Márcia Perez Resende Oliveros

Evolução das mutações de resistência aos inibidores de protease em pacientes infectados pelo HIV-1 subtipo F

Tese apresentada à Faculdade de Medicina da Universidade de São Paulo para a obtenção do título de Doutor em Ciências

Área de concentração: Patologia

Orientador: Dr. Luis Fernandez Lopez

SÃO PAULO

2009

Márcia Perez Resende Oliveros

Evolução das mutações de resistência aos inibidores de protease em pacientes infectados pelo HIV-1 subtipo F

Tese apresentada à Faculdade de Medicina da Universidade de São Paulo para a obtenção do título de Doutor em Ciências

Área de concentração: Patologia

Orientador: Dr. Luis Fernandez Lopez

SÃO PAULO

2009

AGRADECIMENTOS

A Luis Fernandez Lopez pelo apoio, incentivo e confiança.

Ao Dr. Ricardo Sobhie Diaz pela imensa colaboração durante todo o trabalho.

Aos Drs. Maria Cecília Sucupira e Luis Mário Janini pelas preciosas sugestões.

À Dra e amiga de todas as horas Clarice Gameiro da Fonseca Pachi pelo carinho e pela imensa colaboração.

A todos os amigos do Laboratório de Retrovirologia da UNIFESP pela receptividade e disponibilidade ao longo de todo o trabalho.

RESUMO

A elevada variabilidade genética do HIV tipo 1 e a seleção de mutações de resistência às drogas antirretrovirais tem representado um enorme desafio para o sucesso dos esquemas terapêuticos. Muitos estudos têm demonstrado que há padrões específicos de mutações para cada um dos subtipos de HIV-1. Os padrões do subtipo B são os mais bem definidos em função de sua presença majoritária nas epidemias da Europa e Estados Unidos. Contudo, a prevalência dos subtipos não-B, assim como a das formas recombinantes, tem aumentado significativamente em várias regiões. No Brasil, onde os subtipos B e F, e em alguns estados, também o C, coexistem, a presença de recombinantes BF vem ganhando destaque. Os objetivos deste trabalho foram comparar a região da protease do recombinante BF e do subtipo F puro quanto à possível presença de substituições diferentes, analisar o perfil mutacional e a covariação de mutações associadas ao tratamento e ao uso de inibidores específicos na protease do HIV-1 subtipo F e avaliar a reversão de mutações que já haviam sido selecionadas e a seleção de novas mutações após a troca do esquema terapêutico.

Descritores:
1.EPIDEMIOLOGIA E BIOESTATÍSTICA
2. EPIDEMIOLOGIA MOLECULAR
3.RESISTÊNCIA AOS MEDICAMENTOS
4.HIV-1

SUMMARY

The high genetic variability of HIV type 1 and the selection of antiretroviral resistance mutations have represented an enormous challenge to therapeutic schema success. Many studies have demonstrated that there are specific mutations patterns for each of the HIV-1 subtypes. Subtype B mutation profiles are the best studied due to its high presence in Europe and USA epidemics. However, prevalence of non-B subtypes, as well as recombinant forms, has been significantly increasing in many regions. In Brazil, where subtypes B and F, and in some states also C, have been co-circulating, BF recombinants are commonly found. The aims of this work were to compare protease region of BF recombinants to pure subtype F to search for different amino acid substitutions; analyzing mutation profile and co-variation of related-to-treatment mutations; verifying the relationship between the presence of groups of mutations and the use of specific antiretroviral drugs; and evaluating the selection of new mutations after changing therapeutic schema.

Descriptors:
1 EPIDEMIOLOGY AND BIOESTATISTICS
2 MOLECULAR EPIDEMILOGY
3 DRUG RESISTANCE
4 HIV-1

SUMÁRIO

CAPÍTULO 1	INTRODUÇÃO	14
CAPÍTULO 2	OBJETIVOS	20
2.1	OBJETIVOS GERAIS	20
2.2	OBJETIVOS ESPECÍFICOS	20
CAPÍTULO 3	REVISÃO DE LITERATURA	21
3.1	MORFOLOGIA, ORGANIZAÇÃO GENÔMICA, CICLO VIRAL E ORIGEM DE FORMAS RECOMBINANTES	21
3.1.1	CARACTERÍSTICAS ESTRUTURAIS E GENÔMICAS	21
3.1.2	CICLO VIRAL: ETAPA INICIAL (EARLY PHASE)	24
3.1.3	CICLO VIRAL: LATE FASE	26
3.1.4	ORIGEM DE FORMAS RECOMBINANTES	28
3.2	EVOLUÇÃO DA INFECÇÃO PELO HIV, TERAPIA ANTIRRETROVIRAL E FALHA TERAPÊUTICA	31
3.2.1	EVOLUÇÃO DA INFECÇÃO	31
3.2.2	TERAPIA ANTIRRETROVIRAL (TARV)	32
3.2.3	CLASSES DE ANTIRRETROVIRAIS E ESCOLHAS PARA OS ESQUEMAS INICIAIS	33

3.2.4 FALHA TERAPÊUTICA 40

3.2.5 ADERÊNCIA E EMERGÊNCIA DE MUTAÇÕES DE RESISTÊNCIA AOS ANTIRETROVIRAIS 41

3.2.6 QUASISPECIES VIRAIS: COMPLEXO SISTEMA ADAPTATIVO 41

3.2.7 RESISTÊNCIA PRIMÁRIA E RESISTÊNCIA SECUNDÁRIA: MAIS ALGUMAS CONSIDERAÇÕES SOBRE A TRANSMISSÃO DO HIV-1 43

3.2.8 MUTAÇÕES PRINCIPAIS E MUTAÇÕES SECUNDÁRIAS 44

3.2.9 POLIMORFISMOS 45

3.2.10 ROTAS MUTACIONAIS (PATHWAYS) 46

3.3 ANÁLISE MULTIVARIADA 48

3.3.1 MÉTODOS APROPRIADOS PARA CADA SITUAÇÃO 48

3.3.2 ANÁLISE DE CLUSTERS: MATRIZES DE DADOS E MEDIDAS DE DISTÂNCIA ENTRE AS VARIÁVEIS 49

3.3.3 ANÁLISE DE CLUSTERS: COEFICIENTE DE SIMILIARIDADE DE JACCARD 50

3.3.4 ANÁLISE DE CLUSTERS: MATRIZES DE DISTÂNCIA 52

3.3.5 ANÁLISE POR AGRUPAMENTO HIERÁRQUICO (HCA) 53

3.3.6 ANÁLISE POR COMPONENTES PRINCIPAIS (PCA) 58

3.3.7 COMPARAÇÃO ENTRE HCA E PCA 59

CAPÍTULO 4 CASUÍSTICA E MÉTODOS 61

4.1 CASUÍSTICA 61

4.1.1 CASUÍSTICA DA 1ª ETAPA: PERFIL DA PROTEASE DE HIV-1 SUBTIPO F DE RECOMBINANTES BF 61

4.1.2 CASUÍSTICA DA 2ª ETAPA: COVARIAÇÃO DAS MUTAÇÕES NA PROTEASE DO SUBTIPO F, ASSOCIAÇÃO COM O USO DE INIBIDORES DE PROTEASE E ANÁLISE TEMPORAL DAS MUTAÇÕES ASSOCIADAS AO TRATAMENTO 62

4.2 MÉTODOS 63

4.2.1 1ª ETAPA: PERFIL DA PROTEASE DE HIV-1 SUBTIPO F DE RECOMBINANTES BF 63

4.2.2 2ª ETAPA: COVARIAÇÃO DAS MUTAÇÕES NA PROTEASE DO SUBTIPO F, ASSOCIAÇÃO COM O USO DE INIBIDORES DE PROTEASE E ANÁLISE TEMPORAL DAS MUTAÇÕES ASSOCIADAS AO TRATAMENTO 65

CAPÍTULO 5 RESULTADOS 69

5.1 1ª ETAPA: PERFIL DA PROTEASE DE HIV-1 SUBTIPO F DE RECOMBINANTES BF 69

5.1 2ª ETAPA: COVARIAÇÃO DAS MUTAÇÕES NA PROTEASE F, DETERMINAÇÃO DOS AGRUPAMENTOS ESPECÍFICOS (CLUSTERS) PARA ESTE SUBTIPO DE HIV-1, ASSOCIAÇÃO COM O USO DE IPs ESPECÍFICOS E ANÁLISE TEMPORAL DAS MUTAÇÕES ASSOCIADAS AO TRATAMENTO 76

CAPÍTULO 6 DISCUSSÃO 91

6.1 1ª ETAPA: PERFIL DA PROTEASE DE HIV-1 SUBTIPO F DE RECOMBINANTES BF 91

6.2 2ª ETAPA: COVARIAÇÃO DAS MUTAÇÕES NA PROTEASE F, DETERMINAÇÃO DOS AGRUPAMENTOS ESPECÍFICOS (CLUSTERS) PARA ESTE SUBTIPO DE HIV-1, ASSOCIAÇÃO COM O USO DE IPs ESPECÍFICOS E ANÁLISE TEMPORAL DAS MUTAÇÕES ASSOCIADAS AO TRATAMENTO 94

CAPÍTULO 7 CONCLUSÃO 101

CAPÍTULO 8 ANEXOS 103

CAPÍTULO 9 REFERÊNCIAS BIBLIOGRÁFICAS 106

CAPÍTULO 10 APÊNDICE 124

ABREVIAÇÕES

HIV	vírus da imunodeficiência humana (human immunodeficiency virus)
TARV	terapia antiretroviral
CRF	forma recombinante circulante (circulating recombinant form)
URF	forma recombinante única (unique recombinant form)
IP	inibidor de protease
TR	transcriptase reversa
ITRN	inibidor de transcriptase reversa análogo de nucleosídeo
ITRNN	inibidor de transcriptase reversa não análogo de nucleosídeo
RTV	ritonavir
IDV	indinavir
SQV	saquinavir
LPV	lopinavir
ATV	atazanavir
TPV	tipranavir
APV	amprenvir
DRV	darunavir

IP/r	inibidor de protease com RTV utilizado como booster
HCA	análise por agrupamento hierárquico
PCA	análise por componentes principais
ntr	indivíduos não tratados
tr	indivíduos tratados

QUADROS E FIGURAS

FIGURA 3.1.	VISÃO GERAL DA ESTRUTURA DO HIV	22
FIGURA 3.2.	RELAÇÃO GENE/PROTEÍNA VIRAL E ALGUMAS FUNÇÕES EXERCIDAS PELAS PROTEÍNAS VIRAIS	23
FIGURA 3.3.	REPLICAÇÃO DO HIV-1 E ORIGEM DE FORMAS RECOMBINANTES	29
FIGURA 3.4.	FORMAS RECOMBINANTES CIRCULANTES DE HIV-1	30
QUADRO 3.1.	INIBIDORES DE PROTEASE	37
FIGURA 3.5.	MATRIZ DE DADOS PARA ANÁLISE DE CLUSTERS	49
FIGURA 3.6.	MATRIZ DE DADOS BINÁRIOS PARA ANÁLISE DE CLUSTER DAS MUTAÇÕES QUE COMPÕEM AGRUPAMENTOS NO HIV-1	50
FIGURA 3.7A E B.	MATRIZ DE DISTÂNCIA BASEADA NO COEFICIENTE DE SIMILARIDADE DE JACCARD E MATRIZ BASEADA NO COEFICIENTE DE DISSIMILARIDADE PARA O MESMO CONJUNTO DE MUTAÇÕES DA PROTEASE DO HIV-1	53
FIGURA 3.8.	EXEMPLO DE DENDROGRAMA MOSTRANDO OS AGRUPAMENTOS SEGUNDO O NÍVEL DE SIMILARIDADE ENTRE ALGUMAS MUTAÇÕES ASSOCIADAS AO TRATAMENTO NA PROTEASE F DO HIV-1	55
FIGURA 3.9.	ALGORITMO PADRÃO	56
FIGURA 3.10.	PCA PARA AS MUTAÇÕES L10I, K20R, D30N, V32I, D35N, I54V, V82A, N88D	60
FIGURA 5.1.	PROPORÇÃO DE RECOMBINANTES BF NA REGIÃO POL DE ACORDO COM A PRESENÇA DE UMA PROTEASE CLASSIFICADA COMO F, B OU HÍBRIDA BF	70

FIGURA 5.2A E B.	ÁRVORES FILOGENÉTICAS DA REGIÃO DA PROTEASE CONSTRUÍDAS COM SEQUÊNCIAS REFERÊNCIA DOS SUBTIPOS A, B, C, D E F E AS SEQUÊNCIAS DO SUBTIPO F E BFPROF DA BASE DE DADOS DO LABORATÓRIO DE RETROVIROLOGIA DA UNIFESP	71
FIGURA 5.3.	DENDROGRAMA DE SIMILARIDADE ENTRE AS MUTAÇÕES ASSOCIADAS AO TRATAMENTO GERADO POR HCA, COM AVERAGE LINKAGE METHOD	80
FIGURA 5.4.	FREQUÊNCIA RELATIVA DE USO DE CADA IP NOS ESQUEMAS TERAPÊUTICOS DO GRUPO IPTR	82
FIGURA 5.5A E B.	DENDROGRAMAS DE SIMILARIDADE ENTRE AS MUTAÇÕES DO CLUSTER 1 E ENTRE AS MUTAÇÕES DO CLUSTER 5, E VISUALIZAÇÃO DOS SUBGRUPOS	85

LISTA DE TABELAS

TABELA 3.1.	TABELA DE CONTINGÊNCIA 2 X 2 CONTENDO O NÚMERO DE DUPLAS PRESENÇAS (XY), PRESENÇAS DE X E AUSÊNCIAS DE Y (X0), PRESENÇAS DE Y E AUSÊNCIAS DE X (0Y) E DUPLAS AUSÊNCIAS (00)	51
TABELA 5.1.	POSIÇÕES COM DIFERENÇAS SIGNIFICATIVAS NA FREQUÊNCIA DE MUTAÇÕES ENTRE NTRF E NTRBFPROF E ENTRE TRF E TRBFPROF	74
TABELA 5.2.	POSIÇÕES COM DIFERENÇAS SIGNIFICATIVAS NAS FREQÜÊNCIAS DE MUTAÇÕES E ENTRE NTRF E TR F E ENTRE NTR BFPROF E TRBFPROF	75
TABELA 5.3.	MUTAÇÕES POSITIVAMENTE ASSOCIADAS AO TRATAMENTO COM VALORES DE COEFICIENTE DE SIMILARIDADE DE JACCARD ≥0,2 E VALORES DAS PROBABILIDADES CONDICIONAIS P(Y|X) E P(X|Y) PARA O GRUPO IPTR	77
TABELA 5.4.	PARES DE MUTAÇÕES ASSOCIADAS AO TRATAMENTO COM VALORES DO COEFICIENTE DE SIMILARIDADE DE JACCARD ≥0,2 E VALORES DAS PROBABILIDADES CONDICIONAIS P(Y|X) E P(X|Y) PARA O GRUPO IPTR	79
TABELA 5.5.	CORRELAÇÃO ENTRE UMA MUTAÇÃO ESPECÍFICA E O USO DE UM OU MAIS IP NO SUBTIPO F	83
TABELA 5.6.	CORRELAÇÃO ENTRE CLUSTERS DE MUTAÇÕES E IP NA PROTEASE DO HIV-1 SUBTIPO F	86
TABELA 5.7.	MUTAÇÕES ASSOCIADAS AO TRATAMENTO E ESQUEMAS TERAPÊUTICOS DOS PACIENTES COM SEQUÊNCIAS OBTIDAS EM DOIS MOMENTOS DISTINTOS	88

CAPÍTULO 1. INTRODUÇÃO

Desde a descrição e definição do primeiro caso de AIDS nos Estado Unidos (Gotlieb et al, 1981), enormes avanços têm sido registrados no campo da pesquisa biomédica em resposta à pandemia de HIV/AIDS. Em 1985, análises nucleotídicas demonstraram grande semelhança entre três tipos de virus que haviam sido isolados de pacientes aidéticos de diferentes grupos de risco em anos anteriores e, em 1986, o International Committee of Viral Taxonomy (ICTV) os renomeou para human immunodeficiency virus (HIV). Segundo as normas internacionais baseadas principalmente nas propriedades estruturais dos vírus e estabelecidas em 1998 pelo ICTV, os vírus conhecidos como HIV pertencem à família Retroviridae e ao gênero Lentivirus (do latim, lentus = lento). Ainda em 1985, foi desenvolvido o teste diagnóstico específico para anticorpos de HIV e em relativamente pouco tempo estudos moleculares delinearam a estrutura genética, a organização e os mecanismos de replicação viral. Estas informações em conjunto com dados sobre a patogênese do HIV foram fundamentais para facilitar o rápido desenvolvimento de drogas antirretrovirais capazes de limitar a replicação viral e evitar maiores danos ao sistema imunológico.

Atualmente, 25 antirretrovirais estão disponíveis para o tratamento da infecção pelo HIV, o que vem contribuindo para uma substancial melhora da qualidade de vida dos pacientes e para a redução da morbidade e da mortalidade pela AIDS. Um tratamento bem sucedido pode resultar em acentuada redução da carga viral (virions livres no plasma) até níveis indetectáveis (abaixo de 50 cópias/ml), fato que é responsável por reduzir significativamente as chances de transmissão (Relucio & Holodny, 2002).

Apesar dos progressos e inúmeros benefícios trazidos pela terapia antirretroviral (TARV), ainda há uma série de lacunas a serem preenchidas. Apenas 28% das pessoas que necessitam de TARV nos países de baixa e média renda recebem tratamento, e mesmo em países desenvolvidos verificam-se problemas relacionados ao acesso às drogas antirretrovirais (Fauci, 2007). Como conseqüência, a incidência de infecção pelo HIV continua elevada. Segundo a Organização Mundial de Saúde (WHO) e UNAIDS (Joint United Nations Programme on HIV/AIDS), em dezembro de 2007 foram contabilizados em torno de 2,5 milhões de novos infectados e o número de pessoas vivendo com HIV estava em torno de 33 milhões, sendo a grande maioria (mais de 30 milhões) composta por adultos. Ainda segundo a OMS, em 2007 a AIDS provocou 2,1 milhões de mortes.

Segundo estimativas do Programa Conjunto das Nações Unidas sobre HIV/AIDS (WHO- UNAIDS, 2006), existem, no Brasil, cerca de 620 mil pessoas infectadas, o que corresponde a mais de um terço do total de infectados em toda a América Latina. Calcula-se que ocorrem no Brasil, em média, 11 mil óbitos a cada ano (Programa Nacional DST/AIDS, Ministério da Saúde, 2007), com um acumulado de aproximadamente 193 mil óbitos notificados entre 1980 e 2006 (Boletim Epidemiológico Ministério da Saúde, 2007). A taxa de prevalência na população entre 15 e 49 anos tem se estabilizado em torno de 0,61% desde 2000 (Programa Nacional DST/AIDS, Ministério da Saúde, 2007). A tendência à estabilização esconde profundas transformações no perfil socioeconômico da doença. Embora os grupos mais vulneráveis continuem sendo trabalhadores do sexo, homens que fazem sexo com homens e usuários de drogas injetáveis, em grandes centros urbanos, a epidemia de HIV/AIDS tem se alastrado de forma preocupante entre mulheres, idosos, na população

de baixa renda, além de avançar pelo interior (Programa Nacional DST/AIDS, Ministério da Saúde, 2007; Greco & Simão, 2007; Okie, 2006).

Além das questões que envolvem as políticas de prevenção e acesso a tratamento, fatores relacionados às características do vírus representam problemas para a erradicação da pandemia. O genoma do HIV, como o dos demais retrovirus, é composto por RNA e caracteriza-se por enorme diversidade genética, consequência de sua elevada taxa de mutação, ausência de mecanismos eficientes de correção de erro, altíssima taxa de replicação *in vivo* (aproximadamente 10^9 partículas virais por dia) e ocorrência de recombinação como parte dos mecanismos naturais de sua replicação.

Até o momento, foram evidenciados por estudos sorológicos dois tipos antigênicos de HIV: tipo 1 (HIV-1) e tipo 2 (HIV-2). O tipo 1 é o mais virulento e o principal responsável pela pandemia da AIDS (Grant & Cock, 2001), enquanto o HIV–2 é encontrado quase que exclusivamente no oeste da África. Cada um dos tipos, HIV-1 e HIV-2, divide-se em diferentes subtipos que se distinguem por suas organizações genômicas e relações filogenéticas (Hahn et al, 2000). O HIV-1, amplamente distribuído por todos os continentes, apresenta enorme variabilidade e, para descrevê-la, usa-se o conceito de quasispecies (Domingo et al, 1998), que considera o HIV como subpopulações virais, e não como um genoma único.

O HIV tipo 1 divide-se em três grupos distintos, M, N e O, ficando os dois últimos restritos a alguns poucos países do oeste africano. O grupo M, que sozinho responde por mais de 90% dos casos notificados de HIV/AIDS, pode ser subdividido filogeneticamente em vários subtipos (A-D, F-H, J e K), muitas formas recombinantes circulantes (CRF) e inúmeras formas recombinantes únicas (URF).

Estudos epidemiológicos têm demonstrado que a grande diversidade do HIV-1 tem forte associação geográfica e diferentes subtipos predominam em diferentes regiões

(Buonaguro et al, 2007). O Subtipo A e recombinantes A/G predominam no Oeste da África e na África Central. O subtipo de maior prevalência na Europa e Américas é o B. No Sul e Leste da África, bem como na Índia, o subtipo C é o predominante. Na América do Sul, apesar do subtipo B ainda ser o de maior prevalência, a freqüência do F e do C, bem como de formas recombinantes (BF, BC, FC) têm aumentado significativamente (Leal et al, 2008; Sanabani et al, 2006). O subtipo F e recombinantes BF têm sido frequentemente encontrados em muitos países sul americanos (Carr et al, 2001). Na Argentina, foram identificadas duas epidemias (Segura et al, 2007; Avila et al, 2002): subtipo B, responsável pela maioria dos casos em homossexuais masculinos, e recombinantes BF, encontrados principalmente em heterossexuais e usuários de drogas injetáveis. De acordo com alguns estudos, os recombinantes BF são responsáveis por 70% dos casos de HIV/AIDS na Argentina.

No Brasil, onde o subtipo B é o principal responsável pela epidemia, os subtipos F e C, e os recombinantes BF, BC e FC, circulam entre os mesmos grupos de risco, com prevalências que variam conforme a região. O subtipo C é responsável por aproximadamente 50% dos casos de HIV/AIDS na região sul (Brindeiro et al, 2003; Soares et al, 2003). O subtipo F é o segundo em prevalência nos estados do Rio de Janeiro e São Paulo, ambos considerados os epicentros da epidemia de AIDS no Brasil (Sabino et al, 1996; Morgado et al, 1998). O subtipo F é também encontrado em outras regiões brasileiras como Distrito Federal e alguns estados do norte e nordeste (Cerqueira et al, 2004; Cavalcanti et al, 2007; Gadelha et al, 2003). De forma particular, no estado de São Paulo, recombinantes BF tornaram-se mais frequentes que o F puro. Um estudo com amostras provenientes desta região descreveu duas CRFs compostas exclusivamente por B e F (CRF28_BF e CRF29_BF) (Sá-Filho et al, 2006), e mais recentemente, duas novas CRFs também compostas exclusivamente por B e F foram

descritas no Rio de Janeiro (Guimarães et al, 2008). Outro estudo brasileiro (Sá-Filho et al, 2008) analisou o perfil de mutações de resistência na protease e na transcriptase reversa, dois importantes alvos moleculares dos antirretrovirais, dos subtipos B, F e recombinantes BF de amostras da cidade de Santos, uma das mais importantes para a epidemia de HIV-1 no estado de São Paulo. Os resultados sugerem que os mesmos aminoácidos estão sendo selecionados nos três grupos como resultado da pressão exercida pelas drogas. Outro estudo (Leal et al, 2008) demonstrou que cepas de vírus BF têm maior *fitness* quando comparados a cepas de F puros *in vitro*, sugerindo que os recombinantes BF apresentem algum tipo de vantagem.

Mesmo na Europa e Estados Unidos, onde o subtipo B predomina subtipos não-B e formas recombinantes estão se tornando mais freqüentes (Tagliamonte et al, 2006; Tramuto et al, 2007; Holguín et al, 2007). A detecção de subtipos não-B em regiões inicialmente dominadas pelo B indica que a epidemia de HIV-1 está evoluindo e a possibilidade de aparecimento de novas formas recombinantes é alta. Esta evolução torna urgente uma melhor compreensão dos perfis de mutação dos subtipos não-B e formas recombinantes, principalmente pelo fato de que a maior parte do que se conhece sobre resposta ao tratamento e interpretação de testes genotípicos baseia-se no subtipo B.

Trabalhos recentes têm contribuído para elucidar as diferenças entre o subtipo B e os não-B no que se refere aos perfis de mutações de resistência às drogas (Kantor et al, 2005; Shafer et al, 2008; Shafer and Schapiro, 2008). No entanto, estas informações não somente precisam ser complementadas com dados de um número maior de sequências de não-B como também de formas recombinantes cuja prevalência vem aumentando significativamente.

CAPÍTULO 2. OBJETIVOS

2.1. OBJETIVOS GERAIS

- Definição das rotas mutacionais na protease do HIV-1 subtipo F.
- Análise da relação entre o uso de determinado inibidor de protease e a seleção de uma dada mutação ou combinação de mutações.
- Estudo da evolução das mutações associadas à resistência.

2.2. OBJETIVOS ESPECÍFICOS

- Comparação da protease do subtipo F presente no recombinante BF com a protease do subtipo F puro (para a região pol).
- Descrição de mutações polimórficas e não polimórficas associadas à resistência na protease do subtipo F.
- Análise de agrupamentos ou clusters de mutações associadas ao tratamento.
- Análise de agrupamentos ou clusters de mutações associadas ao tratamento com IPs.
- Análise da tendência de mutações principais e acessórias após a troca do esquema terapêutico.

CAPÍTULO 3. REVISÃO DE LITERATURA

3.1. MORFOLOGIA, ORGANIZAÇÃO GENÔMICA, CICLO VIRAL E ORIGEM DE FORMAS RECOMBINANTES.

3.1.1. CARACTERÍSTICAS ESTRUTURAIS E GENÔMICAS

Enquanto um vírus convencional é composto por ácido nucléico e capsídeo protéico, o HIV é considerado um vírus complexo por apresentar um envoltório de natureza lipoglicoprotéica, denominado envelope (Fig. 3.1.). O envelope constitui-se de uma região voltada para o interior do vírus e outra externa, responsável pela afinidade com receptores celulares. Na região externa, que equivale à superfície do vírus, são encontradas 72 projeções contendo trímeros ou tetrâmeros de duas glicoproteínas: gp120 e gp41. A gp 120 ou proteína externa da superficie prende-se ao vírus através de interações com a gp41 ou proteína transmembrana. É na gp120 que são encontrados os sítios de ligação aos receptores celulares e, portanto, importantes domínios a serem neutralizados com a finalidade de se interromper o ciclo viral. O complexo gp120-gp41 encontra-se mergulhado numa dupla cadeia lipídica derivada das membranas da célula hospedeira e onde estão também inúmeras proteínas celulares, incluindo antígenos de histocompatibilidade, actina e ubiquitina (Turner & Summers, 1999).

Internamente, o envelope é composto por uma capa protéica em que a proteína p17 está ligada ao ácido mirístico. Essa capa é responsável por fornecer a matriz, vital para a integridade do vírion (partícula viral completa), além de aparentemente ser necessária para a incorporação das proteínas do envelope nos virions maduros.

Envolto pelo envelope encontra-se o nucleocapsídeo que corresponde ao conjunto formado pelo capsídeo e pelo material genético. O nucleocapsídeo possui um formato cônico, típico dos lentivirus, e é formado pela proteína p24, além das proteínas p6 e p9; essas últimas associadas ao material genético.

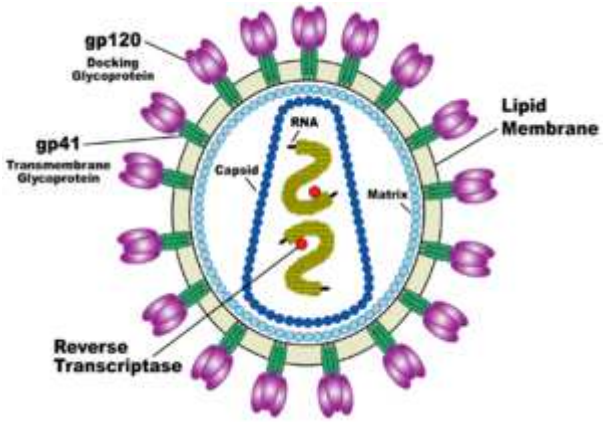


Figura. 3.1. Visão geral da estrutura do HIV (Wikipedia, capturado em 04/04/2007).

Como nos demais retrovírus, o genoma do HIV é composto por duas fitas simples de RNA, positivas e idênticas entre si (Wigg, 2002). Intimamente associadas a essas fitas de RNA, são encontradas três importantes enzimas virais – transcriptase reversa, integrase e protease, as proteínas acessórias nef, vif e vpr além das já mencionadas proteínas do nucleocapsídeo – p6 e p9.

O genoma do HIV tem, aproximadamente, 9,8 Kb, com 9 genes que apresentam diversas possibilidades de processamentos alternativos, o que permite a síntese de um grande número de diferentes polipeptídeos, proteínas e enzimas (Fig. 3.2.). Três desses genes são comuns a todos os retrovirus: gag, pol e env, e codificam importantes enzimas e proteínas participantes diretas da estrutura do vírus ou de seu ciclo. Os demais genes são regulatórios (tat e rev) e acessórios (nef, vif, vpr e vpu) Nas

extremidades 5’ e 3’ encontram-se longas sequências repetidas ou LTRs (*long terminal repeats*).

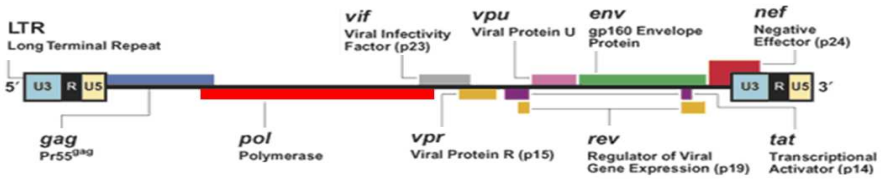


Fig 3.2. Relação gene/proteína viral e algumas funções exercidas pelas proteínas virais (adaptado de Wigg, 2002).

Os retrovirus dependem de recursos da célula hospedeira para a replicação e o acesso a estes recursos pode estar limitado a uma fase particular do ciclo celular. Observações *in vitro* mostram que a expressão gênica do HIV causa a interrupção do ciclo celular na fase G2 de uma grande variedade de células, de linfócitos T a leveduras (Jowett et al, 1995; Elder et al, 2001). Um importante fator celular é a densidade do receptor CD4 e dos co-receptores CXCR4 e CCR5 na superfície das células-alvos. Este conjunto de receptores é necessário para a fusão do envelope viral em células-alvo, podendo ser limitantes para a infecção *in vivo* e *in vitro*. A diminuição de CD4, mediada por proteínas virais Nef, Env e Vpu, e dos co-receptores facilita a replicação viral por interferir em várias conseqüências indesejáveis para o vírus, como indução de sinais pró-apoptose pela CD4, redução da infectividade pela incorporação de receptores de superfície no envelope viral e super infecção pelo vírus HIV (fenômeno que seria tóxico para a célula) (Wildum et al, 2006).

3.1.2. CICLO VIRAL: ETAPA INICIAL (*EARLY PHASE*)

O ciclo replicativo do HIV ocorre em duas etapas. A etapa inicial (*early phase*) envolve os processos de adsorção, fusão, desnudamento, transcrição reversa e integração do genoma viral ao genoma da célula hospedeira (formação do provirus). A etapa final (*late phase*) caracteriza-se pela expressão regulada do genoma proviral integrado e todos os processos que levam à montagem e maturação do vírus, incluindo os próprios eventos de montagem e maturação (Turner & Summers, 1999).

Adsorção, fusão e desnudamento

A adsorção é o processo de interação do envelope viral com receptores específicos da superfície da célula. Participam deste processo as duas glicoproteínas do envelope, gp 120 e gp 41, que formam unidades funcionais triméricas (Greene & Peterlin, 2003). Cada uma dessas unidades é composta por três moléculas da gp120 expostas na superfície do vírion associadas a três moléculas da gp41 na superfície da célula-alvo. Esta unidade induz uma alteração conformacional nas proteínas do envelope que permite ao vírion ligar-se a receptores específicos na superfície da célula. O principal receptor é o CD4, mas, além dele participam co-receptores de quimiocina que podem ser de dois tipos: CCR5 e CXCR4.

A ligação entre gp120, CD4 e o receptor de quimiocina produz uma mudança radical na conformação da gp41. Arranjada como um trímero na membrana viral, esta proteína se lança projetando três domínios de fusão de peptídeos que se prendem à dupla camada lipídica da célula, formando uma estrutura em forma de grampo que une as duas membranas (celular e do envelope). A fusão dessas membranas leva à liberação do conteúdo viral no interior da célula através de um processo denominado desnudamento.

O conteúdo viral que efetivamente entra na célula é composto pelo complexo ribonucleoprotéico (duas fitas de RNA, proteínas estabilizadoras e tRNA), transcriptase reversa, integrase, protease e algumas proteínas acessórias – Nef, Vif e Vpr.

Transcrição reversa e integração do genoma viral na célula hospedeira.

No citoplasma da célula-hospedeira, ocorre o processo de transcrição reversa que consiste na síntese de DNA a partir do RNA genômico viral. Essa reação é catalizada pela enzima Transcriptase Reversa (RT) e utiliza um $_{t}$RNA viral ou celular como primer (Wigg, 2002). Cada fita de RNA é transcrita em uma fita negativa de DNA ($_{c}$DNA). A RT também atua como ribonuclease H degradando o RNA viral, além de conduzir a síntese da fita positiva do DNA.

O DNA viral final é constituído por cadeia dupla e pode ter dois destinos: permanecer no citoplasma na forma circular não integrada ou ser ativamente transportado para o núcleo da célula. No segundo caso, o DNA viral faz parte de um complexo de pré-integração constituído por uma matriz, pela enzima integrase e pela proteína Vpr. A integrase catalisa uma clivagem endonucleolítica chamada processamento da extremidade 3' (*3' end process*) (Integrase. HIV Slideset Download Library, http://clicaloptions.com, capturado em 31/03/2007). Durante este processo, dois nucleotídeos são removidos da extremidade 3' de cada uma das duas fitas de DNA viral.

O próximo passo ocorre após o complexo de pré-integração ter cruzado a membrana nuclear da célula hospedeira, o que propicia o contato com o DNA hospedeiro. A integrase, então, catalisa um processo conhecido como transferência de cadeia de DNA (*DNA strand transfer*). A integrase cliva a cadeia de DNA hospedeiro e une a extremidade 3' do DNA viral à extremidade 5' do DNA hospedeiro.

Finalmente, enzimas celulares reparam a cadeia de DNA, resultando numa total integração do DNA viral ao DNA hospedeiro.

3.1.3. CICLO VIRAL: *LATE FASE*

A expressão gênica tem início com a síntese de transcritos (mRNAs) processados ou não que são transportados por translação para fora do núcleo. As primeiras moléculas de RNA processadas são as que codificam as proteínas regulatórias Tat, Rev e Nef.

A proteína Tat é uma importante ativadora dos processos transcricionais, enquanto Rev é responsável por mediar a saída de mRNAs não processados (idênticos ao RNA viral) e mRNAs com um único processamento. Em ambas as situações, os transcritos contêm informações necessárias para a síntese e empacotamento no citoplasma das proteínas Gag e Gag-Pol e das proteínas Env, Vpu, Vif e Vpr.

Síntese das poliproteínas virais: Env, Gag e Gag-Pol

A síntese da poliproteína precursora das proteínas do envelope – Env (gp 160) ocorre nas membranas do Retículo Endoplasmático Rugoso, utilizando o código processado do gene env. A poliproteína Env sofre uma transformação pós-translacional no retículo e uma clivagem no aparelho golgiensi para produzir o complexo glicoproteína trimérica $(gp41\text{-}gp120)_3$. Este complexo é então transportado para a membrana celular para o início da montagem do vírus (Wigg, 2002).

Moléculas de RNA não processado (idêntico ao RNA viral) são traduzidas em poliproteínas Gag e Gag-Pol nos ribossomos livres.

Montagem da partícula viral

Moléculas de poliproteínas virais (Gag e Gag-Pol) reúnem-se a duas molélulcas de mRNA não processado no citoplasma e migram para as regiões de acúmulo de glicoproteínas do complexo $(gp41+gp120)_3$. O domínio miristilado das poliproteínas Gag direciona uma ligação com a membrana celular e interage com a cauda cioplasmática da gp41. Aproximadamente 1200 a 2000 cópias de Gag se aproximam para formar uma partícula imatura que empacota as duas cópias do mRNA não processado.

Brotamento e maturação

A partícula imatura sai da célula por processo de brotamento. O ambiente extracelular é propício para a auto-ativação da protease. Uma vez ativada, ela promove a clivagem dos precursores Gag e Gag-Pol durante a maturação. A produção de vírus infecciosos depende do *assembly* de vários elementos estruturais fora da célula hospedeira. A ativação da protease dentro da célula levaria à clivagem das poliproteínas e liberação das proteínas estruturais, RT e integrase num ambiente que não permitiria o *assembly*.

A proteína Gag, que foi sintetizada em forma de uma proteína precursora p55, sofre adição de ácido mirístico e clivagens para originar as proteínas p24, p17, p9 e p6 na fase de maturação. A poliproteína precursora Gag-Pol (p160), sofre adição de ácido mirístico e clivagem originando as mesmas quatro proteínas p24, 017, p9 e p6, além das enzimas RT, protease e integrase.

As proteínas estruturais sofrem rearranjos e a partícula viral ou virion assume seu formato característico.

3.1.4. ORIGEM DE FORMAS RECOMBINANTES

A recombinação genética é parte dos mecanismos naturais de replicação retroviral (Clavel et al, 1989). Provavelmente é responsável por mediar o reparo dos genomas defectivos (Boulerice et al, 1991) e pode acelerar a dispersão de mutações benéficas na quasispécie viral. Este potencial de variação pode conferir a capacidade de rapidamente responder às alterações da pressão seletiva (Mouth et al, 1996) exercidas pelo sistema imunológico ou pela terapia antirretroviral.

A formação de retrovirus recombinantes requer que dois vírus infectem uma mesma célula, simultaneamente ou sequencialmente, em múltiplos eventos de transmissão (Fig.3.3.) (Nájera et al, 2002). Outra condição necessária é o co-empacotamento dos dois genomas de RNA em cada virion, o que leva à produção de partículas virais heterozigotas. A formação do genoma recombinante pode ocorrer no próximo ciclo de infecção através da leitura alternada (*alternate jumps*) executada pela transcriptase reversa entre ambos os genomas co-empacotados.

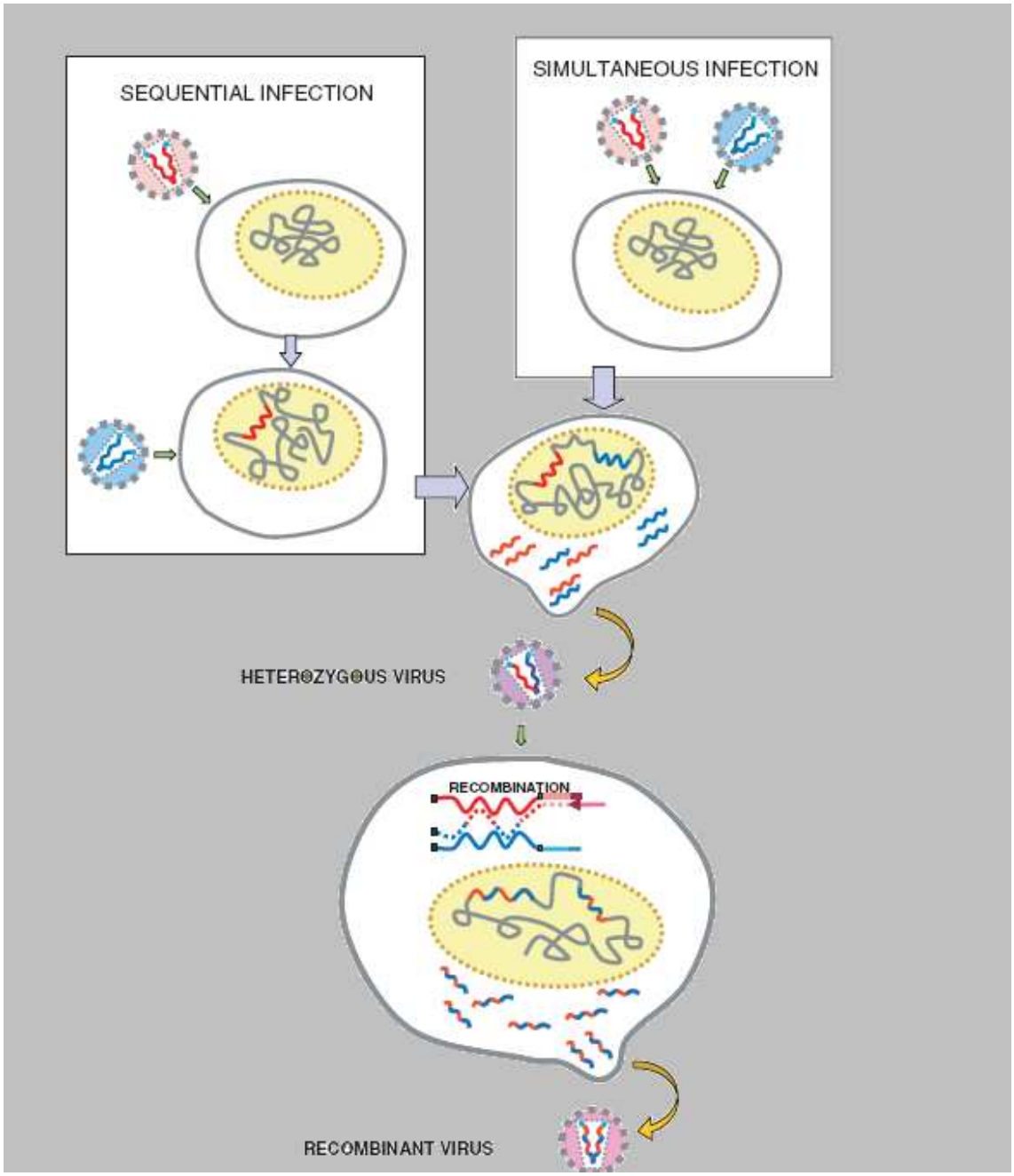


Figura 3.3. Replicação do HIV-1 e origem de formas recombinantes
(Nájera et al, 2002)

Atualmente, inúmeras formas recombinantes circulantes ou CRFs já foram descritas (Fig.3.4.). Para definir uma CRF é necessário que três vírus epidemiologicamente não ligados, com estruturas coincidentes e agrupamento filogenético consistente sejam caracterizados, sendo pelo menos dois em genoma completo. Estas formas recombinantes, bem como as formas recombinantes únicas ou

URF, resultam da presença concomitante de diferentes subtipos numa mesma região geográfica.

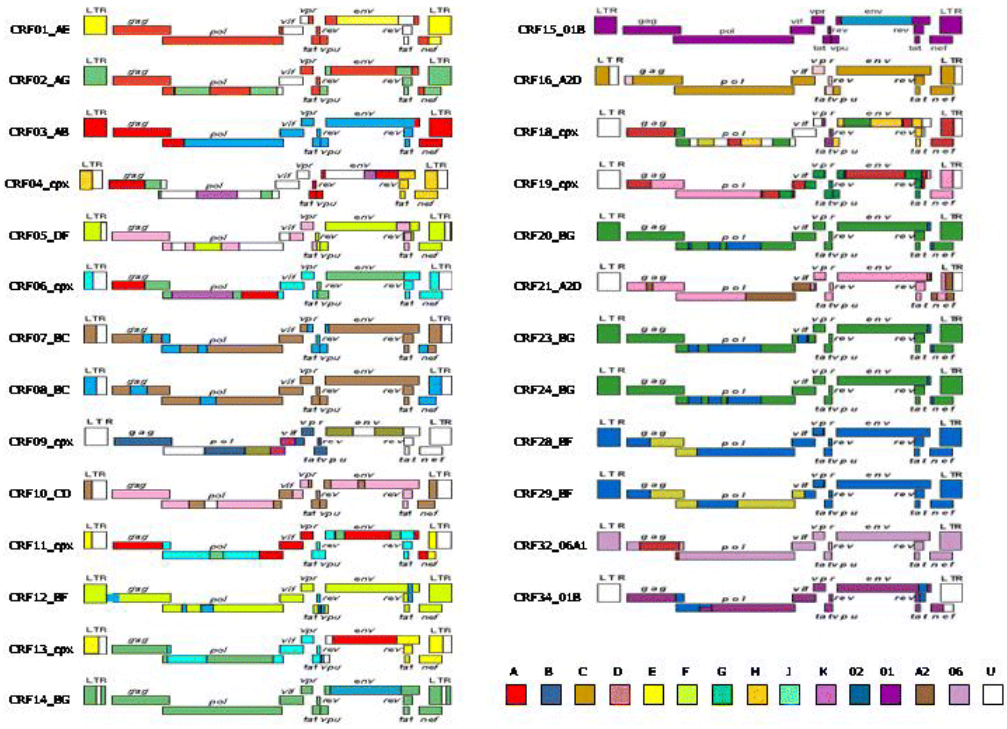


Figura 3.4. Formas recombinantes circulantes de HIV-1 (Buonaguro et al, 2007).

3.2. EVOLUÇÃO DA INFECÇÃO PELO HIV, TERAPIA ANTIRRETROVIRAL E FALHA TERAPÊUTICA.

3.2.1. EVOLUÇÃO DA INFECÇÃO

A infecção pelo HIV pode ser dividida em fase aguda e fase crônica. Uma das principais diferenças entre elas diz respeito à carga viral. Na fase aguda, o valor da carga viral pode atingir 1.000.000 copias/ml. Após um período variável, ocorre um set-point e esse valor geralmente diminui. Quanto mais alta a carga viral no set-point, mais facilmente o indivíduo desenvolve sintomas clínicos (Fraser et al, 2007).

A transmissão está significativamente relacionada ao valor da carga viral. Quando este valor está abaixo de 1.700 copias/ml, as chances de transmissão são muito baixas. Desta forma, a fase aguda assume um importante papel na transmissão, uma vez que a carga viral nesse período é sempre elevada (Wawer et al, 2005). No entanto, como não é raro que indivíduos infectados ainda mantenham cargas virais altas mesmo após a redução característica que sucede o set-point, eventos de transmissão também são possíveis na fase crônica. Outro fator de interferência no processo é a presença de doenças sexualmente transmissíveis, que podem aumentar os níveis de carga viral na região genital sem que ocorra um aumento detectável no sangue. Indivíduos infectados por essas doenças têm altas chances de transmitir o virus mesmo na fase crônica (Chakraborty et al, 2001; Serwada et al, 2003).

A evolução da infecção pelo HIV é marcada por um relevante evento com consequências dramáticas para o indivíduo: a depleção das células T CD4+, alvos principais da infecção viral e que exercem importante papel no estabelecimento e maximização das capacidades do sistema imune. O número de células CD4 pode

diminuir a um ponto tal que sua população fica muito reduzida para reconhecer a completa gama de antígenos que potencialmente poderiam ser detectados, permitindo o aparecimento de infecções oportunistas (Stevenson, 2003).

3.2.2. TERAPIA ANTIRRETROVIRAL (TARV)

O objetivo da TARV na infecção pelo HIV é a redução da capacidade replicativa do vírus ao mais baixo nível possível, prevenindo a infecção de novas células e maiores danos ao sistema imune. Isso significa reduzir a carga viral a níveis abaixo do atual nível de detecção de 50 cópias/ml (Hammer et al, 2006) e manter o nível de supressão pelo maior tempo possível para prevenir infecções oportunistas e elevar a qualidade de vida dos pacientes.

Até recentemente, os dois parâmetros, carga viral e contagem de células T CD4+, eram utilizados para determinar a necessidade de tratamento. O início da terapia era recomendado a todos os pacientes sintomáticos e para todos os não-sintomáticos em que a contagem de células T CD4+ tivesse atingido o nível de 200/µL, ou quando a carga viral estivesse em níveis elevados (>100.000cópias/mm3).

Em 2008, o Ministério da Saúde estabeleceu um novo consenso sobre os parâmetros a serem observados para o início da TARV. As novas determinações baseiam-se no fato de que a contagem de linfócitos T CD4+ estabelece o risco de progresso para AIDS e morte; portanto, é o indicador laboratorial mais importante em pacientes assintomáticos, para definir o momento de iniciar o tratamento. Para esse grupo de pacientes, a carga viral passa a ter maior importância como auxiliar para a decisão sobre o início do tratamento quando a contagem de linfócitos T CD4+ estiver entre 350 e 200 células/mm3. Segundo o atual documento sobre as recomendações para

início da terapia antirretroviral para adultos e adolescentes (Recomendações para Terapia Anti-retroviral em Adultos Infectados pelo HIV: Ministério da Saúde, 2008), a TARV somente deve ser iniciada com as devidas avaliações clínicas e laboratoriais, determinando-se o grau de imunodeficiência existente e o risco de progressão para a doença. Com a utilização da contagem de CD4+ como marcador principal, não há necessidade de avaliações de carga viral plasmática com intervalo inferior a 6 meses para indivíduos que não iniciaram o tratamento e com contagens de CD4 acima de 500 células/mm3. O início da TARV está indicado para pacientes assintomáticos com CD4+ entre 200 e 350cel/mm3 (quanto mais próxima de 200 cel/mm3, maior o risco de progressão para a AIDS, principalmente se associado à carga viral plasmática elevada). A TARV continua indicada para todos os pacientes sintomáticos, independente dos parâmetros laboratoriais.

3.2.3. CLASSES DE ANTIRRETROVIRAIS E ESCOLHAS PARA OS ESQUEMAS INICIAIS

As drogas de ação antirretroviral utilizadas hoje para suprimir a replicação do HIV tem diferentes alvos moleculares e são capazes de afetar o ciclo viral em suas diversas fases. Os antirretrovirais estão divididos em cinco classes, sendo duas de inibidores de transcriptase reversa (ITRN e ITRNN), uma de inibidores de protease, uma de inibidores de entrada e uma de inibidores de integrase. As duas últimas classes contam com dois e um inibidores aprovados, respectivamente, e diversas drogas ainda em fase de testes.

ITRNs

Os inibidores de transcriptase reversa análogos de nucleosídeos (ITRNs) são drogas em que a estrutura química é constituída de uma versão modificada de um nucleosídeo natural. Esses compostos suprimem a replicação dos retrovírus através da interferência na ação da enzima transcriptase reversa. Os inibidores dessa classe são precursores que sofrem trifosforilação pelas enzimas da célula hospedeira. As formas trifosfatadas competem com os deoxinucleosídeos naturais (dNTPs) pela incorporação nas novas cadeias de DNA sintetizadas, onde provocam a terminação prematura destas cadeias. Há vários ITRNs em uso: zidovudina (AZT), didanosina (DDI), zalcitabina (ddC), estavudina (d4t), lamivudina (3TC), abacavir (ABV), e tenofovir (TDF), este último sendo o único análogo de nucleotídeo aprovado e com ação semelhante aos análogos nucleosídicos (http://aids.about.com/od/hivaidslettern/g/nrt.htm).

ITRNNs

Inibidores de transcriptase reversa não análogos de nucleosídeos (ITRNNs) são um grupo com compostos estruturalmente diversos que se ligam ao sítio catalítico da transcriptase reversa do HIV-1. São bastante específicos e não tem ação contra o HIV-2. Como inibidores não competitivos da transcriptase reversa, sua atividade antirretroviral é aditiva ou sinergética com a maioria dos outros antirretrovirais. Entretanto, interações entre drogas devem ditar ajustes de dosagem com inibidores de protease (http://aids.about.com/od/hivaidslettern/g/nrt.htm).

IP

Os inibidores de protease (IPs) pertencem a uma classe de compostos peptidomiméticos que se ligam ao sítio ativo da enzima. A protease do HIV-1, responsável pelo processo pós-translacional das poliproteínas virais Gag e Gag-pol, transformando-as nas proteínas estruturais e enzimas presentes no virion, é uma protease aspártica composta por dois monômeros não covalentemente associados e idênticos entre si de 99 aminoácidos em comprimento. Seu sítio ativo assemelha-se ao de outras proteases aspárticas e contem a tríade Asp-Thr-Gly nas posições 25-27 (Turner & Summers, 1999). O sítio hidrofóbico de clivagem reconhece e cliva nove diferentes seqüências para produzir a matriz, o capsídeo, o nucleocapsídeo e as proteínas p6 da poliproteína Gag, além da protease, TR e integrase a partir da poliproteína Gag-pol (Erickson et al, 1999). A enzima contem uma região flexível conhecida como flap que se fecha sobre o sítio ativo após a ligação com o substrato.

Os IPs bloqueiam a produção de vírus infecciosos de células infectadas através da inibição da clivagem das poliproteínas precursoras necessárias para produzir virions. As partículas produzidas por células tratadas com IPs contem precursores não processados e são não-infecciosas (Peng et al., 1989).

O quadro 3.1 mostra a relação de inibidores de protease em uso e suas especificações. Recomenda-se que os IPs sejam utilizados em combinação com pelo menos outras duas drogas antirretrovirais para um tratamento eficaz da infecção por HIV.

Inibidores de entrada

A entrada do HIV na célula humana necessita de uma sequência de eventos que envolvem a ligação com as proteínas de superfície gp120 ao receptor CD4; alterações conformacionais na gp120, o que aumenta sua afinidade por um correceptor e expõe a gp41; a ligação da gp120 a um correceptor, CCR5 ou CXCR4; a penetração da membrana da célula pela gp41, o que aproxima a membrana do HIV e a célula T promovendo sua fusão; finalmente a entrada nos componentes virais no interior da célula. Os inibidores de entrada atuam interferindo em algum aspecto deste processo (Biswas et al, 2007).

Maraviroc e enfuvirtide são duas drogas dessa classe de antirretrovirais aprovadas para utilização nos esquemas terapêuticos. Maraviroc liga-se ao correceptor CCR5, prevenindo a interação com a proteína gp120 (Pugach et al, 2008). Enfuvirtide, também conhecido como inibidor de fusão, liga-se à proteína gp41 e interfere com sua habilidade para aproximar o envelope viral à membrana celular. Enfuvirtide é indicado para o tratamento anti-HIV em terapia combinada com outros antirretrovirais para pacientes em que os demais tratamentos tenham falhado.

Quadro 3.1. Inibidores de protease.

Nome comercial	**Nome genérico**	**Abreviação**	**Código Experimental**	**Companhia farmacêutica**
Agenerase	amprenavir	APV	141W94 or VX-478	GlaxoSmithKline and Vertex
Aptivus	tipranavir	TPV	PNU-140690	Boehringer Ingelheim
Crixivan	indinavir	IDV	MK-639	Merck & Co
Invirase	saquinavir	SQV	Ro-31-8959	Hoffmann-La Roche
Kaletra	lopinavir + ritonavir	LPV/r	ABT-378/r	Abbott Laboratories
Lexiva	fosamprenavir	FPV	GW-433908 or VX-175	GlaxoSmithKline
Norvir	ritonavir	RTV	ABT-538	Abbott Laboratories
Prezista	darunavir	DRV	TMC-114	Tibotec
Reyataz	atazanavir	ATV	BMS-232632	Bristol-Myers Squibb
Viracept	nelfinavir	NFV	AG-1343	Pfizer

Inibidor de integrase

A integrase é a enzima necessária para que o DNA viral seja inserido com sucesso no DNA da célula hospedeira humana, processo necessário para que o vírus possa utilizar o maquinário celular para produzir novos virions. Inibidores de integrase bloqueiam a ação da enzima e previnem este processo. Em outubro de 2007, raltegravir (Isentress; Merk), uma pequena molécula capaz de ligar-se à integrase, tornou-se a primeira droga desta classe aprovada pelo US FDA para tratamento da infecção por HIV em combinação com outras drogas (Steven et al, 2008). Sua utilização é indicada para pacientes adultos que já tenham sido tratados e com evidências da presença de cepas resistentes a múltiplas drogas antirretrovirais.

Esquema terapêutico inicial

A decisão sobre a necessidade do início do tratamento deve ser seguida por outra, de igual importância: a escolha do melhor esquema terapêutico, ou seja, aquele que dará ao paciente a chance de suprimir a replicação viral e elevar a contagem de CD4+ a níveis que permitam restabelecer as funções imunológicas com um mínimo de efeitos colaterais. Nesse sentido, a escolha do esquema mais adequado precisa ser individualizada e levar em conta diversos fatores, entre eles, dificuldades para aderência e possibilidade de pré-existência de mutações de resistência. Tanto o guia americano contendo indicações para o tratamento de adultos e adolescentes (USA Panel on Antiretroviral Guidelines for Adults and Adolescents, 2008) como o guia brasileiro (Recomendações para Terapia Anti-retroviral em Adultos Infectados pelo HIV: Ministério da Saúde, 2008), recomendam especial cuidado na conscientização dos

pacientes quanto aos riscos de um tratamento não seguido e com interrupções frequentes.

Via de regra, a recomendação dos guias de tratamentos anti-HIV é no sentido da utilização de um esquema baseado na combinação de três drogas. A primeira sugestão é o uso de dois ITRNs e um ITRNN (geralmente, efavirenz). Outra alternativa é a combinação de dois ITRNs e um IP potencializado com RTV (IP/r). Ensaios clínicos randomizados avaliando a terapia inicial em 48 semanas de seguimento mostram equivalência na proporção da resposta virológica ao esquema inicial entre pacientes que recebem 2 ITRN + ITRNN (efavirenz) e 2 ITRN +IP/r (Bartlett et al. AIDS 2006). As exceções parecem ser os esquemas contendo IDV ou NFV. Alguns estudos apontam que a resposta a esquemas iniciais contendo qualquer desses IP é inferior à resposta obtida com esquemas baseados em efavirenz (Staszewski et al., 1999; Robbins et al., 2003) e, por esta razão, ambos os IPs são desaconselhados para esquemas iniciais pelo USA Panel on Antiretroviral Guidelines for Adults and Adolescents.

Um importante fator a se considerar é a dificuldade na manutenção do esquema em função das dosagens, quantidade de pílulas, interação com alimentos e medicamentos. Nesse sentido, esquemas que utilizam 2 ITRN +ITRNN são, em geral, de posologia mais simples, o que provavelmente facilita a adesão ao tratamento. Por outro lado, Os inibidores de protease potencializados com ritonavir (IP/r) oferecem maior barreira genética à resistência do que os inibidores de transcriptase reversa não-análogos de nucleosídeos (ITRNN). Na prática, isso significa que para que se desenvolva resistência a um IP/r, há necessidade de um número maior de mutações do que para o desenvolvimento de resistência a ITRNN. De fato, a resistência a qualquer IP/r resulta do acúmulo de mutações, enquanto apenas uma mutação para ITRNN confere resistência completa ao efavirenz e à nevirapina. O estudo que comparou

lopinavir/r com efavirenz em esquema de terapia inicial mostrou que a falha virológica (decorrente de mutações de resistência) foi mais freqüente com o esquema baseado em efavirenz (Riddler et al., 2006).

3.2.4. FALHA TERAPÊUTICA

O sucesso do esquema terapêutico é verificado através dos dados clínicos do paciente e de coletas periódicas para o monitoramento da carga viral e da contagem de células T CD4. Considera-se que o paciente encontra-se em falha terapêutica quando apresenta sintomas característicos de AIDS e/ou carga viral detectável e baixos níveis de CD4.

Apesar de um elevado número de indivíduos atingir os objetivos da TARV e conseguir manter a supressão viral com consequente decréscimo da carga viral a níveis indetectáveis, muitos não conseguem a mesma resposta. Fala-se em falha virológica primária quando o paciente não atinge satisfatoriamente a supressão viral após alguns poucos meses de tratamento. Em média, entre 10 e 20% dos pacientes que iniciam o tratamento apresentam falha primária (Ministério da Saúde).

Mesmo pacientes com boa resposta inicial podem apresentar falha virológica no futuro. Quando a falha ocorre após um ano de tratamento, é chamada falha virológica secundária. Esse tipo de falha pode atingir de 20 a 50% dos indivíduos com boa resposta inicial.

Indivíduos em falha terapêutica são frequentemente submetidos ao teste de genotipagem. Este teste permite verificar a presença de mutações associadas à resistência aos antirretrovirais nos vírus circulantes (Shafer, 2002). O resultado do teste

de genotipagem permite a escolha de um novo esquema terapêutico para o qual as cepas virais presentes no indivíduo não sejam resistentes.

3.2.5. ADERÊNCIA E EMERGÊNCIA DE MUTAÇÕES DE RESISTÊNCIA AOS ANTIRETROVIRAIS

Muitas razões podem explicar os elevados índices de falha terapêutica. Certamente, um dos mais importantes é a incompleta adesão ao tratamento (Taylor-Castillo et al, 2005). A taxa média de adesão à TARV é de, aproximadamente, 70%, enquanto a ideal para a supressão viral deveria ser, pelo menos, de 95% (Machtinger & Bagsberg, 2005). A não aderência está claramente relacionada ao desenvolvimento de cepas resistentes aos antirretrovirais. Como mencionado anteriormente (capítulo 1), o HIV-1 tem uma alta taxa de mutação e replica-se muito rápido e não possui mecanismos eficientes de correção de erro durante a replicação viral. Mutações podem aparecer facilmente e se uma determinada mutação confere alguma vantagem, pode ser fixada e transmitida à próxima geração. A presença de drogas antirretrovirais representa uma enorme pressão seletiva. Qualquer mutação que possa conferir algum nível de resistência às drogas representa uma grande vantagem ao vírus. Somente a completa supressão viral seria capaz de prevenir o aparecimento e a fixação de mutações de resistência. Isto é possível somente com altos níveis de aderência ao tratamento.

3.2.6. QUASISPECIES VIRAIS: COMPLEXO SISTEMA ADAPTATIVO

As elevadas taxas de replicação e de mutação somadas à baixa atividade de correção de erro no HIV-1 e em outros vírus de RNA resultam numa complexa distribuição de genomas estreitamente relacionados denominada quasispecies (Eigen,

1996). A quasispecies viral comporta-se como um engenhoso sistema adaptativo influenciado pelo “ensemble” de mutantes que num dado tempo compõem a população em evolução. Tal influência resulta da presença de genomas de memória, componentes minoritários dentro do espectro de mutantes que já foram dominantes em outras fases da evolução da quasispecie (Briones et al, 2003).

Segundo Briones e colaboradores (2003), pacientes com HIV-1 apresentam dois tipos de memória de acordo com a dinâmica da quasispecie: memória replicativa e memória celular ou anatômica. A primeira refere-se aos níveis de memória ditados pela competição de crescimento (*growth competition*) de genomas dominantes e de memória. A memória celular ou anatômica deriva da fase integrativa do ciclo do retrovírus (em células contendo DNA proviral, sequências ancestrais de HIV são produzidas e podem entrar numa replicação competitiva com outros genomas de HIV-1).

No HIV-1, os níveis de memória variam de 0,1% a 10% dos componentes do espectro de mutantes (Briones et al, 2003). Estes níveis poderiam conferir vantagens às populações virais para que pudessem responder eficientemente a pressões seletivas previamente experimentadas durante as infecções crônicas, e podem facilitar a re-emergência de genomas virais ancestrais (Garcia-Lerma et al, 2001). Alguns trabalhos apontam a presença de subpopulações minoritárias dentro da quasispecies como um mecanismo que poderia aumentar a emergência de resistência às drogas em alguns pacientes submetidos à TARV (Yeni et al, 2004; Vandamme et al, 2004).

3.2.7. RESISTÊNCIA PRIMÁRIA E RESISTÊNCIA SECUNDÁRIA: MAIS ALGUMAS CONSIDERAÇÕES SOBRE A TRANSMISSÃO DO HIV-1

Em pacientes virgens de tratamento e não infectados com cepas resistentes, duas situações podem ser esperadas: ou não há cepas portadoras de mutações de resistência ou estas cepas já estão presentes, mas constituem populações minoritárias e dificilmente detectadas na quasispecies. Se, por qualquer razão, a supressão viral não for completa, poderá ocorrer a emergência e a fixação de mutações de resistência, fato perfeitamente possível levando-se em conta as características do HIV-1 já citadas (elevada taxa de mutação, velocidade de replicação, ausência de mecanismos eficientes de reparo, recombinação). No caso de já estarem presentes constituindo populações minoritárias, as cepas portadoras de resistência podem ser selecionadas e passarem a constituir as populações dominantes. De qualquer forma, as mutações estariam sendo selecionadas durante o tratamento e a resistência conferida por elas nesta situação denomina-se **resistência secundária** às drogas.

O acesso universal à TARV trouxe nova esperança e melhor qualidade de vida aos portadores de HIV. Contudo, a seleção de mutações de resistência durante o tratamento e o concomitante aumento da carga viral nos indivíduos em falha terapêutica pode aumentar a possibilidade de esses indivíduos transmitirem cepas resistentes. Consequentemente, os novos infectados, embora virgens de tratamento, seriam portadores de vírus já potencialmente resistentes a, pelo menos, algumas das drogas disponíveis. Esta resistência pré-existente e transmitida é chamada de **resistência primária.**

A prevalência da resistência primária tem sido estudada em muitos países. No Brasil, apesar do acesso universal à TARV, as taxas de resistência primária têm se mantido estáveis (Soares et al, 2004; Rodrigues et al, 2006). Isso, no entanto, não é a

regra e altas prevalências de mutações de resistência transmitidas são observadas em outros países (Maslin et al, 2005).

3.2.8. MUTAÇÕES PRINCIPAIS E MUTAÇÕES SECUNDÁRIAS

A elevada variabilidade do HIV-1 torna praticamente impossível a existência de uma cepa selvagem padrão. No entanto, para estudos de resistência aos antirretrovirais, definem-se mutações como diferenças em relação a uma ou várias sequências referência. As referências mais comumente utilizadas são a HXB2 e NL43, além de uma sequência referência consenso com os aminoácidos mais comuns para cada posição em vírus selvagens do subtipo B (consenso do subtipo B). Todas essas sequências são muito semelhantes, divergindo apenas em uns poucos aminoácidos não envolvidos com resistência às drogas. O uso de referências B relaciona-se ao fato de ser este o subtipo de maior prevalência na Europa e EUA. Contudo, a utilização de referências de outros subtipos torna-se necessária, à medida que estes aumentam em prevalência e muito pouco se conhece a respeito de seus perfis mutacionais, principalmente no que se refere à resistência e resposta ao tratamento.

Há diversos painéis e bases de dados que compilam as mutações associadas à resistência às drogas. Todos estes painéis frisam que as mutações descritas em suas tabelas são predominantemente encontradas e estudadas no subtipo B. Clark e colaboradores (2006/2007) mostram uma tabela com 947 mutações em função de combinações de drogas, atualizada até fevereiro de 2007, conseqüência de uma extensa revisão das bases de dados disponíveis.

Apesar da deficiência de dados específicos relacionados aos diferentes subtipos de HIV-1, de um modo geral, a classificação das mutações associadas à resistência às

drogas antirretrovirais empregada para o subtipo B é utilizada para os demais subtipos. Via de regra, são denominadas mutações principais (*major mutations*) aquelas capazes de conferir elevados níveis de resistência a uma ou várias drogas de uma determinada classe. Estas mutações são normalmente selecionadas durante a terapia, principalmente em consequência de baixa aderência, ou de qualquer outro fator que leve à não supressão completa do vírus. A presença delas em cepas selvagens ou em isolados de pacientes virgens de tratamento é rara, pois são responsáveis por alterações conformacionais menos vantajosas para as proteínas onde são encontradas. Durante a terapia, no entanto, constituem vantagem adaptativa. Um segundo grupo de mutações é normalmente selecionado durante o tratamento: as mutações secundárias ou acessórias. Este grupo é constituído por mutações que não conferem níveis significativos de resistência às drogas quando presentes isoladamente. No entanto, quando presentes em grupos, compõem padrões que acompanham mutações principais específicas, compensando a perda de *fitness* que estas geralmente impõem ao vírus. Algumas dessas mutações secundárias, quando em grupo, também são capazes de conferir algum grau de resistência aos inibidores, levando à diminuição da susceptibilidade viral.

A classificação em principal ou secundária pode ser utilizada para ITRNs, ITRNNs e para IPs.

3.2.9. POLIMORFISMOS

Algumas mutações de resistência às drogas ocorrem comumente na ausência da pressão seletiva exercida pelos antirretrovirais e são consideradas polimorfismos naturais. Entre as mutações mais conhecidas deste tipo estão a acessórias aos IPs nas posições 10, 20, 36, 63, 71, 77 e 93 que, dependendo do subtipo, ocorrem entre 5 e

>50% dos indivíduos não tratados. Além destas, outros polimorfismos subtipo-específicos podem ocorrer. A descrição destes polimorfismos é importante para diferenciá-los de mutações não polimórficas que tenham sido transmitidas. Algumas mutações não polimórficas de resistência podem parecer polimórficas em análises que contenham uma grande quantidade de pessoas não tratadas que possam ter sido infectadas com vírus resistentes. Essa é uma ocorrência comum em áreas com acesso a drogas ARV há vários anos. E estas mutações não polimórficas seriam importantes indicadores de resistência transmitida.

3.2.10. ROTAS MUTACIONAIS (*PATHWAYS*)

A pressão seletiva exercida por uma determinada droga antiretroviral, em especial os IP, pode levar o vírus a desenvolver um grupo de mutações específicas associadas à resistência (Molla et al, 1996), constituindo uma espécie de perfil. As mutações que constituem cada perfil vão sendo acumuladas passo a passo seguindo uma rota mutacional (*pathway*) ao longo de vários ciclos replicativos, responsáveis por aumentar gradativamente os níveis de resistência ao inibidor.

Um mesmo inibidor pode levar a rotas mutacionais distintas, ocasionando maiores ou menores dificuldades para a eficiência do tratamento ou mesmo para uma eventual terapia de resgate. Um exemplo da situação descrita é a seleção exercida pelo NFV. Tratamentos à base desse inibidor tendem a selecionar a mutação D30N ou a mutação L90M. Raramente ambas as mutações são observadas ao mesmo tempo. Quando a rota mutacional inclui a mutação D30N, a terapia de resgate em caso de falha terapêutica tem maiores chances, pois esta mutação praticamente não oferece resistência cruzada. Já a rota que envolve a substituição L90M dificulta muito a possibilidade de

uma terapia de resgate uma vez que oferece resistência cruzada a quase todos os antirretrovirais (Diaz, 2004).

3.3. ANÁLISE MULTIVARIADA

3.3.1. MÉTODOS APROPRIADOS PARA CADA SITUAÇÃO

Estabelecer relações e encontrar ou propor leis explicativas é um dos objetivos próprios de qualquer trabalho de natureza científica. Para tanto, há a necessidade de se controlar, manipular e medir as variáveis consideradas relevantes para a compreensão do fenômeno analisado. Em alguns casos, a análise isolada de poucas variáveis é suficiente para permitir inferências sobre a realidade. Quando, no entanto, um fenômeno depende de inúmeras variáveis, não basta conhecer informações estatísticas isoladas. É fundamental conhecer a totalidade dessas informações fornecidas pelo conjunto das variáveis. Nesses casos, análises estatísticas univariadas, como as baseadas em medidas de tendência central (média, mediana) ou medidas de dispersão (variância) são insuficientes ou até inadequadas.

Há inúmeros métodos de análise multivariada com finalidades bem diversas. A análise por agrupamento hierárquico (HCA) e a análise por componentes principais (PCA) fazem parte de uma categoria de métodos conhecidos em conjunto como análises de clusters e são indicados quando o objetivo é verificar como as amostras ou as variáveis se relacionam, ou seja, o quanto são similares (Moita-Neto, 2005).

3.3.2. ANÁLISE DE CLUSTERS: MATRIZES DE DADOS E MEDIDAS DE DISTÂNCIA ENTRE AS VARIÁVEIS

Uma análise de clusters se inicia com a construção de uma matriz de dados (Romesburg, 2004), como exemplificada na figura 3.5, a seguir:

		Objetos				
		1	2	3	4	5
Atributos	1	10	20	30	30	5
	2	5	20	10	15	10

Fig 3.5. Matriz de dados para análise de clusters.

Enquanto as colunas contem dados dos objetos dos quais se deseja estimar a similaridade, as linhas são os atributos ou propriedades desses objetos. Os objetos poderiam ser, por exemplo, diferentes bairros de uma cidade, e os atributos contagens de escolas, estabelecimentos comerciais, residências, farmácias, agências bancárias. A análise de clusters permite que sejam agrupados os objetos com maior grau de similaridade. No caso, bairros com perfis semelhantes podem ser considerados similares e colocados ou classificados num mesmo agrupamento.

Os grupos de mutações com tendência à covariação nos diferentes subtipos de HIV-1 podem ser estudados utilizando-se técnicas de análise de cluster, como as empregadas por Wu e colaboradores (2003) e Rhee e colaboradores (2007). Neste caso, os objetos são as mutações e os atributos correspondem à presença ou ausência de cada mutação nas sequências genômicas virais dos pacientes. A matriz a ser construída baseia-se em dados binários e as contagens são substituídas por zero e 1, que correspondem respectivamente à ausência ou à presença de uma determinada mutação (figura3.6).

Construída a matriz de dados, o próximo passo é a construção de uma segunda matriz, onde são computadas as distâncias entre cada par de objetos (mutações, por exemplo). Existem várias maneiras para calcular a distância entre dois pontos, sendo a mais conhecida e utilizada a chamada distância euclidiana, que corresponde ao sentido trivial de distância no plano. Quando uma determinada variável é tomada como um ponto no espaço de amostras, é possível calcular a distância deste ponto a todos os outros, relembrando que, para duas variáveis, corresponde à aplicação do teorema de Pitágoras (a2=b2+c2: o comprimento da hipotenusa, a, é igual à raiz quadrada da soma dos quadrados dos comprimentos dos catetos, b e c) (Moita-Neto, 2005). Como medida, pode-se utilizar um coeficiente de similaridade ou um coeficiente de dissimilaridade, de acordo com o que se quer enfatizar, pois, quanto menor o valor de um coeficiente de dissimilaridade, maior o grau de similaridade entre dois objetos.

	Mutações							
	L10I	G17E	L19V	K20R	K20T	L24I		T96S
seq1	1	0	0	0	0	1		0
seq2	0	0	0	0	0	0		0
seq3	1	0	0	1	0	0		1
seq4	1	0	0	1	0	0		0
Seqn	1	0	1	1	0	0		0

Figura. 3.6. Matriz de dados binários para análise de cluster das mutações que compõem agrupamentos no HIV-1.

3.3.3. ANÁLISE DE CLUSTERS: COEFICIENTE DE SIMILIARIDADE DE JACCARD

Há uma enorme variedade de coeficientes, cada qual com propriedades e aplicações diversas, que podem ser utilizados para a construção das matrizes de

distância. A escolha do mais apropriado vai depender, entre outras coisas, dos tipos de atributos dos objetos a serem analisados. Os atributos podem ser quantitativos ou qualitativos. Sendo qualitativos, podem ainda ser binários (positivo e negativo, presença e ausência, feminino e masculino) ou com muitos estados (verde, vermelho e amarelo, por exemplo). O cálculo do coeficiente de similaridade para atributos qualitativos binários baseia-se diretamente na matriz de dados (fig 3.6.) e, a partir dela, obtem-se tabelas de contingência 2 x 2 para cada par de objetos (mutação X e mutação Y, por exemplo) (tabela.3.1.). As caselas desta tabela de contingência mostram as contagens de duplas presenças (XY), presença de X e ausência de Y (X0), presença de Y e ausência de X (0Y) e duplas ausências (00).

Tabela 3.1. Tabela de contingência 2 x 2 contendo o número de duplas presenças (XY), presenças de X e ausências de Y, presenças de Y e ausências de X e duplas ausências.

		Mutação X	
		1	0
Mutação Y	1	XY = a	0Y = b
	0	X0 = c	00 = d

Os diversos coeficientes de similaridade para atributos qualitativos correspondem a diferentes funções dos valores a, b, c e d extraídos da Tabela 3.1. Rhee e colaboradores (2007) utilizaram o coeficiente de Jaccard (Jaccard, 1908) para analisar a similaridade entre diferentes mutações no HIV-1 subtipo B. Este coeficiente é calculado como a seguir:

$$J_{(XY)} = a/a+b+c \qquad 0{,}0 \leq J_{(XY)} \leq 1{,}0$$

O coeficiente de Jaccard indica máxima similaridade quando os dois objetos tem valores idênticos, b = c = 0 e J=1, e indica máxima dissimilaridade quando a=0 e J=0. Alguns coeficientes de similaridade incluem d, ou seja, o número de duplas ausências, em seus denominadores. O coeficiente de Jaccard pertence ao grupo de coeficientes que não incluem as duplas ausências, baseando-se na idéia de que um atributo duplamente ausente não contribui para o grau de similaridade entre dois objetos. Para Rhee e colaboradores (2007), este coeficiente utiliza somente aquelas sequências em que pelo menos uma das mutações de um dado par XY esteja presente, ou seja, não sofre a influência da presença de pares raros.

Os valores de J_{XY} permitem a obtenção dos coeficientes de dissimilaridade D para cada par de mutações X e Y através da relação:

$$D_{XY} = 1 - J_{XY}$$

3.3.4. ANÁLISE DE CLUSTERS: MATRIZES DE DISTÂNCIA

Tanto os valores de J_{XY} como os valores de D_{XY} podem ser utilizados para a construção da matriz de distância entre quaisquer mutações X e Y. A fig. 3.7 exemplifica uma matriz de distância baseada no coeficiente de similaridade de Jaccard (figura.3.7.a) para um grupo de mutações da região da protease do HIV-1 e outra baseada no coeficiente de dissimilaridade de Jaccard (figura.3.7.b):

	L10I	G17E	Q18E	L19V	K20R	K20T	L24I	D30N
L10I	1	0.021	0.005	0.016	0.308	0.047	0.214	0.019
G17E		1	0.000	0.000	0.034	0.012	0.045	0.038
Q18E			1	0.000	0.020	0.014	0.018	0.000
L19V				1	0.023	0.025	0.016	0.000
K20R					1	0.000	0.130	0.058
K20T						1	0.017	0.082
L24I							1	0.000
D30N								1

(a)

	L10I	G17E	Q18E	L19V	K20R	K20T	L24I	D30N
L10I	0	0.979	0.995	0.984	0.692	0.953	0.786	0.981
G17E		0	1.000	1.000	0.966	0.988	0.955	0.962
Q18E			0	1.000	0.980	0.986	0.982	1.000
L19V				0	0.977	0.975	0.984	1.000
K20R					0	1.000	0.870	0.942
K20T						0	0.983	0.918
L24I							0	1.000
D30N								0

(b)

Fig.3.7. Matriz de distância baseada no coeficiente de similaridade de Jaccard (a) e matriz de distância baseada no coeficiente de dissimilaridade de Jaccard para o mesmo conjunto de mutações da protease do HIV-1 (b).

3.3.5. ANÁLISE POR AGRUPAMENTO HIERÁRQUICO (HCA)

Através desta técnica, cada amostra é tratada como um ponto no espaço multidiensional descrito pelas variáveis. A técnica também pode ser aplicada invertendo-se o raciocínio: cada variável passa a ser tratada com um ponto no espaço multidimensional descrito pelas amostras. A partir da matriz de similaridade, constrói-se um diagrama denominado dendrograma de similaridade que corresponde ao

agrupamento matemático dos pontos no espaço bidimensional (Fig 3.8). Para se chegar ao dendrograma, inicia-se com cada elemento representando um grupo, e a cada passo, um grupo ou elemento é ligado a outro de acordo com sua similaridade, até o último passo, onde é formado um grupo único com todos os elementos.

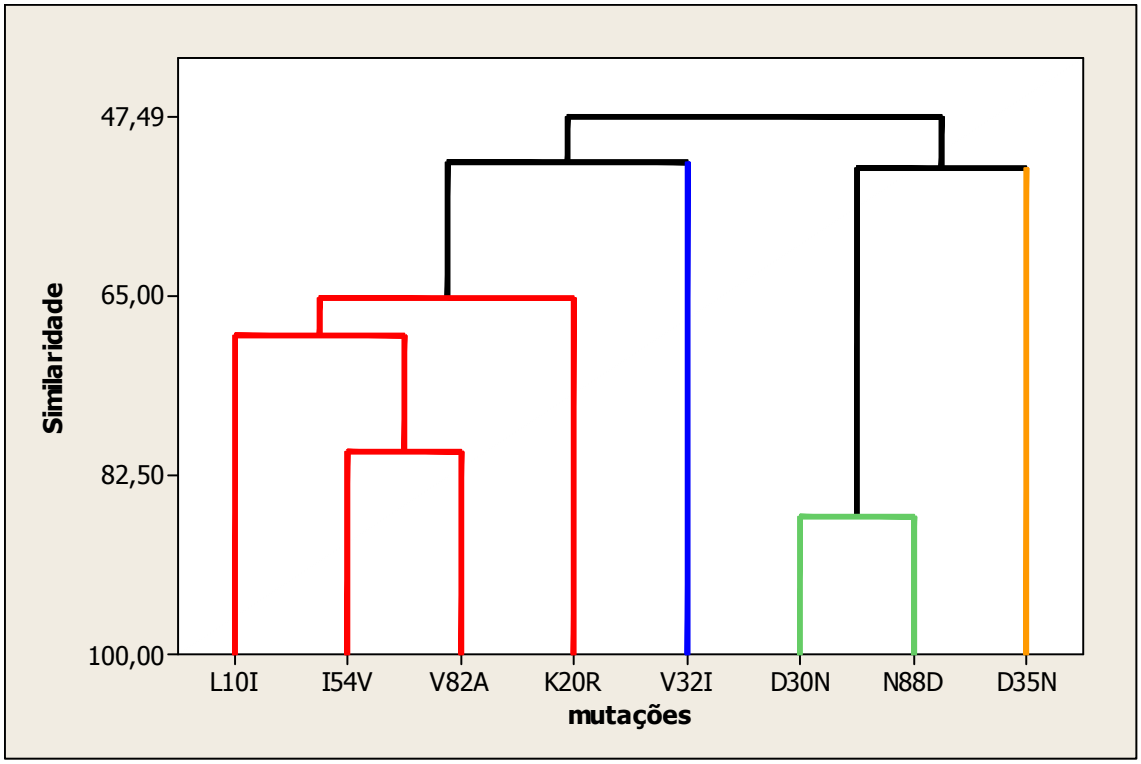


Fig.3.8. Exemplo de dendrograma mostrando agrupamentos segundo o nível de similaridade entre algumas mutações associadas ao tratamento na protease F do HIV-1. Observa-se que as mutações L10I, I54V, V82A e K20R formam um cluster com nível de similaridade em torno de 65,00; enquanto um outro cluster, formado pelas mutações D30N e N88D, apresenta um nível ainda maior de similaridade (>82,50).

Existem várias maneiras de aglomerar matematicamente estes pontos no espaço multidimensional para formar os agrupamentos hierárquicos. Cada um corresponde a um algoritmo específico (ou seja, o modo particular como os cálculos serão feitos pelo computador), que usa as informações da matriz de proximidade para criar um dendrograma de similaridade. A maioria dos métodos parecem ser formulações alternativas de três grandes conceitos de agrupamento aglomerativo (Anderberg, 1973):

1) Métodos de ligação (*single linkage*, *complete linkage*, *average linkage*, *median linkage*);

2) Métodos de centróide;

3) Métodos de soma de erros quadráticos ou variância (método de *Ward*).

De modo geral, os métodos aglomerativos utilizam os passos de um algoritmo padrão, conforme descrito na figura 3.9. A diferença entre eles ocorre no passo 5, ou

seja, a função distância é definida conforme o método escolhido ou mais apropriado (Johnson & Wichern, 1992).

Entrada: Uma base de dados com ***N*** elementos (ou variáveis).
Saída: Um conjunto de grupos.
1. Iniciar com ***N*** grupos, contendo um elemento em cada grupo e uma matriz de similaridade ***DN´N***;
2. Repetir;
3. Localizar a menor distância ***dUV*** (maior similaridade);
4. Atualizar a matriz ***D***, retirando os elementos ***U*** e ***V***;
5. Atualizar a matriz ***D***, adicionando as novas distâncias do grupo ***(U,V)***;
6. Até ***N-1***, quando todos elementos estarão em um único grupo.

Fig. 3.9. Algoritmo padrão

Empiricamente, os melhores resultados para análise de agrupamentos entre as diferentes mutações associadas ao tratamento e a análise de agrupamentos entre diferentes mutações e uso de determinados IPs foram obtidos utilizando-se o método *Average Linkage* ou ligação por média. Nesse método, a função distância é definida por:

$$d_{(UV)W} = (N_u.d_{UW} + N_v.d_{VW})/N_u.N_v$$

onde: N_U e N_V são os números de elementos no grupo *U* e *V*, respectivamente; d_{UW} e d_{VW} são as distâncias entre os elementos *UW* e *VW*, respectivamente. Esse método apresenta como características principais a produção de bons resultados tanto para distâncias Euclidianas quanto para as outras distância, menor sensibilidade a ruídos em relação a outros métodos e tendência a formar grupos com número de elementos similares (Kaufmann & Rousseeuw, 1990).

Através do dendograma e do conhecimento prévio sobre a estrutura dos dados, deve-se determinar uma distância/similaridade de corte para definir quais serão os grupos formados. Essa decisão é subjetiva, e deve ser feita de acordo com o objetivo da análise e o número de grupos desejados. No caso representado pelo exemplo da figura

3.4, há informações a partir de dados empíricos e de literatura sobre perfis de mutações associadas à resistência aos IPs (Wu et al, 2003; Rhee et al, 2007).

A interpretação de um dendrograma de similaridade entre amostras fundamenta-se na intuição: duas amostras próximas devem ter também valores semelhantes para as variáveis medidas. Ou seja, elas devem ser próximas matematicamente no espaço multidimensional. Portanto, quanto maior a proximidade entre as medidas relativas às amostras, maior a similaridade entre elas. O dendrograma hierarquiza esta similaridade de modo que podemos ter uma visão bidimensional da similaridade ou dissimilaridade de todo o conjunto de amostras ou variáveis utilizado no estudo. Quando o dendrograma construído é das variáveis, a similaridade entre duas variáveis aponta forte correlação entre elas dentro do conjunto de dados estudado (Moita-Neto, 2005).

A aplicação da análise de agrupamento hierárquico, quando temos variáveis de escalas diferentes, deve ser precedida por um tratamento prévio dos dados. Quando não é feito o pré-tratamento, as variáveis com valores numéricos mais altos serão mais importantes no cálculo que as variáveis com valores numéricos mais baixos. O pré-tratamento mais comumente empregado é a transformação Z, que transforma as medidas de cada variável de tal modo que o conjunto de dados tenha média zero e variância um. A finalidade deste procedimento é equalizar a importância estatística de todas as variáveis utilizadas. Quando a matriz contem dados binários, esta transformação não é necessária.

3.3.6. ANÁLISE POR COMPONENTES PRINCIPAIS (PCA)

A PCA é uma técnica estatística robusta que pode ser utilizada para reduzir o número de variáveis e para fornecer uma visão estatisticamente privilegiada do conjunto de dados. A PCA permite identificar as variáveis mais importantes no espaço das componentes principais.

A análise de componentes principais consiste em reescrever as variáveis originais em novas variáveis denominadas componentes principais, através de uma transformação de coordenadas. A transformação de coordenadas é um processo trivial quando feito usando matrizes. A transformação matemática das coordenadas pode ser feita de diversas maneiras conforme o interesse.

Os componentes principais são as novas variáveis geradas através de uma transformação matemática especial realizada sobre as variáveis originais. Esta operação matemática está disponível em diversos softwares estatísticos especializados. Cada componente principal é uma combinação linear de todas as variáveis originais. Por exemplo, um sistema com oito variáveis, após a transformação, terá oito componentes principais. Cada uma destas componentes principais, por sua vez, será escrita como uma combinação linear das oito variáveis originais. Nestas combinações, cada variável terá uma importância ou peso diferente (Moita-Neto, 2005).

Duas são as características das componentes principais que as tornam mais efetivas que as variáveis originais para a análise do conjunto das amostras (Prado et al, 2002). As variáveis podem guardar entre si correlações que são suprimidas nas componentes principais. Ou seja, as componentes principais são ortogonais entre si. Deste modo, cada componente principal traz uma informação estatística diferente das outras. A segunda característica importante é decorrente do processo matemático-

estatístico de geração de cada componente que maximiza a informação estatística para cada uma das coordenadas que estão sendo criadas. As variáveis originais têm a mesma importância estatística, enquanto que as componentes principais têm importância estatística decrescente. Ou seja, as primeiras componentes principais são tão mais importantes que podemos até desprezar as demais.

A figura 3.10 ilustra a PCA do mesmo grupo de mutações associadas ao tratamento mostradas na figura 3.8 (L10I, K20R, D30N, V32I, D35N, I54V, V82A e N88D). Observa-se que os agrupamentos neste caso foram bem semelhantes.

3.3.7. COMPARAÇÃO ENTRE HCA E PCA

A análise de componentes principais e a análise de agrupamento hierárquico são técnicas de análise multivariada com fundamentos teóricos bem diferentes, podendo ser aplicadas independentemente. Estas técnicas podem até ser complementares na informação sobre o conjunto de dados, dependendo do sistema analisado. Ambas fornecem a visão mais global possível das amostras dentro do conjunto de dados, conforme as variáveis usadas (Cazar, 2003).

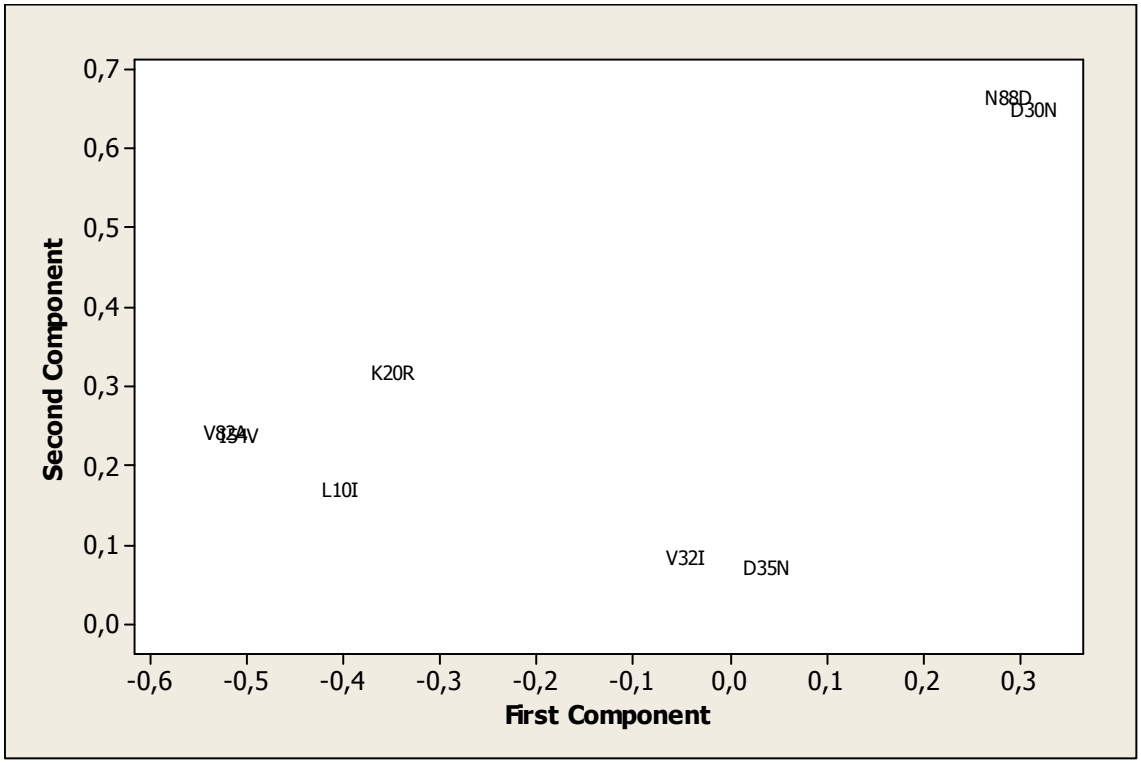


Fig 3.10. PCA para as mutações L10I, K20R, D30N, V32I, D35N, I54V, V82A, N88D. O par de mutações I54V/V82A aparece próximo às mutações L10I e K20R, ao mesmo tempo em que está bem distante do par D30N/N88D.

CAPÍTULO 4. CASUÍSTICA E MÉTODOS

4.1. CASUÍSTICA

O presente estudo analisou sequências correspondentes à região da protease de HIV-1 subtipo F e comparou os resultados aos obtidos com sequências de protease de HIV-1 subtipo B disponíveis em bases de dados públicas e literatura.

As análises compreenderam duas etapas. Na primeira, foi realizado um estudo do perfil da protease encontrada nos recombinantes BF provenientes de pacientes dos estados do Sul e Sudeste do Brasil utilizando sequências selecionadas a partir da base de dados do Laboratório de Retrovirologia da UNIFESP. A segunda etapa envolveu a análise da covariação de mutações na protease do subtipo F e o estudo dos agrupamentos (clusters) envolvendo mutações e inibidores de protease. Para esta segunda etapa, foram utilizadas sequências de protease confirmadas como subtipo F utilizadas na primeira etapa e sequências de protease F obtidas a partir da base de dados de Stanford.

4.1.1. CASUÍSTICA DA 1ª ETAPA: PERFIL DA PROTEASE DE HIV-1 SUBTIPO F DE RECOMBINANTES BF

497 sequencias nucleotídicas do gene pol compreendendo a posição 1 da protease à posição 335 da transcritase reversa (TR) previamente subtipadas como F ou recombinantes BF foram selecionadas da base de dados do laboratório de Retrovirologia da Universidade Federal de São Paulo. Todas as sequências pertenciam a indivíduos em falha terapêutica, moradores em municípios das regiões Sul e, principalmente, Sudeste do Brasil e foram coletadas entre 2002 e 2007. O Histórico terapêutico completo da maioria dos pacientes era conhecido. Outras informações relevantes (idade, gênero, carga viral, contagem de CD4) além do formulário de consentimento de todos os indivíduos envolvidos estavam disponíveis.

Para os indivíduos com mais de uma sequência, apenas a mais recente foi utilizada para as análises da primeira etapa.

4.1.2. CASUÍSTICA DA 2ª ETAPA: COVARIAÇÃO DAS MUTAÇÕES NA PROTEASE DO SUBTIPO F, ASSOCIAÇÃO COM O USO DE INIBIDORES DE PROTEASE E ANÁLISE TEMPORAL DAS MUTAÇÕES ASSOCIADAS AO TRATAMENTO

Participaram desta segunda etapa todas as sequências em que a região da protease foi confirmada como pertencente ao subtipo F através das análises realizadas na primeira etapa, além de sequências de protease do subtipo F oriundas de trabalhos já publicados e disponibilizadas em *Stanford HIV Drug Resistance Database* (http://hivdb.stanford.edu). Também no caso das sequências de Stanford, para os pacientes com mais de uma sequência disponível, apenas a mais recente foi utilizada para efeito das principais análises. Apenas as sequências com histórico terapêutico detalhado e fonte conhecida (plasma ou PBMC) foram consideradas. Havendo múltiplos isolados de uma mesmo paciente, apenas o mais recente foi utilizado. No caso de múltiplos clones, optou-se pelo consenso.

Análises em separado foram realizadas para os grupos de tratados com inibidores de protease (IP_{tr}) e de não tratados com inibidores de protease (IP_{ntr}). Para reduzir a interferência de resistência às drogas transmitida (resistência primária), isolados de indivíduos não tratados que apresentavam duas ou mais mutações principais não polimórficas estabelecidas como relacionadas às drogas foram excluídos.

4.2. MÉTODOS

4.2.1. 1ª ETAPA: PERFIL DA PROTEASE DE HIV-1 SUBTIPO F DE RECOMBINANTES BF

Subtipagem e verificação de recombinantes BF. Foram utilizados três métodos para confirmar a subtipagem prévia das sequências fornecidas pelo Laboratório de Retrovirologia da UNIFESP. Inicialmente, as sequências foram submetidas à ferramenta REGA 2.0 (www.hivdb.stanford.edu) e foram analisadas por *bootscanning* com Simplot software versão 2.5 (Stuart Ray, http://www.med.jhu.edu/deptmed/sray/download). A seguir, foram submetidas ao site brasileiro para subtipagem do HIV (http://clinmaldb.usp.br:8083/hiv/subtype/hivstesubtype.html). A proporção de sequências confirmadas como totalmente F para a fração do gene pol analisada foi comparada à proporção de sequências confirmadas como recombinantes BF para esta mesma região gênica.

Para o caso de recombinantes BF, a análise por *bootscanning* utilizando Simplot também permitiu subdividir as sequências de acordo com a região da protease em três categorias: **protease totalmente F** (BF_{proF}), **protease totalmente B** (BF_{proB}) e **protease híbrida** (BF_{proBF}). A presença de uma protease subtipada como totalmente F ou totalmente B foi também confirmada por análises filogenéticas realizadas com Phylip Inference Package versão 6.68 (http://evolution.gs.washington.edu/phylip.html), com base em alinhamentos gerados com MEGA versão 4 (Tamura et al, 2007). Relações filogenéticas foram estabelecidas usando *neighbor-joining* com Kimura-2-parâmetros para calcular as distâncias genéticas. As sequências referência para as análises filogenéticas foram obtidas de Los Alamos National Laboratory HIV Database (http://www.hiv.lanl.gov). As mesmas sequências referência alinhadas às sequências BF_{proF} foram utilizadas para a inspeção das assinaturas moleculares nucleotídicas do subtipo F.

Com o objetivo de verificar se o mesmo modelo evolutivo poderia ser aplicado à região da protease de sequências F e de sequências de recombinantes BF_{proF}, ModelTest (jModelTest 0.1.1, http://darwin.urvigo.es) foi aplicado separadamente para dois alinhamentos: um compreendendo a região da protease de todas as sequências F (tratados e não tratados com inibidores de protease) e outro compreendendo a região da

protease de todas as sequências BF_{proF} (tratados e não tratados com inibidores de protease). Posições com mutações principais de resistência às drogas de ambos os alinhamentos foram excluídas desta análise para evitar convergência pela pressão seletiva das drogas antirretrovirais.

Comparação do perfil de mutações na protease de sequências subtipadas como F e BF_{proF}. Para analisar o perfil de mutações da protease do subtipo F e dos recombinantes BF, somente as sequências de indivíduos com histórico terapêutico conhecido e confirmadas como F ou recombinantes BF_{proF} (para a região pol) foram consideradas. Estas sequências foram subdivididas em quatro grupos de acordo com o uso ou não de inibidores de protease (IP) em subtipo F não tratados com IP (ntrF), subtipo F tratados com IP (trF), recombinantes BF_{proF} não tratados com IP ($ntrBF_{proF}$) e recombinantes BF_{proF} tratados com IP ($trBF_{proF}$).

Foram construídas duas sequências consenso nucleotídicas da região da protease com Clustall W, MEGA versão 4, uma baseada no alinhamento do grupo ntrF e a outra baseada no alinhamento do grupo $ntrBF_{proF}$. As duas consensos foram comparadas para verificar possíveis diferenças. Contudo, como nenhuma diferença foi constatada, uma consenso única foi utilizada para a obtenção da freqüência de mutações em cada uma das 99 posições da protease nos quatro grupos de sequências (ntrF, trF, $ntrBF_{proF}$ e $trBF_{proF}$). A consenso e os quatro grupos de sequências foram traduzidas com MEGA versão 4.

As freqüências de mutações da protease foram comparadas tanto em ambos os grupos não tratados (ntrF versus $ntrBF_{proF}$) como em ambos tratados (trF versus $trBF_{proF}$) para verificar possíveis diferenças envolvendo a protease do subtipo F e a protease dos recombinantes BF_{proF}. Também foram realizadas comparações entre os grupos de não tratados e tratados (ntrF versus trF, $ntrBF_{proF}$ versus $trBF_{proF}$).

Análises estatísticas. Para a comparação da frequência do subtipo F e dos recombinantes BF entre as amostras analisadas, foi aplicado o teste de proporção baseado na distribuição binomial. O teste qui-quadrado foi usado para verificar diferenças estatísticas entre as freqüências dos grupos de recombinantes BF (BF_{proF}, BF_{proB} e BF_{proBF}). As possíveis diferenças nas proporções das mutações entre os grupos ntrF, $ntrBF_{proF}$, trF_{ntr} e $trBF_{proF}$ foram verificadas através do teste de proporção baseado numa distribuição binomial e o método de Benjamini e Hochberg (Benjamini & Hockberg, 1995) foi aplicado para controlar a taxa da falsa descoberta (FDR: *false*

discovery rate). As analyses estatísticas foram realizadas com MINITAB Release 14 e Microsoft Office Excel 2007. Os resultados foram considerados significantes para $p<0,05$.

4.2.2. 2ª ETAPA: COVARIAÇÃO DAS MUTAÇÕES NA PROTEASE DO SUBTIPO F, ASSOCIAÇÃO COM O USO DE INIBIDORES DE PROTEASE E ANÁLISE TEMPORAL DAS MUTAÇÕES ASSOCIADAS AO TRATAMENTO

Determinação das mutações associadas ao tratamento no subtipo F e análise de covariação de aminoácidos. Para o estudo da covariação de aminoácidos e determinação de clusters para o subtipo F, foram utilizadas todas as sequências com protease F da base de dados do Laboratório de Retrovirologia da UNIFESP (F puros e BF_{proF}) e sequências de protease F de Stanford. Estas sequências foram subdivididas em dois grupos: sequências de indivíduos não tratados com inibidores de protease (IPntr) e sequências de indivíduos que receberam pelo menos um inibidor de protease (IPtr). Para todas as análises, foi utilizada a sequência referência F1.BR.01.01BR125 disponível na base de dados de Los Alamos. Comumente, a interpretação dos testes de genotipagem bem como outras análises, é feita com base em dados do subtipo B. A utilização de uma sequência referência do subtipo F possibilitou verificar as mutações selecionadas pelo tratamento e as diferenças entre o subtipo F e o subtipo B.

As sequências de aminoácidos do grupo IPntr foram alinhadas à referência F1.BR.01.01BR125 e a frequência de mutações foi verificada para cada uma das 99 posições da protease. O mesmo procedimento foi adotado com as sequências do grupo IPtr. Foram definidas como mutações todas as diferenças em relação a essa referência F. A classificação em mutações principais e secundárias quanto à resistência aos inibidores de protease foi feita com base nos critérios adotados por Rhee e colaboradores (2007). Para os propósitos deste trabalho, foram consideradas mutações principais aquelas nas posições 24, 30, 32, 46, 47, 48, 50, 53, 54, 76, 82, 84, 88 e 90. As mutações nas posições 10, 20, 33, 36, 58, 63, 71, 73, 74, 77 e 93 foram consideradas secundárias. As demais, que por ventura pudessem ser apontadas como características do subtipo F, permaneceram como indefinidas ou como sugestivas de ação acessória. Foram considerados polimorfismos naturais as mutações presentes em mais de 0,5% dos indivíduos não tratados, salvo mutações principais que poderiam indicar a presença de

resistência transmitida. Este número foi escolhido para distinguir mutações com maior chance de refletirem a variação natural na protease das mutações raras que podem refletir resistência transmitida, erros de seqüenciamento ou variantes virais não viáveis raras (Rhee et al, 2005).

A associação de uma dada mutação ao tratamento foi verificada através do teste qui-quadrado e do teste exato de Fisher, conforme o caso e o método de Benjamini e Hochberg foi aplicado para controlar a taxa da falsa descoberta. Os testes foram aplicados através da construção de tabelas de contingência 2x2, contendo o número de sequências de pacientes tratados e de não tratados e o número de sequências com e sem uma determinada mutação. Para todas as mutações associadas ao tratamento, foi calculado o coeficiente de Yuli com o objetivo de verificar se uma associação é positiva ou negativa. No caso, associação negativa significou que a frequência de uma determinada mutação diminuiu significativamente com o tratamento.

As mutações que alcançaram significância estatística para associação positiva com o tratamento ($p \leq 0{,}05$) foram utilizadas para a construção de duas matrizes de dados binários (zero e um para ausência e presença, respectivamente, para uma dada mutação). A primeira matriz reuniu os dados do grupo IPntr e a segunda os dados do grupo IPtr. Estas matrizes foram utilizadas para a obtenção dos coeficientes de similaridade de Jaccard (equação 4.1), com o propósito de se verificar a magnitude da associação entre cada par de mutações X e Y e, também, dos coeficientes de dissimilaridade, utilizando-se o programa DendroUPGMA (http://genomes.urv.cat/UPGMA/).

Equação 4.1.

$$J_{XY} = \frac{(XY)}{(XY) + (X0) + (0Y)}$$

Foram consideradas combinações de mutações relevantes e possíveis candidatas à covariação aquelas com valores de $J \geq 0{,}2$. Para verificar qual das duas mutações X ou Y tem maior chance de ser selecionada primeiro, foram calculadas as probabilidades condicionais:

P(probabilidade de X dado Y) = $P(X|Y)$= nº de sequências com mutação X e mutação Y/nº de sequências com mutação Y

e

P(probabilidade de Y dado X) = $P(Y|X)$ = nº de sequências com mutação X e mutação Y/nº de sequências com mutação X.

Segundo Rhee e colaboradores (2007), há uma relação entre o log da razão da probabilidade condicional de duas mutações e o log da razão em que duas mutações se desenvolvem, indicando que a dependência condicional entre mutações está altamente correlacionada com a ordem em que as mutações se desenvolvem, quando ocorrem juntas.

As matrizes de dados binários utilizada anteriormente foram utilizadas para a obtenção de clusters através da análise por agrupamento hierárquico (HCA) e da análise por componentes principais (PCA).

A HCA foi aplicada por clusterização de variáveis utilizando o *average linkage method*, medida de distância por correlação e nível de similaridade de 0,55 para a formação dos clusters. Este método levou à construção de dendrogramas de similaridade envolvendo todas as mutações positivamente associadas ao tratamento no subtipo F.

O PCA foi aplicado com a opção de redução do conjunto de variáveis a dois componentes principais e matriz de correlação, resultando na construção de um gráfico que permite a visualização num espaço bidimensional das variáveis (mutações) e seus agrupamentos. Ambas as técnicas permitem clusterizar diversas variáveis (no caso, cada uma das mutações) formando agrupamentos com alto nível interno de similaridade.

Análise da associação de mutações individuais e agrupamentos de mutações (clusters) com IPs: À matriz de dados binários contendo informações sobre a presença e a ausência das mutações associadas ao tratamento, foram acrescentadas informações sobre a utilização ou não de cada um dos IP por paciente para o grupo de IPtr. Esta nova matriz foi utilizada para a análise de associação entre as mutações selecionadas pelo tratamento e o uso de um determinado IP. Para esta análise, foi aplicada a HCA por clusterização de variáveis utilizando o *average linkage method* e nível de similaridade entre 0,5 e 0,55 para a formação de clusters. Cada cluster principal observado foi analisado em separado, utilizando-se novamente o HCA. Este procedimento permitiu,

em alguns casos, a obtenção de subclusters com alto nível de similaridade interno envolvendo mutações específicas e IPs.

Os resultados obtidos foram comparados a dados disponíveis sobre o subtipo B obtidos de literatura e de bases de dados públicas.

Análise temporal da evolução de mutações associadas ao tratamento na protease do subtipo F: Vinte e dois pacientes da base de dados de Stanford e 17 do Laboratório de Retrovirologia da UNIFESP, totalizando 39 pacientes, possuíam sequências de protease F obtidas em momentos diferentes. Quando disponibilizados, foram obtidos o ano em que se deu o sequenciamento, esquemas terapêuticos e procedência dos pacientes para cada uma das sequências.

Ambas as sequências de cada paciente foram traduzidas e alinhadas à sequência referência F1.BR.01.01BR125. Foram verificadas todas as mutações apontadas como associadas ao tratamento no subtipo F e possível associação com resistência a inibidores de protease por comparação aos painéis disponíveis.

Observou-se a proporção de pacientes em que o número total de mutações aumentou, permaneceu o mesmo e diminuiu da primeira para a segunda sequência. O teste t pareado de *Student* foi aplicado para verificar se houve diferença significativa entre a média do número total de mutações na primeira sequência dos pacientes e a média do número total de mutações na segunda sequência. Da mesma forma, comparou-se a média do número de mutações principais na primeira e na segunda sequências.

Foi feita a verificação da presença das mutações D30N, N88D e L90M, em especial, para os pacientes que utilizaram NFV em seus esquemas terapêuticos.

Para os pares de mutações X e Y associadas ao tratamento no subtipo F e com com $J \geq 0,2$, foram verificadas as frequências com que as duas mutações X e Y apareceram juntas em ambas as sequências, X apareceu primeiro, Y apareceu primeiro, ambas estavam presentes na primeira sequência e desaparecem da segunda e ambas estavam ausentes na primeira e aparecem na segunda. Foi realizada uma busca em cada par de sequências com o objetivo de se verificar a presença dos clusters e subclusters envolvendo mutações e IPs específicos observados nas análises anteriores.

Análises estatísticas: Todas as análises foram realizadas com Microsoft Excel 2007 e MINITAB Release 14 Statistical Software, com resultados considerados significativos para $p \leq 0,05$.

CAPÍTULO 5. RESULTADOS

5.1. 1ª ETAPA: PERFIL DA PROTEASE DE HIV-1 SUBTIPO F DE RECOMBINANTES BF

Subtipagem e verificação de recombinantes BF. 497 sequências da região pol de indivíduos em falha terapêutica previamente subtipados como HIV-1 subtipo F ou recombinantes BF foram analisadas com três ferramentas. Deste total, 489 foram confirmadas como F ou recombinantes BF, seis foram reclassificadas como subtipo B, uma foi considerada DF e uma obteve resultados discordantes com os três métodos utilizados, persistindo a dúvida sobre sua classificação. Considerando as 489 sequências F e BF, 297 eram recombinantes BF enquanto apenas 192 eram do subtipo F (60,7% e 39,3%, respectivamente, *p*=0,000).

Entre as 297 sequências de recombinantes BF para a região pol, 163 apresentavam uma protease totalmente F (BF_{proF}), 98 apresentaram uma protease híbrida e somente 36 apresentaram protease do subtipo B (55,7%, 33,0% e 12,1%, respectivamente; Fig. 5.1.).

Através do alinhamento dos recombinantes BF_{proF} com sequências referência obtidas da base de dados de *Los Alamos National Laboratory* foi possível confirmar a presença das assinaturas moleculares do subtipo F na região da protease.

Para 15 sequências BF_{proF} apontadas pelo Bootscan como tendo uma região de quebra próxima à posição correspondente ao término da protease (ao redor da posição 300 na sequência nucleotídica do gene pol), foi feita a verificação das posições 265, 266 e 267, correspondentes à posição 89 na sequência de aminoácidos. Nesta posição, é comum a substituição da leucina (L) pela metionina (M) no subtipo F. Todas as 15

sequências pertenciam a pacientes tratados. Destas, 6 apresentaram a substituição L89M, 2 apresentaram a substituição L89I, uma apresentou L89T e 5 não tiveram substituição nessa posição. As duas sequências com L89I pertenciam a pacientes tratados exclusivamente com NFV, Essa é uma substituição acessória associada a esse inibidor. Uma única sequência apresentou a mistura M/I, resultado da presença do códon MTG (ATG ou CTG).

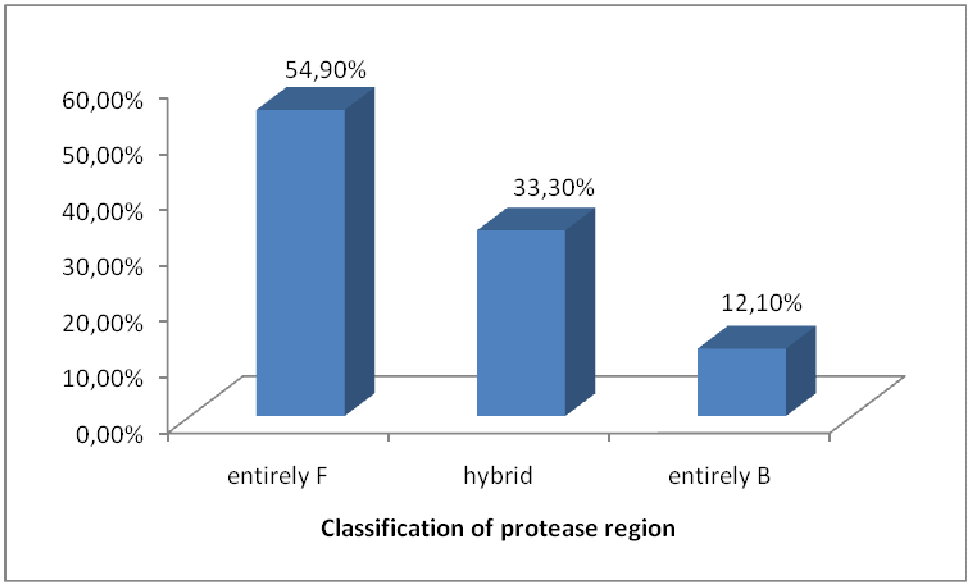


Figura. 5.1. Proporção de recombinantes BF na região pol de acordo com a presença de uma protease classificada como F, B ou híbrida BF (p=0,000).

As análises filogenéticas confirmaram as sequências classificadas como subtipo F e, também, a presença de protease totalmente F nas sequências BF_{proF}. Na figura 5.2. A, observa-se que as sequências formam clusters com as diferentes referências F utilizadas. As demais referências (B, D e C) formam pequenos clusters fora do cluster principal constituído por todas as sequências F (referências e amostras). A figura 5.2.B mostra uma situação semelhante para as sequências BF_{proF}.

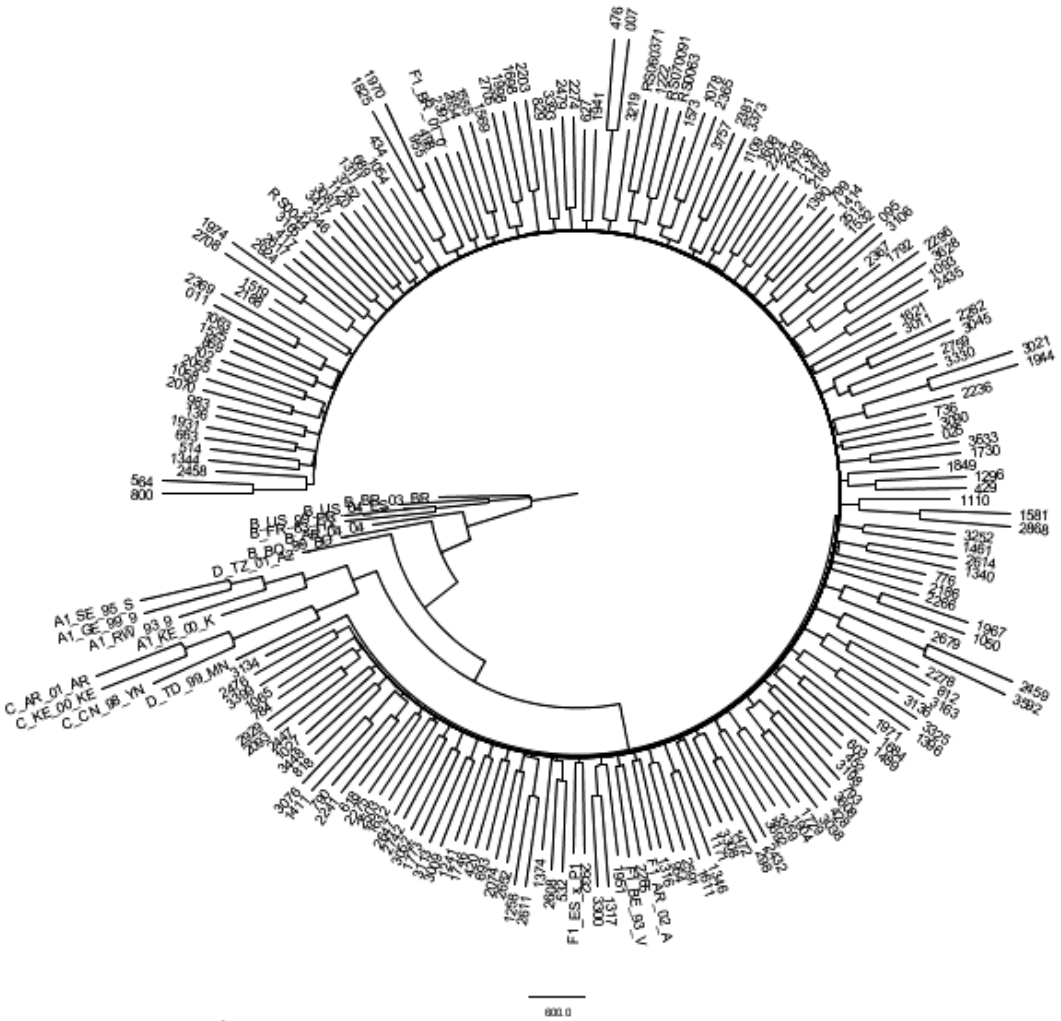


Figura 5.2.A Árvore filogenética da região da protease construída com sequências referência dos subtipos A, B, C, D e F e as sequências do subtipo F da base de dados do Laboratório de Retrovirologia da UNIFESP.

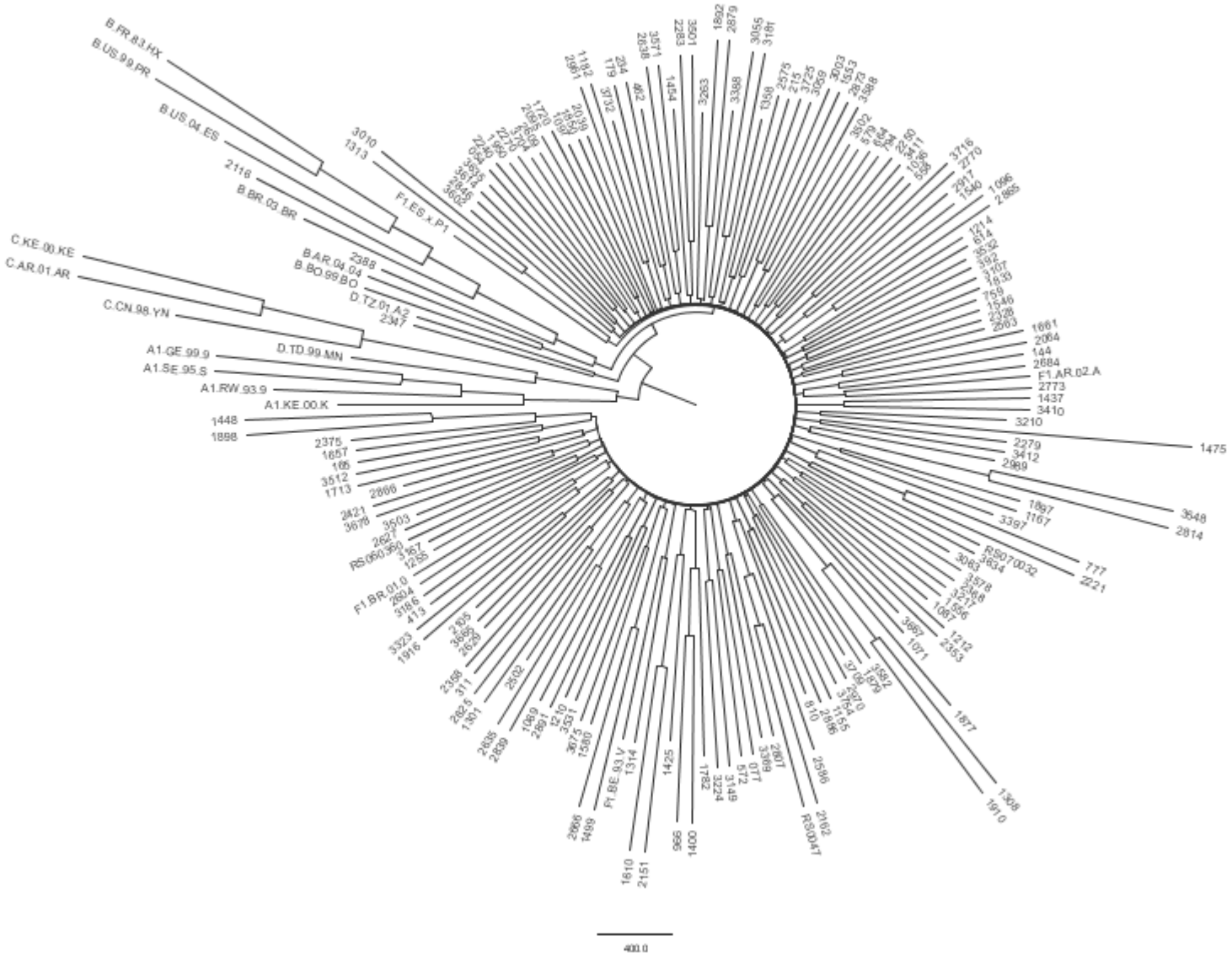


Figura 5.2. B Árvore filogenética da região da protease construída com sequências referência dos subtipos A, B, C, D e F e as sequências BFproF da base de dados do Laboratório de Retrovirologia da UNIFESP.

A aplicação do MODELTEST ao alinhamento da região da protease de sequências F resultou no modelo evolutivo TVM+I+G (Temporal Version Model ou modelo verão temporal, I = invariable sites ou sítios não variáveis, G =distribuição gama), enquanto o resultado para o grupo BF_{proF} foi TPM (Two Phase Model ou modelo em duas fases).

Comparação do perfil de mutações da protease do subtipo F e do recombinante BF_{proF}: Entre as 192 sequências confirmadas como subtipo F, duas pertenciam a pacientes com histórico terapêutico incompleto e foram excluídas das análises que se seguiram. Das restantes 190 sequências, 51 pertenciam a indivíduos não tratados com IP (ntrF) e 139 pertenciam a indivíduos tratados com IP (trF). Das 163 sequências BF_{proF}, 24 pertenciam a indivíduos não tratados com IP ($ntrBF_{proF}$) enquanto 139 pertenciam a indivíduos tratados com IP ($trBF_{proF}$). Cada um dos quatro grupos de sequências já traduzidas (ntrF, trF, $ntrBF_{proF}$ e $trBF_{proF}$) foi alinhado à sequência consenso da protease F e a freqüência de mutações foi verificada para cada uma das 99 posições.

A comparação da proporção de mutações entre a região da protease de ntrF e $ntrBF_{proF}$ revelou cinco diferenças significativas (tabela 5.1.). Mutações nas posições 15 e 67 foram estatisticamente mais comuns no grupo $ntrBF_{proF}$, enquanto mutações nas posições 41, 60 e 72 foram mais frequentes no grupo ntrF. Além dessas posições, outras duas também apareceram com diferenças, mas foram excluídas pelo controle da FDR: 63 e 89. Algumas diferenças foram também observadas entre os dois grupos tratados com IP. A frequência de mutações nas posições 74 e 93 foi significativamente maior no grupo $trBF_{proF}$, enquanto nas posições 16, 20 e 60 a frequência foi mais alta no grupo trF. As posições 14, 24, 47, 50, 54, 67, 76, 88 e 92 foram excluídas pelo controle da FDR. As diferenças observadas, contudo, foram apenas em relação às frequências das mutações. A comparação entre os códons mais frequentes para as substituições mais comuns não revelou diferenças entre os grupos (tabela 5.1.).

Quando a proporção de mutações presentes no grupo ntrF foi comparada à proporção do grupo trF, dez posições alcançaram diferenças significativas (tabela 5.2.):

10, 15, 20, 46, 54, 71, 82, 88, 89 e 90, todas elas com freqüência mais alta no grupo de tratados. Entre $ntrBF_{proF}$ e $trBF_{proF}$, seis posições mostraram diferenças significativas: 10, 20, 54, 72, 82 e 90.

Analisando-se os resultados de todas as comparações realizadas observa-se que na posição 15, a frequência da mutação I15V, já elevada no grupo $ntrBF_{proF}$ mas não no grupo ntrF, permanece alta no grupo $trBF_{proF}$ (não há diferença estatística entre $ntrBF_{proF}$ e $trBF_{proF}$), mas aumenta significativamente no grupo trF. Situação semelhante é observada na posição 72, porém, com frequência de mutação mais elevada no grupo ntF, que permanece elevada no grupo trF, e, baixa frequência de mutação entre os $ntrBF_{proF}$ que aumenta significativamente entre os $trBF_{proF}$. Em ambos os casos, pode estar havendo a seleção ou a permanência das mutações em função de alguma vantagem adaptativa que possam estar conferindo ao vírus na presença do tratamento.

Tabela 5.1. Posições com diferenças significativas na frequência de mutações entre ntrF e $ntrBF_{prof}$ e entre trF e $trBF_{prof}$ ($p \leq 0,05$).

	posição	frequência mais alta	*p*	substituição mais frequente	codons mais frequentes		Stanford	IAS panel
					F	BF		
não tratados	V15	BF	0,000	I	ATA	ATA	*	*
	C67	BF	0,002	S	TGY/WGT	AGT	*	*
	M89	BF	0,002	L	CTG	CTG	*	*
	D60	F	0,004	E	GAG	GAA/GAG	ATV	ATV
	I72	F	0,000	T	ACA	ACA/ACM	*	*
tratados	T74	BF	0,000	S	TCA	TCA	*	*
	I93	BF	0,003	L	CTT	CTT	*	ATV
	G16	F	0,005	E	GAG	GGA/GAG	ATV	ATV
	K20	F	0,004	R/T	AGG	AGG	(a)	ATV,LPV,IDV,TPV
	D60	F	0,005	E	GAG	GAG	ATV	ATV

(a) K20R é considerada uma possível reguladora positiva da protease e K20T é considerada associada à TARV, porém com papel incerto.

Tabela 5.2. Posições com diferenças significativas na frequência de mutações entre ntrF e trF e entre $ntrBF_{proF}$ e $trBF_{prof}$ ($p \leq 0,05$).

posição	subtipo F			BF_{proF}			Stanford	AIAS
	p ntrF x trF	substituições mais frequentes	codons prevalentes	*p* ntrBFproF x trBFproF	Substituições mais frequentes	codons prevalentes		
L10	0,000	V/I	GTA/ATA	0,000	I/V	ATA	todos (exceto DRV)	ATV,FPV, IDV,LPV,TPV(V)
V15	0,000	I	ATA	*	I		sem citação	sem citação
K20	0,000	R/T	AGG/ACG	0,001	R/T	AGG/ACG	(a)	ATV,LPV,IDV,TPV
M46	0,000	L/I	TTG/ATA	*	L/I		ATV,APV,IDV, LPV,NFV,TPV	ATV,APV,IDV,LPV,NFV,TPV(L)
I54	0,000	V	GTC	0,000	V	GTC	todos (exceto DRV)	ATV,APV,IDV,LPV TPV
A71	0,002	V	GTT	*	V		todos (exceto DRV)	ATV,IDV,LPV,NFV SQV
I72	*	T/V		0,006	T/V	ACA	sem citação	sem citação
V82	0,000	A/F	GTC/TTC	0,000	A/F	GCC	Todos (F) ATV,IDV,LPV,NFV(A)	todos (exceto DRV e TPV)
N88	0,003	D/S	GAT/AGT	*	D/S		ATV(S),IDV(S), SQV(S),NFV(D/S)	ATV(S),NFV
M89	0,004	L/I	ATG/ATA	*	L/I		sem citação	sem citação
L90	0,000	M	ATG/ATA	0,000	M	ATG	todos (exceto DRV)	ATV,APV,IDV,LPV,NFV,SQV,TPV

5.1. 2ª ETAPA: COVARIAÇÃO DAS MUTAÇÕES NA PROTEASE F, DETERMINAÇÃO DOS AGRUPAMENTOS ESPECÍFICOS (CLUSTERS) PARA ESTE SUBTIPO DE HIV-1, ASSOCIAÇÃO COM O USO DE IPs ESPECÍFICOS E ANÁLISE TEMPORAL DAS MUTAÇÕES ASSOCIADAS AO TRATAMENTO

Determinação das mutações associadas ao tratamento no subtipo F e análise de covariação de aminoácidos: A junção das sequências de protease F da base de dados do Laboratório de Retrovirologia da UNIFESP com as sequências de Stanford resultou num total de 1265 sequências de protease F, sendo 677 de pacientes não tratados com IPs (IPntr) e 588 de pacientes tratados com IP (IPtr).

A comparação da proporção de mutações entre os grupos IPntr e IPtr através dos testes qui-quadrado e Fisher com a correção pelo método de Benjamini e Hochberg resultou em 60 mutações que alcançaram significância estatística para associação com o tratamento na protease do subtipo F. Destas mutações, 51 apresentaram coeficiente de Yuli positivo (Tabela 5.3.), indicando aumento significativo da frequência da mutação no grupo de tratados em relação ao grupo de não tratados. Entre as 51 mutações positivamente associadas ao tratamento, 13 são polimorfismos (L10I, I13V, K14R, V15I, L19V, K20R, N37D, I62V, L63P, A71T, T74S, M89I, Q92K), entre os quais nove já foram citados como relacionados à resistência a um ou mais IPs ou tem possibilidade de estar associados à resistência (L10I, I13V, K20R, I62V, L63P, A71T, T74S, M89I e Q92K) (Johnson et al, 2008; Shafer & Schapiro, 2008). Além dos polimorfismos associados ao tratamento, outros cinco polimorfismos com elevada prevalência entre os não-tratados que permanece elevada entre os tratados associados à resistência a um ou mais IP foram observados (L10V, K20M, D60E, V77I, I93L).

Tabela 5.3. Mutações positivamente associadas ao tratamento na protease do subtipo F ($p \leq 0,05$).

mutação	*p*	Yule	mutação	*p*	Yule
L10I	0,000	0,63	K55R	0,000	0,97
I13V	0,008	0,21	Q58E	0,000	0,89
R14K	0,002	0,23	**I62V**	0,000	0,37
V15I	0,007	0,18	**L63P**	0,000	0,59
G17E	0,000	1,00	I64V	0,000	0,85
Q18E	0,004	1,00	A71V	0,000	0,95
L19V	0,007	0,61	**A71T**	0,000	0,71
K20R	0,000	0,38	G73S	0,000	0,89
K20T	0,000	0,97	T74P	0,001	1,00
L24I	0,000	1,00	**T74S**	0,000	0,59
D30N	0,000	0,96	L76V	0,000	0,89
V32I	0,000	1,00	V82A	0,000	0,98
L33F	0,000	1,00	**V82I**	0,000	1,00
D35N	0,000	0,85	V82F	0,002	0,85
I36L	0,005	0,83	V82S	0,001	1,00
N37D	0,000	0,32	V82T	0,002	1,00
K43T	0,000	0,83	N83D	0,002	1,00
M46I	0,000	0,94	I84V	0,000	1,00
M46L	0,000	0,98	I85V	0,000	0,92
I47V	0,000	1,00	N88D	0,000	0,96
G48V	0,000	1,00	N88S	0,000	1,00
I50V	0,005	0,83	**M89I**	0,000	0,85
I50L	0,002	1,00	L90M	0,000	0,98
F53L	0,000	1,00	**Q92K**	0,000	0,75
I54V	0,000	0,99	T96S	0,000	0,89
I54A	0,002	1,00			

Em negrito, polimorfismos com fequência acima de 0.5% no grupo de não tratados.

A combinação das 51 mutações significativamente associadas ao tratamento duas a duas resultou em 1275 pares possíveis para os quais foram obtidos os valores do coeficiente de similaridade de Jaccard (*J*) e, também, do coeficiente de dissimilaridade (1-J), tanto para o grupo de IPntr como para o grupo de IPtr. Dos 1275 pares de mutações, 51 apresentaram valores de $J \geq 0,2$ (Tabela 5.4.) no grupo de IPtr. Dos 51 pares obtidos, 10 são formados por duas mutações principais, e 21 são compostos por uma mutação principal e uma acessória ou secundária. Os pares 30N/88D e I54V/V82A, ambos compostos por mutações classificadas como principais, apresentaram os mais altos coeficientes de similaridade (0,60 e 0,57, respectivamente). Para esses e, também, para alguns outros pares, os valores de P(X|Y) e P(Y|X) são

iguais ou muito próximos (Tabela 5.4.), dificultando o estabelecimento de uma ordem cronológica para o aparecimento das mutações. Para muitos pares, no entanto, é possível inferir a ordem para o aparecimento das duas mutações. No par L90M/A71V, por exemplo, a probabilidade de L90M dado a presença de A71V é 0,26 , enquanto a probabilidade inversa (A71V dado a presença de L90M) é 0,14. Segundo o método de Rhee e colaboradores (2007), A71V tem maior probabilidade de ser selecionada primeiro.

Já no grupo de IPntr e considerando-se os mesmos 51 pares de mutações, os valores do coeficiente de Jaccard foram nulos ou inferiores aos obtidos no grupo de IPtr.

A análise por agrupamento hierárquico (HCA) utilizando a matriz de dados contendo informações sobre as 51 mutações associadas ao tratamento do grupo de IPtr, resultou na obtenção de 19 agrupamentos ou clusters mostrados no dendrograma da figura 5.3.

Tabela 5.4. Pares de mutações associadas ao tratamento com valores do coeficiente de similaridade de Jaccard ≥0,2 e valores das probabilidades condicionais P(Y|X) e P(X|Y) para o grupo IPtr.

mut X	mut Y	J	XY	P(Y|X)	P(X|Y)
D30N	**N88D**	0,60	28	0,794	0,794
V82A	**I54V**	0,57	136	0,746	0,701
R14K	K20R	0,45	234	0,463	0,925
V82A	K20R	0,42	128	0,702	0,506
I54V	K20R	0,41	127	0,655	0,502
I54V	L10I	0,40	105	0,541	0,595
V82A	L10I	0,40	102	0,560	0,578
I47V	V32I	0,36	8	0,454	0,680
L90M	L63P	0,34	81	0,530	0,475
I54V	R14K	0,32	168	0,866	0,332
L63P	**I54V**	0,32	87	0,510	0,448
L10I	R14K	0,32	163	0,924	0,322
K20R	L10I	0,31	101	0,399	0,573
V82A	R14K	0,31	160	0,878	0,316
V82A	L63P	0,30	81	0,444	0,475
L90M	I54V	0,30	78	0,510	0,402
L90M	**A71V**	0,29	53	0,347	0,644
R14K	L63P	0,29	151	0,299	0,886
A71V	L63P	0,28	56	0,680	0,328
L63P	L10I	0,28	75	0,440	0,425
A71V	**I54V**	0,27	58	0,705	0,299
L90M	I62V	0,27	58	0,379	0,448
L63P	K20R	0,26	88	0,516	0,348
I62V	**I54V**	0,26	66	0,510	0,340
R14K	I15V	0,26	129	0,256	0,846
K55R	**M46L**	0,25	24	0,454	0,340
I13V	M89I	0,25	12	0,118	0,200
L63P	I62V	0,24	58	0,340	0,448
L90M	R14K	0,24	125	0,818	0,247
V82A	A71V	0,23	50	0,274	0,607
A71V	L10I	0,23	49	0,595	0,278
L90M	L10I	0,23	60	0,392	0,340
I54V	**M46L**	0,23	48	0,247	0,680
V82A	**M46L**	0,23	46	0,252	0,652
M46L	**L24I**	0,22	22	0,312	0,416
L10I	V15I	0,22	47	0,267	0,308
V15I	L63P	0,22	41	0,269	0,241
L63P	**M46L**	0,22	31	0,182	0,439
I54V	**M46I**	0,22	48	0,247	0,628
V82A	**L24I**	0,22	41	0,225	0,775
L24I	L10I	0,21	40	0,756	0,227
L90M	**V82A**	0,21	58	0,379	0,318
I62V	K20R	0,21	66	0,510	0,261
L90M	K20R	0,21	70	0,458	0,277
T74S	**I54V**	0,21	47	0,571	0,242
V82A	T74S	0,21	45	0,247	0,547
I62V	R14K	0,21	107	0,827	0,212
I54V	**L24I**	0,20	41	0,211	0,775
L63P	K55R	0,20	37	0,217	0,699
L90M	T74S	0,20	39	0,255	0,474
T74S	I62V	0,20	35	0,425	0,271

Em negrito encontram-se as mutações principais.

Em cinza, valor mais alto de probabilidade condicional para cada par de mutações X e Y.

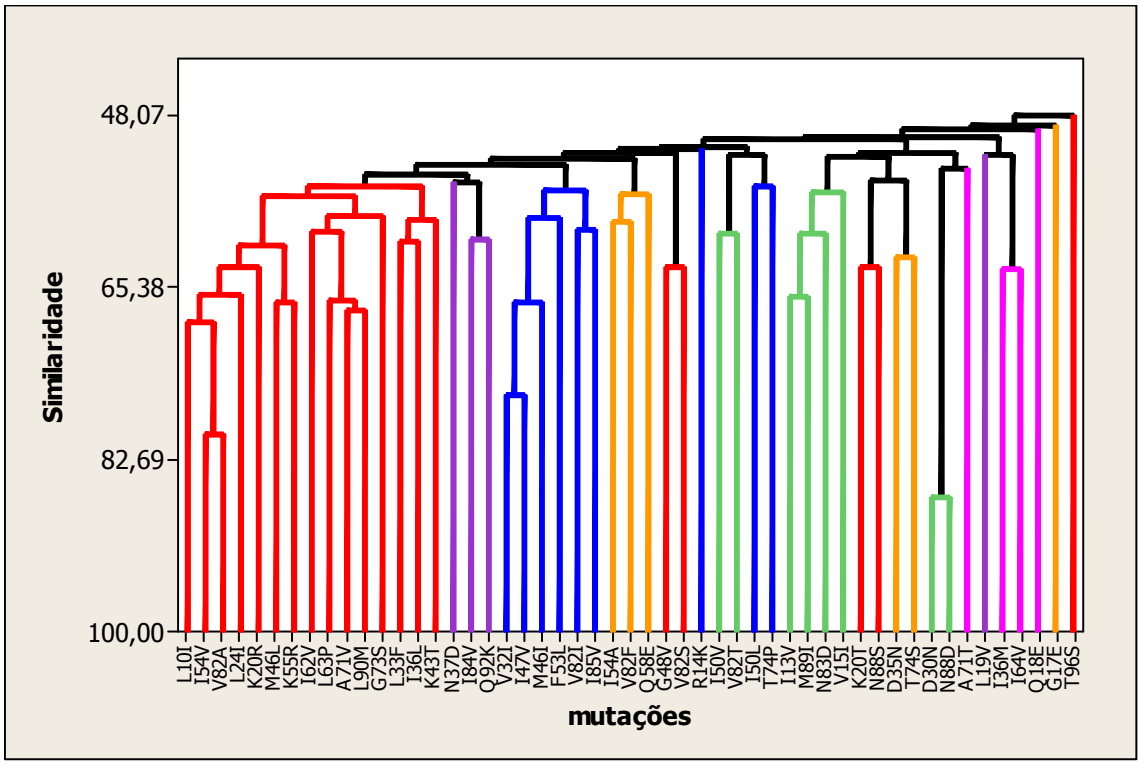


Figura 5.3. Dendrograma de similaridade entre as mutações associadas ao tratamento gerado por HCA, com *average linkage method* e nível de similaridade de 0,55.

Levando-se em conta apenas os clusters compostos por, pelo menos, duas mutações, o número fica reduzido a doze, identificados a seguir:

Cluster 1 L10I K20R L24I L33F I36L K43T M46L I54V K55R I62V L63P A71V G73S V82A L90M

Cluster 2 I13V V15I N83D M89I

Cluster 3 K20T N88S

Cluster 4 D30N N88D

Cluster 5 V32I M46I I47V F53L V82I I85V

Cluster 6 D35N T74S

Cluster 7 I36M I64V

Cluster 8 G48V V82S

Cluster 9 I50V V82T

Cluster 10 I50L T74P

Cluster 11 I54A Q58E V82F

Cluster 12 I84V Q92K

A tentativa de gerar clusters utilizando a matriz de dados do grupo IPntr contendo as mesmas 51 mutações associadas ao tratamento foi frustrada, devido à frequência extremamente baixa ou até mesmo nula da maioria das mutações associadas ao tratamento nesse grupo. Isso gerou um grande número de valores nulos para os coeficientes de similaridade necessários para a montagem dos clusters.

Análise da associação de mutações individuais e agrupamentos de mutações (clusters) com IPs: A análise da associação entre a presença de uma mutação ou grupo de mutações e o uso de um ou mais IP foi realizada apenas com as sequências de indivíduos com esquemas terapêuticos conhecidos. Das 588 sequências de indivíduos

IPtr, 152 traziam a informação sobre uso de IP, mas sem especificar qual ou quais inibidores haviam sido utilizados. A figura 5.4. mostra a proporção do uso de cada IP para os 436 indivíduos com esquemas conhecidos. NFV, IDV e RTV foram as drogas mais freqüentes nos esquemas terapêuticos, não havendo diferenças significativas entre suas proporções (p>0,05). Para o caso da droga RTV, considerou-se seu uso sempre como um IP em separado, como único inibidor ou combinado a outra droga, dada a falta de informação sobre seu uso como *booster*. Apenas para a droga LPV, foi possível considerar a presença da combinação LPV/r (lopinavir combinada a ritonavir como *booster*).

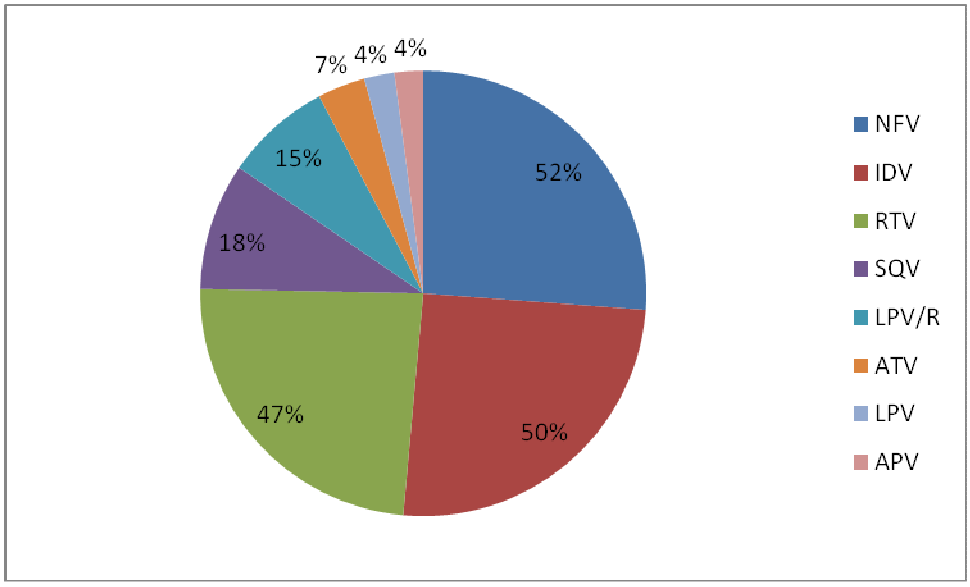


Figura 5.4. Frequência relativa de uso de cada IP nos esquemas terapêuticos do grupo IPtr.

Das 51 mutações associadas ao tratamento no subtipo F, 8 foram analisadas em separado através da HCA para verificação de associação ao uso específico de algum inibidor ou grupo de inibidores: V15I, G17E, Q18E, L19V, K20T, D35N, K55R, T74S, N88S, L89I e Q92K. A escolha destas mutações baseou-se no fato de serem menos frequentes ou mesmo ausentes no subtipo B, tornando as informações disponíveis sobre

elas bastante escassas. A tabela 5.5. resume os agrupamentos significativos (nível mínimo de similaridade para a formação de um cluster de 0,5) envolvendo cada uma dessas mutações e um ou mais IP. Das 12 mutações analisadas, 4 estão aparentemente correlacionadas à presença de NFV (V15I, Q18E, T74S e M89I), um dos IPs mais comuns nos esquemas terapêuticos. LPV com ou sem a presença de RTV, aparece associado a 4 mutações praticamente ausentes no subtipo B (L19V, D35N, K55R e Q92K).

Tabela 5.5. Correlação entre uma mutação específica e o uso de um ou mais inibidores de protease no subtipo F.

mutação	IP	similaridade do cluster
V15I	NFV	0,54
G17E	*	*
Q18E	NFV	0,54
L19V	LPV APV SQV RTV IDV	0,50
D35N	ATV LPV/r	0,54
K55R	LPV/r ATV LPV APV	0,51
T74S	NFV	0,51
M89I	NFV	0,55
Q92K	ATV LPV/r	0,51

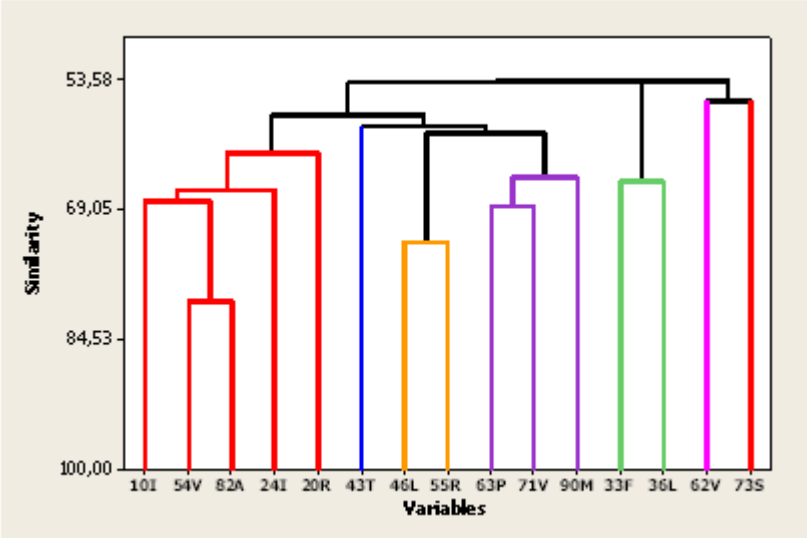
*A mutação G17E não formou nenhum grupamento significativo para o nível de similaridade de 0,5.

Além das 8 mutações individuais, também foram analisados através de HCA cada um dos 12 clusters de mutações citados anteriormente e a correlação com o uso específico de um ou mais IP. A análise em separado de cada um dos clusters resultou na obtenção de subclusters com elevado nível de similaridade interna (similaridade > 0,5).

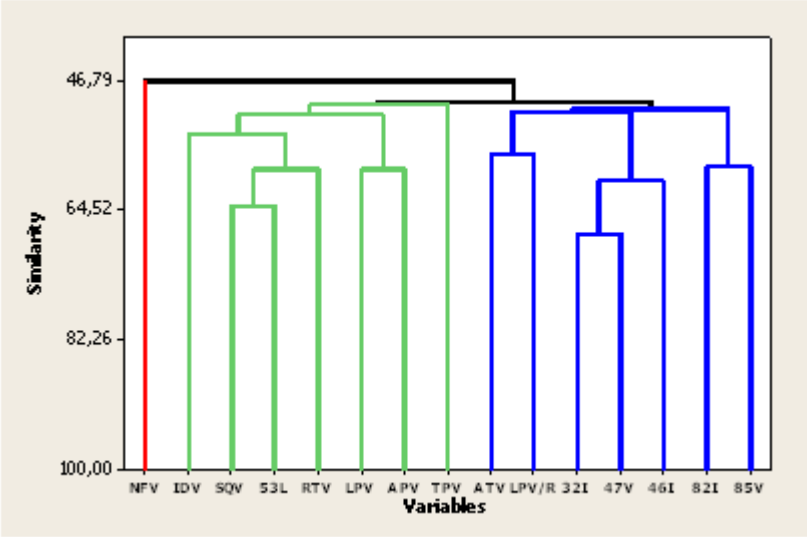
Os clusters 1 e 5 eram formados por 15 e 6 mutações, respectivamente. A análise direta de ambos com os IPs pela técnica de HCA levou à formação de dendrogramas em que a visualização dos agrupamentos ficou prejudicada pela maior similaridade entre as

mutações envolvidas do que entre qualquer mutação e um ou mais IP. Foi necessário subdividir os dois clusters em subgrupos. A subdivisão foi possível agrupando-se as mutações por nível de similaridade. Os subgrupos foram utilizados para a construção dos dendrogramas com os IP. As figuras 5.5.A e B mostram, respectivamente, a subdivisão do cluster 1 em 7 subgrupos e do cluster 5 em dois subgrupos.

A tabela 5.6 mostra a correlação entre clusters e subgrupos de mutações e o uso de determinados inibidores de protease. Para o cluster 1, com exceção das mutações K43T e I62V, cada um compondo um subgrupo, todos os demais subgrupos aparentam estar associados ao uso de LPV. Os IPs IDV, RTV e SQV parecem estar associados a cinco subgrupos do cluster 1 (L10I K20R L24I I54V V82A; K43T; I62V; L63P A71V L90M; G73S).



(A)



(B)

Figura 5.5. Dendrogramas de similaridade entre as mutações do cluster 1 (A) e entre as mutações do cluster 5 (B) e visualização dos subgrupos (nível de similaridade=0,6).

Tabela 5.6. Correlação entre clusters de mutações e IP na protease do HIV-1 subtipo F.

Cluster de mutações	Subgrupo de mutações	Subclusters mutações e IP	similaridade do cluster
cluster 1			
L10I K20R L24I L33F I36L K43T M46L I54V K55R I62V L63P A71V G73S V82A L90M	L10I K20R L24I I54V V82A	LPV APV IDV SQV RTV	0,53
	L33F I36L	LPV APV	0,54
	K43T	RTV SQV IDV	0,55
	M46L K55R	LPV/r APV LPV ATV	0,51
	I62V	TPV IDV SQV RTV	0,50
	L63P A71V L90M	IDV SQV RTV TPV LPV APV	0,52
	G73S	LPV APV IDV SQV RTV	0,52
Cluster 2			
I13V V15I N83D M89I		NFV	0,55
Cluster3			
K20T N88S		NFV	0,61
Cluster 4			
D30N N88D		NFV	0,63
Cluster 5	F53L	SQV TRV IDV LPV APV	0,54
V32I M46I I47V F53L V82I I85V			
	V32I I47V M46I V82I I85V	ATV LPV/r	0,51
Cluster 6			
D35N T74S		NFV	0,52
Cluster 7			
I36M I64V		NFV	0,52
Cluster 8			
G48V V82S		APV IDV SQV RTV	0,53
Cluster 9			
I50V V82T	**I50V**	**LPV APV**	**0,60**
	I50V V82T	LPV APV TPV IDV SQV RTV	0,51
Cluster 10			
I50L T74P	**I50L**	**ATV**	**0,70**
	I50L T74P	ATV LPV/r	0,52
Cluster 11			
I54A Q58E V82F		IDV SQV RTV LPV APV	0,51
Cluster 12			
I84V Q92K		SQV RTV TPV IDV LPV APV	0,50

Vários clusters aparentam estar associados unicamente ao uso de NFV (I13V I15V N83D L89I; K20T N88S; D30N N88D; D35N T74S; I36M I64V), enquanto diversos clusters e subgrupos estão associados a 5 ou mais diferentes IP.

Análise temporal da evolução de mutações associadas ao tratamento na protease do subtipo F: Sequências em dois momentos distintos estavam disponíveis para 39 pacientes. Para um dos pacientes, as informações sobre o esquema terapêutico estavam conflitantes e este paciente foi excluído desta análise.

O alinhamento de ambas as sequências de cada paciente à referência do subtipo F permitiu a verificação da presença de mutações associadas ao tratamento, além da verificação de qual ou quais mutações foram selecionadas, desapareceram ou permaneceram após alterações do esquema terapêutico (tabela 5.7.).

Dos 38 pacientes, 16 aumentaram o número total de mutações associadas ao tratamento, 13 permaneceram com o mesmo número e com as mesmas mutações e apenas 6 tiveram o número reduzido. Outros três pacientes não apresentaram mutações tanto na primeira quanto na segunda sequência. Quando o número médio de mutações associadas ao tratamento no grupo das primeiras sequências foi comparado ao número médio do grupo das segundas sequências, observou-se um aumento significativo ($média_{1^a seq} = 3,8 \pm 3,1$; $média_{2^a seq} = 4,6 \pm 3,6$; $p=0,02$).

Dos 16 pacientes que aumentaram o número de mutações, 10 apresentaram aumento do número de mutações principais, e nos demais o número permaneceu inalterado. A comparação do número médio de mutações principais das primeiras sequências com o número médio das segundas sequências não resultou em diferenças estatísticas significativas, indicando que se não houve aumento, tampouco ocorreu redução, mesmo com a troca do esquema terapêutico. Entre os seis pacientes com

redução no número total de mutações, quatro permaneceram com o mesmo número de mutações principais.

Na tabela 5.7., observa-se que cinco pacientes já apresentavam o par de mutações D30N/N88D na primeira sequência, quatro dos quais tratados com NFV e um com NFV e IDV. Dos cincos pacientes, quatro mantiveram o par D30N/N88D na segunda sequência (dois tomando LPV/RTV, um tomando ATV/RTV e um que permaneceu com NFV). O único paciente que não manteve o par de mutações interrompeu o tratamento com IP. A mutação L90M, geralmente selecionada pelo uso do NFV com mais frequência nos subtipos não-B do que o par D30N/N88D, está presente na primeira sequência de 9 pacientes, dos quais quatro usaram exclusivamente NFV, dois usaram NFV e outros IP e os demais não utilizaram NFV. Apenas em um dos nove pacientes L90M desapareceu.

Tabela 5.7. Mutações associadas ao tratamento e esquemas terapêuticos dos pacientes com sequências obtidas em dois momentos distintos.

Paciente	Origem	Ano	esquemas terapêuticos	Mutações	nº de mutações	nº de mutações principais
AC_pat1	Romênia	2000	IDV	10I 20R 24I 43T 46I 54V 82A	7	3
		2001	SQV,NFV	10I 20R 24I 43T 46I **53L** 54V 82A **84V**	9	4
AC_pat6	Romênia	2000	IDV,RTV,NFV	62V 82I	2	0
		2001	?	20T 62V 82I	3	0
AC_pat7	Romênia	2000	IDV,NFV,SQV, ?	20R 46L 54V 82A	4	3
		2001	?	20R 46L 54V 82A	4	3
AC_pat8	Romênia	2000	IDV,NFV/SQV	18E 20R	2	0
		2001	IDV/RTV, ?	18E 20R	2	0
AC_Pat20	Romênia	2000	IDV	20R	1	0
		2001	IDV/RTV	20R	1	0
AC_Pat21	Romênia	2000	IDV	20R	1	0
		2001	IDV	20R	1	0
AC_Pat23	Romênia	2000	IDV	20R 82A	2	1
		2001	NFV/SQV	20R 82A	2	0
AC_Pat25	Romênia	2000	IDV	20R	1	0
		2001	IDV	20R 48V 82A	3	1

AC_Pat26	Romênia	2000	IDV	74S	1	0
		2001	IDV/RTV	46I 54V 55R 74S 82T 84V	6	2
AC_Pat27	Romênia	2000	IDV	19V 20R 62V	3	0
		2001	IDV	19V 20R 62V	3	0
AC_Pat31	Romênia	2000	IDV	20R 62V	2	0
		2001	NFV/SQV	20R 62V	2	0
AC_Pat32	Romênia	2000	IDV	10I	1	0
		2001	NFV	10I	1	0
AC_Pat33	Romênia	2000	IDV	20R	1	0
		2001	NFV	20R	1	0
CA1317	Eua	2001	IDV,SQV,IDV/R,LPV/SQV	20T 46I 71V 74S 90M	5	2
		2003	LPV	20T 35N 46I 90M	4	2
MB_patBAR	França	1999	RTV,SQV/R,SQV/NFV,SQV/RTV	20R 82A	2	1
		2001	NFV,APV/RTV, LPV	10I 20R 74S 82A	4	1
CA4016	Eua	2002	NFV, nenhum	62V 63P	2	0
		2003	nenhum	62V 63P	2	0
12-1607	Eua	2003	?,TPV/RTV	20R 54V 62V 63P 64V 82T	6	2
		2004	TPV/RTV	20R 33L/F 54V 62V 63P 64V 82T 84I/V	8	4
48-1124	Arg	2003	IP desc	20R 90M	2	1
		2003	TPV/RTV	20R 90M	2	1
48-1068	Arg	2003	IP desc	*	0	0
		2003	TPV/RTV	*	0	0
48-1027	Arg	2003	IP desc	20R 24I 32I 33F 46I 47V 53L 54V 62V 74S 82A 92K	12	6
		2003	TPV/RTV	10I/V 20R 33L/F 46I/M 47I/V 53L 54V 62V 63P 74S 82A 89M/L 90M 92K/Q	14	6
48-8099	Reino unido	2003	IP desc	20R 24I 46L 54V 62V 63P 71V 76V 82A	9	4
		2003	TPV/RTV	20R 24I 46L 54V 62V 63P 71V 76V 82A	9	4
64/2503	Brasil	2002	IDV,NFV	*	0	0
		2005	nenhum	*	0	0
105/3045	Brasil	2002	NFV	62I/V	1	0
		2006	IDV/RTV	35N 46I 54A 58E 62V	5	2
108/3055	Brasil	?	IDV,SQV,RTV	46I 63P 71V 73S 84V	5	2
		2006	nenhum, ATV/RTV	46I 53F/L 63P 71V 73S 84V	6	2
142/784	Brasil	2002	IDV	20R 46L 54V 62V 63P 73S/R 82A 90M	8	4
		2003	LPV/RTV	20R 54V 62V 63P 82A 90M	6	3
304/3300	Brasil	2002	NFV	20R 30N 62V 63P 88D 89I	6	2
		2006	nenhum,LPV/RTV	20R 30N 62V 63P 88D	5	2
421/2886	Brasil	2003	NFV,IDV	20T 30N 43T 54V 63P 64V 74S 88D 89I 90M	10	4
		2006	LPV/RTV	20T 30N 43T 46I/M 54V 63P 64V 74S 82A 88D 89I 90M	12	5

532/2720	Brasil	2003	NFV	30D/N 71T/A 88D	3	2
		2006	?, nenhum	*	0	0
561/3003	Brasil	2003	NFV,nenhum,LPV/RTV	*	0	0
		2006	LPV/RTV	20R 32V/I 43K/T 46I 47V 54V 58E/Q 63P 82A	9	5
574/1210	Brasil	2003	?,RTV	10I 20R 24I 46L 54V 74S 82A	7	3
		2004	nenhum, LPV/RTV	10I 20R 24I 33F 46L 50I/V 54V 74P/S 82A	9	5
580/2759	Brasil	2003	NFV	20R/T 54V 58E/Q 74T/S 90M	5	2
		2006	LPV/RTV	20R 46L 54V 58E 62V 74T/S 90M	7	3
608/1698	Brasil	2003	NFV	20R 54V 62I/V 83D 89I 90M	6	2
		2005	LPV/RTV	20R 46I/M 54V 82A 83D 89I 92K/Q	7	3
1057/3393	Brasil	2004	RTV,LPV/RTV	10I/V 20R 54V 82A	4	2
		2007	nenhum	10I/V 20K/R 54I/V 63P/L 76V/L 82A	6	2
1124/3021	Brasil	2003	IDV,SQV/RTV,NFV ou APV, LPV/RTV	10I 20R 43T 46L 48V 54V 82A 90M	8	4
		2006	LPV/RTV	10I/V 20R 43T 46L 48V 54V 62I/V 74S 82S 90M	10	4
1726/3399	Brasil	2005	NFV	20K/R 30N 36I/M 88D	4	2
		2007	ATV/RTV	20R 30N 88D	3	2
1786/3186	Brasil	2005	NFV,nenhum,LPV/RTV,nenhum	*	0	0
		2006	NFV	*	0	0
1797/3010	Brasil	2005	NFV,LPV/RTV	20T 46L 62I/V 71A/T 90M	5	2
		2006	LPV/RTV	20T 46L 90M	3	2
1802/1951	Brasil	2005	NFV	20R 30N 55R 62V 63P 74P 88D	7	2
		2005	NFV	20R 30N 55R 62V 63P 74P 88D	7	2

Em amarelo, pacientes em que o número de mutações aumentou da 1ª para a 2ª sequência.
Em cinza, pacientes em que o número de mutações diminuiu da 1ª para a 2ª sequência.

CAPÍTULO 6. DISCUSSÃO

6.1. 1ª ETAPA: PERFIL DA PROTEASE DE HIV-1 SUBTIPO F DE RECOMBINANTES BF

O presente trabalho analisou a prevalência e as características dos recombinantes BF na região pol entre 500 sequências previamente subtipadas como F ou recombinantes BF, e também verificou possíveis diferenças entre a protease encontrada no subtipo F e a protease encontrada nos recombinantes BF. Os resultados mostram uma frequência elevada de recombinantes BF na região pol, e, entre os recombinantes, uma elevada frequência da protease F.

O subtipo F é o segundo em prevalência na região sudeste do Brasil e o terceiro na região sul. Nestas localidades, os subtipos B, F e C cocirculam, o que favorece o aparecimento de formas recombinantes. Alguns estudos estimam o início das epidemias pelos subtipos B e F, em 1971 e 1981, respectivamente (Bello et al, 2006; Leal et al, 2008). Os subtipos B, F e C, juntamente aos recombinantes BF, BC e, mais recentemente, FC tem circulado entre os mesmos grupos de risco, mas com diferentes prevalências através das diversas regiões geográficas. Estas diferenças poderiam ser explicadas por efeitos fundadores locais ou através de diferentes entradas de cada um dos subtipos no país. A partir da análise da topologia de árvores filogenéticas geradas de sequências F e B brasileiras da região V3 do gene env, Bello e colaboradores (2006) sugerem que a epidemia de cada subtipo foi o resultado da introdução inicial de um pequeno número de cepas virais geneticamente muito próximas. Isto está em acordo com a prevalência relativamente baixa do subtipo F ao redor do mundo, o que dificulta um número elevado de entradas desse subtipo. Leal e colaboradores (2008) sugerem que os recombinantes BF apareceram quase simultaneamente à introdução do subtipo F. De

acordo com os autores, a rápida e completa assimilação do subtipo F resultou na emergência de recombinantes BF com porções variáveis de fragmentos B e F ao longo do genoma. Além disso, o complexo espectro de URFs poderia ser o resultado de novas recombinações e a completa assimilação do subtipo F pode ter reduzido as chances de este subtipo estabelecer uma epidemia. Outra observação feita pelos autores foi e de que os recombinantes BF apresentaram maior taxa de crescimento e maior capacidade de gerar novas infecções do que o subtipo F puro, o que sugere que a forma recombinante tem *fitness* mais elevado.

As análises realizadas no presente trabalho foram restritas à região pol. Algumas das sequências confirmadas como subtipo F e assim tratados ao longo do estudo podem de fato ser recombinantes BF. Um trabalho realizado na Argentina, onde duas epidemias estão em curso (subtipo B e recombinantes BF), mostrou que, ao serem analisados genomas completos, quase todas as amostras subtipadas como F eram recombinantes BF com grande diversidade de estruturas em mosaico (Thomson et al, 2002). Outro estudo, realizado com amostrar brasileiras, também mostra muitas sequências subtipadas como F que foram reclassificadas como recombinantes BF quando foi feito o sequenciamento quase completo do genoma (Sanabani et al, 2006). A única forma, portanto, de se obter a real classificação das sequências analisadas aqui seria a realização de sequencimentos de genoma completo ou quase completo, o que é praticamente inviável.

A presença de uma alta frequência de protease F entre os recombinantes corrobora algumas análises realizadas anteriormente com amostras brasileiras. Sanabani e colaboradores (2006) estudaram o genoma completo de 14 sequências da cidade de Santos, uma das mais importantes no estado de São Paulo para a epidemia de HIV/AIDS e observaram que, das oito sequências confirmadas como recombinantes BF,

cinco tinham protease F. Outro estudo analisou nove sequências completas de recombinantes BF e mostraram que, apesar da diversidade genética das estruturas em mosaico observadas entre as amostras, todas compartilhavam duas posições de quebra (*breakpoints*) correspondendo à região do gene gag e parte do pol compreendendo a protease e parte da transcriptase reversa. Toda esta região gag-pol foi classificada como F em todas as nove amostras (Sá-Filho et al, 2006). Das quatro CRFs descritas no Brasil (Sá-Filho et al, 2006; Guimarães et al, 2008), apenas a CRF39_BF (Rio de Janeiro) apresenta protease B.

A combinação de linhagens evolutivas diferentes é favorecida porque permite a combinação de propriedades adquiridas independentemente de uma forma rápida. É possível que os recombinantes apareçam como formas dominantes na quasiespecies porque alguns deles combinam características favoráveis presentes em suas linhagens parentais, tornando-se melhor adaptados que os demais componentes da população. Dessa maneira, é razoável se supor que a presença da protease do subtipo F esteja conferindo alguma vantagem quando a combinação ocorre entre os subtipos B e F. Apesar da presença de uma protease B na CRF12_BF, descrita na Argentina, vários estudos vem mostrando que amostras argentinas com pelo menos dois terços da protease classificada como F, de uma forma semelhante ao que se observa em amostras brasileiras, é bastante comum (Thomson et al, 2002; Quarleri et al, 2004).

A elevada prevalência de protease F e, também, o fato de que os recombinantes BF podem apresentar melhor *fitness* que o subtipo F puro, como sugerido por alguns autores (Leal et al, 2008), levantou a possibilidade de que a protease F presente nos recombinantes pudesse ter características próprias, que seriam responsáveis por conduzir um processo evolutivo diferente do apresentado pela protease do subtipo puro. A aplicação do MODELTEST com o grupo de sequências F e com o grupo de

sequências BF_{proF} para obter os melhor modelo evolutivo em cada caso, confirmou essa hipótese. Enquanto para o grupo de sequências F o modelo de melhor ajuste foi TVM+I+G, para o grupo BF_{proF} foi TPM 1uf+I+G. Outra evidência para essa hipótese foram as diferenças significativas observadas no perfil de mutações de ambos os grupos. Uma provável explicação para essas diferenças poderia estar na necessidade de uma melhor conformação espacial da protease em relação aos seus sítios de clivagem. Num genoma híbrido há a probabilidade de pontos de recombinação em diferentes regiões, incluindo as poliproteínas GAG e POL, importantes substratos da protease.

6.2. 2ª ETAPA: COVARIAÇÃO DAS MUTAÇÕES NA PROTEASE F, DETERMINAÇÃO DOS AGRUPAMENTOS ESPECÍFICOS (CLUSTERS) PARA ESTE SUBTIPO DE HIV-1, ASSOCIAÇÃO COM O USO DE IPs ESPECÍFICOS E ANÁLISE TEMPORAL DAS MUTAÇÕES ASSOCIADAS AO TRATAMENTO

Nesta etapa, foi analisado o perfil da protease do subtipo F em relação às mutações selecionadas pelo tratamento, presença de polimorfismos entre essas mutações e covariação das mutações associadas ao tratamento. Avaliou-se ainda a possível associação entre a presença de mutações consideradas raras no subtipo B, e o uso de inibidores de protease, bem como agrupamentos de mutações e sua associação com o uso de determinados IPs. Por último, foi feita a comparação entre duas sequências de protease F obtidas em momentos distintos para 38 pacientes, o que possibilitou avaliar a seleção de novas mutações, a possível reversão de mutações principais e secundárias e quais mutações permanecem após a troca ou trocas do esquema terapêutico.

Análise do perfil de mutações da protease F, polimorfismos, covariação e associação com o uso de IPs: A comparação das frequências de mutações de

indivíduos não tratados e tratados com IP mostrou 51 mutações positivamente associadas ao tratamento, todas em posições previamente descritas (Wu et al, 2003; Kantor et al, 2005; Rhee et al, 2005; Svicher et al, 2005). Algumas delas, no entanto, são consideradas bastante raras no subtipo B: V15I, G16E, G17E, K43T, K55R, T74P/S, N83D, I85V, Q92K e T96S (Shafer e Schapiro, 2008). As mutações K43T, K55R e T74S aparecem com frequências elevadas no grupo trF (6%, 9% e 14%, respectivamente), enquanto T74P, N83D, I85V, Q92K e T96S foram pouco frequentes.

K43T forma um subgrupo dentro do cluster 1 e aparentemente está associada ao uso de RTV, SQV e IDV. K55R compõe um subgrupo com a mutação principal M46L também dentro do cluster 1, mas com chances de associação ao uso de um outro grupo de IP: LPV, APV e ATV. A mutação T74S foi citada num estudo recente como responsável, juntamente com A71T, pela estabilização da mutação M89I, no subtipo G (Gonzales et al, 2008). M89I, por sua vez, em conjunto com a L90M, pode aumentar a resistência aos IP no subtipo G, quando comparada à ação da L90M isolada. Este foi considerado um novo perfil mutacional para resistência às drogas neste subtipo não B (Abecasis et al, 2005). No presente trabalho, observa-se que T74S clusteriza com M89I e A71T para um nível de similaridade entre 0,5 e 0,54, um pouco abaixo do que foi considerado para a montagem dos agrupamentos, além de individualmente clusterizar com NFV, assim como a mutação M89I. O par de mutações T74S e L90M está entre aqueles com valor de Jaccard positivo para o subtipo F, e com maiores chances de que a mutação T74S apareça antes da L90M. As três mutações A71T, T74S e M89I são polimorfismos naturais com freqüências 0,6%, 4% e 0,9%, respectivamente, entre os não tratados.

Sobre a mutação T96S, é importante ressaltar que há chances de sua presença ser resultado de efeito fundador e não de fato uma mutação associada ao tratamento. Todos

os pacientes que continham essa mutação eram provenientes de uma mesma região geográfica. Para distinguir mutações se desenvolvendo em vários indivíduos daquelas que se desenvolvem em um pequeno número de vírus fundadores, é necessário que seja feita a reconstrução de sequências ancestrais em cada nó de uma árvore filogenética para todas as sequências. O número de vezes que cada mutação está prevista para se desenvolver é contado e mutações para as quais vírus fundadores teriam contribuído para ≥50% das ocorrências não seriam consideradas associadas ao tratamento (Rhee et al, 2005).

Diferentes trabalhos têm mostrado que os subtipos não-B podem acumular a maioria das mutações de resistência encontradas no subtipo B tanto na protease como na transcriptase reversa (Kantor et al, 2005; Sá-Filho et al, 2008). Contudo, além das mutações raras no subtipo B que podem aparecer com frequências mais altas nos subtipos não-B, novas mutações subtipo-específicas associadas ao tratamento podem ser encontradas. Esse pode ser o caso das mutações Q18E, L19V e D35N. A posição 19 já havia sido citada por Kantor e colaboradores (2005) como possivelmente associada ao tratamento no subtipo F. No presente trabalho, a mutação L19V alcançou significância estatística para associação com o tratamento e, apesar de não participar de nenhuma combinação ou cluster particular de mutações, aparece associada a vários IPs (LPV, APV, SQV, RTV, IDV).

Das 51 mutações associadas ao tratamento, 13 são polimórficas (L10I, I13V, K14R, V15I, L19V, K20R, N37D, I62V, L63P, A71T, T74S, M89I, Q92K) e entre esses polimorfismos, nove são associados à resistência a um ou mais IPs (L10I, I13V, K20R, I62V, L63P, A71T, T74S, M89I e Q92K). É importante ressaltar que todo o estudo desenvolvido aqui teve como foco as mutações associadas estatisticamente ao tratamento. Contudo, há várias mutações associadas à resistência que não alcançaram

significância estatística para associação ao tratamento provavelmente porque já estavam com uma frequência elevada entre os não tratados, frequência essa que foi mantida elevada nos tratados. É o caso das mutações L10V, K20M, D60E, V77I, e I93L, todas polimórficas (frequências ≥ 0,5% entre não tratados). Se, por um lado a maioria dos estudos *in vitro* e *in vivo* sugerem que os IPs são tão ativos contra os subtipos não-B quanto são contra o B (Parkin et al, 2004; Geretti et al, 2006), alguns estudos mostram que a susceptibilidade aos IP na *baseline* para os subtipos não-B pode ser diferente, justamente pela presença natural de códons associados a algum grau de resistência aos IPs (Abecasis et al, 2006; Holguín e Soriano, 2002,).

Salvo a mutação G17E, associada ao tratamento, mas não associada a qualquer outra mutação ou IP, as demais mutações associadas as tratamento não polimórficas são particularmente importantes, pois devido à sua baixíssima prevalência entre os não tratados (<0,5%), podem ser importantes marcadores específicos de resistência transmitida. São elas: Q18E, K20T, L24I, D30N, V32I, L33F, D35N, I36L, K43T, M46I, I47L, G48V, I50V/L, F53L, I54V/A, K55R, Q58E, I64V, A71V, G73S, T74P, L76V, V82A/I/F/S/T, N83D, I84V, N88D, N88S, L90M e Q92K.

A aplicação do coeficiente de similaridade de Jaccard para duplas de mutações associadas ao tratamento permitiu verificar os pares com maiores chances de covariação no subtipo F. Alguns estudos realizados com o subtipo B utilizaram o coeficiente de correlação binomial *phi* ou a *mutual information* para análises de covariação (Wu et al, 2003; Hoffman et al, 2003). Segundo Rhee e colaboradores (2007), o coeficiente *phi* tem elevada sensibilidade a pares raros de mutações, enquanto a *mutual information* é totalmente insensível a esses mesmos pares. O coeficiente de Jaccard utiliza somente as sequências em que pelo menos uma das mutações do par estudado está presente, não

sendo, portanto, influenciado pelo total de sequências em que nenhuma das mutações está presente.

Entre os 51 pares de mutações com coeficiente de Jaccard > 0,2, 23 pares envolvem uma mutação principal e uma acessória. Os cálculos de probabilidade condicional mostram que, em 13 deles, a mutação principal tem maiores chances de ocorrer primeiro, enquanto em 10, a chance da acessória ocorrer primeiro é maior. A diferença não é estatisticamente significativa, corroborando resultados obtidos com a protease do subtipo B. Estudos recentes observaram que há um processo complexo envolvendo a ordem em que mutações principais e acessórias aparecem na protease B durante a terapia e que a designação de mutação principal como primária e acessória como secundária frequentemente refere-se somente aos seus papéis em relação à resistência e não à ordem em que aparecem (Rhee et al, 2007).

Há sete pares de mutações fortemente correlacionados no subtipo F ($J \geq 0,4$), entre eles, D30N/N88D e I54V/V82A, ambos compostos por duas mutações principais e com os mais altos valores de J (0,60 e 0,57, respectivamente). Comparando-se os valores do coeficiente de Jaccard da protease F com os valores da protease B obtidos de literatura (Rhee et al, 2007), observamos que esses mesmos pares têm os valores mais altos também para o subtipo B (0,59 e 0,51). Para o par D30N/N88D, Rhee e colaboradores (2007) sugerem que a mutação N88D antecede a D30N no processo de seleção, uma vez que P(D30N|N88D) > P(N88D|D30N). Os dados do subtipo F mostram valores iguais para P(D30N|N88D) e P(N88D|D30N), indicando que não deve haver preferência pela seleção primeiro de uma ou de outra.

Entre os 38 pacientes com duas sequências disponíveis,17 apresentaram a mutação I54V ou a V82A ou ambas. Desses 17, 7 (41%) apresentaram as duas mutações em ambas as sequências e nenhum paciente apresentou uma delas na primeira

sequência e ambas na segunda. O cálculo das probabilidades condicionais resultou em valores muito próximos para P(I54V|V82A) e P(V82A|I54V), 0,75 e 0,70, respectivamente. Esse resultado somado à elevada presença de pacientes com as duas mutações em ambas as sequências e à ausência de pacientes com apenas uma delas na primeira sequência e ambas na segunda sugere que no subtipo F talvez não haja um sentido preferencial para o aparecimento deste par de mutações. No subtipo B, a tendência é do aparecimento da mutação V82A primeiro (Rhee et al, 2007).

Outros dois pares fortemente correlacionados na protease F também têm valores de *J* elevados no subtipo B (L10I/I54V e L10I/V82A). Os pares K20R/I54V e K20R/V82A, ambos com $J \geq 0,4$ no subtipo F, apresentam valores bem mais baixos no subtipo B (0,20 e 0,17, respectivamente). L10I e K20R são polimorfismos comuns no subtipo F enquanto V82A e I54V são duas mutações principais. Estas quatro mutações, juntamente com L24I, formam um subgrupo de mutações que clusterizam com vários IPs (LPV, APV, IDV, SQV e RTV).

Além do par D30N/N88D e do grupo de mutações que compõem um cluster com o par I54V/V82A, um terceiro grupo de mutações envolvendo a principal L90M pode ser observado. No subtipo B, o par mais fortemente correlacionado envolvendo essa mutação parece ser L90M/A71V (J=0,38; Rhee et al, 2007); enquanto no F, apesar do par L90M/A71V estar entre os significativos (J=0,29), esta mutação parece mais fortemente correlacionada com L63P e com a principal I54V (J=0,34 e 0,30, respectivamente).

A análise dos dados de pacientes com duas sequências revelou uma possível tendência na protease do subtipo F. Entre os 16 pacientes que apresentaram aumento do número total de mutações, apenas um demonstrou reversão de mutação principal (L90M), apesar da troca de esquema terapêutico. Tampouco entre os 13 pacientes que

mantiveram o número total de mutações inalterado houve reversão de mutações principais. Entre os pacientes em que o número de mutações reduziu da primeira para a segunda sequência, observou-se reversão de mutações acessórias, mas não de mutações principais. O único evento de reversão de mutação principal foi observado para um paciente que, embora aumentasse o número de mutações após a troca do esquema terapêutico, deixou de apresentar a mutação L90M. Esses dados sugerem que as mutações principais, uma vez presentes no subtipo F, tendem a permanecer apesar da troca de esquemas de tratamento.

A maioria dos resultados aqui mostrados e discutidos utilizando-se a técnica HCA também foi obtida de forma muito semelhante pela técnica PCA. Basicamente, os mesmos agrupamentos foram observados, mas muitas vezes com dificuldade de visualização pela quantidade de informação na área do gráfico. Optou-se por mostrar apenas os resultados obtidos pela HCA, pela maior facilidade para a compreensão. Contudo, seria interessante um trabalho mais aprofundado para a comparação entre ambas as metodologias utilizando-se dados de sequências, por exemplo, do subtipo B. Por ser este um subtipo bastante estudado, presta-se melhor a uma comparação de métodos, o que estava além do escopo deste trabalho.

CAPÍTULO 7. CONCLUSÃO

O processo de recombinação como fator fundamental que contribui para o aumento da diversidade viral já foi há muito reconhecido e as formas recombinantes de HIV-1 vem ocupando destaque na pandemia global. Este trabalho corrobora alguns trabalhos anteriores que sugerem que a prevalência do subtipo F puro é significativamente mais baixa que a prevalência do recombinante BF (Sanabani et al, 2006; Sá-Filho et al, 2008). Trouxe também outros dados importantes sobre a epidemia no Brasil: a presença preponderante da protease F entre os recombinantes BF, apesar da grande diversidade de estruturas em mosaico, e o fato de que esta protease F pode estar evoluindo de forma diferente da protease encontrada no subtipo F. Análises de genoma completo poderiam fornecer mais informações sobre *break points*, possíveis *hot spots* e estruturas em mosaico dos recombinantes BF brasileiros.

Este estudo também trouxe avanços no conhecimento sobre polimorfismos naturais, mutações associadas ao tratamento, perfis mutacionais associados ao uso de IPs e dados temporais sobre as mutações associadas ao tratamento na protease do subtipo F. A maioria dos estudos anteriores sobre o subtipo F baseia-se em pequeno número de sequências, o que dificulta análises com suporte estatístico. Neste trabalho foram utilizadas 1265 sequências entre ntr e tr, o que possibilitou resultados estatisticamente significativos.

Embora a grande maioria das mutações associadas ao tratamento e polimorfismos encontrados também esteja presente no subtipo B, características típicas ou até mesmo exclusivas do subtipo F puderam ser apontadas. A partir desses dados, tornam-se possíveis estudos avançados e a utilização de modelos como o proposto por Beerenwinkel e colaboradores (2004), capazes de traçar os passos para o acúmulo de

mutações de resistência na protease e na TR. Até então, todos os modelos contemplaram quase que exclusivamente o subtipo B, em face do pequeno número de dados disponíveis sobre os demais subtipos.

CAPÍTULO 8. ANEXOS

Anexo A. Sequência referência F1.BR.01.01BR125

PQITLWQRPLVTIRVGGQLKEALLDTGADDTVLEDINLPGKWKPKMIGGIGGFIKVKQYDNILIEI
CGHKAIGTVLVGPTPVNIIGRNMLTQIGCTLNF

Anexo B. Números de acesso das sequências

U19416 AF005494 AF076448 AF076441 AF076439 AF079996 AF079995 AF079989 AF079988 AF079987 AF079986 AF079985 AF125284 AF075703 AF221037 AF204042 AF204041 AF204044 AF204043 AF204045 AF204047 AF204049 AF204040 AF204050 AF204048 AF204052 AF204046 AF204051 AF204039 AF247521 AF247520 AJ249236 AJ249237 AJ286971 AJ286972 AJ286973 AJ286974 AJ286975 AJ286976 AJ286977 AJ287042 AJ249238 AF295283 AF295282 AF295299 AF295281 AF295298 AF377956 AF377958 AF259932 AF357611 AF357640 AF357701 AY032519 AF329582 AF329586 AF329587 AF329589 AF329591 AF329598 AF329601 AF329602 AF328562 AF328563 AF328565 AF328566 AF328570 AF328572 AF328573 AF328575 AF328579 AF328586 AF328587 AF328589 AF328591 AF328593 AF328594 AJ012427 AY010407 AY010408 AY010432 AY145824 AY145825 AY145826 AY145818 AY900823 AF467727 AF425459 AF413832 AY051202 AY051201 AY051200 AY051199 AY173957 AY173958 AY900876 AY900890 AY900896 AY900939 AY900940 AY900941 AY900945 AY900947 AF355313 AF332867 AF408627 AF408628 AF408629 AF408630 AF408631 AF408632 AY714439 AY213321 AY213340 AY213449 AY213196 AY213356 AY213227 AY213319 AY213535 AY213215 AY213346 AY213277 AY213527 AY213523 AY213422 AY213549 AY213221 AY213270 AY213326 AF447815 AF447822 AF447826 AF447853 AF447857 AY204311 AY204321 EF548517 EF549324 AF527190 AF527209 AF527238 AF527244 AF527254 AF527311 AF527317 AF527326 AF527346 AF527352 AF527354 AF527360 AF527363 AB097866 AY444279 AY444300 AY444278 AY444303 AY444277 AY444296 AY444297 AY444301 AY444281 AY444298 AY444299 AY444275 AY444276 AY444302 AY444280 AY444274 AY444273 AY444291 AY444290 AY371158 AY371170 AY359692 AY359693 AY359694 AY359698 AY528077 AY569862 AY569869 AY569872 AY569900 AY569901 AY569841 AY569878 AY569882 AY569834 AY569857 AY569888 AY569889 AY569830 AY313300 AY313318 AY313322 AY313304 AY313305 AY313310 AY313311 AY536233 AY536234 AY536235 AY536238 AY900963 AY455782 AY471923 DQ177219 DQ189088 AY611671 AY771589 AY771590 AY248334 AY626953 AY798073 AY643181 AY643185 AY643203 AY643149 AY643172 AY643161 AY643135 AY781128 AY781129 DQ085871 DQ085875 DQ085876 DQ461844 AM279447 DQ443624 DQ443681 DQ394167 DQ394169 DQ383747 DQ383753 DQ383754 DQ206645 DQ206646 DQ206647 DQ206648 DQ206649 DQ206650 DQ206651 DQ206652 DQ206653 DQ206654 DQ206655 DQ206656 DQ206657 DQ206658 DQ206659 DQ206660 DQ206661 DQ206662 DQ206663 DQ345494 DQ345495 DQ345496 DQ345497 DQ345498 DQ345499 DQ345500 DQ345501 DQ345502 DQ345503 DQ345504 DQ345505 DQ070592 DQ070589 DQ070591 DQ070599 DQ070590 AM181794 DQ297242 DQ297266 DQ297200 AM041021 AM041018 AM040990 AM040988 EF368934 EF120076 EF120077 EF120078 EF120079

EF120081 EF120082 EF120083 EF120085 EF120086 EF120089 EF120091 EF120092 EF120093 EF120094 EF120096 EF120100 EF120101 EF120104 EF120106 EF120107 EF120108 EF120109 EF120110 EF120111 EF120113 EF120114 EF120117 EF120118 EF120119 EF120120 EF120121 EF120122 EF120123 EF120127 EF120128 EF120129 EF120130 EF120132 EF120134 EF120137 EF120138 EF120139 EF120140 EF120141 EF120142 EF120143 EF120144 EF120145 EF120146 EF120147 EF120149 EF120151 EF120152 EF120154 EF120156 EF120159 EF120160 EF120161 EF120163 EF120164 EF120166 EF120167 EF120169 EF120170 EF120172 EF120173 EF120174 EF120176 EF120177 EF120178 EF120180 EF120182 EF120188 EF120189 EF120191 EF120192 EF120193 EF120195 EF120196 EF120199 EF120204 EF120205 EF120208 EF120210 EF120211 EF120213 EF120214 EF120216 EF120217 EF120218 EF120220 EF120221 EF120222 EF120224 EF120225 EF120227 EF120228 EF120231 EF120232 EF120233 EF120236 EF120238 EF120240 EF120241 EF120251 EF120252 EF120253 EF120255 EF120256 EF120257 EF120258 EF120260 EF120261 EF120272 EF120275 EF120278 EF120279 EF120280 EF120283 EF120284 EF120285 EF120290 EF120292 EF120293 EF120300 EF120301 EF120303 EF120304 EF120308 EF120309 EF120311 EF120313 EF120321 EF120325 EF120328 EF120340 EF120343 EF120344 EF120348 EF120349 EF120350 EF120352 EF120353 EF120355 DQ899706 DQ899707 DQ899709 EF042629 EF042644 EF042648 EF042650 EF042651 EF042657 EF042665 EF042667 EF042670 EF042671 EF042677 EF042685 EF042688 EF432685 EU312668 EU312690 EU312704 EU312746 EU312756 EU312864 EU312877 EU312884 EU312900 EU312906 EU312918 EU312966 DQ869211 AF251986 AF252092 AF252045 AF221055 AF221052 AF221051 AF221050 AF221049 AF221048 AF221046 AF221045 AF221044 AF221042 AF221041 AF221040 AF221039 AF221035 AF165581 AF165583 AF165584 AF165585 AF369798 AF357747 AY030534 AY031032 AY031100 AY031685 AY032567 AJ401975 AY145817 AY145827 AY145828 AY145819 AY145830 AY900897 AY145811 AY145832 AY900822 AY047419 AF493407 AY900880 AY900887 AY900888 AY677410 AY333463 AY333504 AY333484 AY333471 AY333526 AY333459 AY333523 AY333498 AY333456 AY333460 AY333500 AY333496 AY333503 AY333487 AY333517 AY333452 AY333511 AY333513 AY333519 AY333493 AY333478 AY333521 AY333491 AY333461 AY333488 AY333501 AY333465 AY333507 AY333520 AY333509 AY333453 AY333473 AY333454 AY333457 AY182788 AY182784 AY182818 AY182794 AY182827 AY182829 AY182808 AY798450 EF549325 AY275717 AY275750 AF544464 AY390200 AY390086 AY390088 AY390199 AY313361 AY313342 AY313357 AY313358 AJ577734 AY455778 AY455783 AY455780 AY798416 AY796976 AY796998 AY797026 AY797027 AY798433 AY797454 AY643151 AY441798 AY441820 AY643180 AY643188 AY441814 AY441803 AY643170 AY643139 AY643146 AY643162 AY643156 AY643166 AY643175 AY643158 AY643176 AY643134 AY643159 AY643155 AY643143 AY643168 AY643144 AY643178 AY643169 AY643147 AY643140 AY643163 AY643148 AY643171 AY643165 AY643173 AY643153 AY643174 AY643141 AY643177 AY643179 AY643167 AY643145 AY643142 AY643137 AY643138 DQ879128 DQ879110 DQ877258 DQ877259 DQ877261 DQ877262 DQ877264 DQ877265 DQ877267 DQ877269 DQ877270 DQ877274 DQ877275 DQ877414 DQ877311 DQ877312 DQ877313 DQ877316 DQ877320 DQ877321 DQ877322 DQ877323 DQ877325 DQ877327 DQ877332 DQ877333 DQ877334 DQ877335 DQ877341 DQ877342 DQ877344 DQ877345 DQ877346 DQ877347 DQ877349 DQ877351 DQ877353 DQ877354 DQ877355 DQ877356 DQ877358 DQ877361 DQ877368 DQ877370 DQ877378 DQ877379 DQ877381 DQ877385 DQ877392 DQ877393 DQ877394 DQ877395 DQ877397 DQ877398 DQ877399 DQ877401 DQ877407 DQ877409 DQ877410 DQ877411 DQ877417 DQ877418

DQ877419 DQ877421 DQ877422 DQ877425 DQ877428 DQ877750 DQ877756 DQ877768 DQ877792 DQ877885 DQ877904 DQ877955 DQ878019 DQ878171 DQ878175 DQ878258 DQ878509 DQ878568 DQ878597 DQ878603 DQ878622 DQ878658 DQ878673 DQ878705 DQ878732 DQ878750 DQ878833 DQ878842 DQ878905 DQ877285 DQ877288 DQ877289 DQ877291 DQ877292 DQ877304 DQ877308 DQ877363 DQ877429 DQ877499 DQ877501 DQ877511 DQ877516 DQ877529 DQ877547 DQ877549 DQ877552 DQ877566 DQ877567 DQ877572 DQ877579 DQ877602 DQ877635 DQ877653 DQ877655 DQ877663 DQ877697 DQ877698 DQ877707 DQ877714 DQ877716 DQ877726 DQ877727 DQ877731 DQ877734 DQ879095 DQ358799 DQ358800 DQ358801 DQ358802 DQ358803 DQ358804 DQ358811 DQ358812 DQ899704 DQ899711 EF042646 EF193885 EF193927 EF193945 EF193966 EF449996 EF450023

CAPÍTULO 9. REFERÊNCIAS BIBLIOGRÁFICAS

Abecasis AB, Deforche K, Bacheler LT McKenna P, Carvalho AP, Gomes P, Vandamme A-M; Camacho RJ. Investigation of baseline susceptibility to protease inhibitors in HIV-1 subtypes C, F, G and CRF02_AG. *Antivir Ther* 2006,11: 581–9.

Abecasis AB, Deforchea K, Snoecka J, Bachelerb LT, McKennac P, Carvalho AP, Gomesd P, Camachod RJ and Vandamme AM. Protease mutation M89I/V is linked to therapy failure in patients infected with the HIV-1 non-B subtypes C, F or G. *AIDS* 2005, 19*:*1799–1806

Anderberg MR. *Cluster analysis for applications*. New York: Academic Press,1973.

Avila MM, Pando MA, Carrion G, Martinez Peralta L, Salomon H, Gomez Carrillo M, Sanchez J, maulen S, Hierholzer J, Marinello M, Negrete M, Russel KL, Carr TK. Two HIV-1 epidemics in Argentina: Different genetic subtypes associated with different risk groups. *J Acquir Immune Defic Syndr* 2002;29(4):422–426.

Bartlett JA, Fath MJ, Demasi R, Hermes A, Quinn J, Mondou E, Rousseau F. An updated systematic overview of triple combination therapy in antiretroviral-naive HIV-infected adults. *AIDS* 2006; 20(16):2051-64.

Bello G, Guimarães ML, Morgado MG. Evolutionary history of HIV-1 subtype B and F infections in Brazil. *AIDS* 2006;20:763-768.

Beerenwinkel N, Rahnenführer J, Däumer M, Hoffmann D, Kaiser R, Selbig J, Lengauer T. Learning Multiple Evolutionary Pathwaysfrom Cross-sectional Data. *RECOMB'04* 2004; 27–31.

Benjamini Y, Hockberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statistical Soc Ser B-Methodological. 1995;57:289-300.

Biswas P, Tambussi G, Lazzarin A. Access denied? The status of co-receptor inhibition to counter HIV entry. (abstract page). *Expert Opin Pharmacother 2007;**8** (7): 923–33.*

Boulerice F, Li XG, Lvovich A, Wainberg MA. Recovery of infectious Human immunodeficiency virus type 1 after fusion of defectively infected clones of U-937 cells. *J Virol* 1991;65:5589-5592.

Brindeiro MR, Diaz RS, Sabino EC, Morgado MG, Pires LI, Brigido L, Dantas MC, Barreira D, Teixeira PR, Tanuri A and the Brazilian Network For Drug Resistance Surveillance. Brazilian network for drug resistance surveillance (HIV-BResNet): a survey of chronically infected individuals. *AIDS*. 2003;17:1063-1069.

Briones, C, Domingo, E, Molina-París, C. Memory in retroviral quasispecies: experimental evidence and theoretical model for human immunodeficiency virus. *J Mol Biol* 2003;331:213–229.

Boletim epidemiológico DST/AIDS 2007 Ministério da Saúde; iv(1).

Buonaguro L, Tornesello ML, Buonaguro FM. HIV-1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *J Virol* 2007; 81(19):10209-10219

Carr JK, Avila M, Gomez Carrillo M, et al.: Diverse BF recombinants have spread widely since the introduction of HIV-1 into South America. *AIDS* 2001;15:F41-47

Cavalcanti AMS, Lacerda HR, Brito AM, Pereira S, Medeiros D, Oliveira S. Antiretroviral resistance in individuals presenting therapeutic failure and subtypes of the human immunodeficiency virus type 1 in the Northeast Region of Brazil. *Mem Inst Oswaldo Cruz* 2007;102(7): 785-792

Cazar, R. A. "An Exercise on Chemometrics for a Quantitative Analysis Course". *Journal of Chemical Education* 2003;80(9):1026-1029.

Cerqueira D M, Amorim R M S, Silva R R, Camara G N L, Brígido M M, Martins C R F. Antiretroviral resistance and genetic diversity of human immunodeficiency virus type 1 isolates from the Federal District, Central Brazil. *Mem Inst Oswaldo Cruz* 2004;99(8):877-882.

Chakraborty H, Sena PK, Helmsa RW, Vernazzab PL, Fiscusc SA, Erond JJ, Pattersone BK, Coombsf RW, Kriegerg JN, Cohend MS. Viral burden in genital secretions determines male-to-female sexual transmission of HIV-1: a probabilistic empiric model. *AIDS* 2001;15:621-627.

Clark S, Calef C, Mellors J (2007). Mutations in Retroviral Genes Associated with Drug Resistance pp. 58-158 in HIV Sequence Compendium 2006/2007. Edited by: Thomas Leitner T, Foley B, Hahn B, Marx P, McCutchan F, Mellors J, Wolinsky S, Korber B. Published by: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM. LA-UR 07-4826

Clavel F, Hoggan MD, Willey RL, Strebel K, Martin MA, Repaske R. Genetic recombination of human immunodeficiency virus. *J Virol* 1989;63:1455-1459.

Moita-Neto, 2005criticanarede.com · ISSN 1749-8457 (capturado em 04/05/2009).

Deeks SG, Kar S, Gubernick SI, Kirkpatrick P. Raltegravir. *Nature Reviews Drug Discovery* 2008;7:117-118.

Diaz R S. Guia para o manuseio de testes de resistência anti-retroviral no paciente infectado pelo HIV-1. ABBOTT Laboratórios do Brasil. 2004.

Domingo E, Baranowsky E, Ruiz-Jarajb C, Martin-Hernandez AM, Saiz JC, Escarnis C. Quasispecies Structure and persistance of RNA viruses. *Emerging Infectious Diseases* 1998;4(4):521-527.

Eigen, M. On the nature of virus quasispecies. *Trends Microbiol*. 1996;4:216–218.

Elder, R.T., Yu, M., Chen, M., Zhu, X., Yanagida, M., Zhao, Y. HIV-1 Vpr induces cell cycle G2 arrest in fission yeast (Schizosaccharomyces pombe) through a pathway involving regulatory and catalytic subunits of PP2A and acting on both Wee1 and Cdc25. *Virology* 2001;287 (2), 359–370.

Erickson JW, Gulnik SV, Markowitz M. Protease inhibitors resistance, cross resistance, fitness and the choice of initial and salvage therapies. *AIDS* 1999;13 (suppl A): S189-204.

Fauci AS. 25 Years of HIV/AIDS Scineci: reaching the poor with research advances. *Cell* 2007;2:429-432.

Fraser C, Hollingsworth TD, Chapman R, Wolf F, Hanage WP. Variation in HIV-1 set-point viral load: Epidemiological analysis and an evolutionary hypothesis. *PNAS* 2007;44(104):17441–17446.

Gadelha SR, Shindo N, Cruz JNM, Morgado MG, Galvão-Castro B. Molecular epidemiology of human immunodeficiency virus-1 in the state of Ceará, Northeast, Brazil. *Mem Inst Oswaldo Cruz* 2003;98(4):461-4.

García-Lerma, J.G., Nidtha, S., Blumoff, K.,Weinstock, H., Heneine,W. Increased ability for selection of zidovudine resistance in a distinct class of wild-type HIV-1 from drug-naive persons. *Proc. Natl. Acad. Sci. U. S. A.* 2001;98:13907–13912.

Gathe JC Jr, Ive P, Wood R, Schürmann D, Bellos NC, DeJesus E, Gladysz A, Garris C, YeoJ. SOLO: 48-week efficacy and safety comparison of once-daily fosamprenavir /ritonavir versus twice-daily nelfinavir in naive HIV-1-infected patients. *AIDS* 2004; 18:1529-37.

Gonzalez LMF, Santos AS, Abecasis AB, Van Laethem K, Soares EA, Deforche K, Tanuri A, Camacho R, Vandamme AM and Soares MA. Impact of HIV-1 protease mutations A71V/T and T74S on M89I/V-mediated protease inhibitor resistance in subtype G isolates. *J Antimicrob Chemotherapy* 2008; 61:1201–1204.

Gotlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, Wolf RA, Saxon A. Pneumocystis carinii pneumonia and mucosa candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *New Engl J Med.* 1981;305:1425-1431.

Grant AD, Cock KMD. ABC of AIDS – HIV infection and AIDS in the developing world. *BMJ*. 2001;322:1475-1478.

Greco DB, Simão M. Brazilian policy of universal access to AIDS treatment: sustainaility, challenges and perspectives. *AIDS* 2007; 21:34-45.

Greene WC, Peterlin BM. Molecular Insigths into HIV Biology. HIV InSite Knowledge Base Chapter. February 2003; Content reviewed October 2005. Available from http://hivinsite.ucsf.edu/InSite?page=kb-00&doc=kb-02-01-02 (Arquivo capturado em 13 de abril de 2007).

Guimarães ML, Ever-Silva WA, Couto-Fernandez JC, Morgado MG. Identification of two new CRF_BF in Rio de Janeiro State, Brazil. *AIDS* 2008;22(3):433-5.

Hahn BH, Shaw GM, De Cock KM, Sharp PM. AIDS as a zoonosis: scientific and public health implications. *Science.* 2000;287:607-614.

Hammer SM, Saag MS, Schechter M, Montaner JSG, Schooley RT, Jacobsen DM, Thompson MA, Carpenter CCJ, Fischl MA, Gazzard BG, Gatell JM, Hirsch MS, Katzenstein DA, Richman DD, Vella S, Yeni PG, Volberding PA. Treatment for Adult HIV Infection. 2006 Recommendations of the International AIDS Society–USA Panel. JAMA. 2006;296:827-843.

Hoffman NG, Schiffer CA, Swanstron R. Covariation of amino acid positions in HIV-1 protease. *Virol* 2003;314:536-548.

Holguín A and Soriano V. Resistance to Antiretroviral Agents in Individuals with HIV-1 Non-B Subtypes. *HIV Clin Trials* 2002;3(4):403–411

Holguín A, Pena MJ, Troncoso F and Soriano V. Introduction of non-B subtypes among Spaniards newly diagnosed with HIV type 1 in the Canary Islands. *AIDS Res. Hum Retrovirus* 2007;4:498–502.

ITRN and ITRNN. http://aids.about.com/od/hivaidslettern/g/nrt.htm, capturado em 30/05/2009.

Jaccard P. Nouvelles recherché sur la distribution florale. *Bull Soc. Vaudoise Sci Nat* 1908,44:223-270.

Johnson VA, Brun-Vézinet F, Clotet B, Gunthard HF, Kuritzkes DR, Pillay D, Schapiro JM, Richman DD. Update of the drug resistance mutations in HIV-1:2007. *Top. HIV Med* 2007;15(4):119-125.

Jowett JB, Planelles V, Poon B, Shah NP, Chen ML, Chen lS. The human immunodeficiency virus type 1 vpr gene arrests infected T cells in the G2 + M phase of the cell cycle. *J Virol* 1995 69: 6304–6313.

Kantor R, Katzenstein D A, Efron E, Carvalho A P, Wynhoven, B, Cane P, Clarke J, Sirivichayakul S, Soares M A, Snoeck J, Pillay C, Rudich H, Rodrigues R, Holguin A, Ariyoshi K, Bouzas M B, Cahn P, Sugiura W, Soriano V, Brigido L F, Grossman Z, Morris L, Vandamme Am, Tanuri A, Phanuphak P, Weber J N, Pillay D, Harrigan P R, Camacho R, Schapiro J M, Shafer R W. Impact of HIV-1 subtype and antiretroviral therapy on Protease and reverse transcriptase genotype: results of a global collaboration. *PloS Med*. 2005;2(4):e112.

Kaufman L, Rousseeuw PJ. *Finding groups in data*: an introduction to cluster analysis. New York: Wiley, 1990.

Leal E, Martins L, Janini M, Diaz R. Evolutionary dynamics of HIV-1 BFD and CB recombinants in South America. In Abstract book of 15th Conference on Retroviruses and Opportunistic Infections – CROI. Boston, February 3-6, 2008.

Los Alamos National Laboratory HIV Database http://www.hiv.lanl.gov.

Machtinger, EL, Bangsberg DR. Adherence to HIV Antiretroviral Therapy HIV InSite Knowledge Base Chapter. May 2005. from http://hivinsite.ucsf.edu/InSite?page=kb-00&doc=kb-03-02-09

Maslin J, Rogier C, Caron M, Grandadam M, Koeck JL, Nicand E. Short communication. Antiretroviral drug resistance among drug-naive HIV-1-infected individuals in Djibouti (Horn of Africa). *Antivir Ther.* 2005;10(7):855-9.

Molla A, Kempf D, Korneiva M, Gao Q, Shipper P, Mo H, Markowitz M, Vasavanonda S, Chernyavisky T, Niu P, Lyons N, Hsu A, Granneman G, Ho D, Boucher C, Leonard J, Norbeck D. Ordered accumulation of mutations in HIV protease confers resistance to ritonavir. *Nat Med*. 1996;2:760-766.

Morgado MG, Guimaraes ML, Gripp CB, Costa CI, Neves I Jr, Veloso VG, *et al*. Molecular epidemiology of HIV-1 in Brazil: high prevalence of HIV-1 subtype B and identification of an HIV-1 subtype D infection in the city of Rio de Janeiro, Brazil. Evandro Chagas Hospital AIDS Clinical Research Group. *J Acquir Immune Defic Syndr Hum Retrovirol* 1998; 18:488-494

Mouth L, Corbeil J, Richman DD. Recombination leads to the rapid emergence of HIV-1 dually resistant mutants under selective drug pressure. *Proc Natl Acad Sci USA* 1996;93:6106-6111.

Nájera R, Delgado E, Pérez-Alvarez L, Thomson MM. Genetic recombination and its role in the development of the HIV-1 pandemic. *AIDS* 2002;16(suppl):S3-S16

Okie S. Fighting HIV. Lessons from Brazil. *N Engl J Med* 2006; 354:1977-81.

Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. Department of Health and Human Services. November 3, 2008; 1-139. Available at

http://www.aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf. Capturado em 31/05/2009 p.26-28.

Peng C, HO BK, Chang TW, Chang NT. Role of human immunodeficiency virus type 1-specifc protease in core protein maturation and viral infectivity. *J Virol* 1989,63(6):2550-2556.

Prado PI, Lewinsohn TM, Carmo R L, Hogan D J. "Ordenação Multivariada na Ecologia e seu Uso em Ciências Ambientais." *Ambiente e Sociedade,* Campinas, SP: v.10, p. 69-83, 2002.

Programa Nacional para DST e AIDS, Ministério da Saúde. Epidemiologia. http://www.aids.gov.br/ capturado em 15/Nov/2007.

Pugach P, Ketas TJ, Michael E, Moore JP. Neutralizing antibody anti-retroviral drug sensitivities of HIV-1 isolates resistant to small molcule CCR5 inhibitors. *Virol* 2008;377(2):401-407.

Quarleri JF, Rubio A, Carobene M, Turk G, Vignoles M, Harrigan RP, Montaner JSG, Salomón H and Gómez-Carrillo M. HIV Type 1 BF Recombinant Strains Exhibit Different *pol* Gene Mosaic Patterns: Descriptive Analysis from 284 Patients under Treatment Failure. *AIDS Res. Hum Retrovirus* 2004; 20(10):1100–1107.

Recomendações para Terapia Anti-retroviral em Adultos Infectados pelo HIV:2008/Ministério da Saúde, Secretaria de Vigilância em Saúde, Programa Nacional de DST e Aids. 7a Ed.- Brasília: Ministério da Saúde, 2008.

Relucio K & Holodny M. HIV-1 RNA and viral load. *Clin Lab Med* 2002;22(3):593-610.

Rhee SY, Fessel WJ, Zolopa AR, Hurley R, Liu T, et al. HIV-1 protease and reverse transcriptase mutations: correlations with antiretroviral therapy in subtype B isolates and implications for drug-resistance surveillance. *J Infect Dis* 2005;192:456-465.

Rhee SY, Liu TF, Holmes SP, Shafer RW. HIV-1 subtype B protease and reverse transcriptase amino acid covariation. *PLoS Comput Biol* 2007,3(5):e87. Doi:10.1371/journal.pcbi.0030087

Riddler SA, Haubrich R, DiRienzo G et all. A prospective,randomized, phase III trial of NRTI-, IP-, and NNRTI-sparingregimens for initial treatment of HIV-1 infection: ACTG 5142. XVI International AIIDS Conference. Toronto,Canada, August 2006.

Robbins GK, De Gruttola V, Shafer RW, et al. Comparison of sequential three-drug regimens as initial therapy for HIV-1 infection. *N Engl J Med*, 2003. 349(24):2293-303.

Rodrigues R, Scherer LC, Oliveira CM, Franco HM, Sperhacke RD, Ferreira JL, Castro SM, Stella IM, Brigido LF. Low prevalence of primary antiretroviral resistance mutations and predominance of HIV-1 clade C at polymerase gene in newly diagnosed individuals from south Brazil. *Virus Res*. 2006 Mar;116(1-2):201-7. Epub 2005 Dec 5.

Romesburg HC. Cluster Analysis for Researchers. Lulu. Press, 2004 , pp. 334 (CAP 2 e 10).

Sabino EC, Diaz RS, Brigido LF, Learn GH, Mullins JI, Reingold AL, Duarte AJS, Mayear A,Bush MP. Distribution of HIV-1 subtypes seen in an AIDS clinic in Sao Paulo City, Brazil. *AIDS* 1996; 10:1579-1584.

Sanabani S, Kleine Neto W, Kalmar EMN, Diaz RS, Janini LM, Ester Cerdeira Sabino EC. Analysis of the near full length genomes of HIV-1 subtypes B, F and BF recombinant from a cohort of 14 patients in São Paulo, Brazil. *Infection, Genetics and Evolution* 2006; 6:368–377.

Sanabani S, Kleine-Neto W, Sa- Filho DJ, Diaz RS, Munerato P, Janini LM, Sabino ES. Full-length genome analysis of human immunodeficiency virus type 1 subtype C in Brazil. *AIDS Res Hum Retroviruses* 2006; 22(2):171–176.

Sá-Filho DJ, Soares MS, Candido V, Gagliani LH, Cavaliere E, Diaz RS, Caseiro MM. HIV Type 1 pol Gene Diversity and Antiretroviral Drug Resistance Mutations in Santos, Brazil. *AIDS Res Hum Retrovirus* 2008;24(3), DOI: 10.1089/aid.2007.0203.

Sá-Filho DJ, Sucupira MC, Caseiro MM, Sabino EC, Diaz RS, And Janini LM. Identification of two HIV Type 1 circulating recombinant forms in Brazil. *AIDS Res. Hum Retrovirus* 2006;22(1):1–13

Segura M , Estani SS, Marone R, Bautista CT, Pando MA, Eyzaguirre L, Sánchez JL, Carr JK, Montano, Weissenbacher M, And Ávila MM. Buenos Aires cohort of men who have sex with men: prevalence, incidence, risk factors, and molecular genotyping of HIV Type 1. *AIDS Res. Hum. Retrovirus* 2007;23(11): 1322–1329

Serwadda D, Gray RH, Sewankambo NK, Wabwire-Mangen F, Chen MZ, Quinn TC, Lutalo T, Kiwanuka N, Kigozi G, Nalugoda F, Meehan MP, Ashley Morrow R, Wawer MJ. Human immunodeficiency virus acquisition associated with genital ulcer disease and herpes simplex virus type 2 infection: a nested case-control study in Rakai, Uganda. *J Infect Dis* 2003;188:1492-1497.

Shafer RW. Genotypic testing for human immunodeficiency virus type 1 drug resistance. *Clinical Microbiology Reviews*. 2002;15:247-277.

Shafer RW, Rhee SY, Bennett DE. Consensus Drug Resistance Mutations for Epidemiological Surveillance: Basic Principles and Potential Controversies. *Antiviral Therap* 2008;13 Suppl 2:59-68.

Shafer RW and Schapiro JM. HIV-1 Drug Resistance Mutations: an Updated Framework for the Second Decade of HAART. *AIDS rev* 2008;10:67-84.

Soares EAJM, Santos PR, Pellegrini JA, Sprinz E, Tanuri A, Soares MA. Epidemiologic and molecular characterization of Human Immunodeficiency virus type 1 in Southern Brazil. *J Acquir Immune Defic Syndr*. 2003;34(5):520-526.

Soares MA, Brindeiro RM, Tanuri A. Primary HIV-1 drug resistance in Brazil. *AIDS.* 2004 Jun;18 Suppl 3:S9-13.

Staszewski S, Morales-Ramirez J, Tashima KT, et al. Efavirenz plus zidovudine and lamivudine, efavirenz plus indinavir, and indinavir plus zidovudine and lamivudine in the treatment of HIV-1 infection in adults. *N Engl J Med*, 1999. 341(25):1865-73.

Stanford HIV Drug Resistance Database. http://hivdb.stanford.edu.

Stevenson M. HIV-1 pathogenesis. *Nature Medicine* 2003;9(7):853-860.

Svicher V, Ceccherini-Silberstein F, Erba F, Santoro M, Gori C, Bellocchi MC, Giannella S, Trotta MP, Monforte AA, Antinori A, Perno CF. Novel human immunodeficiency virus type 1 protease mutations potentially involved in resistance to protease inhibitors. *Antimicrob Agents Chemother* 2005;49(5):2015–2025.

Tagliamonte M, Vidal N, Tornesello ML, Peeters M, Buonaguro FM, and Buonaguro L. Genetic and phylogenetic characterization of structural genes from non-B HIV-1 subtypes in Italy. *AIDS Res. Hum Retrovirus* 2006;22(10):1045–1051.

Tamura K, Dudley J, Nei M & Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 10.1093/molbev/msm092.

Taylor-Castillo L, Herrera-Martinez G, Leon-Bratti MP, Boza R, Leon-Rodriguez B, Luftig RB, Visona K. Study of antiretroviral mutants in HIV patients with treatment failures and the effect of risk factors in the virological failures. *Rev Inst Med Trop Sao Paulo.* 2005 Nov-Dec;47(6):327-31.

Tramuto F, Bonura F, Perna AM, Mancuso S, Firenze A, Romano N, and Vitale F For The Group For HIV-1Antiretroviral Studies In Sicily. Genetic Diversity of HIV-1 Non-B Strains in Sicily: Evidence of Intersubtype Recombinants by Sequence Analysis of gag, pol, and env Genes. *AIDS Res. Hum Retrovirus* 2007;23(9):1131–1138.

Thomson MM, Delgado E, Herrero I, Villahermosa ML, Vazquez-de Parga E, Cuevas MT, Carmona R, Medrano L, Perez-Alvarez L, Cuevas L, Najera R. Diversity of mosaic structures and common ancestry of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Argentina revealed by analysis of near full-length genome sequences. *J Gen Virol* 2002;83:107–119.

Turner BG, Summers MF. Structural Biology of HIV. *J Mol Biol*. 1999;285:1-32.

UNAIDS AIDS epidemic update: December 2007. http://data.unaids.org/pub/EPISlides/2007/2007_epiupdate_en.pdf

Velazquez-Campoy A, Todd M J, Vega S, Freire E. Catalytic efficiency and vitality of HIV-1 protease from African viral subtypes. *PNAS.* 2001;98(11):6062-6067.

Vandamme, A.M., et al.,. Updated European recommendations for the clinical use of HIV drug resistance testing. *Antivir. Ther.* 2004;9: 829–848.

Wawer MJ, Gray RH, Sewankambo NK, Serwadda D, Li X, Laeyendecker O, Kiwanuka N, Kigozi G, Kiddugavu M, Lutalo T, Nalugoda F, Wabwire-Mangen F, Meehan MP, Quinn TC. Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, Rakai, Uganda. *J Infect Dis* 2005;191:1403-1409.

Wigg MD. Vírus da imunodeficiência humana. In: Santos NOS, Romanos MTV, Wigg MD. Introdução à virologia humana. Rio de Janeiro: Guanabara Koogan; 2002.p183-97.

Wildum S, Schindler M, Münch J, and Kirchhoff F. Contribution of Vpu, Env, and Nef to CD4 down-modulation and resistance of human immunodeficiency virus type 1-infected t cells to superinfection. *J Virol.* 2006 August; 80(16): 8047–8059.

World Health Organization, Joint United Nations, Programme on HIV/AIDS. Epidemiological fact sheets on HIV/AIDS and sexually transmitted diseases: Brazil 2006 update. Geneva: World Health Organization, Joint United Nations Programme on HIV/AIDS; 2006.

Wu TD, Schiffer CA, Gonzales MJ, Taylor J, Kantor R, Chou S, Israelski D, Zolopa AR, Fessel WJ, Shafer RW. Mutation patterns and structural correlates in human immunodeficiency virus type 1 protease following different protease inhibitor treatments. *J Virol.* 2003;77(8):4836-4847.

Yeni PG, Hammer SM, Hirsch MS, Saag MS, Schechter M, Carpenter CC, Fischl MA, Gatell JM, Gazzard BG, Jacobsen DM, Katzenstein DA, Montaner JS, Richman DD, Schooley RT, Thompson MA, Vella S, Volberding PA. Treatment for adult HIV infection: 2004 recommendations of the International AIDS Society—USA Panel. JAMA 2004;292:251–265.

CAPÍTULO 10. APÊNDICE

Tabela das frequências absolutas de mutações presentes nos grupos de pacientes não tratados e tratados com IP por substituição específica e por posição.

posição	F1.BR.01.01BR125	NTR (n=677)	TOTAL NTR	TR (n=588)	TOTAL TR
1	P	/	/	/	/
2	Q	/	/	/	/
3	I	/	/	I=3	/
4	T	/	/	S=2	/
		/	/	I/N/P=1	/
5	L	/	/	/	/
6	W	/	/	/	/
7	Q	/	/	R=1	/
8	R	/	/	P/Q=1	/
9	P	/	/	/	/
10	L	V=122	183	V=126	312
		I=60	/	I=176	/
		A=1	/	F=8	/
		/	/	S/Y=1	/
11	V	T=1	1	I=5	6
		/	/	L=1	/
12	T	S=37	164	S=18	99
		I=24	/	I=17	/
		K=22	/	E/K=15	/
		N=17	/	N=13	/
		E=16	/	P=7	/
		P=12	/	A=6	/
		A=11	/	M=4	/
		Q=8	/	D/R/Q/V=1	/
		M=7	/	/	/
		R=4	/	/	/
		D/V=2	/	/	/
		H/L=1	/	/	/
13	I	V=79	82	V=98	101
		A/K/L=1	/	A/L/M=1	/
14	R	K=540	554	K=503	529
		T=5	/	N/Q/T=2	/
		N=4	/	I/M=1	/
		E=3	/	/	/
		I=2	/	/	/
		Q/S/V=1	/	/	/
15	V	I=134	136	I=152	153
		G/L=1	/	L=1	/
16	G	E=98	99	E=58	65
		A=1	/	A=7	/
17	G	/	0	E=18	28

		/	/	D=10	/
18	Q	/	0	E=7	13
		/	/	L=3	/
		/	/	H/M/Y=1	/
19	L	I=56	79	I=42	64
		Q=10	/	V=14	/
		T=6	/	Q=3	/
		V=4	/	S/T=2	/
		K/P/R=1	/	P=1	/
20	K	R=173	193	R=253	356
		M=12	/	T=68	/
		I=6	/	M=17	/
		E/T=1	/	I=14	/
		/	/	V=3	/
		/	/	A=1	/
21	E	/	/	K=2	3
		/	/	Q=1	/
22	A	/	/	V=6	7
		/	/	S=1	/
23	L	/	/	I=3	5
		/	/	V=2	/
24	L	/	0	I=51	52
		/	/	F=1	/
25	D	/	/	/	/
26	T	/	/	/	/
27	G	/	/	/	/
28	A	/	/	/	/
29	D	/	/	/	/
30	D	N/T=1	2	N=37	38
31	T	V=1	/	P=1	/
32	V	L=1	1	I=14	28
		/	/	A=3	/
		/	/	T=1	/
33	L	E/F/V=1	3	F=48	59
		/	/	I=9	/
		/	/	V=2	/
34	E	G=2	/	Q=3	6
		D/K=1	/	K/N/R=1	/
35	D	E=80	83	E=36	59
		N=2	/	N=20	/
		I=1	/	H=2	/
		/	/	G=1	/
36	I	M=38	45	M=10	28
		V=3	/	L/V=9	/
		T=2	/	H=1	/
		L/N=1	/	/	/
37	N	D=51	77	D=79	112
		E=11	/	E=17	/
		S=4	/	S=6	/
		T=3	/	T=5	/

		A/K=2	/	K=2	/
		G/L/R/Q=1	/	A/H/Q/Y=1	/
38	L	/	/	M=1	/
39	P	S=22	52	A=1	26
		Q=18	/	Q=9	/
		T=10	/	S/T=8	/
		A/G=1	/	/	/
40	G	/	/	/	/
41	K	R=73	/	/	/
		E/K/N=1	/	/	/
42	W	R=1	/	/	/
43	K	R=17	22	T=34	53
		T=4	/	R=13	/
		P=1	/	I/Q=2	/
		/	/	E/S=1	/
44	P	K=1	/	/	/
45	K	R=26	27	R=17	20
		M=1	/	T=2	/
		/	/	I=1	/
46	M	I=3	4	I=76	146
		L=1	/	L=69	/
		/	/	V=1	/
47	I	G=1	1	V=16	18
		/	/	A=2	/
48	G	/	0	V=23	35
		/	/	M=8	/
		/	/	A/S/Q/W=1	/
49	G	I=1	/	V=1	/
50	I	G/V=1	2	V=9	17
		/	/	L=8	/
51	G	/	/	/	/
52	G	F=1	/	K=1	/
53	F	I=1	1	L=42	43
		/	/	C=1	/
54	I	K/V=1	2	V=192	211
		/	/	A=8	/
		/	/	L=5	/
		/	/	S/T=4	/
		/	/	M=3	/
55	K	F/K/R/V=1	/	R=51	53
		/	/	T=1	/
56	V	/	/	L=1	/
57	K	R=112	114	R=66	66
		Q=2	/	/	/
58	Q	E=2	3	E=27	27
		Y=1	/	/	/
59	Y	D=1	/	F=2	/
		/	/	E/H=1	/
60	D	E=65	68	E=76	77
		N=3	/	K=1	/

61	N	Q=221	332	Q=88	171
		D=81	/	D=57	/
		S=11	/	E=11	/
		E=10	/	S=8	/
		H=7	/	H=6	/
		I/K=1	/	Y=1	/
62	I	V=77	78	V=127	127
		G/Q=1	/	/	/
63	L	T=139	298	P=169	337
		P=65	/	T=77	/
		S=34	/	C=21	/
		A=24	/	S=19	/
		C=24	/	A=18	/
		V=15	/	V=12	/
		H/Q=5	/	Q=9	/
		G/I=4	/	H/I=3	/
		N=2	/	G/R=2	/
		D=1	/	D/N=1	/
64	I	L=12	24	V=30	37
		M=8	/	L=13	/
		V=3	/	M=6	/
		E=1	/	/	/
65	E	D=167	175	D=133	134
		K=3	/	H=1	/
		N=3	/	/	/
		I/Q=1	/	/	/
66	I	C=1	/	F=3	/
		/	/	L=1	/
67	C	S=6	/	Y=5	/
		D=6	/	D/S=3	/
		E/Y=3	/	F/W=2	/
		G=1	/	G/M=1	/
68	G	E/H=1	/	E=7	/
69	H	Q=44	81	Y=17	35
		Y=20	/	Q=9	/
		K=15	/	R=6	/
		C/R=1	/	K=3	/
70	K	R=13	21	R=20	25
		E=4	/	E=4	/
		A/M/Q/T=1	/	Q=1	/
71	A	T=4	8	V=85	97
		V=3	/	T=20	/
		E=1	/	I=7	/
		/	/	L=2	/
		/	/	G=1	/
72	I	T=147	209	T=136	186
		V=46	/	V=25	/
		R=6	/	R=12	/
		E/M=4	/	E=5	/
		G/K=1	/	K/L/M/Q=2	/

73	G	C/S/T/V=1	4	S=15	23
		/	/	C/T=3	/
		/	/	V=2	/
74	T	S=28	34	S=83	99
		A=3	/	P=9	/
		V=1	/	A=7	/
75	V	I/L=1	/	I=1	/
76	L	V=1	0	V=14	16
		/	/	M=2	/
77	V	I=49	51	I=51	51
		G/L=1	/	/	/
78	G	P=1	/	/	/
79	P	S/T=1	/	/	/
80	T	P=1	/	/	/
81	P	V=1	/	/	/
82	V	I=19	25	A=181	240
		A=3	/	I=18	/
		F/L/N=1	/	F=14	/
		/	/	S=13	/
		/	/	T=10	/
		/	/	L=2	/
		/	/	C/M=1	/
83	N	I=1	1	D=8	9
		/	/	S=1	/
84	I	/	0	V=26	27
		/	/	C=1	/
85	I	F/G/N/V=1	0	V=19	20
86	G	R=1	/	K=1	/
87	R	I/K/N=1	/	/	/
88	N	D/L=1	2	D=38	57
		/	/	S=16	/
		/	/	T=2	/
		/	/	G=1	/
89	M	L=343	349	L=236	302
		I=6	/	I=58	/
		/	/	T/V=4	/
90	L	M=2	4	M=150	150
		T=1	/	/	/
91	T	K=1	/	S=2	/
		/	/	N=1	/
92	Q	K=4	9	K=23	24
		I/H/P/R/T=1	/	H=1	/
93	I	L=161	164	L=159	162
		V=2	/	/	/
		G=1	/	/	/
94	G	C=1	/	/	/
95	C	S=26	28	F/L/N=1	3
		T=1	/	/	/
96	T	L=1	2	S=14	14
		S=1	/	/	/

97	L	N=1	/	I=1	/
		F=3	/	/	/
		I=2	/	/	/
98	N	K=6	/	/	/
		H=2	/	/	/
		I/F=1	/	/	/
		/	/	/	/
99	F	/	/	Y=1	/