glicosilação da região C2V3C3 de cada isolado viral.

Por fim, analisou-se a carga viral (cópias ARN/ml) dos indivíduos em função de apresentarem ($>20\%$) ou não atividade neutralizante ($<20\%$). Observou-se que a carga viral é mais elevada em indivíduos sem atividade neutralizante em comparação com os que possuem atividade neutralizante ($\log_{10}$ de cópias de ARN mediana/ml 4,20 vs. 3,00, $P = 0,0391$) (Figura 40A). Estes resultados sugerem que os anticorpos neutralizantes podem contribuir para a redução da carga viral nesta população em estudo.

A carga viral (cópias/ml) dos indivíduos sujeitos a terapêutica antirretroviral e indivíduos *naive*, em função da atividade neutralizante também foi analisada (Figura 40B). Os indivíduos com terapia e que apresentam atividade neutralizante ($>20\%$) possuem uma mediana logarítmica da carga viral mais baixa em comparação com os indivíduos que não apresentam atividade neutralizante ($<20\%$) ($\log_{10}$ de cópias de ARN mediana/ml 2,920 vs. 4,940, $P = 0,0303$). A mesma análise foi realizada relativamente ao número de linfócitos T $CD4^+$ (células/mm³), no entanto, não foram identificadas diferenças significativas ($p>0,05$).

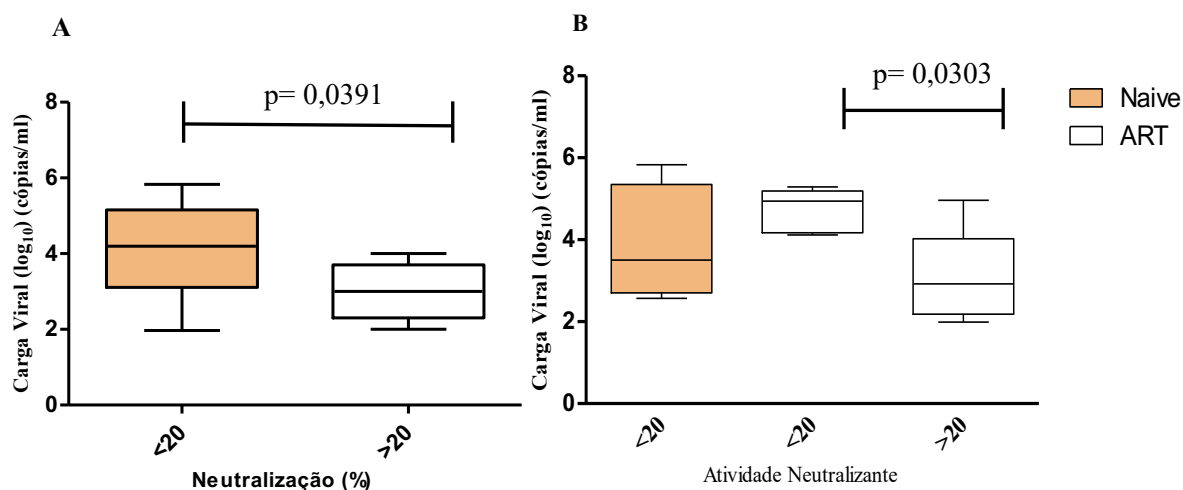


Figura 40 – Carga viral (cópias ARN/ml) dos indivíduos em função de possuírem ou não atividade neutralizante e terapia. A) Carga viral em indivíduos com e sem anticorpos neutralizantes. **B)** Carga viral dos indivíduos naive e com terapêutica com e sem anticorpos neutralizantes. Os valores de carga viral foram convertidos em logaritmos de base 10. As linhas denotam valores medianos, máximos e mínimos. A designação <20 está associado à ausência de atividade neutralizante e >20 está associado à atividade neutralizante superior a 20% (considerada significativa). O teste de Mann-Whitney foi usado para comprar as medianas do $\log_{10}$ da carga viral entre grupos.

4. Discussão e conclusões

Apesar da elevada prevalência [12,6% (11,9%-14,4%)] de VIH-1 em Moçambique [162], existem poucos estudos longitudinais que analisam a diversidade genética e a evolução molecular do vírus, bem como estudos que incluam a identificação de polimorfismos de resistência aos antirretrovirais em uso clínico. Além disso, não existem análises documentadas que caracterizem a resposta em anticorpos neutralizantes e não neutralizantes de indivíduos moçambicanos infetados pelo VIH-1. Neste estudo, abordou-se todos os aspetos destacados, considerados essenciais para um controlo eficaz da pandemia em Moçambique. A análise filogenética mostrou que a maioria dos isolados virais em estudo eram do subtipo C, como relatado em estudos anteriores realizados a sul de Moçambique [159, 163-166]. No entanto, os subtipos A, B, D e CRF02_AG também foram identificados num número significativo de indivíduos. A presença de subtipos não-C bem como de formas recombinantes pode ser explicada pela influência dos países da África Oriental - Quênia, Ruanda, Tanzânia e Uganda - nos quais circulam ativamente vírus de subtipo não-C, principalmente dos subtipos A e D [157,164,167]. Os isolados do subtipo C em circulação em Moçambique são muito diversos em termos genéticos e evolutivos indicando uma introdução recorrente, provavelmente, com origem nos países limítrofes. Estes resultados são consistentes com estudos anteriores que demonstraram a existência de múltiplas sub-linhagens circulantes de vírus do subtipo C na África Austral [227,228]. Uma análise filogeográfica e filodinâmica, que pretendemos realizar em breve, com mais isolados moçambicanos permitirá conhecer com detalhe a data de introdução e a origem dos isolados do subtipo C em circulação em Moçambique. É importante referir que a árvore filogenética da região C2V3C3 apresenta vinte sequências que formam um *cluster* específico com homologia com uma sequência de referência da África do Sul. Este resultado sugere que a origem deste *cluster* de isolados de subtipo C é a África do Sul, o que se justifica pela proximidade geográfica de Maputo com a África do Sul e pelo fluxo migratório laboral entre as regiões, como descrito em estudos anteriores [157,161,164,167]. Por outro lado, a elevada distância genética ($19,7 \pm 0,0266\%$) entre estas sequências de Moçambique e as restantes sequências do subtipo C sugerem a existência de um sub-subtipo do subtipo C [208]. A sequenciação e análise filogenética do genoma integral destes isolados virais será essencial para confirmar esta possibilidade [209]. Como seria de esperar, a região C2V3C3 foi a mais variável entre todos os isolados virais em estudo. Os isolados virais do subtipo C apresentaram entre si, os valores médios mais altos de distância genética, independentemente, da região do genoma em análise.

Esta observação é concordante com estudos anteriores que descrevem uma elevada diversidade intra-subtipo no subtipo C do VIH-1, principalmente em isolados virais da África Austral [227-229]. Os nove vírus recombinantes identificados possuem padrões de recombinação que incluem sempre regiões do genoma subtipo C associados a um ou mais subtipos (A, B, D ou CRF02_AG). Em três dos nove isolados virais o padrão de recombinação inclui três subtipos distintos. No entanto nenhum dos padrões de recombinação encontrados estão documentados na lista de CRF e CRF_cpx já descritas (<http://www.hiv.lanl.gov>).

Em 2004, os regimes de terapia altamente ativa (HAART) foram instituídos em Moçambique de forma gratuita, no entanto, a resistência aos antirretrovirais é um problema comum em todos os países da África Austral e pode comprometer o objetivo da OMS de eliminar o VIH-1 até 2030 [230]. Os países que relatam resultados acerca da prevalência de mutações de resistência adquiridas entre adultos com terapia para o VIH-1, em África, registam entre 50% a 97% de mutações de resistência para os NNRTI e entre 21% e 91% para os NRTI [231]. Neste estudo, analisou-se as mutações na RT, PR e na IN associadas à resistência aos antirretrovirais em uso clínico. As mutações encontradas e associadas à resistência aos NRTI foram na sua maioria (67%) TAMs (D67N, K70R, T215V, K219E) e K65R e M184V e foram identificadas em 15% (4/26) dos isolados virais. Relativamente aos NNRTI foram encontradas quatro mutações principais (K103N, G190A, V106A e Y181C) e quatro mutações acessórias (A98G, E138A, F227L e V108I) em 31% (8/26) dos isolados virais. Como seria de esperar, as mutações de resistência aos NRTI e NNRTI foram identificadas em indivíduos (27%, 7/26) cuja carga viral era > 10000 cópias ARN/ml e número de células T CD4⁺ eram < 200 células/mm³ indicando que estes se encontram infetados com vírus potencialmente resistentes aos regimes terapêuticos de primeira linha definidos em Moçambique. As mutações M184V e K65R foram encontradas em 12% (3/26) dos indivíduos que consequentemente não responderão também ao tratamento de segunda linha 3TC/FTC e TDF. Finalmente, as mutações M184V e K103N foram detetadas num indivíduo *naïve* e são as mutações de resistência primária mais comuns (conferentes de resistência ao tratamento de primeira linha) no continente africano, principalmente na África do Sul [223,224,232]. As mutações encontradas para estas classes de antirretrovirais, são semelhantes às obtidas nos estudos mais recentes realizados em Moçambique, em

indivíduos *naive* [167,168]. Além das mutações de resistências descritas, foram identificados 56 polimorfismos genéticos não associados a resistência direta aos antirretrovirais das classes NRTI e NNRTI. Os polimorfismos encontrados com maior frequência foram identificados em isolados do subtipo C. Como seria de esperar, a maioria dos polimorfismos genéticos encontrados ocorreu no subdomínio *fingers* da RT, e não houve polimorfismos na região corresponde ao centro ativo da enzima (D110, D185, D186) e noutras regiões altamente conservadas (*e.g.* YMDD) [233,234]. Relativamente aos PI não foram encontradas mutações principais; no entanto, foram identificadas quatro mutações acessórias (I47R, Q58E, G73S e L10F) em 9% (3/34) dos isolados virais em estudo. Dois dos isolados identificados com as mutações mencionadas são do subtipo C e A na região corresponde à enzima protease. A ausência de mutações principais para esta classe de antirretrovirais não é inesperado, uma vez que antirretrovirais da classe dos PI não fazem parte dos regimes terapêuticos de primeira linha em Moçambique [235]. Estes resultados são genericamente consistentes com os de estudos anteriores embora mutações acessórias na PR encontradas não sejam frequentemente descritas noutros estudos já realizados em Moçambique e noutras regiões da África Austral [152,159,166,174,175,227]. De notar que a mutação I47R, encontrada em dois isolados virais (subtipos C e A), é rara pelo que é desconhecido o seu impacto na suscetibilidade aos PI. Alguns dos polimorfismos encontrados potenciam a capacidade replicativa dos vírus na sua presença – K20R, M36I, L63P, H69K, V77I, V82I, L89M e I93L [134]. Verificou-se que a maioria dos polimorfismos encontrados ocorrem fora dos domínios funcionais da enzima (centro ativo incluindo a tríade catalítica, *flap* e a zona de dimerização da protease), à exceção de L23F, A28V e V82I que se encontram no centro ativo da enzima e R57K na região *flap*. Apesar de se encontrarem em regiões essenciais, apenas V82I intervém indiretamente num fenótipo resistente [134]. Alguns dos polimorfismos identificados são mais frequentes no subtipo A (L10I, I13V, A28V) e no subtipo C (T12S, I15V, L19I e R41K), coincidente com o descrito em estudos anteriores [174,236]. No contexto apresentado anteriormente, em que 27% dos indivíduos em estudo estão infetados com vírus potencialmente resistentes aos regimes de primeira linha, é essencial que a elevada eficácia dos inibidores da integrase associada à sua elevada barreira genética para a resistência e à progressiva diminuição de custo faz com que esta família de fármacos seja em breve introduzida na primeira linha terapêutica em Moçambique. Por isso é importante a identificação de polimorfismos de resistência aos

INSTI. Desde 2018 que se estuda a possibilidade de integrar INSTI nos regimes terapêuticos do país [173], no entanto na literatura é encontrado apenas um estudo focado na identificação de resistências aos INSTI em Moçambique [176]. No nosso estudo não foram encontradas mutações principais aos INSTI, sendo detetada apenas uma mutação acessória – G163R – num isolado no subtipo C de um indivíduo em regime terapêutico de primeira linha e com uma carga viral >10000 cópias/ml. Geralmente, esta mutação ocorre em conjunto com as mutações primárias, para reconstituir a *fitness* viral, no entanto, neste caso aparece de forma isolada [59,146, 237]. G163R ocorre na região central da integrase, numa posição essencial para a ligação a cofatores celulares de integração (e.g. LEDGF/p75) que intervêm na ligação da integrase à cromatina, permitindo a translocação do ADN proviral do citoplasma para o núcleo [57,58]. Esta mutação confere um baixo nível de resistência a RAL e EVG, no entanto estes fármacos possuem uma baixa barreira genética, ao contrário de DTG e BIC que podem ser usados como alternativa terapêutica neste caso [145,146]. Foram identificados 63 polimorfismos genéticos na integrase, entre os quais quatro associados ao subtipo C (L101I, T125A, V136Q e S283G) e outros quatro associados ao subtipo D (V37I, I113V, K211R e D256E), em concordância com o estudo de Moçambique realizado por Oliveira et al. (2012) [176]. Os polimorfismos identificados no nosso estudo também são semelhantes aos de Rogers, et al. (2018), da África do Sul [213] e sugerem que alguns dos polimorfismos encontrados (por ex. D25E, V31I, I72V, F100Y, L101I, T124A, L234I e S283G) podem afetar a estabilidade estrutural da enzima e influenciar a ligação da mesma ao ADN proviral em detrimento da ligação aos antirretrovirais [213]. Não foram encontradas mutações nos motivos conservados na integrase HCC e DDE. No entanto foi encontrada uma alteração na posição 188 (K188R) em 23% (5/22) dos isolados virais em estudo. O motivo K¹⁸⁶R¹⁸⁷K¹⁸⁸ participa na multimerização da enzima e permite a importação do ADN proviral para o núcleo da célula infetada [59]. Os polimorfismos D167E (14%, 3/22) e K173R (5%, 1/22) também identificados, ocorrem em posições envolvidas no processo de importação do ADN proviral para o núcleo. O polimorfismo D167E está associado a uma maior estabilidade estrutural da enzima em isolados virais não-B [213]. Em conclusão, os nossos dados sugerem que a introdução dos INSTI no programa nacional de tratamento de Moçambique poderia contribuir para o controlo da epidemia de VIH no país.

Os múltiplos potenciais locais de N-glicosilação (NXS/T) contribuem em mais de 50% para a massa molecular da gp120 e cobrem a maior parte da sua superfície [238]. Este escudo de glicanos pode intervir nos mecanismos de escape à neutralização [214,215]. A análise das sequências C2V3C3, evidenciou que o principal N-glicano (N301) da V3 (principal domínio neutralizante do invólucro viral) é maioritariamente conservado nos isolados em estudo (95%, 55/58). O mesmo não se verifica para C2 e principalmente para a C3, em que há variação dos locais de N-glicosilação o que pode ser consequência da pressão exercida pelos anticorpos neutralizantes contra esta região que conduz à evolução da sua matriz de glicanos [238]. O número médio de locais de N-glicosilação é maior em isolados virais do subtipo B, algo que já havia sido observado noutros estudos [225,239,240]. A região V3 de todos os isolados virais em estudo é constituída por 35 aminoácidos e nos subtipos A e C, o motivo GPGQ é altamente conservado enquanto que GPGR é predominante em isolados virais do subtipo B, em concordância com observações de estudos anteriores [217,218]. Foram encontrados 44 polimorfismos no conjunto dos 58 isolados virais analisados para esta região, sendo os mais frequentes: R306S, R315Q, V318Y, T319A, I320T, K322T, N325D e M326I. Os polimorfismos K305R e R308H estão sobretudo presentes no subtipo B e R315Q e A316T no subtipo C. Os polimorfismos K305R, A316T, V318Y e T319A estão associados à resistência ao maraviroc, por aumento da interação entre a *loop* V3 e o co-recetor CCR5 [241-244]. A316T, V318Y e T319A estão ainda associados a um aumento da capacidade replicativa viral [222,244]. Ainda na análise da região V3, como determinante principal do tropismo viral, observou-se que 86% (50/58) dos isolados virais em estudo utilizam potencialmente o co-recetor CCR5 aquando da entrada na célula hospedeira. Os isolados virais que utilizam CXCR4 são do subtipo C o que é interessante uma vez que os isolados X4 são muito raros neste subtipo [221,222,245]. Como seria de esperar, a V3 dos isolados que usam CXCR4 têm uma carga total significativamente superior que os vírus R5, em consequência de um aumento de aminoácidos carregados positivamente [221,222]. As variantes X4 apresentam uma maior variabilidade aminoacídica na V3 em comparação com os isolados virais R5 [221,222]. As posições aminoacídicas em que se observa uma maior variabilidade entre os isolados com diferentes tropismos são 305-307, 317-322 e 325 e 328 (numeração relativa ao isolado HXB2). As mutações T319A, K322D e N325D são mais frequentes em isolados R5 e as mutações K305Q, F317L, K322N e Q328K mais

frequentes em isolados X4 ($p < 0,05$). A ausência de dados referentes à data de infecção dos indivíduos não permite estabelecer uma comparação com estudos anteriores que remetem muitas vezes, para uso do co-recetor X4 em infecções crônicas. No entanto, os nossos dados contribuem para a monitorização de características fundamentais dos isolados virais em circulação em Moçambique.

No nosso estudo realizamos a primeira caracterização da diversidade antigénica de indivíduos moçambicanos infetados pelo VIH-1 com base na resposta de anticorpos contra polipéptidos da região C2V3C3 de diferentes subtipos (B, C, G, H e CRF02_AG). A maioria dos doentes (47/58, 81%) possuem anticorpos de ligação contra pelo menos um dos polipéptidos recombinantes sendo que 77% dos doentes ($N=45$) possuía anticorpos contra o polipéptido CRF02_AG e apenas 45% possuía anticorpos contra o polipéptido C. Este resultado não era esperado uma vez que o subtipo C é predominante em Moçambique, como descrito anteriormente [159, 163-166]. No entanto a análise da sequência aminoacídica do polipéptido CRF02_AG evidenciou conservação das regiões de ligação ao CD4 que são semelhantes à sequência aminoacídica consenso dos isolados virais de Moçambique, bem como dos motivos de N-glicosilação. Estas observações podem indicar que o fragmento de ADN correspondente à região C2V3C3 do vírus do subtipo C, usado para a produção do polipéptido do subtipo C não é representativo dos vírus subtipo C circulantes em Moçambique. A análise filogenética evidenciou uma similaridade entre a sequência do polipéptido recombinante do subtipo C com sequências de isolados virais do subtipo C de Angola. A topologia da árvore mostra uma distância filogenética considerável das sequências de referência de Moçambique e das sequências dos isolados em estudo. Por outro lado, a baixa ligação ao polipéptido B (31%, $N=18$) é consistente com o reduzido número de doentes infetados por este subtipo viral e com as diferenças significativas existentes entre o subtipo B e os restantes subtipos ao nível da região C2V3C3 [225,239,240]. Em contrapartida, os polipéptidos G e C possuem os níveis de reatividade mais elevados. Este resultado pode ser justificado com base na semelhança do motivo GPGQ entre as sequências dos polipéptidos em questão e a maioria das sequências dos isolados em estudo (93%, $N=54$). Também se realizou a primeira avaliação da resposta de anticorpos neutralizantes (NAbs) contra o VIH-1 em indivíduos moçambicanos. Os NAbs contribuem para uma resposta imune eficaz contra o VIH-1 [65,89,106,113,115]. Os principais obstáculos para induzir uma resposta imune contendo

NAbs amplos e potentes, incluem a elevada diversidade genética do invólucro viral bem como a ocultação dos epítomos neutralizantes pelos N-glicanos [246]. No nosso estudo, apenas 32% (N=15) dos indivíduos infetados com VIH-1 possuem anticorpos com atividade neutralizante contra pelo menos um pseudovírus do painel utilizado. Os indivíduos que apresentavam atividade neutralizante (>20%) estavam infetados por isolados virais R5 do subtipo C. Observou-se uma correlação entre o espectro de neutralização (número de vírus neutralizado por cada amostra) e a potência de neutralização (% de neutralização). Como seria de esperar, o vírus mais facilmente neutralizável foi NL4.3 (subtipo B, tier 1), logo seguido pelo TRO11 (subtipo B), BJOX200 (CRF07_BC) e por fim CE1176 (subtipo C). O pseudovírus CE1176 (subtipo C) seguido do BJOX200 (CRF07_BC) foram os mais difíceis de neutralizar não obstante terem sequências C2V3C3 semelhantes às dos vírus dos doentes deste estudo indicando que a atividade neutralizante nestes indivíduos é maioritariamente determinada por epítomos conformacionais que serão diferentes dos existentes nos isolados de referência do painel de neutralização e poderão não estar localizados na região C2V3C3 [247,248]. Estudos anteriores demonstraram que os epítomos neutralizantes da região V3 dos isolados do subtipo B encontram-se, em geral, mais expostos que os epítomos da mesma região no subtipo C, o que impede a acessibilidade de alguns anticorpos [248]. Os nossos resultados são consistentes com isto, uma vez que no nosso estudo o pseudovírus B foi mais facilmente neutralizado que o C. A ausência de informação relativa ao ano de diagnóstico da infeção por VIH-1 dos pacientes é uma limitação do nosso estudo, pois está demonstrado que o tempo de infeção influencia o desenvolvimento da amplitude de neutralização, sendo por isso importante associar essa informação aos resultados obtidos nos ensaios de neutralização [249]. Para completar os nossos estudos seria interessante produzir pseudovírus com Env dos isolados moçambicanos o que permitiria explorar determinantes de neutralização dos doentes moçambicanos para além da região C2V3C3 (região V1V2 e outras) [60,63,250]. A disponibilidade de pseudovírus moçambicanos permitiria ainda testar a suscetibilidade a um painel alargado de anticorpos monoclonais humanos neutralizantes, por forma a perceber quais os anticorpos neutralizantes mais úteis para utilizar na prevenção e tratamento por imunoterapia de indivíduos infetados por VIH-1 em Moçambique [113,115,117, 250]. Neste estudo, e em concordância com outros já publicados, observou-se uma associação inversa entre a potência de neutralização e o número de locais de N-glicosilação encontrados nos isolados virais

[74,189,226,248]. Estes dados sugerem que um maior número de locais de N-glicosilação pode ocultar regiões vulneráveis do invólucro viral impedindo o seu reconhecimento e consequentemente a atuação de anticorpos neutralizantes [248]. Por fim, observou-se que a carga viral foi mais reduzida em indivíduos com atividade neutralizante do que em indivíduos sem atividade neutralizante. Estes dados sugerem que os anticorpos neutralizantes contribuem para a diminuição da replicação viral nesta população em paralelo com os antirretrovirais [113,115,117, 250].

Em resumo, o nosso estudo mostra que a epidemia de VIH-1 em Maputo está a evoluir rapidamente para uma maior complexidade genética e epidemiológica. É necessário analisar e monitorizar as estirpes de VIH-1 circulantes por forma a compreender a sua origem e a sua propagação por todo o país e também o potencial impacto dessa diversidade nos regimes terapêuticos vigentes no país. A análise dos perfis de resistência dos isolados virais de Moçambique mostrou a ineficácia dos regimes terapêuticos de primeira e segunda linha numa percentagem significativa de indivíduos e por isso a necessidade de introduzir novos fármacos no tratamento da infeção por VIH em Moçambique. Neste sentido, os INSTI, principalmente os de segunda geração, podem ser essenciais para a criação de novos regimes terapêuticos eficazes. No entanto, esta introdução requer uma vigilância constante da prevalência de mutações capazes de conferir resistência aos fármacos em questão. Apesar da TARV contribuir para o controlo da pandemia, a acumulação constante de mutações irá, a longo prazo, culminar na ineficácia dos antirretrovirais em uso clínico. Por isso é importante continuar a investir no desenvolvimento de vacinas destinadas a induzir anticorpos amplamente neutralizantes bem como novas abordagens de imunização passiva que envolvem combinações de bNAbs que têm como alvo múltiplos epítomos. No entanto, tanto a imunização ativa como a passiva terão de contornar a diversidade genética do VIH-1 no próprio hospedeiro e a nível global para atingir um nível de eficácia significativo. Os anticorpos amplamente neutralizantes, possivelmente poderão oferecer uma vantagem sobre a TARV tradicional tanto na prevenção como para o tratamento da infeção pelo VIH-1 e no futuro serão imprescindíveis para países como Moçambique.

5. Bibliografia

- [1] - Centers of Disease Control and Prevention (CDC). Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men New York City and California. *MMWR Morb Mortal Wkly Rep.* 1981;31(23):305–8
- [2] - Hymes KB, Greene JB, Marcus A, et al. Kaposi's sarcoma in homosexual men. A report of eight cases. *Lancet.* 1981; 318 (8247): 598-600
- [3] - Gottlieb M.S, Schroff R, Schanker H.M, et al. Pneumocystis carinii Pneumonia and Mucosal Candidiasis in Previously Healthy Homosexual Men. *New England Journal of Medicine.* 1981; 305(24): 1425-1431
- [4] - Siegal FP, Lopez C, Hammer GS, et al. Severe Acquired Immunodeficiency in Male Homosexuals, Manifested by Chronic Perianal Ulcerative Herpes Simplex Lesions. *N Engl J Med.* 1981; 305(24): 1439-1444
- [5] - Centers of Disease Control and Prevention (CDC). Trends Update on Acquired Immune Deficiency Syndrome (AIDS) – United States. *MMWR Morb Mortal Wkly Rep.* 1981; 31(26): 353-361
- [6] - Arthur E, Pitchenik M.D, Margaret A, et al. Opportunistic Infections and Kaposi's Sarcoma Among Haitians: Evidence of a New Acquired Immunodeficiency State. *Annals of Internal Medicine.* 1983; 98(3): 277-284
- [7] - Vicira J, Frank E, Spira TJ, et al. Acquired immune deficiency syndrome in black africans. *Lancet.* 1983; 308: 125-129
- [8] - Gerstoft J, Malchow-Moller A, Bygbjerg I. Severe acquired immunodeficiency European homosexual men. *Br Med J.* 1982; 285(6334): 17-19
- [9] - Nambiar D, Weir A, Aspinall EJ, et al. Mortality and cause of death in a cohort of people who had ever injected drugs in Glasgow: 1982-2012. *Elsevier.* 2014; 147(2015): 215-221
- [10] - Centers of Disease Control and Prevention (CDC). Epidemiologic notes and Reports Reports Pneumocystis carinii Pneumonia among persons with hemophilia A. *MMWR Morb Mortal Wkly Rep.* 1982;31(27):365–7
- [11] - Centers of Disease Control and Prevention (CDC). Immunodeficiency among female sexual partners of males with acquired immune deficiency syndrome (AIDS). *MMWR Morb Mortal Wkly Rep.* 1983;31(52):697–8

- [12] - Barré-Sinoussi F, Chermann JC, Rey F, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Rev Investig Clin.* 2004;56(2):126–9
- [13] - Popovic M, Sarngadharan MG, Read E, Gallo RC. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science.* 1984; 224(4648): 497–500
- [14] - Coffin J, Haase A, Levy JA, et al. Human immunodeficiency viruses. *Science.* 1986; 232(4751): 697
- [15] - World Health Organization (WHO). AIDS action. *WHO report.* 1987; 1: 1-12
- [16] - Clavel F, Guétard D, Brun-Vézinet F, et al. Isolation of a New Human Retrovirus from West African Patients with AIDS. *Science.* 1986; 233: 343-346
- [17] - Clavel F, Mansinho K, Chamaret S, et al. Human immunodeficiency virus type 2 infection associated with AIDS in West Africa. *N Engl J Med.* 1987; 316(19): 1180-1185
- [18] - Hemelaar J. The origin and diversity of the HIV-1 pandemic. *Cell press.* 2012; 18(3): 182-192
- [19] - Locatelli S, Peeters M. Cross-species transmission of simian retroviruses: How and why they could lead to the emergence of new diseases in the human population. *Aids.* 2012; 26(6): 659–73
- [20] - Peeters M, Jung M, Ayoub A. The origin and molecular epidemiology of HIV. *Expert Rev Anti Infect Ther.* 2013;11(9):885–96
- [21] - Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med.* 2011; 1(1): 1–22
- [22] - Santiago ML, Range F, Keele BF, et al. Simian Immunodeficiency Virus Infection in Free-Ranging Sooty Mangabeys (*Cercocebus atys atys*) from the Tai Forest, Cote d'Ivoire: Implications for the Origin of Epidemic Human Immunodeficiency Virus Type 2. *Journal of Virology.* 2005; 79(19): 12515-12527
- [23] - Gao F, Yue L, Robertson DL, et al. Genetic diversity of human immunodeficiency virus type 2: evidence for distinct sequence subtypes with differences in virus biology. *J Virol.* 1994; 68(11): 7433–47
- [24] - Tebit DM, Arts EJ. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *Lancet Infect Dis.* 2011;11(1): 45–56

- [25] - Yamaguchi J, McArthur C, Vallari A, et al. Complete genome sequence of CG-0018a-01 establishes HIV-1 subtype L. *Journal of Acquired Immune Deficiency Syndromes*. 2020; 83(3): 319-22
- [26] - Laboratory LAN. HIV Circulating Recombinant Forms (CRFs). 2020 [cited 2020 Dezembro]; Available from: <https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>
- [27] - UNAIDS. DATA 2020. [cited 2021 June]; Available from: https://www.unaids.org/sites/default/files/media_asset/2020_aids-data-book_en.pdf
- [28] - Hemelaar J, Elangovan R, Yun J, et al. Global and regional molecular epidemiology of HIV-1, 1990–2015: a systematic review, global survey, and trend analysis. *Lancet Infect Dis*. 2018; 19(2): 143-55
- [29] - Bbosa N, Kaleebu P, Ssemwanga D. HIV subtype diversity worldwide. *Curr Opin HIV AIDS*. 2019;14(3):153–60
- [30] - Faria NR, Vidal N, Lourenco J, et al. Distinct rates and patterns of spread of the major HIV-1 subtypes in Central and East Africa. *PLoS Pathog*. 2019; 15(12): 1–23
- [31] - Oster AM, Switzer WM, Hernandez AL, et al. Increasing HIV-1 subtype diversity in seven states, United States, 2006–2013. *Elsevier*. 2017; 27(4):244-251
- [32] - Tumiotto C, Bellecave P, Recordon-Pinson P, et al. Diversity of HIV-1 in Aquitaine, Southwestern France, 2012-2016. *AIDS Res Hum Retroviruses*. 2018; 34(5): 471–473
- [33] - Sallam M, Esbjörnsson J, Baldvinsdóttir G, et al. Molecular epidemiology of HIV-1 in Iceland: Early introductions, transmission dynamics and recent outbreaks among injection drug users. *Infect Genet Evol*. 2017; 49: 157–63
- [34] - Hernandez-Sanchez PG, Guerra-Palomares SE, Ramirez-Garcialuna JL, et al. Prevalence of drug resistance mutations in protease, reverse transcriptase, and integrase genes of north central Mexico HIV isolates. *AIDS Res Hum Retroviruses*. 2018; 34(6): 498–506
- [35] - Aibekova L, Foley B, Hortelano G, Raees M, et al. Molecular epidemiology of HIV-1 subtype A in former Soviet Union countries. *PLoS One*. 2018;13(2):1–10

- [36] - Xiao P, Li J, Fu G, Zhou Y, et al. Geographic distribution and temporal trends of HIV-1 subtypes through heterosexual transmission in china: A systematic review and meta-analysis. *Int J Environ Res Public Health*. 2017;14(7): 1-18
- [37] - Sallam M, Şahin GÖ, Ingman M, et al. Genetic characterization of human immunodeficiency virus type 1 transmission in the Middle East and North Africa. *Elsevier*. 2017; 3(7): 1-19
- [38] - P.Smyth R, Negroni M. A step forward understanding HIV-1 diversity. *Retrovirology*. 2016; 13(27): 1-3
- [39] - Ramirez BC, Simon-Loriere E, Galetto R, Negroni M. Implications of recombination for HIV diversity. *Virus Res*. 2008;134(1–2): 64–73
- [40] - Sanjuán R, Nebot MR, Chirico N, et al. Viral Mutation Rates. *J Virol*. 2010; 84(19): 9733–48
- [41] - Domingo E, Sheldon J, Perales C. Viral Quasispecies Evolution. *Microbiol Mol Biol Rev*. 2012; 76(2): 159–216
- [42] - Sierra S, Kupfer B, Kaiser R. Basics of the virology of HIV-1 and its replication. *J Clin Virol*. 2005; 34(4): 233–44
- [43] - Rajarapu G. Genes and Genome of HIV-1. *Journal of Phylogenetics & Evolutionary Biology*. 2014; 2(1): 1-7
- [44] - Sukosd Z, S.Andersen E, Seemann S, et al. Full-length RNA structure prediction of the HIV-1 genome reveals a conserved core domain. *Nucleic Acids Research*. 2015; 43(21): 10168-10179
- [45] - Li G, Clercq E. HIV Genome-Wide Protein Associations: a Review of 30 Years of Research. *Microbiology and Molecular Biology Reviews*. 2011; 43(5): 678–82
- [46] - ViralZone. HIV Genome. 2019 [cited 2020 Dezembro]; Available from: <https://viralzone.expasy.org/4976>
- [47] - Taveira N, Martin F. O agente. In: Manual sobre a SIDA. editors, Antunes F, Maltez F. 5º. Lisboa; 2018. p. 1–11
- [48] - Los Alamos National Laboratory. HIV Sequence Compendium 2018 [cited 2020 Dezembro]; Available from : <https://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/2018/sequence2018.pdf>

- [49]- Laboratory LAN. HIV sequence database. 2019 [cited: 2020]; Available from: <https://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html>
- [50] - Ghosh A, Osswald H, Prato G. Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS. *J Med Chem.* 2016; 59(11): 5172–5208
- [51] - Konvalinka J, Kräusslich HG, Müller B. Retroviral proteases and their roles in virion maturation. *Virology.* 2015; 1-15
- [52] - Loius J, Aniana A, Weber I, Sayer J. Inhibition of autoprocessing of natural variants and multidrug resistant mutant precursors of HIV-1 protease by clinical inhibitors. *Proceedings of the National Academy of Sciences*, 108(22), 9072–9077
- [53] - Hu Wei-Shau, Hughes S. HIV-1 Reverse Transcription. *Cold Spring Harbor Perspectives in Medicine.* 2012; 2(1): 1-22
- [54] - Tarasova O, Poroikov V, Veselovsky A. Molecular Docking Studies of HIV-1 Resistance to Reverse Transcriptase Inhibitors: Mini-Review. *Molecules.* 2018; 23(5): 11–3
- [55] - Le Grice SFJ. Human immunodeficiency virus reverse transcriptase: 25 years of research, drug discovery, and promise. *J Biol Chem.* 2012; 287(49): 40850–7
- [56] - Métifiot M, Marchand C, Maddali K, Pommier Y. Resistance to integrase inhibitors. *Viruses.* 2010; 2(7): 1347–66
- [57] - Marchand C, Johnson A, Semenova E, Pommier Y. Mechanisms and inhibition of HIV integration. *Drug Discov Today Dis Mech.* 2006; 3(2): 253-260
- [58] - Sala M, Spensiero A, Esposito F, et al. Development and identification of a novel anti-HIV-1 peptide derived by modification of the N-terminal domain of HIV-1 integrase. *Front Microbiol.* 2016; 7(845):1–9
- [59] - Silberstein F, Malet I. Characterization and Structural Analysis of HIV-1 Integrase Conservation. *AIDS Rev.* 2009; 11:17–29
- [60] - Cavrois M, Neidleman J, Santiago ML, et al. Enhanced Fusion and Virion Incorporation for HIV-1 Subtype C Envelope Glycoproteins with Compact V1/V2 Domains. *J Virol.* 2014; 88(4): 2083–94
- [61] - Araújo LAL, Almeida SEM. HIV-1 diversity in the envelope glycoproteins: Implications for viral entry inhibition. *Viruses.* 2013; 5(2): 595–60

- [62] - Rathore U, Saha P, Kesavardhana S, Kumar AA, et al. Glycosylation of the core of the HIV-1 envelope subunit protein gp120 is not required for native trimer formation or viral infectivity. *J Biol Chem*. 2017; 292(24): 10197–219
- [63] - O’Connell RJ, Kim JH, Excler JL. The HIV-1 gp120 V1V2 loop: Structure, function and importance for vaccine development. *Expert Rev Vaccines*. 2014; 13(12):1489–500
- [64] - Antell GC, Dampier W, Aiamkitsumrit B, et al. Utilization of HIV-1 envelope V3 to identify X4- and R5-specific Tat and LTR sequence signatures. *Retrovirology*. 2016; 13(1):1–14
- [65] - Schooten J, Gils M. HIV-1 immunogens and strategies to drive antibody responses towards neutralization breadth. *Retrovirology*. 2018; 15(74): 1-11
- [66] - Rao M, Peachman K, Kim J, et al. HIV-1 Variable Loop 2 and its Importance in HIV-1 Infection and Vaccine Development. *Curr HIV Res*. 2013; 11(5): 427–38
- [67] - Aiamkitsumrit B, Dampier W, Antell G, et al. Bioinformatic Analysis of HIV-1 Entry and Pathogenesis. *Curr HIV Res*. 2014; 12(2): 132–61
- [68] - Beauparlant D, Rusert P, Magnus C, et al. Delineating CD4 dependency of HIV-1: Adaptation to infect low level CD4 expressing target cells widens cellular tropism but severely impacts on envelope functionality. *PLOS pathogens*. 2017; 13(3): 1-42
- [69] - Oppermann M. Chemokine receptor CCR5: Insights into structure, function, and regulation. *Elsevier*. 2004;16(11): 1201–10
- [70] - Ruibal-Ares B, Belmonte L, Bare P, et al. HIV-1 Infection and Chemokine Receptor Modulation. *Curr HIV Res*. 2005; 2(1): 39–50
- [71] - Fouchier RAM, Brouwer M, Broersen SM, Schuitemaker H. Simple determination of human immunodeficiency virus type 1 syncytium- inducing V3 genotype by PCR. *J Clin Microbiol*. 1995; 33(4): 906–11
- [72] - McBurney SP, Ross TM. Viral sequence diversity: Challenges for AIDS vaccine designs. *Expert Rev Vaccines*. 2008; 7(9): 1405–17
- [73] - Shen HS, Yin J, Leng F, et al. HIV coreceptor tropism determination and mutational pattern identification. *Sci Rep*. 2016; 6(21280): 1–11

- [74] - Wood NT, Fadda E, Davis R, et al. The influence of N-linked glycans on the molecular dynamics of the HIV-1 gp120 V3 loop. *PLoS One*. 2013;8(11): 1-9
- [75] - Freed E. HIV-1 assembly, release and maturation. *Nat Rev Microbiol*. 2015;13(8): 484–96
- [76] - Freed EO. HIV-1 replication. *Somat Cell Mol Genet*. 2001; 26(1/6): 13–33
- [77] - Wilen CB, Tilton JC, Doms RW. HIV: Cell Binding and Entry. *Cold Spring Harb Perspect Med*. 2012; 2(1): 1–5
- [78] - Hughes S. Reverse Transcription of Retroviruses and LTR Retrotransposons. *Microbiol Spectr*. 2015; 3(2): 1–40
- [79] - Pereira J. Ciclo biológico do VIH. In: Antunes F, Maltez F, editors. Manual sobre a SIDA. 5ª. Lisboa; 2018. p. 12–20
- [80] - Sertznig H, Hillebrand F, Erkelenz S, et al. Behind the scenes of HIV-1 replication: Alternative splicing as the dependency factor on the quiet. *Virology*. 2018; 516:176–88
- [81] - Rausch JW, le Grice SFJ. HIV Rev assembly on the Rev response element (RRE): A structural perspective. *Viruses*. 2015; 7(6): 3053–75
- [82] - Sakuragi JI. Morphogenesis of the infectious HIV-1 virion. *Front Microbiol*. 2011; 2(242) 1–5
- [83] - Alfadhli A, Staubus A, Tedbury P, et al. Analysis of HIV-1 Matrix-Envelope Cytoplasmic Tail Interactions. 2019; 93(21): 1–10
- [84] - Sundquist WI, Krausslich HG. HIV-1 Assembly, Budding, and Maturation. *Cold Spring Harb Perspect Med*. 2012; 2:1–24
- [85] - Moir S, Chun TW, Fauci AS. Pathogenic mechanisms of HIV disease. *Annu Rev Pathol Mech Dis*. 2011; 6:223–48
- [86] - Cohen M, Shaw G, McMichael A, Haynes B. Acute HIV-1 Infection. *N Engl J Med*. 2011;364(20):1943–54
- [87] - Février M, Dorgham K, Rebollo A. CD4+ T cell depletion in human immunodeficiency virus (HIV) infection: Role of apoptosis. *Viruses*. 2011; 3(5): 586–612

- [88] - Swanstrom R, Coffin J. HIV-1 Pathogenesis :The Virus. *Cold Spring Harb Perspect Med.* 2012; 2: 1–18
- [89] - Overbaugh J, Morris L. The antibody response against HIV-1. *Cold Spring Harb Perspect Med.* 2012; 2(1): 1–17
- [90] - Paiardini M, Müller-Trutwin M. HIV-associated chronic immune activation. *Immunol Rev.* 2013; 254(1): 78–101
- [91] - Naif HM. Pathogenesis of HIV infection. *Infect Dis Rep.* 2013; 5:26–30
- [92] - Centers for Disease Control and Prevention (CDC). About HIV. 2020 [cited: 2020]; Available from: <https://www.cdc.gov/hiv/basics/whatishiv.html>
- [93] - Altfeld M, Gale M. Innate immunity against HIV-1 infection. *Nat Immunol.* 2015;16(6):554–62
- [94] - Chaplin D. Overview of the immune response. *J Allergy Clin Immunol.* 2010;125(2):3–23
- [95] - Walker B, McMichael A. The T-cell response to HIV. *Cold Spring Harb Perspect Med.* 2012;2(11):1–19
- [96] - Swain SL, McKinstry KK, Strutt TM. Expanding roles for CD4 + T cells in immunity to viruses. *Nat Rev Immunol.* 2012;12(2):136–48
- [97] - McMichael A, Rowland-Jones S. Cellular immune responses to HIV. *Nature.* 2001;410: 980-987
- [98] - Miles B, Miller SM, Connick E. CD4 T follicular helper and regulatory cell dynamics and function in HIV infection. *Front Immunol.* 2016;7(659):1–9
- [99] - Cummins NW, Badley AD. Making sense of how HIV kills infected CD4 T cells: implications for HIV cure. *Mol Cell Ther.* 2014;2(20): 1-7
- [100] - Hoffman W, Lakkis FG, Chalasani G. B cells, antibodies, and more. *Clin J Am Soc Nephrol.* 2016;11(1):1-18
- [101] - Regep C, Georges G, Shi J, et al. The H3 loop of antibodies shows unique structural characteristics. *Proteins Struct Funct Bioinformatics.* 2017; 85(7):1311–8

- [102] - Vidarsson G, Dekkers G, Rispens T. IgG subclasses and allotypes: From structure to effector functions. *Front Immunol*. 2014; 5(520):1–17
- [103] - Banerjee K, Klasse PJ, Sanders RW, et al. IgG subclass profiles in infected HIV type 1 controllers and chronic progressors and in uninfected recipients of env vaccines. *AIDS Res Hum Retroviruses*. 2010;26(4):445–58
- [104] - Gao Y, McKay Paul F, Mann Jamie F.S. Advances in HIV-1 Vaccine Development. *Viruses*. 2018; 10 (167): 1-16
- [105] - Jaskova K, Pavlovicova M, Jurkovicova D. B cells - ontogenesis and immune memory development. *Gen Physiol Biophys*. 2014;31(4): 375–82
- [106] - Mouquet H. Antibody B cell responses in HIV-1 infection. *Trends Immunol*. 2014;35(11):549–61
- [107] - Alter G, Moody MA. The humoral response to HIV-1: new insights, renewed focus. *J Infect Dis*. 2010; 202: 1-13
- [108] - Akkaya M, Kwak K, Pierce SK. B cell memory: building two walls of protection against pathogens. *Nat Rev Immunol*. 2020;20(4):229–38
- [109] - Crotty S. A brief history of T cell help to B cells - A timeline of B cell help discoveries. *Nat Rev Immunol*. 2015;15(3):185–9
- [110] - Eivazi S, Bagheri S, Hashemzadeh MS, et al. Development of T follicular helper cells and their role in disease and immune system. *Elsevier*. 2016; 84:1668–78
- [111] - Tomaras GD, Haynes BF. HIV-1-specific antibody responses during acute and chronic HIV-1 infection. *Curr Opin HIV AIDS*. 2009;4(5):373–9
- [112] - Baum L. Role of Humoral Immunity in Host Defense Against HIV. *Curr HIV/AIDS Rep*. 2010;7(1):11–8
- [113] - Yaseen MM, Yaseen MM, Alqudah MA. Broadly neutralizing antibodies: An approach to control HIV-1 infection. *Int Rev Immunol*. 2017;36(1):31–40
- [114] - Sather DN, Carbonetti S, Kehayia J, et al. Broadly Neutralizing Antibodies Developed by an HIV-Positive Elite Neutralizer Exact a Replication Fitness Cost on the Contemporaneous Virus. *J Virol*. 2012;86(23):12676–85
- [115] - Doria-Rose N, Landais E. Coevolution of HIV-1 and broadly neutralizing antibodies. *Curr Opinion*. 2019; 14(4): 286-293

- [116] - Borrow P, Moody MA. Immunologic characteristics of HIV-infected individuals who make broadly neutralizing antibodies. *Immunol Rev.* 2017;275(1):62–78
- [117] - Liu Y, Cao W, Sun M, Li T. Broadly neutralizing antibodies for HIV-1: efficacies, challenges and opportunities. *Emerg Microbes Infect.* 2020;9(1):194–206
- [118] - Escolano A, Gristick HB, Abernathy ME, et al. Immunization expands HIV-1 V3-glycan specific B-cells in mice and macaques. *Nature.* 2019;570(7762):468–73
- [119] - Gao Y, McKay PF, Mann JFS. Advances in HIV-1 vaccine development. *Viruses.* 2018;10(4):1–26
- [120] - Bontempo A, Garcia M, Rivera N, Cayabyab M. A Systematic Approach to HIV-1 Vaccine Immunogen Selection. *AIDS Res Hum Retroviruses.* 2020;1–22
- [121] - Robinson HL. HIV/AIDS Vaccines: 2018. *Clin Pharmacol Ther.* 2018;104(6):1062–73
- [122] - Hsu DC, O’Connell RJ. Progress in HIV vaccine development. *Human Vaccines Immunother.* 2017;13(5):1018–30
- [123] - Valuev-Elliston VT, Kochetkov SN. Novel HIV-1 non-nucleoside reverse transcriptase inhibitors: A combinatorial approach. *Biochemistry.* 2017;82(13):1716–43
- [124] - Zulfiqar HF, Javed A, Sumbal A, et al. HIV Diagnosis and Treatment through Advanced Technologies. *Front Public Heal.* 2017;5(32):1–16
- [125] - Maeda K, Das D, Kobayakawa T, Tamamura H, Takeuchi H. Discovery and Development of Anti-HIV Therapeutic Agents: Progress Towards Improved HIV Medication. *Curr Top Med Chem.* 2019;19(18):1621–49
- [126] - Tang MW, Shafer RW. HIV-1 antiretroviral resistance: Scientific principles and clinical applications. *Drugs.* 2012;72(9):1–25
- [127] - Arts EJ, Hazuda DJ. HIV-1 antiretroviral drug therapy. *Cold Spring Harb Perspect Med.* 2012;2(4): 1-23
- [128] - De Clercq E. Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. *Int J Antimicrob Agents.* 2009;33(4):307–20

- [129] - Matthews T, Salgo M, Greenberg M, et al. Enfuvirtide: The first therapy to inhibit the entry of HIV-1 into host CD4 lymphocytes. *Nat Rev Drug Discov*. 2004;3(3):215–25
- [130] - Westby M, Van Der Ryst E. CCR5 antagonists: Host-targeted antivirals for the treatment of HIV infection. *Antivir Chem Chemother*. 2005;16(6):339–54
- [131] - Markham A. Fostemsavir: First Approval. *Drugs*. 2020; 1: 1-6
- [132] - World Health Organization (WHO). Update of recommendations on first-and second-line antiretroviral regimes. 2019. [cited: June 2021]; Available from: <https://www.who.int/publications/i/item/WHO-CDS-HIV-19.15>
- [133] - Croxford S, Kitching A, Desai S. Mortality and causes of death in people diagnosed with HIV in the era of highly active antiretroviral therapy compared with the general population: an analysis of a national observational cohort. *Lancet*. 2017; 2:35-46
- [134] - Wensing AM, Calvez V, Ceccherini-Silberstein F, et al. 2019 update of the drug resistance mutations in HIV-1. *Top Antivir Med*. 2019;27(3):111–21
- [135] - Stanford University. HIV Drug Resistance Database. 2020 [cited: 2020]; Available from: https://hivdb.stanford.edu/pages/FAQ/FAQ_answers.html
- [136] - Clavel F, Hance A. HIV Drug Resistance. *N Engl J Med*. 2004;3(2):1023–35
- [137] - Ragland DA, Whitfield TW, Lee S, et al. Elucidating the Interdependence of Drug Resistance from Combinations of Mutations. *J Chem Theory Comput*. 2017;13(11):5671–82
- [138] - Clutter D, Jordan M, Bertagnolio S, Shafer R. HIV-1 Drug Resistance and Resistance Testing. *Infect Genet Evol*. 2016;46:292–307
- [139] - Lu L, Yu F, Cai L, Debnath A, Jiang S. Development of Small-molecule HIV Entry Inhibitors Specifically Targeting gp120 or gp41. *Curr Top Med Chem* 2016;16(10):1074–90
- [140] - Bouba Y, Berno G, Fabeni L, et al. Identification of gp120 polymorphisms in HIV-1 B subtype potentially associated with resistance to fostemsavir. *Journal of Antimicrobial Chemotherapy*. 2020; 1-9
- [141] - Shafer RW. Genotypic testing for human immunodeficiency virus type 1 drug resistance. *Clin Microbiol Rev*. 2002;15(2):247–77

- [142] - Stanford University. HIV Drug Database – Resistance Notes 2021 [cited: June 2021]; Available from: <https://hivdb.stanford.edu/dr-summary/resistance-notes>
- [143] - Maldonado JO, Mansky LM. The HIV-1 reverse transcriptase A62V mutation influences replication fidelity and viral fitness in the context of multi-drug-resistant mutations. *Viruses*. 2018;10(7):1–12
- [144] - Shafer RW, Schapiro JM. HIV-1 drug resistance mutations: An updated framework for the second decade of HAART. *AIDS Rev*. 2008;10(2):67–84
- [145] - Marcelin AG, Grude M, Charpentier C, et al. Resistance to integrase inhibitors: a national study in HIV-1-infected treatment-naïve and -experienced patients. *J Antimicrob Chemother*. 2019;74(5):1368–75
- [146] - Anstett K, Brenner B, Mesplede T, Wainberg MA. HIV drug resistance against strand transfer integrase inhibitors. *Retrovirology*. 2017;14(1):1–16
- [147] - Mbisa JL, Martin SA, Cane PA. Patterns of resistance development with integrase inhibitors in HIV. *Infect Drug Resist*. 2011;4(1):65–76
- [148] - Malet I, Thierry E, Wirten M, Lebourgeois S, Subra F, Katlama C, et al. Combination of two pathways involved in raltegravir resistance confers dolutegravir resistance. *J Antimicrob Chemother*. 2015;70(10):2870–80
- [149] - Lefebvre E, Schiffer CA. Resilience to resistance of HIV-1 protease inhibitors: Profile of darunavir. *AIDS Rev*. 2008;10(3):131–42
- [150] - Menéndez-Arias L. Molecular basis of human immunodeficiency virus type 1 drug resistance: Overview and recent developments. *Antiviral Res*. 2013;98(1):93–120
- [151] - Westen GJP, Hendriks A, Wegner JK, IJzerman AP, van Vlijmen HWT, Bender A. Significantly Improved HIV Inhibitor Efficacy Prediction Employing Proteochemometric Models Generated From Antivirogram Data. *PLoS Comput Biol*. 2013;9(2): 1-18
- [152] - Wainberg MA, Brenner BG. The Impact of HIV Genetic Polymorphisms and Subtype Differences on the Occurrence of Resistance to Antiretroviral Drugs. *Mol Biol Int*. 2012;2012:1–10
- [153] - Udeze AO, Olaleye DO, Odaibo GN. Polymorphisms and drug resistance analysis of HIV-1 isolates from patients on first line antiretroviral therapy (ART) in South-eastern Nigeria. *PLoS One*. 2020;15(4):1–12

- [154] - Stanford University. Mutation Prevalence According to Subtype and Treatment. 2020. [cited: 2020]; Available from: <https://hivdb.stanford.edu/cgi-bin/MutPrevBySubtypeRx.cgi>
- [155] - World Health Organization. Regional office for Africa. 2020. [cited: June 2021]; Available from: https://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf
- [156] - World Health Organization. Mozambique – HIV Country Profile 2019 [cited: 2021]; Available from: <https://cfs.hivci.org/country-factsheet.html>
- [157] - Bártoło I, Casanovas J, Bastos R, et al. HIV-1 genetic diversity and transmitted drug resistance in health care settings in Maputo, Mozambique. *J Acquir Immune Defic Syndr*. 2009;51(3):323–31
- [158] - UNAIDS. Mozambique – Country Mozambique 2019 [cited: 2020]; Available from: <https://www.unaids.org/en/regionscountries/countries/mozambique>
- [159] - Parreira R, Piedade J, Domingues A, et al. Genetic characterization of human immunodeficiency virus type 1 from Beira, Mozambique. *Microbes Infect*. 2006;8: 2442–51
- [160] - UNAIDS. Countries 2019 [cited: 2020]; Available from: <https://www.unaids.org/en/regionscountries/countries>
- [161] - Audet CM, Burlison J, Moon TD, et al. Sociocultural and epidemiological aspects of HIV/AIDS in Mozambique. *BMC Int Health Hum Rights*. 2010;10(15): 1-10
- [162] - Nutor JJ, Duodu PA, Agbadi P, et al. Predictors of high HIV+ prevalence in Mozambique: A complex samples logistic regression modeling and spatial mapping approaches. *PLoS One*. 2020;15(6):1–21
- [163] - Vubil AS, Bila DCA, Merks H, Tanuri A, Mabunda NJ, Abreu CM, et al. Evolution of Primary HIV Drug Resistance in a Subtype C Dominated Epidemic in Mozambique. *PLoS One*. 2013;8(7): 1-7
- [164] - Lihana RW, Ssemwanga D, Abimiku A, Ndembu N. Update on HIV-1 diversity in Africa: A decade in review. *AIDS Rev*. 2012;14(2):83–100

- [165] - Abreu CM, Brindeiro PA, Martins AN, Arruda MB, Bule E, Stakteas S, et al. Genotypic and phenotypic characterization of human immunodeficiency virus type 1 isolates circulating in pregnant women from Mozambique. *Arch Virol*. 2008;153(11):2013–7
- [166] - Lahuerta M, Aparicio E, Bardaji A, Marco S, Sacarlal J, Mandomando I, et al. Rapid spread and genetic diversification of HIV type 1 subtype C in a rural area of southern Mozambique. *AIDS Res Hum Retroviruses*. 2008;24(2):327–35
- [167] - A Vubil, Vi Jani, N Mabunda, et al. Genetic Diversity and Transmitted Drug Resistance of HIV-1 Subtypes in Blood Donors Candidates in Northern Mozambique. *J AIDS Clin Res*. 2016;7(10): 1-6
- [168] - Bila DCA, Boullosa LT, Vubil AS, et al. Trends in Prevalence of HIV-1 Drug Resistance in a Public Clinic in Maputo, Mozambique. *PLoS One*. 2015;10(7): 1-12
- [169] - UNAIDS. 90-90-90: Treatment for all [cited: June 2021]; Available from: <https://www.unaids.org/en/resources/909090>
- [170] - Nkengasong J, Adje-Toure C, Weidle P. HIV antiretroviral drug resistance in Africa. *AIDS Rev*. 2004;6:4–12
- [171] - Posse M, Tirivayi N, Saha UR, Baltussen R. The effect of food assistance on adherence to antiretroviral therapy among HIV/AIDS patients in Sofala province, in Mozambique: A retrospective study. *J AIDS Clin Res*. 2013;4(3): 1-6
- [172] - Vermund SH, Blevins M, Moon TD, et al. Poor clinical outcomes for HIV infected children on antiretroviral therapy in rural Mozambique: Need for program quality improvement and community engagement. *PLoS One*. 2014;9(10): 1-9
- [173] - Vaz P, Chris Buck W, Bhatt N, et al. Compromise of Second-Line Antiretroviral Therapy Due to High Rates of Human Immunodeficiency Virus Drug Resistance in Mozambican Treatment-Experienced Children with Virologic Failure. *J Pediatric Infect Dis Soc*. 2019;9(1):6–13
- [174] - Bessong PO. Polymorphisms in HIV-1 subtype C proteases and the potential impact on protease inhibitors. *Trop Med Int Heal*. 2008;13(2):144–51
- [175] - Chang MW, Torbett BE. Accessory mutations maintain stability in drug-resistant HIV-1 protease. *J Mol Biol*. 2011;410(4):756–60

- [176] - Oliveira MF, Ramalho DB, Abreu CM, et al. Genetic diversity and naturally polymorphisms in HIV type 1 integrase isolates from Maputo, Mozambique: Implications for integrase inhibitors. *AIDS Res Hum Retroviruses*. 2012;28(12):1788–92
- [177] - Rupérez M, Pou C, Maculube S, et al. Determinants of virological failure and antiretroviral drug resistance in Mozambique. *J Antimicrob Chemother*. 2015;70(9):2639–47
- [178] - Salazar-Gonzalez J, Bailes E, T.Pham K, et al. Deciphering Human Immunodeficiency Virus Type 1 Transmission and Early Envelope Diversification by Single-Genome Amplification and Sequencing. *Journal of Virology*. 2008; 82(8): 3952-3970
- [179] - QIAGEN – QIAamp Viral RNA Mini Handbook 2020 [cited: 2020]; Available from: <https://www.qiagen.com/gb/resources/resourcedetail?id=c80685c0-4103-49ea-aa72-8989420e3018&lang=en>
- [180] - Bártoło I, Epalanga M, Bartolomeu J, et al. High Genetic Diversity of Human Immunodeficiency Virus Type 1 in Angola. *AIDS Research and Human Retroviruses*. 2005; 21(4): 306-310
- [181] - Bártoło I, Rocha C, Bartolomeu J, et al. Highly divergent subtypes and new recombinant forms prevail in the HIV/AIDS epidemic in Angola: New insights into the origins of the AIDS pandemic. *Elsevier*. 2008; 9(4): 672-682
- [182] - Bioline Meridian Bioscience – MyTaq™ One-Step RT-PCR Kit [cited: 2020]; Available from: <https://www.bioline.com/mytaq-one-step-rt-pcr-kit.html>
- [183] - Nzytech genes & enzymes – NZYTaq II DNA polymerase [cited: 2020]; Available from: <https://www.nzytech.com/products-services/molecular-biology/end-point-pcr/dna-polymerases/mb354/>
- [184] - QIAGEN – QIAGEN LongRange PCR Kit [cited: 2020]; Available from: <https://www.qiagen.com/es/resources/resourcedetail?id=e90c8e4b-f3c4-4876-a320-9890aaec1e17&lang=en>
- [185] - Macherey-Nagel - NucleoSpin Gel and PCR Clean-up, Mini kit for gel extraction and PCR clean up [cited: 2020]; Available from: <https://www.mn-net.com/media/pdf/02/1a/74/Instruction-NucleoSpin-Gel-and-PCR-Clean-up.pdf>
- [186] - Sanger F, Nicklen S, Coulson R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the Nacional Academy of Sciences*. 1977; 74(12): 5463-5467

- [187] - Calado R, Duarte J, Borrego P, et al. A Prime-Boost Immunization Strategy with Vaccinia Virus Expressing Novel gp120 Envelope Glycoprotein from a CRF02_AG Isolate Elicits Cross-Clade Tier 2 HIV-1 Neutralizing Antibodies. *Vaccines*. 2020; 8(171): 1-29
- [188] - Marcelino J, Borrego P, Rocha C, et al. Potent and Broadly Reactive HIV-2 Neutralizing Antibodies Elicited by a Vaccinia Virus Vector Prime-C2V3C3 Polypeptide Boost Immunization Strategy. *Journal of Virology*. 2010; 84(23): 12429-12436
- [189] - Borrego P, Marcelino J, Rocha C, et al. The role of the humoral immune response in the molecular evolution of the envelope C2, V3 and C3 regions in chronically HIV-2 infected patients. *Retrovirology*. 2008; 5(78): 1-12
- [190] - GE Healthcare - Protein A Mag Sepharose™ Xtra Protein G Mag Sepharose Xtra [cited: 2020]; Available from: https://www.cytivalifesciences.co.jp/catalog/pdf/magsepharose_xtra.pdf
- [191] - Darcy E, Fitzgerald J, Danaher M, et al. Purification of Antibodies Using Affinity Chromatography. *Protein Chromatography: Methods and Protocols*. 2016; 305–318
- [192] - Marcelino J, Borrego P, Nilsson C, et al. Resistance to antibody neutralization in HIV-2 infection occurs in late stage disease and is associated with X4 tropism. *AIDS*. 2012; 26: 2275-2284
- [193] - Kielkopf C, Bauer W, Urbatsch, I. Bradford Assay for Determining Protein Concentration. *Cold Spring Harbor Protocols*. 2020; 136-139
- [194] - Montefiori D, deCamp A, Hraber P, et al. Global Panel of HIV-1 Env Reference Strains for Standardized Assessments of Vaccine-Elicited Neutralizing Antibodies. *Journal of Virology*. 2014; 88(5): 2489-2507
- [195] - Srivastava K, Larsson R, Zinkus D, et al. Human Immunodeficiency Virus Type 1 NL4-3 Replication in Four T-Cell Lines: Rate and Efficiency of Entry, a Major Determinant of Permissiveness. *Journal of Virology*. 1991; 65(7): 3900-3902
- [196] - Montefiori D. Measuring HIV Neutralization in a Luciferase Reporter Gene Assay. *HIV Protocols*. 2009; 485: 395-405
- [197] - Katoh K, Rozewicki J, Yamada K. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*. 2019; 20(4): 1160-1166

- [198] - Felsenstein Joseph. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 1985; 39(4): 783-791
- [199] - Hillis D, Bull J. An empirical test for bootscanning as a method for assessing viral sequence relatedness. *System Biol*. 1993; 42: 182-192
- [200] - Kumar S, Tamura K, Nei M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. 2004; 5(2): 150-163
- [201] - Rose P, Korber B. Detecting hypermutations in viral sequences with an emphasis on G -> A hypermutation. *Bioinformatics*. 2000; 16(4):400-401
- [202] - Crooks G, Hon G, Chandonia J, et al. WebLogo: A Sequence Logo Generator. *Genome Research*. 2004; 14(6): 1188-1190
- [203] - Sierra S, Kaiser R, Thielen A, et al. Genotypic coreceptor analysis. *European Journal of Medical Research*. 2007; 12: 453-462
- [204] - Sierra S, Kaiser R, Thielen A, et al. Genotypic coreceptor analysis. *European Journal of Medical Research*. 2007; 12: 453-462
- [205] - Vandekerckhove L, Wensing A, Kaiser R, et al. European guidelines on the clinical management of HIV-1 tropism testing. *The Lancet*. 2011; 11: 394-407
- [206] - Cashin K, Gray L, Harvey K, et al. Reliable Genotypic Tropism Tests for the Major HIV-1 Subtypes. *Scientific reports*. 2015; 5(8543): 1-8
- [207] - Jensen M, Li F, Wout A, et al. Improved Coreceptor Usage Prediction and Genotypic Monitoring of R5-to-X4 Transition by Motif Analysis of Human Immunodeficiency Virus Type 1 *env* V3 Loop Sequences. *Journal of Virology*. 2003; 77(24): 13376-88
- [208] - Meloni S, Kim B, Sankalé J, et al. Distinct Human Immunodeficiency Virus Type 1 Subtype A Virus Circulating in West Africa: Sub-Subtype A3. *Journal of Virology*. 2004; 78(22): 12438-45
- [209] - Kuiken C, Korber B, Shafer R. HIV Sequence Databases. *AIDS Rev*. 2003; 5(1): 52-61
- [210] - Stanford University. Mutation Prevalence According to Subtype and Treatment. 2020 [cited: 2020]; Available from: <https://hivdb.stanford.edu/cgi-bin/MutPrevBySubtypeRx.cgi>

- [211] - Rhee S, Liu T, Kiuchi M, et al. Natural variation of HIV-1 group M integrase: Implications for a new class of antiretroviral inhibitors. *Retrovirology*. 2008; 5(74): 1-11
- [212] - Meixenberger K, Yousef K, Smith M, et al. Molecular evolution of HIV-1 integrase during the 20 years prior to the first approval of integrase inhibitors. *Virology Journal*. 2017; 14(223): 1-13
- [213] - Rogers L, Obasa A, Jacobs G, et al. Structural Implications of Genotypic Variations in HIV-1 Integrase From Diverse Subtypes. *Frontiers in Microbiology*. 2018; 9(1754): 1-9
- [214] - Lumngwena E, Shuping L, Bernitz N, et al. HIV-1 subtype C Envelope function becomes less sensitive to N-glycosylation deletion during disease progression. *BMC Res Notes*. 2019; 12(340): 1-6
- [215] - Wang W, Prochnow C, Truong C, et al. A systematic study of the N-glycosylation sites of HIV-1 envelope protein on infectivity and antibody-mediated neutralization. *Retrovirology*. 2013; 10(14): 1-14
- [216] - Ferreira R, Grant O, Moyou T, et al. Structural Rearrangements Maintain the Glycan Shield of an HIV-1 Envelope Trimer After the Loss of a Glycan. *Scientific Reports*. 2018; 8(15031): 1-15
- [217] - Guo J, Yan Y, Zhang J, et al. Co-receptor tropism and genetic characteristics of the V3 regions in variants of antiretroviral-naive HIV-1 infected subjects. *Epidemiology and Infection*. 2019; 147(181): 1-7
- [218] - Jiang X, Burke V, Totrov M, et al. Conserved structural elements in the V3 crown of HIV-1 gp120. *Nature*. 2010; 17(8): 955-962
- [219] - Jiang X, Feyertag F, Meehan C, et al. Characterizing the Diverse Mutational Pathways Associated with R5- Tropic Maraviroc Resistance: HIV-1 That Uses the Drug-Bound CCR5 Coreceptor. *Journal of Virology*. 2015; 89(22): 11457-72
- [220] - Sede M, Moretti F, Laufer N, et al. HIV-1 Tropism Dynamics and Phylogenetic Analysis from Longitudinal Ultra-Deep Sequencing Data of CCR5- and CXCR4-Using Variants. *PLOS One*. 2014; 9(7): 1-14
- [221] - Coetzer M, Cilliers T, Ping L, et al. Genetic characteristics of the V3 region associated with CXCR4 usage in HIV-1 subtype C isolates. *Elsevier*. 2006; 356: 95-105

- [222] - Matume N, Tebit D, Bessong P. HIV-1 subtype C predicted co-receptor tropism in Africa: an individual sequence level meta-analysis. *AIDS Research and Therapy*. 2020; 17(5): 1-16
- [223] - Hunt G, Ledwaba J, Kalimashe M, et al. Provincial and national prevalence estimates of transmitted HIV-1 drug resistance in South Africa measured using two WHO-recommended methods. *Antiviral Therapy*. 2019; 1359-6535
- [224] - Penrose K, Brumme C, Scoulos-Hanson M, et al. Frequent cross-resistance to rilpivirine among subtype C HIV-1 from first-line antiretroviral therapy failures in South Africa. *Antiviral Chemistry and Chemotherapy*; 2018; 26: 1-8
- [225] - Choisy M, Woelk C, Guégan J, et al. Comparative Study of Adaptive Molecular Evolution in Different Human Immunodeficiency Virus Groups and Subtypes. *Journal of Virology*. 2004; 78(4): 1962-1970
- [226] - Berndsen Z, Chakraborty S, Wang X, et al. Visualization of the HIV-1 Env glycan shield across scales. *PNAS*. 2020; 117(45): 28014-25
- [227] - Gordon M, Oliveira T, Bishop K, et al. Molecular Characteristics of Human Immunodeficiency Virus Type 1 Subtype C Viruses from KwaZulu-Natal, South Africa: Implications for Vaccine and Antiretroviral Control Strategies. *Journal of Virology*. 2003; 77(4): 2587-2599
- [228] - Gatner M, Rocha M, Churchill M, et al. Understanding the mechanisms driving the spread of subtype C HIV-1. *Elsevier*. 2020; 53: 2352-3964
- [229] - Novitsky V, Smith U, Gilbert P, et al. Human Immunodeficiency Virus Type 1 Subtype C Molecular Phylogeny: Consensus Sequence for an AIDS Vaccine Design?. *Journal of Virology*. 2002; 76(11): 5435-5451
- [230] - Macdonald V, Mbuagbaw L, Jordan M, et al. Prevalence of pretreatment HIV drug resistance in key populations: a systematic review and meta-analysis. *Journal of the international AIDS society*. 2020; 23: 1-11
- [231] - World Health Organization (WHO) – HIV drug resistance report 2019. [cited 2021 April]; Available from: <https://www.who.int/hiv/pub/drugresistance/hivdr-report-2019/en/>
- [232] - Onywera H, Maman D, Inzaule S, et al. Surveillance of HIV-1 *pol* transmitted drug resistance in acutely and recently infected antiretroviral drug-naïve persons in rural western Kenya. *PLOS One*. 2017; 1-14

- [233] - Theys K, Abecasis A, Vandamme A. HIV-1 drug resistance: where do polymorphisms fit in?. *Future Microbiology*. 2013; 8(3): 303-306
- [234] - Handema R, Terunuma H, Kasolo F. et al. Prevalence of Drug-Resistance-Associated Mutations in Antiretroviral Drug-Naive Zambians Infected with Subtype C HIV-1. *AIDS Research and Human Retroviruses*. 2003; 19(2): 151–160
- [235] - Pimentel V, Pingarilho M, Alves D. et al. Molecular Epidemiology of HIV-1 Infected Migrants Followed up in Portugal: Trends between 2001–2017. *Viruses*. 2020; 12(3): 1-20
- [236] - Ali A, Bandaranayake R, Cai Y, et al. Molecular Basis for Drug Resistance in HIV-1 Protease. *Viruses*. 2010; 2: 2509-2535
- [237] - Anstett K, Fusco R, Cutillas V, et al. Dolutegravir-Selected HIV-1 Containing the N155H and R263K Resistance Substitutions Does Not Acquire Additional Compensatory Mutations under Drug Pressure That Lead to Higher-Level Resistance and Increased Replicative Capacity. *Journal of Virology*. 2015; 89(20): 10482-88
- [238] - Wei Q, Hargett A, Knoppova B, et al. Glycan Positioning Impacts HIV-1 Env Glycan-Shield Density, Function, and Recognition by Antibodies. *Cell Press*. 2020; 23(101711): 1-18
- [239] - Ping L, Joseph S, Anderson J, et al. Comparison of Viral Env Proteins from Acute and Chronic Infections with Subtype C Human Immunodeficiency Virus Type 1 Identifies Differences in Glycosylation and CCR5 Utilization and Suggests a New Strategy for Immunogen Design. *Journal of Virology*. 2013; 87(13):7218-32
- [240] - Liu Y, Curlin M, Diem K. et al. Env length and N-linked glycosylation following transmission of Human Immunodeficiency Virus Type 1 subtype B viruses. *Virology*. 2008; 374(2): 229-233
- [241] - Yu X, Yuan L, Huang Y. et al. Susceptibility of HIV-1 Subtypes B9, CRF07_BC and CRF01_AE that Are Predominantly Circulating in China to HIV-1 Entry Inhibitors. *PLoS ONE*. 2011; 6(3): 1-8
- [242] - Santoro M, Perno C. HIV-1 Genetic Variability and Clinical Implications. *ISRN Microbiology*. 2013; 1-20
- [243] - Guo H, Liu C, Liu B, et al. Analysis of primary resistance mutations to HIV-1 entry inhibitors in therapy naive subtype C HIV-1 infected mother– infant pairs from Zambia. *J Clin Virol*. 2013; 58(1): 233-239

- [244] - Araújo L, Junqueira D, Medeiros R, et al. Naturally occurring resistance mutations to HIV-1 entry inhibitors in subtypes B, C, and CRF31 BC. *Elsevier*. 2012; 54: 6-10
- [245] - Nontokozo M, Tebit D, Gray L. et al. Next generation sequencing reveals a high frequency of CXCR4 utilizing viruses in HIV-1 chronically infected drug experienced individuals in South Africa. *J Clin Virol*. 2018; 103: 81-87
- [246] - Makhdoomi M, Khan L, Kumar S, et al. Evolution of cross-neutralizing antibodies and mapping epitope specificity in plasma of chronic HIV-1-infected antiretroviral therapy-naïve children from India. *Journal of General Virology*. 2017; 98:1879-1891
- [247] - Lynch R, Shen T, Gnanakaran S, et al. Appreciating HIV Type 1 Diversity: Subtype Differences in Env. *AIDS Research and human retroviruses*. 2009; 25(3): 237-248
- [248] - Gray E, Moore P, Choge I, et al. Neutralizing Antibody Responses in Acute Human Immunodeficiency Virus Type 1 Subtype C Infection. *Journal of Virology*. 2007; 81(12): 6187-6196
- [249] - Stephenson E, Wagh K, Korber B. et al. Vaccines and Broadly Neutralizing
- [250] - Cohen Z, Caskey M. Broadly neutralizing antibodies for treatment and prevention of HIV-1 infection. *Current Opinion in HIV and AIDS*. 2018; 13(4): 366–373

6. Anexos

Anexo 1

Tabela A1 - Antirretrovirais em uso clínico para o tratamento de pacientes seropositivos para VIH-1.

Classes de antirretrovirais	Antirretroviral	Ano de aprovação pela FDA
Inibidores da RT análogos a nucleósidos/nucleótido (NRTI)	Zidovudina (AZT)	1987
	Didanosina (ddi)	1991
	Zalcitabina (ddC)	1992
	Estavudina (d4T)	1994
	Lamivudina (3TC)	1995
	Abacavir (ABC)	1998
	Tenofovir (TDF)	2001
	Emtricitabina (FTC)	2003
Inibidores da RT não análogos a nucleósidos (NNRTI)	Nevirapina (NVP)	1996
	Delavirdina (DLV)	1997
	Efavirenz (EFV)	1998
	Etravirina (ETR)	2008
	Rilpivirina (RPV)	2011
	Doravirina (DOR)	2018
Inibidores de protease (PI)	Saquinavir (SQV)	1995
	Indinavir (IDV)	1996
	Ritonavir (RTV)	1996
	Nelfinavir (NFV)	1997
	Amprenavir (APV)	1999
	Lopinavir (LPV)	2000
	Atazanavir (ATV)	2003
	Fosamprenavir (FPV)	2003
	Tipranavir (TPV)	2005
	Darunavir (DRV)	2006
Inibidores de entrada		
Inibidor de fusão (FI)	Enfuvirtide (T-20)	2003
Antagonista do co-recetor (CA)	Maraviroc (MVC)	2007
Inibidor de ligação ao recetor CD4	Fostemsavir (FTR)	2020
Inibidores de integrase (INSTI)	Raltegravir (RAL)	2007
	Dolutegravir (DTG)	2013
	Elvitegravir (EVG)	2014
	Bictegravir (BIC)	2018

Anexo 2

Tabela A2 - Composição das soluções utilizadas ao longo do trabalho experimental.

Soluções	Composição
Tampão SB 50x	120gramas de Ácido bórico 20gramas de Hidróxido de sódio 1L de água bidestilada pH = 8.00

Anexo 3

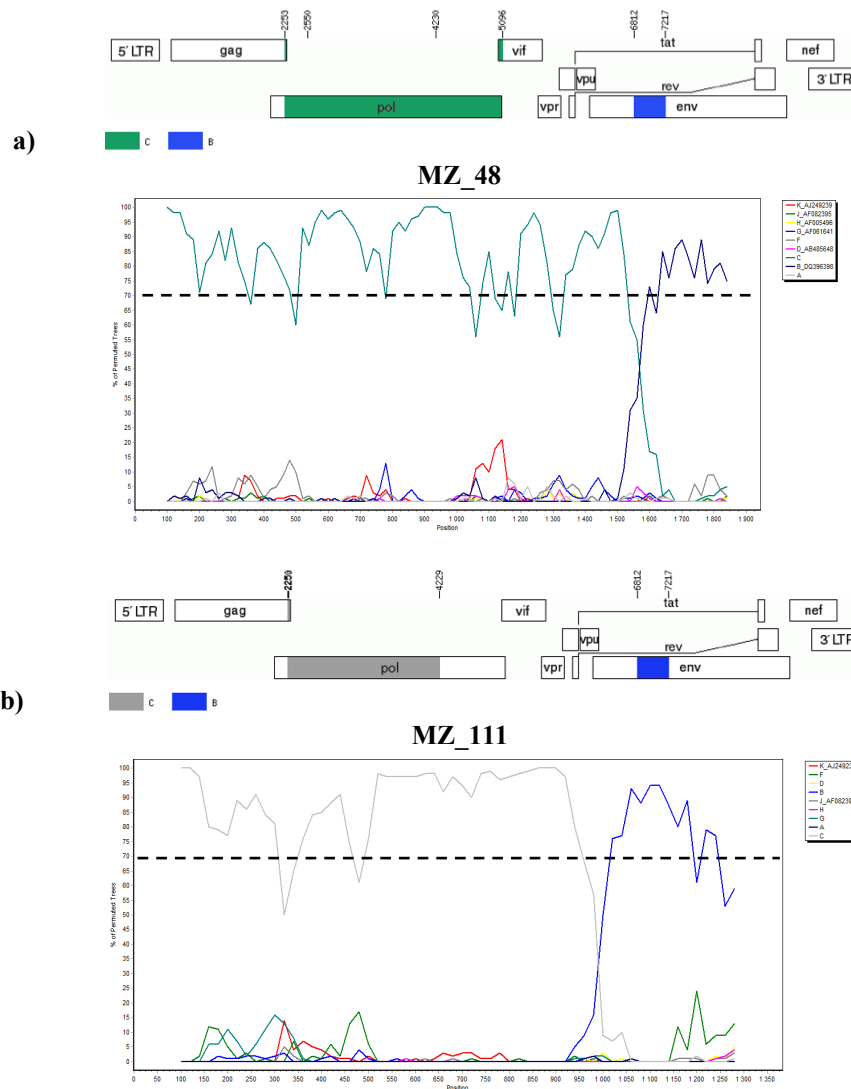


Figura A1 – Gráficos *bootscanning* de MZ_48 e MZ_111. Análise das estruturas mosaico C/C/C/B (PR, RT, IN e C2V3C3) dos isolados virais MZ_48 (a) e C/C/B (PR, RT e C2V3C3) do isolado viral MZ_111 (b) por *bootscanning* (modelo idêntico ao utilizado na construção da árvore filogenética). No eixo das ordenadas encontram-se os valores de *bootstrap* que suportam o agrupamento onde estão inseridas as sequências em estudo com as sequências de referência. O eixo das abcissas representa as posições referentes a cada região do genoma em análise. A linha horizontal assinala o *cut-off* de 70%.

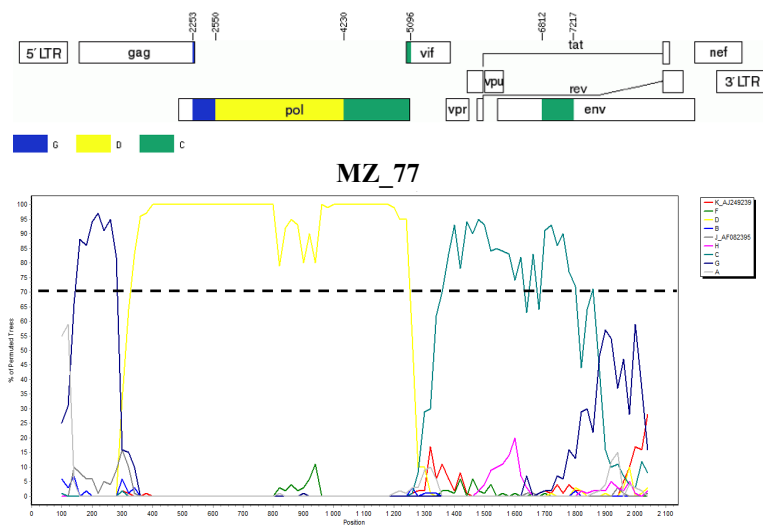
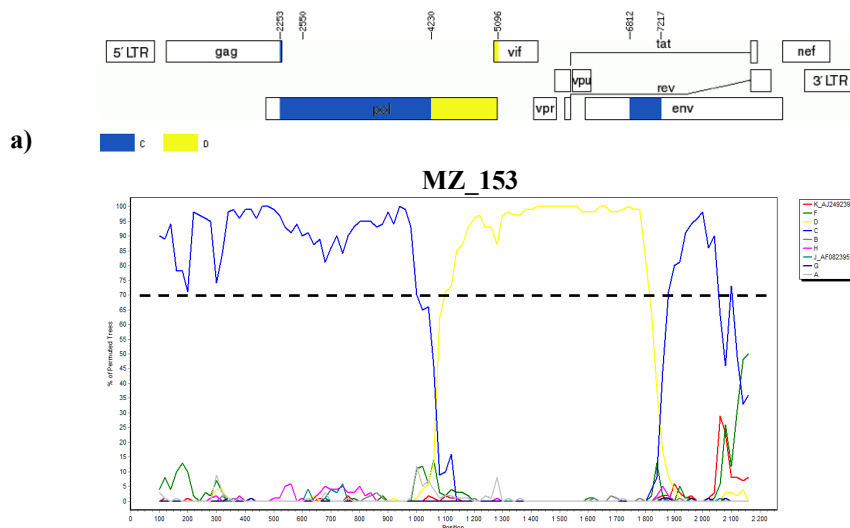


Figura A2 – Gráfico *bootscanning* de MZ_77. Análise das estruturas mosaico CRF02_AG/D/C/C (PR, RT, IN e C2V3C3) do isolado viral MZ_77 por *bootscanning* (modelo idêntico ao utilizado na construção da árvore filogenética). No eixo das ordenadas encontram-se os valores de *bootstrap* que suportam o agrupamento onde estão inseridas as sequências em estudo com as sequências de referência. O eixo das abcissas representa as posições referentes a cada região do genoma em análise. A linha horizontal assinala o *cut-off* de 70%.



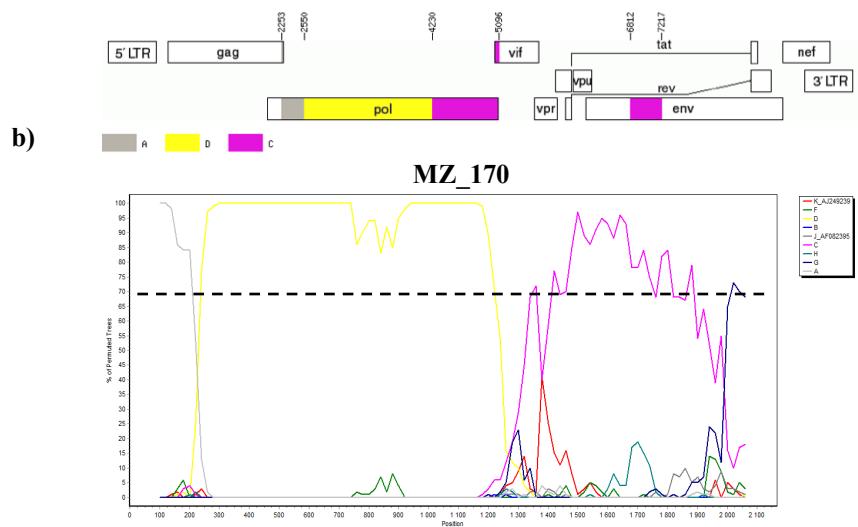
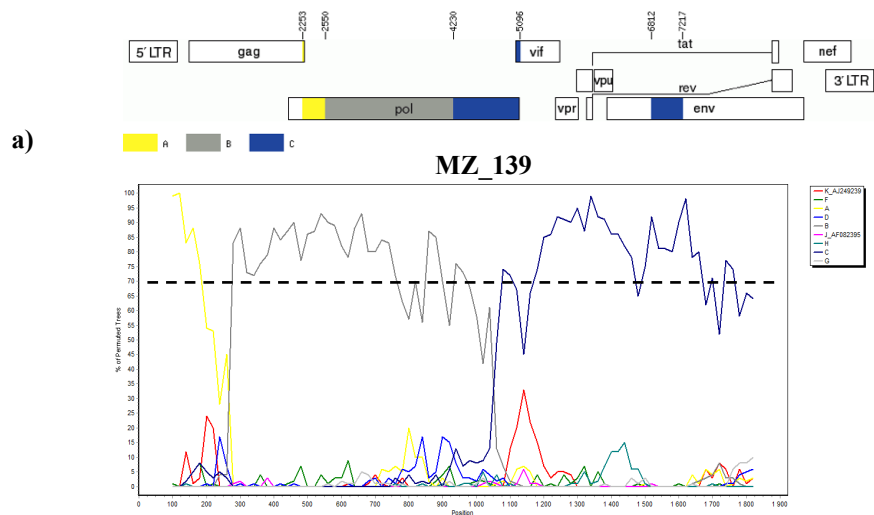


Figura A3 – Gráficos *bootscanning* de MZ_153 e MZ_170. Análise das estruturas mosaico C/C/D/C e A/D/C/C (PR, RT, IN e C2V3C3) dos isolados virais MZ_153(a) e MZ_170(b), respetivamente, por *bootscanning* (modelo idêntico ao utilizado na construção da árvore filogenética). No eixo das ordenadas encontram-se os valores de *bootstrap* que suportam o agrupamento onde estão inseridas as sequências em estudo com as sequências de referência. O eixo das abcissas representa as posições referentes a cada região do genoma em análise. A linha horizontal assinala o *cut-off* de 70%.



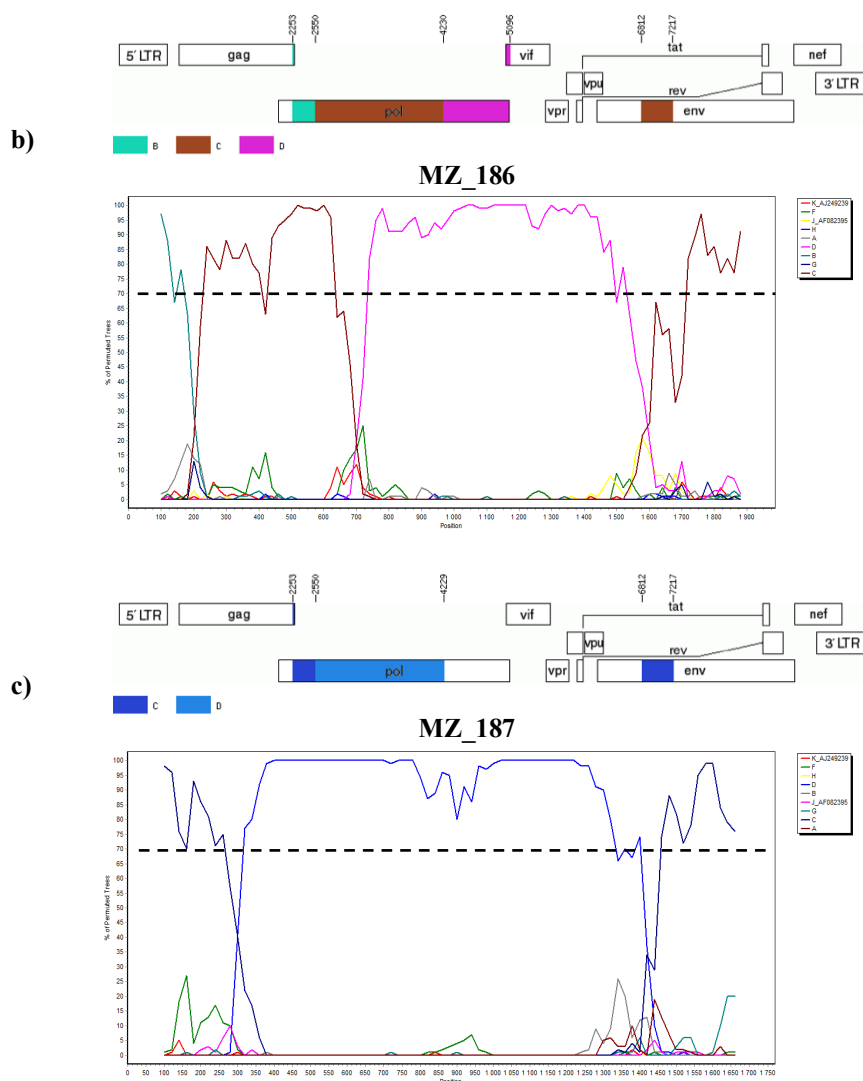


Figura A4 – Gráficos *bootscanning* de MZ_139, MZ_186 e MZ_187. Análise da estrutura mosaico A/B/C/C, B/C/D/C (PR, RT, IN e C2V3C3) dos isolados virais MZ_139(a) e MZ_186(b) e a estrutura mosaico C/D/C (PR, RT e C2V3C3) de MZ_187(c), por *bootscanning* (modelo idêntico ao utilizado na construção da árvore filogenética). No eixo das ordenadas encontram-se os valores de *bootstrap* que suportam o agrupamento onde estão inseridas as sequências em estudo com as sequências de referência. O eixo das abscissas representa as posições referentes a cada região do genoma em análise. A linha horizontal assinala o *cut-off* de 70%.

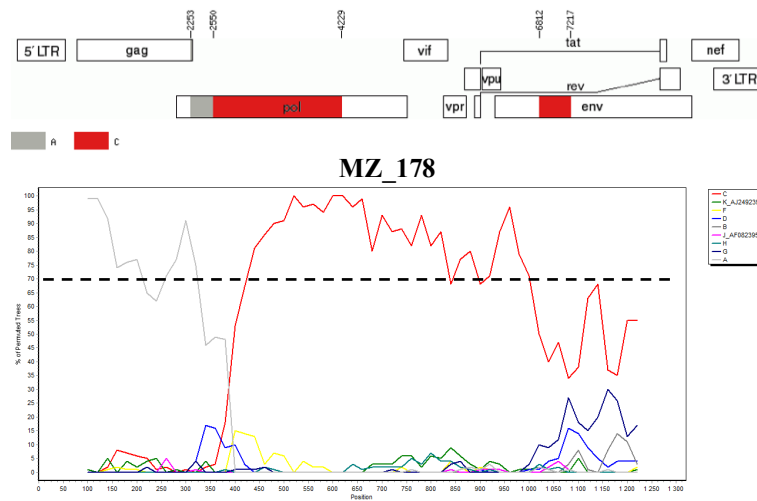


Figura A5 – Gráficos *bootscanning* de MZ_178. Análise da estrutura mosaico A/C/C (PR, RT e C2V3C3) do isolado viral MZ_178 por *bootscanning* (modelo idêntico ao utilizado na construção da árvore filogenética). No eixo das ordenadas encontram-se os valores de *bootstrap* que suportam o agrupamento onde estão inseridas as sequências em estudo com as sequências de referência. O eixo das abcissas representa as posições referentes a cada região do genoma em análise. A linha horizontal assinala o *cut-off* de 70%.

Anexo 4

Tabela A2 - Polimorfismos genéticos identificados nas 26 sequências em estudo não associados a resistência aos inibidores da transcriptase reversa. A tabela apresenta a análise comparativa entre as frequências (%) de cada polimorfismo nos isolados virais em estudo e as encontradas nas sequências de referência da base de dados de Stanford.

Mutação	Isolados MZ (%)			Isolados de Ref. (%)		
	B	C	D	B	C	D
K20R	---	9	---	12	20	8
V35T	---	86	100	10	91	88
E36A	---	32	---	0,2	67	---
T39E	---	82	---	0,2	79	---
E40D	---	9	---	1	6	31
S48T	---	91	---	---	96	3
E53D	---	9	---	1	1	1
V60I	---	45	100	7	40	95
R83K	S	---	---	4	---	---
V118I	---	5	---	48	12	13
D121H	---	5	100	2	3	7
K122E	---	86	100	42	87	80
D123E	---	9	---	26	3	13
D123N	---	9	---	5	17	5
I135R	---	9	---	3	2	5
I135T	---	9	100	24	8	28
T139A	---	9	---	---	2	---
T139K	---	23	---	---	1	1
I142V	S	---	---	3	5	11
T165I	---	18	---	2	13	11
K173A	---	41	---	0,3	58	3
K173N	---	---	100	0,2	0,2	1
K173T	---	27	---	1	25	1
K173V	---	18	---	---	1	---
Q174K	---	64	100	6	46	3
D177E	---	68	100	15	75	92
I178M	---	9	---	8	5	35
V179I	---	5	---	3	2	9

G196E	---	36	---	27	8	3
T200A	---	86	100	13	93	15
I202V	---	18	---	10	15	15
Q207D	---	18	---	2	6	4
Q207E	---	68	100	18	72	69
R211K	S	91	---	27	72	88
F214L	---	18	---	51	31	24
K238R	---	5	---	1	1	---
I244V	---	5	100	0,4	0,2	5
V245Q	---	82	---	1	76	19
E248D	---	18	---	2	12	16
D250E	---	32	100	3	23	86
A272P	---	77	---	36	77	64
K275R	---	9	---	1	7	2
K277R	---	64	---	26	53	54
Q278H	---	9	---	---	8	6
T286A	---	45	100	13	70	50
E291D	---	77	---	2	94	2
V292I	---	77	100	5	92	9
I293V	---	77	100	76	96	73
Y318N	---	5	---	---	---	---
G333E	---	18	---	2	1	3
Q334H	---	23	---	2	26	---
Q334N	---	23	---	6	22	---
G335D	---	50	67	---	79	77
Q336H	S	41	---	---	---	---
Y339F	---	14	33	---	---	---

Anexo 5.A

Tabela A3 - Polimorfismos genéticos identificados nas 26 sequências em estudo não associados a resistência aos inibidores da transcriptase reversa. A tabela apresenta a análise comparativa entre as sequências do subtipo C e não-C com base na aplicação do teste exato de Fisher ou quando aplicável o teste do Chi-Quadrado (χ^2) ($p < 0,05$).

Mutação	Subtipo C (n=22)	Subtipo não-C (n=4)	<i>p</i>
K20R	2	0	>0,9999
V35T	19	3	0,5107
E36A	7	0	0,5463
T39E	18	0	0,0047
E40D	2	0	>0,9999
S48T	20	0	0,0010
E53D	2	0	>0,9999
V60I	10	3	0,5930
R83K	0	1	0,1538
V118I	1	0	>0,9999
D121H	1	3	0,0060
K122E	19	3	0,5107
D123E	2	0	>0,9999
D123N	2	0	>0,9999
I135R	2	0	>0,9999
I135T	2	3	0,0144
T139A	2	0	>0,9999
T139K	5	0	0,5552
I142V	0	1	0,1538
T165I	4	0	>0,9999
K173A	9	0	0,2631
K173N	0	3	0,0015
K173T	6	0	0,5425
K173V	4	0	>0,9999
Q174K	14	3	>0,9999
D177E	15	3	>0,9999
I178M	2	0	>0,9999
V179I	1	0	>0,9999
G196E	8	0	0,2768
T200A	19	3	0,5107

I202V	4	0	>0,9999
Q207D	4	0	>0,9999
Q207E	15	3	>0,9999
R211K	20	1	0,0144
F214L	4	0	>0,9999
K238R	1	0	>0,9999
I244V	1	3	0,0060
V245Q	18	0	0,0047
E248D	4	0	>0,9999
D250E	7	3	0,2642
A272P	17	0	0,0084
K275R	2	0	>0,9999
K277R	14	0	0,0331
Q278H	2	0	>0,9999
T286A	10	3	0,5930
E291D	17	0	0,0084
V292I	17	3	>0,9999
I293V	17	3	>0,9999
Y318N	1	0	>0,9999
G333E	4	0	>0,9999
Q334H	5	0	0,5552
Q334N	5	0	0,5552
G335D	11	2	>0,9999
Q336H	9	1	>0,9999
Y339F	3	1	0,5107

Anexo 5.B

Tabela A4 - Polimorfismos genéticos identificados nas 26 sequências em estudo não associados a resistência aos inibidores da transcriptase reversa. A tabela apresenta a análise comparativa entre as sequências do subtipo D e não-D com base na aplicação do teste exato de Fisher ou quando aplicável o teste do Chi-Quadrado (χ^2) ($p < 0,05$).

Mutação	Subtipo D (n=3)	Subtipo não-D (n=23)	<i>p</i>
K20R	0	2	>0,9999
V35T	3	19	>0,9999
E36A	0	7	0,5396
T39E	0	18	0,0215
E40D	0	2	>0,9999
S48T	0	20	0,2389
E53D	0	2	>0,9999
V60I	3	10	0,2200
R83K	0	1	>0,9999
V118I	0	1	>0,9999
D121H	3	0	0,0004
K122E	3	19	>0,9999
D123E	0	2	>0,9999
D123N	0	2	>0,9999
I135R	0	2	>0,9999
I135T	3	2	0,0038
T139A	0	2	>0,9999
T139K	0	5	>0,9999
I142V	0	1	>0,9999
T165I	0	4	>0,9999
K173A	0	9	0,5292
K173N	3	0	0,0004
K173T	0	6	>0,9999
K173V	0	4	>0,9999
Q174K	3	14	0,5292
D177E	3	15	0,5292
I178M	0	2	>0,9999
V179I	0	1	>0,9999
G196E	0	8	0,5292
T200A	3	19	>0,9999

I202V	0	4	>0,9999
Q207D	0	4	>0,9999
Q207E	3	15	0,5292
R211K	1	21	0,0523
F214L	0	4	>0,9999
K238R	0	1	>0,9999
I244V	3	1	0,0015
V245Q	0	18	0,0215
E248D	0	4	>0,9999
D250E	3	7	0,0462
A272P	0	17	0,0323
K275R	0	2	>0,9999
K277R	0	14	0,0846
Q278H	0	2	>0,9999
T286A	3	10	0,2200
E291D	0	17	0,0323
V292I	3	17	>0,9999
I293V	3	17	>0,9999
Y318N	0	1	>0,9999
G333E	0	4	>0,9999
Q334H	0	5	>0,9999
Q334N	0	5	>0,9999
G335D	2	11	>0,9999
Q336H	1	10	>0,9999
Y339F	1	3	0,4077

Anexo 6

Tabela A5 - Polimorfismos genéticos identificados nas 34 sequências em estudo não associados a resistência aos inibidores da protease. A tabela apresenta a análise comparativa entre as frequências (%) de cada polimorfismo nos isolados virais em estudo e as encontradas nas sequências de referência da base de dados de Stanford. *S representa que está presente na única sequência do subtipo ou CRF em análise.

Mutação	Isolados MZ (%)				Isolados de Ref. (%)			
	A	B	C	CRF02_AG	A	B	C	CRF02_AG
L5F	---	---	3	---	0,2	0,1	1	0,3
L10I	100	---	---	---	100	100	100	100
T12S	---	---	72	---	1	3	40	1
I13V	100	---	3	S	92	25	11	95
K14R	---	---	7	S	35	14	13	38
I15V	---	---	86	---	15	20	69	18
G16E	---	---	3	---	40	9	13	39
G17D	---	---	3	---	1	1	1	---
L19I	---	---	55	---	5	9	49	2
L19T	---	---	34	---	0,2	1	11	1
L19V	---	---	3	---	1	3	11	0,1
K20R	---	---	38	---	13	3	11	5
L23F	---	---	3	---	---	---	1	---
A28V	100	---	---	---	0,1	---	---	---
E35D	100	S	48	S	91	28	21	23
M36I	100	S	90	S	99	22	77	98
N37D	---	---	7	---	18	9	5	17
N37E	---	---	---	S	2	3	1	6
N37K	---	---	7	---	0,4	0,3	19	0,1
N37S	---	---	10	---	1	19	6	7
R41K	100	S	83	---	92	36	70	92
R41I	---	---	3	---	---	0,1	1	---
R41N	---	---	7	---	0,3	0,1	19	0,3
R57K	---	S	7	---	52	15	8	3
D60E	---	---	38	---	8	11	19	1
Q61E	---	---	24	---	4	5	9	1
Q61N	---	---	17	---	0,4	1	1	0,3
I62V	---	---	7	---	8	36	8	1

L63A	---	---	17	---	3	8	4	1
L63P	---	S	45	S	12	36	16	13
L63T	---	---	14	---	15	18	24	1
L63V	---	---	3	---	9	2	13	1
I64L	---	---	7	---	1	6	2	6
E65D	---	---	7	---	5	6	4	1
H69K	100	---	100	S	74	1	81	98
K70Q	---	---	17	---	0,2	1	1	---
I72T	---	---	7	---	2	11	2	2
T74S	---	---	10	---	1	0,3	9	1
V77I	---	---	10	---	7	32	5	2
V82I	---	---	14	---	1	2	4	1
R87K	---	---	3	---	0,1	0,2	0,3	---
L89I	---	---	7	---	1	0,1	5	3
L89M	100	---	48	S	97	3	80	96
I93L	---	S	90	---	9	70	92	1
F99L	---	---	---	S	2	1	1	1

Anexo 7.A

Tabela A6 - Polimorfismos genéticos identificados nas 34 sequências em estudo não associados a resistência aos inibidores da protease. A tabela apresenta a análise comparativa entre as sequências do subtipo C e não-C com base na aplicação do teste exato de Fisher ou quando aplicável o teste do Chi-Quadrado (χ^2) ($p < 0,05$).

Mutação	Subtipo C (n=29)	Subtipo não-C (n=5)	<i>p</i>
L5F	1	0	>0,9999
L10I	0	3	0,0017
T12S	21	0	0,0046
I13V	1	4	0,0075
K14R	2	1	0,3894
I15V	25	0	0,0005
G16E	1	0	>0,9999
G17D	1	0	>0,9999
L19I	16	0	0,0465
L19T	10	0	0,2908
L19V	1	0	>0,9999
K20R	11	0	0,1499
L23F	1	0	>0,9999
A28V	0	3	0,0072
E35D	14	5	0,0526
M36I	26	5	>0,9999
N37D	2	0	>0,9999
N37E	0	1	0,1471
N37K	2	0	>0,9999
N37S	3	0	>0,9999
R41K	24	4	0,0362
R41I	1	0	>0,9999
R41N	2	0	>0,9999
R57K	2	1	0,3894
D60E	11	0	0,1499
Q61E	7	0	0,5585
Q61N	5	0	>0,9999
I62V	2	0	>0,9999
L63A	5	0	>0,9999

L63P	13	2	>0,9999
L63T	4	0	>0,9999
L63V	1	0	>0,9999
I64L	2	0	>0,9999
E65D	2	0	>0,9999
H69K	29	4	0,1471
K70Q	5	0	>0,9999
I72T	2	0	>0,9999

Anexo 7.B

Tabela A7 - Polimorfismos genéticos identificados nas 34 sequências em estudo não associados a resistência aos inibidores da protease. A tabela apresenta a análise comparativa entre as sequências do subtipo A e não-A com base na aplicação do teste exato de Fisher ou quando aplicável o teste do Chi-Quadrado (χ^2) ($p < 0,05$).

Mutação	Subtipo A (n=3)	Subtipo não-A (n=31)	<i>p</i>
L5F	0	1	>0,9999
L10I	3	0	0,0002
T12S	0	21	0,0478
I13V	3	2	0,0017
K14R	0	3	>0,9999
I15V	0	25	0,0140
G16E	0	1	>0,9999
G17D	0	1	>0,9999
L19I	0	16	0,2299
L19T	0	10	0,5388
L19V	0	1	>0,9999
K20R	0	11	0,5349
L23F	0	1	>0,9999
A28V	3	0	0,0002
E35D	3	16	0,2380
M36I	3	28	>0,9999
N37D	0	2	>0,9999
N37E	0	1	>0,9999
N37K	0	2	>0,9999
N37S	0	3	>0,9999
R41K	3	25	>0,9999
R41I	0	1	>0,9999
R41N	0	2	>0,9999
R57K	0	3	>0,9999
D60E	0	11	0,5349
Q61E	0	7	>0,9999
Q61N	0	5	>0,9999
I62V	0	2	>0,9999
L63A	0	5	>0,9999

L63P	0	15	0,2380
L63T	0	4	>0,9999
L63V	0	1	>0,9999
I64L	0	2	>0,9999
E65D	0	2	>0,9999
H69K	3	30	>0,9999
K70Q	0	5	>0,9999
I72T	0	2	>0,9999

Anexo 8

Tabela A7 - Polimorfismos genéticos identificados nas 22 sequências em estudo não associados a resistência aos inibidores da integrase. A tabela apresenta a análise comparativa entre as frequências (%) de cada polimorfismo nos isolados virais em estudo e as encontradas nas sequências de referência da base de dados de Stanford. *S representa que está presente na única sequência do subtipo em análise.

Mutação	Isolados MZ (%)		Isolados de Ref. (%)	
	C	D	C	D
E10D	5	---	33	63
E11D	10	---	15	17
K14R	20	---	33	17
S17N	30	S	23	13
S24N	20	---	8	---
D25E	75	---	71	---
L28I	15	---	1	4
V31I	70	---	76	4
V32I	---	S	2	42
V37I	5	---	6	---
S39C	15	---	2	4
L45Q	10	---	9	---
M50I	45	---	43	---
M50T	25	---	10	---
V54I	20	---	5	---
I60V	5	---	4	---
I72V	50	S	28	58
L74I	5	---	5	---
I84M	50	---	25	---
A91G	10	---	2	4
A91T	5	---	5	---
E96D	25	---	1	---
F100Y	70	---	56	---
L101I	90	---	94	29
K111R	10	---	1	---
T112V	95	S	91	79
I113V	---	S	1	88

S119P	25	---	7	---
T124A	65	S	62	75
T124N	20	---	22	13
T125A	95	---	86	33
K127R	20	---	2	---
K136Q	95	---	73	8
V150A	20	---	1	---
D167E	15	---	19	8
K173R	5	---	7	---
K188R	25	---	14	---
V201I	100	S	99	100
T206S	25	---	9	4
I208M	20	---	1	13
K211R	5	S	10	54
K215N	10	---	9	---
T218I	45	S	49	21
T218L	15	---	11	8
K219N	5	---	2	---
K219Q	5	---	10	---
I220L	5	---	2	4
I220V	5	---	6	---
Q221H	5	---	1	---
N222K	5	---	2	---
L234I	90	S	85	100
L234V	10	---	2	---
D253E	10	---	3	25
S255G	15	---	1	---
D256E	5	S	2	13
A265V	5	---	43	26
I268L	5	---	2	---
R269K	45	---	57	9
D278A	75	---	93	---
S283G	85	---	88	9
Q285R	20	---	---	---
D286E	15	---	---	---
D288G	15	S	---	---

Anexo 9

Tabela A8 - Polimorfismos genéticos identificados nas 22 sequências em estudo não associados a resistência aos inibidores da integrase. A tabela apresenta a análise comparativa entre as sequências do subtipo C e não-C com base na aplicação do teste exato de Fisher ou quando aplicável o teste do Chi-Quadrado (χ^2) ($p < 0,05$).

Mutação	Subtipo C (n=20)	Subtipo não-C (n=2)	<i>p</i>
E10D	1	0	>0,9999
E11D	2	0	>0,9999
K14R	4	0	>0,9999
S17N	6	2	0,1212
S24N	4	0	>0,9999
D25E	15	0	0,0909
L28I	3	0	>0,9999
V31I	14	0	0,1212
V32I	0	2	0,0043
V37I	1	0	>0,9999
S39C	3	0	>0,9999
L45Q	2	0	>0,9999
M50I	9	0	0,4935
M50T	5	0	>0,9999
V54I	4	0	>0,9999
I60V	1	0	>0,9999
I72V	10	2	0,4805
L74I	1	0	>0,9999
I84M	10	0	0,4805
A91G	2	0	>0,9999
A91T	1	0	>0,9999
E96D	5	0	>0,9999
F100Y	14	0	0,1212
L101I	18	0	0,0260
K111R	2	0	>0,9999
T112V	19	1	0,1775
I113V	0	2	0,0043
S119P	5	0	>0,9999
T124A	13	2	>0,9999

T124N	4	0	>0,9999
T125A	19	0	0,0130
K127R	4	0	>0,9999
K136Q	19	0	0,0130
V150A	4	0	>0,9999
D167E	3	0	>0,9999
K173R	1	0	>0,9999
K188R	5	0	>0,9999
V201I	20	2	>0,9999
T206S	5	0	>0,9999
I208M	4	0	>0,9999
K211R	1	2	0,0130
K215N	2	0	>0,9999
T218I	9	2	0,4762
T218L	3	0	>0,9999
K219N	1	0	>0,9999
K219Q	1	0	>0,9999
I220L	1	0	>0,9999
I220V	1	0	>0,9999
Q221H	1	0	>0,9999
N222K	1	0	>0,9999
L234I	18	1	0,2597
L234V	2	0	>0,9999
D253E	2	0	>0,9999
S255G	3	0	>0,9999
D256E	1	2	0,0130
A265V	1	0	>0,9999
I268L	1	0	>0,9999
R269K	9	0	0,4935
D278A	15	0	0,0909
S283G	17	0	0,0433
Q285R	4	0	>0,9999
D286E	3	0	>0,9999
D288G	3	1	0,3377

Anexo 10

N276 N301 N332
 ↓ ↓ ↓
 HXB2_ : AEEEVVIRSVNFTD^{**}NAKTIIVQLN^{*}TSVEIN^{*}CTRPNN^{*}TRKRIRIQ^{***}RGPGRA^{*}FTVTIGK-IGNMRQAH^{*}CNISRAKWNTLTKQIASKLREQFGNNKTIIFKQSSGGDFEIVTHSF^{*}NCGGEFFY

MZ_17 : .T..II.S.E.L.....H..E..N.V.I.....S...--...Q..YAT.DI..DI.....KEN.SR..R.VKE..EKYY-P....Y.NS.....L..TM.....RR....
 MZ_19 :E..N.....E.IA.....RS.H.--...V.YAT.DI..DI.....TE..S..G..VT..K..L.P.....N.....T..TM.....RR....
 MZ_22 :II.S.E.L.....H..E..K.V.I.....S...--...Q..YAT.DI..DI.....KEN.SR..R.VKE..EKYY-P....Y.NS.....L..TM.....RR....
 MZ_24 : ...DI...E.L.N.....H.KDP.R.V.....TSV.--...QT.YAT.DI..DI.....N.SR..QEVTE..KKH.-PG.N.T.TS....L..T.....RR....
 MZ_25 : .G.IM...Q.I.....K.I..V...G...SV.--...Q..FATNEV..DI.....EGD..QA.QNVSK..K.H.-P.....PA....L..TM.....RR....
 MZ_28 :II..E.M...V.....H..E..N.V..H...SV.--...QTIFAT.DI..DI.K.....EGQ..K..QGVKE...KY.-P..N.T.NS.A...L..T.....RR....
 MZ_33 : ...DTA...K..K..T.....FT...R.I.V...DR..TS.L.--...QT.YAM.DI...I.K.....ERN..AA.EG.....-P.....RR....
 MZ_36 : ...IM...E.L...I...H..E...G...S...--...QT.YAT.AI..DI.....GE...R..QGVVE..K.H.-P....A.APP...L..TM.....RR....
 MZ_38 : ...DII...E.L...T.....H..E...V.I...SV.--...QT.YAT.EV..DI.K.....EG..KD..HR.KE..K.I.-K...K.EPH...L..T.....RR....
 MZ_44 : ...DI...E.L.N.....H.KDP.R.V.....TSV.--...QT.YAT.DI..DI.....N.TR..QEVTE..QKH.-PG.N.T.TS....L..T.....RR....
 MZ_45 : .TK.II.S.K.L.....HL.E..N.V.I.....S...--...Q..YAT.DI..DI.....KGN..R..R.VKE..EKYY-P....Y.NS.....L..TM.....RR....
 MZ_48 : .RG...E.LQN.....E.IA.....RS.H.--...YAT.DI..DI.....TE..S..R..VT.....L.P.....N.....T..TM.....RR....
 MZ_55 : .G.IM...E.L.N.....HF.E...S...F--...Q..YAT.DI..DI.....IS...K..Q.VSK..AQH.-P....Q.NS....L..TM.....RR....
 MZ_57 : .TN.TI.S.E.L.....H..E..N.V.V.....S...--...Q..YAT.DI..DI.....EGN.SR..R.VKE..EKYY-P....Y.NS.....L..TM.....RR....
 MZ_62 : .KGDII..E.L.N.....H..E...V...S...SV.--...QTLFAT.NI..DI.....KSN...YRVSE..KGH.-...T.Q.A...L..T.....RR....
 MZ_65 :II..E.L.N.....KT..V...SV.--...Q..YAT.DI..DI.R.....EED...Q.VSR..A.H.-P..S.T.NS.....RR....
 MZ_67 :E..N.....E.IA.....RS.H.--...YAT.DI..DI.....TE..S..R..VT.....L.P.....N.....T..TM.....RR....
 MZ_70 : .RDII...E..K..T.....FTE..K.R.V...DR..TS.L.--...QT.YAM.DI...I.K.....ERD..TA.EG.....-P.....RR....
 MZ_71 :II.S.E.L.....H..E..K.V.....S...--...Q..YAT.DI..DI.....KEN.SR..R.VKE..KKYY-P....Y.NS.....L..TM.....RR....
 MZ_72 : .GGII...E.L.N.....H..D...V...G...TS...--...QT.FAT.EV..DI.....D...K..QEVGK..ATH.-...NS....L..TM.....RR....
 MZ_73 : .G.II...E.L...T.....ET.S.V...G...RS...--...QT.F.TN.I...I.....GGQ..Q...WVRE..KQHY-P..R..K.D.H...L..T.....RR....
 MZ_74 : ...GII...K.L...I...H..E...E...TS...--...Q..Y.T.DIT.DI...Y.S.RTED..A.TAK.V...K.HY-P....APPP...L.VT.....RR....
 MZ_77 : .T..II.STA.....I...H..E...T.V...SV.--...QT.YAT.DI..DI.....T...HLVSK..K.V.QL.G.V.N..A...L..T.....RR....
 MZ_84 : .G.IM...Q.I.....K.I..V...G...SV.--...Q..FATNEV..DI.....EGD..QA.QNVSK..K.H.-P...E..PA....L..TM.....RR....
 MZ_93 : .T..II.S.E.L.....H..E..N.V.I.....S...--...Q..YAT.DI..DI.....KEN.SR..R.VKE..EKYY-P....A.NS....L..TM.....RR....
 MZ_94 : .KRDII...E.L...I.N...VK..K.E...QS.H.--...Q..LY.T-NI...I.K.Y.TLN..SD.E...QNVSK...HL-P.T..E.QP.....T.....RR....
 MZ_95 : ...DII.K.E.L.K.V...H..E...V.I..S...S...--...QT.YAT.EV...I.....NGTA..E..QKVGK.....-P.RK.E..P....L..T.....RR....
 MZ_103 : .N.T..S.E.L.....H..E..N.V.V.....S...--...Q..YAT.DI..DI..Y...EGN.SR..R.VKE..EKYY-P....A.NS....L..TM.....RR....
 MZ_105 : ...NII.KYE..E..T.....H..E..K.E.I..G...SM.--...QT.YAM.DI...I.K...K.NES..YD..RKVSE...H.-P..N.T.NS....L..T.....RR....
 MZ_109 : .KRQ.A.K.Q.M.K...I.....LT..I...GDQ...SMQQ--...QT.YAM.DI...I.K...EL.ER...K.TQL.KQ..LTH.-PQE..KLQ.AV.....RR..S.
 MZ_111 :E..N.....E.IA.....RS.H.--...YAT.DI..DI.....RTE..S..R..VT.....L.P.....N.....T..TM.....RR....
 MZ_113 : ...M...PA.....V..LT..I..I...DQ...SMQQ--...QT.YAM.DI...I.K...E.ER...K.TQL.KQ..LTH.-PQE..KLQ.AV.....RR..S.
 MZ_114 :II..E.L...V...H..K...R...G...RSV.--...Q..YAT.DI..DI...T.N.TD.ED...NVSK..A.H.-P..N.Q.NS.A...L..T.....RR....
 MZ_123 : .K.II.KYE..E..T.....H..E..K.E.I..G...SM.--...QT.YAM.DI...I.K...K.NES..YD..RKVSE...H.-P..N.T.NS....L..T.....RR....
 MZ_127 : .KDII...E.L...T.....H..E..K.E..G...QS.--...QT.YAT.DI..DI...TG..KE.VQR.SK..A.H.-P....AN.....TM.....RR....
 MZ_130 :II.K.E.L.N.....HF.E..K.V...G...SV.--...QV.YAT.DIT.DI...V.EKD..TA.QNVSK..S.H.-P....NS.A...L.VTQ..T...RR....
 MZ_134 : .N.TI.S.E.L.....H..E..N.V.V.....S...--...Q..YAT.DI..DI..Y...ISEGN.SR..R.VKE..EKYY-P....A.NS....L..TM.....RR....
 MZ_135 : .K.II.S.E.L.N...I...H..E..Q...G...S...--...Q..YAT.EV..DP.K...AE...T..TKVKD..ASL.QL.I..V.NS....L..T.....RR....
 MZ_137 : .N.TI.S.E.L.....H..E..N.V.V.....S...--...Q..YAT.DI..DI..Y...EGN.SR..R.VKE..EKYY-P....A.NS....L..TM.....RR....

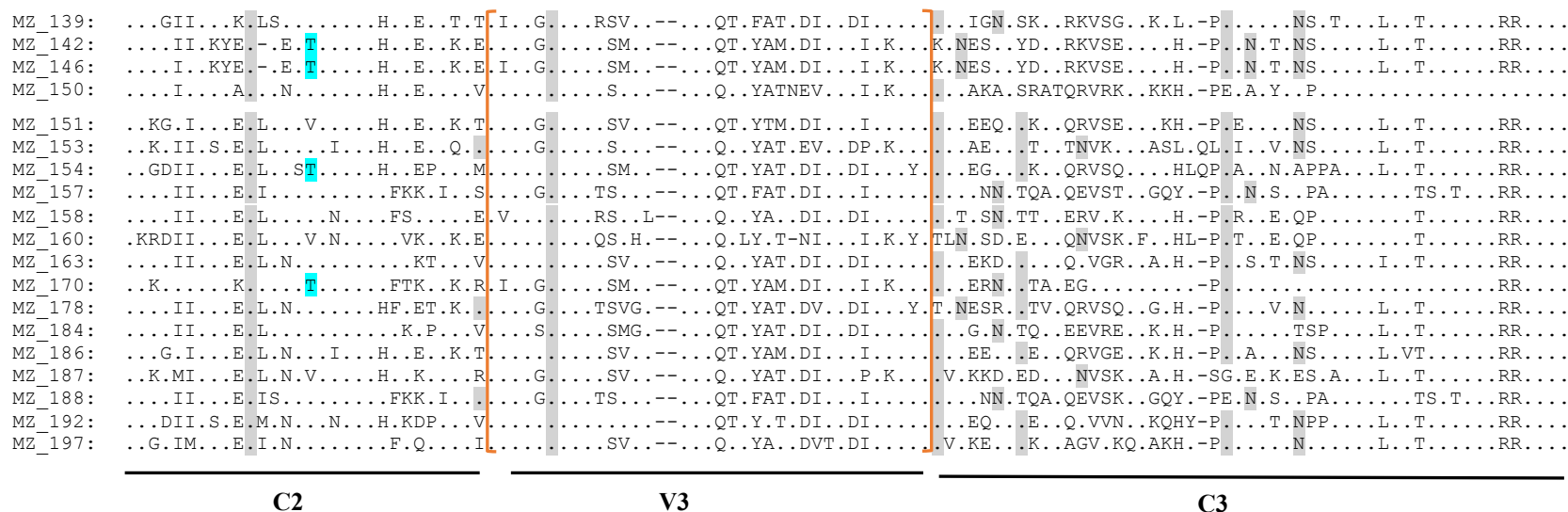


Figura A6 - Alinhamento múltiplo das 58 sequências aminoacídicas da região C2V3C3 dos isolados de VIH-1 de Moçambique em estudo. As sequências alinhadas foram comparadas com a sequência aminoacídica de HXB2, obtida na base de dados Los Alamos HIV database (www.hiv.lanl.gov). Os aminoácidos apresentados em cada posição estão representados pelo código internacional IUPAC. Os triângulos denotam resíduos de cisteína. Os potenciais locais de N-glicosilação (motivo NXS/T, em que X é qualquer aminoácido diferente de prolina) estão representados em cinzento. As posições em que os N-glicanos fazem parte de epítomos amplamente neutralizantes são indicadas: N276, N301 e N332. Os asteriscos são usados para indicar CD4bs.

Anexo 11.A

Tabela A9 - Polimorfismos genéticos identificados nas 58 sequências aminoacídicas da região V3. A tabela apresenta a análise comparativa entre as sequências do subtipo C e não-C com base na aplicação do teste exato de Fisher ou quando aplicável o teste do Chi-Quadrado (χ^2) ($p < 0,05$).

Mutação	Subtipo C (n=53)	Subtipo não-C (n=5)	<i>p</i>
T297I	12	0	0,5726
T297V	8	0	>0,9999
P299H	1	0	>0,9999
N300G	22	0	0,0675
N300S	3	0	>0,9999
N301D	4	0	>0,9999
N302R	2	0	>0,9999
N302Q	2	0	>0,9999
K305R	4	4	0,0008
K305T	9	0	>0,9999
K305Q	3	0	>0,9999
R306S	50	5	0,5849
I307V	17	1	0,5769
I307M	9	0	>0,9999
R308L	2	0	>0,9999
R308G	2	0	>0,9999
R308H	2	4	0,0002
R308Q	2	0	>0,9999
I309L	1	0	>0,9999
I309F	1	0	>0,9999
I309Q	2	0	>0,9999
R315Q	53	1	<0,0001
A316V	1	1	0,1664
A316T	28	0	0,0238

F317I	1	0	>0,9999
F317L	3	0	>0,9999
V318F	9	0	>0,9999
V318Y	43	5	0,2857
T319A	48	5	0,4725
I320T	41	4	0,8923
I320M	11	0	0,5718
G321N	3	0	>0,9999
K322D	38	5	0,1120
K322E	8	0	>0,9999
K322A	1	0	>0,9999
K322N	3	0	>0,9999
I323T	2	1	0,2408
N325D	34	5	0,1025
M326I	50	5	0,5849
M326P	3	0	>0,9999
Q328K	17	0	0,1320
Q328R	1	0	>0,9999
H330Y	8	0	>0,9999

Anexo 11.B

Tabela A10 - Polimorfismos genéticos identificados nas 58 sequências aminoacídicas da região C2V3C3. A tabela apresenta a análise comparativa entre as sequências do subtipo B e não-B com base na aplicação do teste exato de Fisher ou quando aplicável o teste do Chi-Quadrado (χ^2) ($p < 0,05$).

Mutação	Subtipo B (n=4)	Subtipo não-B (n=54)	<i>p</i>
T297I	0	12	0,5707
T297V	0	8	>0,9999
P299H	0	1	>0,9999
N300G	0	22	0,1052
N300S	0	3	>0,9999
N301D	0	4	>0,9999
N302R	0	2	>0,9999
N302Q	0	2	>0,9999
K305R	4	4	0,0002
K305T	0	9	>0,9999
K305Q	0	3	>0,9999
R306S	4	51	0,6283
I307V	0	18	0,2995
I307M	0	9	>0,9999
R308L	0	2	>0,9999
R308G	0	2	>0,9999
R308H	4	2	<0,0001
R308Q	0	2	>0,9999
I309L	0	1	>0,9999
I309F	0	1	>0,9999
I309Q	0	2	>0,9999
R315Q	0	54	<0,0001
A316V	1	1	0,1343
A316T	0	28	0,0452

F317I	0	1	>0,9999
F317L	0	3	>0,9999
V318F	0	9	>0,9999
V318Y	4	44	0,3441
T319A	4	49	0,5244
I320T	4	41	0,2652
I320M	0	11	0,5796
G321N	0	3	>0,9999
K322D	4	39	0,2209
K322E	0	8	>0,9999
K322A	0	1	>0,9999
K322N	0	3	>0,9999
I323T	0	3	>0,9999
N325D	4	35	0,1480
M326I	4	51	0,6283
M326P	0	3	>0,9999
Q328K	0	17	0,1820
Q328R	0	1	>0,9999
H330Y	0	8	>0,9999

Anexo 12

Tabela A11 - Polimorfismos genéticos identificados nas 58 sequências aminoacídicas da região V3 em estudo. A tabela apresenta a análise comparativa das sequências dos isolados de acordo com o seu tropismo viral – CCR5 ou CXCR4 com base na aplicação do teste exato de Fisher ou quando aplicável o teste do Chi-Quadrado (χ^2) ($p < 0,05$).

Mutação	CCR5 (n=50)	CXCR4 (n=8)	<i>p</i>
T297I	11	1	>0,9999
T297V	6	2	0,3030
P299H	1	0	>0,9999
N300G	20	2	0,6971
N300S	2	1	0,3648
N301D	3	1	0,4572
N302R	1	1	0,2589
N302Q	2	0	>0,9999
K305R	7	1	>0,9999
K305T	8	1	>0,9999
K305Q	1	2	0,0472
R306S	49	7	0,2589
I307V	17	1	0,4135
I307M	8	1	>0,9999
R308L	1	1	0,2589
R308G	2	0	>0,9999
R308H	4	2	0,1886
R308Q	2	0	>0,9999
I309L	1	0	>0,9999
I309F	1	0	>0,9999
I309Q	2	0	>0,9999
R315Q	46	8	>0,9999
A316V	2	0	>0,9999
A316T	25	4	>0,9999

F317I	1	0	>0,9999
F317L	0	3	0,0018
V318F	7	3	0,1309
V318Y	43	5	0,1309
T319A	47	5	0,0293
I320T	41	6	0,6391
I320M	10	1	>0,9999
G321N	2	2	0,0876
K322D	43	3	0,0068
K322E	7	1	>0,9999
K322A	1	0	>0,9999
K322N	0	3	0,0018
I323T	3	0	>0,9999
N325D	38	1	0,0011
M326I	47	8	>0,9999
M326P	3	0	>0,9999
Q328K	12	5	0,0401
Q328R	1	0	>0,9999
H330Y	5	3	0,0709

Anexo 13

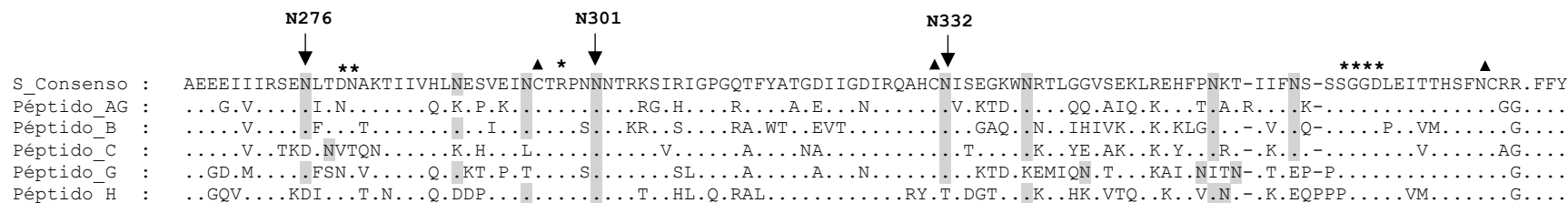


Figura A7 - Alinhamento múltiplo das sequências aminoacídicas da região C2V3C3 dos polipéptidos recombinantes e a sequência consenso dos isolados de VIH-1 de Moçambique em estudo. As sequências alinhadas foram comparadas com a sequência aminoacídica consenso. Os aminoácidos apresentados em cada posição estão representados pelo código internacional IUPAC. Os triângulos denotam resíduos de cisteína. Os potenciais locais de N-glicosilação (motivo NXS/T, em que X é qualquer aminoácido diferente de prolina) estão representados em cinzento. As posições em que os N-glicanos fazem parte de epítomos amplamente neutralizantes são indicadas: N276, N301 e N332. Os asteriscos são usados para indicar CD4bs.

Anexo 14

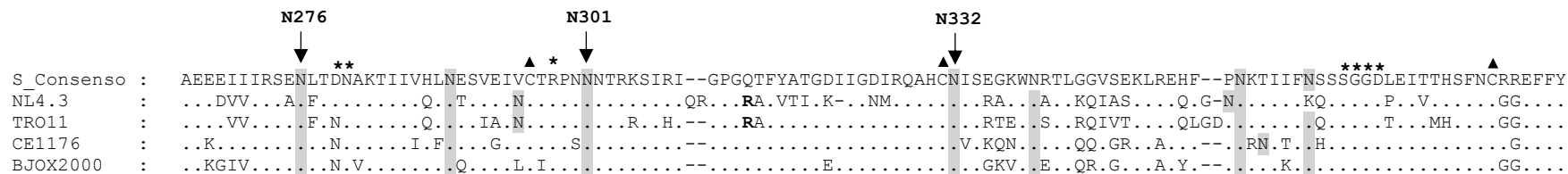


Figura A8 - Alinhamento múltiplo das sequências aminoacídicas da região C2V3C3 dos pseudovírus utilizados nos ensaios de neutralização e a sequência consenso dos isolados de VIH-1 de Moçambique em estudo. As sequências alinhadas foram comparadas com a sequência aminoacídica consenso. Os aminoácidos apresentados em cada posição estão representados pelo código internacional IUPAC. Os triângulos denotam resíduos de cisteína. Os potenciais locais de N-glicosilação (motivo NXS/T, em que X é qualquer aminoácido diferente de prolina) estão representados em cinzento. As posições em que os N-glicanos fazem parte de epítomos amplamente neutralizantes são indicadas: N276, N301 e N332. Os asteriscos são usados para indicar CD4bs.