FUNCTIONAL CHARACTERIZATION OF A SUBTYPE-SPECIFIC NF-κB MOTIF IN HIV-1 SUBTYPE C VIRAL PROMOTER AND ITS ASSOCIATION WITH THE PROXIMAL AND SUBTYPE-SPECIFIC Sp1 SITE

A THESIS

SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

ANJALI VERMA

HIV-AIDS LAB

MOLECULAR BIOLOGY AND GENETICS UNIT

JNCASR

BANGALORE, INDIA

FEBRUARY 2015

A Heartfelt Dedication to.

My family..

DECLARATION

I hereby declare that the matter embodied in the thesis entitled as "**Functional characterization of a subtype-specific NF-κB motif in HIV-1 subtype C viral promoter and its association with the proximal and subtype-specific Sp1 site**" is the result of investigations carried out by me at Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India under supervision of Prof. Udaykumar Ranga. This work in part or full has not been submitted to any other institute or university for the award of any degree or diploma.

In keeping with the general practice in reporting the scientific observations, due acknowledgments have been made wherever the work described is based on the finding of other investigators.

.....

Ms. Anjali Verma

CERTIFICATE

I hereby certify that the matter embodied in the thesis entitled as "**Functional characterization of a subtype-specific NF-κB motif in HIV-1 subtype C viral promoter and its association with the proximal and subtype-specific Sp1 site**" has been carried out by Ms. Anjali Verma at Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India, under my supervision. This work in part or full has not been submitted to any other institute or university for the award of any degree or diploma.

……..

Prof. Udaykumar Ranga

ACKNOWLEDGEMENTS

This doctoral thesis would not have been possible without help and support of the kind people around me. It has been a wonderful experience being a part of the JNCASR.

Now, I take this opportunity to express my deepest sense of gratitude to my advisor Prof. Udaykumar Ranga. First of all, I thank him for allowing me to work in his laboratory. Further, I thank him for his criticism, patience, motivation, enthusiasm and for sharing immense knowledge. His guidance and enthusiasm kept me high throughout the course of study. He gave me enough freedom to think on my own about the problems, new protocols and design experiments in an independent manner. A laboratory facility is the most important requirement, and I thank you for providing everything that was needed to perform an experiment. I thank you for allowing me to take a couple of international tours as scientific conferences.

I thank Prof. Anuranjan Anand and Prof. M.R.S. Rao as the present and past chairman of our department, for their efficient administration and excellent support in all official and technical matters.

My sincere thanks to all faculties of Molecular Biology and Genetics Unit. Special thanks to Prof. Tapas Kumar Kundu, Prof. Hemalata Balaram, Prof. Kaustuv Sanyal, for their excellent teaching during my course work. I also thank other faculties of MBGU, Prof. Manisha Inamdar, Prof. Namita Surolia, Dr. Ravi Manjithiya and Dr. James for creating joyful and scientific atmosphere in MBGU.

My special thanks to Prof. Tapas Kumar Kundu for wonderful suggestions and helpful discussions.

I thank Council of Scientific and Industrial Research (CSIR) for the junior and senior research fellowships. I thank JNCASR for providing funding to attend XII Asian Conference on Transcription (ACT-12), Jeju Island, South Korea (2012) and Indian Council of Medical

Research (ICMR), Department of Biotechnology (DBT) and Centre for International Co-operation in Science (CICS, Chennai) for providing financial assistance to attend Frontiers of Retrovirology held in Cambridge, UK (2013).

I wish to thank my past lab mates Rajesh, Roshan, Shambhu, Dr. Mahesh, Dr. Anand, Dr. Mangai, Rashid, Dr. Venkatesh, Swarupa, Neelakshi, Debanjan, Chetan, Raju, Maunish, Mayank, Saihitha, Mekha, Krupa. They were very co-operative and provided a cordial atmosphere in the lab.

My sincere thank to all my present lab members: Malini, Prabhu, Sutanuka, Arun, Deepak, Shilpee, Chaitra, Shravan, Siddharth, Ria, Surabhi, Disha, Lavanya, Lakkappa, Sridhar for making the lab atmosphere comfortable and pleasant.

My special thanks to Pavithra, Rishi and Rebu for their great help and excellent support in my work.

I sincerely thanks Malini, Prabhu and Sutanuka for the proofreading of my thesis.

I also thank my batchmates and close friends Kalpita, Deeti, Priya for their constant support while my stay in JNC and also for helping me in my project .

My teachers and mentors who have inspired me. I thank my teachers during M.Sc. course Prof. Ajit Sodhi, Prof. Arvind Kayastha, Prof. Anil Kumar Tripathi, Prof. Ashok Kumar, Prof. B.D. Singh, Prof. S.M. Singh, Prof. Jagannadham, Prof. Arvind Kumar, Dr. Ajanta Chadha and all of them who taught me. Special thanks to Prof. Ajit Sodhi for allowing me to work in his lab for my M.Sc. project. What learnt in his lab helped me a lot during my PhD. I thank him for his criticism and motivation.

My special thanks to all my childhood friends Lucky, Neelu, Kalpana and sarika.

Besides this, several people have knowingly and unknowingly helped me in the successful completion of this thesis.

I feel very happy to try to sort of thanking to Mummy, Papa, my brother (Akhil), sister-in-law (Garima, Sashi), niece (Avika), my husband Indar, brother-in-laws (Akash, Prakash and Kamlesh), nephews (Kaushal, Karan and Abhijit), father-in-law and mother-in law.

All my family members for supporting me, for everything I have achieved till today and will in the future.

Finally I am grateful to God, whom I believe in and whose blessings have always been showered upon me.

Thank you all!!

Anjali Verma

Table of contents

Table of contents

Table of contents

Abbreviations

AIDS	Acquired immuno defficiency syndrome
AP-1	Activator protein-1
BSA	Bovine Serum Albumin
CCR5	C-C chemokine receptor 5
CD4	Cluster of differentiation 4
CD8	Cluster of differentiation 8
cDNA	Complementary deoxyribonucleic acid
ChIP	Chromatin Immunoprecipitation
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothretol
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked Immunosorbant assay
EMSA	Electrophoretic mobility shift assay
env	Envelope
gp120	Glycoprotein 120
gp41	Glycoprotein 41
GST	Glutathione S-transferase
HAART	Highly Active Anti-Retroviral Therapy
HBS	HEPES-buffered saline
HDACs	Histone deacetylases
HIV-1	Human immunodeficiency virus-1
IKK	Inhibitory kappaB kinase
I κBα	Inhibitory kappaB alpha
I κBβ	Inhibitory kappaB beta
IL-2	Interleukin-2
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LTR	Long Terminal Repeat
Nef	Negative factor
NFAT	Nuclear Factor of Activated T-cells
NF-κB	Nuclear Fator-Kappa B
PAGE	Poly Acrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PHA-P	Phytohemagglutinin-P
PMA	Phorbol-12-myristate-13-acetate
PMSF	Phenylmethylsulfonyl fluoride
P-TEF b	Positive Transcription Elongation Factor b
RNA	Ribonucleic Acid
RT-PCR	Reverse Transcriptase-Polymerase Chain Reaction
SDS	Sodium dodecyl sulfate

Sp1	Specificity protein 1
TAE	Tris-acetate-EDTA
TAF	TBP-associated factors
TAR	Trans-activating response region
TBE	Tris-borate-EDTA
TBP	TATA-box binding protein
TEMED	N, N, N', N'-Tetramethylethylenediamine
TFBS	Transcription factor bainding site
TNF-α	Tumor Necrosis Factor-α
U3	Unique to 3'
U5	Unique to 5'
UNAIDS	United Nations Programme on Aquired Immune Deficiency Syndrome
USF	Upstream stimulating factor
vpu	Viral protein unknown
vpr	viral protein r

Synopsis of thesis

Title: Functional characterization of a subtype-specific NF-κB motif in HIV-1 subtype C viral promoter and its association with the proximal and subtype-specific Sp1 site

Submitted by

Anjali Verma

Thesis advisor: Prof. Udaykumar Ranga

HIV-AIDS Lab, MBGU

Jawaharlal Nehru Centre for Advanced Scientific Research

Bangalore

Based on genetic variation, Human Immunodeficiency Virus Type 1 (HIV-1) is classified into four distinct groups (M, N, O and P), and the group M into 9 molecular subtypes and numerous circulating recombinant forms (Kuiken *et al*. 2010). The HIV-1 subtypes are unevenly distributed across the world. Amongst the HIV-1 subtypes, subtype C alone is responsible for half of the HIV-1 infections globally and more than 95% of the infections in India (UNAIDS 2013; http://www.unaids.org). The factors that contribute to the widespread expansion of subtype C are not well understood. Diverse viral subtypes differ from one another up to 30-35% in certain gene segments such as the envelope (Chang *et al.* 2000). A genetic variation to this large extent may have a significant impact on the biological properties of the subtypes influencing their relative infectivity and expansion. Subtype-specific genetic variations may underlie the biological differences, although drawing such a correlation between these factors has been always not possible. Subtype-specific molecular variations in elements such as the viral promoter (enhancer and other regulatory elements), regulatory proteins and structural proteins deserve more attention to understand the

differences in such intrinsic biological properties of viral subtypes. Experimental evidence available is often inadequate and conflicting to examine whether the differences observed among the various subtypes at the molecular level could be extrapolated to differences in their biological properties such as the coreceptor preference.

For the present study, we focused on the viral promoter, the long-terminal repeat (LTR), as the viral modulator region, enhancer and the core promoter are characterized by several subtype-specific molecular differences. Such differences have been mapped to many transcription factor binding sites (TFBS) including USF, c-Myb, NF-AT, Ap-1, NF-κB, Sp1, and regulatory elements such as the TATA box and TAR in the LTR (Jeeninga *et al.* 2000; Montano *et al.* 1997). Of these differences, the variations in the genetic sequence and copy number of the NF-κB binding sites are of importance especially for subtype C. While all the HIV-1 subtypes contain two genetically identical NF-κB sites in the viral enhancer (which we refer to as the H-κB site, GGGACTTTCC), with the exception of subtype E that contains only one site (Ranjbar *et al.* 2006;Roof *et al.* 2002), subtype C alone contains three such motifs (Bachu *et al.* 2012;Bjorndal *et al.* 1999;Munkanta *et al.* 2005;Novitsky *et al.* 2002). Importantly, the additional NF-κB motif differs from the standard H-κB site at two positions, thus is genetically distinct (referred to as the C-κB site, GGGGCGTTCC, differences underlined). The viral enhancer comprising of the NF-κB sites is immediately followed by three Sp1 binding sites. The enhancer-proximal Sp1 III site not only plays a more critical role in regulating gene expression from the viral promoter (Burnett *et al.* 2009), but is also characterized by subtype-specific variations (GAGGCGTGGC in subtype B and GAGGTGTGGT in subtype C).

A large number of studies examined the regulation of gene expression form HIV-1 promoter in the context of subtype B (B-LTR). From the previous studies, it is evident that the NF-κB and Sp1 motifs in the viral promoter collectively play a critical role in controlling

not only the gene expression, but also viral latency (Burnett *et al.* 2009). The NF-κB motifs are responsible for the inducible gene expression while the Sp1 motifs regulate the basal level of the transcription. Unlike the many inducible cellular promoters that contain these two important TFBS, HIV-1 LTR is exceptional in not only containing multiple binding elements of each of the two categories, but also positioning the sites in close proximity typically with a spacer of only 2 to 4 bases separating any two sites. Furthermore, in B-LTR, an active functional interaction was demonstrated between the adjacent NF-κB and Sp1 III sites such a way not only the spacing between them, but also the relative orientation of each site with respect to the other cannot be altered suggesting that the host factors must bind the specific elements in a specific orientation (Perkins *et al.* 1993). An orientation- and proximity-dependent interaction between the adjacent NF-κB and Sp1 motifs was subsequently demonstrated by the same research group (Perkins *et al.* 1994).

A large number of reports collectively ascertained the critical role played by the central NF-κB and the Sp1 III motifs in controlling viral gene expression as compared to the flanking sites (one upstream H-κB site and two downstream Sp1 motifs) (Burnett *et al.* 2009). Given the special significance attached to the central NF-κB and the Sp1 III motifs in the HIV-1 LTR, importantly to the established proximity- and orientation-dependent functional interaction between them, it is enigmatic that a genetically different NF-κB site (the C-κB motif) is inserted in subtype C LTR disrupting the original association between the H-κB and the Sp1 III motifs. Additionally, in C-LTR, the Sp1 III motif is subjected to subtype-specific genetic variation (Jeeninga *et al.* 2000) alluding to a newly formed association between the newly inserted C-κB site and the subtype-specific Sp1 motif.

Many previous publications ascribed a quantitative gain-of-function to the subtype C promoter as a consequence of the insertion of the C-κB site by demonstrating enhanced expression of a reporter gene placed under the control of the viral promoter (Kurosu *et al.*

2002;Montano *et al.* 2000;Naghavi *et al.* 1999). These publications, however, did not take into account that the newly inserted NF-κB site is genetically different and that there exists an associated genetic variation in the proximal Sp1 site. We wondered, if quantitative gain-of-function is the only objective, then, why the newly acquired NF-κB site should be a genetic variant. The C-κB site is unique for subtype C, absent from all the other HIV-1 subtypes as well as in HIV-2 and SIV strains, highly conserved among all the subtype C viral strains, invariably placed adjacent to the Sp1III site, not separated from the Sp1 site by the insertion of heterologous sequences and never subjected to sequence deletion suggesting the indispensable nature of this element for subtype C. The indispensable nature of the C-κB site for subtype C LTR and the associated variations in the adjacent Sp1III site collectively allude to the possibility that in addition to the quantitative-gain-of-function, these two genetic elements collectively impart a qualitative-gain-of-function on the viral promoter in a subtype-specific manner.

In this backdrop, the primary objective of the present work is to examine if there exists a functional association between the HIV-1 subtype C unique NF-κB binding site (the C-κB site) and the proximal and subtype-specific Sp1 III motif in C-LTR. A functional association between the two adjacent motifs is expected and remains to be demonstrated. We used a range of experimental strategies to study the functional association between the two regulatory elements in C-LTR.

Chapter 1 presents an overview of the literature that includes an introduction to HIV-1 subtypes and the associated molecular and biological differences, Rel and Sp1 family members and the mechanisms by means of which these families regulate transcription control from the cellular and HIV promoters.

Chapter 2 provides a detailed description of the reagent details, experimental strategies and protocols used in addressing the research questions.

Chapter 3 presents the experimental strategies and the data on the function of the subtype-specific NF-κB and Sp1 sites in regulating viral gene expression in the context of subtype C. Using a parental reporter expression vector that simultaneously expresses secreted guassia luciferase and EGFP under the control of a prototype subtype C LTR (Siddappa *et al.* 2007), we constructed three different panels of reporter vectors. The viral promoter in the reporter vectors was engineered to introduce variations in the relative location, orientation and spacing of the NF-κB and the Sp1 motifs. Using a luciferase reporter assay and flow cytometry analysis of GFP, we demonstrate that the association between the two subtype-specific elements is proximity, orientation and distance dependent. Optimal gene expression from C-LTR requires the presence of the C-κB site and the subtype-specific Sp1 III motif in the central location and other genetic variant motifs cannot serve as functional substitutes for either of the central elements. The substitution of the C-κB site with the variant H-κB motif significantly compromised reporter expression from the viral promoter. Likewise, the Sp1II motif cannot function as a functional substitute for the subtype-specific SP1III motif. The genetically variant Sp1III motifs derived from HIV-1 subtypes A, B, D and A/E, however, can functionally associate with the C-κB motif.

Comparable results were observed with a panel of reporter vectors in which the flanking NF-κB and Sp1 sites were inactivated thus leaving only the central C-κB and Sp1III motifs in the viral promoter suggesting that the flanking TFBS did not influence the gene expression profile significantly. Furthermore, we generated a panel of reporter viral vectors representing the reporter panels constructed above. The viruses were pseudotyped in HEK293 cells (Weinberger *et al.* 2005), viral stocks titered and target cells were infected. The expression of the reporter EGFP expression was evaluated from the integrated provirus using

flow cytometry. The data confirmed the observations made using the reporter vectors above that for optimal gene expression, the C-κB and Sp1III sites must be in close proximity in the natural orientation. In chromatin immuno precipitation (ChIP) analysis, the LTR containing the C-κB and Sp1 III sites primarily bound p50 and Sp1, but little p65 or RNA polymerase in the absence of activation suggesting the occupancy of the C-κB site by the repressive p50 homodimer. Following TNF-α activation, however, the promoter was occupied by p50, p65, Sp1 and significantly higher quantities of RNA polymerase phosphorylated on serine 2 were recruited to the promoter. Collectively, the strategy of the TFBS shuffling suggested that the C-κB and the Sp1 III motifs must be located proximal to each other and the alternate TFBS are not functional substitutes for either of the two original motifs. The specific location, mutual orientation and spacing between the C-κB and Sp1 III sites are critical factors that regulate the biological function of the two elements confirming a functional association between the two important elements in subtype C LTR.

Using gel-shift and super-shift assays, we demonstrate that the C-κB probe can recruit p50-p50 homodimer and p50-p65 heterodimer from Jurkat nuclear extracts under no activation and TNF-α activation conditions, respectively, in a manner comparable to the H-κB probe. Importantly, the C-κB probe (5'-CCACT**GGGGCGTTCC**AGGA-3'), as compared to the H-κB probe (5'-CCACT**GGGACTTTCC**AGGA-3') demonstrates approximately a 2-fold higher affinity for the recombinant p50 protein and for the cellular factors in nuclear extract of Jurkat cells. Furthermore, the C-κB probe, regardless of a debilitating variation as compared to the H-κB probe, also recruits NFAT I and NFAT II cellular factors.

Using the infectious molecular clone Indie-C1, a prototype subtype C strain, we constructed a pair of 'isogenic' viruses that differed from each other with respect to the central NF-κB site alone. The wild type virus contained the natural C-κB motif (HHC) while

the paired variant contained an H-κB motif replacing the C-κB site (HHH). We compared the replication kinetics of the two viruses in the PHA-activated PBMC derived from six different healthy donors, CEM-CCR5 T-cells or CN-2 cells that are PBMC transformed using the human herpes virus (Vella *et al.* 1997). The HHC virus demonstrated the expected proliferation kinetics with the viral replication peaking at day 14-21 in different cells. The HHH virus unexpectedly failed to proliferate in 5 of the 6 PBMC, the CEM-CCR5 T-cells and the CN-2 cells thus demonstrating a severe defect in proliferation. The results were reproducible. The viral stocks were tested before and after the assay and the stocks did not show any difference in the $TCID_{50}$ titers ruling out the possibility of deterioration. In the HHH virus, the H-κB motif was artificially brought in association with the subtype C-specific SP1III motif (HHH:C) suggesting that the H-κB motif and the subtype C-specific Sp1 III sites are probably not compatible with each other. To test this hypothesis, we made an additional virus by substituting the subtype C-specific Sp1III site with the subtype B-specific Sp1 III site (HHH: B) thus bringing the H-κB motif in association with its cognate Sp1 site in the subtype C backbone. In the HHH:B virus, the viral proliferation was restored that was comparable to the wild type virus (HHC:C). An additional HHC:B virus was also constructed and found to be replication competent. In summary, the Sp1III element of subtype B can function with either of the two NF-κB variants (H- or C-κB sites) with comparable efficiency. In contrast, the Sp1III element of subtype C can function only with the natural C-κB element, but not with the heterogeneous H-κB motif although the latter element is functional in a different viral context.

To understand at what level of the viral life cycle the restriction was imposed on the HHH:C virus, CEM-CCR5 T-cells were infected independently with HHC:C and HHH:C isogenic viral strains at an equivalent $TCID_{50}$ titer and viral proliferation was compared at two different viral stages of infection using real-time PCR. We compared the generation of

the reverse transcription products in the cell extract and the formation of proximal *versus* distal viral Transcripts. No significant differences were found between the two viruses at the level of the reverse transcription suggesting comparable magnitude of viral infection by both the viral strains. A significant difference, however, was evident at the level of transcription, at both transcription initiation and elongation suggesting the H-κB site and the subtype C-specific Sp1III motifs are not compatible with each other in collectively coordinating viral transcription. To the best of our knowledge, this work represents the first demonstration that an NF-κB site and an Sp1 motif cannot function in association with each other although these elements are biologically functional in a different viral context. In ChIP, we examined the nature of the cellular complexes recruited by the viral promoters in the above experiment using antibodies specific to p50, p65, Sp1, H3K9acetylation, Pol II S2 or S5 phosphorylation, HDAC1 and 3. The nature of the cellular complexes found at the HHC:C promoter was predominantly transcription promoting (p50, p65, Sp1, H3K9acetylation, Pol II S2 or S5 phosphorylation) while that at the HHH:C was predominantly transcription suppressive (P50, Sp1, HDAC 1 and 3) further confirming the reporter analysis.

Chapter 4 provides a discussion and a summary of the entire work. The results clearly point at a possibility that subtype C uses a mechanism to regulate gene expression in a way different from all other HIV-1 genetic subtypes. Although various other TFBS in the viral promoter must be playing a significant role in regulating the overall gene expression, the association between the centrally located C-κB motif and the Sp1 III site is more critical. Although the two central TFBS belong to different transcription factor families with distinctive biological functions ascribed to each family, given the proximity- and orientation-dependent association between them and importantly, the inability of other variant genetic elements to serve as functional substitutes, the C-κB site and the Sp1III motif must be considered as a single functional unit in C-LTR.

Unlike the previous reports that showed only a quantitative-gain-of-function advantage, our work demonstrates that the acquisition of the C-κB motif (HHC) also leads to qualitative-gain-of-function advantage. Acquisition of an additional H-κB site (HHH) couldn't have provided the qualitative-gain-of-function advantage. Of note, the acquisition of the C-κB site has been also associated with the acquisition of genetic variations within the Sp1III motif. It is quite intriguing to explain which of these two events has preceded the other. The acquisition of subtype-specific Sp1 III element by an ancestral subtype C promoter containing two H-κB motifs (HH) is not practically possible for such a virus cannot enjoy replication competence given the H-κB:Sp1IIIc incompatibility (HH:C). It is therefore reasonable to assume that the subtype-associated variations in C-LTR may have taken place in two sequential steps. First, the ancestral subtype C enhancer (HH) may have acquired an additional and variant C-κB motif first. This specific combination of C-κB motif and the subtype B-specific Sp1III element (HHC:B) should be functionally compatible. The presence of the C-κB site in the C-LTR may have catalyzed subsequent variations in the adjacent Sp1 site as are seen currently (HHC:C). Presently, we do not know if the proposed pathway for the promoter evolution in subtype C is justified given the lack of ancestor subtype C sequences in the public databases representing the intermediary stages of the evolution. Additionally, it is also not clear what were the evolutionary forces that positively selected the transition of the HHC:B promoter to HHC:C despite the fact that the former is as competent as the latter in terms of replication. Furthermore, it is necessary to examine and identify the nature of the cellular factors that regulate the κB-Sp1 incompatibility at the molecular level. The epigentic marks identified in the ChIP analysis showed a predominantly transcription suppressive chromatin in the context of the HHH:C promoter in contrast to the transcription activating chromatin in the HHC:C promoter. It is tempting to propose that the unique H-

κB:Sp1 III_C combination (HHH:C), unlike the other three promoters (HHC:C, HHC:B and HHH:B), recruits one or more suppressive cofactors which modifies the chromatin and suppresses viral transcription. Using double-stranded DNA probes HHC:C and HHH:C, we are presently isolating the cellular factors. The identity of the associated cellular factors will be determined using Mass-spectrometry.

In summary, our work establishes a new found functional association between the subtype-specific variant NF-κB motif and subtype-specific variant Sp1III site in subtype C viral promoter. Our data suggest further that subtype C uses a mechanism to regulate gene expression in a way different from all other HIV-1 genetic subtypes. How the unique genetic elements in subtype C promoter modulate gene expression and viral latency needs additional investigation.

Bibliography

Bachu, M., S. Yalla, M. Asokan, A. Verma, U. Neogi, S. Sharma, R.V. Murali, A.B. Mukthey, R. Bhatt, S. Chatterjee, R.E. Rajan, N. Cheedarla, V.S. Yadavalli, A. Mahadevan, S.K. Shankar, N. Rajagopalan, A. Shet, S. Saravanan, P. Balakrishnan, S. Solomon, M. Vajpayee, K.S. Satish, T.K. Kundu, K.T. Jeang and U. Ranga, 2012. Multiple NF-kappaB sites in HIV-1 subtype C long terminal repeat confer superior magnitude of transcription and thereby the enhanced viral predominance. J.Biol.Chem. 287: 44714-44735. Available from PM:23132857

Bjorndal, A., A. Sonnerborg, C. Tscherning, J. Albert and E.M. Fenyo, 1999. Phenotypic characteristics of human immunodeficiency virus type 1 subtype C isolates of Ethiopian AIDS patients. AIDS Res.Hum.Retroviruses 15: 647-653. Available from PM:10331443

Burnett, J.C., K. Miller-Jensen, P.S. Shah, A.P. Arkin and D.V. Schaffer, 2009. Control of stochastic gene expression by host factors at the HIV promoter. PLoS.Pathog. 5: e1000260. Available from PM:19132086

Chang, S.Y., V. Vithayasai, P. Vithayasai, M. Essex and T.H. Lee, 2000. Human immunodeficiency virus type 1 subtype E envelope recombinant peptides containing naturally immunogenic epitopes. J.Infect.Dis. 182: 442-450. Available from PM:10915074

Jeeninga, R.E., M. Hoogenkamp, M. rmand-Ugon, B.M. de, K. Verhoef and B. Berkhout, 2000a. Functional differences between the long terminal repeat transcriptional promoters of human immunodeficiency virus type 1 subtypes A through G. J.Virol. 74: 3740-3751. Available from PM:10729149

Kuiken, C., Leitner, T., Hahn, B.,Mullins, J., Wolinsky, S., Foley, B., Apetrei, C., Mizrachi, Y., Rambait, A., and Korber, B., 2010. HIV sequence compemndium 2010.

Kurosu, T., T. Mukai, S. Komoto, M.S. Ibrahim, Y.G. Li, T. Kobayashi, S. Tsuji and K. Ikuta, 2002. Human immunodeficiency virus type 1 subtype C exhibits higher transactivation activity of Tat than subtypes B and E. Microbiol.Immunol. 46: 787-799. Available from PM:12516777

Montano, M.A., C.P. Nixon, T. Ndung'u, H. Bussmann, V.A. Novitsky, D. Dickman and M. Essex, 2000. Elevated tumor necrosis factor-alpha activation of human immunodeficiency virus type 1 subtype C in Southern Africa is associated with an NF-kappaB enhancer gain-of-function. J.Infect.Dis. 181: 76-81. Available from PM:10608753

Montano, M.A., V.A. Novitsky, J.T. Blackard, N.L. Cho, D.A. Katzenstein and M. Essex, 1997. Divergent transcriptional regulation among expanding human immunodeficiency virus type 1 subtypes. J.Virol. 71: 8657-8665. Available from PM:9343223

Munkanta, M., R. Handema, H. Kasai, C. Gondwe, X. Deng, A. Yamashita, T. Asagi, N. Yamamoto, M. Ito, F. Kasolo and H. Terunuma, 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: 901-906. Available from PM:16225420

Naghavi, M.H., S. Schwartz, A. Sonnerborg and A. Vahlne, 1999. Long terminal repeat promoter/enhancer activity of different subtypes of HIV type 1. AIDS Res.Hum.Retroviruses 15: 1293-1303. Available from PM:10505678

Novitsky, V., U.R. Smith, P. Gilbert, M.F. McLane, P. Chigwedere, C. Williamson, T. Ndung'u, I. Klein, S.Y. Chang, T. Peter, I. Thior, B.T. Foley, S. Gaolekwe, N. Rybak, S. Gaseitsiwe, F. Vannberg, R. Marlink, T.H. Lee and M. Essex, 2002. Human immunodeficiency virus type 1 subtype C molecular phylogeny: consensus sequence for an AIDS vaccine design? J.Virol. 76: 5435-5451. Available from PM:11991972

Perkins, N.D., A.B. Agranoff, E. Pascal and G.J. Nabel, 1994. An interaction between the DNA-binding domains of RelA(p65) and Sp1 mediates human immunodeficiency virus gene activation. Mol.Cell Biol. 14: 6570-6583. Available from PM:7935378

Perkins, N.D., N.L. Edwards, C.S. Duckett, A.B. Agranoff, R.M. Schmid and G.J. Nabel, 1993. A cooperative interaction between NF-kappa B and Sp1 is required for HIV-1 enhancer activation. EMBO J. 12: 3551-3558. Available from PM:8253080

Ranjbar, S., R. Rajsbaum and A.E. Goldfeld, 2006. Transactivator of transcription from HIV type 1 subtype E selectively inhibits TNF gene expression via interference with chromatin remodeling of the TNF locus. J.Immunol. 176: 4182-4190. Available from PM:16547255

Roof, P., M. Ricci, P. Genin, M.A. Montano, M. Essex, M.A. Wainberg, A. Gatignol and J. Hiscott, 2002. Differential regulation of HIV-1 clade-specific B, C, and E long terminal repeats by NF-kappaB and the Tat transactivator. Virology 296: 77-83. Available from PM:12036319

Siddappa, N.B., V.P. Kashi, M. Venkatramanan, A. Balasiddaiah, N. Jayasuryan, A. Mahadevan, A. Desai, K.S. Satish, S.K. Shankar, V. Ravi and U. Ranga, 2007. Gene expression analysis from human immunodeficiency virus type 1 subtype C promoter and construction of bicistronic reporter vectors. AIDS Res.Hum.Retroviruses 23: 1268-1278. Available from PM:17961115

Vella, C., H. Fickenscher, C. Atkins, M. Penny and R. Daniels, 1997. Herpesvirus saimiri-immortalized human T-cells support long-term, high titred replication of human immunodeficiency virus types 1 and 2. J.Gen.Virol. 78 (Pt 6): 1405-1409. Available from PM:9191937

Weinberger, L.S., J.C. Burnett, J.E. Toettcher, A.P. Arkin and D.V. Schaffer, 2005. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. Cell 122: 169-182. Available from PM:16051143

CHAPTER-1

INTRODUCTION

1.1 Retroviruses

The retroviruses represent a diverse family of RNA containing viruses. In this family of viruses, the RNA is reverse transcribed into double-stranded DNA which subsequently integrates into the genome of the host cell (Coffin,1992;Lee and Coffin,1990). These viruses have been classified into two groups: simple and complex retroviruses. While the simple retroviruses encode only the critical proteins such as gag, pol and capsid, the complex retroviruses also encode additional regulatory proteins including Tat and rev (Figure-1.1) (Shida,2012).

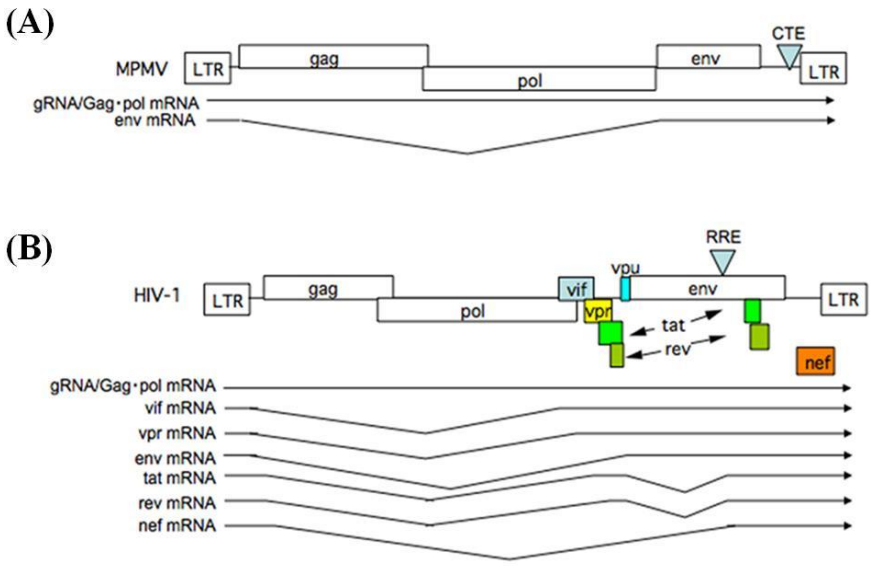


Figure-1.1: (A) The genome of a simple retrovirus encoding three major proteins gag, pol and env. **(B)** The genome of a complex retrovirus coding for regulatory (tat and rev) and accessory proteins (vif, vpr and vpu) in addition to the three elementary proteins. Images adapted from 10.3389/fmicb.2012.00179.

The retroviruses have been broadly classified into three subfamilies: (1) Oncovirus, (2) Lentivirus and (3) Spumavirus (Weiss,1996). The Human immunodeficiency virus (HIV) belongs to the lentiviral family.

1.2 Human immunodeficiency virus (HIV)

HIV, a complex retrovirus, belongs to the lentiviral family. The viral particle is wrapped in a lipid bi-layer membrane derived from the host cell. The envelope consisting of gp120 and gp41 supports the entry of the virion into the host cell. The core of HIV-1 contains two copies of single-stranded RNA molecules (Figure-1.2).

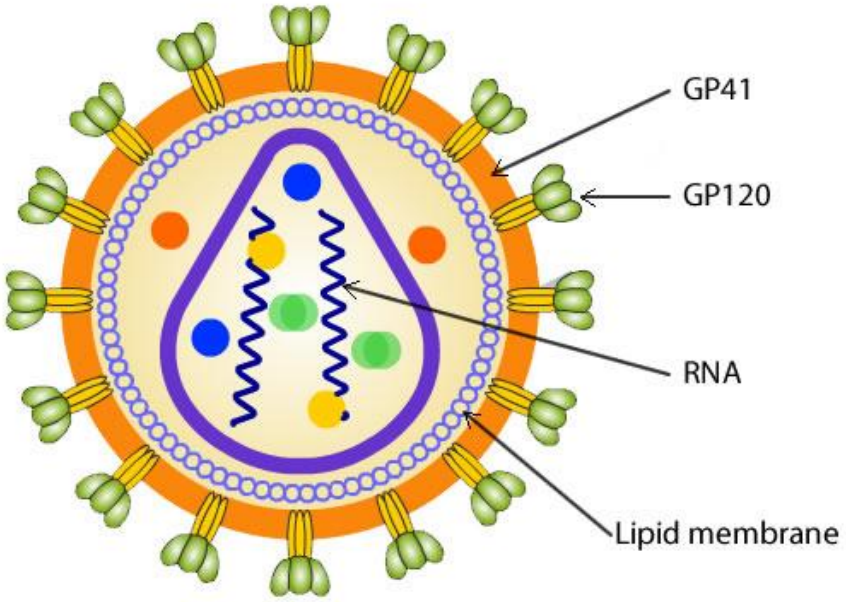


Figure-1.2: Schematic representation of a typical HIV-1 viral particle

1.2.1 Classification of HIV

HIV is notorious for high magnitude of genetic variation. HIV is divided into two types of viruses HIV-1 and HIV-2. Based on genetic variation, HIV-1 is further classified

into four distinct groups (M, N, O and P), and the group M into 9 molecular subtypes and numerous circulating recombinant forms (Kuiken *et al*, 2010) (Figure-1.3). The HIV-1 subtypes are unevenly distributed across the world, with the most widespread subtypes being subtypes C and B. Subtype C is predominant in southern and eastern African countries, India and Nepal. Subtype C is responsible for 50% of infections globally and more than 95% of the infections in India (UNAIDS 2013).

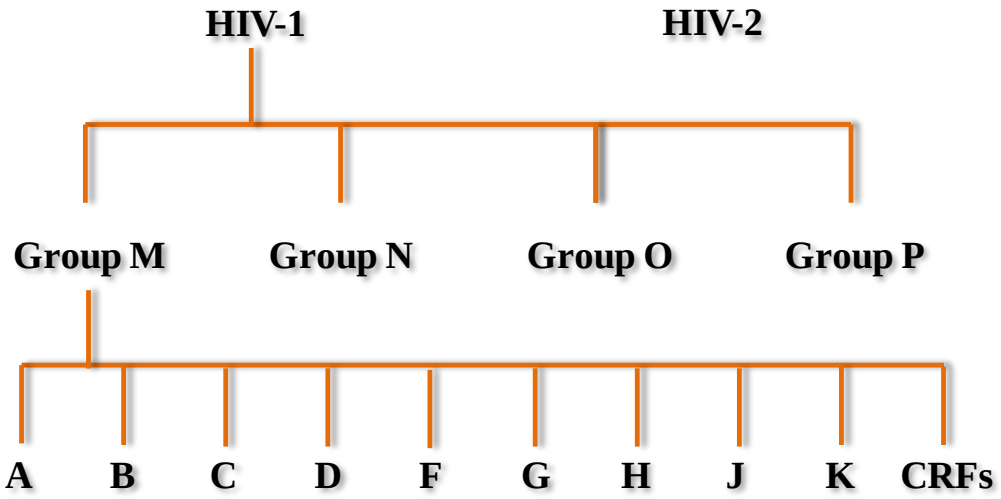


Figure-1.3: The phylogenetic classification of HIV-1

The propensity of subtype C viruses in establishing successful and widespread epidemics might be attributed partly to unique and intrinsic biological properties of these viral strains offering them a selective advantage. Differences in elements such as the viral promoter (enhancer and other regulatory elements), regulatory proteins and structural proteins deserve attention to examine such intrinsic biological properties of viral subtypes that might make them different from one another. The available experimental evidence is often conflicting and inadequate to evaluate whether the differences observed among various

subtypes at the molecular level could be extrapolated to differences in their biological properties.

1.3 Viral replication

HIV-1 contains two copies of the plus-strand RNA. A genome of approximately 9 kb length codes for at least 9 proteins Gag, Pol, Env, Tat, Rev, Nef, Vif, Vpu and Vpr. The CD4 T-lymphocytes and monocytes/macrophages constitute the primary target for the virus (Baxter *et al.* 2014;Nielsen *et al.* 2005). The dendritic cells also support the viral infection (Coleman and Wu, 2009;Manches *et al.* 2014;Wu and KewalRamani, 2006).

The viral infection starts with the interaction of gp120 (the trans-membrane protein of HIV-1) with the CD4 receptor on a T-cell (Kwong *et al.* 1998;Myszka *et al.* 2000). The interaction results in a conformational change in the envelope (Si *et al.* 2004), exposing the binding site for a chemokine co-receptor CXCR4 (X4 isolates) or CCR5 (R5 isolates). The X4 viruses replicate in the CD4 T-cells while the R5 tropic viral isolates infect the monocytic lineage (Schweighardt *et al.* 2004). After membrane fusion, the viral core undergoes a process of unpackaging releasing the contents including the RNA molecules that are converted into a double-stranded cDNA molecule by the reverse transcriptase (RT) (Alexaki *et al.* 2008;Liang *et al.* 2004). The viral RT uses the tRNA-lysine as primer to start the cDNA synthesis using the negative strand of RNA as template. The RT contains three important functions. First, the RNA-dependent DNA polymerase activity, which mediates the conversion of the RNA genome into cDNA. Second, the RNase H activity, which cleaves the RNA genome in the RNA-DNA hybrid. And lastly, the DNA-dependent DNA polymerase activity, which converts the single-stranded cDNA into double-stranded DNA (Nielsen *et al.* 2005). The cDNA, becoming a part of the pre-integration complex (PIC), is transported into

the nucleus by the mediation of multiple and redundant nuclear localization signals (Engelman, 2009;Raghavendra *et al.* 2010). Once in the nucleus, the viral DNA integrates in to the host cell genome by the mediation of viral integrase. Several anti-HIV drugs can inhibit the viral replication by targeting the various steps in HIV-1 lifecycle (Engelman, 2009;Engelman and Cherepanov, 2012;Nielsen *et al.* 2005;Raghavendra *et al.* 2010) (Figure-1.4).

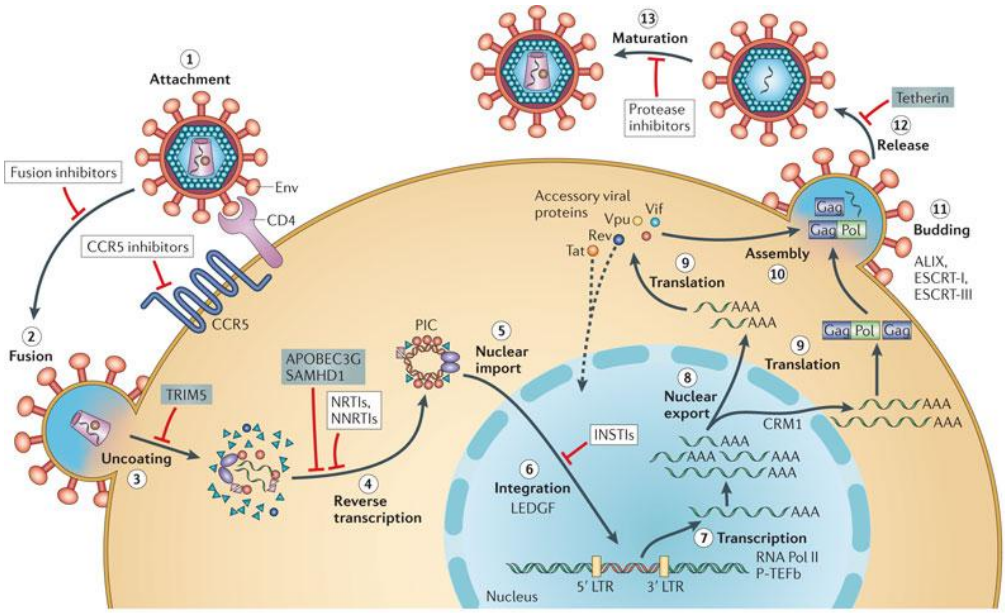


Figure-1.4: A schematic representation of the different stages of the HIV-1life cycle. Image adapted from doi:10.1038/nrmicro2747 with permission from nature publishing group.

1.4 The architecture of the viral promoter

The integrated form of HIV-1, also known as the provirus, is approximately 9.8 kb in length. Both ends of the provirus are flanked by a repeated sequence known as the long terminal repeats (LTR) that are identical genetically. The HIV-1 LTR is divided into three regions: U3, R and U5 containing four functional regions namely: the modulatory region (nt –454 to –

104), an enhancer (nt –105 to –79), a basal or core promoter (nt –78 to –1) and the trans-activation response (TAR) element found within R (nt +1 to +60) (Pereira *et al.* 2000) (Figure-1.5). The first three regions together constitute U3. The modulatory region also contains a negative regulatory element (NRE) between nt –340 and –184. The regulation of HIV-1 gene expression is accomplished by a combination of both cellular and viral factors and mediated by the single viral promoter, the 5'-LTR. HIV-1 gene expression is regulated at both the transcriptional and post-transcriptional levels. The regulation from the viral promoter is tightly controlled by the binding of several cellular transcription factors (TF) to a variety of cis-acting DNA sequences (transcription factor binding sites, TFBS) located within the LTR. These factors include NF-κB, Sp1, c-Myb, NF-AT, USF and many others (Pereira *et al.* 2000) (Figure-1.5).

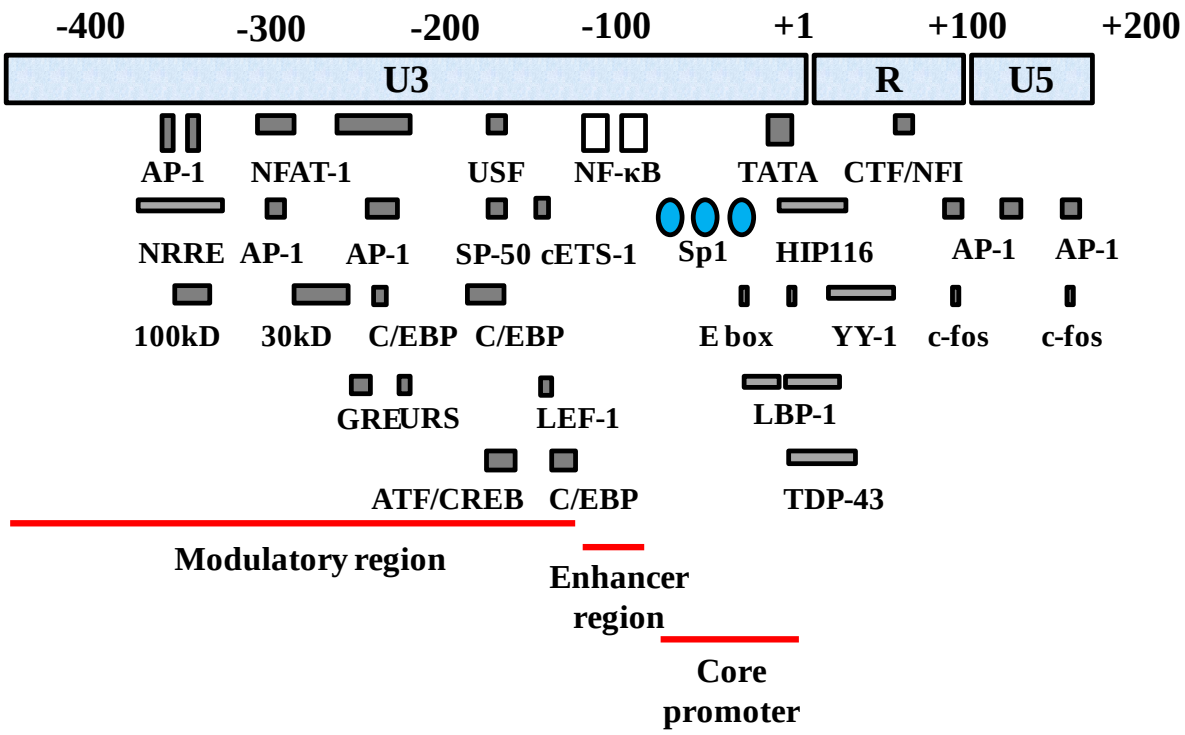


Figure-1.5: The architecture of the typical HIV-1 LTR demonstrating various transcription factor binding sites (TFBS)

The diverse viral subtypes can be phylogenetically classified into distinct groups based on the LTR sequences suggesting that the genetic differences identified in the LTR are non-random and such differences are likely to lead to different biological properties of viral subtypes. There is a considerable level of variation in the TFBS of the subtype LTRs. These variations could have a significant impact on diverse biological properties of the subtypes including the viral replication, infectivity, pathogenesis, viral fitness etc.

1.5 The NF-κB binding-site polymorphism in HIV-1 subtype-C LTR

A comparison of subtype C-LTR (C-LTR) with that of others identifies several distinct differences in the composition of regulatory elements and the TFBS such as the NF-κB, NF-AT, USF, the TATA box, and the TAR region (Jeeninga *et al.* 2000;Montano *et al.* 1997). The enhancer region of the LTR consists of only the NF-κB binding sites. The enhancer in all the HIV-1 subtypes contains two genetically identical NF-κB motifs (referred to here as the H-κB motif) with the exception of subtypes E and C. Subtype E-LTR contains a single H-κB motif (Ranjbar *et al.* 2006;Roof *et al.* 2002), while a large proportion of C-LTRs contains three or four NF-κB sites in contrast to the usual two sites seen in several other subtypes (Bachu *et al.* 2012a;Bachu *et al.* 2012b;Bjorndal *et al.* 1999;Munkanta *et al.* 2005;Novitsky *et al.* 2002). While the two canonical NF-κB motifs of C-LTR are identical to those present in subtype B-LTR (the H-κB motifs), the Sp1 proximal κB-site is characterized by a specific sequence variation unique for subtype C **(**Figure-1.6). We refer to this unique NF-κB site as the C-κB motif that is characterised with variations at positions 4 and 6 **(**Figure-1.6). A large number of subtype C-LTR also contain a fourth NF-κB motif (the F-κB motif), which differs from the canonical H-κB motif by one base-pair at position 10 (C to T)

(Figure-1.6). The prevalence of the four κB viral strains increased from 2% to 35% in last 10 years (Bachu *et al.* 2012a;Bachu *et al.* 2012b).

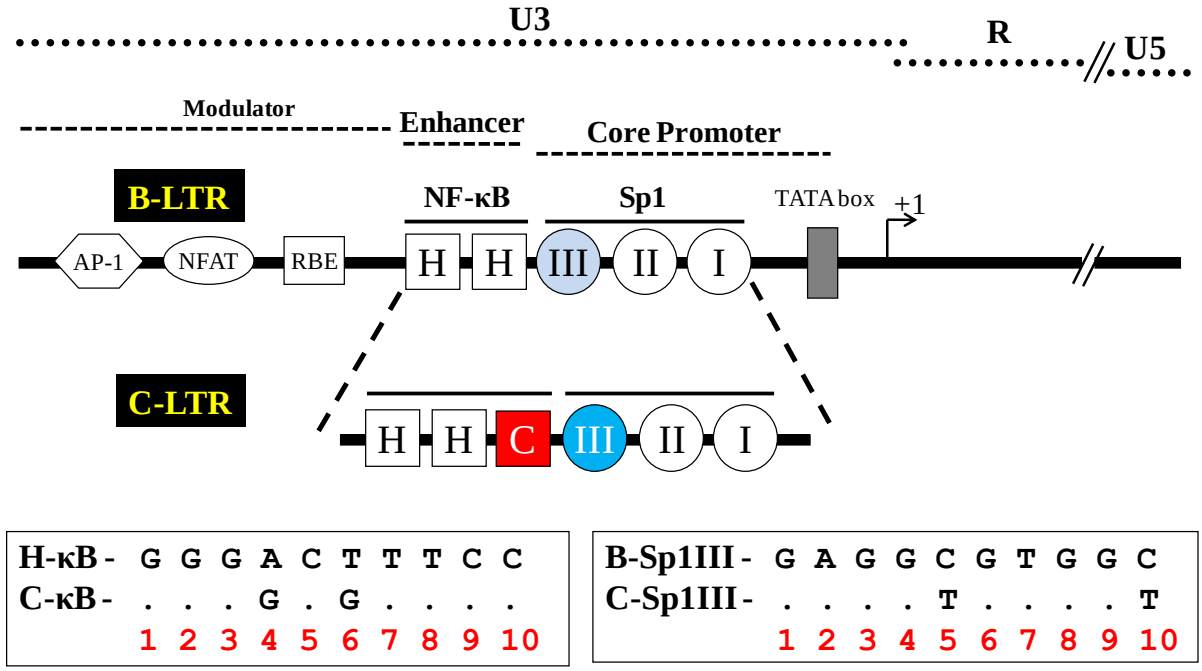


Figure-1.6: Schematic representation of the comparative analysis between the enhancer and core promoter elements of B- and C-LTRs. An outline of the prototype B LTR with basic transcription factors binding sites is shown. The enhancer region of the B-LTR contains two κB motifs (HH), in contrast to C-LTR that contains three κB motifs (HHC). The additional NF-κB motif is genetically distinct and is referred to here as the C-κB site. Note that the enhancer-proximal Sp1 site is also characterized with subtype-specific genetic variations.

1.5.1 The NF-κB family

The NF-κB was first identified as a transcription complex, regulating the expression of immunoglobulin (Ig) light chain gene in B cells (Sen and Baltimore, 1986). The NF-κB family, also called the rel family, consists of five individual members: p50, p52, p65 (RelA), c-Rel and RelB (Hayden and Ghosh , 2008). All the NF-κB members share a common amino-terminal region, called the Rel homology domain (RHD) that regulates many biological functions of the members including their retention in the cytoplasm by the IκB proteins, the nuclear localization, dimerization and the DNA binding properties. Three NF-κB

members- RelA, RelB and c-Rel contain a transcription activation domain (TAD) at the carboxyl-termius, which is necessary for the targeted gene expression (Chen and Greene, 2004;Rothwarf and Karin, 1999;Wan and Lenardo, 2009). With the exception of RelB that can form only hetero-dimers, other rel family members, can form homo- or hetero-dimers (Figure-1.7). Furthermore, the genetic degeneracy of the κB-binding sites adds an additional layer of complexity to gene expression regulation by conferring variable levels of specificity and affinity to the diverse NF-κB dimers (Natoli *et al.* 2005).

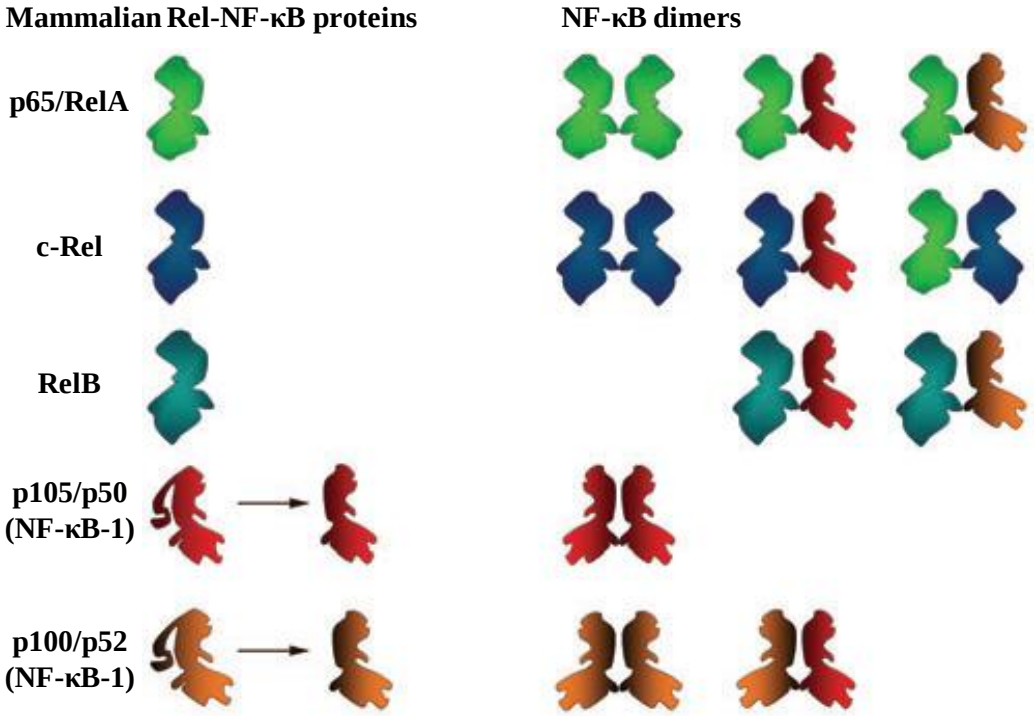


Figure-1.7: The rel family members and the biologically functional dimer protein complexes. Image adapted from doi:10.1038/ni1196 with permission from nature publishing group.

1.5.2 The regulation of the NF-κB signalling

The NF-κB signalling is tightly controlled by a family of NF-κB inhibitory proteins, called the IκB proteins. The IκB family consists of several members, having variable affinities for different NF-κB dimers (Figure-1.8) (Gilmore, 2006). The expression of the IκB proteins is tissue-specific.

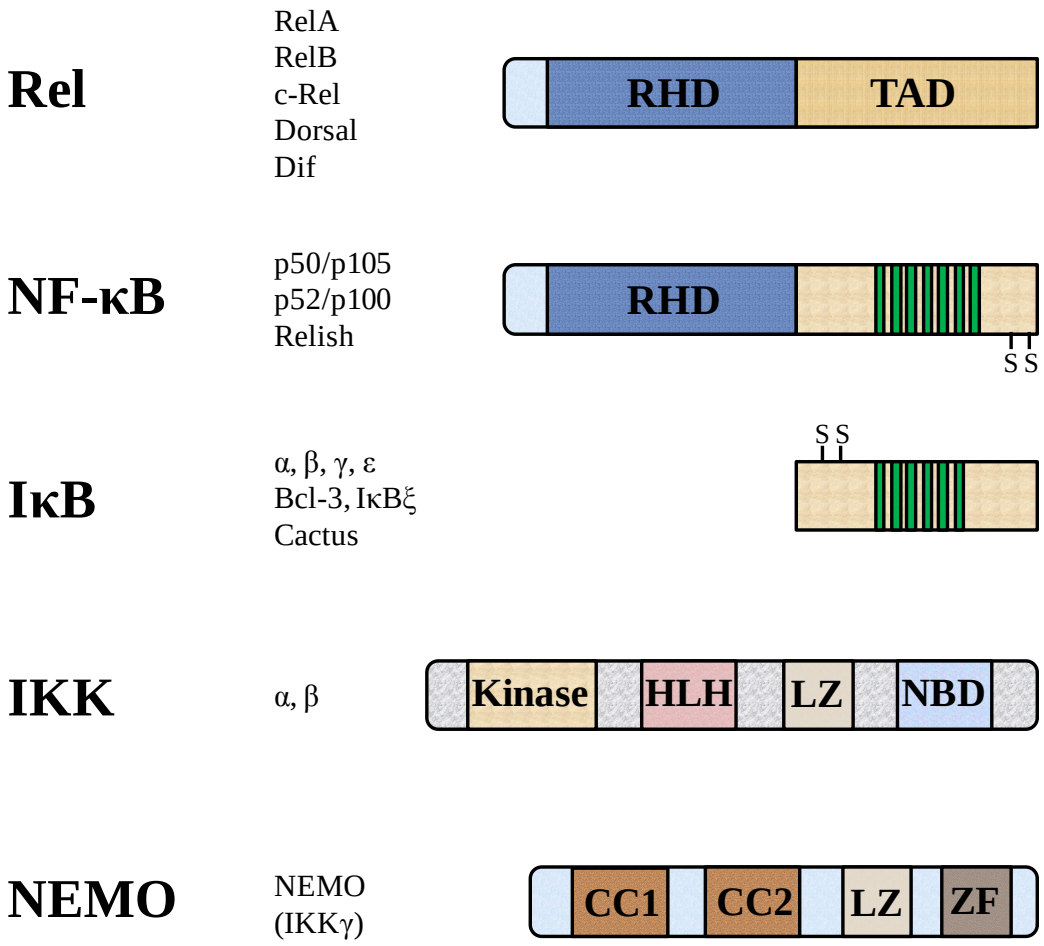


Figure-1.8: The structural domains of the core NF-κB signalling proteins

The NF-κB dimers are typically retained in the cytoplasm in an inactive state complexed with the IκB proteins. NF-κB activation can follow one of two different pathways, leading to the gene expression of the target genes: The canonical or classical pathway and the non-canonical or alternative pathway (Figure-1.9) (Gilmore and Herscovitch, 2006). Both the signalling pathways have an upstream element, the IκB kinase (IKK) complex in common. The IKK complex is made up of catalytic kinase subunits (IKKα and/or IKKβ) and a scaffold sensing protein, called NF-κB essential modulator (NEMO) (Figure-1.8). There are several known modification and subsequent degradation of the IκB members, which regulate the classical and alternative NF-κB signalling pathways (Table-1).

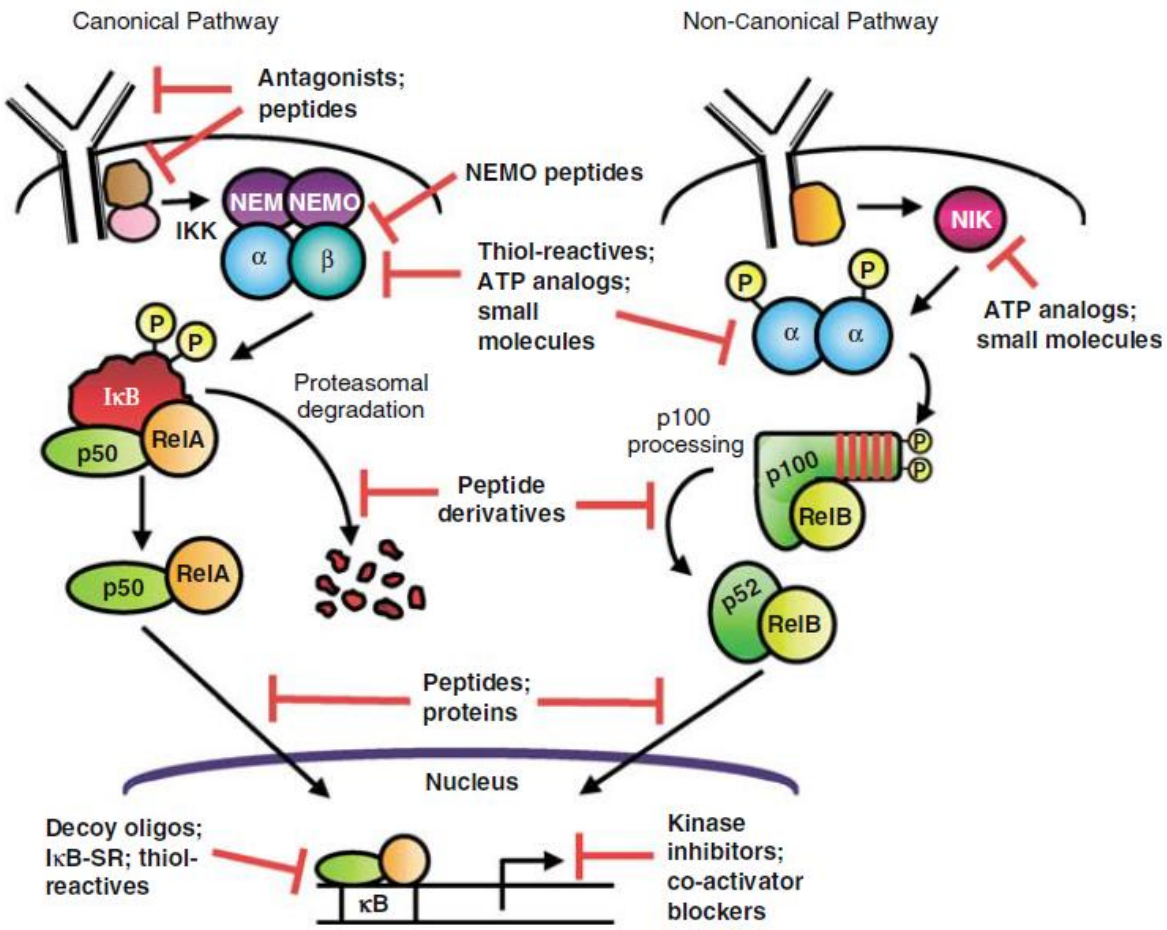


Figure-1.9: The NF-κB signal transduction pathway. Image adapted from doi:10.1038/sj.onc.1209982 with permission from nature publishing group.

Table-1: Post-translational modification of the IκB proteins

IκB family proteins	Target residues	Enzymes	References
Phosphorylation			
IκBα	S32 and S36	IKKβ	(Hayden and Ghosh 2004)
IκBα	S283, S289, T291, S293, and T299	CKII	(Lin et al. 1996; McElhinny et al. 1996; Schwarz et al. 1996)
IκBα	Y42	p56-lck, Syk, and c-Src	(Koong et al. 1994)
	Y42	Unknown	(Schoonbroodt et al. 2000)
	Y42 and Y305	Unknown	(Waris et al. 2003)
IκBα	Unknown	PI3K/Akt	(Sizemore et al. 1999)
IκBβ	S19 and S23	IKKβ	(Hayden and Ghosh 2004)
IκBβ	S313 and S315	CKII	(Chu et al. 1996)
IκBε	S18 and S22	IKKβ	(Hayden and Ghosh 2004)
Bcl-3	S394 and S398	GSK3β	(Viatour et al. 2004)
p100	S99, S108, S115, S123, S866, S870, and S872	IKKα	(Senftleben et al. 2001; Xiao et al. 2001; Xiao et al. 2004)
p105	S927 and S932	IKKβ	(Lang et al. 2003)
p105	S903 and S907	GSK3β	(Demarchi et al. 2003)
p105	S337	PKAc	(Hou et al. 2003)
Ubiquitination			
IκBα	K21 and K22	βTrCP	(Hayden and Ghosh 2004)
IκBβ	K6	βTrCP	(Hayden and Ghosh 2004)
IκBε	K6	βTrCP	(Hayden and Ghosh 2004)
p100	K856	βTrCP	(Amir et al. 2004)
p105	Multiple Ks	βTrCP	(Cohen et al. 2004)
Sumoylation			
IκBα	K21	Unknown	(Desterro et al. 1998)
p100	K90, K298, K689, and K863	Ubc9	(Vatsyayan et al. 2008)
Acetylation			
p100	Unknown	P300	(Hu and Colburn 2005; Deng et al. 2006)
p105	K431, K440, and K441	P300	(Furia et al. 2002; Deng and Wu 2003)
S-nitrosylation			
p105	C62		(Matthews et al. 1996)

IKKβ, IκB kinase β; CKII, Casein kinase II; p56-lck, lymphocyte-specific protein tyrosine kinase; Syk, spleen tyrosine kinase; c-Src, normal cellular Src kinase; PI3K, phosphoinositide 3-kinases; GSK3β, glycogen synthase kinase 3 β; IKKα, IκB kinase α; PKAc, catalytic subunit of protein kinase A; βTrCP, β-transducin repeat-containing protein; Ubc9, E2 small ubiquitin-like modifier (SUMO)-conjugating enzyme Ubc9; p300, E1A binding protein p300.

In brief, IKK complex modulates phosphorylation hence the degradation of the IκB proteins. This results in the release and translocation of the NF-κB dimers into the nucleus (Oeckinghaus and Ghosh, 2009;Sun and Zhang, 2007).

In the canonical pathway, the p50:p65 heterodimer plays a critical role while in the non-canonical pathway the p100:RelB complex is important.

1.6 The Sp1 binding-sites in HIV-1 subtype-C LTR

The core promoter of HIV-1 contains three GC rich motifs, referred to as the Sp1 binding sites. The consensus sequence for these sites is 5'-G Pu G G C Pu G G G N-3'. Of the three Sp1 sites, the site III located proximal to the viral enhancer is genetically distinct in HIV-1 subtypes (Figure-1.10).

```
Subtype C- G A G G T G T G G T
          B- . . . . C . . . . C
          A- . . . . C . . . . .
          D- . . . . C . . . . A
          E- . . . . . . . . . C
```

Figure-1.10: A comparison of the consensus sequence of the Sp1 III sites from the HIV-1 subtypes A, B, C, D and E.

1.6.1 The Sp1 transcription factors

Many cellular promoters contain widely distributed G-rich elements such as the GC-box (GGGGCGGGG) or a related GT/CACCC-box (GGTGTGGGG), recognized by the Sp1 family of transcription factors. Approximately 12,000 Sp1 binding sites are found throughout the human genome. The Sp1 transcription factors are involved in the regulation of various ubiquitous, tissue-specific and viral genes (Deniaud *et al.* 2009;Suske, 1999).

The Sp1 transcription factor was originally identified as a factor that activates transcription through multiple GC-rich elements in the SV40 promoter (Dynan and Tjian, 1983;Gidoni *et al.* 1984;Suske, 1999). The Sp1 family of transcription factors constitutes of three additional members Sp2, Sp3 and Sp4.

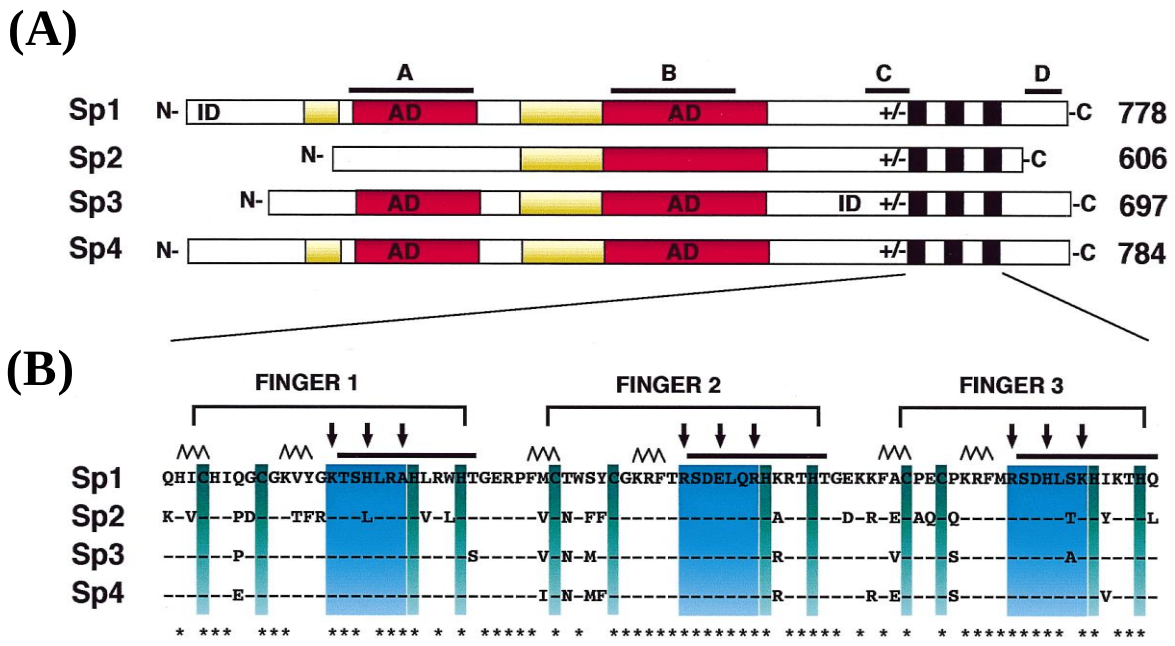


Figure-1.11: Schematic representation of the human Sp1 family members. **(A)** The Sp1 family members contain 4 distinct activation domains (A, B, C, and D) and three Zn-finger DNA binding domains. **(B)** Multiple sequence alignment of the amino acid residues of the three Zn-finger domains of the Sp family members. Image adapted from Suske, G. Gene. 238, 1999.

The members of the Sp1 family possess a similar domain structure (Figure-1.11 A). These proteins contain three Zinc-finger domains (black box) near the C-terminal region. The +/- region preceding the zinc-finger domains, is rich in charged amino acids. The red color box represents the glutamine-rich region and the yellow box the serine/threonine rich region. The activation and inhibitory domains of each Sp1 member are highlighted. All the Sp1 members contain four activation domains A, B, C and D. The domains A and B are rich in

glutamine residues while domain C contains weak basic amino acid residues. The D domain, the shortest of all, located at the carboxy-terminal with no specific recognizable structure assigned to it (Perkins *et al.* 1993). Three zinc-finger domains each made up of 81 amino-acids form the DNA-binding region. The C2-H2 type zinc-finger domains are the most conserved region across the all Sp members (Figure-1.11 B). The specific contact residues, KHA in the first zinc-finger, RER in the second zinc-finger and RHK in the third zinc-finger region are conserved in Sp1, Sp3 and Sp4, but not in Sp2 (Hagen *et al.* 1992;Hagen *et al.* 1994;Suske, 1999). In Sp2, the histidine residue in the first zinc-finger domain is replaced by a lucine (Figure-1.11 B), hence unlike the other members of the Sp1 family, Sp2 binds the GT-rich box, but not the GC-rich box.

A few functions and properties of Sp1 members are listed in the table-2 below:

Table-2: Functions and properties of Sp1 family members

Factor	Accession numbers	Chromosomal localisation	Distribution	Transcriptional properties	Knockout phenotype	Features/ miscellaneous
Sp1	Human: J03133; Mouse: AF062566, AF022363; Rat: D12768	Human: 12q13 (Gaynor et al., 1993; Matera and Ward, 1993); Mouse: 15 (Saffer et al., 1990); Rat: 7q36 (Scohy et al., 1998)	Ubiquitous, developmental variations (Saffer et al., 1991)	Activator (Courey and Tjian, 1988) Synergistic activation (Courey et al., 1989) ; Superactivation (Pascal and Tjian, 1991)	Lethal at embryonic day 10 (Marin et al., 1997)	Two glutamine-rich activation domains (Courey and Tjian, 1988; Gill et al., 1994); phosphorylated (Jackson et al., 1990); glycosylated (Jackson and Tjian, 1988)
Sp2	Human: M97190, D28588	Human: 17q21.3-q22 (Scohy et al., 1998); Rat: 10q31-q32.1 (Scohy et al., 1998);	Various cell lines; tissues unknown (Kingsley and Winoto, 1992)	Unknown	Unknown	Original Sp2 clone incomplete; complete sequence in data bank D28588
Sp3	Human: X68560, S52144; Mouse: AF062567	Human: 2q31 (Kalff-Suske et al., 1996); Rat: 3q24-q31 (Scohy et al., 1998)	Ubiquitous (Hagen et al., 1992) and G. Suske, unpublished	Repressor of Sp1-mediated transcription (Hagen et al., 1994) Activator (Udvadia et al., 1995; Dennig et al., 1996)	Unknown	Two glutamine-rich activation domains; three isoforms (Hagen et al., 1994; Kennett et al., 1997); inhibitory domain (Dennig et al., 1996); translational start site of full length protein unknown
Sp4	Human: X68561, S50516; Mouse: U62522; Rat: U07610	Human: 7p15.3-p21 (Kalff-Suske et al., 1995); Mouse: 12 (Supp et al., 1996); Rat: 6q33 (Scohy et al., 1998)	Predominantly in neuronal cells; also in certain epithelia (Hagen et al., 1992; Supp et al., 1996) and G. Suske, unpublished	Activator (Hagen et al., 1994, 1995)	Growth retardation; males do not breed (Supp et al., 1996)	Two glutamine-rich activation domains (Hagen et al., 1995); entire human genomic sequence in data bank: Accession No. AC004595

While Sp1 and Sp4 predominantly function as mediators of transcription activation, Sp3 represses the Sp-1 mediated transcription although on its own it is an activator. The function of Sp2 is not completely understood.

1.7 HIV-1 transcription

Immediately after HIV-1 infection, only short and fully-spliced RNAs for regulatory proteins (Tat and Rev) are produced. As the infection progresses, larger and incompletely spliced RNA (coding for env, vif, vpr and vpu) are produced. Later, full-length and unspliced RNA (function as the genomic RNA or coding for Gag-Pol poly protein, respectively) are synthesised (Kim *et al.* 1989;Pomerantz *et al.* 1990).

1.7.1 The role of Tat in HIV-1 transcription

Tat is an HIV-1 regulatory protein, which plays a critical role in HIV-1 gene-expression. In the absence of Tat, most of the RNA polymerases stall near the transcription start site leading to the generation of abortive transcripts. Presence of Tat helps the RNA polymerase to read-through the promoter and generate longer transcripts. Tat is associated with TAK (Tat-associated kinase) (Herrmann and Rice ,1995;Yang *et al.* 1996). The kinase subunit of TAK is called CDK9, which is a component of pTEFb (positive transcription elongation factor). Another component of pTEFb complex (CyclinT1) is also found associated with the CDK9, Tat and TAR RNA complex (Figure-1.12) (Karn and Stoltzfus ,2012;Wei *et al.* 1998).

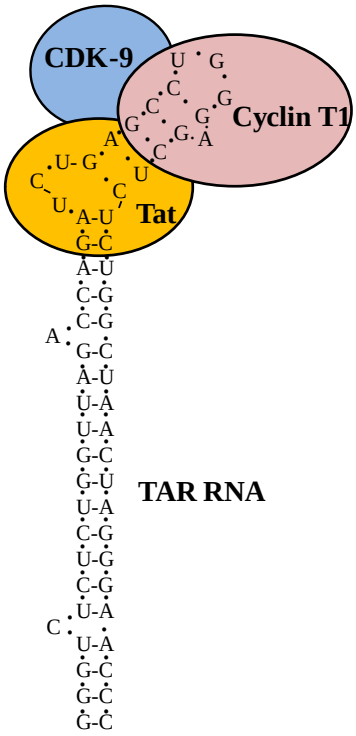


Figure-1.12: A schematic representation of Tat-TAR RNA and pTEFb complex

In the absence of Tat, transcription initiation is strongly induced by the NF-κB factors by recruiting the histone acetyltransferases (HAT). Once TAR element is transcribed, both Tat/pTEFb complex and NELF (negative elongation factor) are recruited to the TAR RNA. This interaction activates the CDK9 kinase (Isel and Karn, 1999) and leads to the hyperphosphorylation of the carboxy-terminal domain (CTD) of RNA polymerase II and NELF. In the absence of Tat, HIV-1 transcription is restricted by NELF (Yamaguchi *et al.* 1999). Phosphorylation of NELF results in its dissociation from the elongation complex. Hyperphosphorylation of RNA polymerase II leads to the increased transcription of the full-length HIV-1 genome.

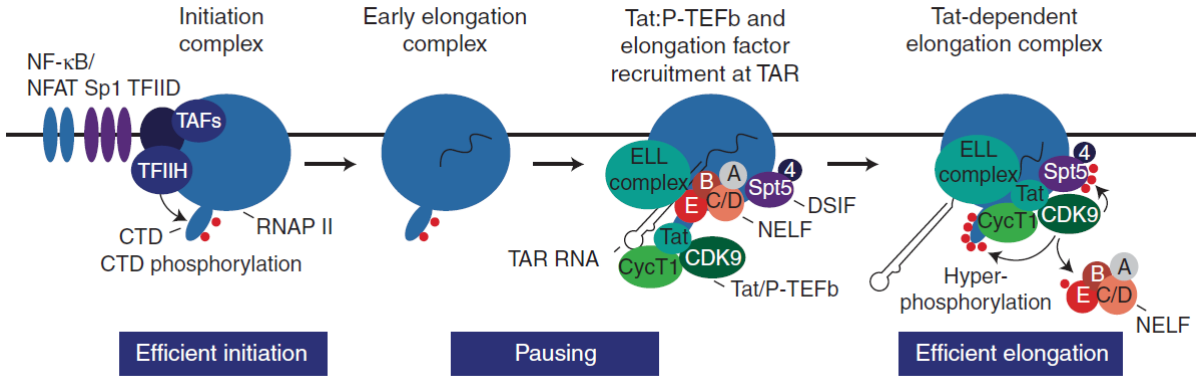


Figure-1.13: NF-κB and Tat activated transcription of HIV-1 provirus. Image adapted from Karn J, *et al*, 2012.

1.7.2 Epigenetic regulation of HIV-1 gene expression

The integration of the proviral DNA into the host genome is a critical step for all retroviruses. Following integration, the viral genome undergoes nucleosomal reorganization which in turn depends on the nature of the surrounding host chromatin. HIV-1 gene expression could also be modulated by the site of the proviral DNA integration. The chromatin remodelling is majorly regulated by post-translational modification of the histones including the acetylation, deacetylation, methylation, phosphorylation, ubiquitylation and sumoylation. The acetylation and deacetylation of the histones are mediated by histone acetyl transferases and histone deacetylases (HDAC), respectively (Colin and Van, 2009). HATs catalyses the decondensation of the chromatin and hence help in the maintenance of active euchromatin, whereas HDACs induce condensation of the chromatin hence maintain the chromatin in repressive heterochromatin forms. Based on the sequence homology and biological functions, HATs are classifieds into different groups: Gcn5-related N-acetyltransferases (GNATs), p300/CBP and the MYST proteins families. HDACs are

categorized into three classes: Class I (HDACs 1, 2, 3 and 8), class II (HDACs 4, 5, 6, 7, 9 and 10) and class III (Sirt1 – Sirt7). Methylation of the chromatin is controlled by histone lysine methyl transferases (HKMTs) and protein arginine methyl transferases (PRMTs). Role of p300/CBP and HDACs 1 and 3 has been majorly studied with respect to HIV-1 transcription (Barton *et al*, 2014;Burnett *et al*, 2009;Marzio *et al*, 1998).

Following integration, the 5' LTR of HIV-1 forms two different nucleosomes, nuc-0 and nuc-1, flanking the viral enhancer region (He *et al.* 2002;Marcello *et al.* 2001;Widlak and Garrard , 1998). Nuc-0 is positioned between -415 to -255 and nuc-1 is located upstream of transcription start site (+10 to +165) (Figure-1.14) (Verdin *et al.* 1993). Several epigenetic modifications play a critical role in the regulation of HIV-1 gene expression. For efficient HIV-1 transcription, nuc-1 remodelling is essential. The remodelling of nuc-1 could be accentuated by the recruitment of HATs by activating transcription factors and/or by Tat. The HIV-1 transcription depends on the dynamic interplay between several transcription factors (activating or repressing), HATs, HDACs, PRMTs and HKMTs.

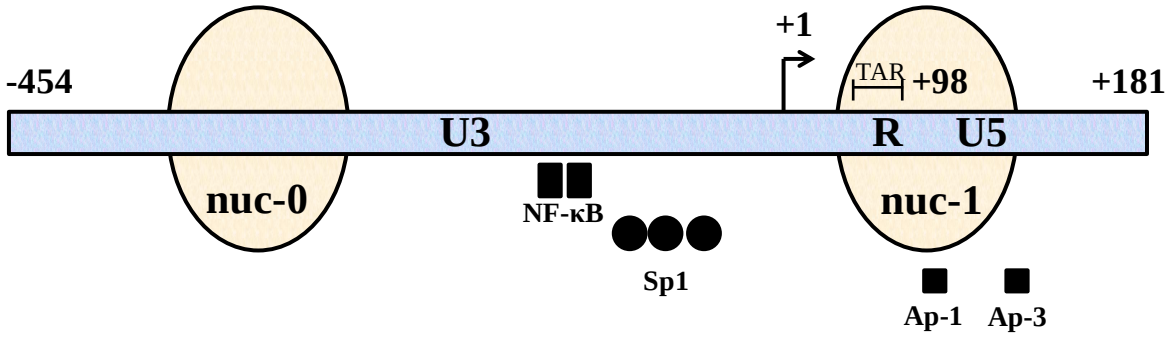


Figure-1.14: Positioning of nucleosome over the HIV-1 LTR

1.8 HIV-1 and latency

The persistence of the latent reservoir is the major barrier in HIV-1 cure. Based on the integration of the virus genome into host chromosome, latency can be classified into two categories: pre- and post-integration latency (Abbas and Herbein , 2012). Partial or complete blockade of the viral life cycle prior to integration of the provirus into the host genome leads to preintegration latency (Han *et al*, 2004). Incomplete reverse transcription or restriction by host-cellular factors such as APOBEC3G can lead to the formation of postintegration latency (Pierson *et al*, 2002). Further, the preintegration latency is not of much clinical significance as the unintegrated forms of viral DNA cannot persist in the cytoplasm for more than 24 h hence cannot account for the long-term latentcy. The integrated form of DNA, however, can remain stable for extended periods in nondividing and metabolically inactive macrophages. Postintegration latency occurs when the viral DNA is integrated into the host genome but remains reversibly silenced transcriptionally (Abbas and Herbein, 2012).

1.9 Bibliography

Abbas, W., G. Herbein, 2012. Molecular Understanding of HIV-1 Latency. Adv.Virol. 2012: 574967. Available from PM:22548060

Alexaki, A., Y. Liu and B. Wigdahl, 2008. Cellular reservoirs of HIV-1 and their role in viral persistence. Curr.HIV.Res. 6: 388-400. Available from PM:18855649

Bachu, M., A.B. Mukthey, R.V. Murali, N. Cheedarla, A. Mahadevan, S.K. Shankar, K.S. Satish, T.K. Kundu and U. Ranga, 2012a. Sequence insertions in the HIV type 1 subtype C viral promoter predominantly generate an additional NF-kappaB binding site. AIDS Res.Hum.Retroviruses 28: 1362-1368. Available from PM:22332607

Bachu, M., S. Yalla, M. Asokan, A. Verma, U. Neogi, S. Sharma, R.V. Murali, A.B. Mukthey, R. Bhatt, S. Chatterjee, R.E. Rajan, N. Cheedarla, V.S. Yadavalli, A. Mahadevan, S.K. Shankar, N. Rajagopalan, A. Shet, S. Saravanan, P. Balakrishnan, S. Solomon, M. Vajpayee, K.S. Satish, T.K. Kundu, K.T. Jeang and U. Ranga, 2012b. Multiple NF-kappaB sites in HIV-1 subtype C long terminal repeat confer superior magnitude of transcription and thereby the enhanced viral predominance. J.Biol.Chem. 287: 44714-44735. Available from PM:23132857

Barton,K.M.; Archin,N.M.; Keedy,K.S.; Espeseth,A.S.; Zhang,Y.L.; Gale,J.; Wagner,F.F.; Holson,E.B.; D.M.Margolis, 2014, Selective HDAC inhibition for the disruption of latent HIV-1 infection, PLoS. One.9 (8). Avaliable from PM:25136952

Baxter, A.E., R.A. Russell, C.J. Duncan, M.D. Moore, C.B. Willberg, J.L. Pablos, A. Finzi, D.E. Kaufmann, C. Ochsenbauer, J.C. Kappes, F. Groot and Q.J. Sattentau, 2014. Macrophage infection via selective capture of HIV-1-infected CD4+ T cells. Cell Host.Microbe 16: 711-721. Available from PM:25467409

Bjorndal, A., A. Sonnerborg, C. Tscherning, J. Albert and E.M. Fenyo, 1999. Phenotypic characteristics of human immunodeficiency virus type 1 subtype C isolates of Ethiopian AIDS patients. AIDS Res.Hum.Retroviruses 15: 647-653. Available from PM:10331443

Burnett, J.C., K. Miller-Jensen, P.S. Shah, A.P. Arkin and D.V. Schaffer, 2009. Control of stochastic gene expression by host factors at the HIV promoter. PLoS.Pathog. 5: e1000260. Available from PM:19132086

Chen, L.F., W.C. Greene, 2004. Shaping the nuclear action of NF-kappaB. Nat.Rev.Mol.Cell Biol. 5: 392-401. Available from PM:15122352

Coffin, J.M., 1992. Retroviral DNA integration. Dev.Biol.Stand. 76: 141-151. Available from PM:1478335

Coleman, C.M., L. Wu, 2009. HIV interactions with monocytes and dendritic cells: viral latency and reservoirs. Retrovirology. 6: 51. Available from PM:19486514

Colin, L., L.C. Van, 2009. Molecular control of HIV-1 postintegration latency: implications for the development of new therapeutic strategies. Retrovirology. 6: 111. Available from PM:19961595

Deniaud, E., J. Baguet, R. Chalard, B. Blanquier, L. Brinza, J. Meunier, M.C. Michallet, A. Laugraud, C. Ah-Soon, A. Wierinckx, M. Castellazzi, J. Lachuer, C. Gautier, J. Marvel and Y. Leverrier, 2009. Overexpression of transcription factor Sp1 leads to gene expression perturbations and cell cycle inhibition. PLoS.One. 4: e7035. Available from PM:19753117

Dynan, W.S., R. Tjian, 1983. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell 35: 79-87. Available from PM:6313230

Engelman, A., 2009. Isolation and analysis of HIV-1 preintegration complexes. Methods Mol.Biol. 485: 135-149. Available from PM:19020823

Engelman, A., P. Cherepanov, 2012a. The structural biology of HIV-1: mechanistic and therapeutic insights. Nat.Rev.Microbiol. 10: 279-290. Available from PM:22421880

Gidoni, D., W.S. Dynan and R. Tjian, 1984. Multiple specific contacts between a mammalian transcription factor and its cognate promoters. Nature 312: 409-413. Available from PM:6095100

Gilmore, T.D., 2006. Introduction to NF-kappaB: players, pathways, perspectives. Oncogene 25: 6680-6684. Available from PM:17072321

Gilmore, T.D., M. Herscovitch, 2006. Inhibitors of NF-kappaB signaling: 785 and counting. Oncogene 25: 6887-6899. Available from PM:17072334

Hagen, G., S. Muller, M. Beato and G. Suske, 1992. Cloning by recognition site screening of two novel GT box binding proteins: a family of Sp1 related genes. Nucleic Acids Res. 20: 5519-5525. Available from PM:1454515

-----, 1994. Sp1-mediated transcriptional activation is repressed by Sp3. EMBO J. 13: 3843-3851. Available from PM:8070411

Hayden, M.S., S. Ghosh, 2008. Shared principles in NF-kappaB signaling. Cell 132: 344-362. Available from PM:18267068

Han,Y.; Lassen,K.; Monie,D.; Sedaghat,A.R.; Shimoji,S.; Liu,X.; Pierson,T.C.; Margolick,J.B.; Siliciano,R.F.; and J.D.Siliciano,2004, Resting CD4+ T cells from human immunodeficiency virus type 1 (HIV-1)-infected individuals carry integrated HIV-1 genomes within actively transcribed host genes. J. Virol. 78: 6122-6133. Available from PM:15163705

He, G., L. Ylisastigui and D.M. Margolis, 2002. The regulation of HIV-1 gene expression: the emerging role of chromatin. DNA Cell Biol. 21: 697-705. Available from PM:12443539

Herrmann, C.H., A.P. Rice, 1995. Lentivirus Tat proteins specifically associate with a cellular protein kinase, TAK, that hyperphosphorylates the carboxyl-terminal domain of the

large subunit of RNA polymerase II: candidate for a Tat cofactor. J.Virol. 69: 1612-1620. Available from PM:7853496

Isel, C., J. Karn, 1999. Direct evidence that HIV-1 Tat stimulates RNA polymerase II carboxyl-terminal domain hyperphosphorylation during transcriptional elongation. J.Mol.Biol. 290: 929-941. Available from PM:10438593

Jeeninga, R.E., M. Hoogenkamp, M. rmand-Ugon, B.M. de, K. Verhoef and B. Berkhout, 2000. Functional differences between the long terminal repeat transcriptional promoters of human immunodeficiency virus type 1 subtypes A through G. J.Virol. 74: 3740-3751. Available from PM:10729149

Karn, J., C.M. Stoltzfus, 2012. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harb.Perspect.Med. 2: a006916. Available from PM:22355797

Kim, S.Y., R. Byrn, J. Groopman and D. Baltimore, 1989. Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. J.Virol. 63: 3708-3713. Available from PM:2760980

Kuiken, C., Leitner, T., Hahn, B.,Mullins, J., Wolinsky, S., Foley, B., Apetrei, C., Mizrachi, Y., Rambait, A., and Korber, B., 2010. HIV sequence compemndium 2010.

Kwong, P.D., R. Wyatt, J. Robinson, R.W. Sweet, J. Sodroski and W.A. Hendrickson, 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 393: 648-659. Available from PM:9641677

Lee, Y.M., J.M. Coffin, 1990. Efficient autointegration of avian retrovirus DNA in vitro. J.Virol. 64: 5958-5965. Available from PM:2173775

Liang, C., J. Hu, R.S. Russell, M. Kameoka and M.A. Wainberg, 2004. Spliced human immunodeficiency virus type 1 RNA is reverse transcribed into cDNA within infected cells. AIDS Res.Hum.Retroviruses 20: 203-211. Available from PM:15018708

Manches, O., D. Frleta and N. Bhardwaj, 2014. Dendritic cells in progression and pathology of HIV infection. Trends Immunol. 35: 114-122. Available from PM:24246474

Marcello, A., M. Zoppe and M. Giacca, 2001. Multiple modes of transcriptional regulation by the HIV-1 Tat transactivator. IUBMB.Life 51: 175-181. Available from PM:11547919

Montano, M.A., V.A. Novitsky, J.T. Blackard, N.L. Cho, D.A. Katzenstein and M. Essex, 1997. Divergent transcriptional regulation among expanding human immunodeficiency virus type 1 subtypes. J.Virol. 71: 8657-8665. Available from PM:9343223

Marzio,G.; Tyagi,M.; Gutierrez,M.I.; and M. Giacca, 1998. HIV-1 tat transactivator recruits p300 and CREB-binding protein histone acetyltransferases to the viral promoter. Proc.Natl.Acad.Sci.U.S.A 95: 13519-13524. Available from PM:9811832.

Munkanta, M., R. Handema, H. Kasai, C. Gondwe, X. Deng, A. Yamashita, T. Asagi, N. Yamamoto, M. Ito, F. Kasolo and H. Terunuma, 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: 901-906. Available from PM:16225420

Myszka, D.G., R.W. Sweet, P. Hensley, M. Brigham-Burke, P.D. Kwong, W.A. Hendrickson, R. Wyatt, J. Sodroski and M.L. Doyle, 2000. Energetics of the HIV gp120-CD4 binding reaction. Proc.Natl.Acad.Sci.U.S.A 97: 9026-9031. Available from PM:10922058

Natoli, G., S. Saccani, D. Bosisio and I. Marazzi, 2005. Interactions of NF-kappaB with chromatin: the art of being at the right place at the right time. Nat.Immunol. 6: 439-445. Available from PM:15843800

Nielsen, M.H., F.S. Pedersen and J. Kjems, 2005b. Molecular strategies to inhibit HIV-1 replication. Retrovirology. 2: 10. Available from PM:15715913

Novitsky, V., U.R. Smith, P. Gilbert, M.F. McLane, P. Chigwedere, C. Williamson, T. Ndung'u, I. Klein, S.Y. Chang, T. Peter, I. Thior, B.T. Foley, S. Gaolekwe, N. Rybak, S. Gaseitsiwe, F. Vannberg, R. Marlink, T.H. Lee and M. Essex, 2002. Human immunodeficiency virus type 1 subtype C molecular phylogeny: consensus sequence for an AIDS vaccine design? J.Virol. 76: 5435-5451. Available from PM:11991972

Oeckinghaus, A., S. Ghosh, 2009. The NF-kappaB family of transcription factors and its regulation. Cold Spring Harb.Perspect.Biol. 1: a000034. Available from PM:20066092

Pereira, L.A., K. Bentley, A. Peeters, M.J. Churchill and N.J. Deacon, 2000. A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. Nucleic Acids Res. 28: 663-668. Available from PM:10637316

Perkins, N.D., N.L. Edwards, C.S. Duckett, A.B. Agranoff, R.M. Schmid and G.J. Nabel, 1993. A cooperative interaction between NF-kappa B and Sp1 is required for HIV-1 enhancer activation. EMBO J. 12: 3551-3558. Available from PM:8253080

Pierson,T.C.; Zhou,Y.; Kieffer,T.L.; Ruff,C.T.; Buck,C.; and R.F.Siliciano, 2000. Molecular characterization of preintegration latency in human immunodeficiency virus type 1 infection. J. Virol. 76: 8518-8531. Available from PM:12163571

Pomerantz, R.J., D. Trono, M.B. Feinberg and D. Baltimore, 1990. Cells nonproductively infected with HIV-1 exhibit an aberrant pattern of viral RNA expression: a molecular model for latency. Cell 61: 1271-1276. Available from PM:2364429

Raghavendra, N.K., N. Shkriabai, R.L. Graham, S. Hess, M. Kvaratskhelia and L. Wu, 2010. Identification of host proteins associated with HIV-1 preintegration complexes isolated from infected CD4+ cells. Retrovirology. 7: 66. Available from PM:20698996

Ranjbar, S., R. Rajsbaum and A.E. Goldfeld, 2006. Transactivator of transcription from HIV type 1 subtype E selectively inhibits TNF gene expression via interference with chromatin remodeling of the TNF locus. J.Immunol. 176: 4182-4190. Available from PM:16547255

Roof, P., M. Ricci, P. Genin, M.A. Montano, M. Essex, M.A. Wainberg, A. Gatignol and J. Hiscott, 2002. Differential regulation of HIV-1 clade-specific B, C, and E long terminal repeats by NF-kappaB and the Tat transactivator. Virology 296: 77-83. Available from PM:12036319

Rothwarf, D.M., M. Karin, 1999. The NF-kappa B activation pathway: a paradigm in information transfer from membrane to nucleus. Sci.STKE. 1999: RE1. Available from PM:11865184

Schweighardt, B., A.M. Roy, D.A. Meiklejohn, E.J. Grace, W.J. Moretto, J.J. Heymann and D.F. Nixon, 2004. R5 human immunodeficiency virus type 1 (HIV-1) replicates more efficiently in primary CD4+ T-cell cultures than X4 HIV-1. J.Virol. 78: 9164-9173. Available from PM:15308712

Sen, R., D. Baltimore, 1986. Inducibility of kappa immunoglobulin enhancer-binding protein Nf-kappa B by a posttranslational mechanism. Cell 47: 921-928. Available from PM:3096580

Shida, H., 2012. Role of Nucleocytoplasmic RNA Transport during the Life Cycle of Retroviruses. Front Microbiol. 3: 179. Available from PM:22783232

Si, Z., N. Madani, J.M. Cox, J.J. Chruma, J.C. Klein, A. Schon, N. Phan, L. Wang, A.C. Biorn, S. Cocklin, I. Chaiken, E. Freire, A.B. Smith, III and J.G. Sodroski, 2004. Small-molecule inhibitors of HIV-1 entry block receptor-induced conformational changes in the viral envelope glycoproteins. Proc.Natl.Acad.Sci.U.S.A 101: 5036-5041. Available from PM:15051887

Sun, X.F., H. Zhang, 2007. NFKB and NFKBI polymorphisms in relation to susceptibility of tumour and other diseases. Histol.Histopathol. 22: 1387-1398. Available from PM:17701919

Suske, G., 1999. The Sp-family of transcription factors. Gene 238: 291-300. Available from PM:10570957

Verdin, E., P. Paras, Jr. and L.C. Van, 1993. Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation. EMBO J. 12: 3249-3259. Available from PM:8344262

Wan, F., M.J. Lenardo, 2009. Specification of DNA binding activity of NF-kappaB proteins. Cold Spring Harb.Perspect.Biol. 1: a000067. Available from PM:20066093

Wei, P., M.E. Garber, S.M. Fang, W.H. Fischer and K.A. Jones, 1998. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. Cell 92: 451-462. Available from PM:9491887

Weiss, R.A., 1996. Retrovirus classification and cell interactions. J.Antimicrob.Chemother. 37 Suppl B: 1-11. Available from PM:8818825

Widlak, P., W.T. Garrard, 1998. Nucleosomes and regulation of gene expression. Structure of the HIV-1 5'LTR. Acta Biochim.Pol. 45: 209-219. Available from PM:9701513

Wu, L., V.N. KewalRamani, 2006. Dendritic-cell interactions with HIV: infection and viral dissemination. Nat.Rev.Immunol. 6: 859-868. Available from PM:17063186

Yamaguchi, Y., T. Takagi, T. Wada, K. Yano, A. Furuya, S. Sugimoto, J. Hasegawa and H. Handa, 1999. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell 97: 41-51. Available from PM:10199401

Yang, X., C.H. Herrmann and A.P. Rice, 1996. The human immunodeficiency virus Tat proteins specifically associate with TAK in vivo and require the carboxyl-terminal domain of RNA polymerase II for function. J.Virol. 70: 4576-4584. Available from PM:8676484

CHAPTER-2

MATERIALS AND METHODS

2.1 Cell Culture

T-cell lines (Jurkat and CEM-CCR5 cells) were cultured in RPMI 1640 medium (Cat. no. R4130, Sigma, St. Louis, USA), supplemented with 10% fetal bovine serum (RM10435, HiMedia Laboratories), 2 mM glutamine (G8540, Sigma, St. Louis, USA), 100 units/ml penicillin G (P3032, Sigma, St. Louis, USA), and 100 μg/ml streptomycin (S9137, Sigma St. Louis, United States). The peripheral blood mononuclear cells (PBMC) were isolated from 10 ml of fresh blood, collected from healthy donors, by density-gradient centrifugation using HiSep LSM-1077 (LS001, HiMedia Laboratories, Mumbai, India). The CD8 cells were depleted from the PBMC using the StemSep Human $CD8^+$ depletion kit (Cat. No. 14662, Stem Cell Technologies, Vancouver, Canada). PBMC were cultured in complete RPMI medium supplemented with 20 U/ml of IL-2 (H7041, Sigma, St. Louis, USA) and 5 μg/ml of PHA-P (L1668, Sigma, St. Louis, USA) for 72 h. PBMC were subsequently used either for nucleofection or viral infection in a medium supplemented with 20 U/ml IL-2 (H7041, Sigma, St. Louis, USA). The human embryonic kidney 293T and 293 cells were grown in complete Dulbecco's modified Eagle's medium (D1152, Sigma St. Louis, USA).

2.2 Construction of the reporter vectors and viral molecular clones

A dual-reporter vector pLTR-sLuc-IRES-GFP that simultaneously expresses two different reporter genes - secreted Gaussia Luciferase (sLuc) and green fluorescent protein (GFP) - was constructed as reported previously from the laboratory (Bachu *et al.* 2012;Siddappa *et al.* 2007). The CMV promoter was subsequently replaced by the full-length HIV-1 subtype C-LTR sequence consisting of 634 bp, amplified from the full-length HIV-1 subtype C molecular clone Indie C1 (Gen Bank Acc No: AB023804) using primers N1515 (5'-

GGATCC*ACGCGT*TGGAAGGGTTAATTTACTG-3') and N1359 (5'-CCGCGG*GAATTC*CTGCTAGAGATTTTCC-3'), between the restriction sites MluI and *Eco*RI. A series of isogenic variant LTR reporter vectors was generated from this parental vector using the overlap PCR strategy. Transcription factor binding sites were inactivated by mutating critical residues using site-directed mutagenesis including the H-κB binding site (GGGACTTTCC to TCTACTTTCC), the C-κB binding site (GGGGCGTTCC to GTCTCGTTCC), Sp1III site (GAGGTGTGGT to GATTTGTGGT), Sp1II site (TGGGCGGGAC to TGTTCGGGAC) and Sp1I site (TGGGAGTGGT to TGTTAGTGGT).

A panel of minimal LTR reporter vectors expressing luciferase was constructed by deleting the 313 bp upstream modulator region from the viral promoter using the BspEI and MluI restriction sites in pLTR-sLuc-IRES-GFP vector. The vectors were self-ligated to generate pmLTR- sLuc-IRES-GFP vectors.

A minivirus reporter vector pLGIT (a kind gift from Dr. David Schaffer, University of California, USA) coexpressed EGFP and Tat under HIV LTR. This vector was modified further in two successive steps to replace Tat and LTR both of subtype B origin with subtype C counterparts. First, the subtype B Tat expression segment was replaced with that of subtype C (BL43, Genebank accession No. FJ765005.1). C-Tat was amplified using pC-Tat.BL43.CS (Ref. Ranga U, et al. 2004) as a template and primers N1140 (5'-TCCAGT*CCACAACCATGG*ATGGAGCCAGTAGATCCTAAC-3') and N1141 (5'-GGGCCC*CTCGAG*CTAGTCGAAGGGGTCTGTCTC-3'). The amplified fragment was cloned between BstXI and XhoI sites on pLGIT to give rise to vector $pLGIT_C$. Second, subtype-B LTR in $pLGIT_C$ vector was replaced with subtype-C LTR (Indie C1, accession no. AB023804). Subtype-C LTR was amplified using primers N1145 (5'-

GGCGCC*CTCGAG*ACGCGTTGGAAGGGTTAATTTACTCC-3') and N1146 (5'-TCTAGA*GTTTAAAC*GCGGCCGCTGCTAGAGATTTTTCCCACACTAC-3) and cloned between XhoI and PmeI to give rise to p_CLGIT vector. Further, the 3'-LTR of pcLGIT vector was subjected to additional inactivation mutations in the NF-κB and Sp1 binding sites to generate a series of vectors with isogenic LTRs.

The viral molecular clones Indie $HHC3_C21$, Indie $HHH3_C21$, Indie $HHC3_B21$ and Indie $HHH3_B21$ were constructed using pIndie FHHC as a backbone reported previously from our laboratory (Bachu *et al.* 2012). Using an overlap PCR with primers N1515 (5'-GGATCC*ACGCGT*TGGAAGGGTTAATTTACTG-3'), and N1516 (5'-TCCCTGTT*CGCCGC*GGACTGCTAGAGATTTTCCACACTACAA-3'), the variant LTRs were generated and substituted for the original FHHC LTR at the 3' end of the virus between the MluI and SacII sites.

2.3 Reporter gene expression analysis

Jurkat cells were transiently transfected with the reporter expression vectors using the Lipofectamine 2000 transfection reagent (Cat. No. 11668-019, Invitrogen BioServices, Bangalore, India). The cells were seeded into a 48-well plate at a density of 0.2 x 10^6 cells/well in 250 µl of RPMI-1640 medium supplemented with 10% fetal calf serum. A plasmid DNA pool of 500 ng, containing 400 ng of one of the reporter plasmids and 100 ng of pGL3 vector expressing Firefly luciferase (Cat. No. E1741, Promega Corporation, USA), included as a control for the transfection efficiency, was prepared in 50 µl of serum-free RPMI medium. One µl of Lipofectamine was diluted with 49 µl of serum-free RPMI medium to prepare the lipid transfection reagent. The plasmid pool was mixed with the lipid reagent and the plasmid-lipid mix was incubated for 20 min at room temperature and then added to

appropriate wells. Eight hours following the transfection, 250 µl of 10% RPMI with or without TNF-α (200 ng/ml; T0157, Sigma) was added to the wells. The plates were incubated and 24 h post activation the reporter gene expression was monitored. BioLux *Gaussia* Luciferase assay kit (Cat. No. E3300L, New England Biolabs, Massachusetts, USA) was used for monitoring the levels of the *Gaussia* Luciferase secreted into the culture supernatant. The luciferase assay was performed using a SpectraMax L Luminescence 96-well Microplate Reader (MDS, Inc Model No: s/n Lu 03094) by mixing 20 µl of the culture supernatant and an equal volume of the 1x BioLux GLuc substrate reagent. The assays were performed in triplicate wells and every experiment was repeated at least two times. Expression of the Firefly luciferase in the cell extracts was measured using the Bright-Glo Luciferase Assay kit, (E2620, Promega Corporation, Wisconsin, USA), as per the manufacturer's instructions. The primary data for each well was normalized for the transfection efficiency.

2.4 Preparation of the viral stocks and the estimation of the viral titer

Viral stocks of different molecular clone were generated in HEK 293T cells using the standard calcium-phosphate protocol (Weinberger *et al.* 2005). Briefly, cells were seeded in a 100 mm dish at a density of 3 $x10^6$ cells and transfected with 20 µg of the viral vector DNA along with 0.2 µg of the CMV-RFP expression vector as an internal control for the transfection efficiency. Six h post-transfection, medium was replaced with complete DMEM. Culture supernatants were harvested at 48 h, filtered through a 0.22 µ filter and stored in a deep-freezer in multiple aliquots of 1 ml. The p24 levels of the viral stocks were measured using a p24 ELISA kit (4^{th} generation p24 ELISA kit, J. Mitra and Co. Pvt Ltd., New Delhi, India), as per the maufacturer's instructions. The tissue culture infectious dose ($TCID_{50}$) titer

of the viral stocks was determined using β-galactosidase assay in TZM-bl cells. Briefly, 10^4 TZM-bl cells were seeded in 100 µl of complete DMEM medium in a flat-bottom 96-well culture plate. After 12 h, 50 µl of serially diluted viral stocks (a serial 4-fold dilution) were added to appropriate wells in complete DMEM medium supplemented with 25 µg/ml of DEAE-dextran (D9885, Sigma). Six h post-infection, medium in the wells was replaced with 100 µl of complete DMEM medium. The plates were incubated at 37°C in the presence of 5% CO_2. To examine β-galactosidase expression, 48 h post-infection, cells in each well were washed using 1X PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 mM KH_2PO_4, pH 7.4). After washing, cells were fixed with 100 µl of a fixing solution (1% formaldehyde and 0.2% glutaraldehyde in 1X PBS) and the plates were incubated for 5 min at room temperature. The cells were washed with 1X PBS and stained with 100 µl of the freshly prepared β-galactosidase staining solution (4 mM potassium ferrocyanide, 4 mM potassium ferricyanide, 2 mM $MgCl_2$, and 1 mM X-gal in PBS) and the plates were incubated for 4 h at 37°C. The wells were washed with 1X PBS and the stained cells were counted manually under a low-resolution microscope. The infectious titer of each viral stock was determined by multiplying the cell-count with the dilution factor. Pseudotyped lentiviral vectors used for the ChIP analysis and flow-cytometry were packaged in HEK 293T cells (Burnett *et al.* 2009). Briefly, cells were seeded in 100 mm culture dishes and transfected using the calcium-phosphate transfection protocol with a total of 20 µg of plasmid DNA pool. The plasmid DNA pool consisted of four different vectors: 10 µg of pcLGIT reporter virus containing one of the 3' LTR variant constructs, 5 µg psPax2, 3.5 µg pVSV-G, and 1.5 µg pCMV-Rev. Culture supernatant was harvested 48 h following the transfecton and stored in a deep freezer in 1 ml aliquots. The p24 levels and the $TCID_{50}$ titers of the viral stocks were determined as described above.

2.5 Protein expression and Purification

The expression plasmid pGEX CD-p50 was a kind gift from Dr. Neil D. Perkins (University of Dundee, Scotland, UK). The p50-GST fusion protein was expressed in *E.coli* BL21 (DE3) cells. The p50-GST fusion protein was purified as described (Bachu *et al.* 2012) with a small modification. Briefly, the cells were grown in 250 ml of the LB broth, supplemented with 100 µg/ml of Ampicilin (A9518, Sigma) for 4 h at 28°C, until the O.D. reached 0.4 at 600 nm. Protein expression was induced for 2 h by supplementing the medium with 1 mM IPTG. The bacterial cells were pelleted and resuspended in 10 ml of homogenization buffer (20 mM HEPES, pH 7.9, 10 mM $MgCl_2$, 20% Glycerol, 0.1% NP-40, 1mM DTT, 0.5mM PMSF and 1X protease inhibitor cocktail (P8454, Sigma) . The resupended cells were sonicated using Vibra cell sonicator (Sonics and Materials Inc., Newtown, U.S.A.) at an amplitude of 35%, with 2 seconds on and 4 seconds off for 2 min. The sonicated bacterial cell lysate was centrifuged at 45,000 g for 30 min. at 4°C and to the supernatant, 400 µl of pre-washed GST-bind resin (70541, Novagen, Madison, USA) were added. The samples were incubated for 2 h at 4°C with constant mixing. The supernatant-resin suspension was centrifuged at 100 g for 5 min and the supernatant was carefully aspirated. The GST-resin pellet was washed three times with 10 ml of homogenization buffer. The washed GST-resin was transferred to a fresh 1.5 ml vial and incubated for 10 min with 200 µl of the elution buffer (50 mM Tris-Cl, pH 7.9 containing 10 mM reduced glutathione). The sample was centrifuged for 5 min at 100 g at 4°C and the clear supernatant was transferred to a fresh vial. The supernatant was dialysed against the total volume of 600 ml dialysis buffer (5 mM HEPES, pH 7.6, 1mM DTT, 1mM EDTA, 10% Glycerol, 1X protease inhibitor cocktail and 0.2 mM PMSF) for 4 h with change in buffer at every hour at 4°C and stored in -20°C deep freezer.

2.6 Electrophoretic mobility gel-shift assay and the supershift assay

Jurkat cells, 5 x 10^6 cells harvested and resuspend in 1 ml of ice cold 1X PBS, were transferred to a 1.5 ml plastic vial and centrifuged at 500 g for 5 min at 4 ºC. The buffer was aspirated and the nuclear extract was prepared using a commercial kit (the cytoplasm and nuclear protein extraction kit, K0311, Fermentas life sciences). The protein concentration of the nuclear extracts was determined using a commercial BCA Protein Assay Kit (Cat. No. 23227, Pierce Biotech, Rockford, USA). Aliquots of 50 µl of the nuclear extract were snap frozen in liquid nitrogen and stored in -80°deep-freezer until use.

A single-stranded DNA probe was radiolabeled using 5 µg of DNA with 1 U of T4 polynucleotide kinase (New England Biolabs, Massachusetts, USA) in a solution containing 2 µCi of [γ-32P] ATP. The reaction mixture was incubated for 1 h at 37°C and the enzyme was heat-inactivated at 65°C for 20 min. The radiolabeled oligonucleotide was then hybridized to 10 µg of the complementary strand in the annealing buffer (50 mM NaCl, 20 mM Tris pH 7.5, 10 mM $MgCl_2$, 20 µM EDTA) for 5 minutes in boiling water followed by cooling at room temperature for 2 h. The radiolabeled annealed double-stranded probe was gel-purified using a 6% (29:1) polyacrylamide gel. The gel was exposed to an X-ray film (GBX-2, Kodak, Colorado, USA) for 1 h to develop the autoradiogram. The developed autoradiogram was superimposed over the gel containing radiolabeled probed and the corresponding gel piece was excised out using a scalpel. The radiolabeled probe was extracted from the excised gel piece using the crush and soak method with minor modifications (Sambrook *et al*, 2001). Briefly, the excised gel-piece was placed in a fresh 1.5 ml plastic vial and crushed into fine pieces using fresh needle. Weight of gel-pieces was calculated and soaked in 2 volume of acrylamide gel-extraction buffer (0.5 M ammonium

acetate, 10 mM magnesium acetate, 1 mM EDTA). Sample was incubated at 37°C for 6-8 h. Sample was vortex briefly and centrifuged at 10,000 rpm for 5 min. Supernatant was transferred to a fresh tube. The gel-pieces were soaked once again in 0.5 volume of acrylamide extraction buffer. The sample was vortexed briefly and the supernatant was transferred to the same tube as in the previous step. Two volumes of 100% ethanol were added to the supernatant and the sample was incubated at -20°C for 1 h. The sample was centrifuged at 10,000 rpm for 5 min. at 4°C. The supernatant was carefully aspirated and 200 µl of 10 mM Tris-EDTA (pH 8.0) was added to the pellet. The probe was again allowed to precipitate with 2 volume of 100% ethanol. The pellet was washed with 1 ml of 70% ethanol, air-dried and resuspenden in 50 µl of 10 mM Tris-EDTA buffer. The counts per million of each probe was determined using a liquid-scintillation counter (Wallac 1409, PerkinElmer, MA, USA). Probes of the following sequences, sense and antisense in that order, were used in the assay. The letters in the upper case of the sequences denote the respective transcription factor binding sites and the lower case letters denote the native flanking sequences: the C-κB probe (5'-ccactGGGGCGTTCCagga-3' and 5'-tcctGGAACGCCCCagtgg-3'), the H-κB probe (5'-ccactGGGACTTTCCagga-3' and 5'- tcctGGAAAGTCCCagtgg-3'), the NFAT binding sequence from the IL-2 promoter (5'-tcgagcccaaagaGGAAAatttgtttcatg-3' and 5'-catgaaacaaatTTTCCtctttgggctcga-3') and a non-specific oligonucleotide sequence (5'-ctactgtctcattaagaa-3' and 5'-ttcttaatgagacagtag-3'). For the EMSA, 25 µg of nuclear extract was mixed with 30,000 CPM of the labeled probe in the binding buffer (10 mM Tris pH 7.9, 150 mM KCl, 1 mM DTT, 0.5 mM EDTA, 0.1 % Triton X-100 and 10% glycerol) and supplemented with 1 µg poly dI-dC per reaction and 40 µg BSA. The binding reaction was incubated for 20 min at room temperature and resolved on a 6% polyacrylamide gel (16 X 16

cm) at 250 V for 2:30 h at 4°C. The competition experiments were performed with 10-, 20- or 50-fold excess of unlabeled oligonucleotides added to the reaction mixtures. For the supershift assay, the nuclear extracts were pre-incubated with 2 μg of affinity-purified rabbit polyclonal antibodies raised against p50, p52, p65, RelB or cRel proteins. For the supershift of NFAT complexes, antibodies specific for NFAT1 (ab2722, Cambridge, UK), NFAT2 (ab2796, Cambridge, UK) and NFAT5 (ab3446, Cambridge, UK) were used. A non-specific antibody raised against HIV-1 p24 (generated in-house) was used as a negative control. Thirty min following the incubation, radiolabeled probes were added to the mix and EMSA was performed as described above. Gels were dried and exposed to a Kodak Biomax film at -80 °C. Quantitative EMSA was performed using jurkat nuclear extract recombinant p50 protein to estimate the affinity of the recombinant p50 protein for the C-κB or H- κB binding sites (Li *et al.* 2004;Dahiya *et al.* 2014;Beinoraviciute-Kellner *et al.* 2005).

2.7 Generation of the cell-pools with stable viral integration

In a volume of 700 μl of complete RPMI supplemented with 25 μg/ml DEAE-Dextran in a 12-well tissue culture plate, 0.3 $x10^5$ Jurkat cells were infected with 0.02 - 0.05 m.o.i of the pseudotyped viruses harbouring one of the LTR variant viral strains. The cells were washed after 6 h of incubation and resuspended in 1 ml of complete RPMI medium. Fifteen days following the viral infection, the cells positive for GFP fluorescence were selected using a flow sorter (BD FACSAriaIII cell sorter, BD biosciences) and the sorted cell-pools were allowed to proliferate and stable GFP expression was confirmed. To confirm comparable level of integration with different viral variants, a quantitative real-time PCR was performed with 100 ng of the genomic DNA extracted from the cells, to measure the mean number of

integrated proviral copies in a fixed quantity of the genomic DNA. A 129 bP fragment spanning in the R region to U5 region (+18 to +147) was amplified using the primers N2208 (5'-GATCTGAGCC(T/C)GGGAGCTCTCTG-3') and N2209 (5'-TCTGAGGGATCTCTAGTTACCAGAGTC-3') in the real-time PCR that has the sensitivity to detect as few copies as 10^1 in the assay. Genomic DNA extracted from Tzm-bL cells that harbour a double copy of the integrated virus per cell was used to construct the standard curve for the quantification of the number of integrated viral copies using the regression analysis. For normalization of the viral real-time PCR data, a house-keeping gene GAPDH was amplified using primers N2232 (5'-GAGCTGAACGGGAAGCTCACTG-3') and N2233 (5'-GCTTCACCACCTTCTTGATGTCATC-3').

2.8 Chromatin Immuno-precipitation assay

The ChIP assay was performed using chromatin extracted from Jurkat cell pools harbouring integrated proviruses containing variations in the NF-κB and/or Sp1 sites. 20 $x10^6$ cells per assay, were incubated in the absence or presence of TNF-α (100 ng/ml) for 4 h. The cells were washed once with 5 ml of 1X PBS, resuspended in 1 ml of plain RPMI supplemented with 1% formaldehyde and agitated in a shaker incubator for 10 min at room temperature. After 10 min, the cross-linking reaction was stopped by adding glycine to a final concentration of 0.125 M. The cells were centrifuged at 3,000 rpm for 5 min at 4^oC, the supernatant was aspirated and the cells were washed once with 1 ml of ice-cold 1X PBS. The cells were resuspended in 0.6 ml of the chilled lysis buffer (1% SDS, 50 mM Tris buffer, pH 8.0, 10 mM EDTA) supplemented with the protease inhibitor cocktail (Cat. No. 11836170001, Roche Applied Science, Indianapolis, USA) and incubated for 10 min on ice. The nuclear suspension was divided into 100 µl fractions and each fraction was transferred to

fresh 1.5 ml vial. Each fraction was subjected to 22 cycles of sonication, at the high-level setting, using 30-second-on and 30-second-off pulses, in ice-chilled water using the Bioruptor plus equipment (Diagenode, Liege, Belgium). The magnitude of DNA shearing was monitored by agarose gel-electrophoresis and an average shearing size of 200 - 600 bp was targeted. . The sonicated lysate was centrifuged at 14,000 rpm for 10 min at 4°C and the supernatant was collected. Immunoprecipitation was performed using 2 μg of one of the antibodies against p50 (ab7971, Abcam), p65 (ab 7970, Abcam), Sp1 (CS200631, Upstate), RNA pol II (C15200004, Diagenode), RNA Pol II CTD phospho S2 (ab5095, Abcam), RNA pol II phosphor S5 (C15200007, Diagenode), HDAC1 (C15410053, Diagenode), HDAC3 (ab 7030, Abcam), H3K9Ac (a kind gift from Prof. Tapas Kumar Kundu) or HIV-1 p24 (as a negative control, generated in-house). The immunoprecipitated DNA was subjected to PCR using the primer pair N1054 (5'-GAAGTATTAAAGTGGAAGTTTGACATTC-3') and N1056 (5'-AGAGACCCAGTACAGGCGAAAAGC-3') that amplified a 240 bp region flanking the κB and Sp1 elements in the LTR. A 239 bp fragment was amplified in the gag region to observe the occupancy of RNA pol II S5 (associated with transcription initiation) using primer pair N2467 (5'-GTAATAGAGGAGAAGGCTTTTAGCC-3') and N2468 (5'-AACCAAGGGGAAGTGACATAGC-3'). Occupancy of RNA pol II S2 was observed by amplifying a 218 bp region in the tat gene using the primer pair N2469 (5'-GATCCTAACCTAGAGCCCTGGA-3') and N2470 (5'-GTTCGGGGTAAGGGTTGCTTTG-3'). The amplified fragments were subjected to agarose-gel electrophoresis to confirm the presence or absence of transcription factors over the transcription factor binding sites.

2.9 Viral proliferation and the replication kinetics

CEM-CCR5 cells or PBMC (0.3 x 10^5) were infected with viruses Indie $HHC3_C21$, Indie $HHH3_C21$, Indie $HHC3_B21$, or Indie $HHH3_B21$ at an approximate 500 IU. The infectious titers of the viral stocks were determined as described above. Prior to viral infection, the PBMC were CD8 cell-depleted and activated with 5 μg/ml PHA in complete RPMI medium supplemented with 20 U/ml of IL-2 for 72 h. Following this, the cells were incubated with the viruses in complete RPMI medium supplemented with 10 μg/ml of DEAE-dextran (Konopka *et al.* 1991;Platt *et al.* 2010) for 6 h at 37°C in the presence of 5% CO_2. Six h following the infection, the cells were washed three times in PBS, resuspended in complete RPMI-1640 medium and incubated. Supernatants from the third wash were saved as the day zero time point for p24 ELISA. The cells were monitored for p24 production at weekly intervals for a period of a month. The p24 ELISA was performed for each time point and the viral growth curve was constructed for each monoinfection.

2.10 The post entry events of the viral infection

Full-length Indie-C1 viruses with variant LTR ($HHC3_C21$, $HHH3_C21$, $HHC3_B21$ and $HHH3_B21$) were allowed to infect the CEM-CCR5 cells ($1x10^6$) at the infectious titre of 1,000 $TCID_{50}$ in a 6-well tissue-culture plate. To analyse the late RT products, the cellular DNA was isolated using the Genelute Blood genomic DNA kit (NA2020, Sigma). Primer pairs N1734 (5'-TGTGTGCCCGTCTGTTGT-3') and N1735 (5'-GAGTCCTGCGTCTAGAGGATC-3') was used to amplify a 145 bp amplicon from the U5-ψ region. A standard curve was constructed using the 10-fold serial dilution of the plasmid Indie-C1 (Copy number ranging from 10^1 - 10^5) using salmon-sperm DNA (5 ng/μl). The real-time PCR was performed using a commercial kit (SensiFAST SYBR Mastermix kit,

London, UK) and with 250 ng of the cellular DNA derived from the infection of each virus for template.

To determine the rate of transcription of the viruses, the real-time PCR analysis was performed for the proximal (TAR) and distal (Tat) transcripts. 48 h post-infection, cells were treated with or without 100 ng/ml of TNF-α for 2 h. The total cellular RNA was extracted from $1x10^6$ cells using Trizol (T9424, Sigma). The RNA were subjected to the cDNA synthesis using a commercial kit (BIO-65042, Tetro cDNA synthesis kit, Bioline, London, UK). The cDNA was diluted 10-fold using salmon sperm DNA (5 ng/µl) and subjected to the real-time PCR analysis. An 89 bp of TAR transcript was amplified using the primer pair N2367 (5'-GGTAGACCAGATCTGAGCC-3') and N2368 (5'-CTCAGAGCACTCAAGGCAAG-3'). The distal transcript of 152 bp fragment was amplified in Tat gene using the primer pair N2270 (5'-TGGAGCCAGTAGATCCTAACCTAGAGCC-3') and N1784 (5'-CTTCGTCGCTGTCTCCGCTTCTTCCTG-3'). For normalization, a transcript from GAPDH gene was amplified using primers N2232 (5'-GAGCTGAACGGGAAGCTCACTG-3') and N2233 (5'-GCTTCACCACCTTCTTGATGTCATC-3'). All the Real-time PCR analyses were performed using the Bio-rad CFX96touch real time PCR machine.

2.11 Bibliography

Bachu, M., S. Yalla, M. Asokan, A. Verma, U. Neogi, S. Sharma, R.V. Murali, A.B. Mukthey, R. Bhatt, S. Chatterjee, R.E. Rajan, N. Cheedarla, V.S. Yadavalli, A. Mahadevan, S.K. Shankar, N. Rajagopalan, A. Shet, S. Saravanan, P. Balakrishnan, S. Solomon, M. Vajpayee, K.S. Satish, T.K. Kundu, K.T. Jeang and U. Ranga, 2012a. Multiple NF-kappaB sites in HIV-1 subtype C long terminal repeat confer superior magnitude of transcription and thereby the enhanced viral predominance. J.Biol.Chem. 287: 44714-44735. Available from PM:23132857

Beinoraviciute-Kellner, R., G. Lipps and G. Krauss, 2005. In vitro selection of DNA binding sites for ABF1 protein from Saccharomyces cerevisiae. FEBS Lett. 579: 4535-4540. Available from PM:16083878

Burnett, J.C., K. Miller-Jensen, P.S. Shah, A.P. Arkin and D.V. Schaffer, 2009. Control of stochastic gene expression by host factors at the HIV promoter. PLoS.Pathog. 5: e1000260. Available from PM:19132086

Dahiya, S., Y. Liu, M.R. Nonnemacher, W. Dampier and B. Wigdahl, 2014. CCAAT enhancer binding protein and nuclear factor of activated T cells regulate HIV-1 LTR via a novel conserved downstream site in cells of the monocyte-macrophage lineage. PLoS.One. 9: e88116. Available from PM:24551078

Konopka, K., L. Stamatatos, C.E. Larsen, B.R. Davis and N. Duzgunes, 1991. Enhancement of human immunodeficiency virus type 1 infection by cationic liposomes: the role of CD4, serum and liposome-cell interactions. J.Gen.Virol. 72 (Pt 11): 2685-2696. Available from PM:1940866

Li, Y., Z. Jiang, H. Chen and W.J. Ma, 2004. A modified quantitative EMSA and its application in the study of RNA--protein interactions. J.Biochem.Biophys.Methods 60: 85-96. Available from PM:15262445

Platt, E.J., S.L. Kozak, J.P. Durnin, T.J. Hope and D. Kabat, 2010. Rapid dissociation of HIV-1 from cultured cells severely limits infectivity assays, causes the inactivation ascribed to entry inhibitors, and masks the inherently high level of infectivity of virions. J.Virol. 84: 3106-3110. Available from PM:20042508

Sambrook, J. and P McCallum, 2001. Working with synthetic oligonucleotide probes. Molecular cloning: A laboratory manual 2: 10.11-10.16.

Siddappa, N.B., V.P. Kashi, M. Venkatramanan, A. Balasiddaiah, N. Jayasuryan, A. Mahadevan, A. Desai, K.S. Satish, S.K. Shankar, V. Ravi and U. Ranga, 2007. Gene expression analysis from human immunodeficiency virus type 1 subtype C promoter and construction of bicistronic reporter vectors. AIDS Res.Hum.Retroviruses 23: 1268-1278. Available from PM:17961115

Weinberger, L.S., J.C. Burnett, J.E. Toettcher, A.P. Arkin and D.V. Schaffer, 2005. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. Cell 122: 169-182. Available from PM:16051143

CHAPTER-3

RESULTS: FUNCTIONAL CHARACTERIZATION OF SUBTYPE-SPECIFIC NF-κB AND SP1 III BINDING SITES

3.1 Introduction

A large number of previous studies examining the regulation of gene expression form HIV-1 promoter used mostly the subtype B-LTR. From these studies, it is evident that the NF-κB and Sp1 motifs in the viral promoter collectively play a critical role in controlling not only the gene expression, but also viral latency. The NF-κB motifs are responsible for the inducible gene expression while the Sp1 motifs regulate the basal level of the transcription. Unlike the many inducible cellular promoters that contain these two important TFBS, HIV-1 LTR is exceptional in not only containing multiple binding element of each of the two categories, but also positioning the sites in close proximity typically with a spacer of only 2 to 4 bases separating any two sites. Furthermore, in subtype B-LTR, an active functional interaction was demonstrated between the centrally located NF-κB and Sp1 III sites such a way that not only the spacing between them, but also the relative orientation of each site with respect to the other cannot be altered suggesting that the host factors must bind the specific elements in a specific orientation (Perkins *et al.* 1993). An orientation- and proximity-dependent interaction between the central NF-κB and Sp1 motifs was subsequently demonstrated by the same research group (Perkins *et al.* 1993;Perkins *et al.* 1994). Surprisingly, the putative DNA-binding domains of the two proteins, the 43 amino-acids of N-terminal domain of p65 subunit and the zinc-finger domain of Sp1, are involved in the protein-protein interaction between the two transcription factors while they are bound to the target DNA (Figure-3.1). Importantly, with the exception of p65, the other members of the Rel family cannot interact with Sp1. Thus, the juxtaposition of p65 bound to the upstream NF-κB-binding site (the H-κB site) and Sp1 bound to the downstream target sequence is

necessary for the two host factors to interact with each other to regulate gene expression from the viral promoter.

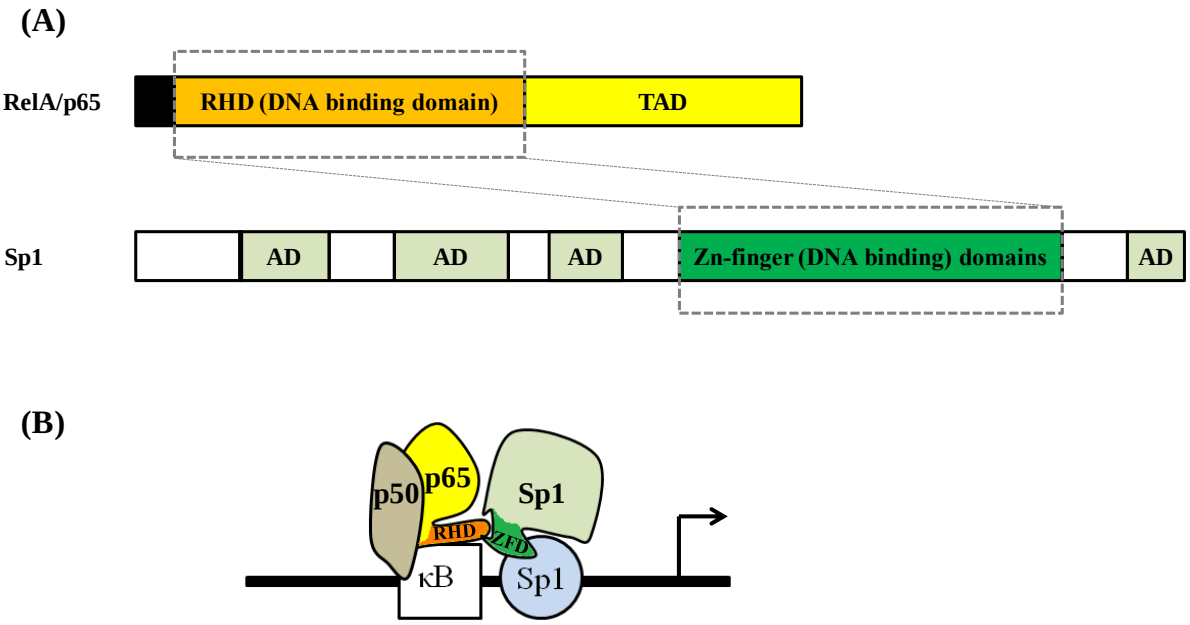


Figure-3.1: A functional association between the NF-κB and proximal Sp1 motifs is critical for viral gene expression. **(A)** A schematic representation of the structure of the p65 and Sp1 proteins. **(B)** p50-p65 hetrodimer and Sp1 factors are recruited to the adjacent NF-κB and Sp1 motifs, respectively. The DNA binding domains of the p65 protein (the RHD, highlighted in orange) and Sp1 factor (the Zn-finger domains, highlighted in green) interact with each other to enhance gene expression from the viral promoter.

A subsequent study by Burnett *et al* examined the relative contribution of each of the two NF-κB and the three Sp1 binding sites to the regulation of gene expression from B-LTR (Burnett *et al.* 2009). Using a reporter virus that coexpressed GFP and Tat, flow cytometry and a panel of variant LTRs that contained mutations in each of the TFBS, they demonstrated that each of the sites differentially contributed to the overall gene expression from the viral

promoter. Importantly, the NF-κB site proximal to the Sp1 region played a stronger activating role than the κB site located upstream showing that the two κB-sites are not absolutely redundant, but play distinct roles supporting each other. Likewise, the Sp1 III site played a more important role than the other two sites downstream, in recruiting p65 and p300 to the promoter. The above studies collectively ascertain the critical role played by the central NF-κB and the Sp1 III motifs in controlling viral gene expression as compared to the flanking sites. Given the special significance attached to the central NF-κB and the Sp1 III motifs in the HIV-1 LTR, importantly to the established proximity- and orientation-dependent functional interaction between them, it is enigmatic that a genetically different NF-κB site (the C-κB motif; GGGGCGTTCC) is inserted in subtype C-LTR (Naghavi *et al.* 1999a). Many previous publications ascribed a quantitative gain-of-function to the C-LTR by demonstrating enhanced levels of gene expression of a reporter gene placed under the control of the viral promoter (Kurosu *et al.* 2002;Naghavi *et al.* 1999b;Perkins *et al.* 1993). These publications, however, did not take into account that the newly inserted NF-κB site must be genetically different and that there exists a coassociated genetic variation in the proximal Sp1 III site.

Of note, a few other studies, however, failed to see enhanced gene expression from the C-LTR as compared to the LTRs of other viral subtypes or binding of transcription factors to the genetically variant C-κB motif (Jeeninga *et al.* 2000;Montano *et al.* 1997) thus questioning the functional significance of the additional NF-κB site. A careful examination of the experimental strategies in the above publications, however, would identify serious experimental flaws including the use of DNA probes that contained incorrect genetic

sequences in the gel shift assays. Lemieux *et al.* used a double-stranded synthetic DNA probe that differed from the authentic sequence at two different sites (5'-GGGGCGGTCT-3', differences underlined) (Lemieux *et al.* 2004). The DNA probe used by Roof *et al.* contained two tandem sequences of the NF-κB site that are separated by a single 'T' residue unlike in the natural context where the spacer consisted of 3 residues (Roof *et al.* 2002). The significance of the spacer length between the two TFBS and the strain caused by the proximity of the binding sites in the cooperative binding of the two sites have been elegantly demonstrated (Stroud *et al.* 2009). Additionally, the probe containing the C-κB site sequences was not used directly for the probe binding, but only in the cold competition with the other probes. In the light of the technical limitations, the inferences drawn by the above reports on the biological significance of the variant NF-κB site in the C-LTR may not be justified. In contrast to the above reports, two other publications demonstrated biological functioning of the C-κB site. Naghavi *et al.* demonstrated strong binding of the cellular factors to the DNA probes from HeLa nuclear extracts (Naghavi *et al.* 1999b). Montano *et al.* showed an efficient competition between the labeled canonical H-κB probe and the cold C-κB sequence in the gel shift assay (Montano *et al.* 1997). Additionally, Siggers *et al.* using protein binding microarrays and diverse NF-κB-binding sequences, demonstrated that a sequence nearly comparable with the HIV-1 subtype C unique C-κB probe (5'- GGGGCGTTCC -3') contained potential to bind NF-κB (Siggers *et al.* 2012;Wong *et al.* 2011).

In this backdrop, the primary objective of the present work is to examine if there exists a new found functional association between the centrally located and subtype C unique NF-κB and Sp1 binding sites in subtype C-LTR. A functional association between the two

centrally located transcription factor binding sites has been demonstrated previously in the context of B-LTR (Perkins *et al.* 1993;Perkins *et al.* 1994). Two pertinent observations regarding C-LTR make the primary objective of the present work warranted. The genetically distinct C-κB motif is inserted invariably between the two upstream H-κB sites and the downstream Sp1 binding elements, thus the C-κB site and the subtype-specific Sp1 III motifs are invariably juxtraposed with each other. Additionally, the proximal Sp1 site (the Sp1 III motif) also is characterised by subtype-specific genetic variations. The subtype C specific C-κB and Sp1 III motifs are highly conserved among subtype C strains (Figure-3.2). Thus in subtype C, a genetically distinct NF-κB motif and a genetically distinct Sp1 site have been brought into close proximity (Figure-1.6). A functional association between these two adjacent motifs is expected and remains to be demonstrated.

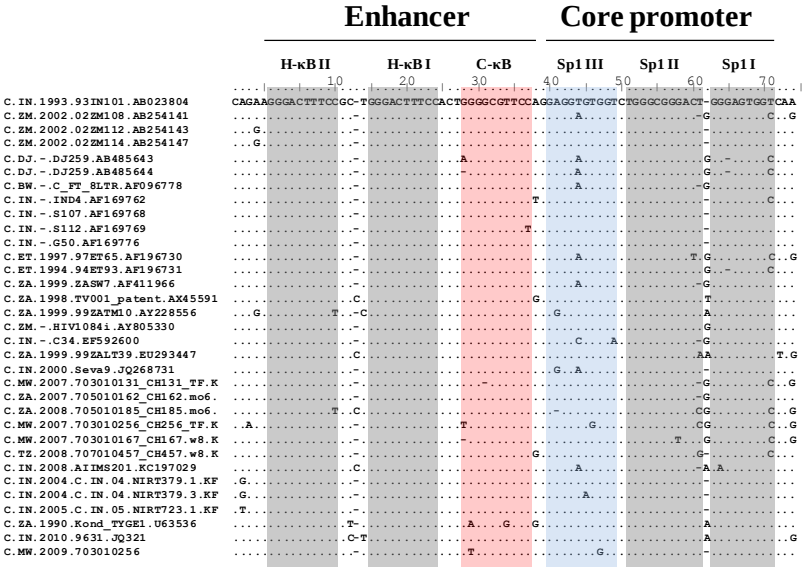


Figure-3.2: Multiple sequence alignment of the enhancer and core promoter sequences of subtype C-LTR. The κB and Sp1 motifs are highlighted. The C-κB site (pink shade) is not only highly conserved among the global subtype C sequences, but is invariably located upstream of the Sp1 III site (blue shade). The sequences deposited between years 1989 to 2010 have been downloaded from hiv.lanl.gov/. Subtype C prototype Indie C1 (AB023804) has been used as the reference sequence. Dots and dashes represent the sequence identity and deletions, respectively.

3.2 Results

3.2.1 Physical proximity between the C-κB and Sp1 III motifs is critical for optimal gene expression from subtype C-LTR

A functional association between the downstream NF-κB binding site (the H-κB site, GGGACTTTCC) and the proximal Sp1 III site (GAGGCGTGGC) was demonstrated previously in the context of subtype B-LTR (Perkins *et al.* 1993). Compared to the B-LTR, there is a genetically variable and additional κB-element (the C-κB site, GGGGCGTTCC) introduced in subtype C enhancer that becomes Sp1 proximal. Importantly, the Sp1 III site also undergoes genetic variations that is subtype specific (GAGGTGTGGT, variations as compared with subtype B underlined). In spite of these two variations in subtype C-LTR, a new found functional association between the proximal κB and Sp1 sites is expected, but remains to be confirmed. To this end, we constructed two different panels of expression vectors under the control of variant C-LTR, using a parental vector pLTR-sLuc-IRES-EGFP reported previously (Bachu *et al.* 2012). All the vectors expressed two different reporter genes simultaneously; a secreted luciferase and EGFP, the expression of either reporter can be measured without terminating the cell culture. The initial round of standardization experiments did not show a significant difference in the profile of the reporter gene expression whether a complete C-LTR or a trimmed LTR that lacked the upstream modulator region (-150 to -463 deleted using Mlu I and BspE I restriction enzymes) was used in the assay (Figure-3.3). We used only the full-length (634 bp) promoter (Indie-C1, accession no.-AB023804) in all the subsequent analyses.

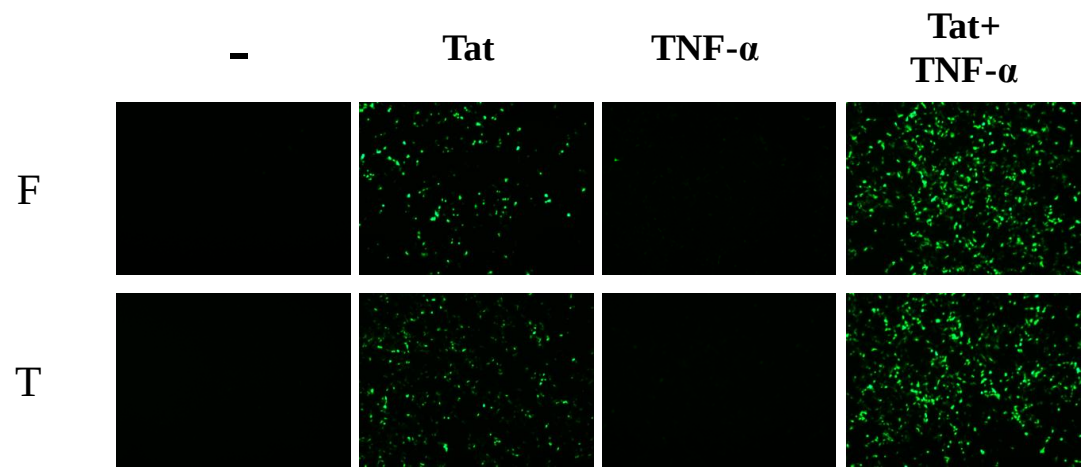


Figure-3.3: A comparison of the full-length and trimmed LTR for the promoter assay**.** The presence or absence of the modulator region does not significantly influence the expression of the reporter genes from the viral promoter. The expression of the EGFP was placed under the control of the full-length (F) LTR or an LTR from which the modulator region (-150 to -463) was deleted (T, trimmed). The vectors were introduced into HEK293 cells and activated with Tat, TNF-α, both or left without activation. The reporter gene (GFP) expression was measured at 24 h after the transfection. Comparable results were obtained at later time points or in other cell lines.

In one of the two panels of the reporter vectors, consisting of three different promoters (HHC, HCH and CHH), keeping the Sp1 III site stable, the C-κB motif was moved progressively upstream, thus increasing the distance between the Sp1 III site and the C-κB element. In the other panel of three vectors, doing the contrary, the C-κB site was kept constant, and the subtype C specific Sp1 III site was moved progressively downstream again increasing the distance between the two elements. Of note, in all the expression vectors, all the 6 TFBS (the 3 each NF-κB and Sp1 sites) remained intact except for the positional shifts.

The vectors of panel-1 containing the positional shift of the C-κB site were introduced into Jurkat or HEK293 cells transiently, the cells were activated with TNF-α (100 ng/ml) or left without activation and the expression pattern of secreted luciferase (Figure-3.4) or GFP (data not shown) was measured 24 h after the activation. The wild type LTR (HHC) demonstrated 588,286.1 ± 428.69 RLU in Jurkat cells or 639,279.0 ± 6,659.5 RLU in HEK293 cells of

luciferase activity that was up regulated approximately 3-fold following TNF-α activation. The progressive upstream shift of the C-kB site to the one (HCH) or the other position (CHH) away from the original location significantly reduced the basal level as well as the induced activities of the promoter (Figure-3.4 A and B). Although the loss in the promoter activity was statistically significant, the gene expression was not completely abrogated. These results suggested that although the H-κB is not an ideal substitute for the natural C-κB site, the former can function from the central location although at a significantly reduced efficiency.

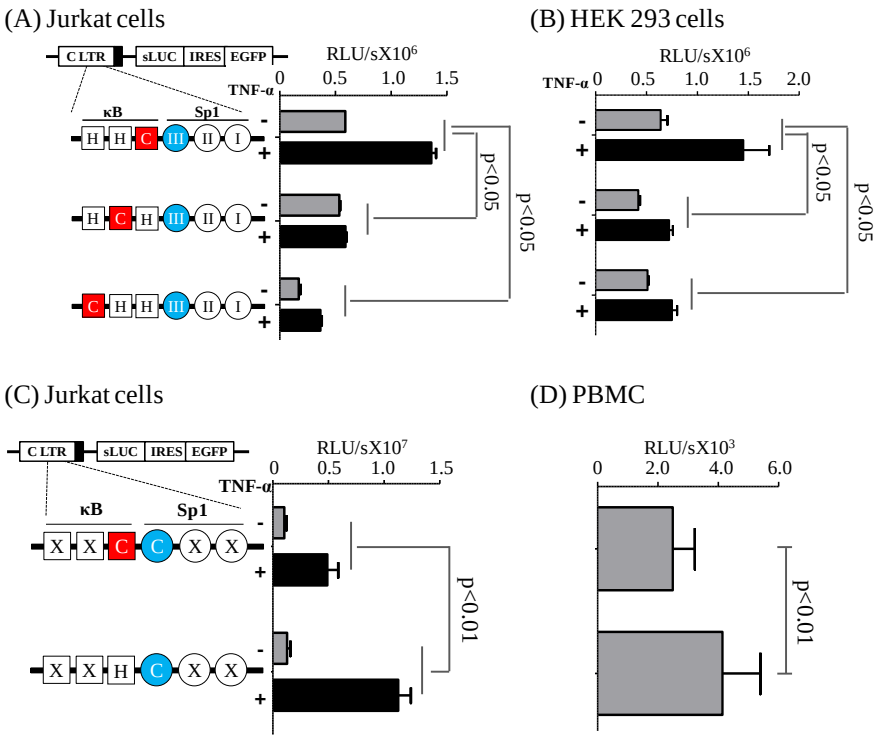


Figure-3.4: The C-κB and the Sp1 III motifs must remain adjacent to each other and neither element could be efficiently substituted by an alternative sequence. **(A)** Jurkat or **(B)** HEK 293 cells were transiently transfected with one of the reporter vectors. Eight h post-transfection, cells were exposed to TNF-α activation (100 ng/ml). The luciferase secretion was evaluated from the culture supernatant at 24 h post-activation. **(C)** Jurkat cells or **(D)** Activated PMBCs were transfected with the vectors (C:C or H:C). The data are representative of three independent experiments. The data are presented as mean relative light units ± S.D. Statistical analysis was performed using two-way ANOVA in the Graph-Pad Prism software.

We generated two additional plasmid vectors in which the two upstream NF-κB sites and the two downstream Sp1 sites were inactivated by introducing point mutations thus leaving only the central NF-κB and Sp1 sites intact. While one of the vectors contained the wild type C:C configuration, in the other one the C-κB site was replaced by the H-κB site (H:C). We wanted to examine if the suboptimal gene expression from the viral promoter where the H-κB site was substituted for the natural C-κB site can be reproducible in the absence of the flanking TFBS. Jurkat cells were transiently transfected, the cells were activated with TNF-α or left without activation and the luciferase activity was determined at 24 h following activation (Figure-3.4 C). The wild type subtype C promoter (C:C) demonstrated 1,193,408.2 ± 222,740.2 RUL/s reporter activity without activation that enhanced approximately 5-fold after the activation ($p<0.01$). The chimera viral promoter where the H-κB element was placed upstream of the Sp1 site (H:C) demonstrated comparable basal level reporter activity (1,479,455.0 ± 321,241.3 RLU/s), as the wild type LTR, which was enhanced approximately 10-fold following the activation. Importantly, the induced reporter activity was superior in the case of the chimera promoter and the difference was statistically significant ($p<0.01$). Similar results were obtained when IL-2-activated PBMC derived from three different healthy subjects were used in the assay (Figure-3.4 D). As in Jurkat cells, the reporter activity was superior from the chimera H:C-LTR as compared to the wild type C:C promoter with the difference being statistically significant ($p<0.01$). These data suggested that under specific conditions, the H-κB site can act as a functional substitute for the native C-κB site.

A loss of promoter activity was observed also from the vectors in panel-2 (Figure-3.5 A and B). A downward shift of the subtype C specific Sp1 III site away from the C-κB site by one or two positions resulted in a severe loss of the reporter gene expression. The loss was so severe suggesting that Sp1 II site is not a functional substitute for Sp1 III in C-LTR. Collectively, the strategy of the TFBS exchange suggested that the C-κB and the Sp1 III motifs must be located proximal to each other and the alternate TFBS are not functional substitutes for either of the two original motifs, especially in the presence of intact flanking sites.

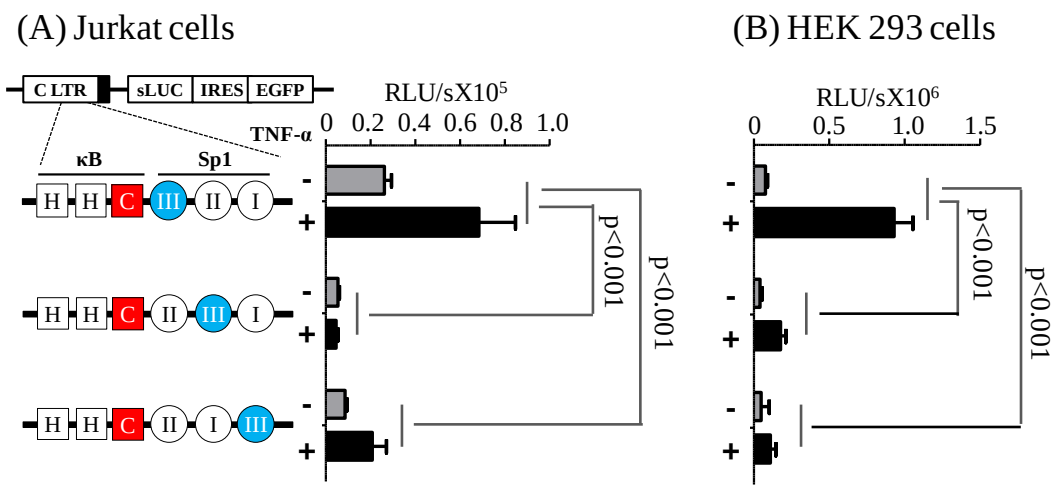


Figure-3.5: The C-κB and the Sp1 III motifs must remain adjacent to each other and neither element could be substituted by an alternative sequence. Jurkat or HEK 293 cells were transiently transfected with one of the reporter vectors. Eight h post-transfection, cells were exposed to TNF-α (100 ng/ml) activation. The luciferase secretion was evaluated from the culture supernatant at 24 h post-activation. The data are representative of three independent experiments. The data are presented as mean relative light units ± S.D. Statistical analysis was performed using two-way ANOVA in Graph-Pad Prism software.

3.2.2 The C-κB motif of HIV-1 subtype C can functionally associate with proximal Sp1 sites of other subtypes

The most unexpected finding of the present work is the inability of the H-κB motif to substitute for the original C-κB site although the former sequence is present in all the HIV-1

promoters and is biologically functional (Figure-3.4 A and B). It was also unexpected that the Sp1 III site of subtype C could not be replaced by Sp1 II sequence of the same viral promoter. In either of the substitutions, the promoter activity was significantly reduced in the LTR containing all the 6 TFBS intact. The core promoter of HIV-1 contains three genetically diverse Sp1 sites of which the Sp1 III site proximal to the NF-κB motifs is the most variable one demonstrating subtype-specific variations (Naghavi *et al.* 1999b). Importantly, experimental evidence also suggests that the Sp1 III site is biologically more important than the other two (Burnett *et al.* 2009;McAllister *et al.* 2000;Yuste *et al.* 2002). Given the biological significance of the centrally located Sp1 III motif and subtype-associated variations within HIV-1 subtypes, we asked if the heterologous Sp1 III sequences, derived from the other HIV-1 subtypes can function in the context of subtype C-LTR. To this end, we generated a panel of reporter vectors by grafting subtype-specific Sp1 III sequences of subtypes A, B, D and E (Figure-3.6 A) at the NF-κB proximal location in C-LTR. The expression of luciferase was determined at 24 h following the TNF-α activation of Jurkat or HEK 293 cells (Figure-3.6 B). The wild type subtype C-LTR demonstrated 22,975.4 ± 7,343.2 and 1.02 $x10^8$ ± 2.0 $x10^7$ RLU/s of reporter activity in Jurkat or HEK293 cells, respectively. This activity was enhanced 4-6 fold following TNF-α activation. Of note, the basal level reporter activity of the wild type subtype C vector was the lowest in both the cell lines. However, the induced reporter activity of all the 5 vectors was comparable suggesting fold transactivation of subtype C-LTR is the highest compared with other viral promoters. In summary, regardless of the genetic variations, the Sp1 III motifs of the five viral subtypes functioned with comparable efficiency suggesting that each of these motifs can complement efficiently with the C-κB site. Unlike the Sp1 II motif of Subtype C (Figure-3.5 A and B),

these heterologous Sp1 III elements can serve as functional substitutes for the subtype C motif.

(A) Sequence comparison of Sp1 III sites

```
Subtype C - G A G G T G T G G T
        A - . . . . C . . . . .
        B - . . . . C . . . . C
        D - . . . . C . . . . A
        E - . . . . . . . . . C
```

(B) Luciferase assay

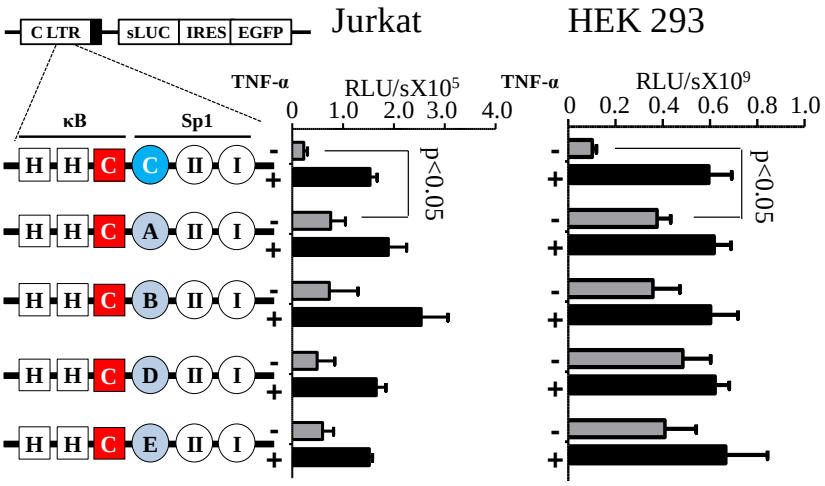


Figure-3.6: The C-κB motif of HIV-1 subtype C can functionally associate with proximal Sp1 sites of other subtypes**. (A)** Sequence comparison of Sp1 III site derived from different HIV-1 subtypes. The sequences under each HIV-1 subtype represent the consensus motif of the subtype.**(B)** Luciferase assay demonstrating that the heterologous Sp1 III sequences of four other viral subtypes (A, B, D and E) can be functional substitutes for the subtype C motif. The Jurkat or HEK293 cells were transiently transfected with one of the reporter vectors. Eight h post-transfection, cells were subjected to TNF-α activation. The luciferase secretion was measured from the culture supernatant at 24 h post activation. The data are presented as mean relative light units ± S.D. Statistical analysis was performed using two-way ANOVA in Graph-Pad Prism software.

3.2.3 The association between the C-κB and Sp1 III motifs is orientation and distance dependent

We generated additional reporter vectors by flipping the C-κB and Sp1 III sites or inserting a 5 residue linker between the two sites (Figure-3.7). The vector in which the positions of the C-κB and Sp1 III sites were interchanged did not show any reporter activity

in Jurakt or HEK 293 cells, as expected (Figure-3.7, middle panel). The DNA-binding domain of the p65 subunit can interact with the DNA-binding domain of the Sp1 only when the NF-κB motif is present upstream of the Sp1 binding motif thus juxtaposing the two factors in a specific fashion (Figure-3.1) as shown previously (Perkins *et al.* 1993). Likewise, when the spacer length between the two sites was increased to 7 bp from the natural 2 bp (GGGGCGTTCC***tgcag***agGAGGTGTGGT, added sequences in the spacer underlined), a complete loss of the reporter activity was observed in both the cell lines (Figure-3.7, lower panel).

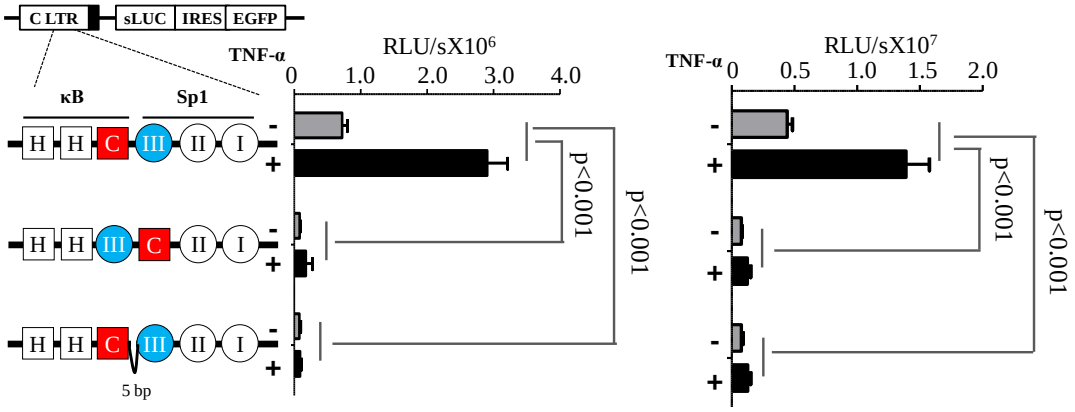


Figure-3.7: The association between the C-κB and Sp1 motifs is orientation and distance dependent. Both Jurkat or HEK293 cells were transiently transfected with one of the reporter vectors. Eight hour post-transfection, cells were activated with TNF-α (100 ng/ml). The luciferase secretion was measured from the culture supernatant at 24 h post activation. The data are representative of the three independent experiments. The data are presented as mean relative light units ± S.D. Statistical analysis was performed using two-way ANOVA in Graph-Pad Prism software.

The data presented above collectively suggest that the specific location, mutual orientation and spacing between the C-κB and Sp1 III sites are critical factors that regulate the biological function of the two elements confirming a functional association between the two important elements in subtype C-LTR. Thus, as expected, the original association

demonstrated previously between the H-κB and Sp1 III elements in B-LTR has been reconfigured in C-LTR to establish a new functional association between the centrally located C-κB motif and the subtype-specific Sp1 III site.

3.2.4 Functional association between the C-κB and Sp1 III motifs in the absence of the flanking NF-κB and Sp1 sites

Although the wild type subtype C-LTR contains three each NF-κB and Sp1 binding sites in the promoter, a direct association is possible only between the centrally located C-κB and Sp1 III sites. In the viral promoter, the multiple NF-κB and Sp1 sites are expected to be functional, collectively regulating the overall gene expression along with the contribution from the other regulatory elements. The relative magnitude of gene expression due to the specific interaction between the adjacent C-κB and Sp1 III sites is likely to be modulated by the interference from the two upstream NF-κB sites and the two downstream Sp1 motifs in subtype C-LTR. To overcome this limitation and to examine the direct influence of the C-κB and Sp1 III elements on gene expression from C-LTR, we introduced inactivating mutations into the flanking κB and Sp1 sequences by site-directed mutagenesis thus leaving only the adjacent C-κB and Sp1 III sites intact. We generated a panel of 4 reporter vectors where the C-κB and/or the Sp1 III sites were intact or both were inactivated (C:C, C:X, X:C and X:X, C before and after the colon represents the subtype-specific NF-κB and Sp1 III sites, respectively) by introducing mutations into critical residues of the corresponding motifs. In all the vectors, the overall length of the nucleotides remained constant, as none of the sequences were deleted, to maintain the DNA helical conformation same as the wild type configuration. In Jurkat cells, the vector that lacked any of the 6 TFBS (X:X, Figure-3.8 A),

or the one that contained only the C-κB site (C:X) demonstrated insignificant levels of the basal reporter activity that was marginally enhanced following TNF-α activation. The vector which contained only the Sp1 III site (X:C) demonstrated 635,758.4 ± 116,208.9 RLU/s of reporter activity without activation and 6-fold enhanced activity (3,919,221.9 ± 664,666.2 RLU/s, $p<0.001$) following activation confirming that the presence of a single Sp1 site in the promoter can restore the transcriptional activity as well as inducibility. The fourth reporter vector containing both the C-κB and Sp1 III sites (C:C) demonstrated the highest magnitude of reporter activity and inducibility. The basal activity of 1,248,381.0 ± 224,166.6 RLU/s in this promoter enhanced to 7,249,027.4 ± 1,268,153.9 RLU/s following TNF-α activation ($p<0.001$). Comparable results were obtained when PBMC from three different subjects were activated and used in the assay (data from one representative donor is shown, Figure-3.8 B).

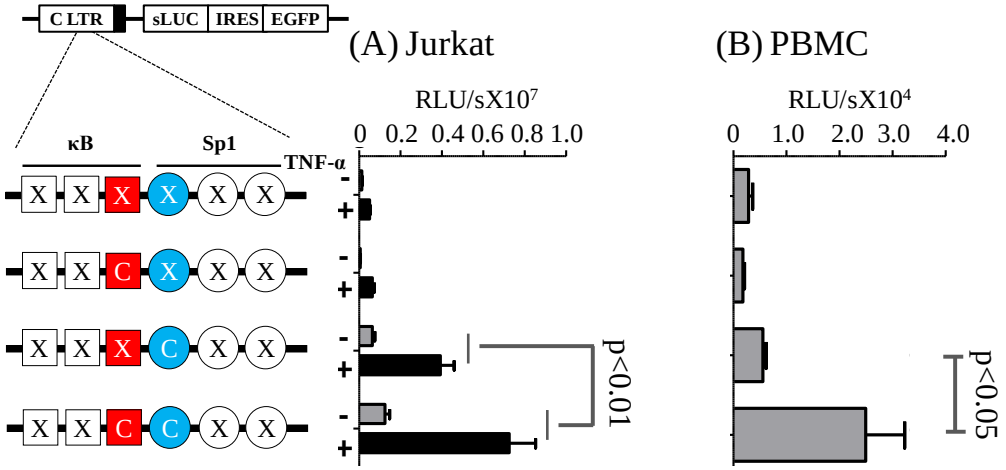


Figure-3.8: Functional association between the C-κB and Sp1 motifs in the absence of the flanking NF-κB and Sp1 sites**.** A panel of reporter expression vectors was generated that contained only the C-κB site (C:X), Sp1 III site (X:C), both (C:C) or none (X:X) intact. While 'C' represents subtype C specific κB or Sp1 III sequences, 'X' denotes a debilitating mutation in the corresponding motif. The luciferase activity was determined in **(A)** Jurkat cells or **(B)** activated PBMC of a healthy subject, 24 h following the transfection. The assay was performed using PBMC derived from three different subjects and the data from one representative subject is shown. The assay was representative of three independent experiments. Two-way ANOVA was used for the statistical analysis.

In summary, the optimal reporter gene expression required the presence of both the κB and the Sp1 elements in the promoter under all the experimental conditions. Furthermore, the reporter expression profile was consistent even in the absence of the flanking κB and Sp1 motifs.

3.2.5 The C-κB probe recruits p50-p65 heterodimer from nuclear extracts

In comparison to the other two canonical H-κB sites, the additional C-κB motif in C-LTR is genetically distinct with variations at positions 4 (A to G) and 6 (T to G) (Figure-1.6). Using gel-shift analysis (cold-competition and super-shift assays), a couple of publications previously claimed the absence of NF-κB binding to the C-κB sequences of subtype C with an underlying implication that the site may not be biologically functional. A close examination of the probes used in these reports reveals technical flaws in the probe sequences or a faulty experimental design (Roof *et al.* 2002;Stroud *et al.* 2009). Importantly, in contrast to the above reports, two other publications demonstrated biological function of the C-κB site (Montano *et al.* 1997;Naghavi *et al.* 1999b).

In the backdrop of the controversy, we set out to examine the ability of the C-κB probe to recruit NF-κB from nuclear extracts. Using double-stranded radio-labelled DNA probes constituting either the canonical H-κB (5'-ccactGGGACTTTCCagga-3', the flanking sequences are in the lower case), the C- κB (5'-ccactGGGGCGTTCCagga-3', the differences are underlined), or a non-specific scrambled (5'-ctactgtctcattaagaa-3') probe sequences, we compared the C- and H-κB probes to recruit cellular factors form Jurkat cell nuclear extract in the presence or absence of TNF-α activation. Distinct cellular complexes were recruited by both H- and C-κB DNA probes under control conditions, and the intensity of the

complexes increased several fold following TNF-α activation (Figure-3.9 A, compare lanes 2 and 3). These specific complexes were out-competed with progressively increasing molar excess of cold probes (10-, 20- and 50-fold) representing H-, or C-κB sequences (Figure-3.9 A, lanes 4–6), but not with a non-specific probe at a 50-fold excess molar concentration (lanes 10). The nature and intensity of the shifted complexes appear to be comparable between the H- and C-κB probes with no significant visible difference. The binding of the nuclear complexes by the C-κB DNA probe was highly reproducible. Having confirmed the functionality of the C-κB probe in EMSA, we wanted to determine the identity of the bound nuclear factors. Using the nuclear extract isolated from TNF-α activated Jurkat cells and immuno-affinity purified rabbit antibodies raised against each of the five Rel family members, we performed a super-shift assay to examine the nature of the host factors binding the DNA probes. The analysis revealed the presence of p50-p65 heterodimers in the complexes, but not any of the other members of the Rel family. The slower moving complexes were evident only in the p50 and p65 lanes (Figure-3.9 B, lanes 4 and 5). The nature of the super-shifted complexes was highly reproducible in Jurkat as well as in HeLa cells (data not shown). The nature of the shifted and super-shifted complexes was comparable between the H- and C-κB probes with a small difference in the intensity of the C-κB probe super-shifted complexes being clearly higher than that of the H-κB probe suggesting a higher affinity of binding of the nuclear complexes by the C-κB element.

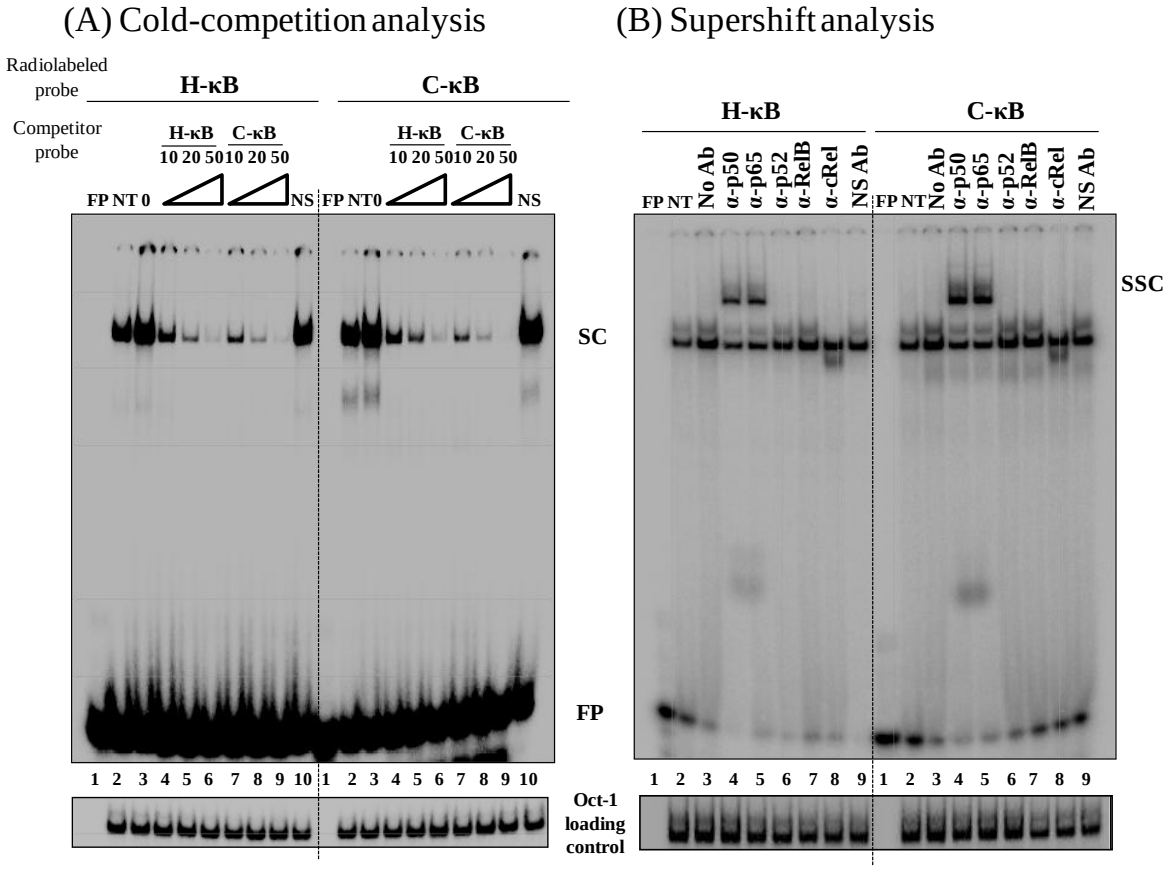


Figure-3.9: The C-κB probe demonstrates the binding of NF-κB. **(A)** Cold-competition analysis. Radio-labelled double stranded probes (40,000 cpm) comprising of the H-κB (left panel) or the C-κB motif (right panel) were incubated with 30 μg of Jurkat nuclear extracts prepared from cells activated or not with TNF-α for 2 h. Cold-competition was performed by incubating the complexes with 10-, 20- or 50-fold molar excess of the cold probes (H-κB or C-κB probe). A nonspecific probe was used only at the highest concentration (50-fold). **(B)** Supershift analysis. The nuclear extract prepared from activated Jurkat cells was incubated with specific antibodies against each of the Rel family members prior to the addition of the radio-labelled H-κB or C-κB probe. Oct-1 was used as a loading control in the experiment. FP-Free probe, SC-shifted complexes, SSC-supershifted complexes, NT- No TNF-α treatment, NS- non-specific probe, NS Ab- non-specific antibody (anti-p24 antibody). Each experiment was repeated a minimum of three times.

3.2.6 The C-κB motif shows higher affinity for NF-κB or recombinant p50 protein

In gel-shift and super-shift assays, we consistently observed that the intensity of the complexes recruited by C-κB motif was always superior to that of the H-κB probe (Figure-3.9 B). To compare the relative strength of binding of the cellular complexes to the H- versus

C-κB probes, we performed a quantitative cold-competition assay (Dahiya *et al.* 2014) using Jurkat nuclear extract. The binding of 30,000 CPM of the labelled H- or C-κB probe to the host factors in 25 μg of nuclear extract was competed with progressively increasing concentration (0-, 4-, 16-, 64-, 128- and 258-fold molar excess) of cold H- or C-κB probes. The band intensity was measured using densitometry and the percent binding was determined at each concentration of the competing probe considering the intensity of the probe binding as 100% in the absence of the competition. The percent binding of the probe was plotted against the concentration of the competing probe. The amount of the competing probe concentration required to reduce probe binding to the host factors to 50% was determined by scatchard plot analysis (Figure-3.10). When the H-κB probe was used in the assay, 13.0 pico moles of the cold H-κB oligo were required for a 50% competition whereas approximately only half of this quantity of the C-κB cold probe, 6.6 pico moles, was sufficient for a comparable level of inhibition (Figure-3.10 A) confirming higher affinity of the C-κB motif for NF-κB. The results were comparable when the C-κB probe binding to the host factors was outcompeted with the H- or C-κB cold probes. For a 50% inhibition of the probe binding to the nuclear complexes, 35.2 and 18.0 p moles of H- and C-κB competitors were required, respectively (Figure-3.10 B).

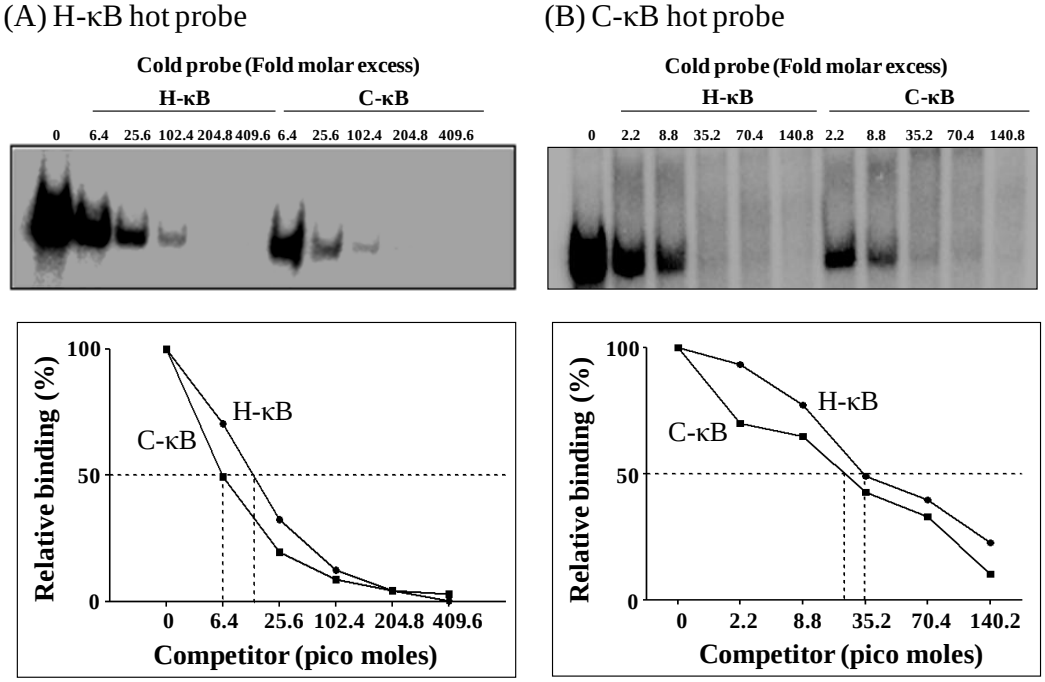


Figure-3.10: The C-κB motif demonstrates a higher binding affinity for cellular factors from Jurkat nuclear extract. 30 µg of the Jurkat nuclear extract was incubated with radio-labelled double-stranded DNA probe (40,000 cpm) containing **(A)** the H-κB or **(B)** C-κB motif. The competition was performed with progressively increasing concentrations of H-, or C-κB cold probes. Densitometry quantitation of the complexes is shown in the lower panel. The data are representative of three independent experiments.

Although the presence of NF-κB in the bound complexes was conclusively demonstrated in the assays described above, the molecular nature of the bound factors is expected to be more complex with the presence of the associated cofactors that may modulate the binding affinity of NF-κB for the cognate binding site. NF-κB DNA probes can bind recombinant p50 protein in the absence of the other host factors. The recruitment of p50 homodimers to NF-κB motifs is critical for the establishment and maintenance of viral latency (Colin and Van ,2009). We therefore compared the binding profile of recombinant p50 to H- and C-κB probes to determine the affinity constant values.

We performed a quantitative EMSA using increasing amounts of recombinant p50 protein (0, 0.6, 1.3, 2.6, 5.2 and 10.4 nM) incubated with a constant amount of H- or C-κB radio-labelled probe (60,000 CPM) (Beinoraviciute-Kellner *et al.* 2005;Li *et al.* 2004). Recombinant p50 protein was purified as per the protocol mentioned in section-2.5 (Figure-3.11). Using densitometry, the quantities of free and complexed probes were quantitated and plotted against the probe concentration used in the assay.

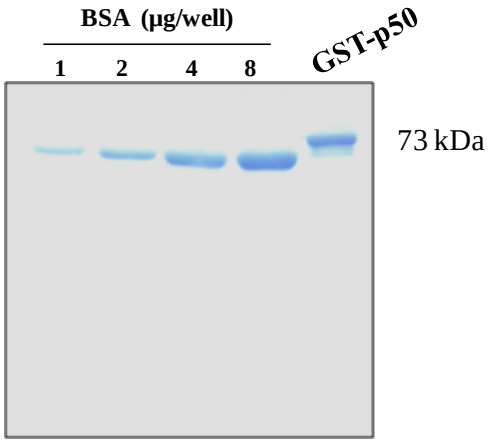


Figure-3.11: Purification of the recombinant GST-tagged p50 protein from the BL-21 strain of *E. coli*.

Using the scatchcard analysis, the dissociation constant (K_d) values for the H- and C-κB probes towards the recombinant p50 protein were found to be 4.6 and 2.8 nM, respectively (Figure-3.12). The data from the gel-shift analyses collectively demonstrates that the C-κB motif, as compared to the H-κB sequence, contains approximately a two-fold higher affinity for NF-κB present in the nuclear extract or for the recombinant p50 homodimer. A higher affinity of binding to the C-κB site for NF-κB in the C-LTR therefore could have a significant impact on the functioning of the viral promoter in subtype C.

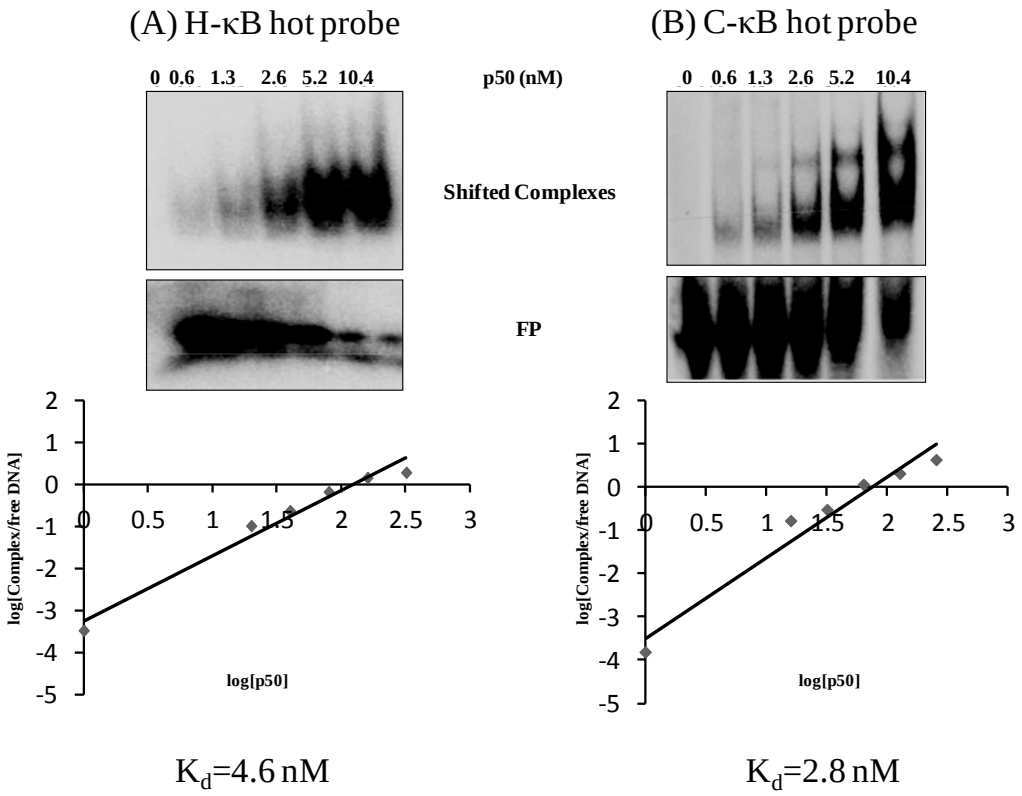


Figure-3.12: The C-κB motif demonstrates a higher binding affinity for the recombinant p50 protein. The recombinant p50 protein, at serially increasing concentration, was incubated with radio-labelled double-stranded probes (60,000 cpm) containing **(A)** the H-κB or **(B)** C-κB motifs. Densitometry quantitation of the complexes is shown in the lower panel. The data are representative of three independent experiments.

3.2.7 The C-κB probe binds NFAT from nuclear extracts

The canonical H-κB motif contains an overlapping binding site for another important family of transcription factors called the the nuclear factor of activated T cells (NFAT). The NFAT family contains four different members (NFAT1-NFAT4) and one calcium-independent Transcription factor (NFAT5) (Hogan *et al.* 2003;Macian *et al.* 2001). The core binding sequence for NFAT is GGAAA (Badran *et al.* 2002;Macian *et al.* 2001). Of the family members, NFAT1 homodimers can occupy the 3' half of the H-κB motif consisting of 5 bases and modulate HIV-1 gene expression (GGGACTTTCC, the NFAT binding location

within the H-κB motif is underlined) (Romanchikova *et al.* 2003;Zhang *et al.* 2012). Due to a partial overlap in the binding site, NFAT and NF-κB could compete with each other leading to mutually exclusive binding and alternate transactivation by these two factors (Pessler and Cron ,2004). While NFAT1 binding to the H-κB site has a negative modulatory effect of NF-κB mediated transactivation from the HIV-1 LTR (Macian *et al.* 2001), NFAT2 binding positively regulates HIV-1 LTR gene expression (Kinoshita *et al.* 1997).

Importantly, the genetic variation at position 6 of the C-κB site (GGGGCGTTCC) is expected to reduce the NFAT binding to only 4 residues in this element as compared to 5 residues in the H-κB motif. With the reduced sequence motif to bind, it is not known if NFAT can still bind the C-κB site of the subtype C-LTR. Using the gel-shift assay, we examined if the DNA probe comprising of the C-κB sequence can bind NFAT from nuclear complexes as compared with the H-κB probe. Double-stranded radio-labelled DNA probe containing the NFAT binding sequence, derived from the IL-2 promoter, was incubated with the nuclear extracts prepared from Jurkat cells induced with PMA (20 ng/ml) and ionomycin (10 nM) for 1 h. The intensity of the shifted complexes increased following the cell induction (Figure-3.13 A, compare lanes 2 and 3). In self competition, the NFAT cold probe outcompeted NFAT probe binding to the nuclear complexes at 5 and 25 molar excess (Figure-3.13 A, compare lanes 3 vs 4 and 5). Interestingly, the H-κB cold probe outcompeted NFAT binding to the probe with a greater efficiency as compared to the NFAT cold probe itself (Figure-3.13 A, compare lanes 4 and 5 with lanes 6 and 7). The C-κB cold probe was as efficient as the H-κB cold probe in inhibiting the binding of the nuclear factors to the NFAT probe (Figure-3.13 A, compare lanes 4 and 5 with lanes 8 and 9). A non-specific

scrambled cold DNA probe, used as negative control, was not able to compete in the assay thus confirming the specificity of the assay (Figure-3.13 A, lanes 10, 11). Collectively, the data confirmed that the C-κB sequence is capable of binding NFAT regardless of the variation at position 6 and apparently a reduced binding footprint.

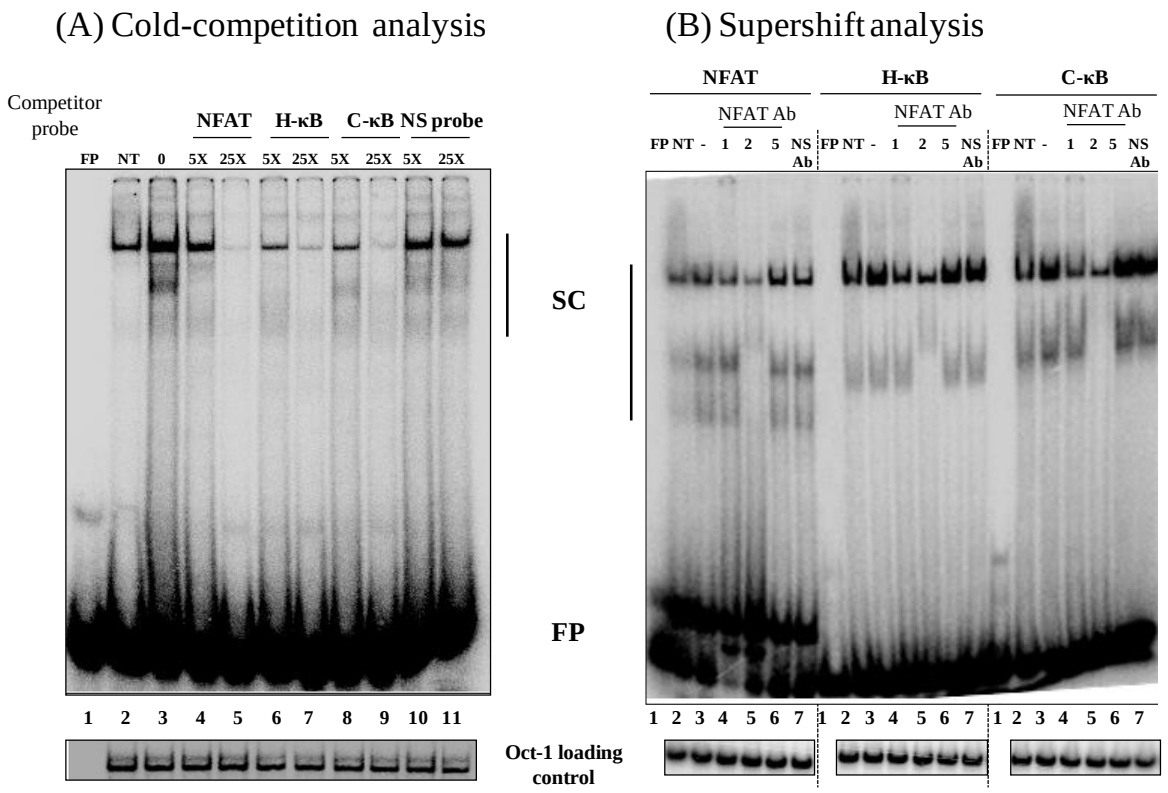


Figure-3.13: The C-κB motif demonstrates the binding of NFAT. **(A)** Cold-competition analysis. A radio-labelled double-stranded probe (40,000 cpm) derived from IL-2 promoter sequence, containing the NFAT binding site was incubated with 30 µg of nuclear extract prepared from Jurkat cells after PMA/ionomycin activation for 1 h or without activation . The competition was performed by incubating the complexes with 5- or 25-fold excess of cold probes (NFAT, H-κB, C-κB, or a Nonspecific probe). **(B)** Supershift analysis. PMA/ionomycin treated Jurkat nuclear extract was incubated with specific antibodies against factors NFAT1, 2 or 5. The nuclear extract was preincubated with the antibodies prior to the addition of the appropriate probe. Oct-1 was used as a loading control. FP- Free probe, SC- shifted complexes, NT- No PMA/ionomycin treatment, NS- non-specific probe, NS Ab- A non-specific anti-p24 antibody. The data is the representation of three experiments.

To find the identity of the specific NFAT family members present in the nuclear complexes bound to the probes, we performed a super-shift analysis using commercial antibodies specific to three of the important members of the family - NFAT1, 2 and 5. The nuclear extract (30 μg total protein) prepared from PMA/ionomycin induced Jurkat cells was preincubated with 2 μg of the antibodies and radio-labelled NFAT, H-κB or C-κB probes were added to appropriate reactions. The resolved nuclear bands clearly demonstrated the presence of NFAT1 and 2, but not 5, in the nuclear complexes recruited by all the three DNA probes including the C-κB probe (Figure-3.13 B; compare lane 3 with 4, 5 and 6). Although a super-shifted complex is not visible under any of the experimental conditions, even with the NFAT positive control probe, the reduced intensity of the bands when specific antibodies are added to the reaction confirmed the presence of the specific host factors (Figure-3.13 B, compare lanes 3 with lanes 4 and 5). Importantly, a non-specific antibody (anti-p24 of HIV-1 raised in our laboratory) used as negative control did not reduce the intensity of the shifted bands (Figure-3.13, compare lanes 3 and 7). In summary, the C-κB motif of subtype C can efficiently bind NFAT hence should be capable of exploiting the NFAT-mediated signalling to modulate gene expression from the viral promoter of subtype C.

3.2.8 Functional association between the C-κB and Sp1 III motifs in the chromtinized viral LTR

The expression analysis of the viral promoter using reporter vectors represents only the transient nature of gene expression without capturing the expression profile of the integrated provirus. To evaluate gene expression from an integrated provirus, we generated a panel of four reporter viruses (X:X, C:X, X:C and C:C) that are analogous to the corresponding reporter vectors lacking the flanking κB and Sp1 elements (Figure-3.8). The

reporter viruses were generated using the pcLGIT vector reported previously that coexpressed EGFP and Tat under the control of the LTR (Bachu *et al.* 2012). Daughter viruses produced from these vectors would copy the 3' U3 into the 5' viral LTR thus placing the expression of the reporter gene EGFP and that of Tat under the control of the engineered LTRs. The profile of gene expression from the viral promoter and the binding of diverse transcription factors to the viral enhancer could therefore be compared following viral integration into the chromatin. Viruses pseudotyped with the VSV-G envelope were produced in HEK 293T cells and used for the infection of Jurkat T-cells at a low m.o.i. of 0.01 to avoid multiple integrations of the cells. Cell pools stably expressing EGFP were sorted 21 days after the viral infection and used in the assay with or without TNF-α activation. The three reporter viruses that lacked C-κB site (X:C), Sp1 III site (C:X) or both (X:X) demonstrated minimal basal level activity that did not increase following TNF-α activation (Figure-3.14 A). Only the reporter virus containing both the motifs (C:C) demonstrated measurable basal level fluorescence (2,127.5 ± 9.2 MFI) that increased 2.2-fold (4,783.0 ± 114.6 MFI, $p<0.05$) following TNF-α activation.

We used a real-time PCR targeting the R-U5 region of the LTR to confirm equivalent levels of viral infection thus ruling out the possibility of differential infection rates underlying the experimental outcome (Figure-3.14 B). The PCR was performed using 100 ng of the genomic DNA extracted from the infected Jurkat cells and by following the protocol as described in the methods (Section-2.7).

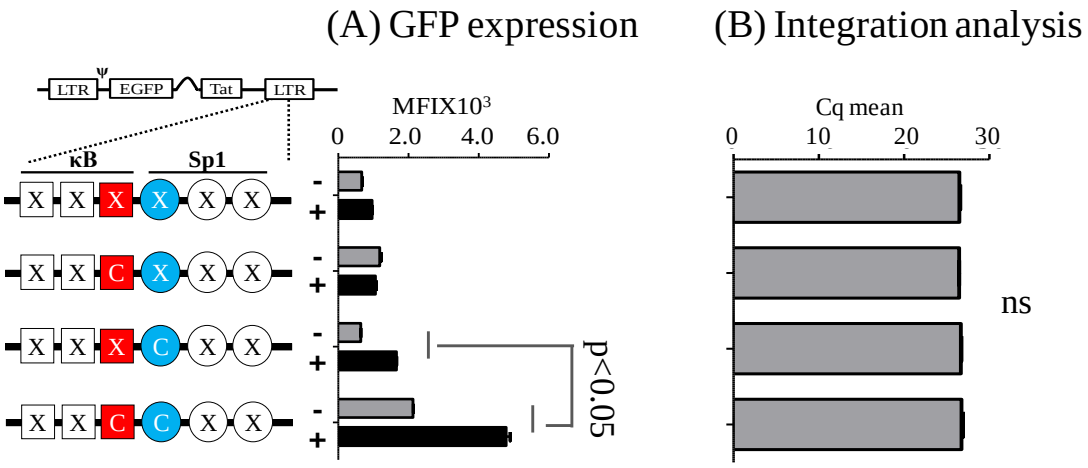


Figure-3.14: Functional association between the C-κB and Sp1 III motifs in the context of the chromatinized virus. **(A)** Flow cytometry analysis to measure GFP expression from promoter variant and pseudotyped reporter viruses. The reporter viruses are pseudotyped with VSV-G envelope and coexpress EGFP and Tat from the LTR. Stable Jurkat cell-pools were generated harbouring the C-LTR containing only the C-κB site (C:X), only the Sp1 III site (X:C), both (C:C) or none (XX) intact in the integrated form. Jurkat cells were activated with TNF-α (100 ng/ml) for 24 h or left without activation and the mean fluroscence intensity (MFI) of the EGFP was measured using the BD flow-cytometer AriaIII. Data are representative of three independent experiments and are presented as a MFI ± S.D. Each assay was performed in duplicate wells. Two-way ANOVA was used for the statistical analysis. **(B)** Real-time PCR analysis confirms comparable levels of integration of the four different reporter viruses in stable Jurkat cell-pools. Genomic DNA was extracted from the cell-pools and subjected to quantitative PCR, amplifying the 129 bp region spanning the R-U5 region. Data are presented as mean Cq ± S.D. Data was normalized against the house-keeping gene, GAPDH.

To correlate the magnitude of gene expression from the viral promoter with the occupancy of the host factors at the promoter, we performed a chromatin immunoprecipitation (ChIP) analysis using genomic DNA extracted from Jurkat cells harbouring integrated LTR X:X and C:C. Fragmented chromatin was immunoprecipitated using antibodies specific for p50, p65, Sp1, RNA polII, RNA polII-Ser-2 or a non-specific antibody against the p24 antigen as a negative control. The X:X viral promoter did not bind

any of the tested host factors with or without TNF-α activation thus ascertaining the lack of viral proliferation. In contrast, the promoter containing intact C-κB and Sp1 III sites, in the absence of activation, bound only p50 and Sp1, but little p65 or RNA polymerase suggesting the occupancy of the C-κB site by the repressive p50 homodimer. Following TNF-α activation, however, the promoter was occupied by p50, p65, Sp1 and significantly higher quantities of RNA polymerase that is phosphorylated on serine 2. Thus following activation, the suppressive p50:p50 homodimer was replaced by a transcription-supportive p50-p65 heterodimer complex at the viral enhancer. Additionally, the presence of the phosphorylated RNA pol II is suggestive of efficient transcription elongation. The relative magnitude of the presence of Sp1 on the promoter was not significantly different under the basal and induced conditions. Sp1 can function as a suppressive factor by recruiting HDACs and as a transcription inducer by recruiting HATs. The ChIP data are not only consistent with the proliferation analysis of the reporter viruses but also confirmed that regardless of the subtype-specific genetic variations that characterize the C-κB and the Sp1 III motifs, these sequences regulate gene expression by recruiting the appropriate host factors (Figure-3.15). Collectively, the data ascertained the pivotal role that the association between the C-κB and Sp1 III sites plays role in regulating gene expression from subtype C viral promoter in the absence or presence of the flanking regulatory elements.

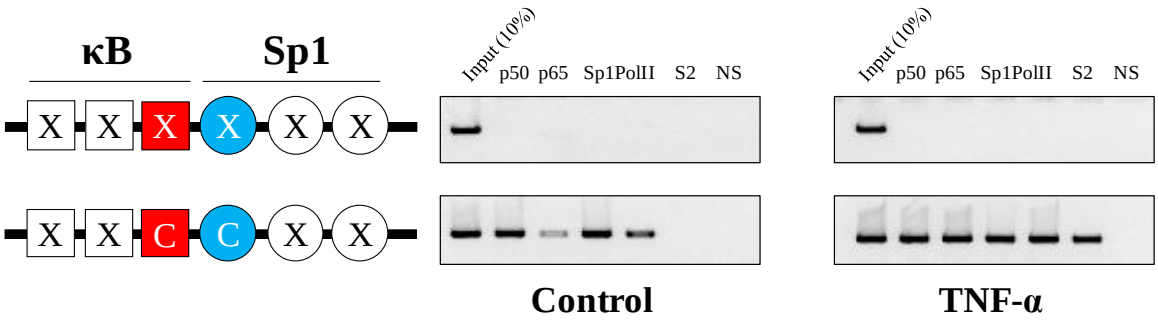


Figure-3.15: ChIP analysis demonstrates the occupancy of transcription factors to TFBS in the variant LTRs. Jurkat stable cell-pools harbouring the LTR (X:X or C:C) were activated with TNF-α for 4 h or left without activation. Cellular complexes were immunoprecipitated using 2 µg of specific antibodies as shown. An antibody against the HIV-1 p24 was used as a negative control in the experiment. One-tenth of the input chromatin was loaded as the input control.

3.2.9 Subtype C specific Sp1 III site functions efficiently in association with the canonical H-κB motif in the context of a reporter virus

The C-κB motif (GGGGCGTTCC) may be considered as the hallmark of subtype C given its exclusive presence restricted only to this subtype and its complete absence from all other HIV-1, HIV-2 and SIV subtypes. All of the non-C HIV-1 subtypes and HIV-2 and SIV contain one or two copies of the generic H-κB site (GGGACTTTCC) constituting the viral enhancer. The C- and H-κB sites differ from each other at two positions. Unlike the Sp1 III motif that is characterised by broader and subtype-specific variations, the Sp1 proximal NF-κB site does not demonstrate a wider and subtype-specific variations with the exception of subtype C. Considering the more universal presence of the H-κB site in the lentiviral enhancers, we asked if this motif can function in associating with subtype C specific Sp1 III motif (GAGGTGTGGT). To this end, we constructed a panel of two reporter viral vectors that contain the Sp1 III site of subtype C with an upstream C- (C:C) or H-κB (H:C) motif and

with the flanking sites mutated (Figure-3.16). Jurkat cells were infected with equivalent titers of the reporter viruses, activated with TNF-α after 3 days and a flow analysis was performed after one day of the activation. The results were consistent with those of the reporter vectors (Figure-3.4 C and D). The chimera H:C-LTR was more responsive to the TNF-α activation as compared to the wild type C:C-LTR. The data collectively confirmed that the H- and C-κB motifs are functional substitutes for each other and that each can complement with the subtype C specific SP1 III site efficiently in the absence of the flanking regulatory sites.

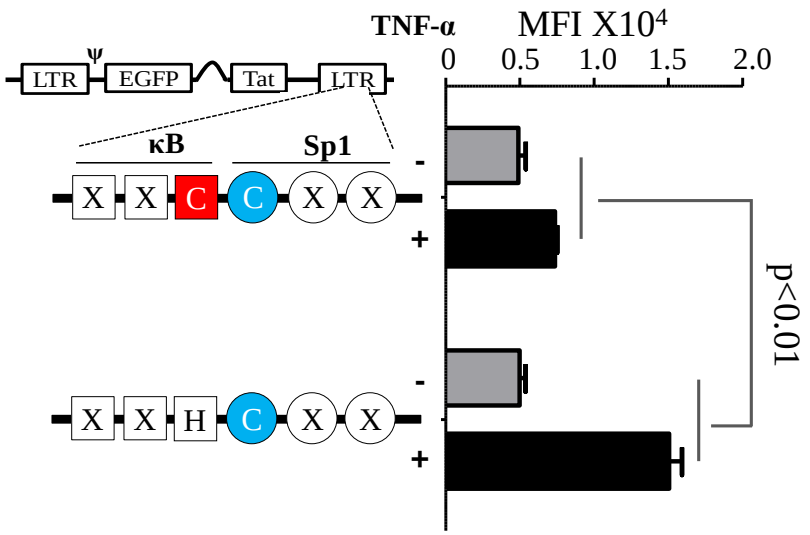


Figure-3.16: The H-κB motif can serve as a functional substitute for the C-κB site in complement with the Sp1 III element of subtype C. Jurkat cells were infected with pseudotyped viruses harbouring LTR (C:C or H:C) driving the expression of EGFP. Three days post-infection cells were activated with TNF-α. One day post-activation, cells were analyzed using the flow-cytometer for the EGFP expression. Data presented here are representative of the three experiments. Two-way ANOVA was used for the statistical analysis.

3.2.10 In a replication competent virus, replication of the HHH virus is severely compromised in target T-cells

The question whether the H-κB site can act as a functional substitute for the C-κB site remains unresolved from the data presented above. When in a reporter expression vector,

the C-κB motif and the upstream H-κB site were swapped (HHC vs HCH), a significant reduction in the transactivation ($p<0.05$) was observed in two different cell lines in a reproducible fashion (Figure-3.4 A and B). The reporter activity reduced further when the C-κB site was shifted further upstream (CHH, $p<0.05$). This result had suggested that the H-κB motif cannot be a functional substitute for the natural C-κB site in subtype C. The subsequent data, however, contradicted this observation in the context of reporter expression vectors as well as reporter viruses. The reporter expression vectors in the second set of experiments, however, contained only the central κB and Sp1III sites and lacked the flanking κB and Sp1 motifs (Figure-3.4 C and D). The H:C LTR, where the H-κB site was experimentally brought in association with the Sp1 III motif of subtype C demonstrated a significantly superior reporter activity ($p<0.01$) as compared with the native C:C LTR. Consistent with this finding, reporter viruses containing only the central κB and Sp1III motifs (H:C, Figure-3.16) too functioned as efficiently as the viral counterparts that contained the natural configuration. The second set of experiments, thus ascertained that the H-κB motif can be as functional as the C-κB site under the above experimental conditions. The reporter viruses are, however, capable of undergoing only one round of infection as they are pseudotyped and lack an inherited envelope. In other words, the gene expression quantitated from the reporter viruses is the outcome of a single-round of infection unlike in the case of infectious viruses that can establish multiple rounds of infection.

To resolve this conflict between the first and second sets of experiments, we constructed a pair of infectious viral clones (HHC and HHH) based on the subtype C prototype molecular clone Indie-C1 (Mochizuki *et al.* 1999). Using an overlap PCR, two

mutations were introduced in the C-κB motif, at positions 4 (G to A) and 6 (G to T) to convert the site into H-κB. The mutated LTR (HHH) was substituted at the 3' end of the Indie molecular clone using unique restriction sites thus generating the HHH Indie molecular clone. Except for the two base-pair variations in the central κB element of the 3' LTR, the two sister molecular clones (HHC and HHH) were identical genetically in the rest of the viral backbone (Figure-3.17). The Sp1 motif proximal to NF-κB site in both the viruses was subtype C specific. Daughter viruses produced from these vectors would copy the 3' U3 element into the 5' viral promoter thus placing the expression of the viral genes under the control of the engineered LTRs. The replication kinetics of the daughter viruses derived from the two viral molecular clones was compared in different cells under diverse experimental conditions.

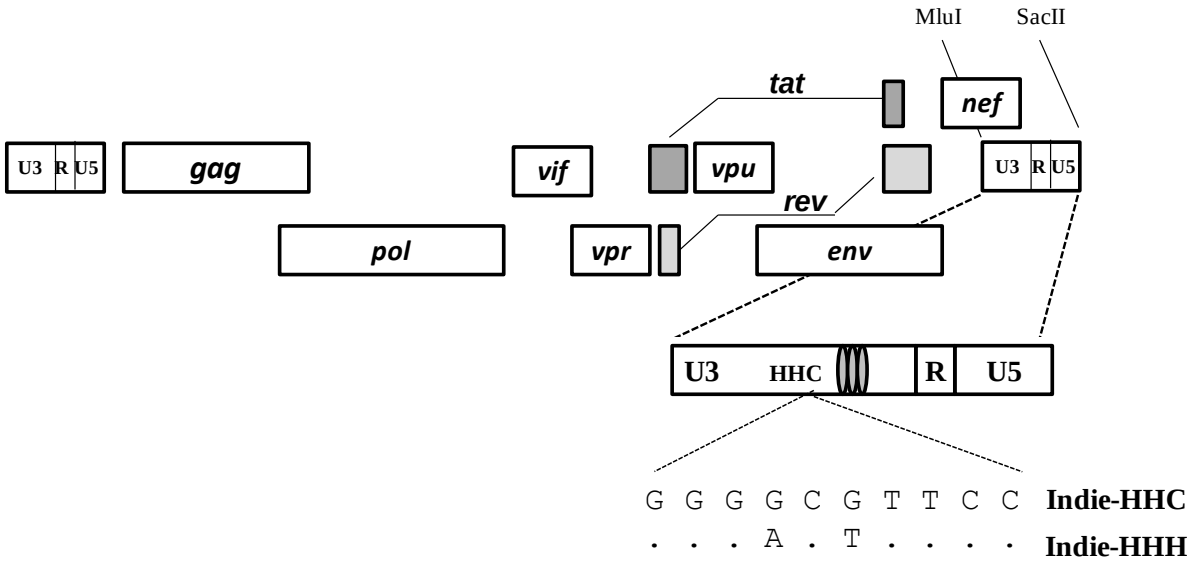


Figure-3.17: Schematic representation of the HIV-1 Indie-HHC and Indie-HHH sister viruses. The two viral clones are genetically identical with the exception of a difference at two positions in the central NF-κB binding site as shown.

Viral stocks from HHC and HHH molecular clones were generated in HEK293T cells and the $TCID_{50}$ titers of the stocks were determined in Tzm-bL cells using standard procedures. Target cells were infected with a fixed titer, 500 $TCID_{50}$, of either of the viruses and viral replication was monitored at weekly intervals for approximately a month by determining p24 quantity in the culture supernatant. Fresh PBMC isolated from different donors were depleted of the $CD8^{+ve}$ cells, activated with PHA (5 μg/ml) and IL-2 (20 U/ml) and used in the assay. The proliferation of the wild type HHC virus demonstrated a normal profile in most of the PBMC derived from 6 different subjects by peaking approximately around day 14 and falling thereafter (Figure-3.18). In contrast, the HHH virus surprisingly failed to show any significant levels of proliferation in 5 of the 6 donor cells. The difference in the proliferation kinetics was significant between the two viruses. Although in the cells of donors-1, 2 and 5, the HHH virus showed a minimal level of replication, the viral proliferation was not restored even after a month in the culture. The replication profile of the two viruses was highly reproducible in CEM-CCR5, a human T-cell line that coexpresses the coreceptor CCR5 along with CD4, hence can be infected by subtype C viral strains that require these two receptors for a productive infection. While the wild type HHC virus demonstrated the standard replication profile, the HHH virus totally failed to proliferate in these cells. The difference in the replication properties could also be recapitulated in CN-2 cells (Figure-3.18) which are human PBMC transformed by herpesvirus saimiri strain, yet retain many biological properties of the primary cells (Vella *et al.* 2002;Zheng *et al.* 2002). In all, the defect in the replication properties of the HHH virus was highly reproducible.

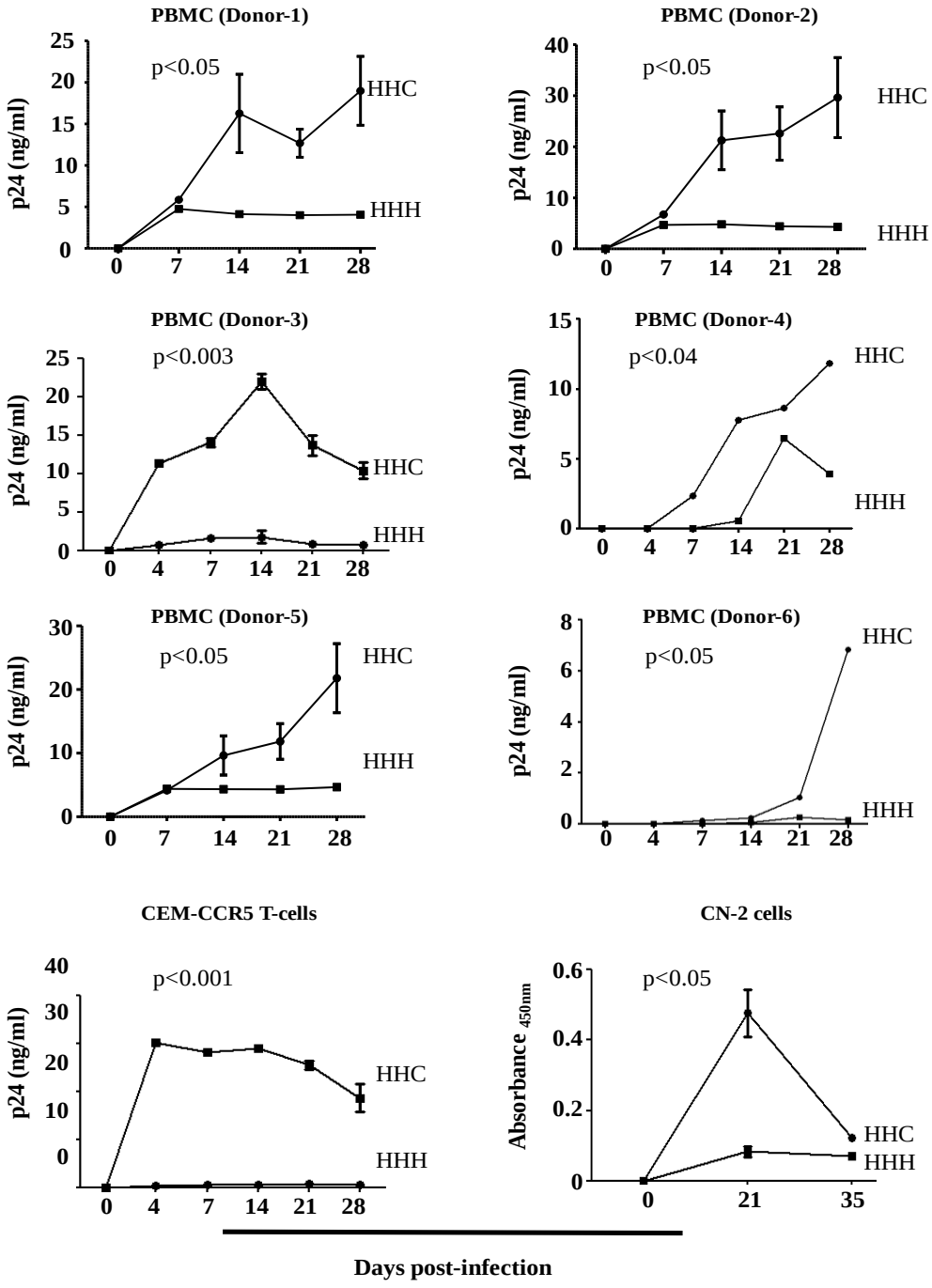


Figure-3.18: Replicative kinetics of the Indie HHC and Indie HHH viruses in different cells. CD8+ cells depleted and mitogen-activated PBMCs, CEM-CCR5 T-cells and herpes virus saimiri strain transformed CN-2 cells were infected with 500 TCID50 of Indie-HHC or Indie-HHH viruses. Secretion of the p24 was monitored at regular intervals for several weeks. The data are presented as a mean of triplicate wells ± S.D. A two-tailed unpaired t-test was used for the stastical analysis.

To rule out the possibility of the viral stock instability, we determined the $TCID_{50}$ titers of the viral stocks, before and after the viral proliferation assay. The viral stocks of both the viral strains HHC and HHH were quite stable before and after the assay (Figure-3.19) and over several months (data not shown). A 4-fold difference in the titres of the viral stocks was evident probably recapitulating the replication differences between the two promoters when the viral stocks were prepared in HEK 293T cells.

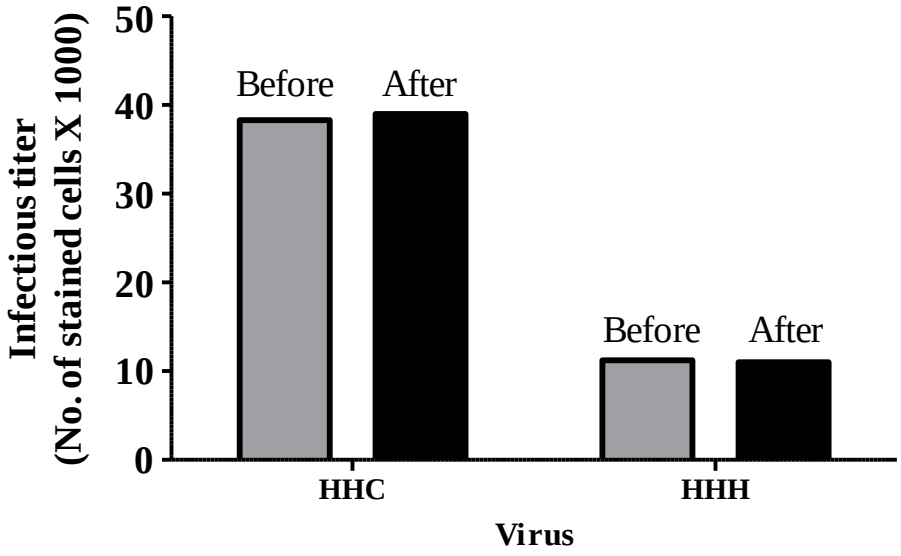


Figure-3.19: Confirmation of the viral titres before and after the experiment for the HHC and HHH viruses. TCID50 was estimated in Tzm-bL cells using β-gal staining. Each bar is representative of the mean of triplicate wells.

The discordant results between the reporter plasmids (Figure-3.4 C and D), reporter viral strains (Figure-3.16) and replication competent viruses (HHC:C and HHH:C, Figure-3.18) was confusing. The reporter viruses contained a minimal LTR that lacked the flanking TFBS mutated. To understand if the discordance can be ascribed to integration status of the

viral LTR as well as the presence of all the 6 TFBS in the viral enhancer:core region, we generated two different and additional panels of expression vectors. One panel contained plasmid reporter vectors expressing GFP and luciferase under the control of two variant LTRs (HHC:C wild type and HHH:C chimera) containing all the three, each NF-κB and Sp1 sites intact. The previous reporter plasmid panel contained three other viral LTRs, but all of them contained the C-κB site (HHC, HCH and CHH, Figure-3.4 A and B). A new viral LTR HHH:C has been introduced in the present panel that lacked the C-κB motif. An analogous panel of reporter viral vector was also constructed (HHC:C and HHH:C) (Figure-3.20). The reporter viruses are capable of infecting the cells for only a single round. Using these two different expression panels that contained an intact viral enhancer:core, we asked if the NF-κB:Sp1 incompatibility could be reproduced when the LTR is in the unintegrated (plasmid vectors) or integrated (reporter viruses) form. Jurkat cells were transfected with either of the two reporter vectors (HHC:C or HHH:C), the cells were treated with TNF-α or left without activation and the luciferase activity was measured at 24 h. In parallel, Jurkat cells were infected at a low m.o.i. 0.01 of VSV-G pseudotyped reporter viruses (HHC:C or HHH:C LTR), three days post infection cells were activated with TNF-α (100 ng/ml) for 24 h or left without activation and the mean fluorescence intensity of GFP was measuredred using flow-cytometry. Regardless of the minor, but significant, differences between the two different LTRs in the basal level promoter expression and induced transactivation, no transcription incompatibility was evident between the H-κB motif and subtype-specific Sp1III site whether in the unintegrated (Figure-3.20 A) or integrated forms (Figure-3.20 B). Collectively, putting all the data together, the transcriptional incompatibility, between the H-κB motif and subtype-specific Sp1III site, thus was consistently manifested only in the context of the infectious and

full-length viruses (Figure-3.18), but not in any other context. The transcriptional restriction was manifested to a smaller extent in plasmid reporter constructs (Figure-3.4 A and B) and not at all in all other experimental formats, plasmid reporter vectors (Figure-3.4 C and D) or reporter viral constructs (Figure-3.16), regardless of whether the viral enhancer:core contained only the two central κB and Sp1III sites or all the six TFBS.

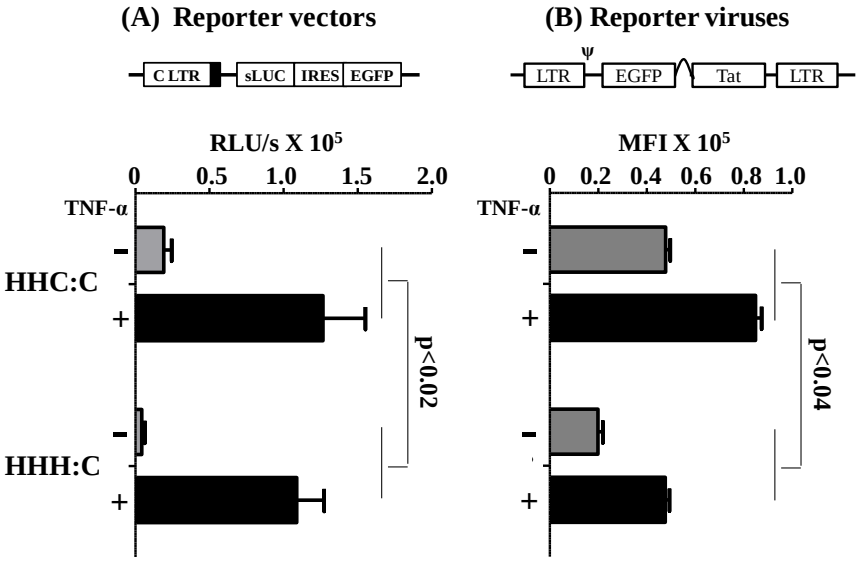


Figure-3.20: Lucifearse assay demonstrating the comparison between the gene expression from HHC:C and HHH:C LTRs. **(A)** Luciferase reporter vectors were transfected in Jurkat cells and cells were activated with TNF-α (100 ng/ml) or left untreated. The data of reporter vectors are representative of two independent experiments. Each assay was performed in triplicate wells. **(B)** Flow-cytometry analysis of Jurkat cells infected with pseudotyped viruses harbouring two different LTRs (HHC:C or HHH:C) driving the expression of EGFP. The Data are the representative of two independent experiments. Two-way ANOVA was used for the statistical analysis.

3.2.11 Substitution of the homologous Sp1 III site restores the replication competence of the HHH viral strain

In Indie-HHH virus, the H-κB site was juxtaposed with the Sp1 III motif of subtype C, a context that is not natural. The Sp1III site of subtype C is characterized by subtype-

specific variations at two positions as compared with the counterpart sequence of subtype B (GAGGCGTGGC in subtype B and GAGGTGTGGT in subtype C, differences underlined). In the Indie-C1 backbone, when the C-κB site was changed into H-κB, by introducing variations at two different positions, the newly created H-κB site was juxtaposed with the subtype C Sp1 III motif. If the artificial combination of the H-κB site and the subtype C Sp1 III motif are not biologically compatible with each other, the restricted replication of the HHH virus in T-cells can be explained. To test the hypothesis that specific combinations of NF-κB and Sp1 elements may not be biologically functional together, we generated two additional viral variants thus increasing the number of the viral strains to four in the panel. A new labelling system was adapted to identify each of the 4 viral strains that contain different combinations of the two different NF-κB sites (H- versus C-κB sites) and the two different Sp1III motifs (B and C unique for subtypes B and C, respectively). In the revised labelling system, the wild type Indie viral strain HHC is labelled as HHC:C, the letter C following the colon representing the subtype identity of the Sp1 motif (Figure-3.21). The variant viral strain HHH that demonstrated restricted proliferation is designated as HHH:C in the revised labelling system. The two new viral strains where the Sp1III site of subtype C was replaced by the subtype B counterpart were designated as HHC:B and HHH:B. We reasoned that if the viral strain HHH:C is devoid of replication fitness due to the incompatible combination of the H-κB motif and the Sp1 III site of subtype C, substituting the Sp1 III motif with that of subtype B (HHH:B) should restore the lost replication competence of the viral strain.

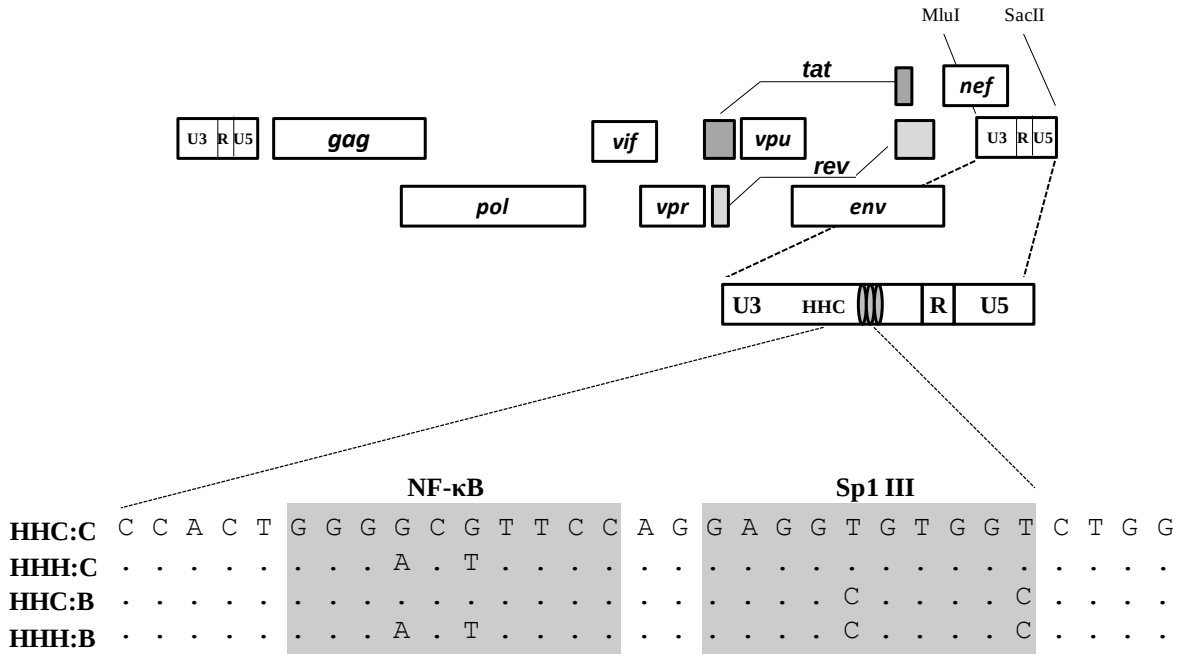


Figure-3.21: A multiple sequence alignment of the central NF-κB and Sp1III motifs. A schematic of the full-length viruses is shown on the top. The lower panel depicts the multiple sequence alignment of the central NF-κB and Sp1III regions of the four replication competent viruses. The viruses have been labelled according to the new labelling system. The letters before the colon represent the NF-κB sites and the letters after the colon the subtype nature of the central Sp1 motif. Two other downstream Sp1 sites are intact, but not shown.

CEM-CCR5 cells were infected with each of the four viral strains independently and the secretion of the viral antigen p24 into the culture medium was determined on days 0, 3 and 7. The differences in the replication kinetics of the viral strains are typically established by day 7 hence we did not monitor the viral proliferation beyond this time point. While the wild type HHC:C viral strain demonstrated a typical replication profile with a progressively increasing p24 concentration at days 3 and 7, the HHH:C virus failed to show a significant level of viral proliferation as was seen previously (Figure-3.18). Importantly, the newly generated HHH:B virus where the central H-κB site was recombined with the homologous Sp1 III site of subtype B demonstrated unrestricted viral proliferation which was as efficient as that of the

wild type HHC:C viral strain (Figure-3.22). Additionally, the fourth virus HHC:B also proliferated as efficiently as the wild type HHC:C viral strain without demonstrating any signs of replication incompetence. Collectively, the data ascertained that the Sp1 III motif of subtype B functioned in combination with either the H- or C-κB elements (HHH:B or HHC:B) with comparable efficiency. In contrasr, the Sp1 III motif of subtype C can function only in association with the C-κB element (HHC:C), but not the H-κB element (HHH:C).

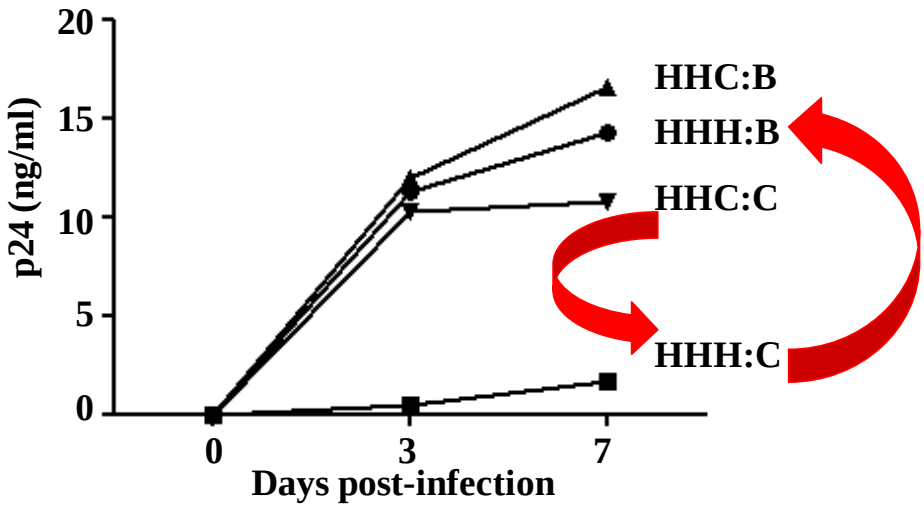


Figure-3.22: Replication kinetics of the four viruses HHC:C, HHH:C, HHC:B and HHH:B in CEM-CCR5 T-cells. CEM-CCR5 cells were infected with 500 TCID50 of one of the four viruses. The secretion of p24 into the medium was monitored for up to a week. The assay was performed in triplicate wells and the data were plotted as the mean ± S.D.

3.2.12 Post-entry events of the viral life cycle

NF-κB plays an important role in the regulation of basal level transactivation from the viral promoter in the absence of Tat. To understand at what level of the viral life cycle the virus containing variations in the NF-κB and Sp1 motifs might achieve replication advantage,

CEM-CCR5 T-cells were infected independently with HHC:C, HHH:C, HHC:B or HHH:B viral strains at equivalent titers (500 $TCID_{50}$ units) (Figure-3.23). Using real time PCR targeting different stages of the viral life cycle, we examined the proliferation of all the four viral strains at two different stages, the generation of the reverse transcription products in the cell extract and the formation of proximal *versus* distal viral transcripts in the nucleus following integration. Twelve hour following the viral infection, the reverse transcription products were probed using real-time PCR that targeted a 142-bp fragment from the U5-ψ region. All the four viral strains demonstrated comparable levels of reverse transcription products suggesting comparable magnitude of infection of the target cells by all the viruses including the HHH:C viral strain that had demonstrated replication deficiency (Figure-3.23 B).

To evaluate the transcripts expressed from the viral promoters, two different PCRs were performed (Figure-3.23 A) one for the proximal viral transcripts within the TAR region and the other for distal transcripts in the Tat gene located ~5.4 kb downstream of the transcription start site (Bachu *et al.* 2012). Additionally, the target cells were activated with TNF-α or left without activation following viral infection. RNA extracted from the cells was subjected for the cDNA synthesis and used in the PCR as template.

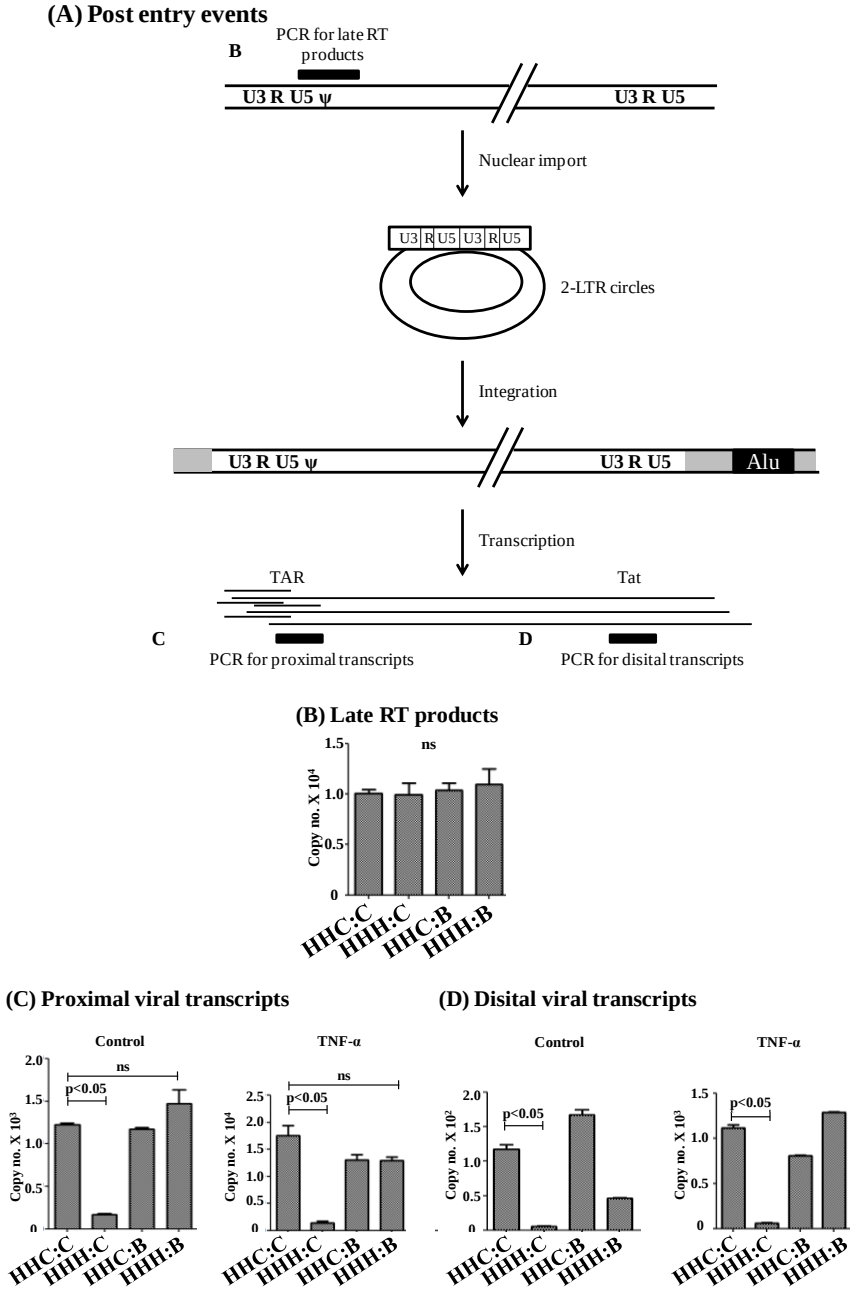


Figure-3.23: Analysis of the post entry events of the four viruses HHC:C, HHH:C, HHC:B and HHH:B in CEM-CCR5 T-cells. **(A)** Schemtaic representation of the post entry events of HIV-1 replication. Filled boxes represent the target sequences for the PCR amplification. **(B)** Analysis of the late RT products. CEM-CCR5 cells were infected with 500 TCID50 of one of the four viruses. 12 h post infection, genomic DNA was extracted and subjected to the qunatitaive PCR. A 145 bp DNA fragment spanning the U5-ψ region was amplified. Quantification of the **(C)** proximal and **(D)** Distal viral transcripts. 48 h post infection, the total cellular RNA was extracted and subjected to the cDNA synthesis. Quantitative real-time PCR was performed with the cDNA as described in section 2.10. The proximal transcripts were detected using primers amplifying a 89-bp region in the TAR region, whereas a 152 bp was amplified in Tat region to detect the distal transcripts approximately 5.4 kb downstream of the transcription start site. Data are presented as the mean copy no. ± S.D. A two-tailed unpaired t-test was used for the statistical analysis.

Significantly higher levels of viral transcripts, from both proximal and distal transcripts, were generated from three of the four viral strains HHC:C, HHC:B and HHH:B in the absence of the TNF-α induction of the cells and in its presence. The HHH:C virus in contrast failed to demonstrate considerable transcription under any of the experimental conditions (Figure-3.23 C and D). The data of the transcription analysis thus are consistent with the viral proliferation kinetics and mapped the replication defect of the HHH:C viral strain to the lack of transcriptional activity from this viral promoter.

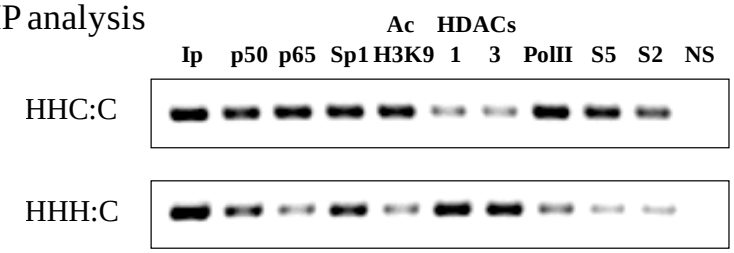
3.2.13 The epigenetic landscape of the viral promoters

The restricted viral proliferation (Figure-3.18 and 3.22) and the compromised transcription activation (Figure-3.23 C and D) from the non-functional HHH:C LTR, as compared to the other three viral promoters, are intriguing. The data collectively suggested the lack of efficient transcription from the HHH:C LTR. It is unexpected that two different transcription factors, NF-κB and Sp1 that function together synergistically regulating transcription form HIV-1 LTR (Perkins *et al.* 1993), fail to demonstrate gene expression when the corresponding binding elements are experimentally positioned together in the HHH:C LTR. Importantly, the two central binding sites show efficient transactivation in other closely related viral LTRs. To the best of our knowledge, this is the first experimental demonstration of a specific combination of NF-κB and Sp1 binding sites not being functional biologically in the context of HIV-1 promoter or in the context of any other promoter.

To gain more insights into the nature of the host factors recruited to the HHH:C LTR, we performed the chromatin immune precipitation analysis targeting several host factors and compared the profile with that of the wild type HHC:C LTR.

(A) ChIP analysis

.



(B) Densitometry analysis

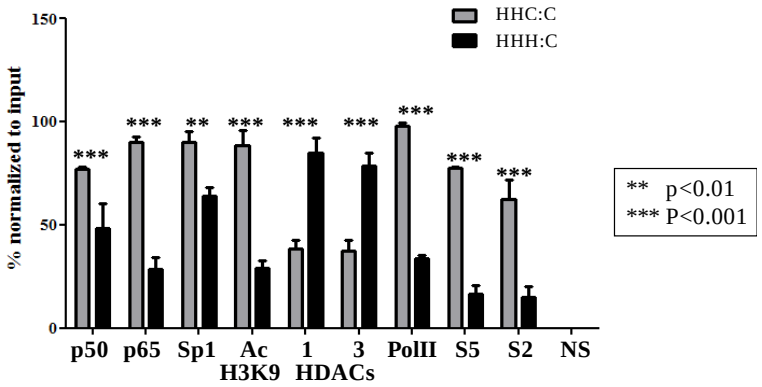


Figure-3.24: ChIP analysis to show the occupancy of different transcription factors and associated activation or repressive marks on HHC:C and HHH:C LTR. **(A)** CEM-CCR5 cells were infected with Indie-HHC:C and Indie-HHH:C viruses at an equivalent $TCID_{50}$. 48 h post infection the ChIP assay was performed. Immunoprecipitation of the complexes was performed with 2 µg of the antibodies as mentioned. An antibody against the p24 was used as a negative control in the experiment. One-tenth of the input chromatin was loaded as input control. **(B)** Densitometry analysis of the data of the top panel. Data for each band are normalized to the input.

CEM-CCR5 cells were infected at equivalent $TCID_{50}$ titers with the wild type Indie HHC:C or Indie HHH:C viral strains. Two days following the infection, a ChIP assay was performed using antibodies against various host transcription factors (p50, p65 and Sp1), chromatin modulators (HDAC1, HDAC3), epigenetic marks (H3K9Ac), RNA pol II or different phosphorylated forms of RNA pol II (Ser-2 or Ser-5). The assay results showed a

transcriptionally active chromatin configuration in the case of the wild type HHC:C LTR and in contrast, a predominantly suppressive chromatin landscape in the case of the restricted HHH:C LTR (Figure-3.24 A). In the HHC:C LTR, a significantly higher levels of p50:p65 heterodimer, acetylated H3K9, pol II and Pol II phosphorylated on serine 2 and serine 5 were identified as compared to the HHH:C promoter. The phosphorylated forms of serine 5 and serine 2 of Pol II are indicative of higher magnitude of transcription initiation and elongation, respectively (Coiras *et al.* 2013). In contrast, the presence of the transcription suppression marks HDAC 1 and 3 was significantly higher on the restricted HHH:C LTR as compared with the HHC:C LTR (Figure-3.24 B). In summary, the analysis of the chromatin configuration using the ChIP assay provided a conclusive experimental evidence that the lack of replication fitness of this viral strain as compared with the wild type HHC:C virus is associated with the lesser recruitment of the transcription machinery to the HHH:C viral promoter .

Based on the data, we propose the following model (Figure-3.25) to explain the selective transcriptional silencing of the HHH:C viral LTR. A unique combination of NF-κB and Sp1 transcription factor complex occupying the central H-κB and the subtype C unique Sp1 III elements of the HHH:C viral promoter selectively recruits one or more cellular cofactors that mediate chromatin suppression by recruiting the HDAC complex. In contrast, the complex occupying the central C-κB and the subtype C unique Sp1 III elements of the wild type HHC:C LTR is qualitatively different than that of the chimera HHH:C LTR and recruits the cofactors that recruit HATs and promote the transcriptional activity from the viral promoter. Efforts are presently in progress to identify the cellular cofactors that are responsible for the transcriptional suppression from the HHH:C viral promoter.

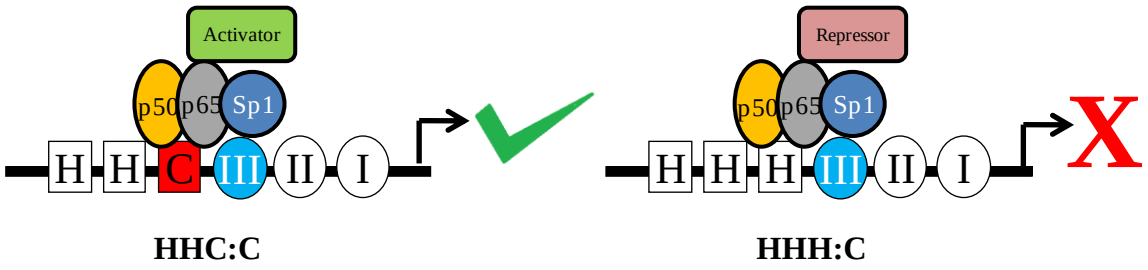


Figure-3.25: A model depicting the transcriptional silencing of the HHH:C LTR in the full-length virus context.

3.3 Bibliography

Bachu, M., S. Yalla, M. Asokan, A. Verma, U. Neogi, S. Sharma, R.V. Murali, A.B. Mukthey, R. Bhatt, S. Chatterjee, R.E. Rajan, N. Cheedarla, V.S. Yadavalli, A. Mahadevan, S.K. Shankar, N. Rajagopalan, A. Shet, S. Saravanan, P. Balakrishnan, S. Solomon, M. Vajpayee, K.S. Satish, T.K. Kundu, K.T. Jeang and U. Ranga, 2012. Multiple NF-kappaB sites in HIV-1 subtype C long terminal repeat confer superior magnitude of transcription and thereby the enhanced viral predominance. J.Biol.Chem. 287: 44714-44735. Available from PM:23132857

Badran, B.M., S.M. Wolinsky, A. Burny and K.E. Willard-Gallo, 2002. Identification of three NFAT binding motifs in the 5'-upstream region of the human CD3gamma gene that differentially bind NFATc1, NFATc2, and NF-kappa B p50. J.Biol.Chem. 277: 47136-47148. Available from PM:12374807

Beinoraviciute-Kellner, R., G. Lipps and G. Krauss, 2005. In vitro selection of DNA binding sites for ABF1 protein from Saccharomyces cerevisiae. FEBS Lett. 579: 4535-4540. Available from PM:16083878

Burnett, J.C., K. Miller-Jensen, P.S. Shah, A.P. Arkin and D.V. Schaffer, 2009. Control of stochastic gene expression by host factors at the HIV promoter. PLoS.Pathog. 5: e1000260. Available from PM:19132086

Coiras, M., M. Montes, I. Montanuy, M.R. Lopez-Huertas, E. Mateos, S.C. Le, M.A. Garcia-Blanco, C. Hernandez-Munain, J. Alcami and C. Sune, 2013. Transcription elongation regulator 1 (TCERG1) regulates competent RNA polymerase II-mediated elongation of HIV-1 transcription and facilitates efficient viral replication. Retrovirology. 10: 124. Available from PM:24165037

Colin, L., L.C. Van, 2009. Molecular control of HIV-1 postintegration latency: implications for the development of new therapeutic strategies. Retrovirology. 6: 111. Available from PM:19961595

Dahiya, S., Y. Liu, M.R. Nonnemacher, W. Dampier and B. Wigdahl, 2014. CCAAT enhancer binding protein and nuclear factor of activated T cells regulate HIV-1 LTR via a novel conserved downstream site in cells of the monocyte-macrophage lineage. PLoS.One. 9: e88116. Available from PM:24551078

Hogan, P.G., L. Chen, J. Nardone and A. Rao, 2003. Transcriptional regulation by calcium, calcineurin, and NFAT. Genes Dev. 17: 2205-2232. Available from PM:12975316

Jeeninga, R.E., M. Hoogenkamp, M. rmand-Ugon, B.M. de, K. Verhoef and B. Berkhout, 2000. Functional differences between the long terminal repeat transcriptional promoters of

human immunodeficiency virus type 1 subtypes A through G. J.Virol. 74: 3740-3751. Available from PM:10729149

Kinoshita, S., L. Su, M. Amano, L.A. Timmerman, H. Kaneshima and G.P. Nolan, 1997. The T cell activation factor NF-ATc positively regulates HIV-1 replication and gene expression in T cells. Immunity. 6: 235-244. Available from PM:9075924

Kurosu, T., T. Mukai, S. Komoto, M.S. Ibrahim, Y.G. Li, T. Kobayashi, S. Tsuji and K. Ikuta, 2002. Human immunodeficiency virus type 1 subtype C exhibits higher transactivation activity of Tat than subtypes B and E. Microbiol.Immunol. 46: 787-799. Available from PM:12516777

Lemieux, A.M., M.E. Pare, B. Audet, E. Legault, S. Lefort, N. Boucher, S. Landry, O.T. van, B. Berkhout, M.H. Naghavi, M.J. Tremblay and B. Barbeau, 2004. T-cell activation leads to poor activation of the HIV-1 clade E long terminal repeat and weak association of nuclear factor-kappaB and NFAT with its enhancer region. J.Biol.Chem. 279: 52949-52960. Available from PM:15466412

Li, Y., Z. Jiang, H. Chen and W.J. Ma, 2004. A modified quantitative EMSA and its application in the study of RNA--protein interactions. J.Biochem.Biophys.Methods 60: 85-96. Available from PM:15262445

Macian, F., C. Lopez-Rodriguez and A. Rao, 2001a. Partners in transcription: NFAT and AP-1. Oncogene 20: 2476-2489. Available from PM:11402342

McAllister, J.J., D. Phillips, S. Millhouse, J. Conner, T. Hogan, H.L. Ross and B. Wigdahl, 2000. Analysis of the HIV-1 LTR NF-kappaB-proximal Sp site III: evidence for cell type-specific gene regulation and viral replication. Virology 274: 262-277. Available from PM:10964770

Mochizuki, N., N. Otsuka, K. Matsuo, T. Shiino, A. Kojima, T. Kurata, K. Sakai, N. Yamamoto, S. Isomura, T.N. Dhole, Y. Takebe, M. Matsuda and M. Tatsumi, 1999. An infectious DNA clone of HIV type 1 subtype C. AIDS Res.Hum.Retroviruses 15: 1321-1324. Available from PM:10505681

Montano, M.A., V.A. Novitsky, J.T. Blackard, N.L. Cho, D.A. Katzenstein and M. Essex, 1997. Divergent transcriptional regulation among expanding human immunodeficiency virus type 1 subtypes. J.Virol. 71: 8657-8665. Available from PM:9343223

Naghavi, M.H., M.O. Salminen, A. Sonnerborg and A. Vahlne, 1999a. DNA sequence of the long terminal repeat of human immunodeficiency virus type 1 subtype A through G. AIDS Res.Hum.Retroviruses 15: 485-488. Available from PM:10195759

Naghavi, M.H., S. Schwartz, A. Sonnerborg and A. Vahlne, 1999b. Long terminal repeat promoter/enhancer activity of different subtypes of HIV type 1. AIDS Res.Hum.Retroviruses 15: 1293-1303. Available from PM:10505678

Perkins, N.D., A.B. Agranoff, E. Pascal and G.J. Nabel, 1994. An interaction between the DNA-binding domains of RelA(p65) and Sp1 mediates human immunodeficiency virus gene activation. Mol.Cell Biol. 14: 6570-6583. Available from PM:7935378

Perkins, N.D., N.L. Edwards, C.S. Duckett, A.B. Agranoff, R.M. Schmid and G.J. Nabel, 1993. A cooperative interaction between NF-kappa B and Sp1 is required for HIV-1 enhancer activation. EMBO J. 12: 3551-3558. Available from PM:8253080

Pessler, F., R.Q. Cron, 2004. Reciprocal regulation of the nuclear factor of activated T cells and HIV-1Genes Immun. 5: 158-167. Available from PM:14762397

Romanchikova, N., V. Ivanova, C. Scheller, E. Jankevics, C. Jassoy and E. Serfling, 2003. NFAT transcription factors control HIV-1 expression through a binding site downstream of TAR region. Immunobiology 208: 361-365. Available from PM:14748509

Roof, P., M. Ricci, P. Genin, M.A. Montano, M. Essex, M.A. Wainberg, A. Gatignol and J. Hiscott, 2002. Differential regulation of HIV-1 clade-specific B, C, and E long terminal repeats by NF-kappaB and the Tat transactivator. Virology 296: 77-83. Available from PM:12036319

Siggers, T., A.B. Chang, A. Teixeira, D. Wong, K.J. Williams, B. Ahmed, J. Ragoussis, I.A. Udalova, S.T. Smale and M.L. Bulyk, 2012. Principles of dimer-specific gene regulation revealed by a comprehensive characterization of NF-kappaB family DNA binding 1. Nat.Immunol. 13: 95-102. Available from PM:22101729

Stroud, J.C., A. Oltman, A. Han, D.L. Bates and L. Chen, 2009. Structural basis of HIV-1 activation by NF-kappaB--a higher-order complex of p50:RelA bound to the HIV-1 LTR 1. J.Mol.Biol. 393: 98-112. Available from PM:19683540

Vella, C., N.N. Zheng, P. Easterbrook and R.S. Daniels, 2002. Herpesvirus saimiri-immortalized human lymphocytes: novel hosts for analyzing HIV type 1 in vitro neutralization. AIDS Res.Hum.Retroviruses 18: 933-946. Available from PM:12230936

Wong, D., A. Teixeira, S. Oikonomopoulos, P. Humburg, I.N. Lone, D. Saliba, T. Siggers, M. Bulyk, D. Angelov, S. Dimitrov, I.A. Udalova and J. Ragoussis, 2011. Extensive characterization of NF-kappaB binding uncovers non-canonical motifs and advances the interpretation of genetic functional traits Genome Biol. 12: R70. Available from PM:21801342

Yuste, E., A. Moya and C. Lopez-Galindez, 2002. Frequency-dependent selection in human immunodeficiency virus type 1 J.Gen.Virol. 83: 103-106. Available from PM:11752706

Zhang, M., A. Clausell, T. Robinson, J. Yin, E. Chen, L. Johnson, G. Weiss, S. Sabbaj, R.M. Lowe, F.H. Wagner, P.A. Goepfert, O. Kutsch and R.Q. Cron, 2012. Host factor transcriptional regulation contributes to preferential expression of HIV type 1 in IL-4-producing CD4 T cells. J.Immunol. 189: 2746-2757. Available from PM:22875803

Zheng, N.N., C. Vella, P.J. Easterbrook and R.S. Daniels, 2002. Selection following isolation of human immunodeficiency virus type 1 in peripheral blood mononuclear cells and herpesvirus saimiri-transformed T cells is comparable. J.Gen.Virol. 83: 1343-1352. Available from PM:12029149

CHAPTER-4

SUMMARY AND DISCUSSION

A single zoonotic transmission event approximately a hundred years ago is believed to have transmitted HIV-1 to human beings (Hemelaar *et al.* 2011). The virus thereafter underwent rapid divergent evolutionary changes at the genetic level as it expanded to different parts of the globe probably in an attempt to adapt to the local differences in the host factor landscape. The divergent evolutionary modifications of HIV-1 was probably the driving force for the origin of the genetically distinct viral subtypes that differ from one another to the extent of 30-35% in specific gene segments such as the envelope. Understanding subtype-specific genetic variations is critical since such variations are likely to influence subtype-associated biological differences that are likely to modulate the relative replication fitness properties of the viral subtypes, in a subtle or profound fashion. The rare switch or lack of switch in the coreceptor use in subtype C infection from CCR5 to CXCR4 is one such example for subtype-associated biological properties that still awaits an explanation (Coetzer *et al.* 2008;Abebe *et al.* 1999;Ping *et al.* 1999).

The long-terminal repeat of HIV-1 varies among the diverse viral subtypes up to 20-25% and is characterized by subtype-associated genetic difference although the overall promoter architecture remains broadly uniform. The molecular features associated within a viral subtype are highly conserved despite genetic variation. Since a single viral promoter drives the expression of all the viral genes in HIV-1, any profound change in the promoter architecture is likely to impact the overall replication fitness of the viral subtype. The LTR of subtype C is believed to be a stronger viral promoter as it is endowed with an enhancer containing three NF-κB binding sites unlike most other HIV-1 subtypes that contain only two such elements. Experimental evidence confirms higher level gene expression from an LTR containing 3 NF-κB binding sites as compared with a promoter containing only 2 such sites

(Abebe *et al.* 1999;Kurosu *et al.* 2002;Montano *et al.* 2000;Naghavi *et al.* 1999). It was proposed (Montano *et al.* 1997;Rodenburg *et al.* 2001) that the stronger viral promoter may have contributed significantly to the relative global expansion of subtype C which by itself is responsible for half of the global HIV-1 infections presently (Hemelaar *et al.* 2011a). Unlike the previous reports that examined only the quantitative aspects of the subtype C promoter, we attempted to evaluate the quantitative as well as the qualitative differences that contribute to viral transcription. Although genetic differences in the centrally located NF-κB and Sp1 elements have been reported previously (Bachu *et al.* 2012;Montano *et al.* 1997;van *et al.* 2004), the present study represents the first analysis to examine the functional implications of these genetic variations in subtype C.

In a typical HIV-1 subtype B LTR, while the cluster of the three Sp1 motifs regulates the basal level gene expression and the two NF-κB site combination controls the induced promoter activity, the centrally located NF-κB and Sp1 III elements synergistically play a more critical role in regulating transcription. In subtype C LTR, in the core of the promoter, a genetically variant NF-κB site and a subtype-specific Sp1III motif are found (Figure-1.6) and a new found functional association between the two variant sites is expected. Using different panels of reporter expression vectors that contained all the 6 TFBS intact or only the core κB and Sp1 sites, we confirmed that the two core TFBS work synergistically in an orientation- , distance-and position-dependent fashion. Importantly, the H-κB site, as compared to the native C-κB motif, demonstrated a suboptimal biological function in association with the subtype C specific Sp1III site (Figure-3.4 A, B) suggesting covariant evolution of the two central TFBS in subtype C. Interestingly, any of the four other variant forms of the Sp1III motif seen in other HIV subtypes (A, B, D and F) could replace the subtype C specific Sp1III

element suggesting that the C-κB motif of subtype can work with any of these Sp1 variants tested (Figure-3.6 B). Collectively, the reporter vector analyses demonstrated superior magnitude of promoter function when the native C-κB and the Sp1III sites were present at the central location of the promoter. The gel-shift and super-shift analyses ascertained that the variant C-κB element is capable of binding the p50:p65 heterodimer (Figure-3.9). The variant C-κB element demonstrated a two-fold higher affinity for the p50-p65 nuclear complexes and for the recombinant p50 protein (Figure-3.10 and Figure-3.12). The dissociation constant of the C- and H-κB motifs for p50 was found to be 2.8 and 4.6 nM, respectively. The C-κB motif recruited NFAT as efficiently as a canonical NFAT probe derived from the IL-2 promoter. Super-shift analysis showed the binding of both NFAT1 and NFAT2 to the C- as well as the H-κB sequences. Since these two NFAT factors show a contrasting effect on the NF-κB binding (Badran *et al.* 2002), the biological significance of the NFAT binding to the NF-κB sites in C-LTR needs additional analysis.

The most unexpected finding of the present study regarding the association between the core TFBS in C-LTR comes from the assay of a panel of four infectious viral strains. The viral strains are genetically identical except for the differences in the core TFBS representing the possible combinations between the two genetically different NF-κB binding sites (C- and H-κB sites) and the two genetically distinct Sp1III motifs (Subtype B and –C associated sequences). While three of the four viral strains (HHC:C, HHC:B and HHH:B) proliferated successfully with comparable efficiency, the fourth virus (HHH:C) where the H-κB motif and the Sp1III site of subtype C were brought together demonstrated severe replication defect Figure-3.18 and Figure-3.22). The substitution of the Sp1III site in the restricted viral strain from C to B (HHH:C to HHH:B) fully restored the severely compromised replication

efficiency alluding to a possible functional incompatibility between the H-κB site and the subtype C unique Sp1 III motif. All the four viral strains, including HHH:C, generated the RT products at comparable levels ruling out the possibility of variable infectivity as the causative factor underlying the replication defect of HHH:C. A direct comparison of the chromatin landscape of the integrated viruses of the wild type HHC:C and the defective HHH:C viral strains provided a conclusive evidence that the HHH:C viral chromatin is predominantly transcription suppressive in nature. To the best of our knowledge, the present study represents the first report of functional incompatibility between a specific NF-κB binding site and a specific Sp-1 binding motif in the context of HIV-1 LTR.

Our data also offer an insight into the specific evolutionary path subtype C may have followed to acquire a stronger viral promoter that contained an additional and variant NF-κB site (the C-κB motif) as compared to other genetic subtypes of HIV-1. A unique subtype C ancestor containing a typical LTR configuration consisting of two identical H-κB sites and the subtype B like Sp1III motif must introduce two or three different variations to transform the viral LTR into the viral promoter (HH:B to HHC:C) as is seen today. These changes consist of the acquisition of the additional and genetically variant NF-κB binding site (the C-κB motif) and the introduction of genetic variations into the Sp1 III site. Although it is difficult to ascertain in which specific order these two or more evolutionary changes were accentuated, acquisition of the variant Sp1III element appears to be more important for subtype C than gaining an additional NF-κB binding site (see below).

We have applied two different assumptions and one observation while predicting the evolutionary path the subtype C ancestor may have followed to introduce the genetic variations into its promoter. First, a virus is likely to follow the shortest evolutionary path

involving the least number of steps to select variations. Second, a virus is more likely to acquire a specific variant transcription factor binding site readily available from a cellular source than altering an existing site into a variant motif especially when this procedure gives the advantage of using fewer evolutionary steps to reach the destination. Lastly, the combination of the H-κB motif and the Sp1 III site of subtype C is non-viable (our observation). Using these guidelines, we predicted six different potential pathways for the subtype C promoter evolution and weighed the merits of each evolutionary path as depicted (Figure-4.1). The evolutionary path-1 predicts that the Sp1 III site was the first to have been subjected to modification. The outcome cannot be a viable option as the resultant LTR HH:C is non-functional (Figure-3.18). The subtype C ancestor therefore could not have attempted to alter the Sp1III site first. The evolutionary paths 2 and 3 predict that initially the central H-κB site was changed into the C-κB motif (HH:B to HC:B). The HC:B virus should be viable. The HC:B ancestor by following path-2 would alter the Sp1III site to generate HC:C intermediate which would acquire an additional H-κB site to make the contemporary C-LTR (HHC:C). Alternatively, the HC:B ancestor could acquire an additional H-κB site first (HHC:B) and alter the Sp1III site to generate the subtype C LTR. The pathways 2 and 3 should be less favoured as these paths not only require the transformation of the existing H-κB site into the C-κB motif, but also need three evolutionary steps to generate subtype C LTR as compared to pathway-6 below. The pathways 4 and 5 predict that the HH:B ancestor initiates the evolutionary process with the acquisition of an additional H-κB element (HHH:B). Continuing through pathway 4 further would have resulted in a replication incompetent viral intermediate HHH:C which is not a viable option, same as pathway-1 above. Alternatively, the ancestor HHH:B could first alter the central H-κB site into C-κB (HHC:B) and then alter the Sp1III site to generate the subtype C LTR. The pathway 5 should be disfavoured for the

same reasons as pathways 2 and 3, although none of the three pathways can be ruled out theoretically. The pathway-6 predicts that the HH:B ancestor acquired an additional C-κB element possibly from a cellular promoter such as the Fas gene that contains an NF-κB element genetically quite similar to the C-κB site (GGGCGTTCC, (Chan *et al.* 1999)) and positions the element immediately upstream of the Sp1III site to create the HHC:B LTR. The HHC:B intermediate though fully functional undergoes additional variations in the Sp1III site to generate the contemporary subtype C LTR HHC:C. Pathway-6 is the most likely evolutionary trajectory the subtype C HH:B ancestor would have adapted as this path needs only two evolutionary steps and doesn't require transformation of an existing TFBS.

It is, however, intriguing why subtype C would attempt to alter the Sp1III site having already added an additional NF-κB binding motif to its enhancer (HHC:B) and as a consequence acquired a functional and stronger promoter. What was the nature of the selection forces that have driven the genetic variations from changing the HHC:B promoter into the HHC:C LTR? It is reasonable to presume that acquisition of the subtype-specific variations in the Sp1III motif is as critical for subtype C promoter as the acquisition of the C-κB site if not less. If the acquisition of a stronger promoter was the only motive, subtype C could have simply duplicated one of the existing H-κB elements (HH:B to HHH:B) or could have acquired an additional variant NF-κB site (HH:B to HHC:B). Subtype C acquires the variant Sp1III site even at the cost that the altered site is biologically compatible only with the C-κB, but not the H-κB site. No such restrictions are seen with the functioning of the C-κB motif which can work with the heterologous Sp1 III motif of 4 different viral subtypes. It therefore appears that the promoter evolution in subtype C was primarily driven to gain the variant Sp1III motif. To retain the Sp1III motif, a variant NF-κB site was necessary. It appears that the

functional incompatibility between the H-κB motif and Sp1III site of subtype C constitutes the primary selection pressure that pushed subtype C into acquiring a genetically variable C-κB element, instead of duplicating the existing H-κB site to enhance the promoter strength.

We, however, do not know if the proposed pathway for the promoter evolution in subtype C is justified given the lack of ancestor subtype C sequences in the public databases representing the intermediary stages of this evolution. Additionally, it is also not clear what were the evolutionary forces that positively selected the transition of the HHC:B promoter to HHC:C despite the fact that the former is as competent as the latter in terms of replication. On reflection, the presence of the C-κB motif in the core of HHC:B virus may have catalyzed the subsequent changes in the core Sp1III site to generate the evolved virus HHC:C. The in vitro replication assay we used in the present study is clearly inadequate to represent the complex natural biological conditions necessary to recapitulate the differences in the replication fitness properties of the two viral strains HHC:B and HHC:C. In other words, it is reasonable to presume that the HHC:C promoter enjoys specific yet to be defined benefits in the natural infection as compared with the HHC:B LTR, due to the pressure of the subtype-specific Sp1 III motif.

Furthermore, it is necessary to examine and identify the nature of the cellular factors that regulate the κB-Sp1 incompatibility at the molecular level. The epigenetic marks identified in the ChIP analysis showed a predominantly transcription suppressive chromatin in the context of the HHH:C promoter in contrast to the transcription activating chromatin in the HHC:C promoter. It is tempting to propose that the unique H-κB:Sp1 III_C combination (HHH:C), unlike the other three promoters (HHC:C, HHC:B and HHH:B), recruits one or more suppressive cofactors which modify the chromatin and suppress viral transcription.

Using double-stranded DNA probes HHC:C and HHH:C, we are presently isolating the cellular factors from the CEM-CCR5 nuclear extracts . The identity of the associated cellular factors will be determined using Mass spectrometry.

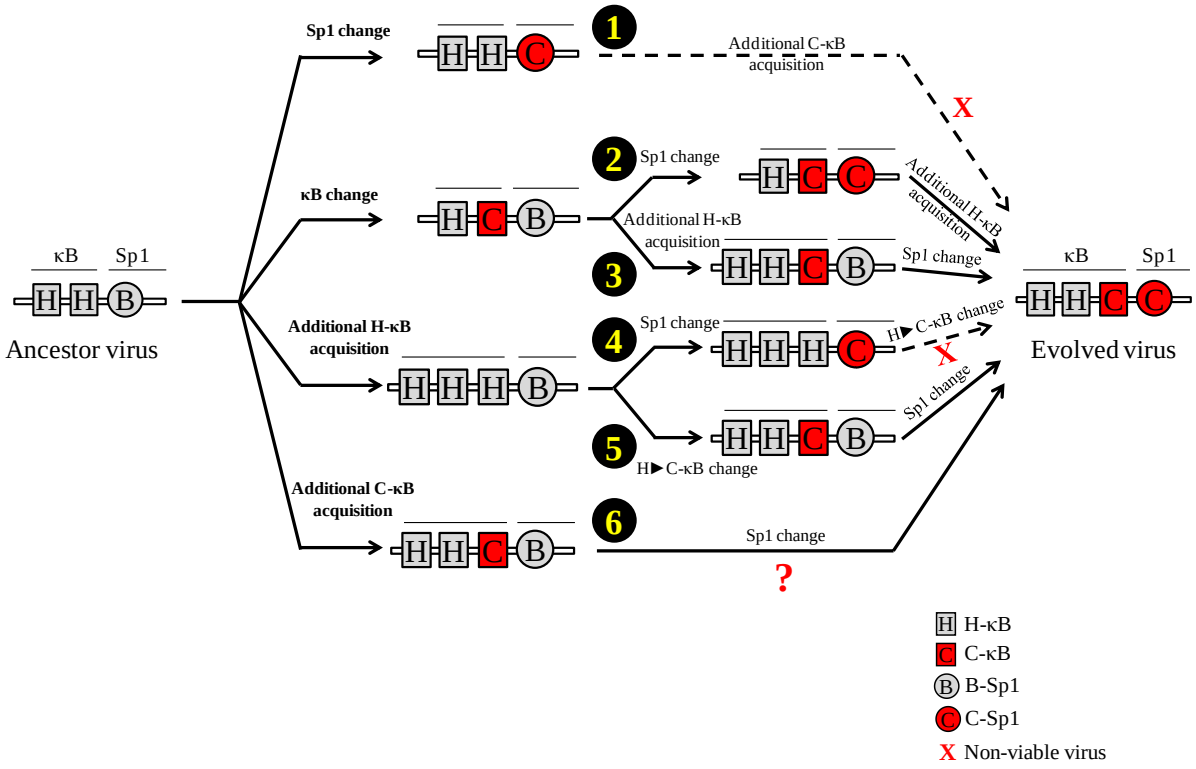


Figure-4.1- A schematic diagram depicting an undefined evolutionary pressure that catalyzes the genetic variations in the Sp1III motif to change this element from 'B' to 'C'.

In summary, our work establishes a new found functional association between the subtype-specific variant NF-κB motif and subtype-specific variant Sp1III site in subtype C viral promoter. Our data suggest further that subtype C uses a mechanism to regulate gene expression in a way different from all other HIV-1 genetic subtypes. How the unique genetic elements in subtype C promoter modulate gene expression and viral latency needs additional investigation.

Bibliography

Abebe, A., D. Demissie, J. Goudsmit, M. Brouwer, C.L. Kuiken, G. Pollakis, H. Schuitemaker, A.L. Fontanet and T.F. Rinke De Wit, 1999. HIV-1 subtype C syncytium- and non-syncytium-inducing phenotypes and coreceptor usage among Ethiopian patients with AIDS. AIDS 13: 1305-1311. Available from PM:10449282

Bachu, M., S. Yalla, M. Asokan, A. Verma, U. Neogi, S. Sharma, R.V. Murali, A.B. Mukthey, R. Bhatt, S. Chatterjee, R.E. Rajan, N. Cheedarla, V.S. Yadavalli, A. Mahadevan, S.K. Shankar, N. Rajagopalan, A. Shet, S. Saravanan, P. Balakrishnan, S. Solomon, M. Vajpayee, K.S. Satish, T.K. Kundu, K.T. Jeang and U. Ranga, 2012. Multiple NF-kappaB sites in HIV-1 subtype C long terminal repeat confer superior magnitude of transcription and thereby the enhanced viral predominance. J.Biol.Chem. 287: 44714-44735. Available from PM:23132857

Badran, B.M., S.M. Wolinsky, A. Burny and K.E. Willard-Gallo, 2002. Identification of three NFAT binding motifs in the 5'-upstream region of the human CD3gamma gene that differentially bind NFATc1, NFATc2, and NF-kappa B p50. J.Biol.Chem. 277: 47136-47148. Available from PM:12374807

Chan, H., D.P. Bartos and L.B. Owen-Schaub, 1999. Activation-dependent transcriptional regulation of the human Fas promoter requires NF-kappaB p50-p65 recruitment. Mol.Cell Biol. 19: 2098-2108. Available from PM:10022897

Coetzer, M., R. Nedellec, J. Salkowitz, S. McLaughlin, Y. Liu, L. Heath, J.I. Mullins and D.E. Mosier, 2008. Evolution of CCR5 use before and during coreceptor switching. J.Virol. 82: 11758-11766. Available from PM:18815295

Hemelaar, J., E. Gouws, P.D. Ghys and S. Osmanov, 2011a. Global trends in molecular epidemiology of HIV-1 during 2000-2007. AIDS 25: 679-689. Available from PM:21297424

Kurosu, T., T. Mukai, S. Komoto, M.S. Ibrahim, Y.G. Li, T. Kobayashi, S. Tsuji and K. Ikuta, 2002. Human immunodeficiency virus type 1 subtype C exhibits higher transactivation activity of Tat than subtypes B and E. Microbiol.Immunol. 46: 787-799. Available from PM:12516777

Montano, M.A., C.P. Nixon, T. Ndung'u, H. Bussmann, V.A. Novitsky, D. Dickman and M. Essex, 2000. Elevated tumor necrosis factor-alpha activation of human immunodeficiency virus type 1 subtype C in Southern Africa is associated with an NF-kappaB enhancer gain-of-function. J.Infect.Dis. 181: 76-81. Available from PM:10608753

Montano, M.A., V.A. Novitsky, J.T. Blackard, N.L. Cho, D.A. Katzenstein and M. Essex, 1997. Divergent transcriptional regulation among expanding human immunodeficiency virus type 1 subtypes. J.Virol. 71: 8657-8665. Available from PM:9343223

Naghavi, M.H., S. Schwartz, A. Sonnerborg and A. Vahlne, 1999. Long terminal repeat promoter/enhancer activity of different subtypes of HIV type 1. AIDS Res.Hum.Retroviruses 15: 1293-1303. Available from PM:10505678

Ping, L.H., J.A. Nelson, I.F. Hoffman, J. Schock, S.L. Lamers, M. Goodman, P. Vernazza, P. Kazembe, M. Maida, D. Zimba, M.M. Goodenow, J.J. Eron, Jr., S.A. Fiscus, M.S. Cohen and R. Swanstrom, 1999. Characterization of V3 sequence heterogeneity in subtype C human immunodeficiency virus type 1 isolates from Malawi: underrepresentation of X4 variants. J.Virol. 73: 6271-6281. Available from PM:10400718

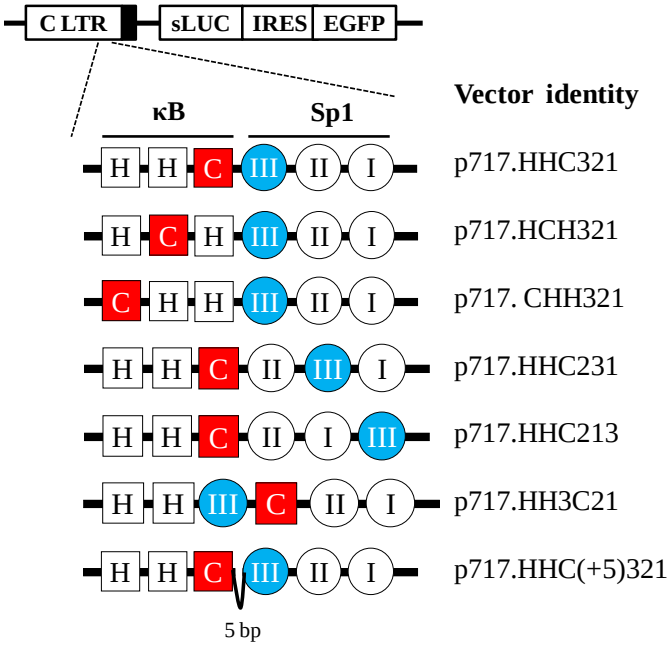
Rodenburg, C.M., Y. Li, S.A. Trask, Y. Chen, J. Decker, D.L. Robertson, M.L. Kalish, G.M. Shaw, S. Allen, B.H. Hahn and F. Gao, 2001. Near full-length clones and reference sequences for subtype C isolates of HIV type 1 from three different continents. AIDS Res.Hum.Retroviruses 17: 161-168. Available from PM:11177395

van, O.T., J. Kamoschinski, R.E. Jeeninga and B. Berkhout, 2004. The human immunodeficiency virus type 1 promoter contains a CATA box instead of a TATA box for optimal transcription and replication. J.Virol. 78: 6883-6890. Available from PM:15194764

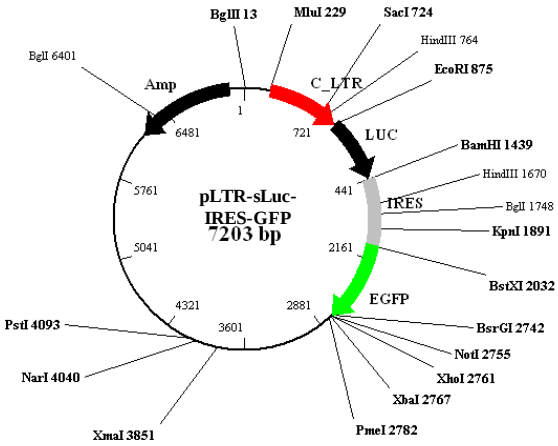
Isogenic variants of pcLIG vector-1

Plasmid No. - p717 and variants
Lab - Prof. Udaykumar Ranga
HIV-AIDS Lab, MBGU
Backbone origin - pcDNA3.1_Luc_IRES_EGFP
Constructed by - Anjali Verma, Pavithra. R
Construction date - August 2012

Key words:
Dual reporter vector, C-LTR, IRES, EGFP, Secretary luciferase, NF-κB, Sp1, mutageneis



Plasmid diagnosis:
Enzyme : HindIII
Fragments (bp) : 6,897
906



Notes:

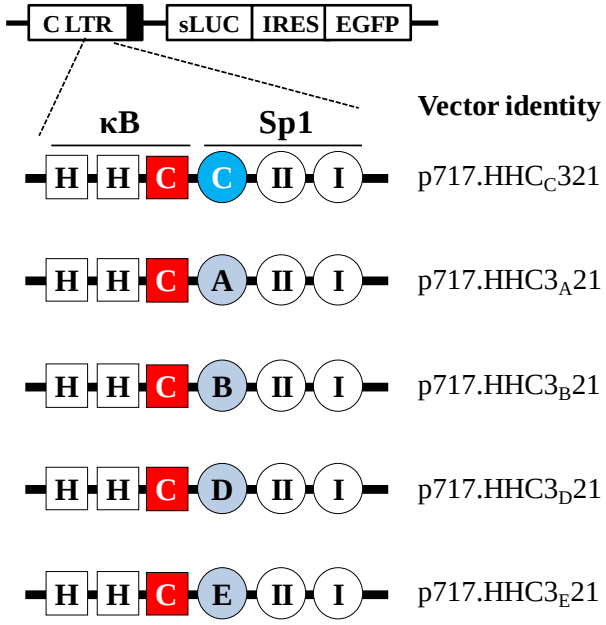
1. The p717 series under C-LTR derived from HIV-1 subtype C molecular clone pIndie (AB023804), was subjected to targeted mutations in κB and/or Sp1 sites.
2. Isogenic LTR variants were generated using overlap PCR. Outer primers N1515 (MluI) and N1359 (EcoRI) were used for the cloning.
3. The original CMV promoter in the vector pcDNA3.1_Luc_IRES_GFP (p450) was replaced with the isogenic LTRs using MluI and EcoRI to generate the LTR reporter vectors.
4. These vectors are dual reporter vectors containing secretary luciferase as well as EGFP.

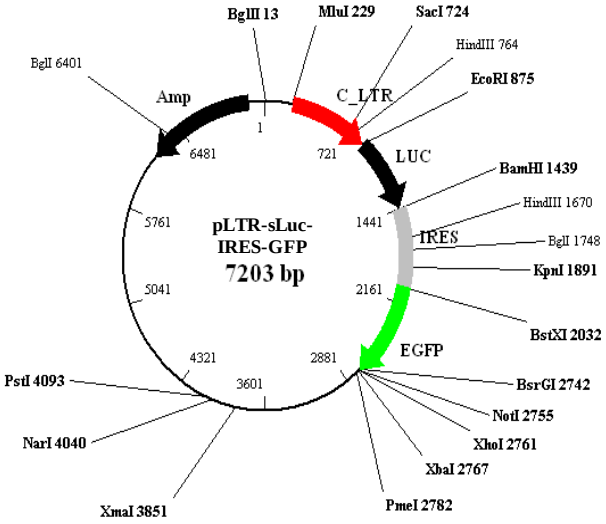
Isogenic variants of pcLIG vector - 2

Plasmid No. - p717 and variants
Lab - Prof. Udaykumar Ranga
HIV-AIDS Lab, MBGU
Backbone origin - pcDNA3.1_Luc_IRES_EGFP
Constructed by - Anjali Verma, Pavithra. R
Construction date - August 2012

Key words:
Dual reporter vector, C-LTR, IRES, EGFP, Secretary luciferase, Sp1III, mutageneis

Plasmid diagnosis:
Enzyme : HindIII
Fragments (bp) : 6,897
906





Consensus sequence of Sp1III sites from various subtypes, which were replaced at Sp1III sites in HIV-1 subtype C LTR

Sub C	G A G G T G T G G T
Sub A	. . . . C . . . . .
Sub B	. . . . C C
Sub D	. . . . C . . . . A
Sub E	 C

Notes:

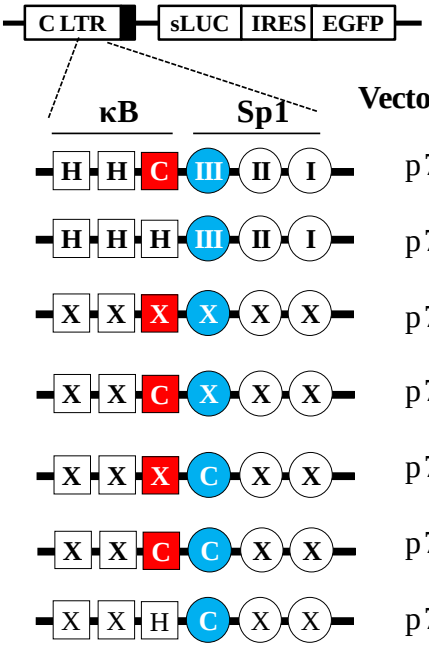
1. This p717 series under C-LTR derived from HIV-1 subtype C molecular clone pIndie (AB023804), was subjected to targeted mutations in Sp1 III site. Subtype C-unique NF-κB site is placed in the association of different Sp1III sites, differentially conserved in some of the major HIV-1 ubtypes.
2. Isogenic LTR variants were generated using overlap PCR. Outer primers N1515 (MluI) and N1359 (EcoRI) were used for the cloning
3. The original CMV promoter in the vector pcDNA3.1_Luc_IRES_GFP (p450) was replaced with the isogenic LTRs using MluI and EcoRI to generate the LTR reporter vectors.
4. These vectors are dual reporter vectors containing secretary luciferase as well as EGFP.

Isogenic variants of pcLIG vector -3

Plasmid No. - p718 and variants
Lab - Prof. Udaykumar Ranga
HIV-AIDS Lab, MBGU
Backbone origin -- pcDNA3.1_Luc_IRES_EGFP
Constructed by - Anjali Verma, Pavithra. R
Construction date - August 2012

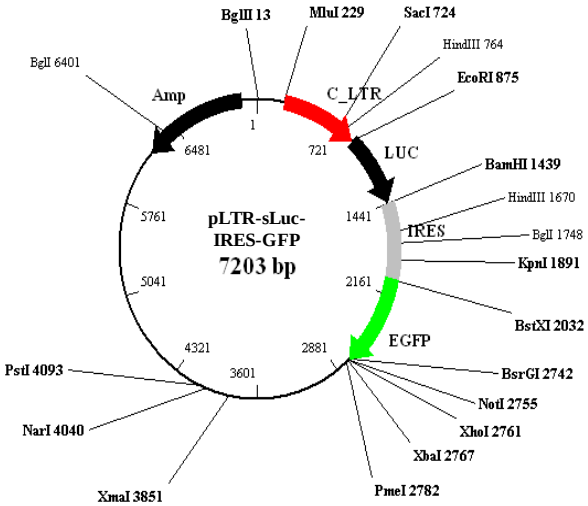
Key words:
Dual reporter vector, C-LTR, IRES, EGFP, Secretary luciferase, NF-κB, Sp1, mutageneis

Plasmid diagnosis:
Enzyme : HindIII
Fragments (bp) : 6,897
906



Mutations to create inactivation of H-κB and C-κB sites

H-κB	G G G A C T T T C C
H-κBmut	T C T - - - - - - -
C-KB	G G G G C G T T C C
C-Kbmut	- T C T - - - - - -



Notes:

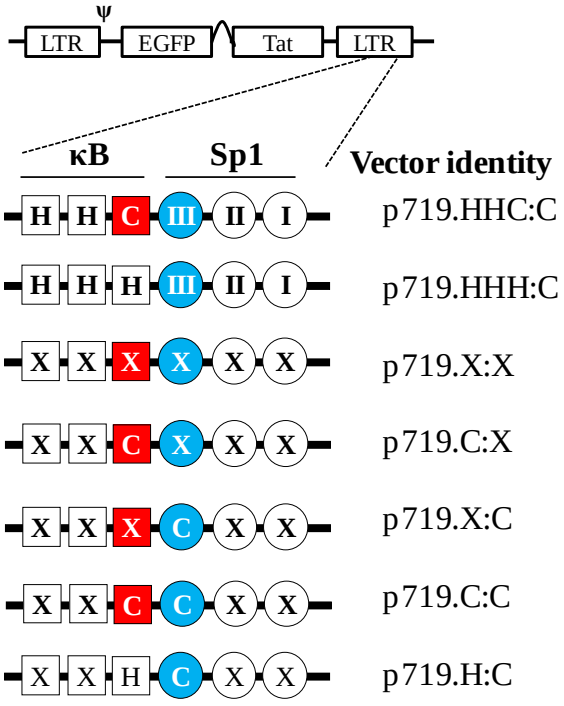
1. This series of vectors differs from that of p717 in lacking of the κB and Sp1 sites other than the one proximal to each other. The LTRs contain only one each κB and Sp1III sites.
2. The p718 series under C-LTR derived from HIV-1 subtype C molecular clone pIndie (AB023804), was subjected to targeted mutations in κB and/or Sp1 sites.
3. Isogenic LTR variants were generated using overlap PCR. Outer primers N1515 (MluI) and N1359 (EcoRI) were used for the cloning
4. The original CMV promoter in the vector pcDNA3.1_Luc_IRES_GFP (p450) was replaced with the isogenic LTRs using MluI and EcoRI to generate the LTR reporter vectors.
5. These vectors are dual reporter vectors containing secretary luciferase as well as EGFP.

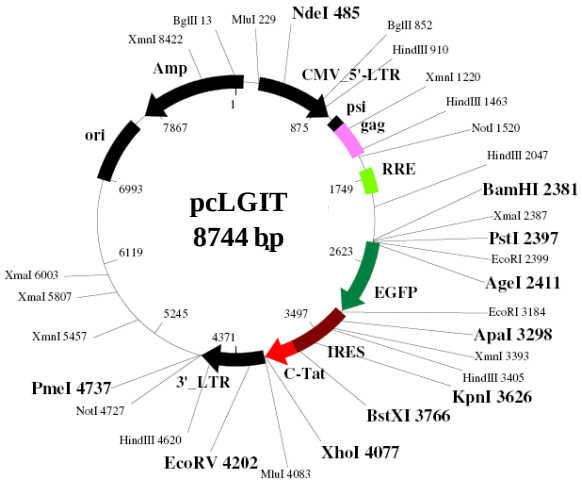
Isogenic variants of pcLGIT vector

Plasmid No. - p719 and variants
Lab - Prof. Udaykumar Ranga
HIV-AIDS Lab, MBGU
Backbone origin - pLGITc
Constructed by - Anjali Verma, Pavithra. R
Construction date - August 2012

Key words:
Mini-reporter viral vectors, C-LTR, IRES, EGFP, C-Tat, NF-κB, Sp1III, mutagenesis

Plasmid diagnosis:
Enzyme : BglII
Fragments (bp) : 4,191, 3,375, 839, 507





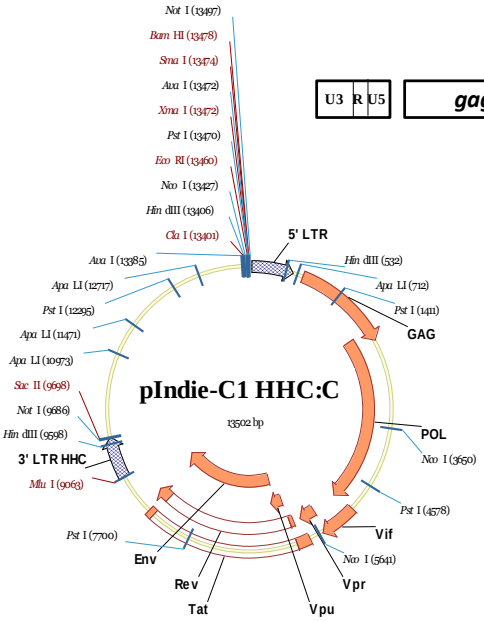
Notes:

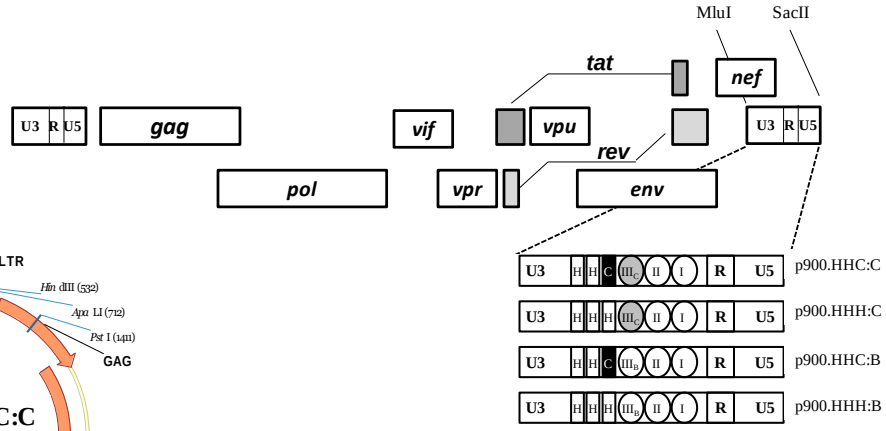
1. This series of viral vectors expresses a reporter mini-HIV with EGFP and C-tat under the control of HIV-1 LTR on its promoter variants.
2. B LTR of p701 was replaced with C-LTR derived from pIndie HIV-1 subtype C molecular clone (AB023804). Isogenic LTR variants were generated using overlap PCR. Outer primers N1145 (XhoI) and N1146 (PmeI) were used for the cloning.

Isogenic variants of pIndie-C1 vector

Plasmid No. - p900 and variants
Lab - Prof. Udaykumar Ranga
HIV-AIDS Lab, MBGU
Backbone origin - pIndie-FHHC
Constructed by - Anjali Verma
Construction date - August 2014

Key words:
C-LTR, NF-κB, Sp1, mutageneis

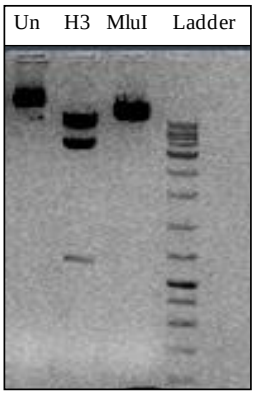




Notes:

1. The p900 series under C-LTR derived from HIV-1 subtype C molecular clone pIndie (AB023804), was subjected to targeted mutations in κB and/or Sp1 sites.
2. Sequence of the H-/C-κB and Sp1III$_B$ / Sp1III$_C$ has been mentioned .
3. Isogenic LTR variants were generated using overlap PCR. Outer primers N1515 (MluI) and N1516 (SacII) were used for the cloning.
4. The 3' BL-42 FHHC LTR in the vector pIndie-FHHC (p450.4) was replaced with the isogenic LTRs using MluI and SacII to generate the isogenic LTR vectors.
5. These vectors can be used to make fll-length replication competent virus.
6. These vector are helpful in the study of replication kinetics.

Plasmid diagnosis: Hind III
9,063 bp
3,806 bp
638 bp



Manuscripts

1. **Verma A**, Rajgopalan P, Lotke R, Varghese R, Ranga U. A functional association between subype-specific NF-κB and Sp1 motifs of HIV-1 subtype C viral promoter. Manuscript in preparation.

2. Mishra M, Verghese R, **Verma A**, Das S, Aguiar R S, Tanuri A, Mahadevan A, Shankar S K, Satishchandra P, Ranga U. Genetic diversity and proviral DNA load in different neural compartments of HIV-1 subtype C infection. Journal of Neurovirology. *In press*.

3. Bachu M, Yalla S, Asokan M, **Verma A**, Neogi U, Sharma S *et al*.: Multiple NF-kappaB sites in HIV-1 subtype C long terminal repeat confer superior magnitude of transcription and thereby the enhanced viral predominance**. *J Biol Chem*** 2012, 287: 44714-44735.