Genetic Diversity and Naturally Polymorphisms in HIV Type 1 Integrase Isolates from Maputo, Mozambique: Implications for Integrase Inhibitors

Michelli F. Oliveira,¹ Dulce B. Ramalho,^{1,2} Celina M. Abreu,¹ Adolfo Vubil,² Nédio Mabunda,² Nalia Ismael,² Cidia Francisco,² Ilesh V. Jani,² and Amilcar Tanuri¹

Abstract

HIV proviral DNA integration into the host chromosome is carried out by integrase becoming an important target antiretroviral therapy. Raltegravir was the first integrase inhibitor approved for use in HIV therapy and elvitegravir is in the late phase of clinical development; both show good results in monotherapy studies and may be used worldwide for rescue therapy. In this work we analyzed 57 integrase sequences obtained from samples from drug-naïve and first line regime-failing patients from Maputo, Mozambique, to evaluate the presence of natural polymorphisms and resistance mutations associated with raltegravir and elvitegravir. No major mutations conferring resistance to integrase inhibitors were found and polymorphic accessory mutations were solely observed in low frequency among subtype C sequences—L74M (3.4%), T97A (1.8%), and E157Q (1.8%)—suggesting that this new antiretroviral drug class will be effective in Mozambique providing a good perspective to the introduction of this class of drugs in that country.

THE AIDS PANDEMIC IS ONE of the main challenges to public health with about 33.2 million people infected and 2.1 million deaths globally up to the end of 2008 according to estimative from the Joint United Nations Program on HIV/AIDS (UNAIDS).¹ Mozambique is a limited resources country in sub-Saharan Africa and faces a serious HIV epidemic with a prevalence of 11.5% of adults aged 15–49 years old and 1.4% of children aged 0–14 years old as of 2009.² As used in most of the sub-Saharan Africa region, antiretroviral (ARV) treatment in Mozambique consists of the use of two nucleoside analogue and one nonnucleoside reverse transcriptase inhibitor as first line therapy; the use of a protease inhibitor is prescribed for patients failing first line therapy as suggested by the World Health Organization recommendation.³ ARV treatment became free and publicly available in 2004 in Mozambique; by 2010, 218,991 people were receiving anti-HIV therapy.⁴

Throughout the world, the intensive use of ARV drugs has reduced the mortality and mobility of HIV/AIDS patients by blocking the viral replication that reduces viral loads in patients' plasma and helps to maintain the patient's immune system. However, the emergence of failure during treatment associated with nonadherence, drug toxicity, and the emergence of drug-resistant viral variants represents an important

public health problem requiring drug substitutions and alternative regimens. So the development of new drugs acting on alternative steps of the virus replication cycle has become necessary in order to increase the efficiency of rescue regimens. In this context, integrase (IN), the enzyme responsible for the integration of viral cDNA into the host genome, has been a new important target for the development of new ARV drugs and therapy.

Integrase is a 288-amino acid enzyme, codified by the 3' portion of the HIV *pol* gene and composed of three domains: an N-terminal domain (NTD), a catalytic core domain (CCD), and a C-terminal domain (CTD). The NTD (1–50 residues) possesses the H(12)H(16)C(40)C(43) motif, which is a zinc ligand and has evolved with multimerization of the enzyme. The CCD (51–211 residues) contains the active site of the enzyme with the conserved residues D(64)D(116)E(152), which bind divalent metals (Mn^{2+} and Mg^{2+}) that act as cofactors. The CTD (212–288 residues) interacts with the reverse transcriptase and binds to DNA in an unspecific way.⁵

Integration of the HIV-1 genome into host chromosomes is a multistep process. First, the enzyme recognizes specific sequences of long terminal repeats (LTRs) and forms a stable complex with the proviral DNA within the preintegration complex (PIC). The 3' processing step consists of the removal

¹Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

²Instituto Nacional de Saúde, Maputo, Mozambique.

of dinucleotide GT from the each 3' end of viral DNA generating reactive 3' hydroxyl DNA ends. Subsequently, the pre-integration complex migrates toward the nucleus where integrase catalyzes the strand transfer step by the ligation of the 3'-OH DNA ends to the 5' phosphate of the host DNA. Finally, repair host enzymes remove gaps generated during the process.⁵

Raltegravir (RAL; Isentress, Merck & Co., Inc., Whitehouse Station, NJ) was the first IN inhibitor approved by the U.S. Food and Drug Administration in October 2007 and elvitegravir (ELV; Gilead Sciences, Foster City, CA) is in phase III clinical trials. Both compounds block the viral replication by inhibiting the DNA strand-transfer reaction with 50% inhibitory concentrations in the nanomolar range and possess good safety profiles, being rarely associated with severe adverse cases.⁵ Both have shown powerful results in short-term monotherapy studies and may be an alternative for HIV-infected patients, mainly those who have selected resistant viral variants to drugs targeting the reverse transcriptase and protease. However, *in vitro* and *in vivo* studies have described the occurrence of resistance mutations selected by the RAL and ELV that appear mainly near the CTD of integrase.⁵ In the Stanford Database algorithm (<http://hivdb.stanford.edu>), there are at least 28 positions listed for resistance mutations in the IN gene and these are described as being associated with a decreased susceptibility to these inhibitors. The main resistance pathways involve positions 143, 148, and 155 of the enzyme, which are associated with specific secondary and accessory mutations, some of which appear as polymorphic mutations.⁵

Here we evaluated the genetic diversity and natural polymorphisms in HIV-1 IN sequences from naive and first line regimen-failing patients from Maputo, Mozambique, and, for the first time, we examined the prevalence of raltegravir and elvitegravir resistance-associated mutations in the isolates circulating in the Maputo population to predict the effectiveness of these new drugs in that country as an alternative ARV treatment.

This study included a total of 57 plasma samples collected from patients attending the Alto-Maé Health Center, Maputo, Mozambique from December 2009 to August 2010. Thirty-nine of these were drug-naïve patients and 18 were first line ARV treatment-experienced patients. This study was approved by the ethics commit (number 307/CNBS). Genomic DNA was extracted from the buffy coat from patients using a DNA Extraction Column kit (QIAamp DNA Mini Blood Kit; QIAGEN, Valencia, CA) in the automated QIACube according to the manufacturer's instructions.

The amplification of the fragment content of the whole IN gene (1147 base pairs, 4068 to 5214 position from pNL43) was performed by a nested polymerase chain reaction (PCR) protocol using primers targeting conserved regions that flank the IN gene. PCR products from 57 samples were purified using Microcon Centrifugal Filter Devices (Millipore) and subsequently submitted to sequencing using the ABI Big Dye terminator system (v3.1) in an ABI 3130 Genetic Analyzer automated sequencer (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions. Nucleotide sequences were manually edited using the SeqMan program (DNASTAR suite).

For subtyping, sequences were submitted to analysis on the Rega (<http://www.bioafrica.net/rega-genotype/html/>

subtypinghiv.html) algorithm. All nucleotide sequences were translated to amino acids sequences and aligned using the Clustal W method (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>).

Consensus sequences were generated for Mozambique C, and alternatively for naive and ARV-failing Mozambique C sequences using the BioEdit Sequence Alignment Editor (version 5.0.6, Carlsbad, CA), including the most frequent amino acid per position. Additionally, 72 subtype C sequences from South Africa were obtained from GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>; access numbers GQ872465 to GQ872538) and were used to create a South Africa C consensus. Positions where two amino acids occurred at frequencies higher than 40% show both amino acids at the consensus (the first letter seen at the position of the consensus represents the amino acid that is most prevalent). Finally, all consensus sequences were aligned with the B and C consensus obtained from the Los Alamos Database (LASDB; www.hiv.lanl.gov) using BioEdit and then compared for molecular characterization. All Mozambique sequences were extensively analyzed for the presence of molecular signatures, polymorphic mutations, and resistance mutations.

The complete sequence of the IN gene was obtained from 57 patients' buffy coat and the nucleotide sequences of IN genes varied from 867 to 876 base pairs in correct open reading frames. The subtyping analysis using the Rega algorithm revealed that 92.9% sequences (53 sequences) clustered to subtype C and 1.7% were subtype A1 (1 sequence; JQ670837). This is in accordance with previous work showing that subtype C is the most frequent subtype in this African region.⁶ The remaining sequences (JQ670795, JQ670835, and JQ670824) were determined to be recombinants and were further analyzed in the RIP 3.0 (Recombinant Identification Program) program from the Los Alamos HIV Database (LAHDB; web site <http://www.hiv.lanl.gov>), which indicated that all were URF composed by C and G subtype sequences (data not shown).

An extensive analyses of all 53 translated subtype C sequences was performed and revealed enzyme length variations from 288 to 292 amino acids, due to an insertion of four amino acids—Q(N/G)ME—in the enzyme CTD found in seven isolates. This insertion has been reported previously in South Africa sequences.^{7–9} CTD has been described as the nonspecific DNA binding domain⁵ and this interaction is required for IN 3' processing and strand-transfer activities. Moreover, CTD binds to RT and it seems to be necessary for reverse transcription.¹⁰ Thus, the importance of this unusual insertion to viral functions needs to be investigated further.

Figure 1 summarizes the distribution of amino acids among sequences of complete subtype C IN where the population mixture was considered and represented for each position. We have found no amino acid substitutions in the conserved motifs HHCC and DDE as well as for the majority of positions interacting with the LEDGF/p75 factor (W131, W132, Q164, Q168, E170, L172, and K173) confirming the importance of these codons for enzyme activity. Analyzing subtype C sequences, we found only three conservative amino acid substitutions within the LEDGF/p75 interaction positions: V165I representing 3.7% of sequences (2 sequences), R166K representing 1.8 % of sequences (1 sequence), and D167E representing 21.4 % of sequences (12 sequences).

FIG. 1. Distribution of Mozambique subtype C variants. The subtype C consensus from the Los Alamos Database (LASDB) (<http://hivdb.stanford.edu>) is shown. Numbers below the sequence represent the amount of times the amino acids were found among sequences.

Among non-C subtype sequences, the D167E substitution was found in the subtype A1 sequence (JQ670837) and one CG recombinant (JQ670824).

To identify the natural polymorphisms (frequency higher than 40%) and molecular signatures of subtype C from Mozambique, we performed an alignment of IN consensus sequences (288 pb) and compared this with the subtype B and C consensus obtained from LASDB and the South Africa subtype C consensus obtained from GenBank. Considering the cut-off of 40% of variation, four positions were selected with two amino acids in Mozambique subtype C: 72 (V = 58.1% and I = 42.8%), 100 (Y = 55.5% and F = 44.4%), 265 (A = 53.4% and V = 46.5%), and 269 (R = 59.6% and K = 40.3%). Additionally, positions 50 (M = 50.0% and I = 38.3%) and 218 (I = 45.6% and T = 38.5%) showed amino acids with variation frequencies higher than 30% (Fig. 1). All 16 polymorphic codons (25, 31, 50, 100, 101, 112, 124, 125, 136, 201, 218, 234, 265, 269, 278, 283) previously showed by Rhee and collaborators¹¹ were observed in Mozambique C sequences when compared with the subtype B consensus confirming these sites as subtype C signatures. However, an extra polymorphic site was located at codon 72, where the frequency of amino acid valine was 58.1% suggesting that this could be considered a signature of subtype C IN from Mozambique (Fig. 2).

Additionally, the Mozambique C consensus of ARV naive sequences failing first line therapy was generated in comparison and no differences were observed, except at positions 100 and 269 of the naive Mozambique C consensus, which

presented only one amino acid with a frequency of more than 40% at these positions (Y and R, respectively).

A high homology was observed when Mozambique and LASDB subtype C consensus sequences were compared, showing only a few differences occurring only at positions 50 and 72, 100, 218, 265, and 269 due to the presence of more than one prevalent amino acid in the Mozambique C consensus. High homology was also observed when we compared Mozambique subtype C consensus and South Africa subtype C consensus suggesting that IN sequences from both countries are closely related.

The amino acid sequence alignment was also extensively analyzed for the identification of primary mutations, non-polymorphic and polymorphic mutations associated with resistance to RAL and ELV in all sequences obtained from Mozambique. The sequences were analyzed for the presence of listed mutations in the Stanford Database (access in 03/10/2011) and described as conferring resistance *in vitro* and *in vivo* by RAL (T66A/K, L68V/I, L74M, E92Q, Q95K, T97A, F121Y, E138K/A, G140S/A/C, Y143C/R/H, Q148H/K/R, N155H/S, E157Q, G163R/K) and ELV (H51Y, T66A/I/K, L68V/I, E92Q, Q95K, T97A, H114Y, F121Y, E138K/A, G140S/A/C, P145S, Q146P, S147G, Q148H/K/R, S153Y, N155H/S, E157Q, R263K). No primary resistance mutations were associated with RAL and ELV except for three of the previously described codons associated with resistance to RAL and ELV (74, 97, and 157) where a very low frequency was found and these were distributed among four of the subtype C isolates

FIG. 2. Alignment of HIV-1 integrase (IN) consensus sequences from LASDB (<http://hivdb.stanford.edu>), Mozambique, and South Africa. Mozambique consensus sequences from naive and antiretroviral (ARV)-experienced patients were also represented for comparison. Positions with more than one amino acid with a frequency higher than 40% were both represented. The arrows indicate HHCC and DDE motifs. Positions associated with resistance to integrase inhibitors are underlined.

studied. The L74M substitution was detected in two specimen sequences (3.4%), one naive patient and the other in an ARV-failing patient. Interestingly, we also found two polymorphisms (I and L) not related to IN inhibitor resistance. Of note, L74M is a naturally polymorphic accessory mutation that cannot confer resistance alone, but only in combination with N155H it reduces RAL susceptibility.^{12,13} One sequence from a naive patient (1.8% of sequences) presented T97TA, which is also an accessory polymorphic mutation often coselected with the Y143 mutation complex reducing susceptibility to RAL.¹⁴ Of note, the E157Q substitution was detected in 1.8% of isolates studied, in an ARV-failing patient. In fact, E157Q is a minimally polymorphic mutation selected *in vitro* by ELV and *in vivo* by RAL but has little or no effect on IN inhibitors.

Our results showed the absence of primary resistance mutations in IN of Mozambique isolates suggesting this new antiretroviral drug class will be effective in this country.

Acknowledgments

We are thankful to the laboratory and hospital staff of the Centro de Saúde do Alto Maé. Special thanks to Carla Wale, Lídia Boullosa, Salomão Mujui, and to the staff of the Departamento de Imunologia of the INS for their collaboration in this project. The study was supported by the Brazilian National Research Council, CNPq (Grant 471299/2009-0) and CAPES, National AIDS Program/Ministerio da Saude do Brasil, and UFRJ. The sequences generated in this study were deposited in GenBank under accession numbers JQ670788–JQ670817, JQ670819–JQ670845.

Author Disclosure Statement

No competing financial interests exist.

References

1. AIDS Epidemic Update, 2009: <http://www.unaids.org/en/dataanalysis/epidemiology/2009aidsepidemicupdate/>.
2. Instituto Nacional de Saúde (INS), Instituto Nacional de estatística (INE) e ICF Macro. 2010. Inquérito nacional de prevalência, riscos comportamentais e informação sobre o HIV e SIDA em Moçambique 2009. Calverton, Maryland, EUA: INS, INE e ICF Macro. ([WWW.measuredhs.com](http://www.measuredhs.com)).
3. World Health Organization: http://whqlibdoc.who.int/publications/2010/9789241599764_eng.pdf.
4. Dados de TARV Nacionais, Ministério de Saúde, Direcção de assistência médica, Mozambique. http://www.misau.gov.mz/pt/hiv_sida/dados_tarv_nacionais_de_nov_2010_a_abril_2011
5. Métifiot M, Marchand C, Maddali K, and Pommier Y: Resistance to Integrase Inhibitors. *Viruses* 2010;2:1347–1366.
6. Hemelaar J, Gouws E, Ghys PD, and Osmanov S: Global trends in molecular epidemiology of HIV-1 during 2000–2007. *AIDS* 2011;25(5):679–689.
7. Hunt GM, Papathanasopoulos MA, Gray GE, and Tie-messen CT: Characterisation of near-full length genome sequences of three South African human immunodeficiency virus type 1 subtype C isolates. *Virus Genes* 2003;26:49–56.
8. Tzitzivacos DB, Tiemessen CT, Stevens WS, and Papathanasopoulos MA: Viral genetic determinants of nonpro-

- gressive HIV-1 subtype C infection in antiretroviral drug naive children. *AIDS Res Hum Retroviruses* 2009;25:1141–1148.
9. Fish MQ, Hewer R, Wallis CL, Venter WD, Stevens WS, and Papathanasopoulos MA. Natural polymorphisms of integrase among HIV type 1-infected South African patients. *AIDS Res Hum Retroviruses* 2010;26(4):489–493.
  10. Zhu K, Dobard C, and Chow SA: Requirement for integrase during reverse transcription of human immunodeficiency virus type 1 and the effect of cysteine mutations of integrase on its interactions with reverse transcriptase. *J Virol* 2004;78: 5045–5055.
  11. Rhee SY, Liu TF, Kiuchi M, Zioni R, Gifford RJ, Holmes SP, and Shafer RW: Natural variation of HIV-1 group M integrase: Implications for a new class of anti-retroviral inhibitors. *Retrovirology* 2008;5:74.
  12. Reuman EC, Bachmann MH, Varghese V, Fessel WJ, and Shafer RW: Panel of prototypical raltegravir-resistant infectious molecular clones in a novel integrase-deleted cloning vector. *Antimicrob Agents Chemother* 2010;54(2):934–936.
  13. Kobayashi M, Nakahara K, Seki T, Miki S, Kawauchi S, Suyama A, *et al.*: Selection of diverse and clinically relevant integrase inhibitor-resistant human immunodeficiency virus type 1 mutants. *Antiviral Res* 2008;80(2):213–222.
  14. Fransen S, Gupta S, Danovich R, Hazuda D, Miller M, Witmer M, Petropoulos CJ, and Huang W: Loss of raltegravir susceptibility by human immunodeficiency virus type 1 is conferred via multiple nonoverlapping genetic pathways. *J Virol* 2009;83(22):11440–11446.

Address correspondence to:

Amílcar Tanuri
 Laboratório de Virologia Molecular
 Instituto de Biologia
 Departamento de Genética
 Universidade Federal do Rio de Janeiro (UFRJ)
 Av. Brigadeiro Trompowsk s/nº, CCS, Bloco A,
 Sala 121, Cidade Universitária
 CEP: 21941-970 Rio de Janeiro, RJ
 Brazil
 E-mail: atanuri@biologia.ufrj.br