EUROPEAN JOURNAL OF BIOMEDICAL AND PHARMACEUTICAL SCIENCES
http://www.ejbps.com

ISSN 2349-8870
Volume: 7
Issue: 7
306-318
Year: 2020

ANALYSIS OF THE BLUEPRINT OF HIV

***Kushal Nandi, Pritam Bakshi, Sandip Sarkar, Subhanjana Guha, Shreya Sarkar, Maitri Sahoo, Souradip Mandal, Tanistha Mumtaz Chatterjee, Barnamoni Saha, Anirbita Ghosh, Dr. Dhrubo Jyoti Sen, Shaktijit Das, Dr. Partha Ranjan Bhattacharya and Dr. Beduin Mahanti**

Department of Pharmaceutical Chemistry & Pharmacology, School of Pharmacy, Techno India University, Salt Lake City, Sector-V, EM-4, Kolkata-700091, West Bengal, India.

***Corresponding Author: Kushal Nandi**
Department of Pharmaceutical Chemistry & Pharmacology, School of Pharmacy, Techno India University, Salt Lake City, Sector-V, EM-4, Kolkata-700091, West Bengal, India.

Article Received on 07/05/2020 **Article Revised on 27/05/2020** **Article Accepted on 17/06/2020**

ABSTRACT
The Human immunodeficiency viruses (HIV) are two species of Lentivirus (a subgroup of retrovirus) that infect humans. Over time, they cause acquired immunodeficiency syndrome (AIDS), a condition in which progressive failure of the immune system allows life-threatening opportunistic infections and cancers to thrive. Without treatment, average survival time after infection with HIV is estimated to be 9 to 11 years, depending on the HIV subtype. In most cases, HIV is a sexually transmitted infection and occurs by contact with or transfer of blood, pre-ejaculate, semen, and vaginal fluids. Research has shown (for both same-sex and opposite-sex couples) that HIV is untransmittable through condom less sexual intercourse if the HIV-positive partner has a consistently undetectable viral load. Non-sexual transmission can occur from an infected mother to her infant during pregnancy, during childbirth by exposure to her blood or vaginal fluid, and through breast milk. Within these bodily fluids, HIV is present as both free virus particles and virus within infected immune cells. HIV infects vital cells in the human immune system, such as helper T cells (specifically CD4+ T cells), macrophages, and dendritic cells. HIV infection leads to low levels of CD4+ T cells through a number of mechanisms, including pyroptosis of abortively infected T cells, apoptosis of uninfected bystander cells, direct viral killing of infected cells, and killing of infected CD4+ T cells by CD8+ cytotoxic lymphocytes that recognize infected cells. When CD4+ T cell numbers decline below a critical level, cell-mediated immunity is lost, and the body becomes progressively more susceptible to opportunistic infections, leading to the development of AIDS.

KEYWORDS: Lentivirus, Retrovirus, CD4+, CD8+, Macrophage, Lymphocytes, RNA genome.

INTRODUCTION

HIV is a member of the genus Lentivirus, part of the family Retroviridae. Lentiviruses have many morphologies and biological properties in common. Many species are infected by lentiviruses, which are characteristically responsible for long-duration illnesses with a long incubation period. Lentiviruses are transmitted as single-stranded, positive-sense, enveloped RNA viruses. Upon entry into the target cell, the viral RNA genome is converted (reverse transcribed) into double-stranded DNA by a virally encoded enzyme, reverse transcriptase, that is transported along with the viral genome in the virus particle. The resulting viral DNA is then imported into the cell nucleus and integrated into the cellular DNA by a virally encoded enzyme, integrase, and host co-factors. Once integrated, the virus may become latent, allowing the virus and its host cell to avoid detection by the immune system, for an indeterminate amount of time. The HIV virus can remain dormant in the human body for up to ten years after primary infection; during this period the virus does not cause symptoms. Alternatively, the integrated viral DNA may be transcribed, producing new RNA genomes and viral proteins, using host cell resources, that are packaged and released from the cell as new virus particles that will begin the replication cycle anew.[1]

Classification

Table-1: Class of HIV-1 & 2.

Comparison of HIV species				
Species	**Virulence**	**Infectivity**	**Prevalence**	**Inferred origin**
HIV-1	High	High	Global	Common chimpanzee
HIV-2	Lower	Low	West Africa	Sooty mangabey

Two types of HIV have been characterized: **HIV-1 and HIV-2**. HIV-1 is the virus that was initially discovered and termed both lymphadenopathies associated virus (LAV) and human T-lymphotropic virus 3 (HTLV-III). HIV-1 is more virulent and more infective than HIV-2, and is the cause of the majority of HIV infections globally. The lower infectivity of HIV-2, compared to HIV-1, implies that fewer of those exposed to HIV-2 will be infected per exposure. Due to its relatively poor capacity for transmission, HIV-2 is largely confined to West Africa.

The human immunodeficiency viruses are approximately 100 nm in diameter. It has a lipid envelope, in which are embedded the trimeric transmembrane glycoprotein gp41 to which the surface glycoprotein gp120 is attached. These two viral proteins are responsible for attachment to the host cell and are encoded by the *env* gene of the viral RNA genome. Beneath the envelope, is the matrix protein p17, the core proteins p24 and p6 and the nucleocapsid protein p7 (bound to the RNA), all encoded by the viral *gag* gene. Within the viral core, lies 2 copies of the ~10 kilobase (kb) positive- sense, viral RNA genome (i.e. it has a diploid RNA genome), together with the protease, integrase and reverse transcriptase enzymes. These three enzymes are encoded by the viral *pol* gene. There are several other proteins coded for by both HIV-1 and HIV-2, with various regulatory or immuno- modulatory functions, including *vif* (viral infectivity protein), *vpr* (viral protein R), *tat* (transactivator of transcription), *rev* (regulator of viral protein expression) and *nef* (negative regulatory factor). An additional protein found in HIV-1 but not HIV-2 is *vpu* (viral protein U). Similarly, *vpx* (viral protein X) is found in HIV-2 and not HIV-1. The molecular structure of the viral spike has now been determined by X-ray crystallography and cryo-electron microscopy.

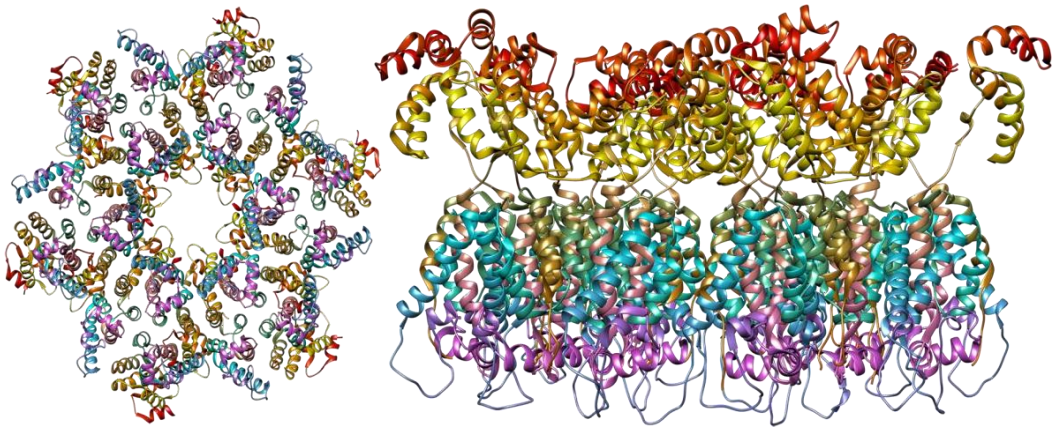

Figure-1: Structure of the immature HIV-1 capsid in intact virus particles.

Human Immunodeficiency Virus genome mainly consists of nine genes, three of them are essential for reproduction and infection and other six are regulatory genes.

HIV has several major genes coding for structural proteins that are found in all retroviruses, and several non-structural (accessory) genes that are unique to HIV. The **gag** gene provides the basic physical infrastructure of the virus, and **pol** provides the basic mechanism by which retroviruses reproduce, while the others help HIV to enter the host cell and enhance its reproduction. Though they may be altered by mutation, all of these genes except **tev** exist in all known variants of HIV.[2]

Gag (group-specific-antigen): codes for the **gag** polyprotein, which is processed during maturation to MA (matrix protein, p17); CA (capsid protein, p24), SP1 (spacer peptide 1, p2); NC (nucleocapsid protein, p7); SP2 (spacer peptide 2) and p6.

Pol: codes for viral enzymes reverse transcriptase, integrase and HIV protease.

Env: codes for gp160, the precursor to gp120 and gp41, proteins embedded in the viral envelop which enable the virus to attach to and fuse with target cells.
Trans-activators: **tat, rev, vpr**
Other regulators: **Vif, nef, vpu**

Tev: This gene is only present in a few HIV-1 isolates. It is a fusion of parts of the tat, env and rev genes, and codes for a protein with some of the properties of tat, but little or none of the properties of rev.

It is a cis-acting RNA element identified in the genomes of the retroviruses Human Immunodeficiency Virus (HIV) and Simian immunodeficiency virus (SIV). It is involved in regulating encapsidation of the retroviral RNA, an essential step in replication.

RNA secondary structure: The HIV genome consists of two identical single-stranded RNA molecules that are enclosed within the core of the virus particle.

The genome of the HIV provirus also known as proviral DNA, is generated by the reverse transcription of the viral RNA genome into DNA, degradation of the RNA

and integration of the double-stranded HIV DNA into the human genome.

The DNA genome is flanked at both ends by LTR (long terminal repeat) sequences. The 5′ LTR region codes for the promoter for transcription of the viral genes. In the direction 5′ to 3′ the reading frame of the gag gene follows, encoding the proteins of the outer core membrane (MA, p17), the capsid protein (CA, p24), the nucleocapsid (NC, p7) and a smaller, nucleic acid-stabilising protein. The gag reading frame is followed by the pol reading frame coding for the enzyme protease (PR, p12), reverse transcriptase (RT, p51) and RNase H (p15) or RT plus RNase H (together p66) and integrase (IN, p32). Adjacent to the pol gene, the env reading frame follows from which the two envelope glycoproteins gp120 (surface protein, SU) and gp41 (transmembrane protein, TM) are derived.

In addition to the structural proteins, the HIV genome codes for several regulatory proteins: Tat (transactivator protein) and Rev (RNA splicing-regulator) are necessary for the initiation of HIV replication, while the other regulatory proteins Nef (negative regulating factor), Vif (viral infectivity factor), Vpr (virus protein r) and Vpu (virus protein unique) have an impact on viral replication, virus budding and pathogenesis HIV-2 codes for Vpx (virus protein x) instead of Vpu, which is partially responsible for the reduced pathogenicity of HIV-2.

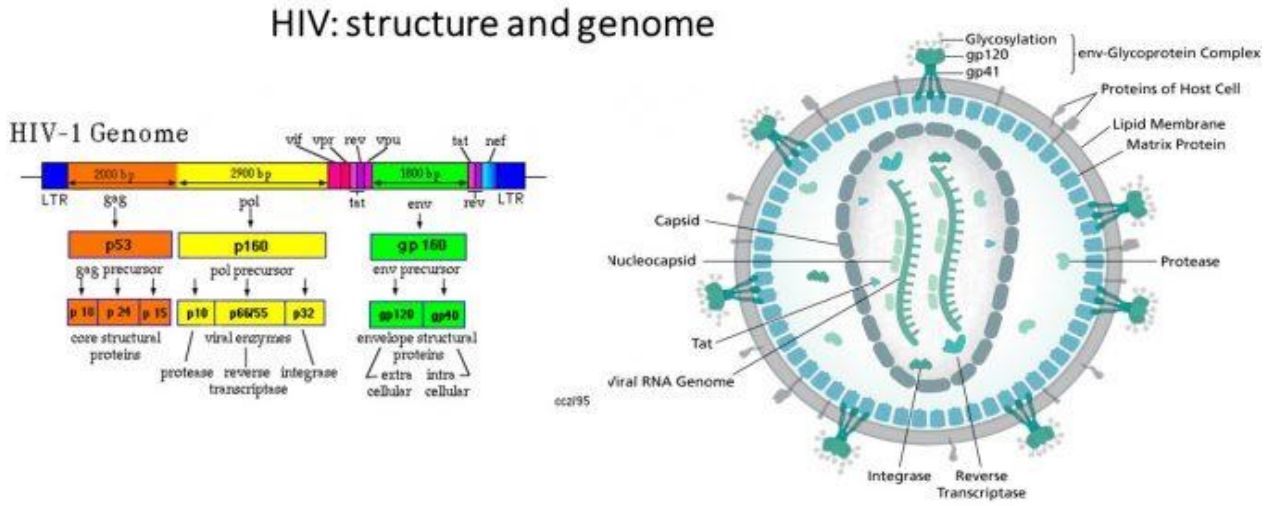

Figure-2: HIV-1 genome.
LTR = long terminal repeat; gag = group-specific antigen; pol = polymerase; env = envelope.

The biased neucleotide composition of HIV Genome: The RNA genome of the lentivirus family of retroviruses, including human immunodeficiency virus (HIV), contains an above average percentage of adenine (A) nucleotides, while being extremely poor in cytosine (C). The early genes such as tat, rev, nef, and the untranslated 5' leader RNA are less A- rich than the pol gene.[3]

The base composition of the HIV genome is surprisingly stable over time, varying less than 1% per base per isolate whether originating from early or later years of the epidemic and regardless of HIV-1 group or subtype.

The genome and proteins of HIV (human immunodeficiency virus) have been the subject of extensive research since the discovery of the virus in 1983. "In the search for the causative agent, it was initially believed that the virus was a form of the Human T-cell leukaemia virus (HTLV), which was known at the time to affect the human immune system and cause certain leukaemia's. However, researchers at the Pasteur Institute in Paris isolated a previously unknown and genetically distinct retrovirus in patients with AIDS which was later named HIV." Each virion comprises a viral envelope and associated matrix enclosing a capsid, which itself encloses two copies of the single-stranded RNA genome and several enzymes. The discovery of the virus itself occurred two years following the report of the first major cases of AIDS-associated illnesses.

Genome Organism: The HIV genome consists of two identical single-stranded RNA molecules that are enclosed within the core of the virus particle. The genome of the HIV provirus, also known as pro-viral DNA, is generated by the reverse transcription of the viral RNA genome into DNA, degradation of the RNA

and integration of the double-stranded HIV DNA into the human genome. The DNA genome is flanked at both ends by LTR (long terminal repeat) sequences. The 5′ LTR region codes for the promotor for transcription of the viral genes. In the direction 5′ to 3′ the reading frame of the *gag* gene follows, encoding the proteins of the outer core membrane (MA, p17), the capsid protein (CA, p24), the nucleocapsid (NC, p7) and a smaller, nucleic acid-stabilising protein. The *gag* reading frame is followed by the *pol* reading frame coding for the enzyme's protease (PR, p12), reverse transcriptase (RT, p51) and RNase H (p15) or RT plus RNase H (together p66) and integrase (IN, p32). Adjacent to the *pol* gene, the *env* reading frame follows from which the two envelope glycoproteins gp120 (surface protein, SU) and gp41 (transmembrane protein, TM) are derived. In addition to the structural proteins, the HIV genome codes for several regulatory proteins: Tat (trans activator protein) and Rev (RNA splicing-regulator) are necessary for the initiation of HIV replication, while the other regulatory proteins Nef (negative regulating factor), Vif (viral infectivity factor), Vpr (virus protein r) and Vpu (virus protein unique) have an impact on viral replication, virus budding and pathogenesis HIV-2 codes for Vpx (virus protein x) instead of Vpu, which is partially responsible for the reduced pathogenicity of HIV-2. The genome structure of the immunodeficiency viruses of chimpanzees (SIVcpz) and gorillas (SIVgor) is identical to that of HIV-1.

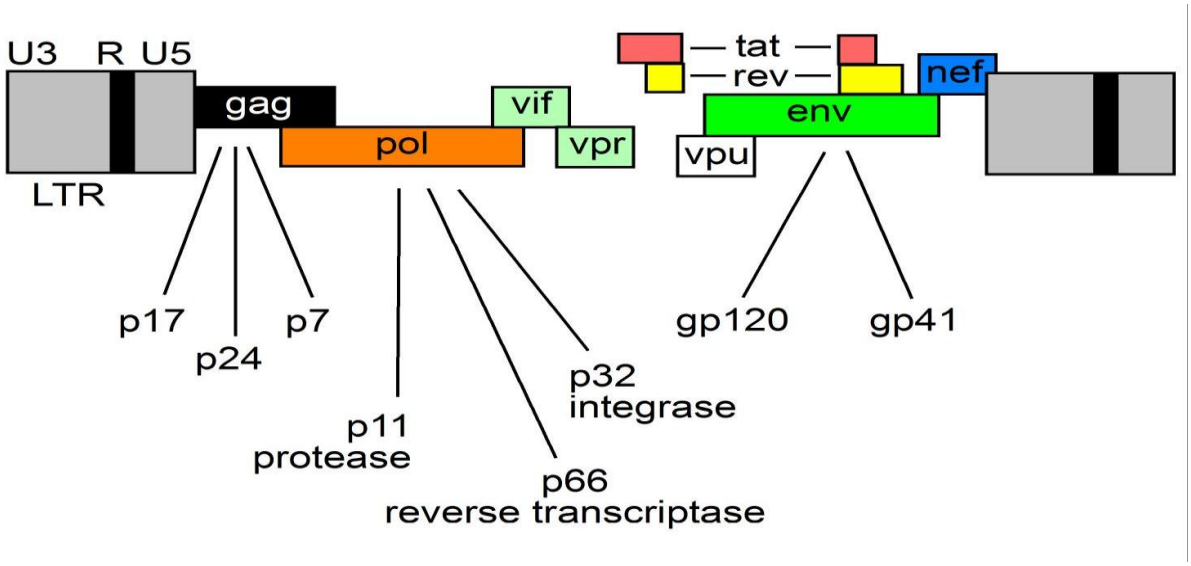

Figure-3: Structure and organization of the HIV-1 genome. Shown are the reading frames of the genes coding for structural and regulatory proteins, LTR = long terminal repeat; *gag* = group-specific antigen; *pol* = polymerase; *env* = envelope. In the case of the regulator genes, the proteins of *tat* and *rev* are composed of two gene regions. In HIV-2, *vpx* corresponds to the *vpu* gene. The 5´ and 3´ LTR nucleic acid sequences are not translated into protein. The genome consists of 9,200-9,600 nucleotides in the case of HIV-1 and approximately 9,800 nucleotides in the case of HIV-2.

Table-2: HIV genome protein encoding.

Proteins encoded by the HIV genome			
Class	**Gene name**	**Primary protein products**	**Processed protein products**
Viral structural proteins	*gag*	Gag polyprotein	MA, CA, SP1, NC, SP2, P6
	pol	Pol polyprotein	RT, RNase H, IN, PR
	env	gp160	gp120, gp41
Essential regulatory elements	*tat*	Tat	
	rev	Rev	
Accessory regulatory proteins	*nef*	Nef	
	vpr	Vpr	
	vif	Vif	
	vpu	Vpu	

Accessory regulatory proteins: vpr (lentivirus protein R): Vpr is a virion-associated, nucleocytoplasmic shuttling regulatory protein. It is believed to play an important role in replication of the virus, specifically, nuclear import of the pre-integration complex. Vpr also appears to cause its host cells to arrest their cell cycle in the G2 phase. This arrest activates the host DNA repair machinery which may enable integration of the viral DNA. HIV-2 and SIV encode an additional Vpr related protein called Vpx which functions in association with Vpr.[4]

Vif: Vif is a highly conserved, 23 kDa phosphoprotein important for the infectivity of HIV-1 virions depending on the cell type. HIV-1 has been found to require Vif to synthesize infectious viruses in lymphocytes, macrophages, and certain human cell lines. It does not appear to require Vif for the same process in HeLa cells or COS cells, among others.

Nef: Nef, negative factor, is a N-terminal myristoylated membrane-associated phosphoprotein. It is involved in multiple functions during the replication cycle of the virus. It is believed to play an important role in cell apoptosis and increase virus infectivity.

Vpu: (Virus protein U) - Vpu is specific to HIV-1. It is a class I oligomeric integral membrane phosphoprotein with numerous biological functions. Vpu is involved in CD4 degradation involving the ubiquitin proteasome pathway as well as in the successful release of virions from infected cells.

Tev: This gene is only present in a few HIV-1 isolates. It is a fusion of parts of the tat, env, and rev genes, and codes for a protein with some of the properties of tat, but little or none of the properties of rev.

RNA Secondary Structure Of HIV

Introduction: An RNA secondary structure determined by SHAPE analysis has shown to contain three stem loops and is located between the HIV protease and reverse transcriptase genes. This cis regulatory RNA has been shown to be conserved throughout the HIV family and is thought to influence the viral life cycle.

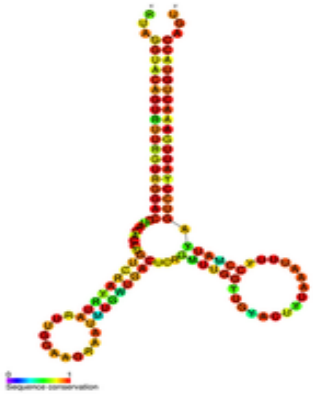

Figure-4: RNA Secondary structure of HIV.

A detailed secondary structure model has been proposed for the entire HIV-1 genome, based on the SHAPE technique (selective 2′-hydroxyl acylation analysed by primer extension) and a thermodynamic structure prediction algorithm. The SHAPE technique probes the accessibility and flexibility of the 2′ hydroxyl groups in the RNA chain, and thereby detects the engagement of nucleotides in secondary and, to some extent, tertiary structures. This first complete structural model confirmed many previously suggested structures, but also featured a large number of previously uncharacterized secondary structures, which remain to be experimentally confirmed.

RNA that encode protein junctions. ribosome processivity is inhibited by cycloheximide and sites preferentially occupied by the ribosome are detected as **stops to primer extension** in an in vitro translation.

Ribosome pause sites are statistically over represented at the MA-CA and CA-NC junctions in Gag and at the sequences encoding the cyclophilin loop in CA.

Conversely, ribosome pause sites are under represented in flanking, but unstructured, regions of the HIV RNA ($p = 0.018$). These experiments thus strongly support the model that mRNA structure over a region spanning 60–100 nucleotides specifically modulate ribosome processivity at protein domain junctions.

RNA secondary structure model for HIV-1: Comprehensive SHAPE reactivity information can also be used to determine a nucleotide-resolution secondary structure model for the entire NL4-3 HIV-1 genome.

SHAPE reactivities are converted to free energy change terms and used to constrain a thermodynamic folding algorithm 22,23. The final result is a thermodynamically favoured structural model highly reflective of the experimental SHAPE data, at single nucleotide resolution. For example, most nucleotides assigned to single-stranded regions are reactive towards SHAPE (red, orange, and green nucleotides) whereas, base-paired nucleotides are predominantly unreactive (black nucleotides and inset).

The HIV-1 genome is less structured than ribosomal RNA but, similarly, contains multiple independent RNA folding domains that extend from the overall genomic backbone. These domains include both small stem-loops plus roughly 21 large and complexly folded structures. Although many genome regions are highly structured, only 7 helices spina complete turn of an 11 bp RNA duplex. The largest paired region, devoid of bulges, is

the structured RNA element that bridges the coding junction between the RT and RNase H-folding domains.

This helix is 19 bp long, contains a non-canonical G-A base pair and is thus shorter than the 30-base pair length competent to induce the interferon response 24. The HIV-1 genome structural model provides a robust starting point for identifying previously unrecognized functional elements and long-range RNA interactions. SHAPE reactivities describe a well-formed stem 3′ to the signal peptide (SP) coding region in the Env protein. This stem (the SP-stem) is evolutionarily conserved reinforcing an important biological role. The signal recognition particle (SRP) binds the nascent Env SP and translocates the cytoplasmic ribosome elongation complex to the rough endoplasmic reticulum where translation of gp120 and gp41 continue 25. RNA-induced translational pausing occurs as the ribosome unwinds highly structured RNA, typically located 6-7 nucleotides downstream of the A-site18.

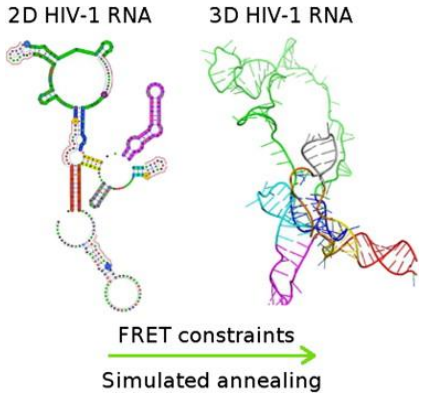

Figure-5: Comparison between 2D HIV-1 RNA & 3D HIV-1 RNA.

In this conformation when the final tRNAAla from SP and the first tRNAThr of gp120 are in the P-and A-sites (boxed nucleotides). Ribosomal attenuation or pausing at the SP-stem provides additional time for SRP recruitment and subsequent translocation of the elongation complex to the endoplasmic reticulum.[5]

The SHAPE-constrained secondary structure is also informative for previously identified regulatory motifs.

In HIV-1, pro and pol gene products are translated when the ribosome undergoes a –1 register shift from the gag to pol reading frames.

Frameshifting occurs at a slippery sequence **(U UUU UUA)** and is enhanced by a downstream RNA structure. These elements are typically drawn as a single stranded slippery sequence and a 12 bp stem-loop26. Direct analysis of intact genomic RNA shows that the gag-pol frameshift signal is one of a 3-helix structure. The slippery sequence pairs to form one of the three helices (P2). These two helices are stabilized by an anchoring helix (P1) that creates this discrete structural element. This three-helix junction structure is conserved among HIV-1 group M sequences. Most RNA viruses require a complex pseudo knotted structure to induce ribosomal Frameshifting. The three-helix junction may function, in part, to slow translation before the ribosome encounters P3, facilitating the prerequisite pause necessary for frameshifting. The three-helix junction model may also explain why changing the slippery site to sequences which allow alternate tRNA pairing and enhance frameshifting in other RNA viruses eliminates frameshifting in HIV-128.

In the SHAPE-directed model, changes to the slippery sequence compromise base pairing in the conserved P2 helix. Unstructured motifs and insulator.

V3 Loop: The **third variable loop** or **V3 loop** is a part or region of the Human Immunodeficiency Virus. The **V3 loop** of the virion's envelope glycoprotein, gp120, allows it to infect human immune cells by binding to a cytokine receptor on the target human immune cell, such as a CCR5 cell or CXCR4 cell, depending on the strain of HIV. The envelope glycoprotein (Env)gp 120/41 is essential for HIV-1 entry into cells. Env serves as a molecular target of a medicine treating individuals with HIV-1 infection, and a source of immunogen to develop AIDS vaccine.

Role of the V3 Loop in Human Immunodeficiency Virus Type 1 Envelope Glycoprotein Function: Mutations within the principal neutralizing determinant (the V3 loop) of the HIV-1 surface envelope glycoprotein gp120 block or greatly reduce the ability of the HIV-1 envelope glycoprotein to induce cell fusion in CD4+ HeLa T4 cells while keeping its CD4 binding ability. However, when either cysteine or both cysteines forming the V3 disulphide (-S-S-) bridge were mutated, the resultant glycoprotein could not mediate cell fusion, undergo proteolytic processing, or bind CD4. To investigate the role that the V3 loop plays in gp160 processing and CD4 binding, we deleted the entire V3 loop region of the HIV-1 env gene. The resultant glycoprotein could not mediate cell fusion in the HeLa T4 cell line and no proteolytic processing of gp160 or CD4 binding could be detected. To test whether any domain of the V3 loop is involved in attaining the proper envelope glycoprotein conformation required for

proteolytic processing and CD4 binding, we introduced a series of deletions into the coding region of the V3 loop. Most of the residues within the V3 loop could be removed while retaining gp160 processing and CD4 binding. The results indicate that the cysteines that form the V3 loop or the disulphide bond itself are important for proper envelope glycoprotein folding and processing.[6]

Role of HIV-1 Envelope V3 Loop Cleavage in Cell Tropism: In order for an HIV-1 virion to infect a host cell, the gp120 subunit of envelope, the only viral protein exposed on the surface of the virion, first attaches to the cellular receptor CD4 and subsequently to a chemokine coreceptor, either CCR5 or CXCR4. Tropism (movement of the virion) for these coreceptors is principally determined by the V3 loop, which is a 35-residue region bound by a disulphide bond and exposed on the surface of gp120. While V3 is known to be critical for coreceptor binding, V3-coreceptor interactions and the structural basis for CCR5 or CXCR4 specificity remain elusive. We hypothesized that subdomains within V3 determine coreceptor tropism. Using the dual-tropic virus HIV-1R3A, which uses both CCR5 and CXCR4, we made an extensive panel of small V3 deletions and evaluated these mutants for their effects on tropism and sensitivity to coreceptor antagonists. Small deletions on either side of the V3 base abrogated R5 tropism. Only a deletion of residues 9 to 12 in the V3 stem ablated X4 tropism. Remarkably, this R5-tropic mutant showed complete resistance to a panel of CCR5 antagonists, including vicriviroc and aplaviroc. V3 differentially affect R5 and X4 tropism and modulate sensitivity to coreceptor antagonists. Small molecule CCR5 inhibitors are being used clinically to treat HIV-1-infected patients, so in order to study mechanisms of resistance to this new drug class.

CCR5 Deletion: The CCR5 protein belongs to the beta chemokine receptors family of integral membrane proteins. It is a G protein–coupled receptor which functions as a chemokine receptor in the CC chemokine group.

CCR5's cognate ligands include CCL3, CCL4 (also known as MIP 1α and 1β, respectively), and CCL3L1. CCR5 furthermore interacts with CCL5 (a chemotactic cytokine protein also known as RANTES).

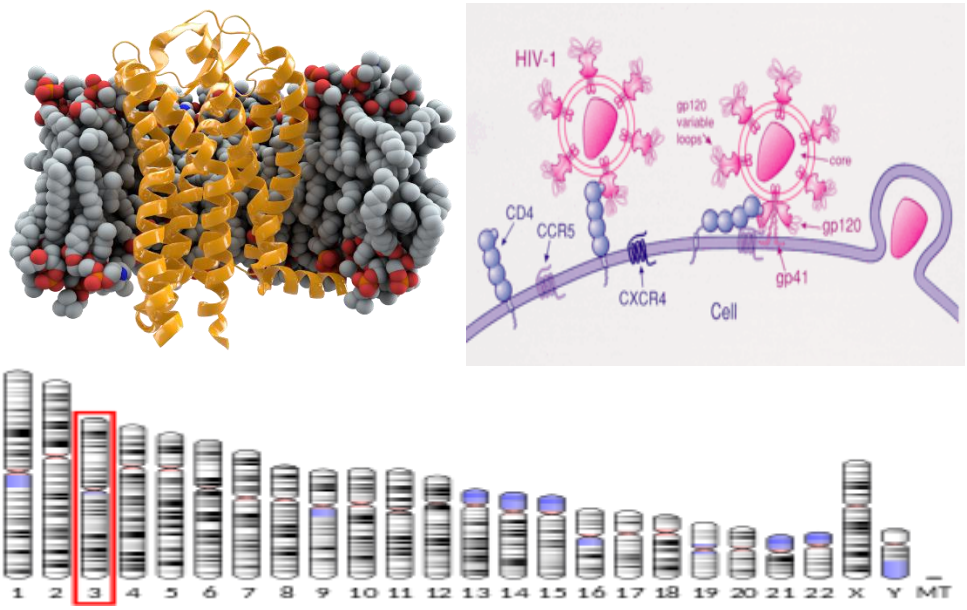

Figure-6: CC Chemokine receptor bed.

CCR5 is predominantly expressed on T cells, macrophages, dendritic cells, eosinophils, microglia and a subpopulation of either breast or prostate cancer cells. The expression of CCR5 is selectively induced during the cancer transformation process and is not expressed in normal breast or prostate epithelial cells. Approximately 50% of human breast cancer expressed CCR5, primarily in triple negative breast cancer. CCR5 inhibitors blocked the migration and metastasis of breast and prostate cancer cells that expressed CCR5, suggesting that CCR5 may function as a new therapeutic target. Recent studies suggest that CCR5 is expressed in a subset of cancer cells with characteristics of cancer stem cells, which are known to drive therapy resistance, and that CCR5 inhibitors enhanced the number of cells killed by current chemotherapy. It is likely that CCR5 plays a role in inflammatory responses to infection, though its exact role in normal immune function is unclear. Regions of this protein are also crucial for chemokine ligand binding, the functional response of the receptor, and HIV co-receptor activity.[7]

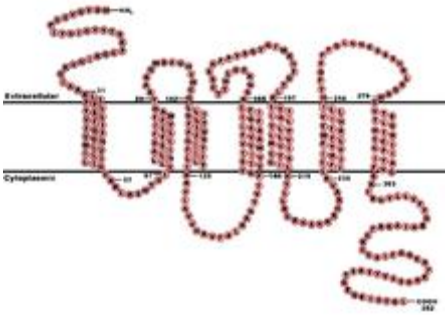

Figure-7: Protein sequence.

Primary Protein Sequence: HIV-1 most commonly uses the chemokine receptors CCR5 and/or CXCR4 as co-receptors to enter target immunological cells. These receptors are located on the surface of host immune cells whereby they provide a method of entry for the HIV-1 virus to infect the cell. The HIV-1 envelope glycoprotein structure is essential in enabling the viral entry of HIV-1 into a target host cell. The envelope glycoprotein structure consists of two protein subunits cleaved from a Gp160 protein precursor encoded for by the HIV-1 *env* gene: the Gp120 external subunit and the Gp41 transmembrane subunit. This envelope glycoprotein structure is arranged into a spike-like structure located on the surface of the virion and consists of a trimer of Gp120-Gp41 hetero-dimers. The Gp120 envelope protein is a chemokine mimic. Though it lacks the unique structure of a chemokine, it is still capable of binding to the CCR5 and CXCR4 chemokine receptors. During HIV-1 infection, the Gp120 envelope glycoprotein subunit binds to a CD4 glycoprotein and a HIV-1 co-receptor expressed on a target cell, forming a heterotrimeric complex. The formation of this complex stimulates the release of a fusogenic peptide, causing the viral membrane to fuse with the membrane of the target host cell. Because binding to CD4 alone can sometimes result in gp120 shedding, gp120 must next bind to co-receptor CCR5 in order for fusion to proceed. The tyrosine-sulphated amino terminus of this co-receptor is the "essential determinant" of binding to the gp120 glycoprotein. The co-receptor also recognizes the V1-V2 region of gp120 and the bridging sheet (an antiparallel, 4-stranded β sheet that connects the inner and outer domains of gp120). The V1-V2 stem can influence "co-receptor usage through its peptide composition as well as by the degree of N-linked glycosylation." Unlike V1-V2 however, the V3 loop is highly variable and thus is the most important determinant of co-receptor specificity. The normal ligands for this receptor, RANTES, MIP-1β, and MIP-1α, are able to suppress HIV-1 infection *in-vitro*. In individuals infected with HIV, CCR5-using viruses are the predominant species isolated during the early stages of viral infection, suggesting that these viruses may have a selective advantage during transmission or the acute phase of disease. Moreover, at least half of all infected individuals harbour only CCR5-using viruses throughout the course of infection.

CCR5 is the primary co-receptor used by gp120 sequentially with CD4. This bind results in gp41, the other protein product of gp160, to be released from its metastable conformation and insert itself into the membrane of the host cell. Although it has not been confirmed, binding of gp120-CCR5 involves two crucial steps: 1) The tyrosine-sulphated amino terminus of this co-receptor is an "essential determinant" of binding to gp120 (as stated previously) 2) Following step 1., there must be reciprocal action (synergy, intercommunication) between gp120 and the CCR5 transmembrane domains. CCR5 is essential for the spread of the R5-strain of the HIV-1 virus. Knowledge of the mechanism by which this strain of HIV-1 mediates infection has prompted research into the development of therapeutic interventions to block CCR5 function. A number of new experimental HIV drugs, called CCR5 receptor antagonists, have been designed to interfere with binding between the Gp120 envelope protein and the HIV co-receptor CCR5. These experimental drugs include PRO140 (CytoDyn), Vicriviroc (Phase III trials were cancelled in July 2010) (Schering Plough), Aplaviroc (GW-873140) (GlaxoSmithKline) and Maraviroc (UK-427857) (Pfizer). Maraviroc was approved for use by the FDA in August 2007. It is the only one thus far approved by the FDA for clinical use, thus becoming the first CCR5 inhibitor. A problem of this approach is that, while CCR5 is the major co-receptor by which HIV infects cells, it is not the only such co-receptor. It is possible that under selective pressure HIV will evolve to use another co-receptor. However, examination of viral resistance to AD101, molecular antagonist of CCR5, indicated that resistant viruses did not switch to another co-receptor (CXCR4), but persisted in using CCR5: they either bound to alternative domains of CCR5 or to the receptor at a higher affinity. However, because there is still another co-receptor available, it is probable that lacking the CCR5 gene does not make one immune to the virus; it would simply be more challenging for the individual to contract it. Also, the virus still has access to CD4. Unlike CCR5, which is not required (as evidenced by those living healthy lives even when lacking the gene as a result of the delta32 mutation), CD4 is critical in the body's immune defence system. Even without the availability of either co-receptor (even CCR5), the virus can still invade cells if gp41 were to go through an alteration (including its cytoplasmic tail) that resulted in

the independence of CD4 without the need of CCR5 and/or CXCR4 as a doorway.[8]

CCR5-Δ32: CCR5 Δ32 is a 32-base-pair deletion that introduces a premature stop codon into the CCR5 receptor locus, resulting in a non-functional receptor. CCR5 is required for M-tropic HIV-1 virus entry. Individuals homozygous (denoted Δ32/Δ32) for CCR5 Δ32 do not express functional CCR5 receptors on their cell surfaces and are resistant to HIV-1 infection, despite multiple high-risk exposures. Individuals heterozygous (+/Δ32) for the mutant allele have a greater than 50% reduction in functional CCR5 receptors on their cell surfaces due to dimerization between mutant and wild-type receptors that interferes with transport of CCR5 to the cell surface. Heterozygote carriers are resistant to HIV-1 infection relative to wild types and when infected, heterozygotes exhibit reduced viral loads and a 2-3-year-slower progression to AIDS relative to wild types. Heterozygosity for this mutant allele also has shown to improve one's virological response to anti-retroviral treatment. CCR5 Δ32 has an (heterozygote) allele frequency of 10% in Europe, and a homozygote frequency of 1%.

Recent research indicates that CCR5 Δ32 enhances cognition and memory. In 2016, researchers showed that removing the CCR5 gene from mice significantly improved their memory. CCR5 is a powerful suppressor for neuronal plasticity, learning, and memory; CCR5 over-activation by viral proteins may contribute to HIV-associated cognitive deficits.

Evolutionary history and age of the allele: The CCR5 Δ32 allele is notable for its recent origin, unexpectedly high frequency, and distinct geographic distribution, which together suggest that (a) it arose from a single mutation, and (b) it was historically subject to positive selection.

Two studies have used linkage analysis to estimate the age of the CCR5 Δ32 deletion, assuming that the amount of recombination and mutation observed on genomic regions surrounding the CCR5 Δ32 deletion would be proportional to the age of the deletion. Using a sample of 4000 individuals from 38 ethnic populations, Stephens et al. estimated that the CCR5-Δ32 deletion occurred 700 years ago (275-1875, 95% confidence interval). Another group, Libert et al. (1998), used microsatellite mutations to estimate the age of the CCR5 Δ32 mutation to be 2100 years (700-4800, 95% confidence interval). On the basis of observed recombination events, they estimated the age of the mutation to be 2250 years (900-4700, 95% confidence interval). A third hypothesis relies on the north-to-south gradient of allele frequency in Europe, which shows that the highest allele frequency occurred in the Nordic countries and lowest allele frequency in southern Europe. Because the Vikings historically occupied these countries, it may be possible that the allele spread throughout Europe due to the Viking dispersal in the 8th to 10th centuries. Vikings were later replaced by the Varangians in Russia, which may have contributed to the observed east-to-west cline of allele frequency.

HIV-1 was initially transmitted from chimpanzees (*Pan troglodytes*) to humans in the early 1900s in Southeast Cameroon, Africa, through exposure to infected blood and body fluids while butchering bushmeat. However, HIV-1 was effectively absent from Europe until the 1980s. Therefore, given the average age of roughly 1000 years for the CCR5-Δ32 allele, it can be established that HIV-1 did not exert selection pressure on the human population for long enough to achieve the current frequencies. Hence, other pathogens have been suggested as agents of positive selection for CCR5 Δ32, including bubonic plague (*Yersinia pestis*) and smallpox (*Variola* major). Other data suggest that the allele frequency experienced negative selection pressure as a result of pathogens that became more widespread during Roman expansion. The idea that negative selection played a role in the allele's low frequency is also supported by experiments using knockout mice and Influenza A, which demonstrated that the presence of the CCR5 receptor is important for efficient response to a pathogen.

Evidence for a single mutation: Several lines of evidence suggest that the CCR5 Δ32 allele evolved only once. First, CCR5 Δ32 has a relatively high frequency in several different European populations but is comparatively absent in Asian, Middle Eastern and American Indian populations, suggesting that a single mutation occurred after divergence of Europeans from their African ancestor. Second, genetic linkage analysis indicates that the mutation occurs on a homogenous genetic background, implying that inheritance of the mutation occurred from a common ancestor. This was demonstrated by showing that the CCR5 Δ32 allele is in strong linkage disequilibrium with highly polymorphic microsatellites. More than 95% of CCR5 Δ32 chromosomes also carried the IRI3.1-0 allele, while 88% carried the IRI3.2 allele. By contrast, the microsatellite markers IRI3.1-0 and IRI3.2-0 were found in only 2 or 1.5% of chromosomes carrying a wild-type CCR5 allele. This evidence of linkage disequilibrium supports the hypothesis that most, if not all, CCR5 Δ32 alleles arose from a single mutational event. Finally, the CCR5 Δ32 allele has a unique geographical distribution indicating a single Northern origin followed by migration. A study measuring allele frequencies in 18 European populations found a North-to-South gradient, with the highest allele frequencies in Finnish and Mordvinian populations (16%), and the lowest in Sardinia (4%).[9]

Positive selection: In the absence of selection, a single mutation would take an estimated 127,500 years to rise to a population frequency of 10%. Estimates based on

genetic recombination and mutation rates place the age of the allele between 1000 and 2000 years. This discrepancy is a signature of positive selection.

It is estimated that HIV-1 entered the human population in Africa in the early 1900s; however symptomatic infections were not reported until the 1980s. The HIV-1 epidemic is therefore far too young to be the source of positive selection that drove the frequency of CCR5 Δ32 from zero to 10% in 2000 years. Stephens, et al. (1998), suggest that bubonic plague (*Yersinia pestis*) had exerted positive selective pressure on CCR5 Δ32. This hypothesis was based on the timing and severity of the Black Death pandemic, which killed 30% of the European population of all ages between 1346 and 1352. After the Black Death, there were less severe, intermittent epidemics. Individual cities experienced high mortality, but overall mortality in Europe was only a few percent. In 1655-1656 a second pandemic called the "Great Plague" killed 15-20% of Europe's population Importantly, the plague epidemics were intermittent. Bubonic plague is a zoonotic disease, primarily infecting rodents, spread by fleas, and only occasionally infecting humans. Human-to-human infection of bubonic plague does not occur, though it can occur in pneumonic plague, which infects the lungs. Only when the density of rodents is low are infected fleas forced to feed on alternative hosts such as humans, and under these circumstances a human epidemic may occur. Based on population genetic models, Galvani and Slatkin (2003) argue that the intermittent nature of plague epidemics did not generate a sufficiently strong selective force to drive the allele frequency of CCR5 Δ32 to 10% in Europe.

To test this hypothesis, Galvani and Slatkin (2003) modelled the historical selection pressures produced by plague and smallpox. Plague was modeled according to historical accounts, while age-specific smallpox mortality was gleaned from the age distribution of smallpox burials in York (England) between 1770 and 1812. Smallpox preferentially infects young, pre-reproductive members of the population since they are the only individuals who are not immunized or dead from past infection. Because smallpox preferentially kills pre-reproductive members of a population, it generates stronger selective pressure than plague. Unlike plague, smallpox does not have an animal reservoir and is only transmitted from human to human. The authors calculated that if plague were selecting for CCR5 Δ32, the frequency of the allele would still be less than 1%, while smallpox has exerted a selective force sufficient to reach 10%.

The hypothesis that smallpox exerted positive selection for CCR5 Δ32 is also biologically plausible, since poxviruses, like HIV, enter white blood cells using chemokine receptors. By contrast, *Yersinia pestis* is a bacterium with a very different biology. Although Europeans are the only group to have subpopulations with a high frequency of CCR5 Δ32, they are not the only population that has been subject to selection by smallpox, which had a worldwide distribution before it was declared eradicated in 1980. The earliest unmistakable descriptions of smallpox appear in the 5th century A.D. in China, the 7th century A.D. in India and the Mediterranean, and the 10th century A.D. in southwestern Asia. By contrast, the *CCR5 Δ*32 mutation is found only in European, West Asian, and North African populations. The anomalously high frequency of CCR5 Δ32 in these populations appears to require both a unique origin in Northern Europe and subsequent selection by smallpox.

Potential costs: CCR5 Δ32 can be beneficial to the host in some infections (e.g., HIV-1, possibly smallpox), but detrimental in others (e.g., tick-borne encephalitis, West Nile virus). Whether CCR5 function is helpful or harmful in the context of a given infection depends on a complex interplay between the immune system and the pathogen.

In general, research suggests that the CCR5 Δ32 mutation may play a deleterious role in post-infection inflammatory processes, which can injure tissue and create further pathology. The best evidence for this proposed antagonistic pleiotropy is found in flavivirus infections. In general, many viral infections are asymptomatic or produce only mild symptoms in the vast majority of the population. However, certain unlucky individuals experience a particularly destructive clinical course, which is otherwise unexplained but appears to be genetically mediated. Patients homozygous for CCR5 Δ32 were found to be at higher risk for a neuroinvasive form of tick-borne encephalitis (a flavivirus). In addition, functional CCR5 may be required to prevent symptomatic disease after infection with West Nile virus, another flavivirus; CCR5 Δ32 was associated with early symptom development and more pronounced clinical manifestations after infection with West Nile virus.[10]

This finding in humans confirmed a previously observed experiment in an animal model of CCR5 Δ32 homozygosity. After infection with West Nile Virus, CCR5 Δ32 mice had markedly increased viral titres in the central nervous system and had increased mortality compared with that of wild-type mice, thus suggesting that CCR5 expression was necessary to mount a strong host defence against West Nile virus.

PROTEIN STRUCTURAL DISORDER OF THE ENVELOPE OF V3 CONTRIBUTES TO THE SWITCH IN HUMAN IMMUNOVIRUS-1: Human immunodeficiency virus type 1 (HIV-1) envelope gp120 is partly an intrinsically disordered (unstructured/disordered) protein as it contains regions that do not fold into well-defined protein structures. These disordered regions play important roles in HIV's life cycle, particularly, V3 loop-dependent cell entry,

which determines how the virus uses two coreceptors on immune cells, the chemokine receptors CCR5 (R5), CXCR4 (X4) or both (R5X4 virus). Most infecting HIV-1 variants utilise CCR5, while a switch to CXCR4-use occurs in the majority of infections. Why does this 'rewiring' event occur in HIV-1 infected patients? As changes in the charge of the V3 loop are associated with this receptor switch and it has been suggested that charged residues promote structure disorder, we found that the intrinsic disorder of the V3 loop is permissive to sequence variation thus contributing to the switch in cell tropism. To test this, we use three independent data sets of gp120 to analyse V3 loop disorder. We find that the V3 loop of X4 virus has significantly higher intrinsic disorder tendency than R5 and R5X4 virus, while R5X4 virus has the lowest. These results indicate that structural disorder plays an important role in HIV-1 cell tropism and CXCR4 binding.

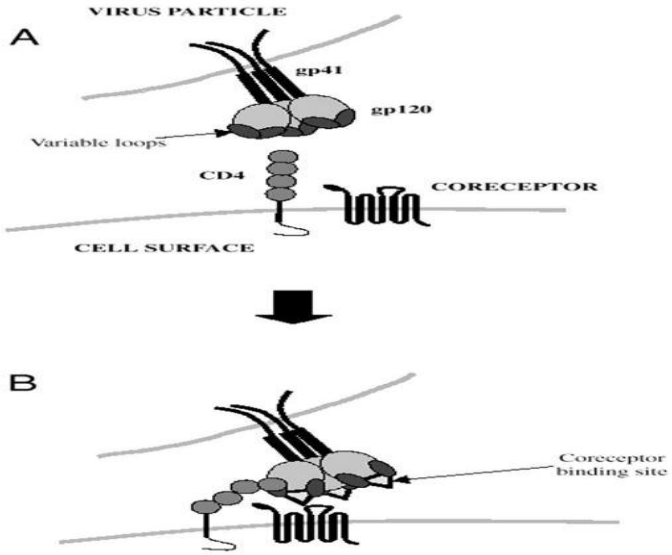

Figure-8: Virus receptor.

Receptor interactions involved in HIV entry. (A) HIV virion binds CD4. (B) CD4 binding induces conformation changes in gp120 that result in the movement of the variable loops and exposure of the co-receptor binding site.

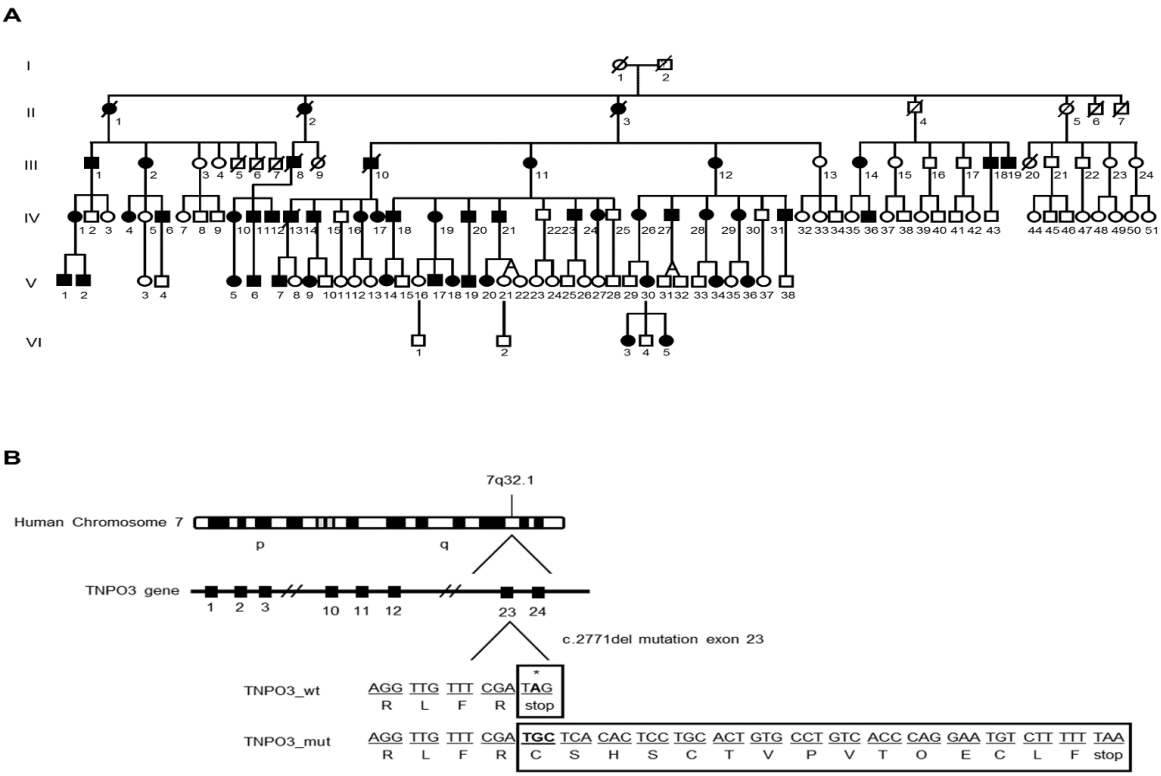

Figure-9: HIV mutation.

TNPO3 Mutation: TNPO3 gene encodes a protein called Transportin-3 in humans. The encoded protein is involved in HIV-1 infection, apparently through interaction with the HIV-1 capsid protein. TNPO3 is also known as IPO12; TRNSR; LGMD1F; LGMDD2; MTR10A; TRN-SR; TRN-SR2. The protein encoded by this gene is a nuclear import receptor for serine/arginine-rich (SR) proteins such as the splicing factors SFRS1 and SFRS2. TNPO3 has been known to be a pivotal factor in the infection by the human immunodeficiency virus (HIV-1), the causative agent of AIDS. It was seen that a relationship between a genetic defect in *TNPO3* gene and a muscle disease named Limb Girdle Muscular Dystrophy 1F (LGMD1F), with an autosomal dominant transmission, was noted. In LGMD1F patients, there is heterozygous single nucleotide deletion in the *TNPO3* gene that generates a TNPO3_mut protein. As a result, the cells from patients with this mutation in TNPO3 are resistant to HIV-1 infection in vitro. It was seen that, with an in vivo situation in which the genetic defect that causes this rare disease confers resistance to HIV infection. Therefore, *TNPO3* mutation assists us in understanding the pathogenesis of both diseases. Cells from LGMD1F patients can be used to understand the mechanisms of action of TNPO3 in HIV infection and to design new therapeutic strategies for the treatment of both diseases. Thus, HIV-1 can be used to understand the physio pathological mechanisms resulting from the mutation in TNPO3 that causes Limb Girdle Muscular Dystrophy 1F(LGMD1F).

Creating Genetic Resistance for HIV

CCR5 deletion: C-C chemokine receptor type 5, also known as CCR5 or CD195, is a protein on the surface of white blood cells that is involved in the immune system as it acts as a receptor for chemokines. This is the process by which T cells are attracted to specific tissue and organ targets. Many strains of HIV use CCR5 as a co-receptor to enter and infect host cells. A few individuals carry a mutation known as CCR5-Δ32 in the CCR5 gene, protecting them against these strains of HIV. In humans, the CCR5 gene that encodes the CCR5 protein is located on the short (p) arm at position 21 on chromosome 3. A cohort study, from June 1981 to October 2016, looked into the correlation between the delta 32 deletion and HIV resistance, and found that homozygous carriers of the delta 32 mutation are resistant to M-tropic strains of HIV-1 infection. Certain populations have inherited the Delta 32 mutation resulting in the genetic deletion of a portion of the CCR5 gene.

TNPO3 mutation: In 2019, it was discovered that the mutation of TNPO3 that is the cause of type 1F Limb-girdle muscular dystrophy (LGMD1F), also causes innate resistance to HIV-1. The gene TNP03 was known to be involved into virus transportation into the infected cells. Blood samples from a family affected by LGMD1F showed a resistance to HIV infection. While the CCR5Δ32 deletion blocks the entry of virus strains that use the CCR5 receptor, the TNPO3 mutation causing LGMD1F blocks the CXCR4 receptor, making it effective on different HIV-1 strains, due to HIV tropism.

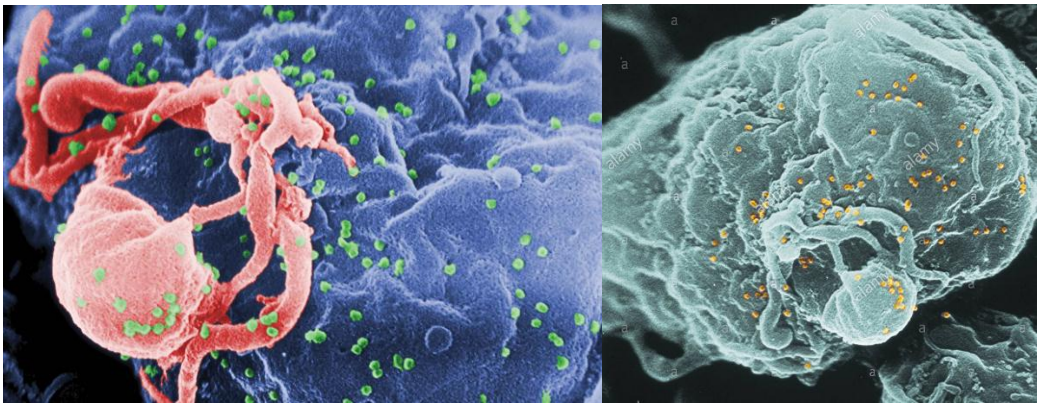

Figure-10: HIV genome [Scanning Electron Micrograph].

Cytotoxic T-lymphocytes: Cytotoxic T-lymphocytes (CTLs) provide a protective reaction against HIV when consistent exposure to the virus is present. Sex workers are found to have these CTLs within genital mucus, preventing the spread of HIV within heterosexual transmission. While creating a protective seal, CTLs become ineffective when lapses in HIV exposure occur, which leads to the possibility of CTLs only being an indicator of other genetic resistances towards HIV, such as immunoglobulin A responses within vaginal fluids.

African nonhuman primates: Chimpanzees in African countries have been found to develop AIDS at a slower rate than humans. This resistance is not due to the primate's ability to control the virus in a manner that is substantially more effective than humans, but rather because of the lack of tissues created within the body that typically progress HIV to AIDS. The chimpanzees also lack CD4 T cells and immune activation that is required for the spread of HIV.

Creating genetic resistance: While antiretroviral therapy (ART) has slowed the progression of HIV among patients, gene therapy through stem cell research gave resistance to HIV. One method of genetic modification is through the manipulation of hematopoietic stem cells, which replaces HIV genes with engineered particles that attach to chromosomes. Peptides are formed that prevent HIV from fusing to the host cells and therefore stops the infection from spreading. Another method used by the Kiem lab was the release of zinc finger nuclease (ZFN), which identifies specific sections of DNA to cause a break in the double helix. These ZFNs were used to target CCR5 in order to delete the protein, halting the course of the infection. Alternatively, to gene therapy, medication such as maraviroc (MVC) is being used to bind with CCR5 particles, blocking the entry of HIV into the cell. While not effective with all types, MVC has been proven to decrease the spread of HIV through monotherapy as well as combination therapy with ARTs. MVC is the only CCR5 binding drug approved for use by the Food and Drug Administration, the European Commission, and Health Canada.

CONCLUSION

The HIV genome encodes a small number of viral proteins invariably establishing cooperative associations among HIV proteins and between HIV and host proteins, to invade host cells and hijack their internal machineries. As a known example, the HIV envelope glycoprotein GP120 is closely associated with GP41 for viral entry. From a genome-wide perspective, a hypothesis can be worked out to determine whether 16 HIV proteins could develop 120 possible pairwise associations either by physical interactions or by functional associations mediated via HIV or host molecules. Here, we present the first systematic review of experimental evidence on HIV genome-wide protein associations using a large body of publications accumulated over the past 3 decades. Of 120 possible pairwise associations between 16 HIV proteins, at least 34 physical interactions and 17 functional associations have been identified. To achieve efficient viral replication and infection, HIV protein associations play essential roles (e.g., cleavage, inhibition, and activation) during the HIV life cycle. In either a dispensable or an indispensable manner, each HIV protein collaborates with another viral protein to accomplish specific activities that precisely take place at the proper stages of the HIV life cycle. In addition, HIV genome-wide protein associations have an impact on anti-HIV inhibitors due to the extensive cross talk between drug-inhibited proteins and other HIV proteins. Overall, this study presents for the first time a comprehensive overview of HIV genome-wide protein associations, highlighting meticulous collaborations between all viral proteins during the HIV life cycle.

REFERENCES

1. Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vézinet-Brun F, Rouzioux C, Rozenbaum W, Montagnier L (1983). "Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)". Science, 220(4599): 868–71.
2. Gallo RC, Sarin PS, Gelmann EP, Robert-Guroff M, Richardson E, Kalyanaraman VS, Mann D, Sidhu GD, Stahl RE, Zolla-Pazner S, Leibowitch J, Popovic M (1983). "Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS)". Science, 220(4599): 865–7.
3. Churi, C.; Ross, M. W. (2015). "Hiv/Aids". In Whelehan, P.; Bolin, A. (eds.). The international encyclopedia of human sexuality. Wiley. ISBN 9781405190060. OCLC 949701914.
4. Centers for Disease Control (1981). "Pneumocystis pneumonia--Los Angeles". MMWR. Morbidity and Mortality Weekly Report, 30(21): 250–2. PMID 6265753.
5. Centers for Disease Control (CDC) (1981). "Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California" (PDF). MMWR. Morbidity and Mortality Weekly Report, 30(25): 305–8.
6. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Swanstrom R, Burch CL, Weeks KM (2009). "Architecture and secondary structure of an entire HIV-1 RNA genome". Nature, 460(7256): 711–6.
7. Li G, De Clercq E (S2016). "HIV Genome-Wide Protein Associations: a Review of 30 Years of Research". Microbiology and Molecular Biology Reviews, 80(3): 679–731.
8. Singleton, P.; Sainsbury, D., eds. (2006). "Hiv". Dictionary of microbiology & molecular biology (3rd ed.). Hoboken, NJ: Wiley. ISBN 9780470035450. OCLC 71223221.
9. Montagnier L (1999). "Human Immunodeficiency Viruses (Retroviridae)". Encyclopedia of Virology (2nd ed.). pp. 763–774.
10. Lu K, Heng X, Summers MF (2011). "Structural determinants and mechanism of HIV-1 genome packaging". Journal of Molecular Biology, 410(4): 609–33.