Recombination events and epitope prediction in HIV-1 strains from Southwest Cameroon

By

Bixa O. Ogola

Student Number: 11530098

Dissertation submitted in fulfilment of the requirements for the award of Master of Science degree in Microbiology

To the Department of Microbiology,

School of Mathematical and Natural Sciences

University of Venda, South Africa

Supervisor: Prof. Pascal O. Bessong (University of Venda)

Co-supervisor: Dr. Denis M. Tebit (University of Virginia)

April 2016

Abstract

BACKGROUND: HIV has a heterodiploid genome and reverse transcriptase (RT) that has recombinogenic propeties which promotes high diversity. This diversity is illustrated by the types, groups, subtypes and growing number of circulating recombinant viruses. Viral diversity may impact on viral fitness, diagnosis, disease progression, therapy management, as well as vaccine design. Cameroon is known to be a "hotspot" for high HIV diversity with a predominance of the CRF_02 AG strain. This study was conducted on viruses from Southwest Region of Cameroon with the aim of describing novel recombinant viruses. Doing this could reveal information on the complexity of the identified unclassified strains as well as predict what impact they may have on the course of managing them.

METHOD: The study setting was the Mutengene Baptist hospital in the Southwest Region of Cameroon. Study subjects included individuals from different locales in the Southwest and Littoral Regions of Cameroon. Approval was obtained from the Ethics Committee of the Cameroon Baptist Health Board (Cameroon) and the Research Ethics Committee of the University of Venda, South Africa. Whole blood was collected from 107 HIV-1 infected individuals of whom 83 were drug naïve and 24 were drug experienced. Continuous partial p17 and partial p24 subgenomic regions was PCR amplified from proviral DNA, sequenced using the Sanger protocol, genotyped using MEGA and analyzed for recombination patterns using RIP and jpHMM. CTL epitope prediction was done using NETCTLpan and MHC-1 tools using HLA alleles with the highest frequencies among the coastal people of Cameroon. Virus BM189 was selected, due to its unique recombination pattern, for near-full length genome amplification, followed by next generation sequencing, and genotyped using MEGA and analysed for recombination using, Simplot, RIP and jpHMM. Drug resistance profiles of viral populations at >20%, <5% and <1% were determined using different interpretation algorithms. CTL epitopes within *gag* and *env* genes were predicted.

RESULTS: Fifty two of 107 specimens were successfully amplified and 39 were successful sequenced. The majority of the 39 sequences (74.4%) were subtype A-related strains (n=29/39), 12.8% were CRF01_AE (n=5/39), 7.7% were subtype G-related (n=3/39) and 5.1%

were sub-subtype F2-related (n=2/39). From the 39 *gag* sequences, it was observed that most of subtypes A related, G related and CRF01_AE related sequences were possible recombinants based on the high amino acid variation to what was expected from their pure subtypes/CRF. Twenty of the 39 samples had partial *pol* and *env* sequences available from previous analysis and analysis of the subgenomes revealed that the dominant strain (50.0%) was CRF02_AG (n=10/20), 35.0% were URFs (n=7/20), 10.0% were CRF22_01A1 (n=2/20) and 5.0% were subsubtype F2 (n=1/20). One of the URFs, Virus BM189, which showed subtype F2 for partial *gag* and partial *pol* and subtype A1 for partial *env* analysis was selected and succesfully amplified for near-full length genome analysis. The virus was obtained from a 28 year old unmarried female with residence in Mutengene, with no history of anti-retroviral treatment, and classified as AIDS stage 3. The near full length sequence (9077 nucleotides) was determined to be a second generation recombinant of subtypes F2/CRF01_AE/A1/F1. It was made up of mostly sub-subtype F2, CRF01_AE within 407 nucleotides of gp120 of *env*, sub-subtype A1 for rest of *env* and F1 for most of *nef*.

Drug resistance profile for BM189 showed that viral populations >20% within the quasispecies were susceptible to all known anti-retroviral drugs. Nevertheless, both <5% and <1% viral populations had M184V mutation which is associated with high level resistance to Emtricitabine (FTC) and Lamivudine (3TC), and K103N associated with high level resistance to Efavirenz (EFV) and Nevirapine (NVP). The majority population (>20%) were predicted to be a R5 tropic virus with susceptibility to entry inhibitors, whereas the minority populations had G36S mutation which is associated with resistance to the entry inhibitor Enfuvirtide (INN).

Partial *gag* epitope prediction showed numerous epitopes that were good binders for the HLAs used. P24 epitope TSTLQEQIGW was predicted with a 20% frequency among CRF01_AE related sequences and was restricted by HLA-B*58:01, which has been associated with slow disease progression. P17 epitopes WPFRNSRRM and YLIQQPSIVY, restricted by HLA-B*35:01, was predicted within CRF01_AE related sequences. This allele has been associated with rapid disease progression. BM189 *gag* and *env* epitope prediction showed epitopes that were strong binders for HLA-B*58:01. *Env* epitopes (HSFTCGGEFFY and FAILKCNDAEF) were predicted to be strong binders of HLA-B*35:01 and gag epitope (MYSPVSIL) was predicted to be strong binders for HLA-C*04:01. HLA-B*35:01 and HLA-C*04:01 have been

associated with rapid disease progression. HLA typing was not done on the study participants, hence the presence of these epitopes cannot adequeately tell the nature of disease progression. HLA typing on the participants is recommended to determine if these associations held true for the study participants. Furthermore, other phenotypic studies such as interferon-gamma enzyme-linked immunospot (IFN-γ ELISPOT) need to be done to determine if these predictions will hold true.

In conclusion, the findings from this study supports the on-going genetic diversification of HIV in Cameroon; and that recombinants viruses are dominating the epidemic. Furthermore, it highlights the importance of next-generation sequencing as a genotypic tool for drug resistance testing for minority populations. Numerous epitopes were predicted, which would serve in future phenotypic studies.

Keywords: HIV, recombination events, cytotoxic T-cell epitopes, Southwest Cameroon.