HIV-1 subtype C LTR Sp1IIIT5A mutant enhances transcription activity and Sp1 binding affinity

Submitted by: Nomcebo Emelda Mtshali

218002912

Supervised by: Dr Paradise Madlala

Submitted in fulfillment of the requirements for the degree of

Master of Medical Science in Virology School of Laboratory Medicine and Medical Sciences (Virology) in the School of Laboratory Medicine and Medical Sciences, HIV Pathogenesis Programme, Doris Duke Medical Research Institute, Nelson R. Mandela School of Medicine, College of Health Sciences, University of KwaZulu-Natal, Durban

2024

PREFACE

The experimental procedures presented in this thesis were performed at the Hasso Plattner Research Laboratory, Doris Duke Medical Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa. This study was performed under supervision of Dr. Paradise Madlala. This dissertation represents the original work of the author and has not been submitted for any degree or examination at any other university. Where the work of others has been used, the authors have been duly acknowledged.

[redacted]

Nomcebo Emelda Mtshali (Student)

[redacted]

Dr. Paradise Madlala (Supervisor)

PLAGIARISM DECLARATION

I, Nomcebo Emelda Mtshali, declare that

1. The research reported in this thesis, except where otherwise indicated, is my original research.

2. This thesis has not been submitted for any degree or examination at any other university.

3. This thesis does not contain other persons' data, pictures, graphs, or other information, unless specifically acknowledged as being sourced from other persons.

4. This thesis does not contain other persons' writing unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a. Their words have been re-written, but the general information attributed to them has been referenced

b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks and referenced.

5. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Initials and Surname: NE Mtshali

Signature----------

Date: 27 November 2024

DEDICATION

I dedicate this thesis to God, for granting me strength and wisdom and to my parents, whose unwavering support, encouragement, and belief in my abilities have been a constant source of inspiration throughout my academic journey.

ACKNOWLEDGMENTS

Sincere thanks to my supervisor, Dr. Paradise Madlala (HIV Pathogenesis Programme, Doris Duke Medical Research Institute, Nelson R Mandela School of Medicine, University of KwaZulu-Natal). I further wish to express my gratitude to Prof. Thumbi Ndung'u who is the Scientific Director of the HIV Pathogenesis Programme, Doris Duke Medical Research Institute, Nelson R Mandela School of Medicine, University of KwaZulu-Natal. I thank the collaborators of my Masters project Prof. Mahmoud Soliman (Pharmaceutical Science, University of KwaZulu-Natal), Dr. Rene Khan (School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of KwaZulu-Natal, Durban), and Mr Ibrahim Oluwatobi Kehinde (Pharmaceutical Science, University of KwaZulu-Natal) for motivation, guidance and support with this project. I am grateful that I was afforded the opportunity to learn and work in the HIV Pathogenesis Programme laboratory.

Many thanks to my parents, sisters, brother, and friends for their prayers, support and encouragement.

The staff and students at HIV Pathogenesis Programme (HPP) for their assistance and guidance.

ETHICAL APPROVAL

The ethics approval for this study was obtained from the Biomedical Research Committee of the Nelson R. Mandela School of Medicine, University of KwaZulu Natal, Durban, South Africa. The ethics approval reference number is as follows: BREC/00006386/2023.

PRESENTATIONS

Nomcebo Mtshali, Ibrahim Kehinde, Rene Khan, Mahmoud Soliman, and Paradise Madlala: Oral presentation on: HIV-1 subtype C LTR Sp1IIIT5A mutant enhances transcription activity and Sp1 binding affinity in the University of KwaZulu-Natal Westville Campus, Annual Research Day, Durban (South Africa) (2nd Prize).

Nomcebo Mtshali, Ibrahim Kehinde, Rene Khan, Mahmoud Soliman, and Paradise Madlala: Oral presentation on: HIV-1 subtype C LTR Sp1IIIT5A mutant enhances transcription activity and Sp1 binding affinity in the University of KwaZulu-Natal Medical School Campus, College of Health Science symposium, Durban (South Africa).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AIDS: Acquired immune deficiency syndrome

ART: Antiretroviral therapy

cDNA: complementary Deoxyribonucleic Acid

CRF: Circulating recombinant form

CTD: C-terminal domain

BBB: Blood-brain barrier

Bp: Base Pairs

CCR5: C-C chemokine receptor type 5

CNS: Central Nervous System

Copies/ml: Copies Per Millilitre

CSF: Cerebrospinal Fluid

CXCR4: C-X-C chemokine receptor type 4

DMEM: Dulbecco Modified Eagle Medium

DNA: Deoxyribonucleic acid

dNTP: Deoxyribonucleotide triphosphate

E-box: Enhancer box

EDTA: Ethylenediaminetetraacetic acid

Env: Envelope Glycoprotein

Gag: Group Specific Antigen

GM: Growth Media

HEPES 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HIV-1: Human immunodeficiency virus type 1

HIV-2: Human immunodeficiency virus type 2

HPP: HIV Pathogenesis Programme

IN: Integrase

LB: Lysogeny Broth

LTR: Long Terminal Repeat

Luc: Luciferase

NF-κB: Nuclear Factor Kappa-Light-Chain-Enhancer of Activated B Cells

NaOAc: Sodium Acetate

NFAT: Nuclear factor of Transcription

PBS: Primer binding site

PCR: Polymerase chain reaction

PLHIV: People living with HIV

P-TEFb: Positive transcription elongation factor-b

PCP: Pneumocystis Pneumonia

Pol: Polymerase

Pol II: RNA Polymerase II

PR: Protease

R: Repeat region

RLU: Relative light unit

RNA: Ribonucleic acid

RNase H: Ribonuclease H

RNA pol II: RNA Polymerase II

Rev: Regulator of Virion Protein

RRE: Rev response element

RT: Reverse Transcriptase

RTC: Reverse transcription complex

Sp1: Specificity protein 1

Sp1III5A: Adenine mutation at position 5 of Sp1 III site

Sp1III5T: Thymine mutation at position 5 of Sp1 III site

SVG: SV40-transformed human glial cells (Astrocytes)

Tat: Trans-activator of Transcription

TBP: Tata Binding Protein

TFs: Transcription factors

TFBS: Transcription factor binding site

U3 region: Unique 3 region

U5: Unique 5 region

U3R- Unique 3 Repeat region

URFs: Unique recombinant forms

ABSTRACT

Background: Human immunodeficiency virus type 1 subtype C (HIV-1C) long terminal repeat (LTR) variants play a crucial role in influencing transcriptional activation and disease outcomes. However, it is unclear how much effect Sp1IIIT5A has on HIV-1C LTR transcription activity. To address this gap, the current study investigates the impact of a specific single mutation, T5A, which has been reported to affect the HIV-1 subtype B (HIV-1B), within the Specific Protein 1 (Sp1) III motif on HIV-1C LTR transcription activity and Sp1 binding affinity.

Methods: HIV-1C LTR consensus and LTR mutant exhibiting a single T5A mutation into the Sp1III motif were (LTR Sp1IIIT5A) independently cloned into the pGL3 Luciferase basic reporter vector and sequenced to confirm the presence of the mutation. Consensus LTR-pGL3 and LTR Sp1IIIT5A-pGL3 recombinants were transfected into SVG and Jurkat cell lines to determine their transcription activity. The Sp1 transcription factor expression levels in these cell lines, were quantified by western blotting assay. Additionally, we modeled the crystal structures of the HIV-1C LTR and Sp1 protein, and performed docking calculations using various web servers, including HDOCK, HADDOCK, and pyDockDNA. Lastly, molecular dynamics simulations were carried out to evaluate the stability and binding affinity of Sp1 to LTR consensus and LTR Sp1IIIT5A.

Results and Discussion: Our data show that LTR exhibiting Sp1IIIT5A mutation was successfully cloned into pGL3 basic reporter vector. The Sp1III5A mutation significantly enhances basal transcription activity compared to the canonical Sp1III motif in SVG cells (p<0.00001) and Jurkat cells (p=0.00521). This suggest that LTR Sp1III5A mutation is associated with enhanced transcription activity. Similarly, the Sp1IIIT5A mutation exhibited significantly increased Tat-mediated HIV-1C LTR transcription activity in SVG cells (p<0.00001) and Jurkat cells (p<0.00001). This further reinforces the notion that Sp1III5A enhances HIV-1C LTR transcription activity. Notably, transcription activity exhibited by LTR Sp1IIIT5A was more pronounced in SVG then in Jurkat cell line. Furthermore, the expression levels of the Sp1 transcription factor were comparable (p=0.08140) between SVG and Jurkat cell lines, suggesting that differences in transcriptional activity observed are not attributable to variations in Sp1 expression levels. Moreover, the binding affinity analysis revealed that the Sp1IIIT5A mutation exhibits stronger interactions with the Sp1 transcription factor than its canonical counterpart, with interaction values of -332.7, -174.6, and -279.2 kcal/mol compared to -311.4, -157.0, and -247.3 kcal/mol for the canonical sequence. These suggest that the

Sp1IIIT5A mutation not only potentially enhances transcriptional efficiency but also stabilizes the binding of Sp1 transcription factor.

Conclusion: The HIV-1 LTR Sp1IIIT5A mutation enhances transcription activity and binding affinity of Sp1 compared to the Sp1III canonical sequence in both SVG and Jurkat cell lines. Future studies should investigate the effect of Sp1IIIT5A mutation on the HIV-1 subtype C latency profile.

CHAPTER 1: INTRODUCTION

1.1. Background

The rates of clinical disease progression measured by plasma viral load in people living with HIV (PLWH) who are antiretroviral therapy naïve vary significantly (Shoko and Chikobvu, 2019). The underlying mechanisms for this heterogeneity remain poorly understood. The regulatory elements of viruses could play an important role in conferring differences in replication fitness. The 5' long terminal repeat (LTR) is the viral promoter that drives gene transcription and is essential for the viral life cycle (Jeeninga *et al.*, 2000, Opijnen *et al.*, 2004). The HIV-1 5' LTR is composed of three regions: Unique 3 (U3), Repeat (R) and Unique 5 (U5) regions, with the U3 region further subdivided into modulatory, core-enhancer and core-promoter domains (reviewed in (Groen and Morris, 2013). Each of these domains contains distinct transcription factor binding sites (TFBSs); notably, the core-promoter domain features a TATA box, two E-boxes, and three Specific Protein 1 (Sp1) binding motifs (reviewed in (Groen and Morris, 2013).

Among these transcription factors, only a select few, including nuclear factor-kappa B binding site (NF-κB) and Sp1, have been identified as critical for initiating HIV-1 transcription (Qu *et al.*, 2016). Sp1 is a transcription factor family, which participates in the regulation of cellular gene expression and is quired to initiate HIV-1 transcription by binding to the Sp1 motifs located within the 5' LTR (Gaynor, 1992b). Subsequently, Sp1 interacts with TATA-binding protein (TBP) and some TBP-associated factors such as TAF110, which are components of the general RNA Polymerase II (RNAPII) transcriptional factor D (Guermah *et al.*, 1998, Hoey *et al.*, 1993). Therefore, Sp1 is an essential transcription factor that initiates basal transcription by recruiting the RNAPII-dependent transcriptional machinery to the HIV-1 promoter 5' LTR.

Basal transcription is characterized by the production of short abortive viral mRNA transcripts during elongation in the absence of the trans-activator of transcription (Tat) (Roebuck and Saifuddin, 2018, Gatignol *et al.*, 2018). Tat is a potent activator that enhances transcription by binding to the TAR hairpin at the 5′ end of newly synthesized RNA transcripts (Das *et al.*, 2011, Gatignol *et al.*, 2018). In the absence of Tat, the short transcripts predominate, whereas in the presence of Tat, there is a dramatic increase in the levels of long transcripts (Wu and Marsh, 2003, Das *et al.*, 2011). Thus, Tat participates in a positive feedback mechanism that ensures

elevated levels of HIV-1 transcription. Taken together, these studies suggest that HIV-1 LTR transcription activity and viral replication are partly dependent on interactions of the Sp1 motifs with the family of transcription factors.

The genetic variation within the HIV genome has been extensively studied (Nonnemacher *et al.*, 2004, Munkanta *et al.*, 2005, Li *et al.*, 2012, Gray *et al.*, 2013, Madlala *et al.*, 2023). Genetic variation within the HIV-1 viral genome is a naturally occurring process driven by the low fidelity of reverse transcriptase, coupled with the selective pressures brought about within the host such as antiretroviral therapy, immunological pressures, viral recombination events, host-cell phenotype, and rates of virus production (Coffin, 1995, Li *et al.*, 2011). These events result in single nucleotide polymorphisms (SNPs) throughout the genome including the promoter region, LTR. The genetic variation occurring within LTR binding sites where host transcription factors and viral regulatory proteins bind, significantly influences how the LTR regulates viral transcription. This suggests that variations in promoter regions may lead to differential transcriptional regulation or subtle changes in transcription rates. Notably, slight alterations in the LTR can impact disease progression (Nonnemacher *et al.*, 2004, Schiralli Lester and Henderson, 2012, Miller-Jensen *et al.*, 2013). An analysis in HIV-infected patients revealed that the C-to-T change at the Sp1 site III is positively correlated with disease progression (Nonnemacher *et al.*, 2004). Furthermore, this C-to-T SNP has been demonstrated to affect transcriptional activity (Shah *et al.*, 2014a), although these findings were primarily observed within the context of HIV-1 subtype B (HIV-1B) consensus.

The LTRs of HIV-1 subtype B contain two NF-κB binding sites, while the LTRs from subtype C, which is the predominant clade in Africa, possess an additional NF-κB-like binding site and demonstrate greater activity in activating viral gene transcription (Montano *et al.*, 1997, Monty *et al.*, 2000, Roof *et al.*, 2002, Hemelaar *et al.*, 2006). The emergence of an additional NF-κB binding site in the LTRs of subtype C has been observed in patients experiencing therapeutic failure in Brazil and Mozambique. This phenomenon may partly elucidate the enhanced fitness of subtype C and its replacement of subtype B in these regions (Boullosa *et al.*, 2014). A systematic genetic screen for HIV LTR promoter elements revealed that single point mutations at the fourth guanine nucleotide of the Sp1 site III and the second adenine of the TATA box increased the likelihood of transitioning from latency to productive HIV-1 infection (Miller-Jensen *et al.*, 2013). Recent findings have shown that genetic variation within the HIV-1 subtype C transmitted/founder (T/F) LTR influences transcriptional activation and clinical

outcomes (Madlala *et al.*, 2023). Analysis indicated that HIV-1C T/F LTR variants exhibited variability within the Sp1III binding site, with an adenine replacing thymine at position 5 of the Sp1III motif (Sp1IIIT5A) being more prevalent (Madlala *et al.*, 2023). This mutation was shown to be prevalent in approximately 61% of transmitted/founder LTRs analyzed in HIV-1C (Madlala *et al*., 2023), which is responsible for a 50% significant proportion of global infections, particularly in southern Africa. However, it is unclear how much effect Sp1IIIT5A has on HIV-1C LTR transcription activity. Previous research has demonstrated that Tripartite Motif (TRIM22) proteins can suppress HIV-1 LTR-driven transcription by inhibiting Sp1 binding to the promoter (Turrini *et al.*, 2015). Despite this, limited studies have investigated the binding affinity of the Sp1 transcription factor to the promoter. Consequently, this study aims to examine the effects of the HIV-1C Sp1 IIIT5A mutation on LTR transcription activity, Sp1 transcription factor expression levels, and the ability of the Sp1 transcription factor to bind to the mutated Sp1 IIIT5A motif.

1.2. Aim:

To investigate the effect of HIV-1C Sp1 IIIT5A mutation on LTR transcription activity, Sp1 transcription factor expression levels, and the ability of Sp1 transcription factor to bind to Sp1IIIT5A mutated motif.

1.3. Hypothesis:

The Sp1 IIIT5A mutation may mediate HIV-1C LTR transcription activity and the ability of the Sp1 transcription factor to bind to Sp1IIIT5A mutated motif.

1.4. Objectives:

a) To introduce Sp1IIIT5A mutation on the HIV-1C consensus LTR sequence using molecular techniques such as cloning of a gene block.
b) To assess the effect of Sp1IIIT5A mutation on the HIV-1C LTR transcription activity in Jurkat and SVG cell lines compared to wild type.
c) Determine the expression levels of Sp1 transcription factor in Jurkat and SVG cell lines using the Western blot assay.
d) To characterize the binding affinity of the Sp1 transcription factor to mutated Sp1IIIT5A versus wild type by performing molecular docking and dynamics simulation.

CHAPTER 2: LITERATURE REVIEW

2.1. The human immunodeficiency virus (HIV) origin

The human immunodeficiency virus (HIV) was identified in 1983 by the esteemed scientist Luc Montagnier at the Pasteur Institute in Paris as the causative agent behind acquired immunodeficiency syndrome (AIDS) (Barré-Sinoussi *et al.*, 1983). HIV-1 continues to be a significant worldwide health concern with approximately 39.9 million individuals living with HIV-1, 1.3 million new HIV infections, and 630,000 AIDS-related deaths globally in 2023 (UNAIDS., 2024) (Figure 1A). However, this is still far below the numbers at the height of the epidemic in the 1990s, where UNAIDS estimates showed up to three million new infections per year. In the 2000s, up to 1.9 million people died each year in connection with AIDS. Most recently, the coronavirus has made it more difficult to combat the HIV epidemic that has been with us since the 1980s (Figure 1B).

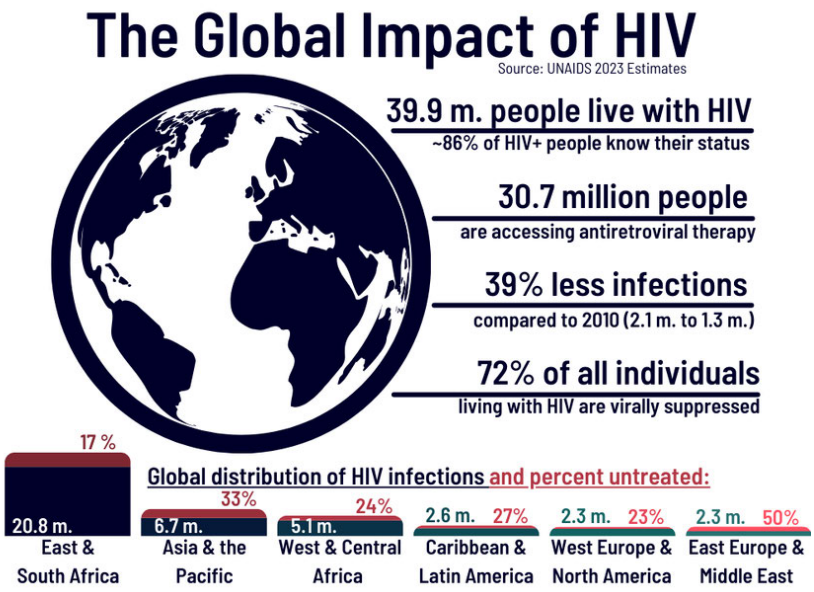


Figure 1A: **The summary of the global HIV epidemic**. Approximately 39.9 million people were living with HIV/AIDS, 1.3 million newly infected people which is 39% less compared to 2010 and 630,000 AIDS-related deaths in 2023 (UNAIDS., 2024).

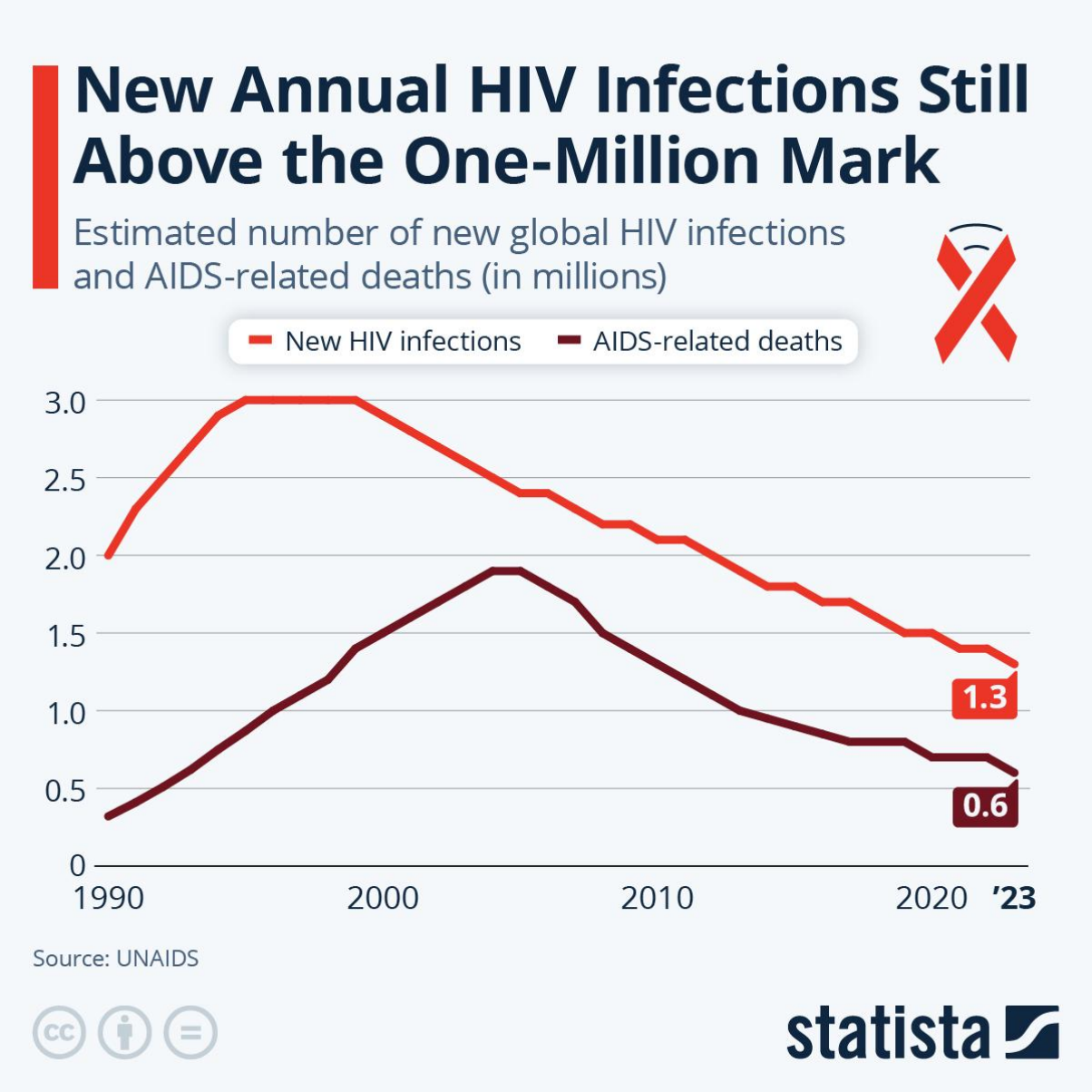


Figure 1B: Timeline depicting change in the HIV pandemic statistics overtime. In 2023, around 1.3 million people were newly infected with HIV and 600,000 died as a result of AIDS. However, this is still far below the numbers at the height of the epidemic in the 1990s, where UNAIDS estimates showed up to three million new infections per year. In the 2000s, up to 1.9 million people died each year in connection with AIDS. Most recently, the coronavirus has made it more difficult to combat the HIV epidemic that has been with us since the 1980s. (UNAIDS., 2024)

2.2. The HIV geographical distribution and HIV-1 subtypes

HIV is divided into two genetically distinct types namely HIV type 1 (HIV-1) and HIV type 2 (HIV-2) (Nyamweya *et al*., 2013). While HIV-2 is primarily found in West Africa, HIV-1 is present worldwide (Gao *et al*., 1999, Lemey *et al*., 2003). HIV-1 mutates extensively and shows high genetic diversity, posing significant challenges for effective surveillance and disease

control whereas HIV-2 has a lower transmission rate than HIV-1. HIV-1 is categorized into four groups: Group M (major), group O (outlier), group N (non-major and non-outlier) (Hemelaar *et al.*, 2020), and a recently discovered group P in Cameroon (Vallari *et al*., 2011). The majority of global HIV-1 infections are caused by group M viruses, while groups O and N are responsible for a small proportion of infections in central Africa. Group M is further divided into nine subtypes, labelled with the letters A–D, F–H, J, and K, along with numerous circulating recombinant forms (CRFs) and unique recombinant forms (URFs) (Hemelaar *et al.*, 2020) (see Figure 2). The CRFs are recombinant HIV-1 genomes that have infected at least three epidemiologically unrelated individuals. CRFs are named consecutively according to an internationally recognized nomenclature, with 106 distinct CRFs identified to date (Njai *et al.*, 2006, Tee *et al.*, 2008, Hemelaar *et al.*, 2019). URFs are recombinant HIV-1 sequences that do not show evidence of further transmission.

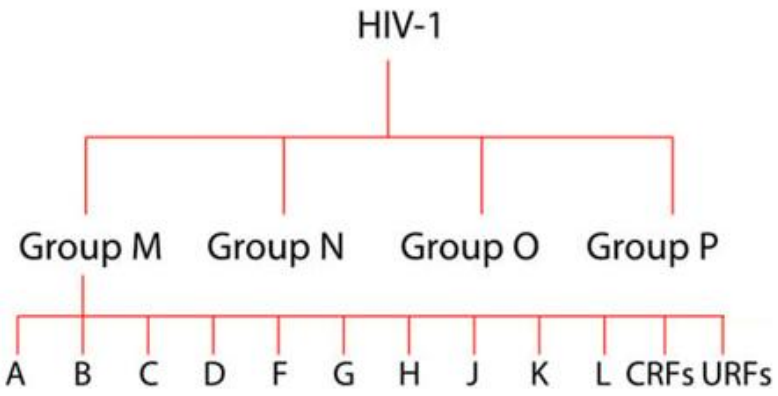


Figure 1: Schematic representation of the evolutionary relationships of human immunodeficiency virus type 1 (HIV-1) groups, subtypes, and recombinant forms. Worldwide proportions of HIV-1: subtype C (46.6%), B (12.1%), A (10.3%), G (4.6%), D (2.7%), F, H, J, K (together 0.9%), circulating Recombinant Forms (CRFs) 16.7%. N, O, and P groups are extremely rare. (Giovanetti *et al.*, 2020)

The geographic distribution of HIV-1 subtypes is not uniform. Subtype B is widespread in most parts of Europe and the Americas, while a diverse range of subtypes circulate in West and Central Africa (Hemelaar *et al.*, 2019). Subtype C accounts for around 50% of global HIV-1 infections and is the most prevalent strain in India and Ethiopia, while also being the dominant variant in Southern Africa (Hemelaar *et al.*, 2020) (Figure 3). Subtype D is prevalent in East and Central Africa, with sporadic cases seen in Southern and Western Africa. Subtypes F, H, J,

and K collectively occur in Central, Southern, and West Africa (Hemelaar *et al.*, 2006, Hemelaar *et al.*, 2019). Subtypes A, B, C, and G were accountable for 10,3%, 12.1%, 46,6%, 2.7%, and 4.6% of all global HIV-1 infections, respectively. Subtypes F, H, J, and K combined accounted for 0·9%. All CRFs accounted for 16·7% and URFs constituted 6·1% resulting in recombinants accounting for 22·8% of all global HIV-1 infections.

The distribution of HIV-1 subtypes and recombinants changes over time in countries, regions, and globally. At a global level during 2005–15, subtype B increased, subtypes A and D were stable, and subtypes C, G and CRF02_AG decreased. CRF01_AE, other CRFs, and URFs increased (Hemelaar *et al.*, 2019, Hemelaar *et al.*, 2020). It has also been shown that the most prevalent CRFs in the global HIV-1 epidemic are CRF01_AE and CRF02_AG. The CRF01_AE is a recombinant of a subtype A and a putative extinct subtype E ancestor, containing subtype A *gag, pol,* and subtype E *vif, vpr, env, nef,* and LTR. CRF02_AG is a recombinant of subtypes A and G, bearing subtype A *gag, env*, subtype A/G *pol, tat, rev, nef,* and subtype G LTR (Janssens *et al.*, 2000, Zhou *et al.*, 2020). Consequently, this study focuses on subtype C LTR due to its worldwide predominance, particularly in Sub-Saharan Africa.

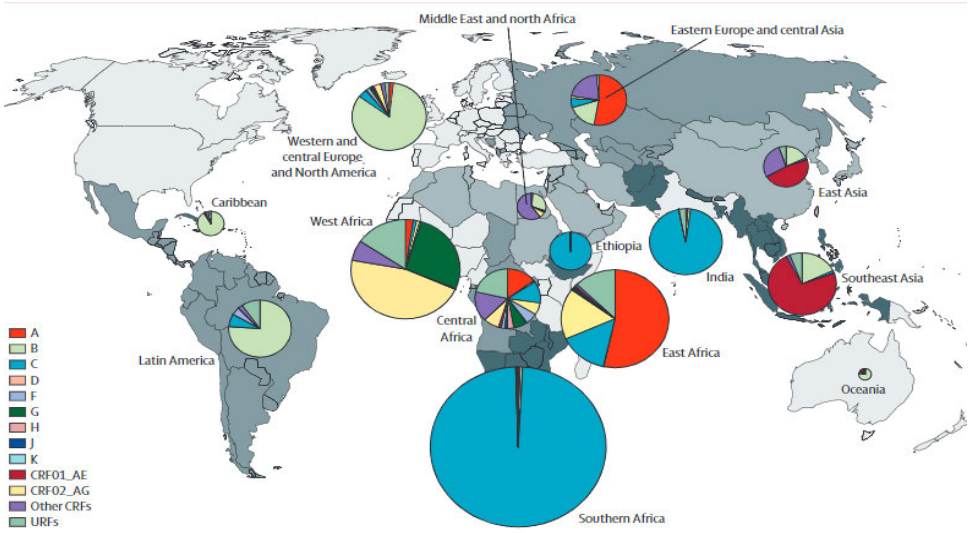
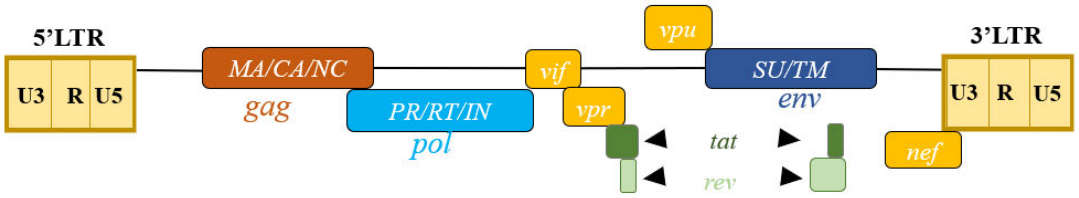


Figure 2: World map illustrating the prevalence of HIV-1 group M subtypes within each region. Pie charts show the percentage of each subtype that circulates within a region and the size of each pie chart represents the total number of infections in that region. Each subtype is colour coded. This map was adapted from subtype prevalence data from (Hemelaar et al., 2019).

2.3. HIV-1 genome and proteome

HIV-1 belongs to the Lentivirus genus within the Retroviridae family. The HIV-1 genome consists of two identical single-stranded RNA molecules that are enclosed within the core of the virus particle. The integrated genome of the HIV-1, also known as proviral DNA, is generated by the reverse transcription of the viral RNA genome into DNA, degradation of the RNA and integration of the double-stranded HIV DNA into the human genome. The DNA genome is flanked at both the 5' and 3' ends by LTR (long terminal repeat) sequence (figure 4). The 5′ LTR is the promotor that drives gene transcription. s. In the direction of 5′ to 3′ the reading frame of the group-specific antigen (*gag*) gene follows, encoding the proteins of the outer core membrane (MA, p17), the capsid protein (CA, p24), the nucleocapsid (NC, p7) and a smaller, nucleic acid-stabilising protein. The *gag* open reading frame is followed by the polymerase (*pol*) open reading frame coding for the enzymes protease (PR, p12), reverse transcriptase (RT, p51) and RNase H (p15) or RT plus RNase H (together p66) and integrase (IN, p32). Adjacent to the *pol* gene, is the envelope *(env*) open reading frame from which the two envelope glycoproteins gp120 (surface protein, SU) and gp41 (transmembrane protein, TM) are derived. In addition to the structural proteins, the HIV-1 genome codes for several regulatory proteins: Tat (trans-activator protein) and Rev (RNA splicing-regulator) are necessary for regulating HIV-1 replication, while the other regulatory proteins Nef (negative regulating factor), Vif (viral infectivity factor), Vpr (virus protein r) and Vpu (virus protein unique) have an impact on viral replication, virus budding and pathogenesis (Frankel and Young, 1998, Van Heuvel *et al.*, 2022).



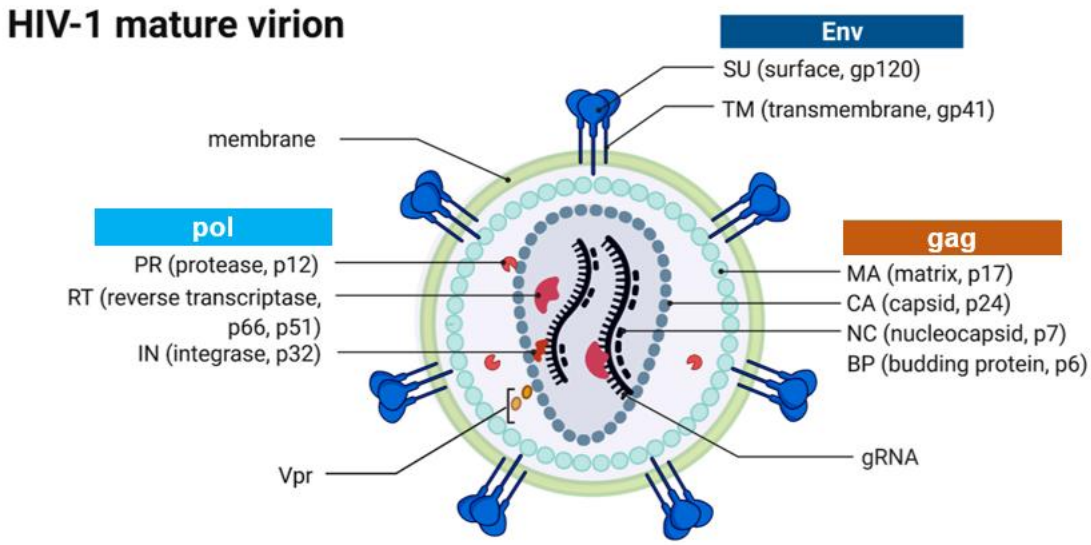


Figure 3: HIV-1 genome and virion structure. (Top) The HIV-1 genome's genetic layout is depicted, highlighting the coding frames for various structural, regulatory, and accessory proteins. (Bottom) The mature virion, enveloped by a lipid bilayer membrane derived from the host cell, exhibits a spherical shape and bears envelope glycoprotein trimers. The inner layer of the membrane serves as an anchor for MA proteins originating from Gag, and it also accommodates Vpr and PR. The central region of the virion contains the capsid, which houses two copies of gRNA, RT, and IN. The gRNA is stabilized by NC proteins. (Adapted from Figure 1 of (Van Heuvel et al., 2022).

2.4. The human immunodeficiency virus type 1 (HIV-1) replication cycle

2.4.1 The viral cell entry

Viral entry is a complex multi-step process that offers multiple possibilities for therapeutic intervention (Engelman and Cherepanov, 2012). Either directly or following unspecific binding of HIV-1 to its target cell, the infection process is initiated by the interaction of the external viral glycoprotein gp120 with the cellular CD4 receptor (Engelman and Cherepanov, 2012). CD4 receptor binding induces conformational changes in the *env* trimer that allow the interaction of gp120 with either the CXCR4 (X4) or CCR5 (R5) co-receptor. Co-receptor interaction induces additional conformational changes that allow the gp41 transmembrane protein, which is usually hidden by the gp120, to insert its hydrophobic fusion peptide into the cell membrane to make the first direct contact between the virus and its target cell (Didigu and Doms, 2012). Thereafter, the trimeric gp41 complex forms a helical bundle structure that pulls the cellular and viral membranes together, thus allowing virion fusion and the release of the contents of the virus particle into the cell.

2.4.2. Uncoating and Reverse transcription

Uncoating refers to the disassembly of the viral capsid before the import of the viral genome into the nucleus. As the capsid moves through the cytoplasm, it is assisted by dynein and kinesin 1 along the microtubules toward the nucleus (Engelman and Cherepanov, 2012). Following this, the capsid attaches its narrow end to the nuclear pore complex (NPC), interacting with the NPC-associated proteins Nup358 and Nup62 (Zila *et al.*, 2021). . After fusion, the process called reverse transcription (RT) is activated. Activation of RT takes place in the cytoplasm. HIV RT transcribes the single-strand HIV RNA genome into DNA (complementary DNA or cDNA). In parallel to DNA synthesis, the RNA strand is degraded enzymatically by RNase H, followed by the conversion of single-stranded cDNA into double-stranded DNA (proviral DNA) by the DNA-dependent DNA polymerase activity of RT (reviewed in (Kirchhoff, 2013)). This proviral DNA is transported via nucleopores into the cell nucleus in the form of a complex consisting of the integrase (IN) and linear or circular proviral DNA.

2.4.3. DNA Nuclear Import and Integration

After the viral DNA is synthesized by reverse transcription in the cytoplasm, it stably associates with IN, MA, NC, and some host cell proteins as a high-molecular-weight NPC. The NPC also known as pre-integration complex (PIC) is later transported to the nucleus for subsequent integration. This process leads to the establishment of a proviral state where the viral DNA is part of the host cell's DNA. It has been revealed that HIV-1 prefers to integrate into active genes (Schröder *et al.*, 2002, Mitchell *et al.*, 2004). One important factor that determines HIV-1 integration preferences is the LEDGF/p75 cellular transcriptional coactivator which is capable of binding both PIC-associated HIV-1 IN (Cherepanov *et al.*, 2003) and chromatin (Maertens *et al.*, 2003) and is considered the major tethering factor in HIV-1 integration.

2.4.4. Transcription

The integrated HIV-1 provirus serves as a template for the transcription of both viral messengers and genomic RNA by the cellular RNA polymerase II. Proviral transcription is initiated by the viral promoter, which is located in the U3 region of the 5' LTR and active in many cell types (Dutilleul *et al.*, 2020). The 5'LTR will be further discussed in subsequent section. Viral gene expression is strictly dependent on cellular transcription factors. These transcription factors (such as NF-κB, Sp1, and TATA-box binding protein (TBP)) are crucial

for initiating HIV-1 transcription, while others (including NFAT, C/EBP, Est-1, and AP-1) modulate promoter activity without being essential for viral transcription (Colin *et al.*, 2014). Initially, the transcriptional output is low because elongation of viral transcripts is very inefficient and the viral trans-activator protein Tat is required for effective viral gene expression (Karn and Stoltzfus, 2012). To increase transcriptional processivity, Tat binds to a specific sequence in the R region of the 5' LTR, named the trans-acting response (TAR) element. This effect is dependent on the cellular human positive transcriptional elongation factor b (pTEFb) which interacts with a critical protein-RNA complex consisting of the viral trans-activator Tat and the emerging HIV-1 RNA TAR. *Tat* allows the efficient synthesis of full-length HIV transcripts. More than 25 different mRNAs in three size classes are generated by alternative splicing: (i) unspliced RNA serving as genomic RNA or to produce the *Gag* and *Gag-Pol* precursors; (ii) singly spliced RNA encoding *vif, vpr, vpu,* and *env*; or (iii) fully spliced RNA expressing *tat, rev,* and *nef* (Kharytonchyk *et al.*, 2016). Transport of unspliced and partially spliced mRNAs from the nucleus to the cytoplasm is mediated by the viral Rev protein which interacts with the rev-responsive element (RRE) in the viral RNA and the cellular export factor Crm1 to connect these viral RNAs to the export machinery.

2.4.5. Translation

The fully processed mRNAs are utilized in the translation of the accessory protein Nef, as well as the regulatory proteins Tat and Rev (Nawroth *et al.*, 2014). Nef enhances viral infectivity by modifying signal pathways, reducing the expression of certain cell surface proteins like CD4 and major histocompatibility complex-I, and promoting viral transcription through NF-κB (Ganser-Pornillos *et al.*, 2008). This leads to a shift in viral protein expression towards those necessary for generating new virions (Kirchhoff, 2013).

2.4.6. Assembly, Budding and Maturation

HIV-1 virion assembly occurs at the plasma membrane (Figure 5), within specialized membrane microdomains. The HIV-1 Gag (and Gag-Pro-Pol) polyprotein itself mediates all of the essential events in virion assembly, including binding the plasma membrane, making the protein-to-protein interactions necessary to create spherical particles, concentrating the viral Env protein, and packaging the genomic RNA via direct interactions with the RNA packaging sequence. These events all appear to occur simultaneously at the plasma membrane, where conformational change(s) within Gag couples' membrane binding, virion assembly, and RNA packaging. Although Gag itself can bind membranes and assemble into spherical particles, the

budding event that releases the virion from the plasma membrane is mediated by the host ESCRT (endosomal sorting complexes required for transport) machinery (Hurley and Cada, 2018).

The viral particle undergoes a maturation process where its structural components, genetic material (gRNAs), and enzymes are rearranged, resulting in the creation of an infectious virion. This maturation is triggered by the self-activation of the PR enzyme, which then systematically cleaves the precursor proteins Gag and Gag-Pol. This releases the viral enzymes PR, RT, and IN, along with the structural proteins p17-MA, p24-CA, and p7-NC (Ganser-Pornillos *et al.*, 2008, Pornillos and Ganser-Pornillos, 2019). These structural alterations are crucial for the virion to become infectious. The NC protein tightly binds to the gRNA dimer, enhancing the stability of the connection between the two gRNA molecules (Mouhand *et al.*, 2020). The CA proteins assemble around the NC: gRNA complex, encapsulating the viral genome as well as RT and IN. The division of Gag into its subunits enhances the fusogenicity of the incorporated Env trimers. This marks the conclusion of the productive cell infection by the HIV-1 virion, which is now equipped for a new replication cycle (Van Heuvel *et al.*, 2022).

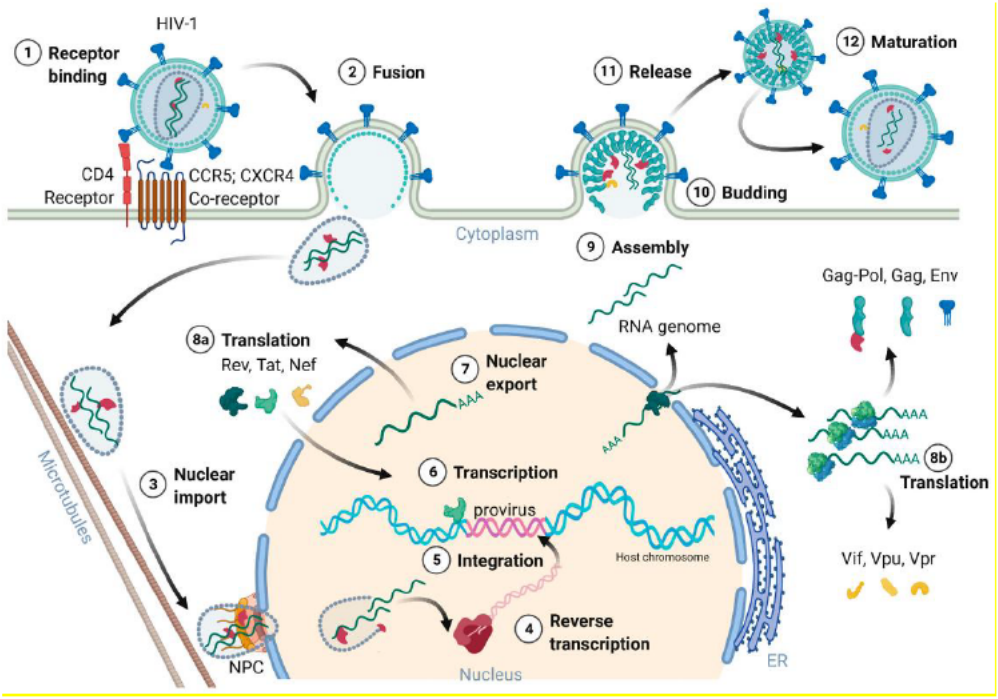


Figure 4: schematic overview of the HIV-1 replication cycle. (1) the HIV-1 infection begins with the binding of envelope glycoproteins gp120-SU to primary CD4 receptor and chemokine co-receptors (CCR5 or CXCR4) on the host cell surface. (2) the virion's membrane envelope then fuses with the cellular membrane, releasing the viral capsid into the cytoplasm. (3) The capsid travels along the microtubules to the nucleus. The capsid docks to the nuclear pore complex (NPC) and passes through

the pore into the nucleus. (4) The capsid partially uncoats during nuclear cell entry and the reverse transcription of the viral gRNA into the provirus is completed inside the nucleus. (5) The integrase together with co-factors promotes the integration of the provirus into highly active chromosomal regions of the host genome. (6) Tat activates gene transcription of the provirus. (7) the *rev* recruits several host proteins to export the intron-containing viral mRNAs. (8a/8b) Viral mRNA translation occurs within the cytoplasm, first, *rev, tat* and *nef* are expressed. Signal peptide-containing proteins such as Vpu and Env enter the endoplasmic reticulum (ER) for further posttranscriptional modifications. Glycosylated env passes through the Golgi apparatus and is cleaved by the cellular furin-like proteases into gp120-SU and gp41-TM. (9) Two gRNAs, *gag*, *pol*, *env* and assemble into nascent HIV-1 particles at the cell membrane. (10) Immature HIV-1 particles bud from the cell membrane. (11) Immature HIV-1 particles are released from the host cell. (12) During maturation, *gag* and *pol* precursor proteins are cleaved by the viral protease into their subunits MA, CA and NC as well as the viral enzymes PR, RT and IN. Upon finalization of the maturation, the newly formed HIV-1 virions are prepared for the next host cell infection, reinitiating a new replication cycle (Van Heuvel *et al.*, 2022).

2.5. The structure and function of HIV-1 provirus long terminal repeat (HIV-1 LTR)

The subsequent paragraph will discuss the structure of the HIV-1 LTR and the functions of the LTR domains and TFBSs.

2.5.1. The structure of HIV-1 LTR

The 9.2 kb HIV-1 RNA becomes the HIV-1 DNA after reverse transcription, which is then integrated into the host DNA through an integration process (Caputi, 2011). Similar to typical cellular mRNA, HIV RNA contains both a 5′UTR and a 3′UTR. However, there are key differences: (1) both the 5′ and 3′ UTRs of HIV-1 are longer than those found in cellular mRNA, and (2). The 5′UTR becomes the 5′LTR and the 3′UTR becomes the 3′LTR during reverse transcription (Zhang and Crumpacker, 2022). Upon successful integration, this process results in the formation of the provirus. The provirus retains all the conserved regulatory sequences in both the 5′ and 3′UTRs, but also gains a new U3 region at the start of the 5′LTR and a new U5 region at the end of the 3′LTR (see Figure 6). This gives the provirus its characteristic long terminal repeat (LTR) structure (Zhang and Crumpacker, 2022), producing a longer DNA sequence than the 5′ and 3′UTR of gRNA. A provirus has two identical enhancer-promoter sections flanking its genome with a 5' and 3' LTR symmetry.

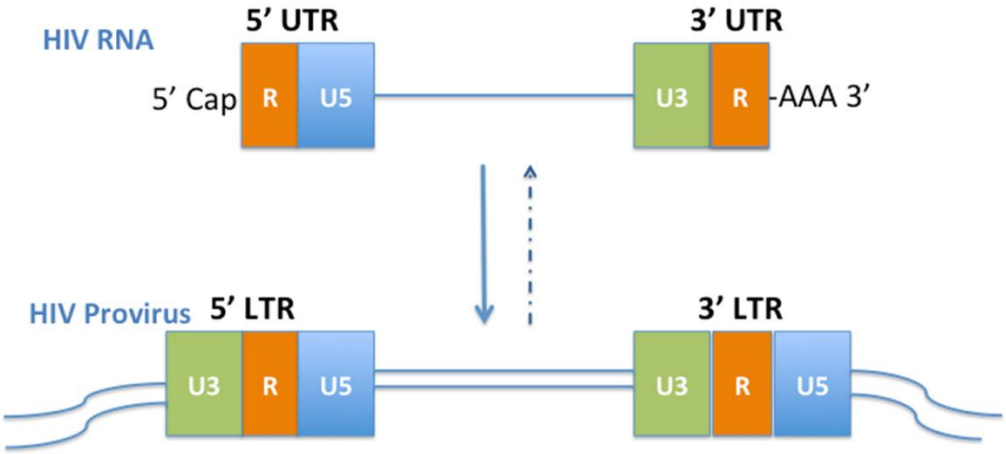


Figure 5: The HIV UTR and LTR. U: unique element. R: repeat element. Provirus is a metastable stage in the retroviral lifecycle. A breach of the metastable equilibrium depends on the host cell signals not the virus. Adapted from (Zhang and Crumpacker, 2022).

The 5' LTR of the HIV-1 RNA The U3 region is further divided into the functional core-promoter, enhancer and modulatory domains (Figure 7). These domains contain crucial TFBSs that bind an array of host and viral factors found in the core-promoter, core-enhancer and regulatory domains (Pereira *et al.*, 2000b). The core promoter domain is comprised of a specialized initiator element, a canonical TATA element, Enhancer box (E-box) and three tandem Sp1 binding sites. The core enhancer domain contains a subtype-specific number of inducible transcriptional activator nuclear factor kappa-light-chain enhancers of activated B cells (NF-kB). The core regulatory domain contains activator protein-1 (AP1) and nuclear factor of activated T-cells (NFAT) binding sites. The R region includes the mRNA initiation site (+1) and ends at a polyadenylation termination site. It also element contains the *trans*-acting responsive region (TAR), forming an RNA stem-loop structure for translation. The U5 element contains the polyadenylation signal (poly-A) and regulatory regions for reverse transcription and is known to separate the R region from the tRNA primer binding site used to initiate the reverse transcription.

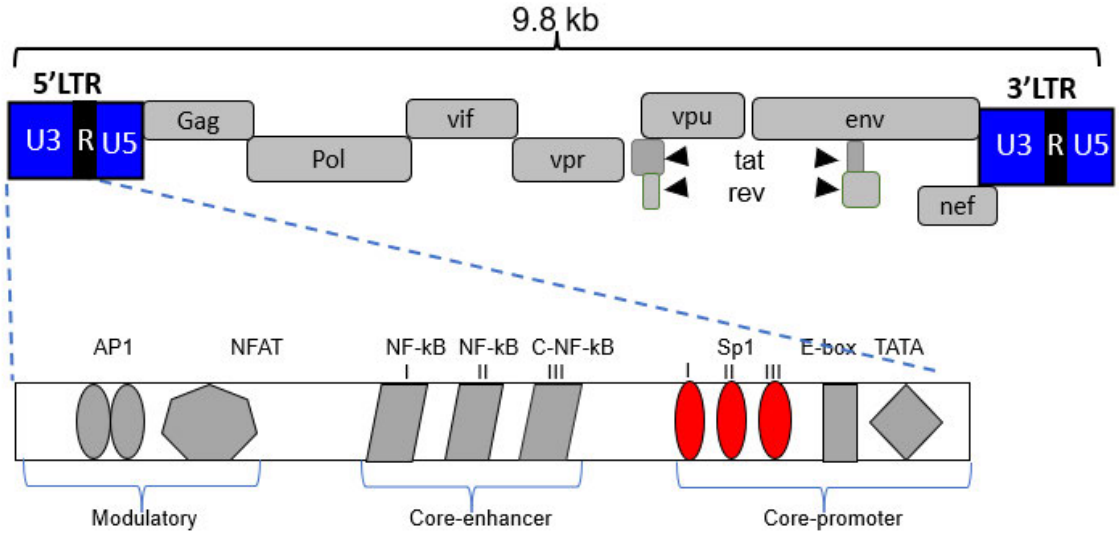


Figure 6: The genome of HIV-1 and structure of the HIV-1 LTR. The HIV-1 DNA genome is 9.8 kb in size and is flanked by identical long terminal repeat (LTR) sequences at both the 5' and 3' ends. The HIV-1 LTR is approximately 640 base pairs long and is divided into three regions: the unique 3' (U3), repeat (R), and unique 5' (U5) regions. The U3 region is further divided into three domains: the core promoter, core enhancer, and modulatory domain. The core promoter contains a TATA box, E-box, an initiator element, and three binding sites for the transcription factor Sp1. The core enhancer domain includes NF-κB binding sites, while the modulatory domain contains binding sites for activator protein 1 (AP-1), nuclear factor of activated T cells (NFAT), and CCAAT enhancer-binding protein (C/EBP).

2.5.2. The function of HIV-1 LTR

Among transcription factors in HIV-1 LTR, only a few have been described as key elements to initiate HIV-1 transcription including NF-kB, Sp1 and the TATA-box binding protein (TBP) (Munkanta *et al.*, 2005, Nonnemacher *et al.*, 2004). Other transcription factors like NFAT, C/EBP, Est-1, and AP-1) modulate the promoter activity without being essential for viral transcription. The TFBS within the U3 region of the HIV-1C 5'LTR plays a role in regulating both the basal and trans-activated transcription of the HIV-1 genome (Gray *et al.*, 2013). Specifically, the Sp1 motif polymorphisms are known to reduce both basal and Tat-activated transcription in CNS-derived HIV-1C 5'LTR sequences relative to lymphoid-derived HIV-1C 5'LTRs of people living with HIV-1 (Gray *et al*. 2015). After cellular stimulation, it has been shown that Sp1 acts synergistically with NF-kB via protein-protein interactions to activate viral transcription (Perkins *et al.*, 1993).

Many studies have been conducted based on different transcription factors and binding sites (Perkins *et al.*, 1993, Bachu *et al.*, 2012, Shah *et al.*, 2014b, Gaynor, 1992a), but few have been conducted on how Sp1 promotes HIV-1C LTR transcription activity (Shah *et al.*, 2014b, Miller-Jensen *et al.*, 2013). The binding of cellular Sp1 transcription factors to the three tandem Sp1

sites on the HIV-1 core promoter activates the Sp1 sites, thus enhancing the assembly of the preinitiation complex (Harrich et al., 1989), which is stabilized by transcription factor (TF) IIA and TFIIB and further completed by the recruitment of RNA Polymerase II-TFIIF, TFIIE and TFIIH (Thomas and Chiang, 2006). This study focuses on Sp1 mutation that has been previously shown to affect HIV-1 LTR transcription activity. The CSF-derived T-cell-derived Sp1III5A mutations were also examined and showed to induce HIV-1 LTR transcription activity (W Ntshangase et al, unpublished).

2.5.3. The HIV-1 LTR variation

The long terminal repeat sequences of the viral subtypes are highly diverse, differing up to 20–25% between subtypes (De Baar *et al.*, 2000) and has been shown to have a significant impact on promoter/ activity, viral replication kinetics, virulence, and dissemination (Jeeninga *et al.*, 2000). A comparison of HIV-1C LTR with that of others identified several distinct differences in the composition of the TFBS, including NF-kB (Munkanta *et al.*, 2005), NFAT, upstream stimulatory factor (Naghavi *et al.*, 1999), and other regulatory elements such as the Sp1, and the TAR region (Qu *et al*., 2016). Of these variations in the LTR, subtype-specific patterns within the enhancer element, exclusively consisting of the NF-kB motifs, are important given the profound impact NF-kB has on gene expression regulation from the viral promoter. The enhancer in most of the viral subtypes, including the prototype subtype B virus, consists of two identical and canonical NF-kB motifs except for subtypes A/E and C (Verhoef *et al.*, 1999). While in sub-type A/E, have one NF-kB site (Verhoef *et al.*, 1999), subtype C contains three NF-kB sites (Montano *et al.*, 1997, Naghavi *et al.*, 1999) with a minority of these LTRs containing a fourth motif that is either a canonical NF-kB site or an NF-κB-like site, which enhances viral transcription and replicative fitness (Papathanasopoulos *et al.*, 2002, Bachu *et al.*, 2012). The HIV-1 subtype C-LTR containing three or more NF-kB sites is associated with a stronger transactivation activity compared to the LTRs of other subtypes containing two NF-kB sites (Figure 7) (Monty *et al.*, 2000, Bachu *et al.*, 2012). Thus, unlike other HIV-1 subtypes, the HIV-1 subtype C LTR has a significant variation of NF-kB site number and sequence changes within these sites. Duplication of viral enhancer/promoter elements has been generally associated with an increase in the likelihood of a promoter being transcriptionally active (Miller-Jensen *et al.*, 2013) or with an increase in the level of transcriptional activity (Monty *et al.*, 2000).

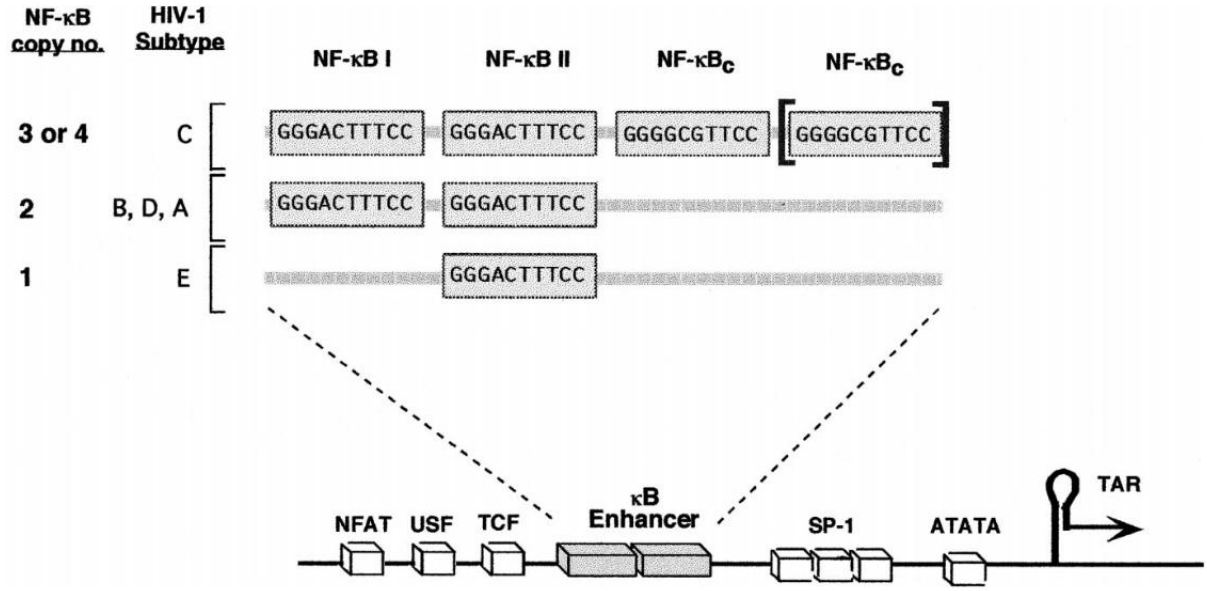


Figure 7: Schematic illustration of NF-kB enhancer region of 5 major human immunodeficiency virus type 1(HIV-1) subtypes (A–E). HIV-1 subtype C in southern Africa contains extra NF-kB binding sites (3or4). Subtypes HIV-1 B, -A, and -D contain 2 tandem sites, whereas HIV-1E isolates contain 1 NF-kB site (Monty et al., 2000).

2.5.4. Transcription activity of HIV-1 LTR

2.5.4.1. Basal transcription activity of HIV-1 LTR

HIV-1 gene expression is mainly regulated at the transcriptional level by hijacking the cellular RNA polymerase II (RNAPII) machinery, cellular host transcription factors that recognized multiple *cis*-regulatory elements along the LTR and through the recruitment of chromatin-modifying enzymes (Colin *et al.*, 2014). HIV-1 transcription initiates at the U3/R junction in the 5′ LTR and is regulated by cellular transcription factors that binds to the 5′ LTR through their specific binding sites (figure 7). Functional regions of 5'LTR contain numerous cis-regulatory elements that are recognized by both constitutive transcription factors such as Sp1 and Oct-1 (o-binding protein 1) and NF-κB, AP-1, and NFAT that are able to enhance or repress HIV-1 transcription (reviewed in (Jeon, 2013)) (Figure 7). The outcome of viral transcriptional activity is dependent on the balance between activators and/or repressors binding to the 5′LTR.

2.5.4.2. Tat-mediated transcription of HIV-1 LTR

The ability of the HIV-1 LTR to act as an inducible promoter was first described by Sodroski et al. (1985) who noted that the induction, or "transactivation", of transcription was due to the presence of a novel transactivating factor, Tat. In the absence of Tat, the short transcripts

predominate, whereas in the presence of Tat there is a dramatic increase in the levels of long transcripts (Wu and Marsh, 2003)). Thus, Tat participates in a positive feedback mechanism that ensures high levels of HIV-1 transcription following the activation of cells carrying HIV-1 proviruses. The simplest explanation for the ability of Tat to stimulate transcription from viral LTRs that carry TAR elements is that Tat binds directly to TAR RNA (Roy *et al.*, 1990, Karn, 1999). Briefly, the transcriptional elongation step is mainly regulated by the human positive transcriptional elongation factor b (P-TEFb), which interacts with the critical protein-RNA complex composed of the viral transactivator Tat and the nascent HIV-1 RNA TAR (Karn, 1999). The HIV-1 Tat protein is the strongest HIV-1 activator and promotes the RNAPII escape from the HIV-1 promoter to allow an efficient transcriptional elongation (D'Orso, 2024).

2.6. Methodology Review

2.6.1. Sanger sequencing of HIV-1 LTR mutations

The Next-generation sequencing methodology is used for wildly used (reviewed in (Behjati and Tarpey, 2013)), However, this is not used for sequencing the LTR. The main disadvantage of NGS in the clinical setting is putting in place the required infrastructure, such as computer capacity and storage, and also the personnel expertise required to comprehensively analyse and interpret the subsequent data. In addition, the volume of data needs to be managed skilfully to extract the clinically important information in a clear and robust interface. However, this study uses sanger sequencing method which has served as the cornerstone for genome sequence production since 1977 (Sanger *et al.*, 1977). Sanger sequencing provides highly accurate results for identifying DNA sequences, is cost-effective and simpler to perform compared to NGS, which requires more resources and setup (reviewed in (França *et al.*, 2002)). Its ability to correctly identify single nucleotide polymorphisms (SNPs), insertions, and deletions is unparalleled, especially for smaller DNA regions.

2.6.2. Transcription assays

Transfection is a cutting-edge technique for introducing foreign nucleic acids into eukaryotic cells (Kim & Eberwine, 2010; Chow *et al.*, 2016). Methods to assess transcriptional activity include Nuclear Run-On and Run-Off Assays, which evaluate the activity of RNA polymerase on specific genes by isolating nuclei and resuming transcription in vitro using labeled nucleotides. These nucleotides are incorporated into new RNA, which can then be detected or

quantified (Khraiwesh, 2011). However, these assays are technically complex, low-throughput, and often involve radioactive labeling.

The current study employs a luciferase assay, a sensitive, quantitative, and flexible method for measuring transcriptional activity (Stepanenko and Heng, 2017). The luciferase assay operates on the principle that the luciferase enzyme, typically derived from fireflies (Photinus pyralis), reacts with its substrate, luciferin, in the presence of ATP and oxygen to produce light. This light emission is directly proportional to the amount of luciferase produced by the cells, which reflects the transcriptional activity of a target gene (Smale, 2010). Many studies have used the luciferase assay under the control of LTR (Schwartz *et al.*, 1990, Chande *et al.*, 2012, Le Hingrat *et al.*, 2020, Madlala *et al.*, 2023).Transfection approaches are categorized as stable or transient. Stable transfection achieves long-term expression by integrating foreign DNA into the host genome or maintaining it as an episomal element (Lufino *et al.*, 2008). Transient transfection, by contrast, does not involve integration into the host genome (Riedl *et al.*, 2018) and can be performed using plasmids (Nejepinska *et al.*, 2012) or oligonucleotides. This method offers a low-background, high-signal-to-noise system that is particularly effective for studying transcription factors, regulatory elements, and gene expression. Additionally, its suitability for high-throughput screening and dual-reporter systems makes it highly valuable in research and drug discovery applications.

2.6.3. Expression levels of proteins using western blotting

Various techniques have been employed to assess protein expression, including quantitative PCR (qPCR) and reverse transcription qPCR (RT-qPCR). Both qPCR and RT-qPCR focus on measuring mRNA levels to estimate gene expression at the transcript level (Forero *et al.*, 2019). In RT-qPCR, mRNA is converted into complementary DNA (cDNA), which is then amplified in real-time to quantify the expression of specific genes. This method is highly sensitive, quantitative, and allows for either relative or absolute quantification, but it only measures mRNA, which may not always correlate directly with protein levels.

In contrast, the current study uses Western blotting (WB), a crucial technique for detecting proteins, especially those present in low abundance. WB involves transferring protein patterns from a gel to a microporous membrane. Since its inception, protein blotting has evolved to include various transfer methods (reviewed in (Bass *et al.*, 2017)). The principle of WB lies in detecting proteins through the binding of antibodies (Abs) to specific antigen portions or

epitopes. This interaction is highly specific, occurring between the antigen's epitope and the antibody's paratope region (Kurien *et al.*, 2011). Typically, secondary antibodies (2°Ab) are used to detect the antigen bound by a primary antibody (1°Ab), as primary antibodies are usually not conjugated to labels for detection. Secondary antibodies are conjugated to reporter functions like horseradish peroxidase (HRP) or fluorophores, allowing detection through imaging devices. To account for variations in sample preparation and loading, normalization using housekeeping proteins (HKPs) such as GAPDH or β-actin is essential, as these proteins serve as internal controls assuming their expression remains stable under the experimental conditions (Welinder and Ekblad, 2011).

2.6.4. Molecular docking and dynamics simulation

Various methods are used to evaluate compound binding affinities, including homology modeling, comparative modeling, Free Energy Perturbation (FEP) and Thermodynamic Integration (TI) and molecular docking and dynamics simulation. Homology modeling predicts a protein's 3D structure using known structures of homologous proteins through tools like MODELLER, SWISS-MODEL, and I-TASSER (Hasani and Barakat, 2017). While helpful when the target structure is unknown, its accuracy depends on the template and is less suitable for detailed binding studies. The FEP and TI calculate the binding free energy changes upon ligand binding or modification with higher accuracy (reviewed in (Chipot, 2023)). It uses physics-based methods such as GROMACS (with plugins), AMBER, and NAMD software to simulate changes in energy as molecules interact or transform. This method provides highly accurate binding energy estimates, good for detailed affinity predictions. However, it is computationally expensive and time-consuming; often requiring high-performance computing resources. This study used molecular docking and dynamic simulation which uses 3D structure of the protein, typically obtained from databases like the Protein Data Bank (PDB) (reviewed in (Vidal-Limon *et al.*, 2022)). It analyzes the top-ranked poses by examining binding orientation, interaction types, and energy scores and visualizes the docking interactions to identify key residues involved in binding, which can guide further refinements m. Combining docking and MD provides a robust strategy in drug design (Mandlik and Singh, 2016). Typically, docking generates an initial complex model, which can then be refined and validated through MD simulation. This approach allows researchers to evaluate the stability of docked poses, study ligand-induced conformational changes in the receptor and explore the binding site dynamics, potentially uncovering additional binding features.

2.7. HIV-1 latency

HIV-1 latency refers to a state in which the virus remains dormant within infected cells, particularly resting memory CD4+ T cells, without producing infectious virions. This latent infection allows HIV-1 to persist in the host despite the use of antiretroviral therapy (ART), which effectively suppresses active viral replication but does not eliminate the latent reservoirs. After infection, HIV-1 integrates its genome into the host cell's DNA. In resting CD4+ T cells, this integrated viral DNA exists in a transcriptionally silent state, often due to epigenetic modifications that inhibit gene expression (Siliciano and Greene, 2011, Battistini and Sgarbanti, 2014). Key transcription factors necessary for HIV gene expression, such as NF-κB and NFAT, are often sequestered in the cytoplasm or are present at suboptimal levels in resting T cells. This prevents the activation of viral transcription from the long terminal repeat (LTR) region (reviewed in (Dufour *et al.*, 2020)). Current strategies aimed at eradicating HIV focus on "shock and kill" approaches, which aim to reactivate latent virus to make it visible to the immune system while simultaneously targeting and eliminating these cells. However, the complexity of latency mechanisms complicates these efforts (Khanal *et al.*, 2021). Research into the molecular mechanisms governing latency inform novel therapeutic strategies that specifically target these pathways.

CHAPTER 3: METHODS AND MATERIALS

3.1. Cloning of the HIV-1 LTR Sp1III5A gene block into pGL3 luciferase reporter vector

3.1.1. Restriction Digestion of the pGL3 plasmid and HIV-1 LTR Sp1III5A gene block

The HIV-1 LTR Sp1IIIT5A gene block, identified as #IDT4601567 was purchased from WhiteSci IDT. The HIV-1C LTR Sp1IIIT5A consists of U3 (455bp), R (96 bp), and U5 (83bp) (Pereira *et al.*, 2000a, Jeeninga *et al.*, 2000), with U3 further containing elements of promoters, enhancers, and multiple transcription factor binding sites (TFBS) (Mori and Valente, 2020). There are four main regions according to function, relative to the transcription start site (TSS) at +1 in HIV-1 LTR: the modulatory region (−455 to −104), the enhancer element (−109 to −79), the core promoter (−78 to −1), and the TAR region (+1 to +60) (Mori and Valente, 2020). The gene block was econstituted by dissolving in PCR-grade water. The concentration of the solution was determined to be 10.7ng/µL using a NanoDrop (Thermo Fisher Scientific, MA, USA). A total amount of 100ng/µL HIV-1 LTR Sp1III5A gene block and the pGL3 vector were digested using KpnI and HindIII (New England Biolabs, Massachusetts, USA) restriction enzymes independently. This resulted in the creation of complementary ends and the linearization of pGL3 plasmid. Each digestion reaction involved 1 Unit of each restriction enzyme (KpnI or HindIII), 1X CutSmart Buffer (New England Biolabs, Massachusetts, USA) and DEPC-treated water, making a final reaction volume of 10µL. These digestion reactions were performed in a thermocycler at 37°C for 1 hour, followed by heat inactivation at 80°C for 20 minutes. Subsequently, the 5µL of restriction digestion reaction of the 5'LTR-pGL3 recombinant was electrophoresed on a 1% agarose gel for 1 hour at 120 Volts and 1Kb DNA Ladder (Thermo Fisher Scientific, Invitrogen, MA, USA) to confirm the size of the restricted plasmid DNA. Following, 1uL of the reaction was loaded in a separate well as a control for visualization under ultraviolet (UV) light to detect the plasmid, while the remaining 9µL reaction was loaded in a different well for gel extraction. The section of the gel containing the 9µL reaction was not exposed to UV light to preserve the integrity of the plasmid.

3.1.2. Gel Extraction

The gel purification process for the linearized pGL3 plasmid was conducted using the GeneJet Gel Extraction Kit (Thermo Fisher Scientific, Invitrogen, MA, USA) according to the manufacturer's instructions. Briefly, an empty 1.5mL microcentrifuge tube was weighed to record its weight. Subsequently, the linearized pGL3 plasmid was extracted from the gel using

a gel extractor. The gel slice containing DNA fragment or band corresponding to the linearized pGL3 plasmid was then transferred into the pre-weighed 1.5mL microcentrifuge tube and weighed again. The weight of the DNA fragment was then determined by subtracting the weight of the empty tube from the weight of the tube containing the gel slice containing DNA fragment or band corresponding to the linearized pGL3 plasmid. This weight was used to calculate the required volume of binding buffer for the gel purification in the next step, where 1µL of binding buffer was equivalent to 1mg of the weight of the gel slice containing DNA fragment or band corresponding to the linearized pGL3 plasmid.

The volume of binding buffer needed was added to the microcentrifuge tube containing the gel slice, and this mixture was incubated at 60°C in a dry heat block until the gel completely dissolved in the binding buffer. Subsequently, the solubilized gel solution was transferred to a GeneJet purification column, which was then centrifuged at 13,000rpm for 1 minute to facilitate DNA binding to the silica membrane in the spin column. The resulting filtrate was discarded, and 700µL of Wash Buffer (used to remove debris or contaminants) was introduced to the column, followed by another centrifugation at 13,000rpm for 1 minute. The resulting filtrate was also discarded, and an additional spin at 13,000rpm for 1 minute was performed to eliminate excess buffer from the column. Following this, the QIAquick spin column was placed in a new 1.5mL microcentrifuge tube, and 30µL of elution buffer was added to elute the DNA. This mixture was incubated for 1 minute and subsequently centrifuged at 13,000rpm for 1 minute, resulting in the collection of purified linearized empty pGL3 plasmid, which was later quantified using the Nanodrop spectrophotometer.

3.1.3. Ligation of the digested pGL3 plasmid and HIV-1 LTR Sp1III5A gene block

The process involved cloning the 5'LTR insert, which had been digested, into the linearized pGL3 expression vector. Specifically, the linearized pGL3 plasmid and the LTR insert (Sp1III5A gene block) were ligated using T4 DNA Ligase (New England Biolabs, Massachusetts, USA). A ligation ratio of 1:27 (LTR (640bp): pGL3(4818)), which had been determined as the optimized ratio through the NEB ligation calculator (NEBio Calculator v1.15.0) available online, following the manufacturer's instructions. The ligation reaction mixture was composed of 100ng of insert DNA, 50ng of plasmid DNA, 1X T4 DNA ligase Buffer, and 1U of T4 DNA ligase and DEPC-treated water was added to achieve a final reaction volume of 15µL. Subsequently, the ligation reaction was incubated in a thermocycler at 25°C for 10 minutes and then heat-inactivated at 65°C for 20 minutes. Afterward, the tube containing

the ligation mixture, now comprising the HIV-1 LTR Sp1III5A-pGL3 recombinant plasmid, was placed back on ice in preparation for the subsequent transformation reaction.

3.1.4. Transformation of the HIV-1 LTR Sp1III5A-pGL3 recombinant plasmid into JM109 E. coli competent cells

The recombinant plasmid HIV-1 LTR Sp1III5A-pGL3 was transformed into competent JM109 E. coli cells (New England Biolabs, Massachusetts, USA) following the provided instructions. In short, 5µL of the ligation mixture was combined with 50µL of pre-chilled JM109 cells and gently mixed. This mixture was then kept on ice for 5 minutes to promote close interaction between the recombinant plasmid and the cells. Subsequently, the plasmid-cell combination was briefly heat shocked in a pre-warmed 42°C water bath for 30 seconds, allowing the DNA to enter the cells through the disrupted membrane. Following this, the heated mixture was returned to ice for 2 minutes to retain the plasmids within the bacteria. Next, 450µL of SOC media (a microbial growth medium) was introduced to the mixture and it was incubated at 37°C for 1 hour while being gently shaken at 230rpm. Following the 1-hour shaking period, 100µL of the mixture was spread onto an LB agar plate containing ampicillin, using a surface spreader. The plates were then left to incubate at 37°C for 16-20 hours overnight. As the pGL3 plasmid carries an ampicillin-resistant gene, the bacterial cells that have taken up the plasmid become ampicillin-resistant, allowing them to thrive on the agar plate and form colonies.

3.1.5. Screening of positive clones through colony PCR of the HIV-1 LTR Sp1III5A-pGL3 recombinant plasmid

After the overnight incubation, the agar plates were examined to verify the presence of bacterial growth. These plates showed numerous individual colonies, indicating a successful transformation. To confirm the cloning success, a colony PCR was conducted to identify the presence of the 5'LTR insert in single colonies. This was done using a previously optimized in-house procedure. In detail, each colony was picked from the ampicillin agar plates using a sterile pipette tip and gently streaked onto a separate ampicillin agar plate to create a master plate for future inoculation. Afterward, the colony was mixed with 10µL of DEPC-treated water in a 96-well plate. This plate with diluted colonies was subjected to a 95°C incubation in a thermocycler for 10 minutes to lyse the bacterial colonies through heating to release DNA for PCR.

Once the plasmids were lysed, PCR was carried out using the following conditions: pre-denaturation at 94°C for 5 minutes, 25 cycles of denaturation at 94°C for 15 seconds, annealing at 55°C for 30 seconds, 68°C synthesis for 30 seconds, extension at 68°C for minutes and finally held at 4˚C. To elaborate, the PCR mixture consisted of 0.2µM 5'KpnI and 3'HindIII primers (5'T7LTRKpn15'taatacgactcactatagggTTGGTACCTTTAAAAGAAAAGGGGGGAC-3'_nt 9064 to 9085) and reverse primer (3'Sp6LTRHindIII5'-atttaggtgacactatagAGCTTTATTGAGGCAAGCTTAGTGGGTT-3'_nt9622-9593), 1X High Fidelity PCR Buffer, 0.2mM dNTP mix, 2mM MgSO4, and 1U Taq DNA Polymerase (ThermoFisher Scientific, Invitrogen™, USA) and DEPC-treated water was added to reach a final PCR reaction volume of 9µL. Subsequently, 1µL of the boiled recombinant plasmid (DNA template) was added to yield a final PCR reaction of 10µL. Following this, 2µL of the colony PCR products were assessed using 1% agarose gel electrophoresis as mentioned in section 1.1. A positive cloning result was determined by observing a single band at approximately 640bp when compared to a 1Kb DNA Ladder from Invitrogen (ThermoFisher Scientific, Massachusetts, USA). This indicated the successful cloning of the HIV-1 LTR Sp1III5A into the pGL3 plasmid.

3.2. Sequencing of the recombinant HIV-1 LTR Sp1III5A-pGL3

The positive clones were then subjected to Sanger sequencing, followed by a sequencing clean-up and analysis.

3.2.1. Sequencing

Once a positive clone was confirmed, the positive colony PCR products, as previously described, underwent sequencing using the Big Dye Terminator v3.1 Sequencing Kit (Thermo Fisher Scientific, Massachusetts, USA), following the manufacturer's instructions. To obtain the nucleotide sequence of the recombinant HIV-1 Sp1III5A-pGL3, separate sequencing mixtures were prepared for each sample, employing both the forward (5'KpnI) and reverse (3'HindIII) primers, resulting in forward and reverse sequences for each sample. Each sequencing mixture comprised 1X Big Dye sequencing Buffer (2µL), 3.2µL of either the forward or reverse primer, 0.4µL of light-sensitive Big Dye v3.1 enzyme, 1µL of the colony PCR product (diluted to 20ng/µL) as the DNA template, and 3.4µL of DEPC-treated water, totaling a final reaction volume of 10uL. The 96-well plate containing the sequencing reaction

was subjected to the following thermocycler conditions: initial denaturation at 96˚C for 1 minute, followed by 25 cycles of denaturation at 96˚C for 10 seconds, annealing at 50˚C for 5 seconds, extension at 50˚C for 4 seconds, and a final hold at 4˚C. The sequencing plates were shielded from light exposure.

3.2.2. Sequencing clean-up

Post-cycle-sequencing reaction, the sequences underwent a sequencing clean-up process to eliminate unwanted primers, dNTPs, or enzymes. This was carried out according to a previously established in-house protocol. In brief, 1µL of 125mM ethylenediaminetetraacetic acid (EDTA) was added to each well of the 96-well plate containing the sequencing products, followed by the addition of 26µL of 3M Sodium Acetate solution (NaOAE) with a pH of 5.2, mixed with 100% cold ethanol to facilitate the DNA precipitation. The 96-well plate was then sealed with an adhesive plastic seal, briefly vortexed, and wrapped in foil before being left to incubate at room temperature for 10 minutes. Afterward, the plate was centrifuged at 4000rpm for 2 minutes to pellet out the precipitated DNA. Following centrifugation, the supernatant was removed by inverting the plate onto tissue paper and centrifuging at 150x gravity (g) for 5 minutes. To further process, 35µL of 70% ethanol was added to each well to wash the pellet, and the plate was centrifuged at 3000rpm for 5 minutes. Once again, to discard the supernatant, the plate was inverted on tissue paper and centrifuged at 150xg for 1 minute. Post-centrifugation, the 96-well plate was dried in the thermocycler at 50°C for 5 minutes to desiccate the DNA pellet and eliminate any residual supernatant. Finally, the plate was sealed with an adhesive foil and stored in the -20°C freezer until analysis in the ABI 3130XI Genetic Analyzer (Applied Biosystems).

3.2.3. Sequencing analysis

After processing the sequencing reaction in the Genetic Analyzer, forward and reverse primer sequences for the recombinant HIV-1 LTR Sp1III5A-pGL3 were obtained. These sequences underwent a series of steps including cleaning, editing, and assembly using Sequencher 5.1 software (Gene Codes in Ann Arbor, Michigan, USA). Subsequently, the refined HIV-1 Sp1III5A-pGL3 sequences were inputted into Jalview v2.10.5, developed by the University of Dundee, for a multiple sequence alignment. This was performed in comparison with the HIV-1 subtype C 5’LTR consensus sequence from Zambia (accession number: AF127567.1). The aligned sequences were then analyzed using BioEdit Sequence Alignment Editor v7.2.5 to identify any potential Sp1III5A mutations.

3.3. Maxiprep of positive clones

Using the master plate, a single positive colony from each sample was streaked and introduced into 200mL of Lysogeny Broth (LB Broth) with 200µg/mL of ampicillin in a 1:1 ratio. This mixture was then left to incubate at 37°C for an overnight period of 16-18 hours, with agitation at 230rpm in a shaking incubator. The following day, the LB broth culture displayed a cloudy appearance, confirming bacterial growth. It was then refrigerated (4-8°C) for 2 hours. Subsequently, the cultures were moved into 50mL falcon tubes and subjected to centrifugation at 4000rpm for 30 minutes at 4°C. After centrifugation, the resulting liquid above the cell pellet was poured off. The plasmid DNA clones were subsequently collected and purified using the Qiagen Plasmid Maxiprep Kit (QIAGEN, Hulsterweg, Netherlands) following the provided instructions. To achieve this, the cell pellet was loosened by suspending it in 10mL of Buffer P1. Following this, Buffer P2, which induces cell lysis, was added and the contents were mixed vigorously through inversion. The mixture was then left to incubate for 5 minutes at room temperature. Afterward, 10mL of Buffer P3, a neutralization buffer that leads to DNA precipitation, was added, mixed through inversion, and left to incubate on ice for 30 minutes. This was followed by centrifugation at 4000rpm for 1 hour at 4°C. After centrifugation, which resulted in the separation of DNA (in the supernatant) from cell debris (forming a white fluffy coat), the resulting supernatant was delicately extracted from the tube using a serological pipette, taking care not to include any cell debris. This supernatant was then added to a QIAGEN-tip that had been prepared with 10mL of Buffer QBT (equilibration buffer). The supernatant was allowed to flow through the resin on the tip under the force of gravity, which caused the plasmid DNA to bind to the QIAGEN-tip. The QIAGEN-tip 500 underwent two washes with 30mL of Buffer QC (a wash buffer for removing debris or contaminants) on the vacuum manifold.

Following this, the QIAGEN-tip was placed into a new 50mL conical falcon tube using the blue tip holder, and the DNA was released using 15mL of prewarmed (65°C) Buffer QF (elution buffer). After elution through gravity flow, 10.5mL of isopropanol was introduced to the eluted DNA for DNA precipitation. This mixture was then subjected to centrifugation at 4000rpm for 2 hours at 4°C. After centrifugation, the resulting liquid above the pellet was discarded, and 5mL of 70% ethanol (used for washing the precipitated DNA) was added to resuspend the pellet. This was followed by centrifugation at 4000rpm for 1.5 hours. Next, the liquid above the pellet was discarded, and the pellet itself was air-dried for 8-10 minutes before being finally dissolved in 100µL of DEPC-treated water. The purified Sp1III5A-pGL3 recombinant plasmids

were subsequently quantified using a Nanodrop spectrophotometer and employed in subsequent experiments.

3.4. Transfection of HIV-1 LTR Sp1III5A-pGL3 into Jurkat and astrocyte cell lines

3.4.1. Tissue Culture

The Sp1 motif polymorphisms are known to reduce both basal and Tat-activated transcription in CNS-derived HIV-1C 5'LTR sequences relative to lymphoid-derived HIV-1C 5'LTRs of people living with HIV-1 (Gray *et al*. 2015) which is why astrocytes and Jurkat cells were chosen for this particular study.

3.4.1.1. Preparation of Astrocytes

The Astrocyte (SVG) cells are a type of adherent cells chosen for their representation of central nervous system (CNS) cells. These cells were cultured in a growth medium composed of Dulbecco modified Eagle medium (DMEM) (Gibco BRL, Life Technologies, Cergy Pontoise, France) supplemented with 10% FBS (Gibco BRL, Life Technologies, Cergy Pontoise, France), 0.5% Pen-Strep (Sigma–Aldrich, St. Louis, MO, USA), 1% HEPES (Gibco BRL, Life Technologies, Cergy Pontoise, France), 1% L-Glut (Gibco BRL, Life Technologies, Cergy Pontoise, France), and 1% light-sensitive N2 supplement (Sigma–Aldrich, St. Louis, MO, USA).

In short, a vial containing SVG cells was taken out of the liquid nitrogen tank and thawed. This was done by placing the vial in a 37˚C water bath until only a small amount of ice remained, ensuring it was not fully immersed in water and not left for more than 2 minutes. After thawing, the vial was sterilized with 70% isopropanol, and the cell mixture was swiftly transferred into a 50 mL tube using a serological pipette along with 30mL of pre-warmed DMEM. The mixture was then centrifuged at 290 x g for 10 minutes to remove DMSO, and the resulting liquid was discarded. Subsequently, the cell pellet was re-suspended in 5mL of fresh, pre-warmed growth medium, moved to a T75 culture flask, and supplemented with an additional 10mL of pre-warmed growth medium before being placed in an incubator at 37°C for 24 hours. Thereafter, the original growth medium was replaced with 15mL of freshly warmed medium. The flask was then returned to the incubator for a period of 2-3 days. After 3 days, the culture flask containing cells was examined using a fluorescent imager to evaluate growth, which showed a 90-100% coverage. The cells were subsequently divided by pouring off the supernatant and washing the cell layer with 5mL of PBS.

Additionally, the cells were treated with 3mL of 0.25% Trypsin-EDTA (GIBCO BRL, Life Technologies, Cergy Pontoise, France) and left to incubate at room temperature for 30 seconds. Following this, the trypsin solution was removed, and the flask was further incubated at 37°C in 5% CO_2 for 4 minutes. Immediately after, the flask was replenished with 10mL of the growth medium, and 1mL of the cell suspension was placed into an Eppendorf tube for cell counting. The cell count was carried out. Briefly, the cells in the flask were combined using a serological pipette by pipetting up and down. Subsequently, 1mL of this mixture was transferred into an Eppendorf tube. Another Eppendorf tube received 10µL of Trypan blue (BIO-RAD Laboratories, Hercules, CA). Then, 10µL of the cell suspension from the first tube was extracted and blended with the 10µL of Trypan blue in the second tube. From that mixture, 10µL of the trypan blue cell suspension mixture was dispensed onto one side of a slide for cell counting, while an additional 10µL was deposited onto the other side. The slide was then placed into the TC20™ Automated Cell Counter (BIO-RAD Laboratories, Hercules, CA) to ascertain both the cell count per mL and viability. Subsequently, the cells were seeded into new T75 culture flasks at a density of $1x10^6$ cells/mL and with a confluency of 90% or higher, in 10mL of pre-warmed growth medium prior to incubation at 37°C in 5% CO_2 for 2-3 days. The cells were maintained until the 20^{th} passage, after which a fresh vial was requested.

3.4.1.2. Preparation of Jurkat cells

Jurkat cells, which represent those found in the peripheral compartment, were cultured in a growth medium comprising RPMI 1640 (Gibco BRL, Life Technologies, Cergy Pontoise, France) supplemented with 10% Fetal bovine serum (FBS) (Gibco BRL, Life Technologies, Cergy Pontoise, France), 5% Penicillin Streptomycin (Pen-Strep) (Sigma–Aldrich, St. Louis, MO, USA), 1% HEPES (Gibco BRL, Life Technologies, Cergy Pontoise, France), and 1% L-Glutamine (L-Glut) (R10 medium) (Sigma-Aldrich, St. Louis, MO, USA). This medium is referred to as R10.

To thaw the frozen Jurkat cells, a vial was taken from the liquid nitrogen tank and placed in dry ice. The vial was then thawed in a prewarmed water bath at 37˚C with gentle swirling until only a small amount of ice remained. This process took no longer than 2 minutes, and the vial was kept from being fully immersed in water. After thawing, the vial was sanitized with ethanol, and the cell suspension was swiftly pipetted into a 15mL tube containing 9mL of pre-warmed growth medium (R10). The tube was centrifuged for 10 minutes at 290 x g (Eppendorf centrifuge 5810R, Merck, Germany) to eliminate dimethyl sulfoxide (DMSO). The resulting supernatant was discarded, and the cell pellet was re-suspended by pipetting up and down in

10mL of pre-warmed growth medium before being centrifuged again for 10 minutes at 290 x g. Following this, the supernatant was discarded, and the cell pellet was gently resuspended by pipetting up and down in 2mL of pre-warmed growth medium. Next, the 2mL cell suspension was transferred into a T-25 culture flask, which already contained 3mL of pre-warmed growth medium, resulting in a total volume of 5mL. The day after, the cell culture was supplemented with an additional 5mL of pre-warmed growth medium, bringing the total volume to 10mL.

On the subsequent day, using a serological pipette, the cells were moved from the T25 flask to a T75 flask and supplemented with 5mL of pre-warmed growth medium. The flask was then incubated for 24 hours at 37°C with 5% carbon dioxide (CO_2). After this incubation period, the T75 cell culture flask was examined on the Zoe^{TM} fluorescent imager (Bio-Rad, Hercules, California, USA) to assess cell growth quality before the splitting process. The cells were counted, aiming for 95-100% confluency, employing the same method as was done with the astrocytes. After this procedure, the cells from the flasks were transferred into fresh T75 culture flasks, with approximately 35.0 x 10^4 cells in 25mL of pre-warmed growth medium. They were then incubated for 24 hours at 37°C in an environment containing 5% CO_2. These cells were consistently cultured by passaging them once a week, up until passage 20. After this point, a new vial of fresh cells was obtained for continued experimentation.

3.4.2. Transfection into Jurkat cells

To evaluate the transcriptional activity of Sp1III5A, the HIV-1 LTR Sp1III5A-pGL3 modified plasmid was transfected into Jurkat cells using a previously fine-tuned in-house method. In a nutshell, Jurkat cells were placed in a 24-well tissue culture plate with a density of 500,000 cells per well in 400µL of RPMI lacking antibiotics but supplemented only with 10% FBS to support cell growth. Before seeding, the transfection mixture was prepared. Essentially, 1µL of Sp1III5A -pGL3 recombinant (300ng) and 1µL of either empty pCTat (100ng) (a vector lacking Tat) for basal transcription or pCTat for tat-mediated transcription were blended with 48µL of serum-free RPMI for each well. The lipid-based transfection agent was created by combining 1µL of polyethyleneimine (PEI) and 49µL of serum-free RPMI for each well and left at room temperature (RT) for 5 minutes. The mixture of Sp1III5A -pGL3 recombinant and either empty pCTat or pCTat, along with the lipid-based transfection agent, were combined and allowed to incubate (RT, 20 minutes). Subsequently, 100µL of the plasmid and lipid mixture were introduced into the respective wells and incubated (at 37°C, 5% CO_2, 12 hours). Following this 12-hour period, the cells were rinsed twice with 5mL of 0.1M PBS to eliminate the lipid complexes. After each rinse, the cells were spun at 15000rpm for 10 minutes and the

supernatant was removed. The pellet was then re-suspended in 500µL of R10 media and incubated (37°C, 5% CO_2, 12 hours).

3.4.3. Luciferase Assay in Jurkat cells

After incubation of Jurkat cells with DNA (Section 3.4.2), the luciferase assay was carried out according to the guidelines provided by the manufacturer, using the Bright-Glo Luciferase Assay System (Promega, Madison, United States). To begin, the entire contents from each well were individually moved into a 15mL tube and then subjected to centrifugation at 15000rpm for a duration of 10 minutes. The supernatant was removed, leaving approximately 150µL, which was thoroughly mixed with a pipette. Following this, 100µL of this mixture, along with 100µL of Bright-Glo (E2610, Promega), was dispensed into the respective wells of the 24-well plate. The plate was then covered with foil and incubated at room temperature for 2 minutes. Thereafter, 150µL of the mixture was transferred into a black 96-well plate and subsequently analyzed using the Victor Nivo Multimode plate reader (PerkinElmer, Massachusetts, USA).

3.4.4. Transfection into astrocyte (SVG) cells

To assess the transcription activity of the Sp1III5A, the HIV-1 LTR Sp1III5A-pGL3 recombinant plasmid was transfected into SVG cells using a previously optimized in-house protocol. Briefly, the SVG cells were seeded into a 24-well tissue culture plate at a density of 400 000 cells per well in 1.5mL of antibiotic-free DMEM and incubated (37°C, 5%CO_2, 48 hours) to reach 85% confluency. To enhance the effectiveness of transfection, certain measures were implemented in these trials. Initially, a cell sample solely comprising SVG cells was included as a control. For the basal and tat-mediated transcription, empty pCTat (vector without tat) and pCTat (expression vector) were used, respectively. These tests were conducted three times to ensure consistency.

In preparation for the transfection mixture, 4µL of Sp1III5A -pGL3 recombinant (300ng) and 4µL of either empty pCTat (100ng) or pCTat (100ng) were blended with 248µL of serum-free RPMI for each well. The lipid-based transfection agent was created by combining 10µL of polyethyleneimine (PEI) and 240µL of serum-free RPMI for each well and left at room temperature (RT) for 5 minutes. The mixture of Sp1III5A -pGL3 recombinant and either empty pCTat or pCTat, along with the lipid-based transfection agent, were combined and allowed to incubate (RT, 30 minutes). Subsequently, 500µL of the plasmid and lipid mixture were introduced into the respective wells and incubated for 6 hours. Following incubation, the

plasmid lipid mixture was removed and replaced with fresh 1.5mL cell media without antibiotics and incubated for 20-24 hours.

3.4.5. Luciferase assay in astrocytes

After incubation of SVG cells with DNA (Section 3.4.4), the luciferase assay was carried out according to the guidelines provided by the manufacturer, using the Bright-Glo Luciferase Assay System (Promega, Madison, United States). Following this, 100µL of the whole content which was thoroughly mixed with a pipette along with 100µL of Bright-Glo (E2610, Promega), was dispensed into the respective wells of the 96-well plate. The plate was then covered with foil and incubated at room temperature for 2 minutes. Thereafter, 150µL of the mixture was transferred into a black 96-well plate and subsequently analyzed using the Victor Nivo Multimode plate reader.

3.5. Western blotting

Protein expression of specificity protein transcription factor (Sp1, S9809) was determined using Western blotting assay. The cultured Jurkat and astrocyte cells were separately lysed in a cocktail of protease and phosphatase inhibitors supplemented with Cytobuster™ protein extraction Reagent (#71009-3, Sigma-Aldrich, St. Louis, Missouri, USA). The cells were stored on ice for 30 minutes before being scraped mechanically. Cellular lysates were centrifuged (10000xg, 10 minutes, 4°C, using Eppendorf, Hamburg, Germany) and the supernatants with the crude protein extract were aspirated into fresh 1.5ml micro-centrifuge tubes and kept on ice.

The crude protein was quantified using the bicinchoninic acid (BCA) assay. The astrocyte and Jurkat proteins were then standardized to concentrations of 1.166mg/mL and 0.854mg/mL, respectively and the standardized samples were produced in Laemmli buffer [dH_2O, 0.5MTris–HCl (pH 6.8), glycerol, 10%SDS, 5%beta-mercaptoethanol, and 1%bromophenol blue] by boiling for 5minutes at 100°C. Using the Bio-Rad compact power supply, the protein samples and the molecular weight marker were electrophoresed (1h, 150V) in sodium dodecyl sulphate polyacrylamide gels (10%resolving gel and 4%stacking gel) and transferred to a nitrocellulose membrane using the Bio-Rad Transblot® Turbo™ Transfer System ((Bio-Rad, Hercules, California, USA). The membranes were blocked for 1 hour (RT) with gentle shaking in 1:5 000 (5%BSA/non-fat dry milk (NFDM) in Tris-buffered saline with 0.05%Tween 20 (TTBS; 150mMNaCl, 3mMKCl, 25mMTris). The membranes were probed with primary antibody

(anti-Sp1); 1:1,000 in 5%BSA/TTBS (Cell Signalling Technology, Danvers, Massachusetts, USA) for overnight (4°C).

Membranes were rinsed five times in TTBS for five times (10 minutes each) before being incubated with the secondary antibody for 1 hour. The Clarity Western ECL Substrate Kit (Bio-Rad, Hercules, California, USA) was used to visualize protein bands and were detected using the Molecular Imager® ChemiDoc^{TM} XRS Bio-Red Imaging System (Hercules, California, USA). The membranes were quenched with 5%H_2O_2 (30 minutes, 37°C), washed three times with TTBS (10 minutes, RT), blocked in 5%BSA in TTBS (1h, RT), and re-probed with HRP-conjugated secondary antibody for the house-keeping protein, β-actin (Sigma Aldrich, St. Louis, Missouri, USA) after detection (1:5,000 dilutions in 5%BSA, 30 minutes, RT). Images were captured and analyzed by dividing the band intensity of each sample by the respective loading controls to determine relative band density (RBD).

3.6. Statistical analysis

Statistical analysis for the transcription activity and western blotting data was performed using GraphPad Prism 9 software for Windows (GraphPad Software, San Diego California USA). Statistical significance was assessed using the unpaired t-test with Welch's correction (data expressed as mean ± standard deviation (SD)) or the one-way analysis of variance (ANOVA) in association with the Bonferroni test for multiple group comparison. The data obtained was deemed statistically significant with a p-value < 0.05.

3.7. Computational Methodology

The molecular docking and dynamics simulation was conducted to access the binding affinity of the Sp1 transcription factor to HIV-1C LTR Sp1IIIT5A and HIV-1C wild type.

3.7.1. Modeling and Validation of the LTR and Protein (Specificity protein 1)

Owing to the absence of a resolved crystal structure for the LTR (nucleotide) and the specificity protein 1, the nucleotide was modeled using two different webservers: model.it® Server (http://pongor.itk.ppke.hu/dna/model_it.html#/modelit_form) (Vlahovičcek & Pongor, 2000) and Supercomputing Facility for Bioinformatics & Computational Biology webserver (http://www.scfbio-iitd.res.in/software/drugdesign/bdna.jsp#). The web servers are used to model nucleotide sequences (wild type and mutant) that integrate computational biology and

structural modeling. This is done by inputting a nucleotide sequence of the DNA in the correct format (A, T, C, G) and then selecting parameters like DNA type and structural conformation. The server generates a 3D model, often allowing for energy minimization and refinement to ensure stability. The structure obtained from both servers shows a similar orientation (visually and sequentially) (shown in Figure S1 in the supplementary material). The protein structure was modeled using the amino acids sequence generated from the Genbank database (accession code: ACR22508) (https://www.ncbi.nlm.nih.gov/genbank/)(Sayers *et al*., 2022). The accession code corresponded to an Alphafold-predicted protein structure on the UniProt database (Bateman et al., 2023; Coudert et al., 2023) and AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/)(Jumper *et al*., 2021; Varadi et al., 2024) with the access code C4PGM0 shown in Figure 13A (cyan). The Protein was further modeled (homology modeling) using the SWISS-MODEL (https://swissmodel.expasy.org/) (Waterhouse *et al*., 2018). The homology model of Specificity protein 1 demonstrated an excellent sequence coverage of 100% and sequence similarity of 96.66%, with the highest sequence identity being 94.95%. The model quality was affirmed by GMQE and QMEANDisCo global model evaluation scores of 0.51 and 0.39, respectively, suggesting a highly reliable structure. The predicted 3D structure showed predominant regions of high and a few regions of low confidence, indicated by blue and red colors, respectively (Figure 13B). Additionally, the QMEAN Z-score of -9.28, which is below to zero, further indicated the stability and accuracy of the model (Figure 13C). The spatial distribution of the modeled structure was consistent with that of a non-redundant set of PDB structures, reflecting a good local similarity to the target (Figures 13D and 13E). The Ramachandran plot analysis further supported the model high quality, showing a MolProbity score of 1.65, a clash score of 0.23, and 82.83% of residues in the favored regions with 6.06% outliers. Moreover, there were only 3.02% rotamer outliers, 9 C-Beta deviations, 4 bad bonds out of 2227, and 56 bad angles out of 3009 (Figure 13F) (Chen *et al*., 2010). These findings collectively underscore the robustness and reliability of the homology model for Specificity protein 1. A sequence alignment analysis was performed on the two structures (the AlphaFold predicted structure (C4PGM0) and the homology modeled structure) using CLUSTAL O (1.2.4) multiple sequence alignment (Sievers & Higgins, 2018) which indicates that both structures have similar sequence orientation (Figure S2). The two structures were analyzed visually using Chimera (Pettersen *et al*., 2004) and Discovery Studio (Yue-Dong & Jing-Fei, 2010) and it was observed that the homology result has a well-oriented secondary structure compared to the AlphaFold prediction. The homology model was selected for further analysis.

3.7.2. Molecular Docking Calculation

The docking calculations were done using three different web servers namely; HDOCK (http://hdock.phys.hust.edu.cn/) (Yan *et al.*, 2020), HADDOCK (https://rascar.science.uu.nl/haddock2.4/) (Honorato *et al.*, 2024), and pyDockDNA (https://model3dbio.csic.es/pydockdna) (Rodríguez-Lumbreras *et al.*, 2022) webservers. HDOCK is particularly useful for protein-DNA docking due to its efficient combination of template-based modeling and free docking, allowing it to explore a wide range of potential binding poses. This versatility can help identify multiple viable interactions, which is important when dealing with complex biomolecular systems. HADDOCK, on the other hand, emphasizes the incorporation of experimental data into the docking process. If you have any structural or biochemical data about the protein-DNA interaction, HADDOCK can utilize that information to refine the docking results. This data-driven approach often leads to more accurate and reliable models, particularly in scenarios where you have specific binding site information. pyDockDNA is tailored specifically for DNA and RNA interactions, focusing on the unique characteristics of nucleic acids. Its algorithms take into account the distinct binding geometries and electrostatic properties involved in DNA-protein interactions, which can enhance the accuracy of the docking predictions. Using these three web servers, gives insight for comparison and validate your docking results, ensuring good complexity of the protein-DNA (SP1-LTR) interactions. The structure of the protein and the LTR (both mutant and wild type) were supplied to the webservers in PDB format and the results were visualized using Chimera and Discovery studio.

3.7.3. Molecular Dynamics Simulation

The Amber 18 PMEMD engine GPU version was used to run a 100 ns molecular dynamics (MD) simulation on SP1 complexes of the LTR wild and mutant forms (Zhang & Lu, 2023). The Amber18 Leap module neutralized the complexes, and an orthorhombic TIP3P water box was used for simulation. Using the conjugate gradient algorithm, energy minimization in 10,000 iterations with a 500 kcal/mol^2 constraint (Case et al., 2023). The complexes were heated progressively over 10 ps from 0 to 300 K using static atoms and volumes to determine the heating process. The Berendsen barostat mechanism, which had a harmonic restriction of 10 kcal/mol·Å^2 and a collision frequency of 1.0 ps, kept the system pressure at 1 bar. The operating temperature (300° K) was maintained during the 10 ns, 5,000,000 steps systems equilibration. An isobaric isothermal ensemble (NPT) was maintained by maintaining a constant number of atoms and pressure (Santos *et al.*, 2019). To keep the pressure at 1 bar, the

Berendsen barostat was used. Hydrogen atom bonds were created using the SHAKE algorithm. To examine stability, the AMBER18 GPU PTRAJ module computes the root mean square deviation (RMSD) after analyzing the coordinates and trajectories every 1 ps (Salomon-Ferrer *et al*., 2013).

CHAPTER 4: RESULTS

4.1. Introduction of Sp1III5A mutation into the HIV-1 subtype C LTR

The previous study conducted by our group demonstrated that certain elements, namely RBE III, TATA Box, NF-κB I, NF-κB II, Sp1I, Sp1II, and the E-Box TFBS, remained relatively unchanged within the LTR regions (Madlala *et al.*, 2023). In this study, the HIVC LTR Sp1IIIT5A gene block was inserted into the pGL3 vector, as outlined in Section 3. Colony PCR analysis confirmed the successful presence of the LTR (Figure 10A), as evidenced by a 640 bp band corresponding to its expected length. Consistent with previous reports (Nonnemacher *et al.*, 2004, Shah *et al.*, 2014a), our group observed noticeable variation within the Sp1III motif displaying a higher level of genetic variability at positions 2 and 5, involving adenine (A) and thymine (T) mutations in the LTR variants (Madlala *et al.*, 2023).

Figure 10A: **The colony PCR confirming the insertion of LTR.** The LTR is 640bp long and the PCR results obtained showed bands at 640bp when compared to 1Kb ladder/marker. This indicates that the insertion of LTR was successful.

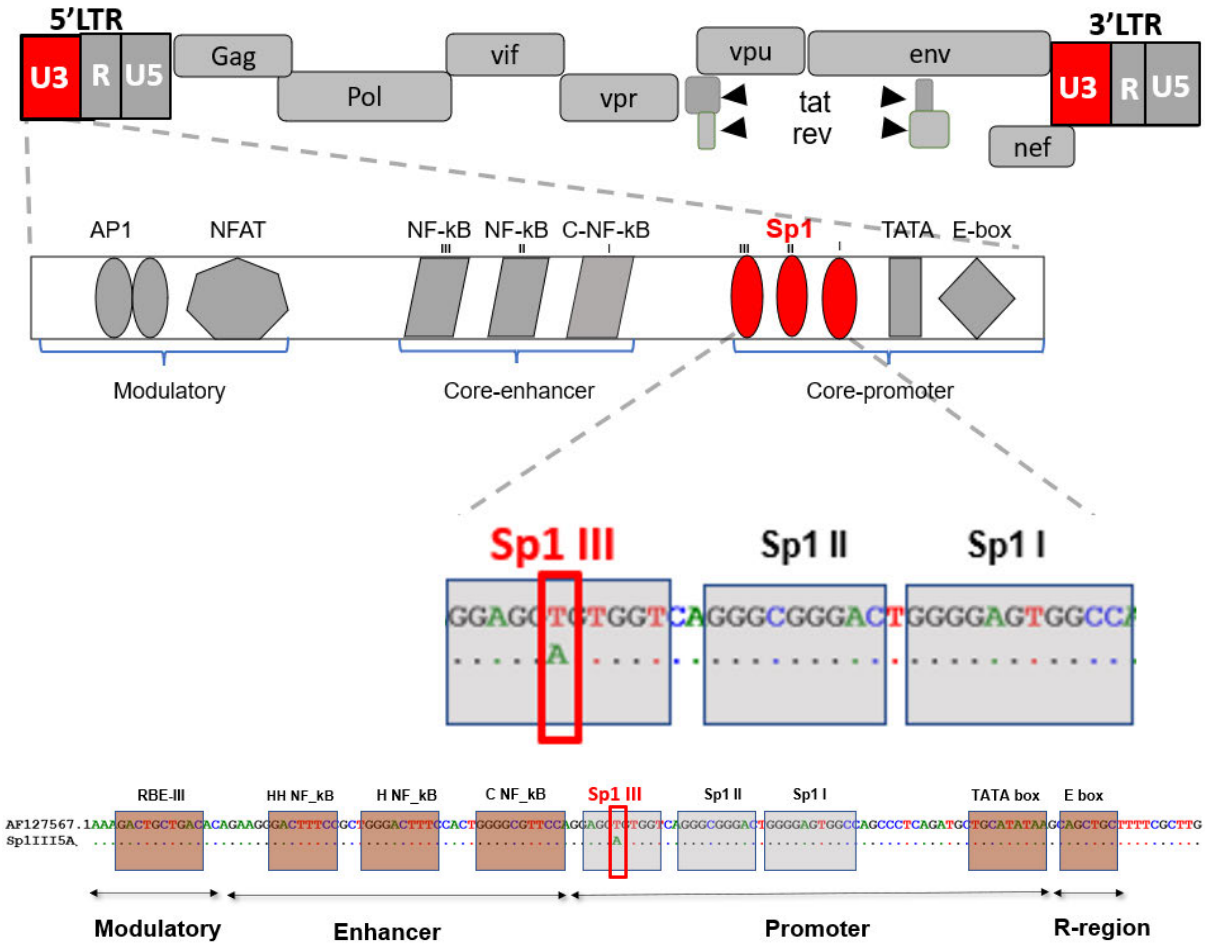


Figure 8B: Sequence alignment of CSF-derived HIV-1C LTR with A mutation at position 5 of Sp1 III site (Sp1III5A). The LTR consists of modulatory, core-enhancer and core-promoter regions. The CSF-derived HIV-1C LTR with Sp1III5A was aligned with a Zambian subtype C reference (AF127567.1). The A mutation was successfully introduced at position 5 of the Sp1 III site located in the core-promoter region.

We hypothesize that the Sp1 IIIT5A mutation may mediate HIV-1C LTR transcription activity and the ability of the Sp1 transcription factor to bind to Sp1 IIIT5A mutated motif. To address this hypothesis, we Synthesized the A mutation at position 5 of Sp1 site III HIV-1 subtype C LTR, cloned this mutated LTR into the pGL3 plasmid and this was confirmed through PCR (Figure 10A) and sequenced it to confirm the presence of this mutation. Our data show that the adenine mutation was successfully synthesized at position 5 of HIV-1 LTR Sp1 III binding site (see Figure 10B).

4.2. Transfection of HIV-C LTR sequence exhibiting Sp1III5A and/or Sp1III5T into Astrocytes and Jurkat cells

As outlined in the literature review, the TFBS within the U3 region of the HIV-1C 5'LTR plays a role in regulating both the basal and trans-activated transcription of the HIV-1 genome (Gray *et al.*, 2013). Specifically, the Sp1 motif polymorphisms are known to reduce both basal and Tat-activated transcription in CNS-derived HIV-1C 5'LTR sequences relative to lymphoid-derived HIV-1C 5'LTRs of people living with HIV-1 (Gray *et al*. 2015) which is why astrocytes and Jurkat cells were chosen for this particular study. Previous studies showed that C-to-T (Shah *et al.*, 2014b) and T-to-A single nucleotide polymorphism (SNP) at position 5 of the Sp1III motif affect transcription activity, in addition to other mutations. However, it was not known whether it was a single nucleotide mutation or a combination of mutations that impacted the transcription activity. This study hypothesized that Sp1 IIIT5A mutation may mediate HIV-1C LTR transcription activity. To address this hypothesis, a transfection assay was performed to assess the transcription activity of HIV-1C LTR exhibiting Sp1III5A compared to Sp1III5T in SVG (red) and Jurkat cells (blue) (Figure 11).

Our data show that basal transcription activity of HIV-1C LTR sequence exhibiting Sp1III5A (p<0.00001) was significantly higher compared to Sp1III5T in SVG (Figure 11A). Consistently, Tat-mediated transcription activity of HIV-1C LTR exhibiting Sp1III5A was significantly induced in SVG cells compared to HIV-1C LTR exhibiting Sp1III5T (p<0.00001)) (Figure 11B).

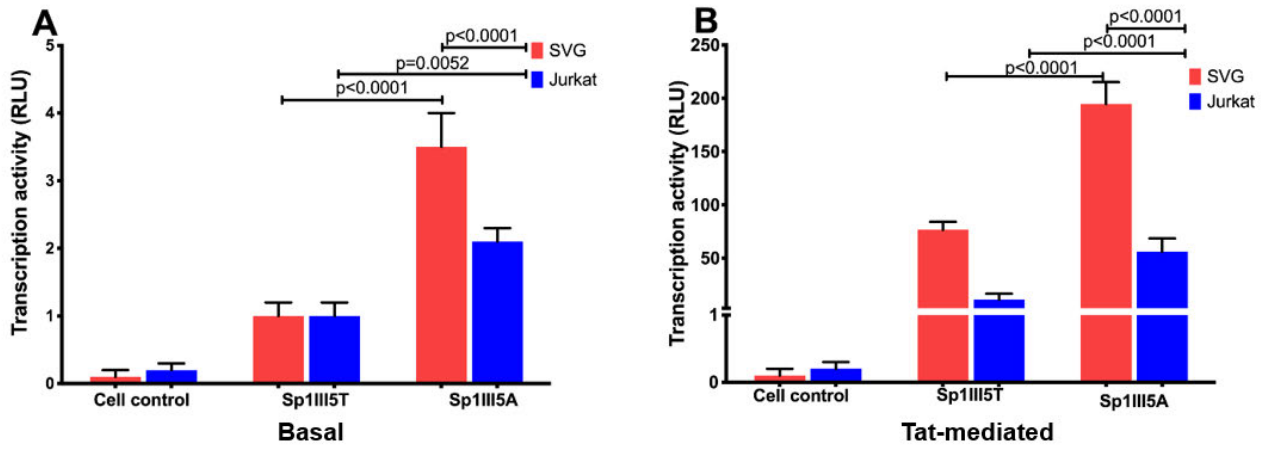


Figure 9: The transcription activity of HIV-1 subtype C LTR in Astrocytes and Jurkat cell lines. (A) Basal transcription activity of the HIV-1C LTR: Sp1III5A (p<0.00001) exhibited a significantly higher transcription activity than Sp1III5T (p=0.00521) under basal conditions in SVGs (red) and Jurkat cell line (blue). **(B) Tat-mediated transcription activity of the HIV-1C LTR sequences:** Similarly,

Sp1III5A enhanced Tat-mediated transcription activity in both SVGs (red; $p<0.00001$) and Jurkat cell line (blue; $p<0.00001$). Notable, Sp1III5A exhibited significantly higher transcription activity than Sp1III5T ($p<0.00001$) in both SVG (red) and Jurkat cells (blue), as determined by the Chi-square test.

Although HIV-1 infects a various cell type in the host organism, CD4+ T cells are preferentially targeted (reviewed in (Vanegas-Torres and Schindler, 2024)). Next, we undertook to investigate the transcription activity of HIV-1C LTR exhibiting Sp1III5A in the Jurkat T cell line, which is an immortalized line of human T lymphocytes compared to HIV-1C LTR exhibiting Sp1III5T. Our data indicate that the HIV-1C LTR containing the Sp1III5A mutation exhibited a significant increase in basal transcriptional activity compared to the LTR with the Sp1III5T in Jurkat cells ($p=0.00521$) (Figure 11A). Similarly, the HIV-1C LTR with the Sp1III5A mutation demonstrated a significant elevation in Tat-mediated transcriptional activity relative to the Sp1III5T variant in Jurkat cells ($p<0.00001$) (Figure 11B). Notably, the Sp1III5A mutation resulted in significantly enhanced basal and Tat-mediated transcriptional activity in both SVG and Jurkat cells ($p<0.00001$), However, the transcription activity was more pronounced in SVGs.

4.3. Expression levels of Specificity protein 1 (Sp1) transcription factor in astrocytes and Jurkat cells

The HIV-1 LTR Sp1 III5A variant displayed markedly greater transcription activity compared to HIV-1 LTR exhibiting Sp1 III5T (Figure 11). Notably, SVG cells were associated with high transcription activity compared to Jurkat cells. To further explore these differences, we assessed Sp1 transcription factor expression levels in SVG cells and Jurkat cells through western blotting. The expression of Sp1 in SVG cells did not show a significant increase compared to Jurkat cells ($p=0.0814$) (Figure 12). However, we observed a trend that may correspond to the enhanced transcriptional activity in SVG cells.

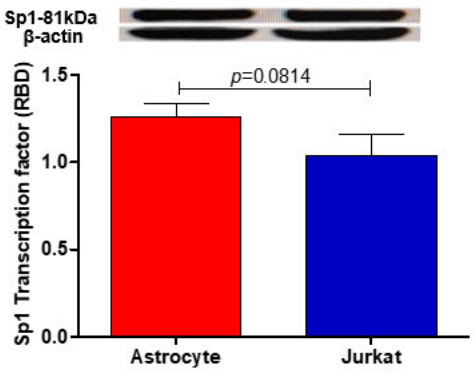


Figure 10: The expression levels of Sp1 transcription factor in astrocyte and Jurkat cells. The Sp1 expression levels in SVG and Jurkat cells showed no significant difference (p=0.0814).

4.4. Molecular Docking and Dynamics Simulation

4.4.1. Modeling and Validation of the LTR and Protein (Specificity protein 1)

Owing to the absence of a resolved crystal structure for the LTR (nucleotide) and the specificity protein 1, the nucleotide was modeled using two different webservers (Figure 13): model.it® Server (http://pongor.itk.ppke.hu/dna/model_it.html#/modelit_form) (Vlahoviček & Pongor, 2000) and Supercomputing Facility for Bioinformatics & Computational Biology webserver (http://www.scfbio-iitd.res.in/software/drugdesign/bdna.jsp#), as previously described in section 3.7.1. The findings collectively underscore the robustness and reliability of the homology model for Specificity protein 1. A sequence alignment analysis was performed on the two structures (the AlphaFold predicted structure (C4PGM0) and the homology modeled structure) using CLUSTAL O (1.2.4) multiple sequence alignment (Sievers & Higgins, 2018) which indicates that both structures have similar sequence orientation (Figure S2). The two structures were analyzed visually using Chimera (Pettersen et al., 2004) and Discovery Studio (Yue-Dong & Jing-Fei, 2010) and it was observed that the homology result has a well-oriented secondary structure compared to the AlphaFold prediction. The homology model was selected for further analysis.

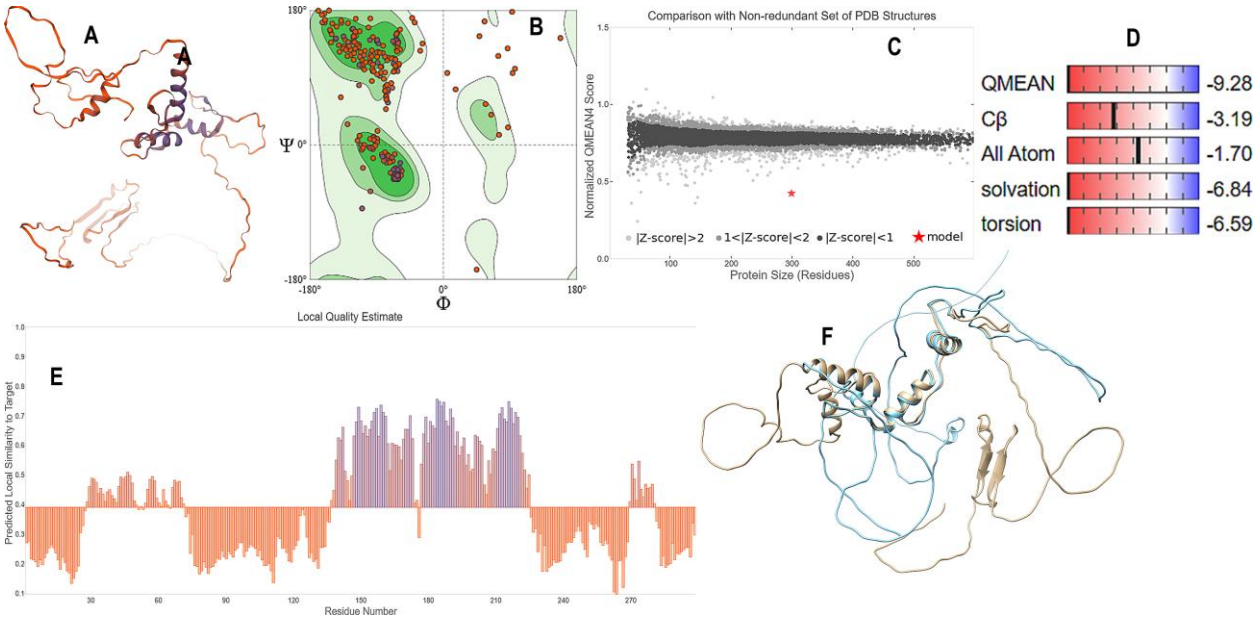


Figure 13: Validation of homology model from SWISS-MODEL web server showing the computed three-dimensional (3D) homology model colored by the confidence gradient scale (A), The Ramachandran plot analysis of the modeled structure (B), Graph of Comparison with a non-redundant set of PDB structures (C), Global quality estimate scores (D), Local quality estimate graph (E), and Alignment analysis of C4PGM0 and model structure (F).

4.4.2. Docking Calculations

The molecular docking and dynamics simulation were performed to evaluate the interaction of Sp1 and LTR. The nucleotide (LTR) was modeled using two different webservers: model.it® Server (http://pongor.itk.ppke.hu/dna/model_it.html#/modelit_form) (Vlahovičcek & Pongor, 2000) and Supercomputing Facility for Bioinformatics & Computational Biology webserver (http://www.scfbio-iitd.res.in/software/drugdesign/bdna.jsp#) as shown in the Figure S3. Both web servers use different algorithms in modeling nucleotides but the structures obtained from both servers show a similar orientation (visually and sequentially). The LTR model with Supercomputing Facility for Bioinformatics & Computational Biology webserver and model.it® Server were represented with cyan helix and brown helix respectively.

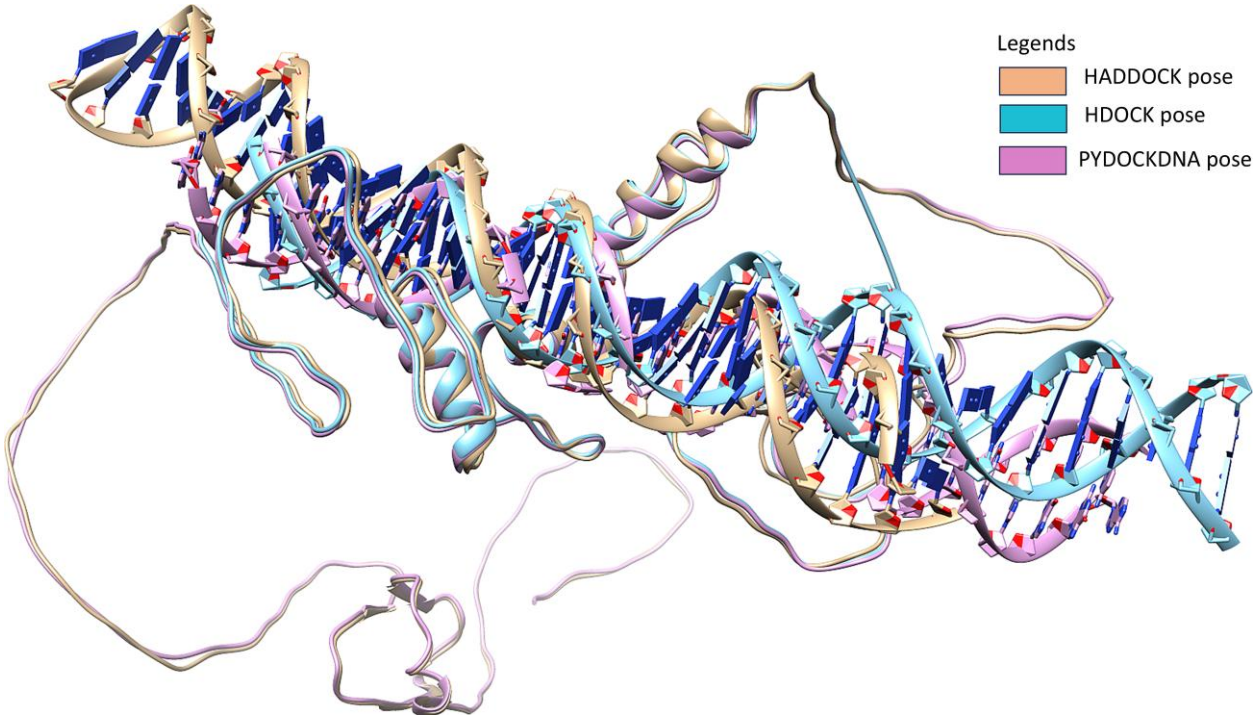


Figure S3: Alignment of the docking pose of the best pose of the Sp1-LTR complexes from three web servers.

The docking calculations aim to predict how a protein interacts with the nucleotides (LTR) by simulating their binding modes and identifying the most likely conformations for complex formation. The results (Table 1) typically provide insights into the binding affinity, binding interface, and specific amino acid-nucleotide interactions**.**

Table 1: The binding affinity of the wild and Mutant types of the LTR against the Specificity protein 1

Webserver	**Binding Affinity (kcal/mol)**	
	Mutant Type (A)	Wild Type (T)
HDOCK	-332.7	-311.4
HADDOCK	-174.6	-157.0
pyDockDNA	-279.2	-247.3

4.4.3. Molecular Dynamic Simulation

Molecular interactions, conformational flexibility, and environmental factors interact to control protein-DNA stability during molecular dynamics (MD) simulations. The integrity of the

complex is maintained by important stabilizing interactions such as van der Waals forces, hydrophobic contacts, electrostatic interactions, and hydrogen bonds. These connections can be strengthened or weakened by dynamic conformational changes in both DNA and proteins. Temperature and solvent content are two other environmental variables that affect stability. To effectively describe these dynamics, the simulation duration and force field selection are crucial. Important information about the stability and functional mechanisms behind protein-DNA interactions can be obtained by analyzing measures such as root mean square deviation (RMSD) as shown in Figure 14.

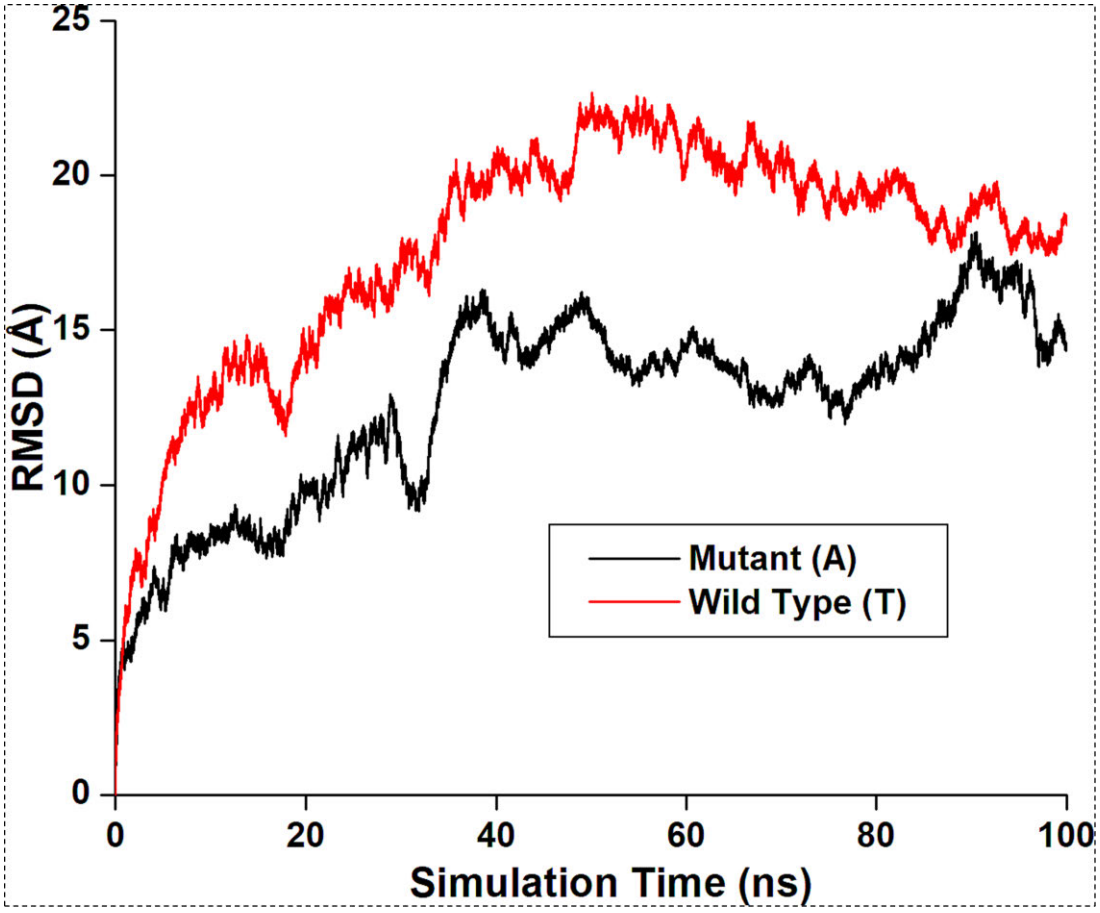


Figure 14: The Root Mean Square Deviation (RMSD) plot of the SP1 in complex with LTR Mutant and Wild type.

4.4.3. Molecular Interaction

The interactions of the Sp1 protein with both the mutant and wild-type Long Terminal Repeat (LTR) sequences were accessed (Figure 15). Protein-DNA molecular interactions are essential for the regulation of gene expression, DNA replication, repair, and overall genomic stability.

These interactions can be sequence-specific, where proteins such as transcription factors recognize specific DNA base pairs, or non-specific, where proteins interact primarily with the DNA backbone. Key forces, including hydrogen bonds, electrostatic forces, and van der Waals interactions, help stabilize these complexes. The table (Table 2) below highlights the active residue of Sp1 protein interacting with the mutant and the wild type of LTR.

Table 2: Interaction of the active site residue of SP1 with the wild type and Mutant

Protein Residues	**Wild type (T)**	**Mutant(A)**
LYS153	Charge-Charge	N/A
HIE156	Conventional hydrogen bond	N/A
ARG168	Salt-bridge, Charge-charge	Salt-bridge, Charge-charge
LYS179	Charge-charge	Charge-charge
THR182	Carbon hydrogen bond and Conventional hydrogen bond	N/A
ARG183	Conventional hydrogen bond	Conventional hydrogen bond and Pi-Cation
ASP185	Conventional hydrogen bond	N/A
GLU186	N/A	Carbon hydrogen bond
ARG189	N/A	Conventional hydrogen bond
LYS207	N/A	Salt-bridge, Charge-charge
ARG211	N/A	Conventional hydrogen bond
HIE214	Conventional hydrogen bond	Carbon hydrogen bond and Pi-Pi Stacked
LYS217	N/A	Pi-Alkyl
ARG275	Charge-charge	N/A
GLY298	Carbon hydrogen bond,	N/A

Charge-charge interactions, such as those involving residues like LYS153, LYS179, and ARG168, are crucial for establishing strong electrostatic attractions between the positively charged protein residues and negatively charged regions of the DNA, significantly affecting the initial binding affinity (Oduro-Kwateng *et al.*, 2024). The presence of conventional hydrogen bonds involving residues such as HIE156, ARG183, and ASP185 indicates specific

and directional interactions that help stabilize the protein-DNA complex. Notably, ARG183 shows a dual role in the mutant with both a conventional hydrogen bond and a pi-cation interaction, suggesting enhanced stability and specificity compared to the wild type. The stability of the protein-DNA complex is facilitated by salt bridges, which are established by residues such as ARG168 and LYS207 and offer robust interactions crucial to preserving structural integrity (Oduro-Kwateng *et al.*, 2024). A combination of stacking and hydrogen bonding interactions may favorably stabilize the mutant, increasing its overall binding effectiveness, according to differences in interaction patterns between the mutant and wild type, such as those seen with THR182, ARG189, and HIE214. The molecular interaction of the mutant LTR showed a better interaction compared to the wild type against the Sp1 protein (Figure 15).

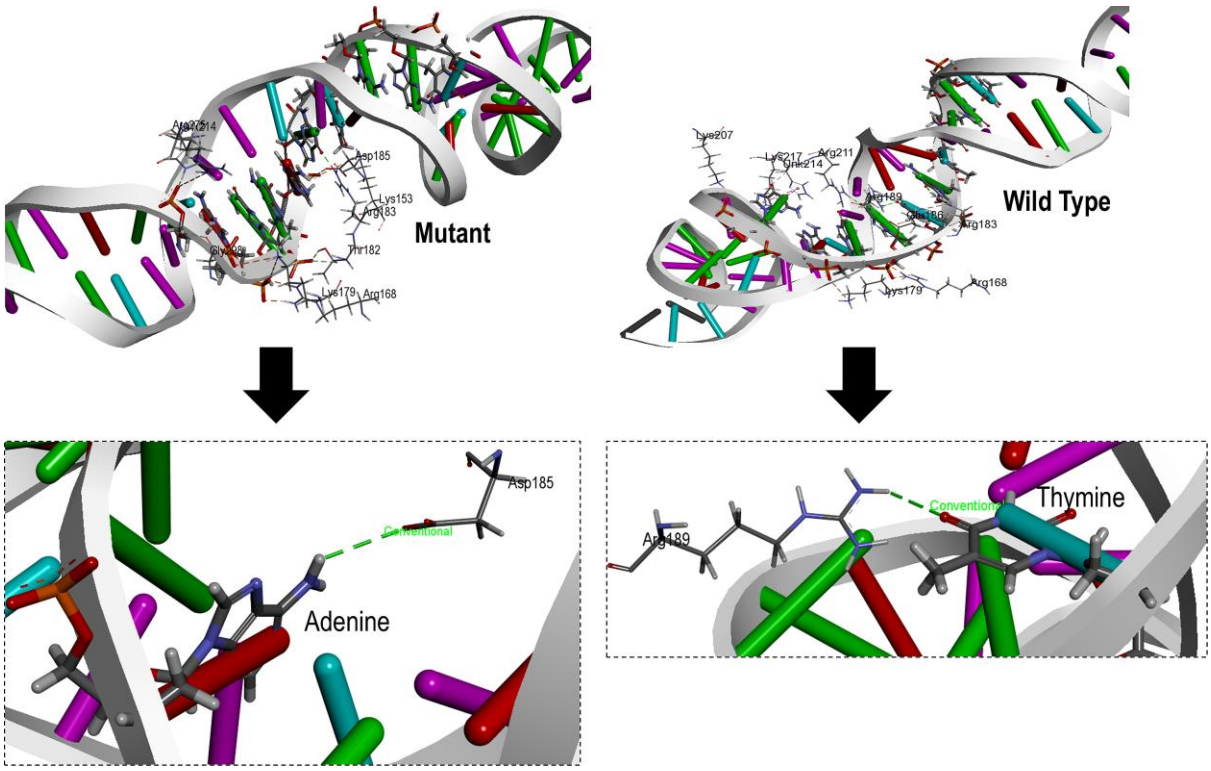


Figure 15: Interaction of LTR (mutant and wild type) with SP1 protein, displaying interaction of the mutant residue (Thymine to Adenine) at the binding site of LTR.

These docking calculations, molecular dynamics simulation, and molecular interaction all together suggest that the Adenine mutation in the LTR binding site, strengthens binding to Sp1, potentially increasing the stability or specificity of the protein-DNA complex. This could suggest the induced transcription activity in Sp1III5A compared to wildtype.

CHAPTER 5: DISCUSSION

Genetic variation within the LTR of the HIV-1 genome translates into differential functional activity and disease outcome (Madlala *et al.*, 2023). In this study, we investigated the effect of Sp1 IIIT5A mutation on the ability of Sp1 site III to bind Sp1 transcription factor and mediate HIV-1C LTR transcription activity. We hypothesize that the Sp1III5A mutation may influence the ability of Sp1 III binding site to bind Sp1 transcription factor and promote higher transcription activity and Sp1 binding affinity. The results from this study supports this hypothesis. Our data showed that the Sp1III5A mutation was successfully introduced into the HIV-1 subtype C 5'LTR sequence and was compared with the HIV-1 subtype C 5'LTR consensus sequence from Zambia (accession number: AF127567.1) which contained thymine (T) at position five of Sp1 III binding site (Sp1III5T). As outlined in the literature review, the TFBS within the U3 region of the HIV-1C 5'LTR plays a role in regulating both the basal and trans-activated transcription of the HIV-1 genome (Gray *et al.*, 2013). The Sp1 motif polymorphisms are known to reduce both basal and Tat-activated transcription in CNS-derived HIV-1C 5'LTR sequences relative to lymphoid-derived HIV-1C 5'LTRs of people living with HIV-1 (Gray *et al.*, 2013, Verma, 2015). This distinction informed the selection of SVG and Jurkat cells for the study. Notably, our data showed a significant basal and Tat-activated transcription activity induction in HIV-1C 5'LTR sequences. Of particular interest, the Sp1III5A mutation significantly enhanced basal transcription activity in SVG cells compared to the canonical Sp1III5T motif. This indicates a robust transcriptional response to the Sp1III5A mutation. Conversely, Jurkat cells exhibited a significantly increased basal transcription activity; however, the effect is less pronounced than in SVG cells. Similarly, the Sp1IIIT5A mutation enhanced Tat-mediated HIV-1C LTR transcription activity in both cell lines. In SVG cells, the enhancement was significant, indicating a strong response to Tat in conjunction with the Sp1III5A mutation. Likewise, Jurkat cells showed a significantly increased Tat-mediated transcription activity with the Sp1III5A mutation. This suggests that the Sp1IIIT5A mutation may confer a more pronounced advantage in CNS-derived cells compared to lymphoid-derived cells, highlighting the importance of cellular context in HIV-1 transcriptional regulation. This further reinforces the notion that Sp1III5A enhances HIV-1C LTR transcription activity. This contrasts with earlier studies showing that individual mutations in Sp1 binding sites generally lead to minimal changes in basal and Tat-induced transcriptional activation (Harrich *et al.*, 1989). Moreover, this research indicated that simultaneous mutations in multiple Sp1 binding sites (e.g., I and II, I and III, or II and III) also result in only minor reductions in transcriptional

activity (Harrich *et al.*, 1989). It has also been revealed that the depletion of Sp1 core promoter led to suppressed viral gene expression (Qu *et al.*, 2016). The enhanced activity observed with the Sp1III5A mutation suggests that targeted modifications to Sp1 binding sites may influence viral gene expression and replication in a manner that warrants further exploration.

We further hypothesized that the enhanced transcriptional activity associated with the Sp1III5A variant might be linked to the expression levels of the Sp1 transcription factor. Our data demonstrated that the expression levels of the Sp1 transcription factor were comparable between SVG and Jurkat cell lines, suggesting that differences in transcriptional activity observed are not attributable to variations in Sp1 expression levels. Our findings, in conjunction with previous research demonstrating Sp1's ability to autoregulate its transcription by binding to its promoter and repressing gene expression within a chromatin context (Deniaud *et al.*, 2009), offer important insights into the nuanced role of Sp1 in transcriptional regulation. This implies that Sp1III5A mutation might influence transcriptional regulation through other mechanisms, such as enhanced binding affinity to specific LTR sequences. We then hypothesized that the induced transcription activity by the mutant, Sp1III5A compared to wildtype may be due to the binding affinity between Sp1 transcription factor and HIV-1C LTR DNA. The docking calculation provided distinctions in the binding affinities of wild-type (T) and mutant (A) forms of an LTR sequence when interacting with Specificity protein 1 (Sp1), assessed using three web servers: HDOCK, HADDOCK, and pyDockDNA. Across all three platforms, the mutant type shows consistently stronger binding affinities than the wild type, as indicated by the more negative values, which reflect a more favorable interaction. Specifically, HDOCK reports the highest difference, with a binding affinity of -332.7 kcal/mol for the mutant versus -311.4 kcal/mol for the wild type, demonstrating a 21.3 kcal/mol stronger interaction in favor of the mutant. HADDOCK also supports this trend, showing affinities of -174.6 kcal/mol for the mutant and -157.0 kcal/mol for the wild type, with a difference of 17.6 kcal/mol. Finally, pyDockDNA indicates a similar result with affinities of -279.2 kcal/mol for the mutant and -247.3 kcal/mol for the wild type, yielding a 31.9 kcal/mol difference. These values together suggest that the Adenine mutation in the LTR binding site, strengthens binding to Sp1, potentially increasing the stability or specificity of the protein-DNA complex. The stronger binding affinity observed for the mutant type could have biological significance, potentially influencing gene regulation by enhancing protein-DNA binding interactions. Differences in binding affinity values across web servers may be attributed to the unique

algorithms and docking approaches each platform employs, but the consistent trend of enhanced binding for the mutant type supports the conclusion that this mutation favorably impacts binding strength. The docking pose of the three web servers was examined and an alignment analysis was performed which showed a very low deviation (less than 0.7 Å (RMSD)) as shown in Figure S3.

The stability of the SP1 protein interaction with LTR in molecular dynamics (MD) simulations was quantitatively evaluated using metrics such as root mean square deviation (RMSD) (Figure 13). RMSD values of 13.02608 Å for the mutant LTR and 17.96259 Å for the wild-type LTR indicate that the stability of the two complexes and structural dynamics differ (Sargsyan *et al.*, 2017). Throughout the simulation, the mutant LTR appears to maintain a more stable conformation, as indicated by its reduced RMSD value of 13.03 Å. This suggested that the mutant LTR and the Sp1 protein interact more favorably, resulting in a tighter binding environment and fewer conformational alterations. Such stability contributes to the general robustness of the mutant complex and could result from improved specific interactions or fewer destabilizing influences. These differences suggest functional implications since the mutant LTR may facilitate more efficient transcriptional activation or repression by the SP1 protein due to its wider variety of stabilizing contacts. On the other hand, the wildtype changed interaction profile can result in a lower binding affinity, which would then affect how transcripts are regulated. In contrast, the higher RMSD value for the wild-type LTR (17.96 Å) indicates greater conformational variability during the simulation. This increased flexibility could suggest that the wild-type LTR undergoes more significant structural rearrangements, which reflect a less stable interaction with the Sp1 protein. This variability can result from multiple factors, such as weaker interactions, the presence of alternative binding modes, or inherent flexibility in the DNA structure.

The interactions highlighted in Figure 14 provide valuable insights into the binding dynamics of the Sp1 protein with both the mutant and wild-type LTR sequences. The molecular interaction of the mutant LTR showed a better interaction compared to the wild type against the Sp1 protein. A study utilizing inducible time-course ChIP-seq methods demonstrated that Sp1 exhibits different binding dynamics at target sites within the human genome (Hasegawa and Struhl, 2021). It was observed that Sp1 can reach maximal binding levels rapidly at some

sites, while at others, it displays biphasic kinetics characterized by slower binding rates. Variations in binding dynamics are shaped by chromatin states and indicate that specific sequence contexts can strongly influence Sp1's binding affinity and transcriptional effects (Hasegawa and Struhl, 2021). Another study highlighted that Sp1 can bind not only to canonical double-stranded DNA motifs but also to non-canonical structures such as G-quadruplexes with high affinity. This indicates that Sp1's binding capabilities extend beyond traditional DNA motifs, emphasizing its versatility as a transcription factor (Raiber *et al.*, 2012). The ability of Sp1 to interact with diverse DNA structures may resemble the enhanced binding observed with the Sp1IIIT5A mutation.

CHAPTER 6: CONCLUSION

This study highlights that the Sp1III5A mutation significantly enhances both basal and Tat-mediated transcriptional activity. This suggests that genetic variation within the HIV-1 subtype C LTR impacts the transcription rate. Additionally, stronger binding interactions of the Sp1IIIT5A variant with the Sp1 transcription factor may contribute to its enhanced transcriptional activity. Future research should focus on elucidating the mechanisms by which this mutation affects viral behavior, particularly in relation to latency profiles and reactivation dynamics. Understanding how the Sp1IIIT5A mutation interacts with host cellular factors involved in latency could inform potential therapeutic strategies that target these regulatory pathways to mitigate HIV-1 pathogenesis and address the challenges posed by latent reservoirs. By integrating insights from both transcriptional regulation and latency mechanisms, we can develop more effective interventions aimed at achieving long-term control or eradication of HIV-1 infection.

CHAPTER 7: LIMITATIONS

This study primarily focused on the Sp1IIIT5A mutation without exploring its prevalence or effects in other HIV-1 subtypes, such as subtype B, which is one of the most extensively studied and prevalent subtypes globally. Investigating the mutation's frequency and functional implications in diverse subtypes could offer broader insights into its role in viral pathogenesis. Although the study suggests that stronger binding interactions with the Sp1 transcription factor contribute to enhanced transcriptional activity, the precise molecular mechanisms underlying these interactions and the interaction with other LTR transcription factor binding sites remain unclear. The experiments were conducted using a T cell line representative of peripheral cells, which provided a controlled environment to study the effects of the Sp1IIIT5A mutation. However, this approach does not capture the full spectrum of cellular environments relevant to HIV-1 infection. Specifically, the lack of investigation in monocytes/macrophages, which are key targets for HIV-1 and play a critical role in viral reservoirs and immune responses, limits the study's scope. Including these cell types would have provided a more comprehensive understanding of the mutation's impact across different immune cell populations. Further research is needed to elucidate how this mutation affects viral behavior, particularly concerning latency and reactivation dynamics. Additionally, this study does not provide longitudinal data on how the Sp1III5A mutation influences transcriptional activity over time or its impact on viral load and disease progression in infected individuals. Long-term studies are necessary to understand fully its implications for HIV-1 pathogenesis.

REFERENCES

BACHU, M., YALLA, S., ASOKAN, M., VERMA, A., NEOGI, U., SHARMA, S., MURALI, R. V., MUKTHEY, A. B., BHATT, R., CHATTERJEE, S., RAJAN, R. E., CHEEDARLA, N., YADAVALLI, V. S., MAHADEVAN, A., SHANKAR, S. K., RAJAGOPALAN, N., SHET, A., SARAVANAN, S., BALAKRISHNAN, P., SOLOMON, S., VAJPAYEE, M., SATISH, K. S., KUNDU, T. K., JEANG, K.-T. & RANGA, U. 2012. Multiple NF-κB Sites in HIV-1 Subtype C Long Terminal Repeat Confer Superior Magnitude of Transcription and Thereby the Enhanced Viral Predominance. *Journal of Biological Chemistry,* 287**,** 44714-44735.

BARRÉ-SINOUSSI, F., CHERMANN, J. C., REY, F., NUGEYRE, M. T., CHAMARET, S., GRUEST, J., DAUGUET, C., AXLER-BLIN, C., VÉZINET-BRUN, F., ROUZIOUX, C., ROZENBAUM, W. & MONTAGNIER, L. 1983. Isolation of a T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS). *Science,* 220**,** 868-871.

BASS, J. J., WILKINSON, D. J., RANKIN, D., PHILLIPS, B. E., SZEWCZYK, N. J., SMITH, K. & ATHERTON, P. J. 2017. An overview of technical considerations for Western blotting applications to physiological research. *Scandinavian Journal of Medicine & Science in Sports,* 27**,** 4-25.

BATTISTINI, A. & SGARBANTI, M. 2014. HIV-1 latency: an update of molecular mechanisms and therapeutic strategies. *Viruses,* 6**,** 1715-58.

BEHJATI, S. & TARPEY, P. S. 2013. What is next generation sequencing? *Archives of disease in childhood - Education & practice edition,* 98**,** 236-238.

BOULLOSA, J., BACHU, M., BILA, D., RANGA, U., SÜFFERT, T., SASAZAWA, T. & TANURI, A. 2014. Genetic diversity in HIV-1 subtype C LTR from Brazil and Mozambique generates new transcription factor-binding sites. *Viruses,* 6**,** 2495-504.

CAPUTI, M. 2011. The regulation of HIV-1 mRNA biogenesis. *RNA processing,* 1.

CHANDE, A. G., BABA, M. & MUKHOPADHYAYA, R. 2012. Short communication: a single step assay for rapid evaluation of inhibitors targeting HIV type 1 Tat-mediated long terminal repeat transactivation. *AIDS Res Hum Retroviruses,* 28**,** 902-6.

CHEREPANOV, P., MAERTENS, G., PROOST, P., DEVREESE, B., VAN BEEUMEN, J., ENGELBORGHS, Y., DE CLERCQ, E. & DEBYSER, Z. 2003. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J Biol Chem,* 278**,** 372-81.

CHIPOT, C. 2023. Free energy methods for the description of molecular processes. *Annual Review of Biophysics,* 52**,** 113-138.

COFFIN, J. M. 1995. HIV Population Dynamics in Vivo: Implications for Genetic Variation, Pathogenesis, and Therapy. *Science,* 267**,** 483-489.

COLIN, L., VERDIN, E. & VAN LINT, C. 2014. HIV-1 chromatin, transcription, and the regulatory protein Tat. *Methods Mol Biol,* 1087**,** 85-101.

D'ORSO, I. 2024. The HIV-1 transcriptional program: from initiation to elongation control. *Journal of Molecular Biology***,** 168690.

DAS, A. T., HARWIG, A. & BERKHOUT, B. 2011. The HIV-1 Tat protein has a versatile role in activating viral transcription. *Journal of virology,* 85**,** 9506-9516.

DE BAAR, M. P., DE RONDE, A., BERKHOUT, B., CORNELISSEN, M., VAN DER HORN, K. H., VAN DER SCHOOT, A. M., DE WOLF, F., LUKASHOV, V. V. & GOUDSMIT, J. 2000. Subtype-specific sequence variation of the HIV type 1 long terminal repeat and primer-binding site. *AIDS Res Hum Retroviruses,* 16**,** 499-504.

DENIAUD, E., BAGUET, J., CHALARD, R., BLANQUIER, B., BRINZA, L., MEUNIER, J., MICHALLET, M.-C., LAUGRAUD, A., AH-SOON, C., WIERINCKX, A., CASTELLAZZI, M., LACHUER, J., GAUTIER, C., MARVEL, J. & LEVERRIER, Y.

2009. Overexpression of Transcription Factor Sp1 Leads to Gene Expression Perturbations and Cell Cycle Inhibition. *PLoS ONE,* 4**,** e7035.
DIDIGU, C. A. & DOMS, R. W. 2012. Novel Approaches to Inhibit HIV Entry. *Viruses,* 4**,** 309-324.
DUFOUR, C., GANTNER, P., FROMENTIN, R. & CHOMONT, N. 2020. The multifaceted nature of HIV latency. *Journal of Clinical Investigation,* 130**,** 3381-3390.
DUTILLEUL, A., RODARI, A. & VAN LINT, C. 2020. Depicting HIV-1 Transcriptional Mechanisms: A Summary of What We Know. *Viruses,* 12**,** 1385.
ENGELMAN, A. & CHEREPANOV, P. 2012. The structural biology of HIV-1: mechanistic and therapeutic insights. *Nat Rev Microbiol,* 10**,** 279-90.
FORERO, D. A., GONZÁLEZ-GIRALDO, Y., CASTRO-VEGA, L. J. & BARRETO, G. E. 2019. qPCR-based methods for expression analysis of miRNAs. *Biotechniques,* 67**,** 192-199.
FRANÇA, L. T., CARRILHO, E. & KIST, T. B. 2002. A review of DNA sequencing techniques. *Quarterly reviews of biophysics,* 35**,** 169-200.
FRANKEL, A. D. & YOUNG, J. A. T. 1998. HIV-1: Fifteen Proteins and an RNA. *Annual Review of Biochemistry,* 67**,** 1-25.
GANSER-PORNILLOS, B. K., YEAGER, M. & SUNDQUIST, W. I. 2008. The structural biology of HIV assembly. *Curr Opin Struct Biol,* 18**,** 203-17.
GATIGNOL, A., DUARTE, M., DAVIET, L., CHANG, Y.-N. & JEANG, K.-T. 2018. Sequential steps in Tat trans-activation of HIV-1 mediated through cellular DNA, RNA, and protein binding factors. *Gene expression,* 5**,** 217.
GAYNOR, R. 1992a. Cellular Transcription Factors Involved in the Regulation of HIV-1 Gene Expression. *AIDS,* 6.
GAYNOR, R. 1992b. Cellular transcription factors involved in the regulation of HIV-1 gene expression. *Aids,* 6**,** 347-63.
GIOVANETTI, M., CICCOZZI, M., PAROLIN, C. & BORSETTI, A. 2020. Molecular epidemiology of HIV-1 in African countries: a comprehensive overview. *Pathogens,* 9**,** 1072.
GRAY, L. R., COWLEY, D., CRESPAN, E., WELSH, C., MACKENZIE, C., WESSELINGH, S. L., GORRY, P. R. & CHURCHILL, M. J. 2013. Reduced Basal Transcriptional Activity of Central Nervous System-Derived HIV Type 1 Long Terminal Repeats. *AIDS Research and Human Retroviruses,* 29**,** 365-370.
GROEN, J. N. & MORRIS, K. V. 2013. Chromatin, non-coding RNAs, and the expression of HIV. *Viruses,* 5**,** 1633-45.
GUERMAH, M., MALIK, S. & ROEDER, R. G. 1998. Involvement of TFIID and USA components in transcriptional activation of the human immunodeficiency virus promoter by NF-kappaB and Sp1. *Mol Cell Biol,* 18**,** 3234-44.
HARRICH, D., GARCIA, J., WU, F., MITSUYASU, R., GONAZALEZ, J. & GAYNOR, R. 1989. Role of SP1-binding domains in in vivo transcriptional regulation of the human immunodeficiency virus type 1 long terminal repeat. *Journal of Virology,* 63**,** 2585-2591.
HASANI, H. J. & BARAKAT, K. 2017. Homology modeling: an overview of fundamentals and tools. *Int. Rev. Model. Simul,* 10**,** 1-14.
HASEGAWA, Y. & STRUHL, K. 2021. Different SP1 binding dynamics at individual genomic loci in human cells. *Proceedings of the National Academy of Sciences,* 118**,** e2113579118.
HEMELAAR, J., ELANGOVAN, R., YUN, J., DICKSON-TETTEH, L., FLEMINGER, I., KIRTLEY, S., WILLIAMS, B., GOUWS-WILLIAMS, E. & GHYS, P. D. 2019. Global

and regional molecular epidemiology of HIV-1, 1990-2015: a systematic review, global survey, and trend analysis. *Lancet Infect Dis,* 19**,** 143-155.

HEMELAAR, J., ELANGOVAN, R., YUN, J., DICKSON-TETTEH, L., KIRTLEY, S., GOUWS-WILLIAMS, E., GHYS, P. D., ABIMIKU, A. L. G., AGWALE, S., ARCHIBALD, C., AVIDOR, B., BARBÁS, M. G., BARRE-SINOUSSI, F., BARUGAHARE, B., BELABBES, E. H., BERTAGNOLIO, S., BIRX, D., BOBKOV, A. F., BRANDFUL, J., BREDELL, H., BRENNAN, C. A., BROOKS, J., BRUCKOVA, M., BUONAGURO, L., BUONAGURO, F., BUTTÒ, S., BUVÉ, A., CAMPBELL, M., CARR, J., CARRERA, A., CARRILLO, M. G., CELUM, C., CHAPLIN, B., CHARLES, M., CHATZIDIMITRIOU, D., CHEN, Z., CHIJIWA, K., COOPER, D., CUNNINGHAM, P., DAGNRA, A., DE GASCUN, C. F., DEL AMO, J., DELGADO, E., DIETRICH, U., DWYER, D., ELLENBERGER, D., ENSOLI, B., ESSEX, M., GAO, F., FLEURY, H., FONJUNGO, P. N., FOULONGNE, V., GADKARI, D. A., GAO, F., GARCÍA, F., GARSIA, R., GERSHY-DAMET, G. M., GLYNN, J. R., GOODALL, R., GROSSMAN, Z., LINDENMEYER-GUIMARÃES, M., HAHN, B., HAMERS, R. L., HAMOUDA, O., HANDEMA, R., HE, X., HERBECK, J., HO, D. D., HOLGUIN, A., HOSSEINIPOUR, M., HUNT, G., ITO, M., BEL HADJ KACEM, M. A., KAHLE, E., KALEEBU, P., KALISH, M., KAMARULZAMAN, A., KANG, C., KANKI, P., KARAMOV, E., KARASI, J.-C., KAYITENKORE, K., KELLEHER, T., KITAYAPORN, D., KOSTRIKIS, L. G., KUCHERER, C., LARA, C., LEITNER, T., LIITSOLA, K., LINGAPPA, J., LINKA, M., LORENZANA DE RIVERA, I., LUKASHOV, V., MAAYAN, S., MAYR, L., MCCUTCHAN, F., MEDA, N., MENU, E., MHALU, F., MLOKA, D., MOKILI, J. L., MONTES, B., MOR, O., MORGADO, M., MOSHA, F., MOUSSI, A., MULLINS, J., NAJERA, R., NASR, M., NDEMBI, N., NEILSON, J. R., NERURKAR, V. R., NEUHANN, F., NOLTE, C., NOVITSKY, V., NYAMBI, P., OFNER, M., PALADIN, F. J., PAPA, A., PAPE, J., PARKIN, N., PARRY, C., PEETERS, M., PELLETIER, A., PÉREZ-ÁLVAREZ, L., PILLAY, D., PINTO, A., QUANG, T. D., RADEMEYER, C., RAIKANIKODA, F., RAYFIELD, M. A., REYNES, J.-M., RINKE DE WIT, T., ROBBINS, K. E., ROLLAND, M., ROUSSEAU, C., SALAZAR-GONZALES, J., SALEM, H., SALMINEN, M., SALOMON, H., SANDSTROM, P., SANTIAGO, M. L., SARR, A. D., SCHROEDER, B., SEGONDY, M., SELHORST, P., SEMPALA, S., SERVAIS, J., SHAIK, A., SHAO, Y., SLIM, A., SOARES, M. A., SONGOK, E., STEWART, D., STOKES, J., SUBBARAO, S., SUTTHENT, R., TAKEHISA, J., TANURI, A., TEE, K. K., THAPA, K., THOMSON, M., TRAN, T., URASSA, W., USHIJIMA, H., VAN DE PERRE, P., VAN DER GROEN, G., VAN LAETHEM, K., VAN OOSTERHOUT, J., VAN SIGHEM, A., VAN WIJNGAERDEN, E., VANDAMME, A.-M., VERCAUTEREN, J., VIDAL, N., WALLACE, L., WILLIAMSON, C., WOLDAY, D., XU, J., YANG, C., ZHANG, L. & ZHANG, R. 2020. Global and regional epidemiology of HIV-1 recombinants in 1990–2015: a systematic review and global survey. *The Lancet HIV,* 7**,** e772-e781.

HEMELAAR, J., GOUWS, E., GHYS, P. D. & OSMANOV, S. 2006. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS,* 20**,** W13-W23.

HOEY, T., WEINZIERL, R. O., GILL, G., CHEN, J. L., DYNLACHT, B. D. & TJIAN, R. 1993. Molecular cloning and functional analysis of Drosophila TAF110 reveal properties expected of coactivators. *Cell,* 72**,** 247-60.

HURLEY, J. H. & CADA, A. K. 2018. Inside job: how the ESCRTs release HIV-1 from infected cells. *Biochemical Society Transactions,* 46**,** 1029-1036.

JANSSENS, W., SALMINEN, M. O., LAUKKANEN, T., HEYNDRICKX, L., VAN DER AUWERA, G., COLEBUNDERS, R., MCCUTCHAN, F. E. & VAN DER GROEN, G.

2000. Near Full-Length Genome Analysis of HIV Type 1 CRF02. AG, Subtype C and CRF02. AG Subtype G Recombinants. *AIDS research and human retroviruses,* 16**,** 1183-1189.

JEENINGA, R. E., HOOGENKAMP, M., ARMAND-UGON, M., BAAR, M. D., VERHOEF, K. & BERKHOUT, B. 2000. Functional Differences between the Long Terminal Repeat Transcriptional Promoters of Human Immunodeficiency Virus Type 1 Subtypes A through G. *Journal of Virology,* 74**,** 3740-3751.

JEON, K. W. 2013. International review of cell and molecular biology.

KARN, J. 1999. Tackling tat. *Journal of molecular biology,* 293**,** 235-254.

KARN, J. & STOLTZFUS, C. M. 2012. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harb Perspect Med,* 2**,** a006916.

KHANAL, S., SCHANK, M., EL GAZZAR, M., MOORMAN, J. P. & YAO, Z. Q. 2021. HIV-1 Latency and Viral Reservoirs: Existing Reversal Approaches and Potential Technologies, Targets, and Pathways Involved in HIV Latency Studies. *Cells,* 10**,** 475.

KHARYTONCHYK, S., MONTI, S., SMALDINO, P. J., VAN, V., BOLDEN, N. C., BROWN, J. D., RUSSO, E., SWANSON, C., SHUEY, A., TELESNITSKY, A. & SUMMERS, M. F. 2016. Transcriptional start site heterogeneity modulates the structure and function of the HIV-1 genome. *Proceedings of the National Academy of Sciences,* 113**,** 13378-13383.

KHRAIWESH, B. 2011. Using nuclear run-on transcription assays in RNAi studies. *RNAi and Plant Gene Function Analysis: Methods and Protocols***,** 199-209.

KIRCHHOFF, F. 2013. HIV life cycle: overview. *Encyclopedia of AIDS***,** 1-9.

KURIEN, B. T., DORRI, Y., DILLON, S., DSOUZA, A. & SCOFIELD, R. H. 2011. An overview of Western blotting for determining antibody specificities for immunohistochemistry. *Signal transduction immunohistochemistry: methods and protocols***,** 55-67.

LE HINGRAT, Q., VISSEAUX, B., BERTINE, M., CHAUVEAU, L., SCHWARTZ, O., COLLIN, F., DAMOND, F., MATHERON, S., DESCAMPS, D. & CHARPENTIER, C. 2020. Genetic Variability of Long Terminal Repeat Region between HIV-2 Groups Impacts Transcriptional Activity. *Journal of Virology,* 94.

LI, L., AIAMKITSUMRIT, B., PIRRONE, V., NONNEMACHER, M. R., WOJNO, A., PASSIC, S., FLAIG, K., KILARESKI, E., BLAKEY, B., KU, J., PARIKH, N., SHAH, R., MARTIN-GARCIA, J., MOLDOVER, B., SERVANCE, L., DOWNIE, D., LEWIS, S., JACOBSON, J. M., KOLSON, D. & WIGDAHL, B. 2011. Development of co-selected single nucleotide polymorphisms in the viral promoter precedes the onset of human immunodeficiency virus type 1-associated neurocognitive impairment. *Journal of NeuroVirology,* 17**,** 92-109.

LI, L., DAHIYA, S., KORTAGERE, S., AIAMKITSUMRIT, B., CUNNINGHAM, D., PIRRONE, V., NONNEMACHER, M. R. & WIGDAHL, B. 2012. Impact of Tat Genetic Variation on HIV-1 Disease. *Advances in Virology,* 2012**,** 1-28.

LUFINO, M. M., EDSER, P. A. & WADE-MARTINS, R. 2008. Advances in High-capacity Extrachromosomal Vector Technology: Episomal Maintenance, Vector Delivery, and Transgene Expression. *Molecular Therapy,* 16**,** 1525-1538.

MADLALA, P., MKHIZE, Z., NAICKER, S., KHATHI, S. P., MAIKOO, S., GOPEE, K., DONG, K. L. & NDUNG'U, T. 2023. Genetic variation of the HIV-1 subtype C transmitted/founder viruses long terminal repeat elements and the impact on transcription activation potential and clinical disease outcomes. *PLOS Pathogens,* 19**,** e1011194.

MAERTENS, G., CHEREPANOV, P., PLUYMERS, W., BUSSCHOTS, K., DE CLERCQ, E., DEBYSER, Z. & ENGELBORGHS, Y. 2003. LEDGF/p75 is essential for nuclear and

chromosomal targeting of HIV-1 integrase in human cells. *J Biol Chem,* 278**,** 33528-39.

MANDLIK, V. & SINGH, S. 2016. Molecular docking and molecular dynamics simulation study of inositol phosphorylceramide synthase–inhibitor complex in leishmaniasis: Insight into the structure based drug design. *F1000Research,* 5.

MILLER-JENSEN, K., SKUPSKY, R., SHAH, P. S., ARKIN, A. P. & SCHAFFER, D. V. 2013. Genetic selection for context-dependent stochastic phenotypes: Sp1 and TATA mutations increase phenotypic noise in HIV-1 gene expression. *PLoS Comput Biol,* 9**,** e1003135.

MITCHELL, R. S., BEITZEL, B. F., SCHRODER, A. R., SHINN, P., CHEN, H., BERRY, C. C., ECKER, J. R. & BUSHMAN, F. D. 2004. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol,* 2**,** E234.

MONTANO, M. A., NOVITSKY, V. A., BLACKARD, J. T., CHO, N. L., KATZENSTEIN, D. A. & ESSEX, M. 1997. Divergent transcriptional regulation among expanding human immunodeficiency virus type 1 subtypes. *J Virol,* 71**,** 8657-65.

MONTY, CHRISTIAN, NDUNG'U, T., BUSSMANN, H., VLADIMIR, DICKMAN, D. & ESSEX, M. 2000. Elevated Tumor Necrosis Factor–α Activation of Human Immunodeficiency Virus Type 1 Subtype C in Southern Africa Is Associated with an NF-κB Enhancer Gain-of-Function. *The Journal of Infectious Diseases,* 181**,** 76-81.

MORI, L. & VALENTE, S. T. 2020. Key players in HIV-1 transcriptional regulation: targets for a functional cure. *Viruses,* 12**,** 529.

MOUHAND, A., PASI, M., CATALA, M., ZARGARIAN, L., BELFETMI, A., BARRAUD, P., MAUFFRET, O. & TISNÉ, C. 2020. Overview of the Nucleic-Acid Binding Properties of the HIV-1 Nucleocapsid Protein in Its Different Maturation States. *Viruses,* 12.

MUNKANTA, M., HANDEMA, R., KASAI, H., GONDWE, C., DENG, X., YAMASHITA, A., ASAGI, T., YAMAMOTO, N., ITO, M., KASOLO, F. & TERUNUMA, H. 2005. Predominance of three NF-kappaB binding sites in the long terminal repeat region of HIV Type 1 subtype C isolates from Zambia. *AIDS Res Hum Retroviruses,* 21**,** 901-6.

NAGHAVI, M. H., SCHWARTZ, S., SÖNNERBORG, A. & VAHLNE, A. 1999. Long terminal repeat promoter/enhancer activity of different subtypes of HIV type 1. *AIDS Res Hum Retroviruses,* 15**,** 1293-303.

NAWROTH, I., MUELLER, F., BASYUK, E., BEERENS, N., RAHBEK, U. L., DARZACQ, X., BERTRAND, E., KJEMS, J. & SCHMIDT, U. 2014. Stable assembly of HIV-1 export complexes occurs cotranscriptionally. *RNA,* 20**,** 1-8.

NEJEPINSKA, J., MALIK, R., MORAVEC, M. & SVOBODA, P. 2012. Deep Sequencing Reveals Complex Spurious Transcription from Transiently Transfected Plasmids. *PLoS ONE,* 7**,** e43283.

NJAI, H. F., GALI, Y., VANHAM, G., CLYBERGH, C., JENNES, W., VIDAL, N., BUTEL, C., MPOUDI-NGOLLE, E., PEETERS, M. & ARIËN, K. K. 2006. The predominance of Human Immunodeficiency Virus type 1 (HIV-1) circulating recombinant form 02 (CRF02_AG) in West Central Africa may be related to its replicative fitness. *Retrovirology,* 3.

NONNEMACHER, M. R., IRISH, B. P., LIU, Y., MAUGER, D. & WIGDAHL, B. 2004. Specific sequence configurations of HIV-1 LTR G/C box array result in altered recruitment of Sp isoforms and correlate with disease progression. *J Neuroimmunol,* 157**,** 39-47.

ODURO-KWATENG, E., ALI, M., KEHINDE, I. O., ZHANG, Z. & SOLIMAN, M. E. 2024. De Novo Rational Design of Peptide-Based Protein–Protein Inhibitors (Pep-PPIs) Approach by Mapping the Interaction Motifs of the PP Interface and Physicochemical

Filtration: A Case on p25-Cdk5-Mediated Neurodegenerative Diseases. *Journal of Cellular Biochemistry,* 125**,** e30633.

OPIJNEN, T. V., JEENINGA, R. E., BOERLIJST, M. C., POLLAKIS, G. P., ZETTERBERG, V., SALMINEN, M. & BERKHOUT, B. 2004. Human Immunodeficiency Virus Type 1 Subtypes Have a Distinct Long Terminal Repeat That Determines the Replication Rate in a Host-Cell-Specific Manner. *Journal of Virology,* 78**,** 3675-3683.

PAPATHANASOPOULOS, M. A., CILLIERS, T., MORRIS, L., MOKILI, J. L., DOWLING, W., BIRX, D. L. & MCCUTCHAN, F. E. 2002. Full-length genome analysis of HIV-1 subtype C utilizing CXCR4 and intersubtype recombinants isolated in South Africa. *AIDS Res Hum Retroviruses,* 18**,** 879-86.

PEREIRA, L. A., BENTLEY, K., PEETERS, A., CHURCHILL, M. J. & DEACON, N. J. 2000a. A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. *Nucleic Acids Res,* 28**,** 663-8.

PEREIRA, L. A., BENTLEY, K., PEETERS, A., CHURCHILL, M. J. & DEACON, N. J. 2000b. SURVEY AND SUMMARY A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. *Nucleic acids research,* 28**,** 663-668.

PERKINS, N. D., EDWARDS, N. L., DUCKETT, C. S., AGRANOFF, A. B., SCHMID, R. M. & NABEL, G. J. 1993. A cooperative interaction between NF-kappa B and Sp1 is required for HIV-1 enhancer activation. *Embo j,* 12**,** 3551-8.

PORNILLOS, O. & GANSER-PORNILLOS, B. K. 2019. Maturation of retroviruses. *Curr Opin Virol,* 36**,** 47-55.

QU, D., LI, C., SANG, F., LI, Q., JIANG, Z.-Q., XU, L.-R., GUO, H.-J., ZHANG, C. & WANG, J.-H. 2016. The variances of Sp1 and NF-κB elements correlate with the greater capacity of Chinese HIV-1 B′-LTR for driving gene expression. *Scientific reports,* 6**,** 34532.

RAIBER, E.-A., KRANASTER, R., LAM, E., NIKAN, M. & BALASUBRAMANIAN, S. 2012. A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Research,* 40**,** 1499-1508.

RIEDL, S. A. B., KAISER, P., RAUP, A., SYNATSCHKE, C. V., JÉRÔME, V. & FREITAG, R. 2018. Non-Viral Transfection of Human T Lymphocytes. *Processes,* 6**,** 188.

ROEBUCK, K. A. & SAIFUDDIN, M. 2018. Regulation of HIV-1 transcription. *Gene expression,* 8**,** 67.

ROOF, P., RICCI, M., GENIN, P., MONTANO, M. A., ESSEX, M., WAINBERG, M. A., GATIGNOL, A. & HISCOTT, J. 2002. Differential regulation of HIV-1 clade-specific B, C, and E long terminal repeats by NF-κB and the Tat transactivator. *Virology,* 296**,** 77-83.

ROY, S., DELLING, U., CHEN, C.-H., ROSEN, C. & SONENBERG, N. 1990. A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. *Genes & development,* 4**,** 1365-1373.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences,* 74**,** 5463-5467.

SARGSYAN, K., GRAUFFEL, C. & LIM, C. 2017. How molecular size impacts RMSD applications in molecular dynamics simulations. *Journal of chemical theory and computation,* 13**,** 1518-1524.

SCHIRALLI LESTER, G. M. & HENDERSON, A. J. 2012. Mechanisms of HIV Transcriptional Regulation and Their Contribution to Latency. *Mol Biol Int,* 2012**,** 614120.

SCHRÖDER, A. R., SHINN, P., CHEN, H., BERRY, C., ECKER, J. R. & BUSHMAN, F. 2002. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell,* 110**,** 521-9.

SCHWARTZ, O., VIRELIZIER, J. L., MONTAGNIER, L. & HAZAN, U. 1990. A microtransfection method using the luciferase-encoding reporter gene for the assay of human immunodeficiency virus LTR promoter activity. *Gene,* 88**,** 197-205.

SHAH, S., ALEXAKI, A., PIRRONE, V., DAHIYA, S., NONNEMACHER, M. R. & WIGDAHL, B. 2014a. Functional properties of the HIV-1 long terminal repeat containing single-nucleotide polymorphisms in Sp site III and CCAAT/enhancer binding protein site I. *Virology Journal,* 11**,** 92.

SHAH, S., ALEXAKI, A., PIRRONE, V., DAHIYA, S., NONNEMACHER, M. R. & WIGDAHL, B. 2014b. Functional properties of the HIV-1 long terminal repeat containing single-nucleotide polymorphisms in Sp site III and CCAAT/enhancer binding protein site I. *Virol J,* 11**,** 92.

SHOKO, C. & CHIKOBVU, D. 2019. A superiority of viral load over CD4 cell count when predicting mortality in HIV patients on therapy. *BMC Infectious Diseases,* 19.

SILICIANO, R. F. & GREENE, W. C. 2011. HIV latency. *Cold Spring Harb Perspect Med,* 1**,** a007096.

SMALE, S. T. 2010. Luciferase assay. *Cold Spring Harb Protoc,* 2010**,** pdb.prot5421.

STEPANENKO, A. A. & HENG, H. H. 2017. Transient and stable vector transfection: Pitfalls, off-target effects, artifacts. *Mutation Research/Reviews in Mutation Research,* 773**,** 91-103.

TEE, K. K., PYBUS, O. G., LI, X.-J., HAN, X., SHANG, H., KAMARULZAMAN, A. & TAKEBE, Y. 2008. Temporal and Spatial Dynamics of Human Immunodeficiency Virus Type 1 Circulating Recombinant Forms 08_BC and 07_BC in Asia. *Journal of Virology,* 82**,** 9206-9215.

TURRINI, F., MARELLI, S., KAJASTE-RUDNITSKI, A., LUSIC, M., VAN LINT, C., DAS, A. T., HARWIG, A., BERKHOUT, B. & VICENZI, E. 2015. HIV-1 transcriptional silencing caused by TRIM22 inhibition of Sp1 binding to the viral promoter. *Retrovirology,* 12.

VAN HEUVEL, Y., SCHATZ, S., ROSENGARTEN, J. F. & STITZ, J. 2022. Infectious RNA: Human Immunodeficiency Virus (HIV) Biology, Therapeutic Intervention, and the Quest for a Vaccine. *Toxins,* 14**,** 138.

VANEGAS-TORRES, C. A. & SCHINDLER, M. 2024. HIV-1 Vpr Functions in Primary CD4+ T Cells. *Viruses,* 16**,** 420.

VERHOEF, K., SANDERS, R. W., FONTAINE, V., KITAJIMA, S. & BERKHOUT, B. 1999. Evolution of the human immunodeficiency virus type 1 long terminal repeat promoter by conversion of an NF-kappaB enhancer element into a GABP binding site. *J Virol,* 73**,** 1331-40.

VERMA, A. 2015. *Functional characterization of a subtype-specific NF-kB motif in HIV-1 subtype C viral promoter and its association with the proximal and subtype-specific Sp1site.* Jawaharlal Nehru Centre for Advanced Scientific Research.

VIDAL-LIMON, A., AGUILAR-TOALÁ, J. E. & LICEAGA, A. M. 2022. Integration of molecular docking analysis and molecular dynamics simulations for studying food proteins and bioactive peptides. *Journal of Agricultural and Food Chemistry,* 70**,** 934-943.

WELINDER, C. & EKBLAD, L. 2011. Coomassie staining as loading control in Western blot analysis. *Journal of proteome research,* 10**,** 1416-1419.

WU, Y. & MARSH, J. W. 2003. Gene transcription in HIV infection. *Microbes and Infection,* 5**,** 1023-1027.

ZHANG, J. & CRUMPACKER, C. 2022. HIV UTR, LTR, and Epigenetic Immunity. *Viruses,* 14**,** 1084.

ZHOU, J., LI, M., MIN, C., MA, Y., SHAO, Y. & XING, H. 2020. Near full-length genomic characterization of a novel HIV-1 circulating recombinant form (CRF106_cpx) identified among heterosexuals in China. *AIDS Research and Human Retroviruses,* 36**,** 875-880.

ZILA, V., MARGIOTTA, E., TUROŇOVÁ, B., MÜLLER, T. G., ZIMMERLI, C. E., MATTEI, S., ALLEGRETTI, M., BÖRNER, K., RADA, J., MÜLLER, B., LUSIC, M., KRÄUSSLICH, H.-G. & BECK, M. 2021. Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *Cell,* 184**,** 1032-1046.e18.

APPENDICES (SUPPLEMENTARY MATERIAL)

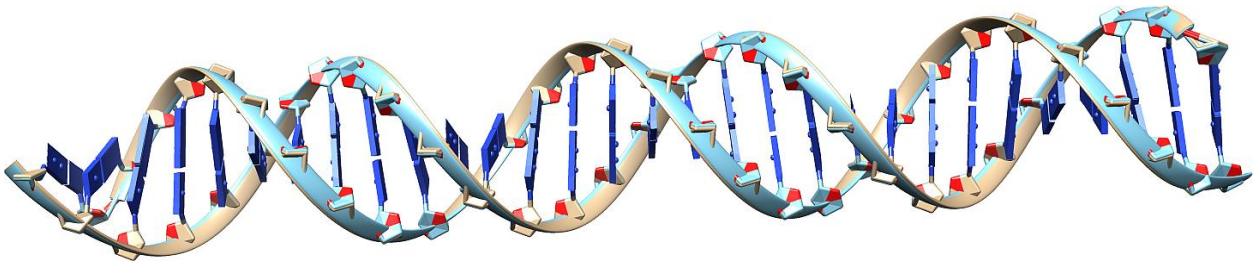


Figure S1: Alignment of the two modeled LTR nucleotide structure. The nucleotide (LTR) was modeled using two different webservers: model.it® Server (http://pongor.itk.ppke.hu/dna/model_it.html#/modelit_form) (Vlahoviček & Pongor, 2000) and Supercomputing Facility for Bioinformatics & Computational Biology webserver (http://www.scfbio-iitd.res.in/software/drugdesign/bdna.jsp#) as shown in the Figure S1. Both web servers uses different algorithms in modelling nucleotides but the structures obtained from both servers show a similar orientation (visually and sequentially). The LTR model with Supercomputing Facility for Bioinformatics & Computational Biology webserver and model.it® Server were represented with cyan helix and brown helix respectively.

```
CLUSTAL O(1.2.4) multiple sequence alignment


ACR22508.1      MQGVSLGQTSSSNTTLTPIASAASIPAGTVTVNAAQLSSMPGLQTINLSALGTSGIQVHP    60
pdb:A           MQGVSLGQTSSSNTTLTPIASAASIPAGTVTVNAAQLSSMPGLQTINLSALGTSGIQVHP    60
                ************************************************************

ACR22508.1      IQGLPLAIANAPGDHGAQLGLHGAGGDGIHDDTAGGEEGENSPDAQPQAGRRTRREACTC    120
pdb:A           IQGLPLAIANAPGDHGAQLGLHGAGGDGIHDDTAGGEEGENSPDAQPQAGRRTRREACTC    120
                ************************************************************

ACR22508.1      PYCKDSEGRGSGDPGKKKQHICHIQGCGKVYGKTSHLRAHLRWHTGERPFMCTWSYCGKR    180
pdb:A           PYCKDSEGRGSGDPGKKKQHICHIQGCGKVYGKTSHLRAHLRWHTGERPFMCTWSYCGKR    180
                ************************************************************

ACR22508.1      FTRSDELQRHKRTHTGEKKFACPECPKRFMRSDHLSKHIKTHQNKKGGP-VALSVGTLPL    239
pdb:A           FTRSDELQRHKRTHTGEKKFACPECPKRFMRSDHLSKHIKTHQNKKGGPGVALSVGTLPL    240
                ************************************************ **********

ACR22508.1      DSGAGSEGSGTATPSALITTNMVAMEAICPEGIARLANSGINVMQVADLQSINISGNGF     298
pdb:A           DSGAGSEGSGTATPSALITTNMVAMEAICPEGIARLANSGINVMQVADLQSINISGNGF     299
                ************************************************************
```

Figure S2: CLUSTAL O (1.2.4) multiple sequence alignment of the AlphaFold predicted structure (ACR22508.1) and the homology modeled structure.

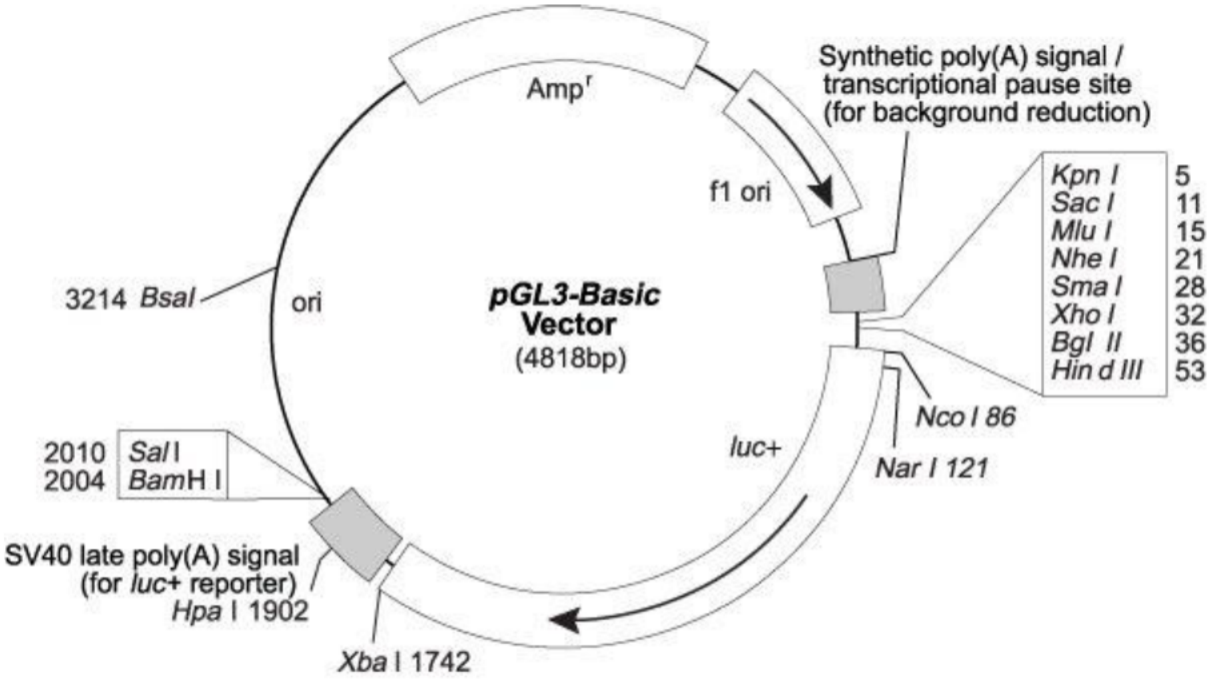


Figure S4: **The PGL3 basic vector** map. The pGL3-Basic vector contains a modified firefly luciferase gene (luc+) and a gene conferring ampicillin resistance in E. coli (Amp). The arrow within luc+ indicates the direction of transcription. (uploaded by Dylan Glubb)

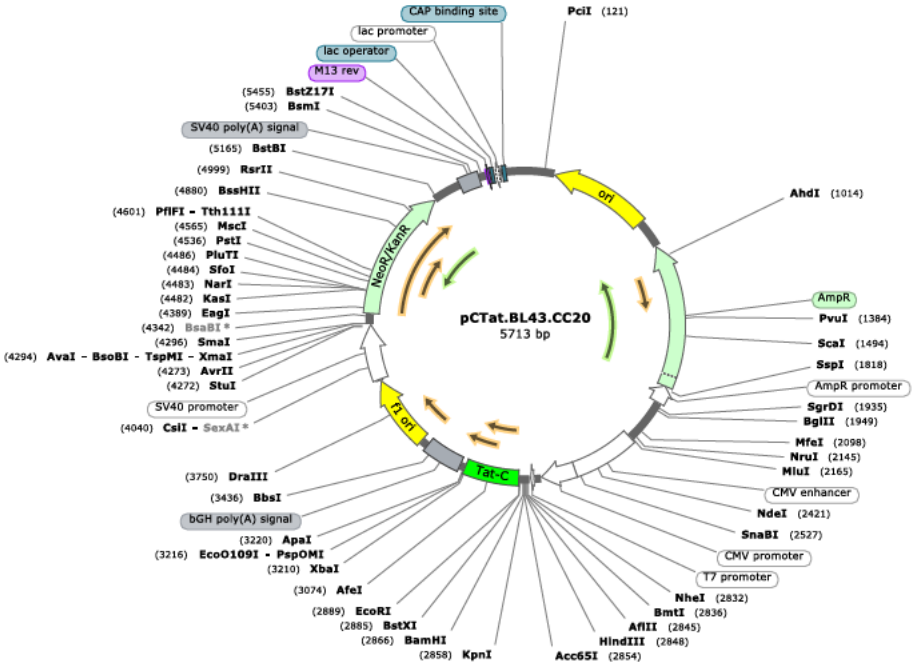


Figure S5: **The pcTat expression vector map**, showing key functional elements including the promoter, Tat-C gene, antibiotic resistance markers (such as AmpR), and origins of replication. Restriction enzyme sites are indicated for cloning purposes.