Genetic and Molecular Characterization of Gag Gene of Human Immunodeficiency Virus

By
Syed Ahsan Shahid

Registration no: 02272013023

Department of Biotechnology
Faculty of Biological Sciences
Quaid-i-Azam, University
Islamabad, Pakistan
2022

Genetic and Molecular Characterization of Gag Gene of Human Immunodeficiency Virus

By
Syed Ahsan Shahid

Supervised by
Dr. Muhammad Ali

A thesis submitted in the partial fulfillment of the requirements for the degree of

MASTER OF PHILOSOPHY

In

BIOTECHNOLOGY

Department of Biotechnology, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad, Pakistan.

2022

In the name of Allah, The Most Gracious, The Most Merciful

Certificate of Approval

This is to certify that the Department of Biotechnology, Faculty of Biological Sciences, Quaid-i-Azam University Islamabad, Pakistan accepts the dissertation entitled "**Genetic and Molecular Characterization of Gag Gene of Human Immunodeficiency Virus**" submitted by **Syed Ahsan Shahid** in its present form as satisfying the dissertation requirement for the Degree of Master of Philosophy in **Biotechnology**.

Supervisor _____

Dr. Muhammad Ali

Assistant Professor

Department of Biotechnology

Quaid-i-Azam University, Islamabad

External Examiner _____

Dr. Shaheen Shahzad

Associate Professor

Department of Biotechnology and Bioinformatics.

Islamic International University, Islamabad.

Chairperson Javaria Qazi 07/11-22

Dr. Javaria Qazi

Associate Professor

Department of Biotechnology

Quaid-i-Azam University, Islamabad

Dated **07-11-2022**

DECLARATION OF ORIGINALITY

I hereby declare that the work ***"Genetic and Molecular Characterization of Gag Gene of Human Immunodeficiency Virus"*** accomplished in this thesis is the result of my own research carried out in ***Infectious Diseases and Molecular Pathology laboratory***, Department of Biotechnology, Quaid-i-Azam University, Islamabad, Pakistan. This research study has not been published previously nor does it contain any material from the published resources that can be considered as a violation of international copyright law. Furthermore, I also declare that I am aware of the term "copy right" and plagiarism. If, any copyright violation is found in this research work, I will be responsible for the consequence of any such violation.

Signature: 

Syed Ahsan Shahid.

Date: _____

DEDICATION

I dedicate this dissertation, with all my heart, to my beloved father, mother, brothers, sisters, my fiancé and my respected supervisor. Without their support, a bunch of sincere prayers, and sacrifices it would not have been possible for me to accomplish my work.

Table of Contents

ACKNOWLEDGEMENT

Millionth gratitude to Allah Almighty, the Most Beneficent and the most merciful who granted us health and ability to seek knowledge and explore some of the many aspects of His creation and bestowed us with the potential to bring this research work to its successful completion. Countless mercy on the Holy Prophet Mohammad (S.A.W).

I am grateful to our Chairman, Prof. Dr. Bilal Haider Abbasi for encouragement, cooperation and affectionate treatment during my study. I am thankful to my supervisor Dr. Muhammad Ali, for his personal interests, valuable guidance, constructive criticism, encouragement, and continuous support during the whole span of my research work. His helping attitude, practical knowledge and patronizing concern provide me the facilities of the center in a proper, suitable and helping manner.

I wish to express my sincere words of thanks to my beloved friends Hamad Khan, Hamza-ur-Rehman, Ahmad Jamal, Muzamil Shinwari, Muhammad Ishaq, Zain Asif Alvi, Muhammad Kamran Afridi, Salman Khan, Muhammad Idrees, Shahrukh Awan, Amen Jehangir, Zenab Syed, Qurat-ul-Ain, Saman Zahoor, and Sara Fatima for their attention, unceasing encouragement, friendly advice, and continuous support during my study. Additionally, I would like to express my deep feelings of respect to my lab fellows Sidra Rahman, Lubna Rahman, Sajid Ali Shah, Tazeen, Adnan Zeb, Hanbal Ahmad Khan, Asim Saeed, Asiya Nawaz, Iffat Irshad, Ahsan Saeed, and Naveera Khan for their aspiring guidance and valuable suggestions during my laboratory experiments.

How can I forget the aspiring guidance and help of my seniors throughout my research, Sana Ullah, Misbah-ud-din, Muhammad Haris, Rizwan Abbas and Ibrar Khan. They have been very cooperative and kind during my research period.

This thesis owes a lot to my beloved Parents, Siblings, Grandparents and Fiancé. Without their support, constant words of encouragement, inspiration, suggestion, guideline every single day of my life, and their constant push for me to excel, this thesis would not have succeeded on time

May Allah bless you all with eternal happiness and success! Ameen.

LIST OF ABBREVATIONS

%	Percentage
°C	Centigrade
µl/µg	Microliter/Microgram
aa	Amino acid
bp	Base pair
HIV	Human Immunodeficiency Virus
AIDS	Acquired Immuno-Deficiency Syndrome
cDNA	Complementary DNA
DNA	Deoxyribonucleic acid
RT-PCR	Real Time- Polymerase Chain Reaction
EDTA	Ethylene Diamine Tetra Acetate
ELISA	Enzyme Linked Immunosorbent Assay
CTL	Cytotoxic T Lymphocytes
HTL	Helper T Lymphocytes
IFN	Interferon
KP	Khyber Pakhtunkhwa
3D	Three Dimensional
Min	Minute
ml	Milliliter
PCR	Polymerase Chain Reaction
RNA	Ribonucleic Acid
ICT	Inter Coombs Test
RT	Reverse Transcriptase
Sec	Seconds
UTR	Untranslated Regions
LTR	Long Terminal Repeats
NFAT	Nuclear Factor of Activated T-cells
NF-kB	Nuclear Factor kappa B
HAART	Highly Active Anti-retroviral Therapy
TNF-alpha	Tumor Necrosis Factor alpha
NF	Nuclear Factor

LIST OF FIGURES

LIST OF TABLES

Abstract

Human Immunodeficiency Virus (HIV) which is the major cause of a condition called Acquired Immunodeficiency Syndrome (AIDS). It is a single-stranded RNA virus from *Retroviridae* family and genus Lentivirus. Global burden of HIV has been increasing and World Health Organization (WHO) recently (as of 2019) revealed a staggering increase as 1.7 million new cases and 0.69 million deaths have been reported. Pakistan is among those countries where HIV prevalence is increasing. In this current study, we isolated blood samples from high-risk behavior individuals of HIV infection referred to some cities of Pakistan like Islamabad, Lahore, Peshawar and Kohat. The HIV positive individuals were confirmed by Real Time-Polymerase Chain Reaction (RT-PCR). The Real Time PCR based positive samples were further processed for Gag gene amplification following nucleotide sequencing through Sanger method. The evolutionary history was inferred by using the Jukes-Cantor model and Neighbor-Joining method. Sequences analyzed in the current study (isolates QAU-HIV1-AH1 and QAU- HIV1-AH2) clustered with a previously reported reference sequence (accession No. KX232595). In the present study, we report several amino acid substitutions in functionally important domains of Gag gene. Key substitutions in the isolate QAU-HIV1-AH1 are W212M, P328Q, N392S, E474K, D479T, and Q480R. Whereas, amino acid substitutions found in isolate QAU-HIV1-AH2 are L321M, L322Q, V323I, N325I, A326V, N327R, N374Y, K413Q and S477C position of the latterly analyzed reference sequence (KX232595). On the basis of current results, we suggest that the mentioned substitutions are linked with structural variation in the protein and could alter protein's function(s). The Gag gene of the viral isolates from Pakistani patients need to be further explored for possible drug resistance associated substitutions.

Key words: HIV, Immunodeficiency, RNA, Infection, Gag polyproteins, Gag gene.

CHAPTER 1

INTRODUCTION

1. Introduction

Human Immunodeficiency Virus (HIV), a single stranded RNA virus of the genus Lentivirus in *Retroviridae* family (Ali et al., 2017). HIV is the causative agent of chronic life threatening and dangerous condition called Acquired Immuno-Deficiency Syndrome (AIDS) (Jacobson et al., 1993). HIV/AIDS is still a serious health problem despite the good progress in diagnosis, prevention, and treatment sector (Zak-Place & Stern, 2004). World health organization (WHO) by the end of 2019, reported that HIV had infected estimated 38 million people globally (Tappuni, 2020). In 2019, 1.7 million individuals were infected by HIV, with 0.69 million deaths worldwide (Freiberg et al., 2021). The clinical symptoms of the HIV depends on the stage of infection. In the initial stages, headache, rashes, sore throat, and fever. Later, the immune system weakens and thus develop signs and symptoms like weight loss, swollen lymph nodes, diarrhea, cough, fever and without treatment of the said conditions may lead to severe illnesses like cancer specifically Kaposi's sarcoma and lymphomas, tuberculosis and severe bacterial infections (Yarchoan & Uldrick, 2018).

The first HIV case in Pakistan was reported in 1987, which were increasing continuously. Currently, the HIV prevalence rate is less than 0.7% among general population of Pakistan. Pakistan is one of the few countries of Asia, in which the HIV incident cases are increasing significantly every year after its first case reported in 1987 (Haq et al., 2020). In Pakistan, the HIV epidemic is highly heterogeneous with a large diversity in the transmission dynamics at local, regional and national level (Raza et al., 2020). Moreover, demographic, economic and social changes in Pakistan create dynamism in the trajectory and proximate drivers of epidemic. According to the National AIDS Control Program (NACP), Pakistan has 43,964 registered HIV cases till January 5th, 2020 (Haq et al., 2020). In 2019, a high prevalence of more than 13% HIV cases have been reported from Kot Imrana, a village of district Sargodha from the province Punjab, Pakistan, which was only 1.29% earlier in 2018 (Wahid, 2019). In Pakistan, the highest HIV cases (38.4%) were reported in people who inject drugs (PWID) followed by transgender sex workers (7.5%), sex workers male (5.6%) and female (2.1%) (Haq et al., 2020).

Punjab accommodates an approximately 50% population of the country and has the highest number of HIV positive cases (Ali et al., 2017). In 2017, a study was conducted in several major cities of province Punjab, Pakistan (Sargodha, Dera Ghazi Khan, Chiniot, Multan and Rawalpindi)

to investigate the seroprevalance of HIV. The results revealed that out of 2,062 individuals, 329 (16%) were HIV positive. Among all the positive individuals, 51.4% were males, while 31.3% were females and 17.3% were transgender. The highest number of positive cases among these were from Multan district (22%) followed by Sargodha (21%), D.G Khan (Dera Ghazi Khan) (21%), Rawalpindi (19%) and Chiniot (17%) (Karim et al., 2020). Looking to all these numbers, which is highly threating to Pakistani as well as international community; we need rapid diagnostic services, awareness campaigns and the establishment of treatment centers throughout the country.

HIV infection could be detected via several diagnostic assays such as ICT method (Inter coombs test), Immunochromatographic Techniques (Peeling, Mabey, Herring, & Hook, 2006), Enzyme Linked Immunosorbent Assay (Richardson & Page, 2018) and Nucleic Acid Testing (NAT). ICT and ELISA is an easy technique for the detection of HIV and is used for routine screening testing. Although ELISA is cost-effective, but several limitations are associated with ELISA based detection. The major issue that is associated with ELISA and ICT is false positivity (Shafiee et al., 2015). Therefore, other confirmative assays are recommended such like quantitative RT-PCR (Real Time-Polymerase Chain Reaction) and Radio Immunoassay (RIA). Other limitations associated with ELISA and ICT methods are the false negative results during the early phase of HIV infection in which antibodies are not in detectable range (Shafiee et al., 2015).

PCR is an important tool that is used for the amplification and detection of nucleic acids. Currently, real-time techniques (PCR-based HIV RNA assays) are in concentration with constantly detecting fluorescence emanated during every cycle of PCR (Bustin & Mueller, 2005). HIV-1 RNA estimation is achieved during first phase of PCR, as compared to conventional PCR or other technologies, and provide more reliable results. Owing to new developed assays, the detection number has been improved much higher than old technologies (-50 copies/ml). Quantification of target sequence by real-time PCR depends on the reduction of fluorescence signals or the measurement of accumulation during the amplification cycle. Detection and amplification of samples are done simultaneously (Patel, 2021).

The genetic material of HIV includes two copies of positive sense single-stranded RNA. These strands are about 10,000 nucleotides in length. In nine overlapping ORF (Open Reading Frames) about fifteen viral proteins are encoded. HIV have structural genes as well as regulatory and accessory genes. These are all processed by three different precursors by post translational

cleavage to yield matrix proteins *p17*, core proteins *p24, p7, p6* and *p15* (gag), the enzymes like protease (PR), integrase (IN), and reverse transcriptase (RT) are encoded by the *pol* gene (Nunzio, 2018). The envelope glycoproteins *gp41* and *gp120* that forms the receptors present on the cell surface (Gelderblom, Özel, & Pauli, 1989). The *TAT* and *rev* are essential regulatory genes. *TAT* gene encodes for *TAT* proteins very soon after infection and it encourages the expression of other genes of HIV. *Rev* gene encodes for *rev* proteins that allows the transfer of the processed messenger and genomic RNA from nucleus to the cytoplasm. The other accessory genes are the *Viral Protein-U (VPU), the Viral Protein-R (VPR), Viral Infectivity Factor (VIF), Viral Protein X (VPX)* and *Nef.*

The HIV-1 precursor protein Gag is a 35kDa protein (also called p55) which is cleaved into structural proteins by a viral protease (34kb) present at the 5' end of the pol gene. When the cleavage of the Gag proteins occurs it is divided into matrix protein (p17), capsid protein (p24), nucleocapsid protein (p7), another small protein (p6), and two small peptides (p2 and p1) of unknown functions, are also part of the viral core structural proteins. These proteins are present in all retroviruses and in both types of Human Immunodeficiency Virus (HIV-1 and HIV-2) (Pillay et al., 2009). Gag has a role in the outer protein coat assembly and viral particle synthesis. Clavel and colleagues reported that it also has a role in HIV resistance to protease inhibitors (Clavel & Mammano, 2010).

Growing numbers and low awareness of HIV/AIDS is the cause of persistent HIV epidemics in Pakistan. Risk factors including unscreened blood, and unprotected intercourse makes people more prone to spread of HIV to extremely dangerous levels (Khan et al., 2010). The government's health policy must be proactive in tackling this precarious health threat. Most of the Pakistani population is safe from HIV but it doesn't take long for infectious disease to spread across border. Epidemics of HIV have emerged a few times therefore, infections can spread to a peak rise in no time so strict rules and regulations must be set up by the regulatory authorities to control the spread of the AIDS.

To control the extent of AIDS in Pakistan, the government needs not only to spread awareness at national level but practical steps such as wide-spread screening should be taken for population at high risk of AIDS. The stigma regarding HIV/AIDS in public could be eliminate if well-known public figures such as politicians and sports celebrities discuss it on bigger platforms. Once the discrimination against HIV/AIDS is neutralized, Pakistan will be able to win against this disease

relatively easily. Action however must be taken at present, as quick steps taken now can help prevent a wide-spread HIV epidemic in Pakistan.

The current study was aimed to examine the HIV infection in the high-risk behavior individuals who were referred for HIV testing. Further, to confirm the HIV positive (ICT based) samples by highly specific RNA quantification RT-PCR and to explore the Gag gene of HIV isolates for nucleotide sequencing, its phylogenetic analysis and mutational analysis in Pakistani population.

1.1 Aims and objectives

1. To perform nucleotide sequencing and phylogenetic analysis of Gag gene of HIV.
2. To determine amino acids substitution in Gag proteins in chronically HIV infected patients and its impact on structure and function.

CHAPTER 2

LITERATURE REVIEW

2. Literature Review:

2.1. Human Immunodeficiency Virus (HIV)

Human Immunodeficiency Virus (HIV) causes a chronic and life threatening condition called Acquired Immuno-Deficiency Syndrome (AIDS). It belongs to the subfamily *Lentivirinae* of family *Retroviridae*. All viruses from this family contains RNA as their genetic material and a reverse transcriptase enzyme for the reverse transcription of its RNA to DNA causing infection in the host. HIV is divided into two types that are HIV-1 and HIV-2. (Gao et al., 1999). The potent causative agent among these two is HIV-1 which has infected millions of people throughout the world. While HIV-2 infections have been limited and reported in some areas of Western and Central Africa. Both viruses consist of 2 identical copies of single-stranded RNA molecules which includes regulatory, structural, and accessory genes. The structural genes consists of *gag, pol* and *env* genes, the regulatory genes consists of *tat* and *rev*, and the accessory genes comprises *vif, vpu, vpr*, and *Nef* (Freed, 1998). The basic structure of HIV-1 and HIV-2 are same, however there is some difference in the organization of their genome. However, HIV-2 is less lethal than HIV-1. The HIV targets immune system of host like the T helper cells, macrophages or dendritic cells by different mechanisms like apoptosis or direct viral killing that mainly cause reduction in immunity, due to which it increased susceptibility to opportunistic infections like Hepatitis-C virus, hepatitis-B Virus, sexually transmitted diseases and a variety of co-infections. (Zahra et al., 2021)

2.2. History of Human Immunodeficiency Virus (HIV):

This disease was present in Central and Eastern Africa for a very long time. The serological proof of HIV was found in samples that were stored in 1959 from humans in Zaire (Jin et al., 1994). HIV-1 is thought to be originated from chimpanzee (*Pan troglodytes*) in Central and West Africa which has been possibly transmitted to Humans through hunting and slaughtering of meat. Moreover, HIV-2 is thought to be originated from West African Green monkeys and proved to be less virulent than HIV-1 but AIDS can be caused by both types of HIV (Jin et al., 1994). The CDC (Centers for Disease Control) on June 5th 1981, published a report about five healthy young homosexual men present in Los Angeles who got pneumonia (*Pneumocystis carinii*), two of which had died. A remarkable reduction in cellular immune response was observed during examination of those patients and it was common among all of them. The media influenced the disease as Gay-related immune deficiency (GRID). Similar cases frequently started to report from Western

Europe, and they were characterized by compromised immune system that leads to opportunistic infections and rare malignancies. These infections were not only limited to men who have sex with men (MSM), but also targeted people belonging to other groups like injecting drug users and blood transfusion patients. Later on, this newly recognized entity was termed Acquired Immune Deficiency Syndrome (AIDS) by CDC (Center for Disease Control) in 1982 (Harden, 2009). Luc Montagnier and colleagues in January 1983, at the Pasteur institute (France) isolated a potential causative agent from cultured T lymphocytes that were derived from cervical lymphadenopathy patient, which was considered as a sign for precursor of AIDS (Papadopulos-Eleopulos et al., 2004). This new retrovirus initially named LAV (lymphadenopathy-associated virus) (Barré et al., 1983). Robert Gallo and colleagues in May 1984, found enough evidences to convince the scientific and medical communities that it was a new virus, later it was renamed as human T-lymphotropic virus type III (HTLV-III). This new virus was considered as the etiological agent for the AIDS epidemic situation (Gallo et al., 1984). Gallo's findings were confirmed by Jay and levy in August 1984, they reported the isolation of a retrovirus named as AIDS-associated retrovirus (ARV) in 22 AIDS patients (Leyv et al., 1984).

Sequencing of the entire genome from ARV, LAV and HTLV-III viruses was done in February 1985, which defined that they were all the decedents of the same virus (Ratneir et al., 1985). In 1986, the name human immunodeficiency virus (HIV) was proposed by the Committee on the Taxonomy of Viruses and were reported as the AIDS causing agent. Another retrovirus was isolated from a patient at West African Lisbon hospital (Claevel et al., 1986), which showed similar clinical symptoms with AIDS but no antibodies were found against HIV. This leads to the discovery of a new virus and named as human immunodeficiency virus type 2 (HIV-2), while the original/old HIV was renamed as human immunodeficiency virus type 1 (HIV-1). In 2008, Francoise Barré-Sinoussi and Luc Montagnier (Pasteur Institute) were awarded the Nobel Prize in Medicine for the HIV-1 discovery. HIV has rolled out to almost all parts of the world and its prevalence is highest in Sub-Saharan Africa where the major cause of death is AIDS (Quinn, 2008). CDC (1992) reported AIDS as the major cause of death for U.S. men between the age of 25 to 44 and that year 2,332 cases reported in San Francisco (Padamsee, 2020).

2.3. World-wide distribution of HIV-1

2.3.1. Africa

HIV-1 was initially originated from Chimpanzees in West of Central Africa and became one of the largest global public health issue that almost cost 33 million lives till date (WHO). It has been assessed that more than 0.2 billion people in Africa are living with HIV/AIDS and annually 0.73 million new HIV-1 infections occur due to lack of testing facilities. The high mutation rate and fast replication cycle cause more genetic variation in HIV-1, which cause the assembly of multiple viral variants in an infected patient during a single day. Therefore, the diagnoses, care, treatment and interventions against the HIV-1 subtypes and new recombinant strains which are circulating through African countries pose to be a serious problem (Jaffar et al.,2004). Approximately 20% of adults in South African regions are living with HIV-1 and this is the highest HIV-1 prevalence worldwide. Collective analysis regarding the distribution of HIV-1 strains in Africa (1990–2020), the most predominated is the subtype B in Northern Africa. CRF02_AG has been the most prevalent in Central and Western Africa, while in Southern and Eastern Africa, subtype C has predominated (Giovanetti et al., 2020).

2.3.2. Europe.

Although Europe had been kept isolated from HIV pandemic due to its strict policies and restrictions regarding contact with foreigners, they implemented strict laws to bring social control. HIV-1 was introduced during early-1980s in Western and Central Europe as subtype B, which is the major clade till date (Beloukas et al., 2016). Though, the subtype A (AFSU) being a predominant strain introduced in Eastern Europe like Russia and Former Soviet Union (FSU) countries in the mid-1990s, and it transmitted mostly in PWID (People Who Inject Drugs) (Hamers & Downs, 2003). During recent years, the non-B subtypes prevailed and other recombinant strains in the West and Central Europe caused a complex tapestry. These are more often associated with heterosexuals, immigrants, and females. Subtype-A in Albania, Cyprus and Greece while subtype G in Portugal. A single and unique case of sub-subtype F1 in Romania has predominated throughout the epidemic (Abecasis et al., 2013). By the end of 2001, cumulative total of 251,237 HIV infection cases reported in both the East and Central European regions. And a major increase in number of HIV diagnoses from 234 cases (1994) to 99,499 (2001), among those increased HIV

cases most of them were injecting drug users and majority of those infected heterosexually are women (Hamers & Downs, 2003).

2.3.3. North and South America

During late 1980's HIV has been detected in the North and South America. WHO reported that over 1.5 million people in mid-1993 were infected with HIV/AIDS in North America (Merson, 1993) and infections were on a rise of over 1.5 million cumulative adult infections. Majority of them were due to homosexual intercourse and drug injections (Merson, 1993). The most widespread viral strains in North America is subtype B (Oster et al., 2017). Moreover, it also dominates in South America, while subtypes C, subtype F, subtype CRF12_BF and the related B/F recombinants strains have been reported in Argentina, Uruguay, and Bolivia (Thomson et al., 2000). In South Brazil an increase in pure subtype C has also been found among co-infected patients (Avanzi et al., 2017).

2.3.4. Asia

In Asia, HIV was reported quite late as compared to the rest of the world, the first case reported in 1984, yet the potential threat of epidemic was not considered until the wide spread of HIV in Burma, India, Cambodia and Thailand during early 1990s. The disease prevention and control stories at the national level in those areas remain rare (Ruxrungtham et al., 2004). As of many other areas, Asia has been reported to be a "hotbed" for its diversity of recombinant viral strains (Lemey et al., 2003). The Asian regions were temporarily divided into three groups for HIV prevalence: The first group include those areas which were hit hard and early like in some states of India, Burma, Cambodia and Thailand the HIV prevalence exceeds 1% in adults (Brown et al., 1994). The second group comprises those areas which were in transition and growing noticeably in the 5 years including China, Nepal, Indonesia and Vietnam. The third one comprises those who were having low infection levels such as: South Korea, Bangladesh, Philippines, Laos and South Korea (Dore et al., 1996). In Asian countries, the HIV epidemic situation have been driven largely by sex workers, injecting drug, man having sex with man and heterosexual transmission (Beyrer et al., 2000; Mertens & Low-Beer, 1996). In Russia and other Soviet Union countries, subtype A predominates even though in Kyrgyzstan, CRF02_AG prevails (Aibekova et al., 2018). In China, the subtype CRF01_AE and CRF07_BC were found dominant (Chu et al., 2017; Indriati et al.,

2018). CRF01_AE in South-east Asia was the most common while subtype C dominates in India (Jenness et al., 2017).

Pakistan stood among the list of those 4 countries (Afghanistan, Bangladesh and Philippines) where HIV infection has been increased since 1990. By November 1996, a total of 64 AIDS cases reported to National AIDS Program. Though, it has been estimated by WHO that the actual HIV infected cases were 40,000 by the end of 1994 when they considered those cases that were in diagnosis phases, under-reporting and delayed reporting were considered (Khawaja et al., 1997). It was estimated that about 150,000 adults and children were infected with HIV in Pakistan by the end of 2017, representing 127% rise of HIV positive cases from 2010 onwards. About 3500 (2%) of the total were children younger than 15 years (Azim et al., 2008; Mir et al., 2020). The HIV outbreak in Pakistan has been condensed between population that involves PWID (people who inject drugs), MSM, males, females, transgender and sex workers (Azim et al., 2008). The prevalence of HIV epidemic in big cities showed the majority of infection ratio among PWID which is about 35%. In sex workers the HIV prevalence was less i.e., about 1% for female prostitutes and 7% for transgender sex workers (Khan et al., 2010). Large number of pediatric cases of HIV were diagnosed in Karachi (April 2019). It was found that those 46 HIV positive children were having thalassemia and underwent blood transfusion and other treatments at local medical center in Larkana (Mir et al., 2020). HIV-1 subtype A was the only subtype of HIV prevalent in Pakistan (Rai et al., 2010).

2.4. Morphology of Human Immunodeficiency Virus (HIV)

2.4.1. General structure

By using negative staining of electron microscopy, HIV-1 and HIV-2 has been studied. A large amount of 130-200 nm particles containing a 130-nm-long by 30-70 nm-wide core of HIV-2 has been observed in culture. Morphologically the core is pear or conical shaped. Some of those particles displayed a small fringe like structures that may perhaps be seen to contain a regular organization of repeating subunits. While the HIV-1 cultures were found to have same particles but in much lower numbers. HIV-1 comprises an ambiguous projective structure, HIV-2 comprises a knob-like and SIV a projection-like structure. The virus consists a fringed envelope, the inner membrane, and the core, that contains the nucleoid. (Chrystie & Almeida, 1988).

2.4.2. Envelope

The HIV (1 and 2) has a double layered envelope with projections on its surface. These projections are affected by the outer environment of the cell and have an important role in the interaction and entry with host proteins during adsorption, fusion and entry of HIV into the host membrane. It has a 4.8nm in diameter head, 3.2nm long and 1.0nm wide stalk. The length of HIV varies with respect to the visible width of the envelope (Goto et al., 1998).

2.4.3. Core

The core of the HIV has been studied by negative staining techniques and ultrathin sectioning despite it has not been fully clarified yet. The core of HIV basically have 2 dimensional shapes which are bar like, triangular and sectorial (Gelderblom et al., 1988). The presence of the protein p24 in HIV-1 core has been found. The core of HIV may be made of about 380 units, even though the size of these cores may differ (Ehrlich et al., 2001). Some images of the ultrathin microscopy revealed some cores might be connected via a narrow end to the envelope while some might be isolated from the envelope (Goto et al., 1998).

2.4.4. Matrix

Matrix basically has ground equivalent substances which supports envelope and the core. Bouillant and Becker in 1984 explained the matrix as an electron-lucent space or an electron-lucent layer (Goto et al., 1998). Without matrix the virus particle may seem like a tear drop. The computer analysis techniques found that the matrix of HIV is an icosahedral structure (Marx et al., 1988).

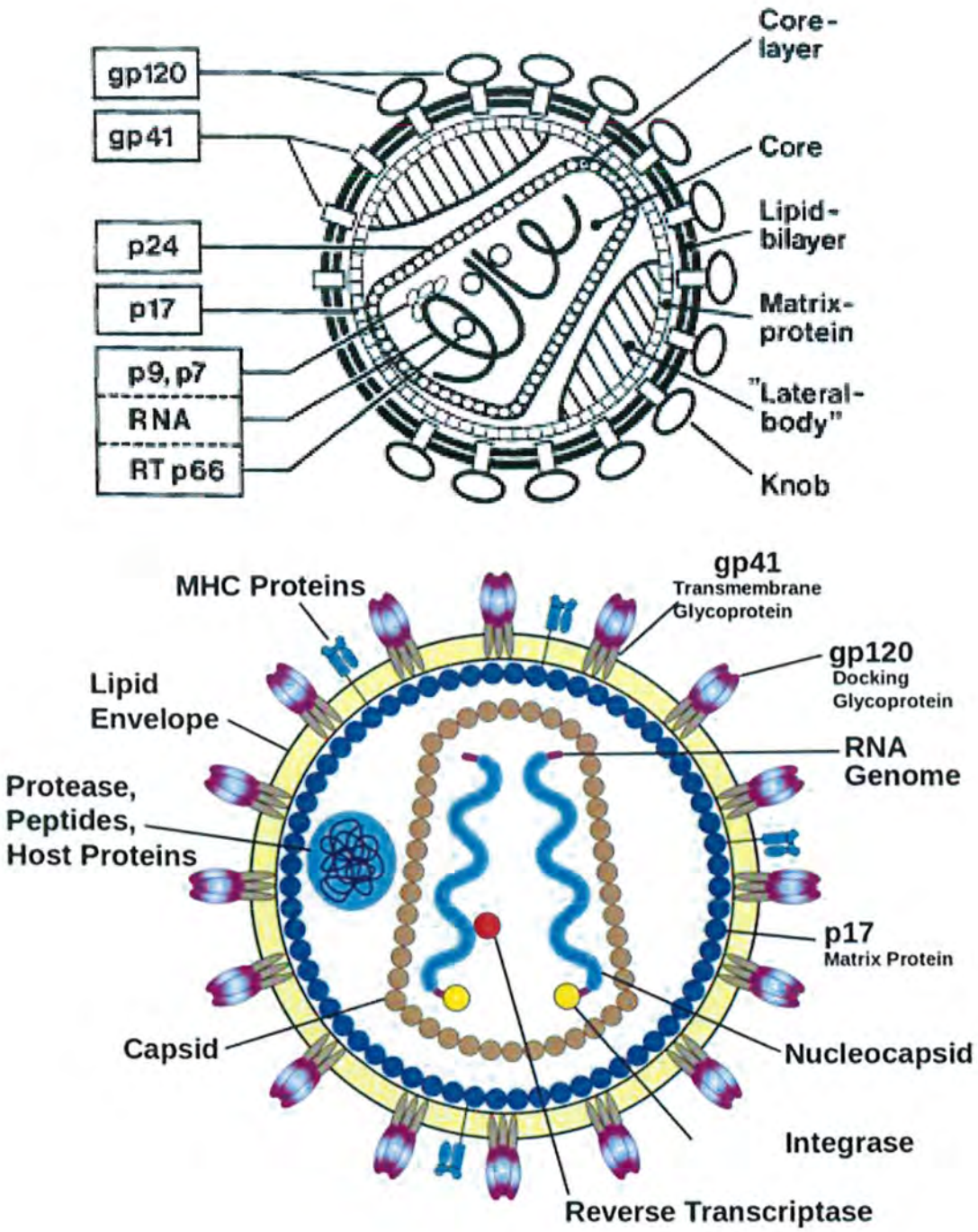


Figure 2.1. Morphology of Human Immunodeficiency Virus (Gelderblom et al., 1988).

2.5. Genome organization of HIV

The human immunodeficiency virus has approximately 10kb genomic length and comprises of two copies of positive sense, unspliced, 5′ capped and 3′ -polyadenylated single stranded RNA molecules and these RNA molecules are about 10,000 nucleotides in length each. HIV-1 has been categorized into groups, subtypes and sub-subtypes, the categorization is based on genetic variation and phylogenetic analysis. The three groups such as M, N and O and these groups further divided as subtypes. The M group subtypes are A,B,C,D, F,G,H, J and K (Rajarapu, 2014). They have nine overlapping open reading frames (ORF) which codes for about fifteen viral proteins. The HIV comprises of structural, regulatory and accessory genes. The *Gag, Pol* and *env* are structural genes which codes structural proteins, these genes can be found in all retroviruses. These are all processed by three different precursors by post translational cleavage to yield matrix proteins *p17*, core proteins *p24, p7, p6* and *p15* (gag), the pol gene codes for the important regulatory enzymes like reverse transcriptase (RT), protease (PR) and integrase (IN) (Nunzio, 2018). The envelope glycoproteins *gp41* and *gp120* that forms the receptors are present on the surface of cell (Gelderblom et al., 1989). The *TAT* and *rev* are essential regulatory genes. *TAT* gene codes for *TAT* protein very soon after infection and it encourages the expression of other genes of HIV. *Rev* protein is encoded by the *rev* gene that allows the export of the appropriately processed mRNA and genomic RNA from nucleus to the cytoplasm. The other accessory genes are the *Viral Protein-U (VPU), the Viral Protein-R (VPR), Viral Infectivity Factor (VIF), Viral Protein-X (VPX)* and *Nef*. The *VPR* proteins have a very important role in the cell cycle arrest as it facilitates the viral cDNA into the nucleus of the host cell, where the viral integration into the host genome completes. After that the release of the virus particle is done by the *VPU* proteins and another small proteins, *VIF* protein increases the infectiveness of viral progeny by targeting the antiviral activity of the human enzyme *"APOBEC3G"* or *A3G* by targeting it for cellular degradation and ubiquitination (Rücker, Grivel, Münch, Kirchhoff, & Margolis, 2004). The negative regulatory protein *(NEF)* has many functions like down regulation and receptor expression of CD4 on the surface of the infected T-cell, it also enhances the release of the virions and improves the viral replication and takes control of the cellular signal transduction of the host cell (Lusic & Siliciano, 2017).

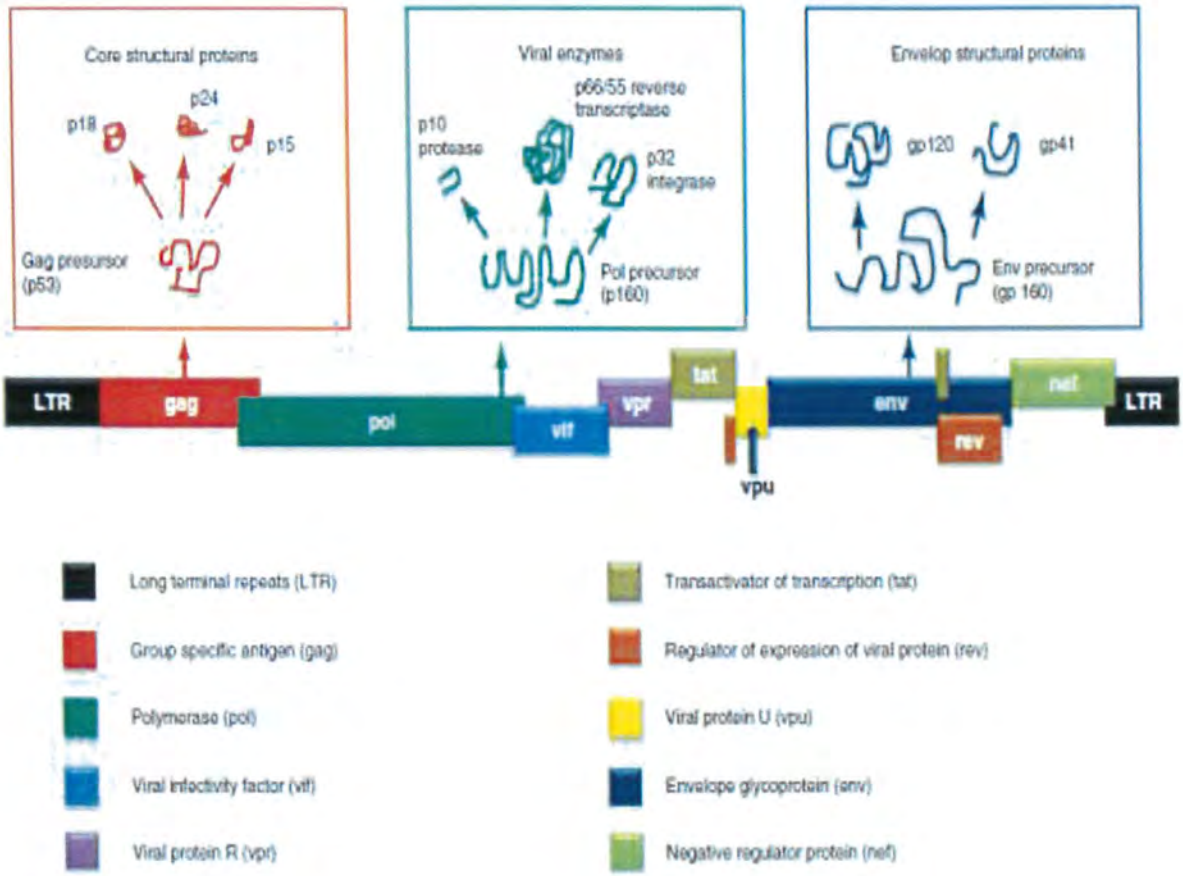


Figure 2.2. Genomic Organization of HIV (Rajarapu, 2014).

2.6. The Life Cycle of HIV

The HIV replication cycle consists a series of steps, as viruses do not have their metabolic machinery. HIV depends on the host cells for their reproduction and thus they are called the intracellular parasites.

2.6.1. Viral entry

The macrophages, dendritic or T-helper cells are of major importance in regulation of the humoral and cellular response against any pathogen. The main targets of the HIV are these regulatory cells that comprises CD4+ helper T cells. Immature dendritic cells, macrophages and other T-cell subsets are also the targets of HIV but these cell types have no role in the virus replication (Kirchhoff, 2013). Thus the targeting and destruction of the CD4 and T cells makes the body unable to protect itself against HIV or any other pathogen. The CD4 cells have affinity for HIV envelope spike proteins that is *gp120* (glycoprotein 120) and *gp41* (glycoprotein 41), CD4 receptors along with co-regulators and chemokine receptor *(CCR5)* binds with *gp120* of HIV (Freed, 2001). This binding of the host cells to the HIV causes conformational changes in *gp41* that brings virus closer to the membrane and thus membrane fusion occurs between the host's cell membrane and the lipid bilayer of HIV, which allows the viral core to enter into the cytoplasm (Zhou et al., 2004). As soon as the viral core enters the host cell the capsid degrades and un-coating of the virus takes place, after that the viral genome manipulates the transcriptional and translational machinery of the host to replicate itself (Kirchhoff, 2013).

2.6.2. Reverse Transcription

Once the virus fuses with the host cell, its genetic material enters the host cell. The HIV genome comprises of two positive sense single stranded RNA molecules protected by nucleocapsid. Once the fusion occurs, both the single-stranded viral RNAs transcribes into linear double-stranded DNAs by a process called reverse transcription (RT) (Kirchhoff et al,. 2013). This process consists of the generation of mRNA from nuclear DNA, the RNAase H activity is used to degrade RNA hybrid which are formed as a result of reverse transcription. The cDNA synthesis is followed by its export across nuclear membrane for protein synthesis. This whole mechanism is completed by an enzyme named reverse transcriptase. This characteristic is followed by all the other retroviruses and this whole process of reverse transcription involves a series of very complex events (Bour et al., 1995).

2.6.3. Integration

Once the linear ds DNA is created and its transportation occurs across the nuclear membrane, the insertion of the viral genome into the host cell occurs for productive infection and its gene expression. Integrase enzyme is responsible for the activity of the integration of viral DNA into the host DNA (Anderson & Maldarelli, 2018). The host cell remains infected for the rest of its life and the integrated virus or viral genome in the host genome is called the provirus. The viral genome integrated in the host behaves fundamentally as a cellular gene. This proviral DNA performs the replication process with the host DNA. The viral infection spreads by infecting new cells or by multiplication of the proviral DNA (Chiu & Davies, 2004). The integration is quite a critical step in retroviral replication because mutants are commonly unable in spreading infections. In some cells, the integrated provirus can stay quiet/silent for many years and this can pose a major hurdle in virus eradication. Due to its inactive form, the immune system cannot recognize it and thus unable to eliminate it. Integrase enzyme has three domains: the C-terminal domain, a central core and the N-terminal with a zinc finger (Goodsell, 2015).

2.6.4. Gene Expression

Once the viral genome is integrated into the host genome, the provirus works as a template for the formation of the viral RNAs which codes for all the structural, accessory and regulatory proteins of HIV. It works as a template for the transcription of viral messenger RNAs and genomic RNA by utilizing the polymerase enzyme of the cellular machinery of the host (Wu & Marsh, 2003). The initiation of the transcription process of the provirus takes place by the viral promoter that resides the U3 region of the 5′ LTR and contains many cis-acting molecules that are essential for the formation of RNA. This gene expression of the virus extremely depends on the cellular transcription factor (*NFAT and NF-kB*) (Gaynor, 1992). The transcriptional output is low in the start due to the inefficient viral transcripts and the viral trans-activator protein *TAT* is needed for effective viral gene expression. To increase the transcriptional process the *TAT* binds to a specific site (trans-acting response, *TAR*) of the R region in the 5′ LTR. As a result large amount of three different types of viral RNAs are formed. Small, spliced mRNAs which codes for *Rev, TAT* and *Nef*. The partially spliced mRNAs which codes for *Env, VIF, VPR* and *VPU* proteins and the un-spliced mRNAs which acts as *Gag, Pol* and *Gag* polyprotein precursors (Gatignol, 2007).

The *TAT* proteins holds a very important role in the viral transcription RNA elongation, the unspliced viral RNA is transported to the cytoplasm by *Rev* while the *Nef* proteins performs several functions including the one which makes the infected cell less visible to the host's immune system. The *TAT* and *Rev* permits the formation of full length mRNAs which encodes the formation of the Gag and Gag precursors that are processed for the synthesis of major structural and enzymatic proteins (Jeang et al., 1999).

2.6.5. Viral Particle Production

After the formation of different viral proteins assembly, the formation of the HIV virus is initiated. The main performer of the viral assembly is the *Gag* precursor polyprotein. The end domain of the *Gag* (*p6* part) and the cellular *Tsg101* protein are involved in this step, they have elements which targets the viral particles to the cell membrane, support *Gag-Gag* interactions, promote budding from the cell, and encapsulate the HIV genome for association with the *Env* viral glycoproteins (Von Schwedler et al., 2003).

2.6.6. Maturation

Once the viruses are released in immature (non-infectious form) out of the host plasma membrane, they are morphologically characterized by a thick layer of *Gag* and *Gag-polyprotein* precursors. Shortly, the proteases of the virus activates after budding and this cleaves the *Gag* and *Gag-Pol* precursors and the maturation of their final components occurs. As a result, the morphology of the virus converted to a doughnut shape (Hikichi & Freed, 2021).

2.7. Transmission of HIV

The HIV transmission depends on the biological properties, levels/concentration of the virus in the exposed body fluid, and the nature of the host vulnerability both at the cellular and immunological levels. The initial infection by HIV can be done even by a single virus which means that a single virus can be window of opportunity for AIDS. HIV is transmitted through sexual contact, blood transfusion, disclosure of infected blood, it can also be transmitted from mother to her child, organ transplants and through intravenous drug uses (Rothenberg et al., 1998). The transmission of HIV through sexual contact is about 75% to 80% of all HIV infections worldwide. Many diseases which are transferred through sexual contact can enhance the risk of HIV because huge amount of CD4

cells are present in the genital areas in sexually transmitted infections. The chance of HIV is higher when an uninfected female has a sexual intercourse with infected male (Royce et al., 1997).

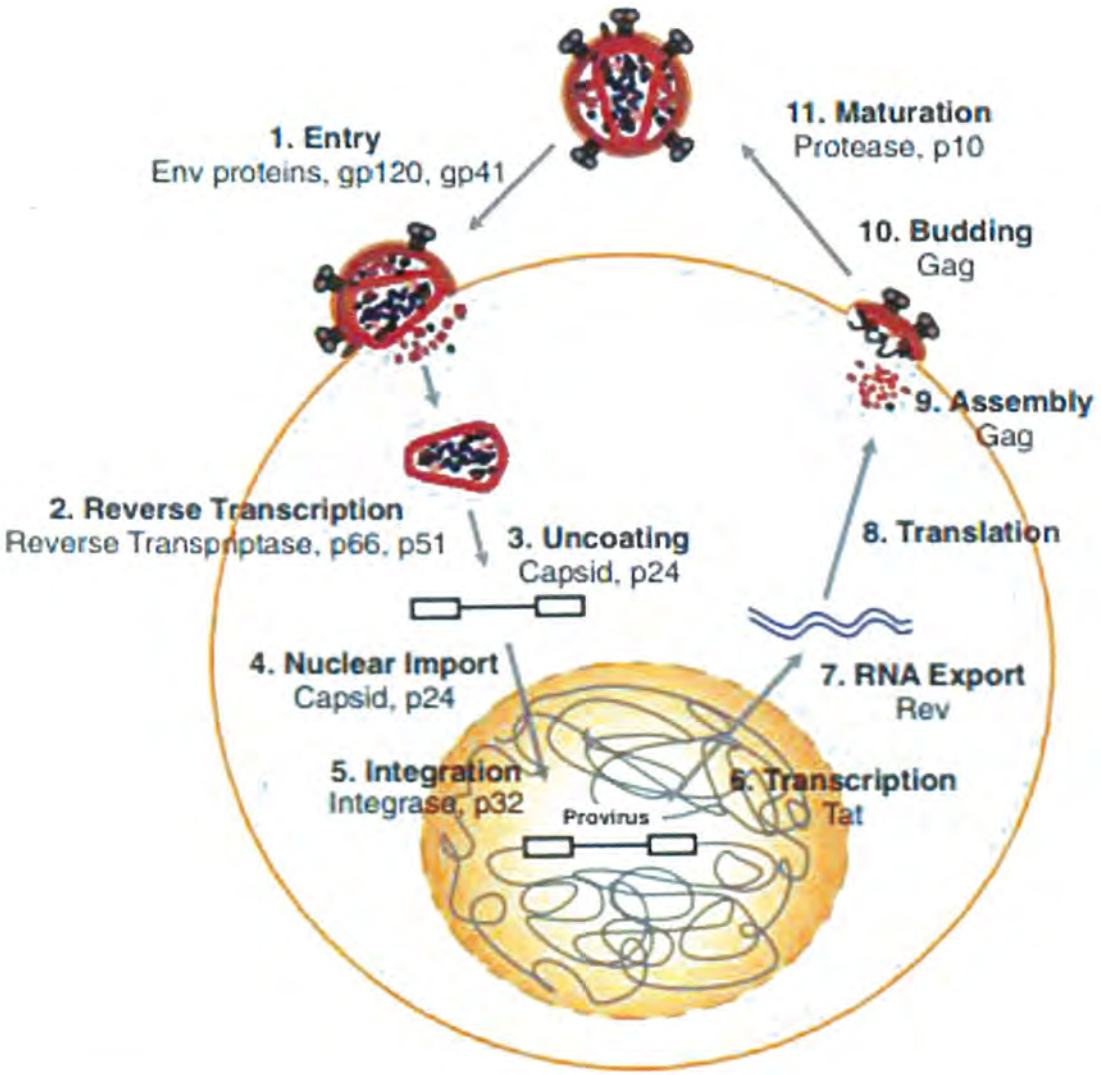


Figure 2.3. Overview of the viral replication cycle (Kirchhoff, 2013).

2.8. Pathogenesis

The two major factors that promotes the HIV pathogenesis are cellular tropism that regulates viral phenotype and viral entry associated to receptors and co-receptors (Staszewski et al., 1999). The initial three months of HIV infection is asymptomatic with no sign of illness along with slight changes in the host immune system. Later on, due to sero-conversion specific antibodies for HIV can be detected in individuals. The progression of disease from initial infection is slow, it takes a few years for primary infection to develop into immunosuppression and advanced HIV (Wodarz & Nowak, 2002).

The primary phase of the HIV pathogenesis can be summarized in three steps: the attachment to the host cell, fusion into the host cells and nucleocapsid entry, each of these depends on the type of virus and entry cells involved (Krakauer & Nowak, 1999). During this phase the individual looks healthy physically and the HIV virus replicates actively in their blood stream and lymph nodes. As a result, if viral load burst in their body the infected individual suffers from progressive damage to the immune system. Within days to months this viral load may decrease and can remain low for months or years and causing susceptibility to the opportunistic pathogens or infections in late phases, after that a rise in viral load occurs and the symptoms starts with onset of AIDS. If the viral load in plasma is higher and the CD4+T count is lower than 200 mm in infected individuals they could be declared as AIDS patients (Saez-Cirion & Manel, 2018).

The HIV weakens the immune system by two basic mechanisms, one is by destroying the CD4 and T cells which ultimately leads to immunodeficiency. The virus attaches to the CD4+ T cells through its *gp41* and *gp120* proteins (surface proteins) to gain entry into the host cell. The second way through which HIV gain access to cells is through the chemokine receptor which is an essential co-receptor for HIV-1 (Goto et al., 1998). Two main chemokine receptors that can be used by virus to enter the cell are *CCR5* and *CXCR4* or Fusin (Speck et al., 1998).

Based on cellular tropism, the HIV strains can be subdivided into three main groups, that are referred to as macrophage-tropism (M-tropic) and are known as *R5* viruses with a non-syncytium-inducing phenotype that infect *PBMC* (peripheral blood mononuclear cells), monocytes macrophages, and T lymphocytes through *CCR5* and not T-cell lines, these are present through all the infection stages (Naif, 2013). The second one is T-cell line tropism (T-tropic) which are known as *X4* viruses that induces syncytium phenotype and infect T lymphocytes and T-cell lines by using

CXCR4 as their principle co-receptor, but these cannot infect monocytes and macrophages, and this take time for AIDS progression. Individuals who do not have *CCR5* expression are actually resistant to *R5* viral infection but they are susceptible to *X4* viruses (Stevenson, 2003). The third type of strain is dual tropic HIV that utilizes the *CCR5* and *CXCR4 (R5X4)* by which both monocytes/ macrophages and T cell lines are infected, therefore these have populations of both NS1 and S1 phenotypes. Most of the cells like blood and tissue dendritic cells, macrophages, and T lymphocytes show co-expression of *CCR5, CXCR4*, and chemokine receptors. Hence, they are highly susceptible to HIV (Lackner et al., 2012). Some of the other chemokine co-receptors can also act as sites of attachment (primary or secondary) for both HIV-1 and HIV-2 isolates *like CCR2b* and *CCR3* but commonly they are not involved in infection. Furthermore, besides CD4 Galactosyl ceramide (*GalC*) can also be a major binding site for HIV-1 infection, in brain, vagina and bowel areas (Levy, 2009). Moreover, if the HIV makes a complex with antibodies they can easily gain access into the T cells, macrophages and other immune cells through complement receptors and *Fc* domains (Coffin & Swanstrom, 2013).

The major reservoirs of HIV-1 are the macrophages which can be found in seropositive individual's brain, lymph, lungs, skin, blood, nodes and bone marrow. These are the main agents of productive infection in brain and in infected individuals the large number of viral particles are found in their lymph nodes as these are the main production factory of the immune components (Derdeyn & Silvestri, 2005).

As the acute HIV infection progresses to AIDS the immune system starts to deregulate, and individual becomes more susceptible to opportunistic infections. This is supplemented by the rise in infected macrophages and CD4+T cells. Although *HAART* (highly active antiretroviral therapy) is known for minimizing the opportunistic infections as immune reconstitution occurs but *HAART* is not universally easily available nor fully successful (Yeni, 2006). Opportunistic infections cause the up regulation of viral co-receptors, activation of *NF* and the production of *TNF-alpha* which leads to the activation of HIV replication in infected persons. Along with this T cells apoptosis also act as an antiviral strategy to get rid of infected lymphoid cells which results in the failure of encountering the persistent HIV infection in tissue macrophages and monocytes (Montoya et al., 2006).

2.9. HIV variability

The most influential weapon of HIV is its variability through which the virus overcomes the host's immune response, effects of drugs and vaccines. The variability of HIV is the result of 3 unusual features:

1) The error prone mechanism of reverse transcriptase enzyme. Its reverse transcriptase enzyme comprises of multitasking abilities that is DNA-dependent polymerase activities and RNA-dependent polymerase activities. It also performs the RNAase-H activity that precisely degrades the RNA strand of hybrids of DNA/RNA and after each replication round there is one substitution per HIV genome. The nucleotide substitutions rate presented by reverse transcriptase is approximately 10^{-4} per nucleotide per cycle of replication.

2) The replication mechanism is very fast it has been reported that the HIV-1 has a fast turnover of approximately 10^{9} virions each day being generated in an infected individual.

3) The occurrence of recombination between 2 or more HIV. The process strongly contributes to increased levels of multiple drug resistance. Recombination by intermolecular and intramolecular jumps between the two single strands of the HIV marks a major issue of multi-drug resistance (Santoro & Perno, 2013).

2.10. Antiretroviral therapies (ART)

For HIV treatment there are 24 ARV (antiretroviral) drugs which are in six classes. In these six classes, nucleoside and nucleotide reverse transcriptase inhibitors (*NNRTIs*) are Abacavir, Di-danosine, Emtricitabine, Lamivudine, Stavudine, Tenofovir alafenamide, Tenofovir disoproxil fumarate, Zidovudine. One CC chemokine receptors 5 (*CCR5*) antagonist Maraviroc, one fusion inhibitor Enfuvirtide and integrase inhibitors Bictegravir, Cabotegravir and rilpivirine, Cabotegravir, Dolutegravir, Elvitegravir, Raltegravir (McManus et al., 2019). These drugs prevent HIV in different stages of its replication. Current treatment is *HAART* (highly active antiretroviral therapy) which is including three drugs to two classes of antiretroviral agents (McManus et al., 2019).

The mutation rate of HIV is very high as per replication cycle nearly one nucleotide mutation has reported. However, people are mostly infected with one or few original cones. Each day about 10 viruses are formed in untreated patients, which results in unlimited virus variants, known as quasi-

species. The capacity to form new variants permits the HIV-1 evades the host's immune system and encourages the development of ARV drug resistance (Sigal & Baltimore, 2012).

2.11. Vaccine development

Due to the incredible capacity of HIV to escape immune burden and number of strain variations, vaccine development has been very slow. Thousands of healthy humans contributed voluntarily to the testing of 30 candidate vaccines in 1987 (WHO). The degree of trouble in the formation of an effective vaccine for AIDS is still very challenging.

CHAPTER 3

MATERIAL AND METHODS

3. Materials and Methods

3.1. Sample Collection

In this study, serum samples from Human Immunodeficiency Virus (HIV) positive patients were acquired from hospitals and HIV/AIDS centers at Islamabad, Lahore, Peshawar, and Kohat. A total of 18 HIV positive individuals serum were collected, of which 4 were children under the age of 15 years (1 male and 3 females), 2 were adult females, and 12 were adult males.

3.2. Inclusion Criteria

HIV positive samples confirmed by PCR were included. Patients with chronic HIV infection, some naïve and some on different antiviral therapies.

3.3. Exclusion Criteria

Those patients who were negative for HIV-RNA were excluded from the study.

3.4. Primers designing for Gag gene of HIV

We synthesized previously reported primers (Tcherepanova et al., 2008) for sequence amplification. The primers sets were confirmed through Primer3 (http://bioinfo.ut.ee/primer3/) which is an online tool; and were further validated/confirmed using NCBI primer designing online tool (Primer Blast) (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Primers were then modified using recently reported nucleotide sequences. For this purpose, Human Immunodeficiency Virus HIV-1 sequences reported from Pakistan (particularly Accession no. KX232594; isolate HIV-1_PK001) was considered as a reference sequences. Name, oligonucleotide sequence, and product size of the primers having maximum *in-silico* specificity are listed below (Table 3.2).

3.5. RNA Extraction

HIV RNA extraction was performed by using automatic extractor TANBead Nucleic Acid Extractor, Model no. SLA-32 (Taiwan). The TANBead auto extractor plate was used for loading of samples into the auto extractor. Carefully removed the aluminum foil on the Auto Plate and 300 µl HIV positive serum was added in to the columns of Auto Plate under the BSL-2/B2 biosafety cabinet (Esco Life sciences). As the volume ratio of serum and lysis buffer mixture was made sure

to be maintained as 300 µl: 600 µl. After that the Auto Plate was placed to the plate holder of auto stage in the extractor. It was made sure that the missing corner of base faced toward the lower left. The program was set to Virus-auto a4 and the extraction process was initiated. When the extraction process completed, the plate was removed from the Auto Plate holder carefully and micropipette was used to transfer the extracted nucleic acid from column to autoclaved eppendorf tubes. The extracted product was immediately processed for cDNA synthesis or were stored immediately at -80°C before processing it for cDNA synthesis.

3.6. HIV complementary DNA (cDNA) synthesis

Complementary DNA (cDNA) were synthesized using the extracted RNA as a template. First strand cDNA Synthesis Kit (RevertAid, ThermoScientific) was used. The components for cDNA synthesis were as follows: 8µl of extracted RNA along with 2µl Random hexamer or gene specific reverse primers, 2µl double distilled water, 2µl dNTP (10 mM), 4µl 5x reaction buffer, 1µl RevertAid RT enzyme (200 U/µL) and 1µl RiboLock (20U/µL) RNase Inhibitor were mixed gently and total reaction volume adjusted to 20 µl as shown in Table 3.2. All the components of reaction mixture were mixed and incubated for 5min on 25°C, followed by 42°C for 60 min, 45 °C for 30 minutes and finally reaction was completed at 70°C for 5 mins.

Table 3.1. Conditions optimized for HIV complementary DNA (cDNA) synthesis.

Serial no	Reagents	Quantity
1	5x reaction buffer	4µl
2	Random hexamer primers/ R. Primer	2µl
3	10 mM dNTP Mix	2µl
4	Template RNA	8µl
5	ddH_2O	2µl
6	RiboLock (20U/µL)	1µl
7	RevertAid RT enzyme (200 U/µL)	1µl
	Total volume	**20µl**

3.7. PCR amplification of HIV Gag gene

For HIV Gag gene amplification nested PCR was performed and PCR Master Mix (2X Phusion High-Fidelity, ThermoScientific) was used. The reaction mixture of PCR first round contained 2X Phusion High Fidelity Master Mix 10µl, external Gag forward primer (10µM) 1.5µl, external Gag reversed primer (10µM) 1.5µl, cDNA 6µl and double distilled water 1µl which makes a total of 20µl reaction as given in table 3.3. The cyclic condition for PCR were as follows: 96°C for 5mins following 35 cycles of 96°C for 45 seconds, 51°C for 30 seconds, 72°C for 45 seconds and final extension at 72°C for 10 min as shown in Figure 3.1. The reaction was hold at 4°C for infinity.

The second round of PCR was performed by using nested primers and the template used was the first round product. Second round PCR contained the same reaction reagents as in first round with slight modification such as internal Gag forward and reverse primers and 3µl first round PCR product were used as template and 4µl ddH_2O was added as given in Table 3.3. Moreover, the PCR cyclic conditions with changes in annealing temperature to 50 °C as presented in Figure 3.2. The total reaction volume for both rounds were 20µl.

The primers specifically targeted and amplified approximately 1500 nucleotides long fragment. The amplified product was confirmed through gel electrophoresis as represented in figure. 4.1.

Table 3.2. Primer Name, oligonucleotide sequence, and PCR product size for Gag gene amplification of HIV-1

S. no	Primer name	Oligonucleotide seq.	PCR Products Size	
1	Gag	F. 5′-AATTTTGACTAGCGGAGGC-3′	1st round ~1600bp	(Tcherepanova et al., 2008)
2		R. 5′-TTTGGTTTCCATCTTCCTGG-3′		
3	Gag	*F. 5′-AGATGGGTGCGAGASCGT-3′	2nd round ~1500bp	(Tcherepanova et al., 2008)
4		R. 5′-GCTCCTGTATCTAATAGAGC-3′		

* Modified after using recently reported reference sequences

Table 3.3. PCR 1st round reaction mix.

Serial no.	Reagents	Amount
1	2X Phusion HF Master Mix	10µl
2	External Gag Forward Primer	1.5µl
3	External Gag Reverse Primer	1.5µl
4	Template (cDNA)	6µl
5	Double distilled water	1µl
	Final volume	**20µl**

Table 3.4. PCR 2nd round reaction mix.

Serial no.	Reagents	Amount
1	2X Phusion HF Master Mix	10µl
2	Internal Gag Forward Primer	1.5µl
3	Internal Gag Reverse Primer	1.5µl
4	Template (1^{st} round product)	3µl
5	ddH_2O	4µl
	Final volume	**20µl**

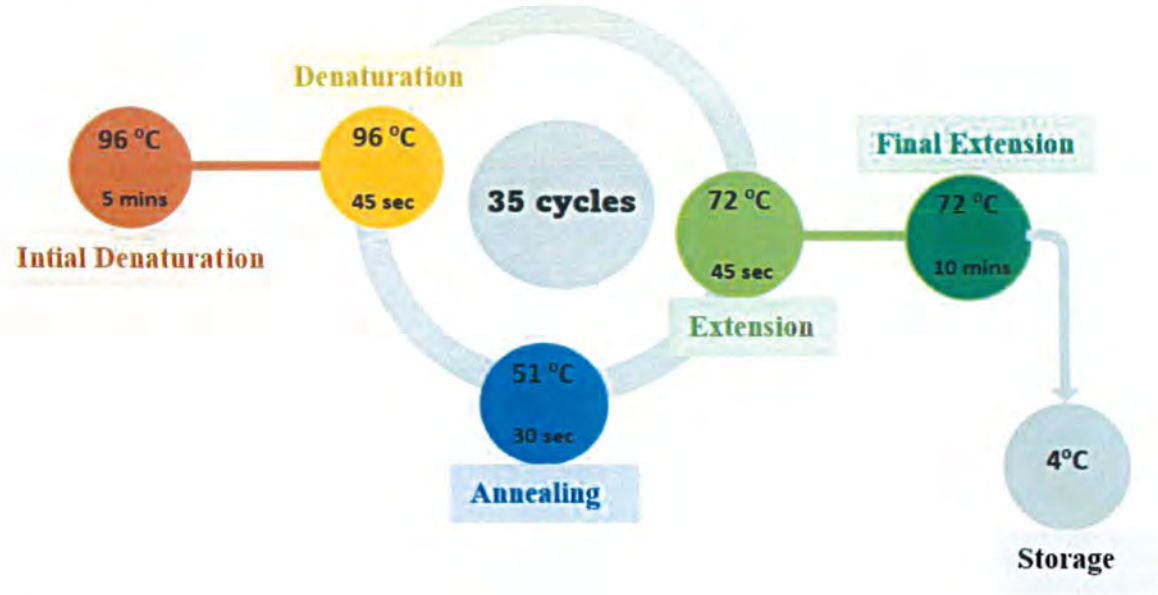


Figure 3.1. Cyclic conditions for 1st round of PCR. The conditions were optimized using gradient PCR

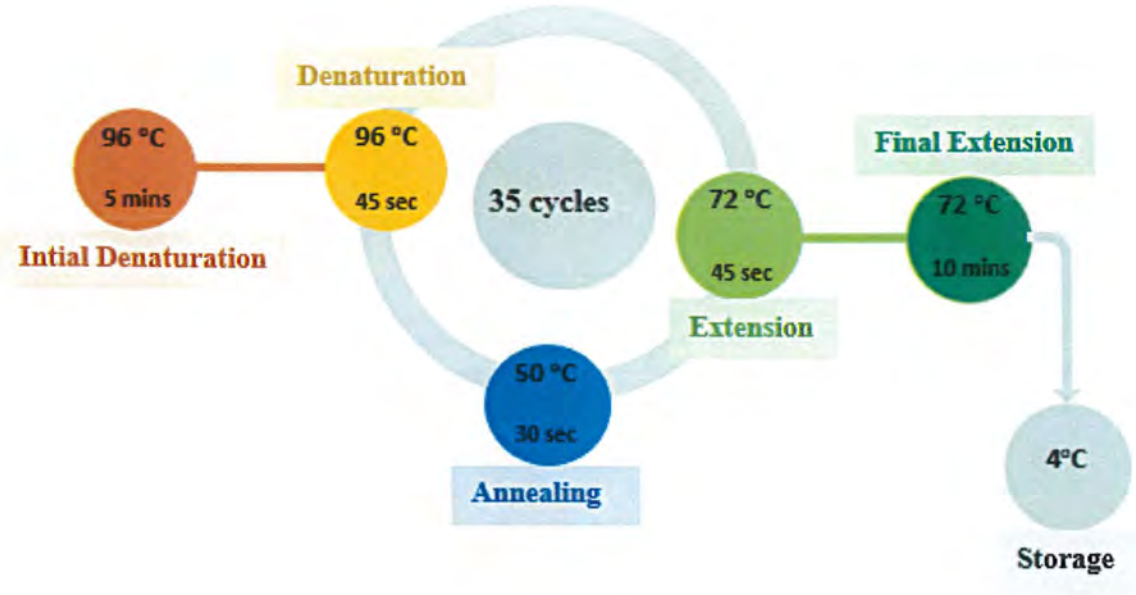


Figure 3.2. Cyclic conditions for 2nd round of PCR. The conditions were optimized using gradient PCR

3.8. Detection of PCR products by gel electrophoresis

For confirmation of amplified PCR products they were run on 1.5% agarose gel. The gel was prepared in a 40ml (1X TAE) buffer by liquefying 0.6 g of agarose powder. The conical flask was then placed for 30 seconds in a microwave oven to heat the crude powder and then cooled down. 4µl Ethidium bromide was added to 1.5% agarose solution after cooling and poured off to gel casting tray which was already affixed with the appropriate comb and get until solidifications. 1X TAE buffer was also used as a running buffer in the gel tank. The nested PCR product was blended with a 6X loading dye and loaded into the wells of the gel. 100bp DNA gene ruler (Thermo scientific) was loaded in a separate well for comparison of the size of amplified PCR products. After that the gel was run at 90V for 30mins, once the gel electrophoresis process completes gel was observed in UV trans-illuminator for visualization of amplified DNA.

Table 3.5. Components for gel electrophoresis procedure

S. No.	Components (1.5% gel)	Amount
01.	1 X TAE buffer	40 ml
02.	Agarose	0.6 g
03.	Ethidium bromide solution	4 µl

3.9.DNA extraction from agarose gel

The amplified fragment (excised from the gel) was purified using Gene JET Gel Extraction Kit (ThermoScientific). The target DNA gel fragment was excised with a smooth razor blade and put in a pre-weighted autoclaved microfuge tube. Digital balance was used weighing the excised gel. After that the binding buffer was added to the tube as 1:1 ratio (weight:volume) i.e., 50µl Binding buffer was added to 50mg of gel slice and heat up the gel slice and binding buffer mixture at 60°C for 10 min using heat block. Then transferred upto 800µl solubilized gel solution to the column and centrifuge for 1min on 12000 rpm. After discarding the flow-through, 100µl more binding buffer was supplemented for sequencing purposes to the column. The mixture was centrifuged again on 12000 rpm for 60 seconds and discarded the flow-through. 700µl Wash Buffer was added and centrifuged for 1min on 12000 rpm and discard flow-through. The empty columns were then centrifuged for 1min on 12000 rpm to eliminate remaining wash buffer. The column was subsequently transferred to a sterile microfuge tube and 35µl Elution buffer was added. This was centrifuged on 12000 rpm for 1min, column removed, and purified DNA fragment stored for further processing at -20 °C. For confirmation, 5µl of obtained DNA was again run-on agarose gel (1.5%) and visualized via UV trans-illuminator (Figure 3.3).

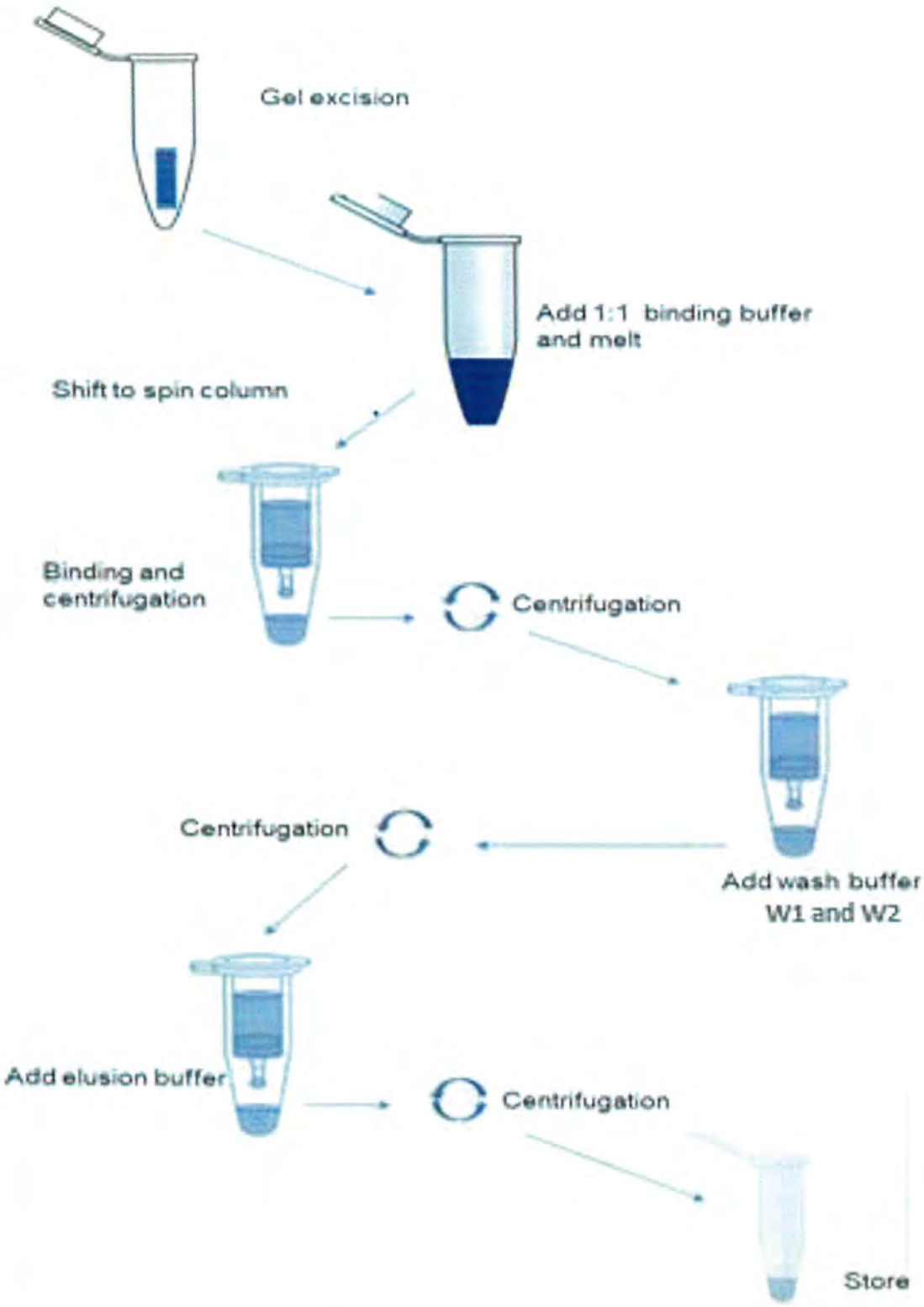


Figure 3.3. Gel purification of the amplified PCR product

3.10. Sequencing of HIV Gag gene

Sanger sequencing of the gel purified PCR product was done with Gag specific internal primers separately (forward and reverse). The sequencing results analysis were done through a sequence alignment software called BioEdit. Nucleotide sequences obtained were in the form of "FASTA" format for further analysis.

3.11. Phylogenetic analysis

For finding similarities of the current sequences of Gag gene with other reported sequences of HIV-1 from Pakistan and other regions of the world, nucleotide BLAST of the sequences was done. Available nucleotide sequences of Gag gene of HIV-1 were saved from the Genbank NCBI. The saved sequences were then aligned using ClustalW of BioEdit. The tree was constructed from sequences of the current study and from other 50 retrieved sequences of Gag-HIV-1. The evolutionary history was concluded by means of the neighbor joining method and Jukes-Cantor Model (Abecasis et al., 2018). All the evolutionary history analyses were conducted in MEGA 11 (Tamura et al., 2013) .

3.12. Translation of nucleotide sequences

Nucleotides sequences of Gag obtained in the form of "FASTA" format were translated into their respective protein sequences through the help of an online tool. For execution of this job, ExPASy Translate tool was used (Artimo et al., 2012).

3.13. Alignment of HIV

The nucleotide sequences of the samples were then translated to amino acid sequences using online server ExPASy (https://web.expasy.org/translate/). The alignment of amino acid sequences was executed using CLC (Yamaguchi, 2018). The sequences aligned with a reference protein sequence of Human Immunodeficiency virus (HIV-1 isolate_PK002, accession no. KW252395). Amino acid substitutions were then analyzed for positions that have been described in the previous literature.

3.14. Conserved regions identification

Conserved regions were identified using BioEdit by aligning both the isolate protein sequences with reference protein sequences of HIV-1 from different regions.

3.15. Prediction of Secondary structures

To predict secondary structure of HIV-Gag SOPMA was used (Geourjon & Deleage, 1995).

3.16. Protein Structure Modeling and Visualization

For tertiary structure modeling Robbeta server (http://robetta.bakerlab.org) was used (Kim, 2004). Robetta server is based on the comparative modeling technique to generate structures. The model structure of the given amino acid is formed by using PSI BLAST, FFAS03 or 3D-Jury. The model was obtained in the form of a "PDB file" for further analysis. For molecular visualization and structure comparison, PyMOL was used (Yang et al., 2012).

CHAPTER 4

RESULTS

4. Results

Total 25 HIV suspected samples were collected from Islamabad, Lahore, Peshawar, and Kohat hospitals and HIV/AIDS centers. Of the total 25 suspected samples, 18 were confirm HIV positive. The patients of the study were of different age ranging from 08-45 years, 4 were children under the age of 15 years (3 females and 1 male), 12 adult males and 2 adult females. Out of the total positive samples 5 samples were collected from Islamabad, 8 from Peshawar, 3 from Kohat and 2 from Lahore, Punjab.

4.1. Gel electrophoresis and Nucleotide Sequencing result

Gag gene of HIV-1 was amplified using nested PCR and the product of approximately 1500 base pairs was obtained using two sets of gene specific primers. These were external forward/reverse primers and internal forward/reverse primers. The agarose gel showing approximately 1500 bp of HIV-1 Gag gene is given in Figure 4.1.

Each sample after amplification was purified from excised gel using gel extraction kit and were further processed for Sangers sequencing. Chromas was used for refining sequences (Zajec, 1986). Two sequences of Gag HIV-1 were obtained in current study and further analysis were done through bioinformatics tools.

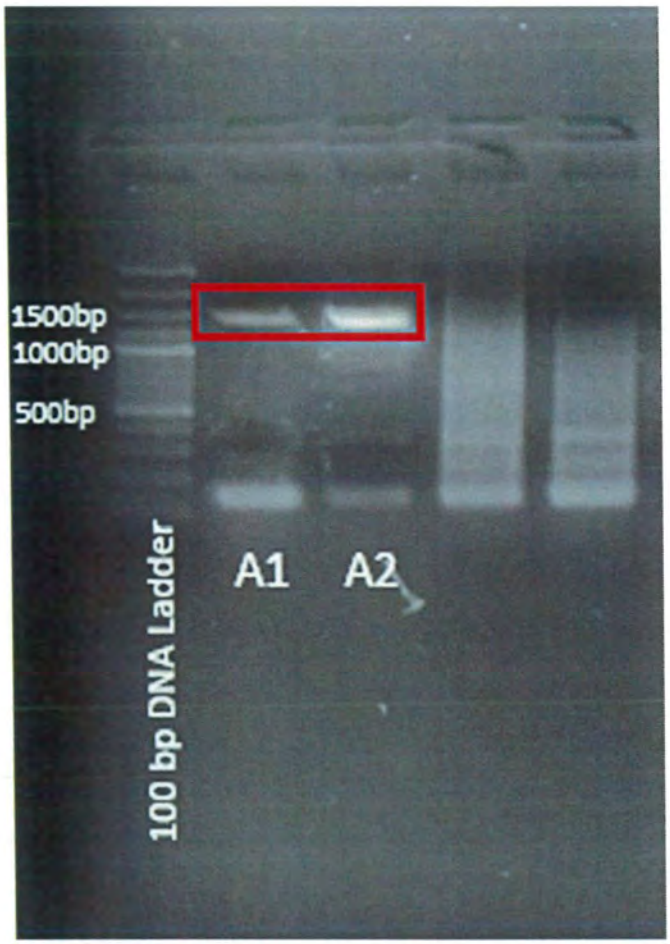


Figure 4.1. PCR Amplified products. Lane 1: 100-bp DNA gene ruler, Lane 2 (A1) and 3 (A2): Positive samples showing Gag HIV-1 specific bands of approximately ~ 1500-bp highlighted in red box.

4.2. Sequencing Results

Figure 4.2 Nucleotide sequencing results of Gag partial gene of HIV-1

4.3. Phylogenetic analysis

For understanding the genetic variability, sequencing of the mentioned gene of HIV-1 samples was performed. A total of 50 reference sequences of the aforementioned type of HIV have been acquired from NCBI database. Neighbor-Joining Method was used for finding the evolutionary history. The highest log-likelihood tree is displayed in the results. The percentage of trees is shown next to the branches where by the linked taxa grouped together. The evolutionary distances were computed by using Jukes-Cantor model. 52 nucleotide sequences were included in this valuation. Software package used for the evolutionary analysis was MEGA 11. Sequences of the current study (QAU-HIV1-AH1, QAU-HIV1-AH2) make a cluster with previously reported sequence from Pakistan (KX232595). The study demonstrates that HIV-1 Gag gene isolates clustered together with reference sequences from Pakistan and are phylogenetically distinct from those reported from African countries like Kenya, Uganda, and Zambia *etc.* which depicts that they might have been originated from above-mentioned sequence recently. They could be thought-out as emerging viral isolates. The taxa of current study viral isolates are labelled as red, while other isolates of HIV-1 Gag gene, from Pakistan and other regions are presented in black color as shown in the Figure 4.3.

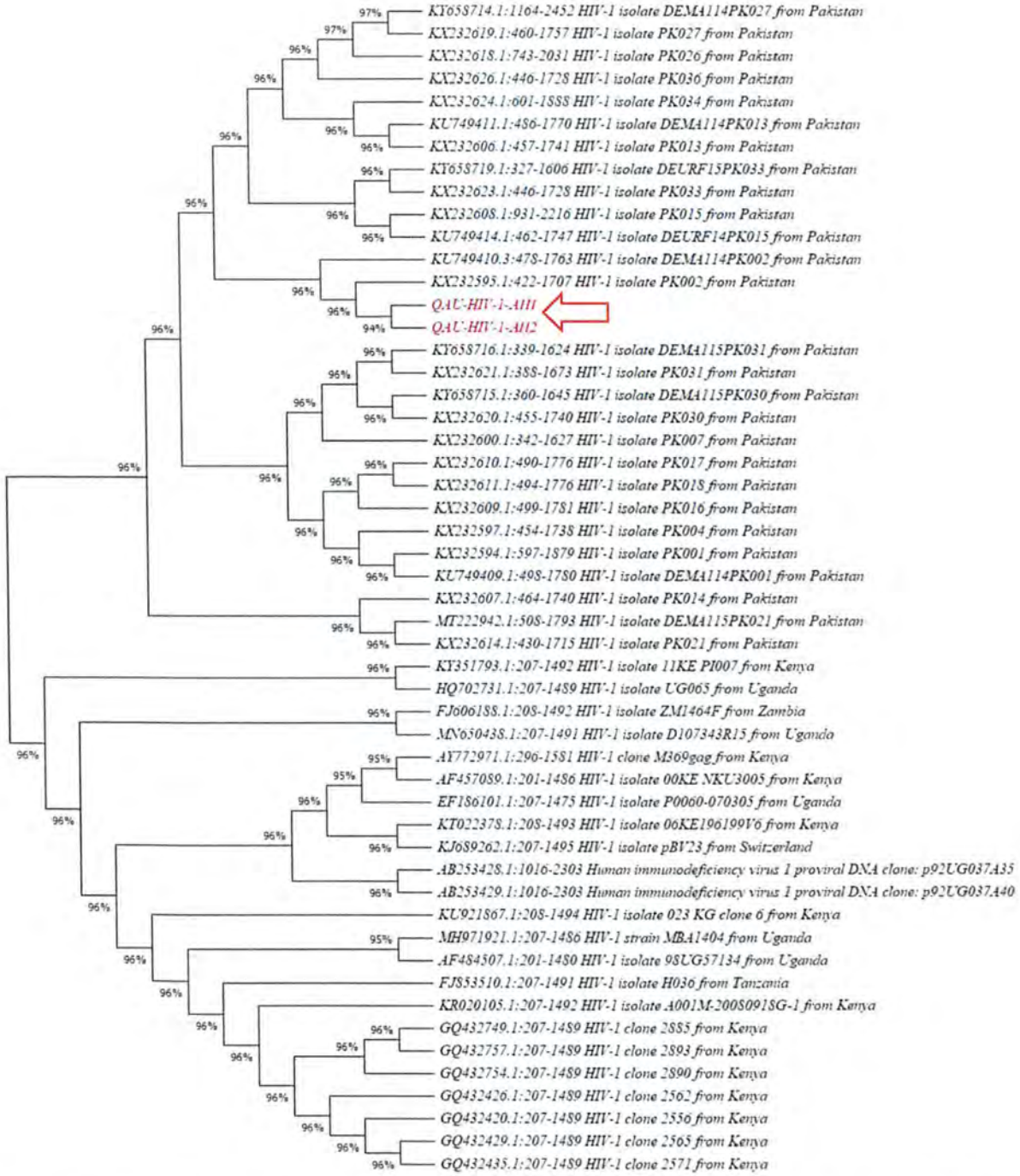


Figure 4.3. Phylogenetic analysis of HIV-1 isolates. Current study (isolates, n=2) with formerly reported Gag gene sequences from Pakistan and other parts of the world (reference sequences, n=50). The MEGA 11 package was used for construction of phylogenetic tree.

4.4. Amino acids substitutions in HIV-1 Gag.

Both the nucleotide sequences of the study were translated to amino acid sequences by using online server ExPASy, following the alignment of resultant amino acids sequences (using reference sequence from NCBI.GenBank) through BioEdit. The significant amino acid substitutions found in isolate QAU-HIV1-AH1, Gag protein are W212M, P328Q, N392S, E474K, D479T, and Q480R. On the other hand, amino acid substitutions found in isolate QAU-HIV1-AH2, Gag protein are L321M, L322Q, V323I, N325I, A326V, N327R, N374Y, K413Q and S477C position of the latterly analyzed query sequence (Figure 4.4, Table 4.1 and Table 4.2).

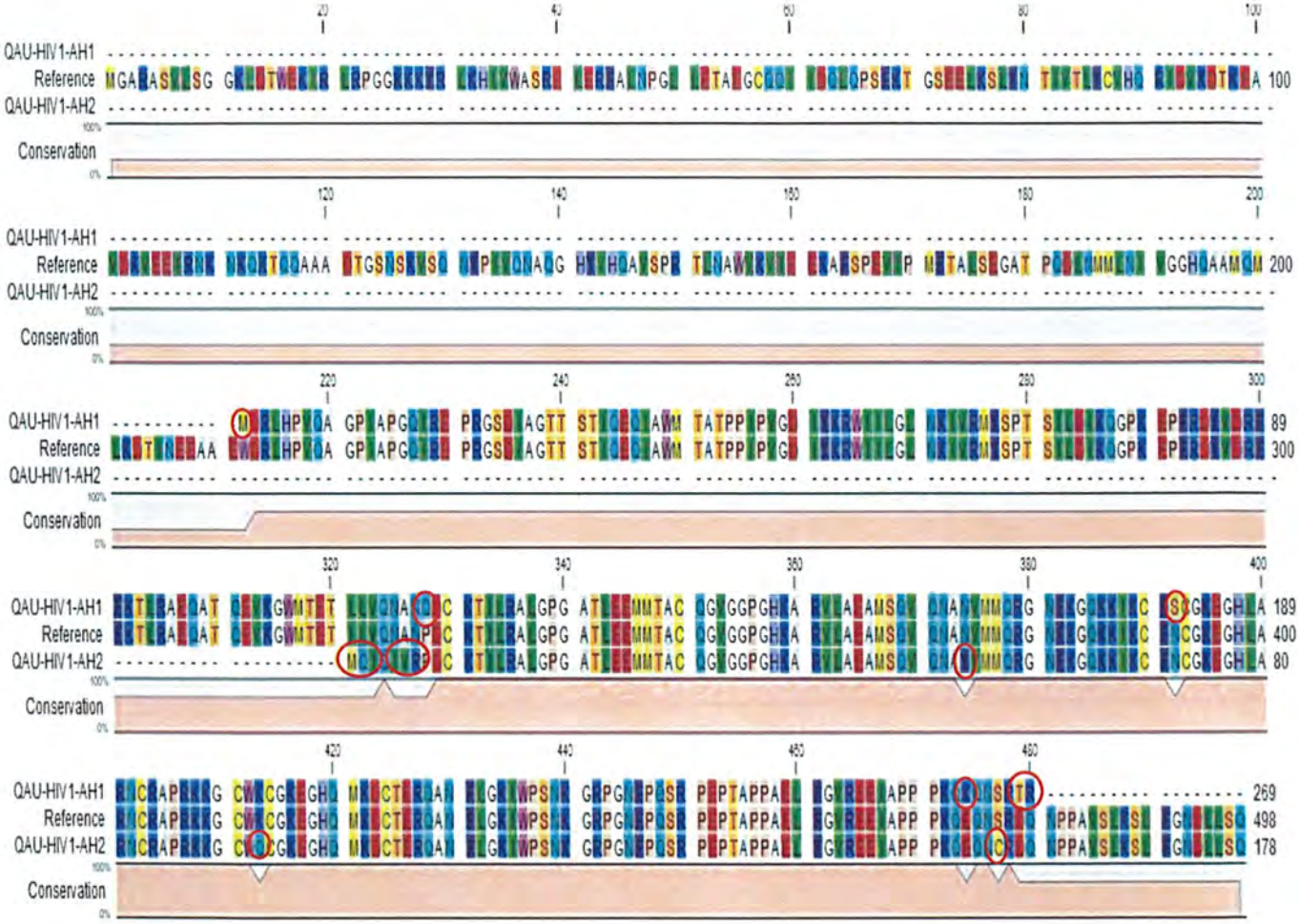


Figure 4.4. Alignment of amino acid sequences of the query sequences (HIV isolates QAU-HIV1-AH1 and QAU- HIV1-AH2) with the reference sequence of HIV-1 isolate_PK002, accession number KX232595.

Table 4.1. Amino acid substitutions present in Gag protein of Human Immunodeficiency Virus (HIV-1), (isolate QAU-HIV1-AH1).

Position	QAU-HIV1-AH1					
	Wild- type	Polarity	Charge	Variant	Polarity	Charge
212	W (Tryptophan)	Non-polar	Neutral	M (Methionine)	Non-polar	Neutral
328	P (Proline)	Non-polar	Neutral	Q (Glutamine)	Polar	Neutral
392	N (Asparagine)	Polar	Neutral	S (Serine)	Polar	Neutral
474	E (Glutamic acid)	Polar	Neutral	K (Lysine)	Polar	Positive
479	D (Aspartic acid)	Polar	Negative	T (Threonine)	Polar	Neutral
480	Q (Glutamine)	Polar	Neutral	R (Arginine)	Polar	Positive

Table 4.2. Amino acid substitutions present in Gag protein of Human Immunodeficiency Virus (HIV-1), (isolate QAU-HIV1-AH2).

Position	QAU-HIV1-AH2					
	Wild- type	Polarity	Charge	Variant	Polarity	Charge
321	L (Leucine)	Non-polar	Neutral	M (Methionine)	Non-polar	Neutral
322	L (Leucine)	Non-polar	Neutral	Q (Glutamine)	Polar	Neutral
323	V (Valine)	Non-polar	Neutral	I (Isoleucine)	Non-polar	Neutral
325	N (Asparagine)	Polar	Neutral	I (Isoleucine)	Non-polar	Neutral
326	A (Alanine)	Non-polar	Neutral	V (Valine)	Non-polar	Neutral
327	N (Asparagine)	Polar	Neutral	R (Arginine)	Polar	Positive
374	N (Asparagine)	Polar	Neutral	Y (Tyrosine)	Polar	Neutral
413	K (Lysine)	Polar	Positive	Q (Glutamine)	Polar	Neutral
477	S (Serine)	Polar	Neutral	C (Cystine)	Polar	Neutral

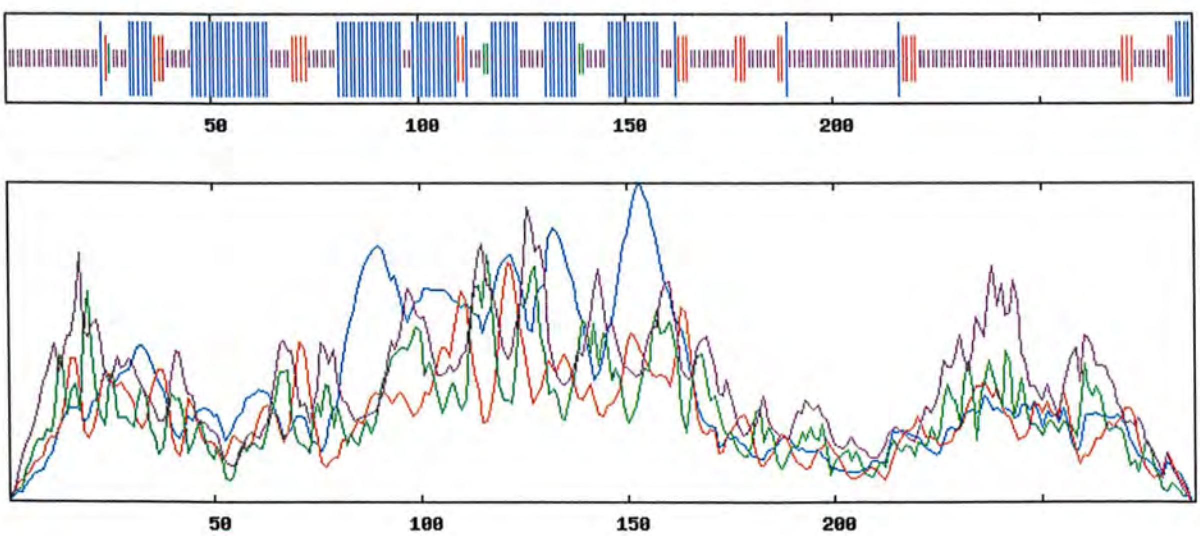
4.5. Secondary structure prediction

For the prediction of Secondary structure of HIV-Gag protein, the online server SOPMA (Geourjon, 1995) was used. HIV-1 isolate QAU-HIV1-AH1 has an alpha helix of 31.36%, extended strand of 9.41%, Beta turn of 1.74% and Random coils of 57.49%. While HIV isolate QAU-HIV1-AH2 has an alpha helix of 24.16%, extended strand of 7.30%, Beta turn of 5.06% and Random coils of 63.48% (Figure 4.5 and 4.6). It is supposed that conserved regions are existing which maintained the structure of the Gag.

```
        10        20        30        40        50        60        70
         |         |         |         |         |         |         |
MDRLHPVQAGPIAPGQIREPRGSDIAGTTSTIQEQIAWMTATPPIPVGDIYKRWIILGLNKIVRMYSPTS
hcccccccccccccccccccccccchetcccchhhhhheeecccccchhhhhhhhhhhhhhhhhhccccc
ILDIKQGPKEPFRDYVDRFFKTLRAEQATQEVKGWMTETLLVQNANQDCKTILRALGPGATLEEMMTACQ
eeeecccccccchhhhhhhhhhhhhhhhcchhhhhhhhhhheehccctthhhhhhhcccccchhhhhhhht
GVGGPGHKARVLAEAMSQVQNANVMMQRGNFKGQKKIKCFSCGKEGHLARNCRAPRKKGCWKCGKEGHQM
tccccchhhhhhhhhhhhhcccheeeccccccccccceeeccccccceehcccccccccccccccccccc
KDCTERQANFLGKIWPSNKGRPGNFPQSRPEPTAPPAELFGVREEIAPPPKQKQNSRTRTHLQFPSNHSL
cccccchheeeeccccccccccccccccccccccccccccccccccccccccccceeeccccccc
ATTSLSQ
ceehhhh
```

```
Sequence length :   287
```

```
SOPMA :
    Alpha helix        (Hh) :    90 is  31.36%
    3₁₀ helix          (Gg) :     0 is   0.00%
    Pi helix           (Ii) :     0 is   0.00%
    Beta bridge        (Bb) :     0 is   0.00%
    Extended strand    (Ee) :    27 is   9.41%
    Beta turn          (Tt) :     5 is   1.74%
    Bend region        (Ss) :     0 is   0.00%
    Random coil        (Cc) :   165 is  57.49%
    Ambiguous states   (?)  :     0 is   0.00%
    Other states            :     0 is   0.00%
```



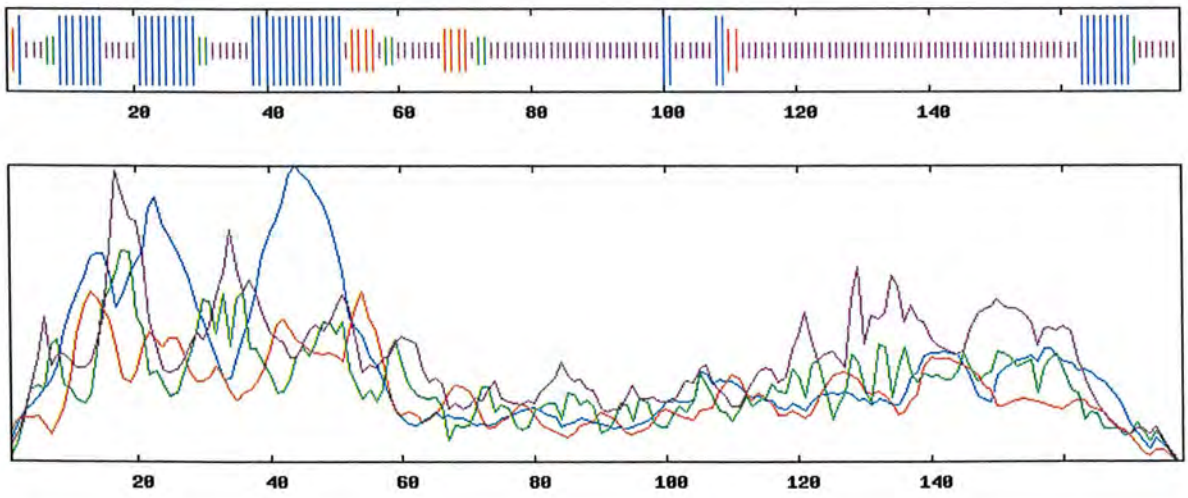

```
Parameters :
   Window width         :  17
   Similarity threshold :   8
   Number of states     :   4
```

Figure 4.5. Secondary structure of QAU-HIV1-AH1 Gag protein. Different motifs and parts of Gag *i.e.* Alpha helix, Beta turn, extended strand, and random coils are presented along with their percentage.

```
        10        20        30        40        50        60        70
         |         |         |         |         |         |         |
MQIQIVRPDCKTILRALGPGATLEEMMTACQGVGGPGHKARVLAEAMSQVQNAYVMMQRGNFKGQKKIKC
eeehccctthhhhhhhccccchhhhhhhhhttccccccchhhhhhhhhhhhhhhceeeeecttcccccccccee
FNCGKEGHLARNCRAPRKKGCWQCGKEGHQMKDCTERQANFLGKIWPSNKGRPGNFPQSRPEPTAPPAEL
ecttcccccccccccccccccccccccccchhccccchheecccccccccccccccccccccccccccc
FGVREEIAPPPKQEQNCRDQNPPAVSLKSLFGNDLLSQ
ccccccccccccccccccccccchhhhhhhhhtcccccc
```

```
Sequence length :   178
```

```
SOPMA :
   Alpha helix      (Hh) :    43 is  24.16%
   3₁₀ helix        (Gg) :     0 is   0.00%
   Pi helix         (Ii) :     0 is   0.00%
   Beta bridge      (Bb) :     0 is   0.00%
   Extended strand  (Ee) :    13 is   7.30%
   Beta turn        (Tt) :     9 is   5.06%
   Bend region      (Ss) :     0 is   0.00%
   Random coil      (Cc) :   113 is  63.48%
   Ambiguous states (?)  :     0 is   0.00%
   Other states          :     0 is   0.00%
```




```
Parameters :
   Window width         :  17
   Similarity threshold :   8
   Number of states     :   4
```

Figure 4.6. Secondary structure of QAU-HIV1-AH2 Gag protein. Different motifs and parts of Gag *i.e.* Alpha helix, Beta turn, extended strand, and random coils are presented along with their percentage.

4.6. Protein modeling and validation

3D proteins structures were generated by Robetta for two amino sequences of isolate QAU-HIV1-AH1 and isolate QAU-HIV1-AH2. Structural comparison of the variants with the reference protein sequence was made possible through visualization of protein models. The assessment of substitutions were made easier and clear through structure predication (Figure 4.7, 4.8). The substitutions are highlighted in amino acid sequences of both variants with respect to their sites compared to reference sequences. Moreover, the conserved sites are also highlighted in sequences with respect to the reference sequences. PyMOL tool was used for structural comparison of both the variants with the reference sequence.

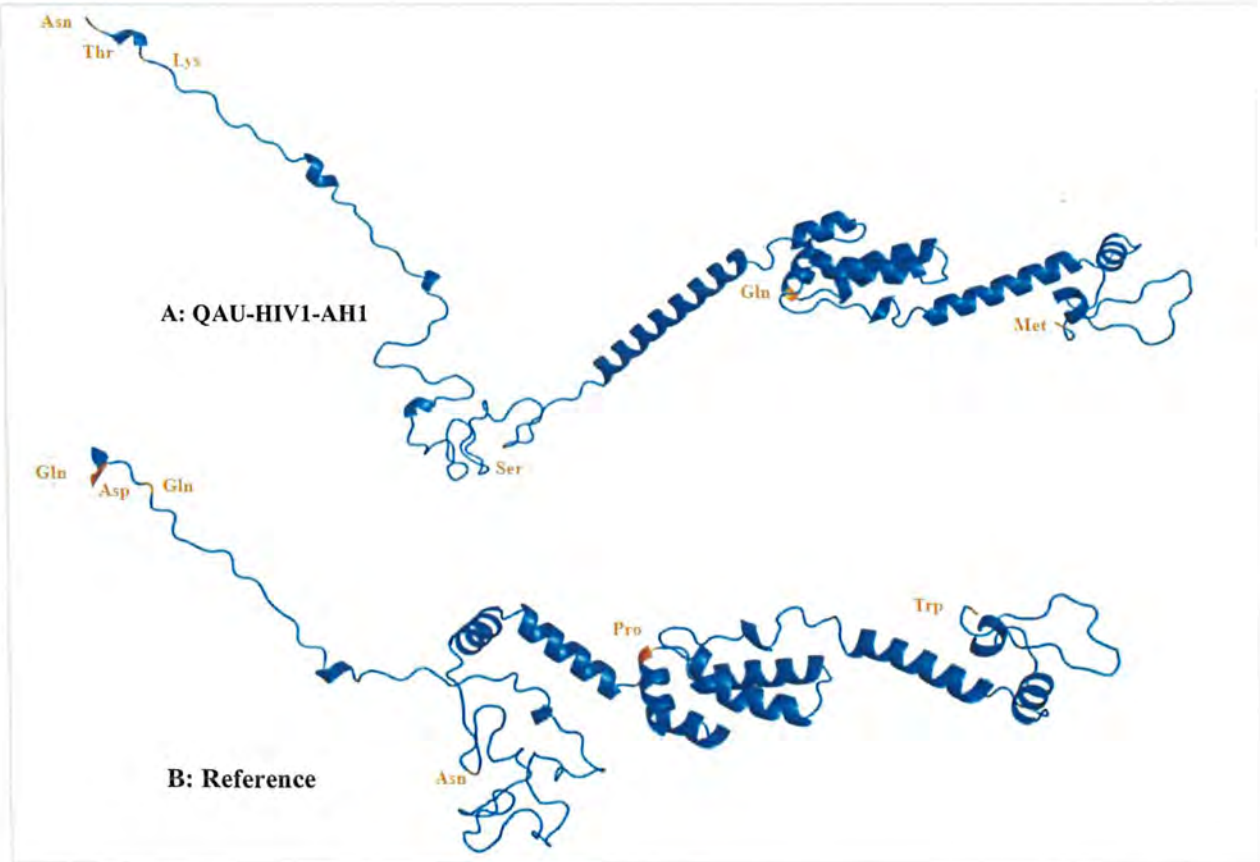


Figure 4.7. A). 3D structure (3D) of HIV-1 Gag protein of isolate QAU-HIV1-AH1. B). 3D structure of HIV-1 Gag protein of Reference isolate.

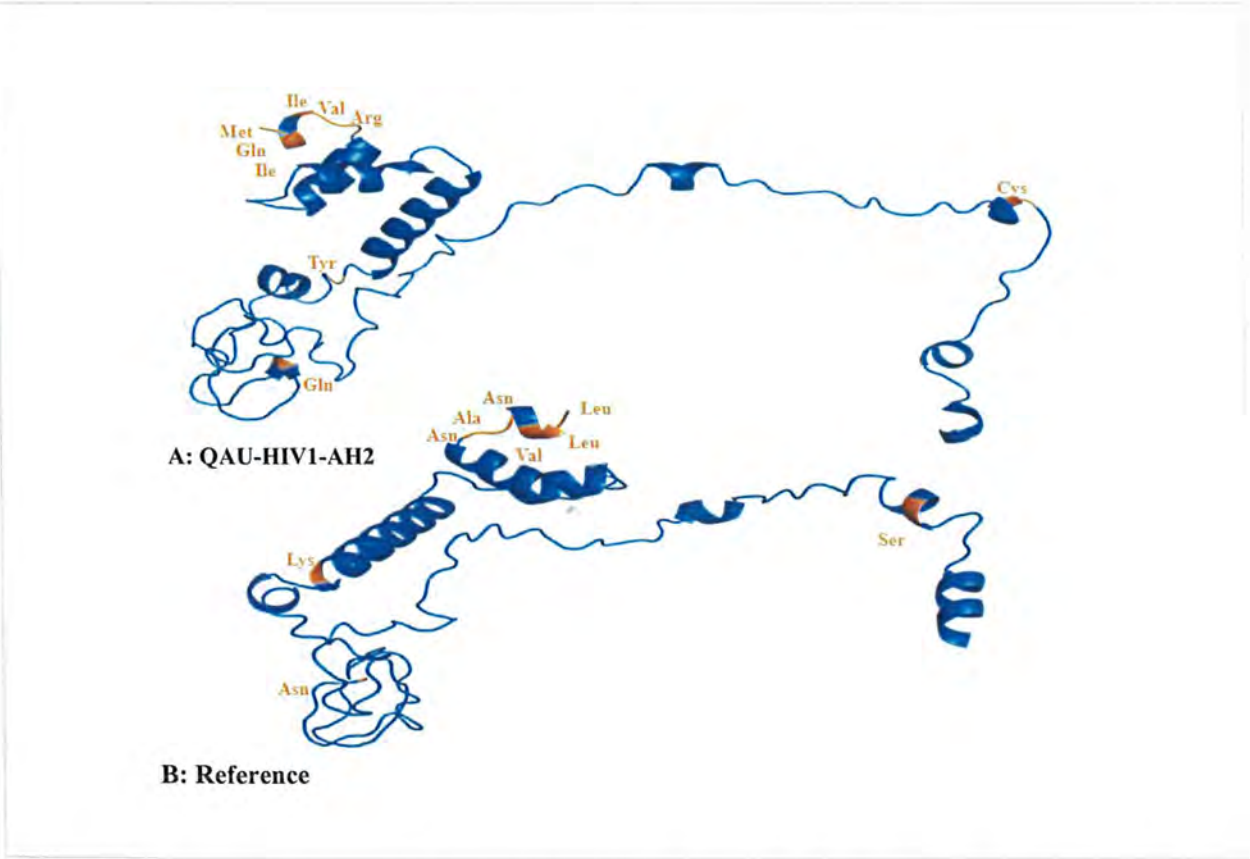


Figure 4.8. A). 3D structure (3D) of HIV-1 Gag protein of isolate QAU-HIV1-AH2. B). 3D structure of HIV-1 Gag protein of Reference isolate.

4.7. Conserved regions

Conserved regions were identified using BioEdit by aligning both the isolate protein sequences with thirty reference protein sequences of HIV-1 from different regions. Two conserved regions were identified in isolate QAU-HIV1-AH1. Region 1 consensus sequence was identified as 49-*IYKRWIILGLNKIVRMYSP*-67 with segment length of 19. The region 2 consensus sequence was identified as 75-*QGPKEPFRDYVDRFFK*-90 with the segment length of 16 shown in figure 4.9. While the isolate QAU-HIV1-AH2 showed no conserved regions with reference sequences.

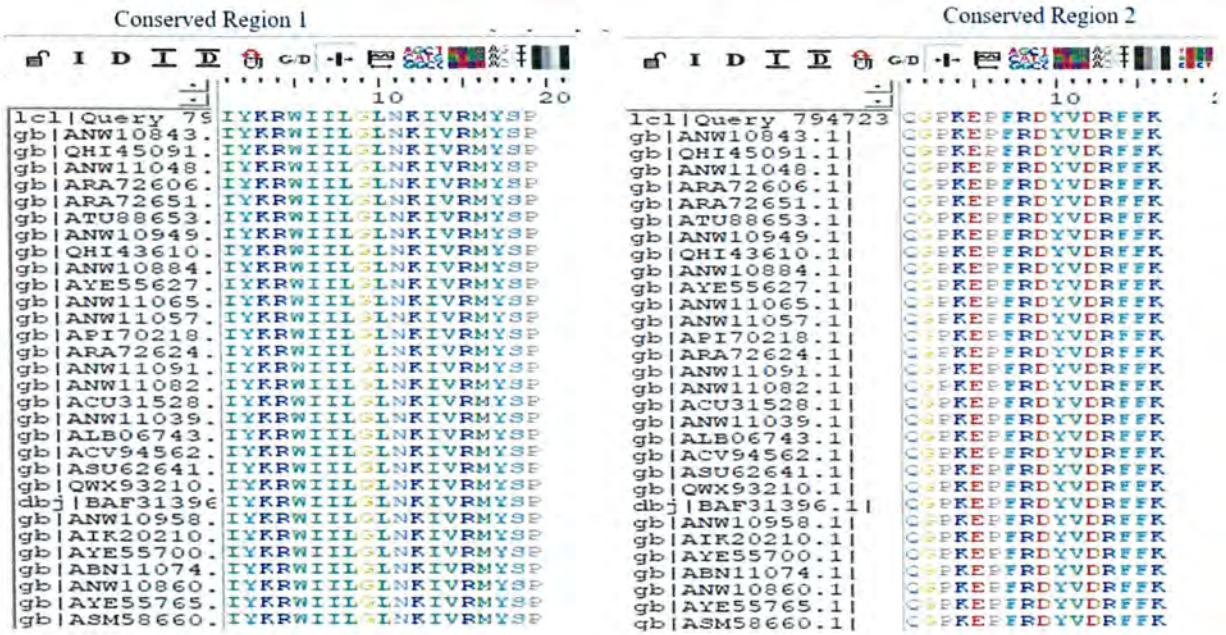


Figure 4.9. Conserved region 1 and 2 of the query sequence QAU-HIV1-AH1 with 30 reference sequences.

CHAPTER 5
DISCUSSION

5. Discussion

The HIV epidemic needs significant attention regarding its control and prevention as its burden is increasing around the globe. In year 2018, the number of new infected HIV cases reported were more than 1.9 million globally, and increasing every year (Rotheram et al., 2018). The significant increase in HIV cases might be due to lack of knowledge/safe practices/attitude about HIV infection and its asymptomatic nature in the early stages of infection. Individuals during the severe phase of infection could be at higher risk of secondary transmission of virus with a very high viral load (Cohen, 2010). The current study was performed from November 2021 to July 2022, the participants of the study were the high-risk behavior presenting individuals and referred to the physician for HIV testing. The participants included different age groups ranging from eight years children to 45 years adults. Among the 25 suspected patients 18 were tested positive for HIV infection. Most of the individuals represented general symptoms such as pale color, fatigue, muscles ache, swollen lymph nodes, fever, and mouth ulcers. The samples were initially checked through HIV-ICT testing and then subjected to confirmation through RT-PCR. As HIV-ICT could have false positivity, therefore all the positive samples were confirmed by a highly sensitive RNA quantification RT-PCR technique. After confirmation, those positive samples were processed to assess HIV-Gag gene and its mutations through conventional PCR amplification and nucleotide sequencing.

The overall prevalence in the general population of Pakistan has been reported to be 0.1 % (Rabold et al., 2021). HIV prevalence in different age groups in the current study showed the most affected age group observed was 24 to 30 years old followed by 31 to 45 years (Figure 5.1). These current results are in agreement with the former studies from the country (Haq et al., 2020; Wazir et al., 2008). The highest prevalence among the adult age group could be linked to the lack of awareness, use of contaminated syringes, and unsafe sexual interactions (Farid et al., 2009). Similarly, the gender wise prevalence revealed that the male individuals were more infected as compared to females (Figure 5.1). This could be positively linked to the lifestyle of the male individuals. Practices such as male circumcision with unhygienic blades, unsterilized blades usage by barbers' use of injecting drugs, unsafe sexual practices, piercing, and major exposure of males to environment and unhygienic activities posed to be a higher prevalence of HIV infection in male individuals (Haq et al., 2020).

Figure 5.1. Bar graph and table showing age and gender wise distribution of the patients (n=18).

Haq and colleagues (Haq et al., 2020) reported prevalence of 0.73% of HIV infection among the blood donors in Peshawar. However, the highest infection rate previously reported in Pakistan was from Sargodha district representing 13.3% prevalence (Wahid, 2019). The HIV infection in Pakistan is increasing and the health sector must intensify their efforts to manage and control the increase in HIV infection on immediate basis.

The confirmed HIV positive samples were processed for genetic and molecular assessment of HIV. Various viral genome regions (including the Gag gene) were targeted for amplification. Gag proteins are regulatory and structural polyproteins of HIV and all other retroviruses. Initially it was understood as a simple scaffolding protein which forms the viral core, but it was revealed that the Gag can recognize the genomic RNA and proteins of both virus and host as it transports to the cell membrane. Gag is responsible for forming the high ordered structures which are required for the assembly, maturation and budding of new infectious particles (Bell & Lever, 2013). It comprises of p15, p18 and p24 proteins and it has been reported that a point mutation in any of these. Gag proteins may lead to HIV-1 resistance to protease inhibitors (Dam et al., 2009). HIV Gag region

mutations has not been studied significantly in Pakistani isolates and therefore should be further studied and analyzed.

Phylogenetic analysis of HIV provides a valuable platform for molecular analysis, ancestral studies, performing genetic assessments, and guidance for the treatment. The evolutionary genetics and hierarchal relationship data is also provided. Nucleotide sequences of the current study were assessed for phylogenetic and evolutionary relationships with other sequences of HIV-1 Gag gene of Pakistan and other parts of the world. Phylogenetic trees of current sequences were constructed by MEGA 11 software, version 11 (Tamura, 2021). The evolutionary relationship was concluded by using the Neighbor-Joining method and Jukes-Cantor Model. The query sequences (isolates QAU-HIV1-AH1 and QAU-HIV1-AH2) clustered with a local sequence (KX232595). As shown in figure 4.3, isolates from the current study cluster together with the reference sequences from Pakistan and are phylogenetically distinct from those reported from African countries like Kenya, Uganda, and Zambia etc. All Pakistani isolates makes a separate cluster and are phylogenetically distinct from isolates of other regions of the world. Further studies are needed to understand the evolutionary relationship of Pakistani sequences with other sequences of the world.

Nucleotide substitutions in the wild-type codon leads to antiviral drug resistance in HIV (Shafer et al., 2000). Usually drug resistance associated mutations occur in more than one codon. Production of molecular target drugs that can specifically suppress HIV transmission and replication is possible due to recent developments molecular biology (Wensing et al., 2019).

HIV is hard to treat because its genetic material persistently sustain variations in the form of substitution or deletions (Ueda et al., 2019). For the eradication of this disease a global strategy has been adopted for reduction of new cases of HIV, mortality reduction, providing of affordable treatments and awareness (WHO). Insights into the global distribution, clinical information and genotype diversity is required in order to effectively tackle this problem (Kagan et al., 2019).

HIV has a very diverse pattern of molecular and genetic distribution both worldwide and locally. We studied the HIV-1 subtype-A sequences which is the most prevalent in Pakistan (Shah et al., 2011). Various studies conducted earlier has depicted this phenomenon about the prevalence in Pakistan (Khan et al., 2018).

Analysis of local and worldwide consensus sequences shows that Gag isolate of this study has relative conserved regions like other reference isolates. It consists of important motifs that are involved in key functions such as the ability to recognize genomic RNA of both virus and the host proteins as it transports to the cell membrane. It also forms structures that are required for the functioning of new virus particles. Amino acid substitutions in Gag region could play significant role in resistance to protease inhibitors and severity of the disease (Lv & Wang, 2015).

5.1. Amino acid substitution analysis of HIV-1 Gag isolates (QAU-HIV-AH1 and QAU-HIV1-AH2).

Amino acid substitution analysis is very important as it gives information about a particular change in amino acid will have difference in the structure and function of the virus and its role in resistance against drugs.

When alignment of the query sequence isolates (QAU-HIV1-AH1 and QAU-HIV1-AH2) with the reference isolate (HIV-1 isolate_PK002, accession number KX232595), we noted several amino acid substitutions (Figure 4.4, Table 4.1 and 4.2). Details of the corresponding substitutions are as follows:

5.1.1. Isolate QAU-HIV-AH1 amino acid substitutions.

5.1.1.1. W212M, P328Q and N392S.

Tryptophan a neutral, non-polar and hydrophobic amino acid present at the position 212 of the reference sequence is substituted to Methionine, which is also a neutral, non-polar and hydrophobic amino acid on the same position at the query sequence of the said isolate. Tryptophan, can replace any other aromatic amino acid residues, but it has a unique size and chemistry, so replacement by anything could affect the assembly and functioning of protein (Salzwedel et al., 1999). Methionine usually is non-reactive, and thus could be rarely directly involved in protein function. As of other hydrophobic amino acids, it might have a role in binding or recognition of hydrophobic ligands. So, with this substitution the structure of protein have been slightly changed (compared to the reference protein) but the binding ability of both amino acid share same properties so it might not have a significant impact on its binding. Although, the tryptophan share a unique property of the binding and envelop-mediated fusion of the virus to the host (Salzwedel et al., 1999).

Proline a neutral, non-polar and hydrophobic amino acid at position 328 of the reference sequence is substituted to a neutral, polar and hydrophilic amino acid Glutamine on the same position at the query sequence of the said isolate. Proline being less reactive is often involved in protein active or binding site (Betts et al., 2003), while Glutamine has a role in active sites of protein or its binding sites. The side chain (polar) is responsible for connections with other charged atoms or polar side groups (Russell, 1998). The substitution discussed might affect normal function of protein as both the amino acids have almost different roles.

Asparagine which is a polar, neutral, and hydrophilic amino acid at position 392 of the reference sequence is substituted to another neutral, polar and hydrophilic amino acid Serine on the same position at the query sequence of the said isolate. Asparagine have a key role in active site of protein or its binding sites (Coren et al., 2007) whereas, Serine are known for their role as protein functional centers and protein active or binding sites (Simon et al., 2002). So, the said substitution could not have any effect on the protein normal function as both amino acids are related to same roles.

5.1.1.2. E474K, D479T, and Q480R

Glutamate a hydrophilic, polar and neutral amino acid at position 474 of the reference sequence is substituted to another polar, positively charged and hydrophobic amino acid Leucine on the same position at the query sequence of the said isolate. Glutamate have a major role in binding sites or active sites of protein, they also have a role of providing a catalytic site for proteins such as proteases or lipases (Betts & Russell, 2003). Lysine are also involved in active and binding site of protein (Stites, 1997). These amino acids share almost same properties with different nature so their binding sites and structure could not be disturbed.

Aspartate which is a hydrophilic, negatively charged and polar amino acid at position 479 of the reference sequence is substituted to a hydrophobic, neutral and polar amino acid Threonine on the same position at the query sequence of the said isolate. Aspartate is likely to have role in active or binding sites of protein (Tantillo et al., 1994), while threonine occurs in protein surfaces or within interior part of the protein. They are present in protein functional sites, which are fairly reactive and form hydrogen bonding with different polar substrates (Frankel & Young, 1998). Both of these amino acids have different natures, functions and role so a substitution of these two results in the changing of structure and function of the site.

At position 480 of the reference sequence the hydrophilic, neutral and polar amino acid Glutamine substituted to Arginine a hydrophilic, positively charged and polar amino acid on the same position at the query sequence of the said isolate. Glutamine has a key role in active or binding sites of protein. The polar side chain is responsible for binding with other charged atoms or polar side groups (Russell, 1998), whereas Arginine also lies in the active site of protein or its binding sites and they may also interact with other polar side groups (Borders Jr et al., 1994). The substitute amino acids both share similar properties so they did not have an impact on their regulatory functions.

5.1.2. Isolate QAU-HIV-AH2 amino acid substitutions.

5.1.2.1. L321M, L322Q, V323I and N325I.

Leucine a non-polar, neutral, and hydrophobic amino acid at position 321 of the reference sequence substituted to another same nature acid Methionine on the same position at the query sequence of the said isolate. Leucine is non-reactive and rarely involved in catalysis of proteins but plays a role in substrate recognition and binding of hydrophobic ligands like lipids (Yu et al., 2017). Methionine is also non-reactive amino acid and share same properties as of Leucine. Like other hydrophobic amino acids, its role is in recognition or binding of hydrophobic ligands (Salzwedel et al., 1999). So, with this substitution no significant change in the structure and function of the protein occur as both the amino acids share similar properties.

At position 322 of the reference sequence a non-polar, hydrophobic, and neutral amino acid Leucine substituted which is hydrophilic, polar and neutral amino acid Glutamine on the same position at the query sequence of the said isolate. Leucine a non-reactive amino acid is rarely involved in catalysis but also plays a role in substrate recognition and binding of hydrophobic ligands such as lipids (Yu et al., 2017) while Glutamine is involved in protein binding or its active sites. The polar side chains are responsible for interactions with other charged atoms or polar side groups (Russell, 1998). Both amino acids are totally of different nature, polarity and share different characteristics which may have an impact on its function and reduces the membrane-binding capacity of the precursor protein (González & Affranchino, 1998).

Valine is replaced by Isoleucine at position 323 of the reference protein. Both are hydrophobic, non-polar and neutral in nature. Both Valine and Isoleucine possess low reactivity and are rarely

involved in functions. Though they could play its part in substrate recognition (Palanisamy et al., 2017).

Asparagine which is a neutral, polar and hydrophilic in nature present at position 325 of the reference sequence is substituted to a hydrophobic, non-polar, neutral amino acid Isoleucine on the same position at the query sequence of the said isolate. Asparagine have a key role in protein active or binding sites (Coren et al., 2007) whereas, Isoleucine has low reactivity and hardly involved in functions. Although it could have a role in substrate recognition. So, a change in reactivity and binding/active site could occur.

5.1.2.2. A326V, N327R, N374Y, K413Q, and S477C.

At position 326 of the reference sequence a non-polar, hydrophobic, and neutral amino acid Alanine substituted to a hydrophobic, non-polar, neutral amino acid Valine on the same position at the query sequence of the said isolate. Both Alanine and Valine possess same properties and are involved in protein functions. Although they could have a role in substrate recognition (Kempf et al., 2001).

Asparagine which is polar, neutral and hydrophilic amino acid at position 327 of the reference sequence is substituted to Arginine which is a hydrophilic, positively charged, polar amino acid on the same position at the query sequence of the said isolate. Asparagine have an important role for active or binding sites of protein and the polar side groups are involved for interactions and binding with other polar or charged groups. While, Arginine also have a role in active sites of protein and they can also interact better with other charged atoms or polar side groups (Jenkins et al., 2007). Both the amino acids share same properties and functions so no significant change could occur.

At position 374 of the reference sequence a neutral, hydrophilic, and polar amino acid Asparagine changed to a different nature polar, hydrophobic, neutral amino acid Tyrosine on the same position at the query sequence of the said isolate. Asparagine have a key role in binding sites of protein and its polar side groups has a role in binding with other polar charged groups. On the other hand, tyrosine is more of associated with binding of non-carbon atoms. Common role for tyrosine is phosphorylation (Shimizu et al., 2008). So, the tyrosine substitution might promote the co- receptor activity of the protein.

At position 413 of the reference sequence a positively charged, polar and hydrophilic amino acid Lysine changed to another hydrophilic, neutral and polar amino acid Glutamine on the same position at the query sequence of the said isolate. Lysine and Glutamine both often exist in protein active sites and could better interact with other charged atoms or polar side groups (Waksman & Sansom, 2005). As, both amino acids are of same nature and share same properties so no significant change in function could occur.

Serine a neutral, polar and hydrophilic amino acid at position 477 of the reference sequence is substituted to cysteine another polar but hydrophobic amino acid on the same position at the query sequence of the said isolate. Serine and Cysteine are known for their role as protein functional centers and protein active or binding sites (Simon et al., 2002). They are also known for their protease activities. The substitution of these amino acids may cause change of structural and functional activities of the proteins and their binding sights to the respective residues.

Conclusion

A staggering amount of 39 million cases of HIV has been reported worldwide by World health organization. In Pakistan, the prevalence of HIV is still increasing and seems uncontrollable till date. For the complete eradication of HIV, the understanding of nucleotide sequence variability and its consequent amino acid substitutions is very important. In the current study, HIV Gag gene was amplified, sequenced and its phylogenetic analysis conceals that the current sequences (isolate QAU-HIV1-AH1 and QAU-HIV1-AH2) grouped together with KX232595. This study also reports several amino acid substitutions in functionally and structurally important domains of Gag gene. Further analyses are required to analyze and understand whether these substitutions are associated with antiretroviral drug resistance or not. And we further suggest viral isolates from the Pakistani patients' needs to be studied for possible drug resistant mutations to prescribe specific antiretroviral therapies.

REFERENCES

REFERENCES

Abecasis, A. B., Pingarilho, M., & Vandamme, A.-M. (2018). Phylogenetic analysis as a forensic tool in HIV transmission investigations. *Aids, 32*(5), 543-554.

Abecasis, A. B., Wensing, A. M., Paraskevis, D., Vercauteren, J., Theys, K., Van de Vijver, D. A., . . . Beshkov, D. (2013). HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics. *Retrovirology, 10*(1), 1-13.

Aibekova, L., Foley, B., Hortelano, G., Raees, M., Abdraimov, S., Toichuev, R., & Ali, S. (2018). Molecular epidemiology of HIV-1 subtype A in former Soviet Union countries. *PloS one, 13*(2), e0191891.

Ali, M., Nadeem, M., Numan, M., Khalil, A. T., Maqbool, K., Yousaf, M. Z., . . . Idrees, M. (2017). Thirty years of HIV in Pakistan: a systematic review of prevalence and current scenario. *Future Virology, 12*(10), 609-623.

Anderson, E. M., & Maldarelli, F. (2018). The role of integration and clonal expansion in HIV infection: live long and prosper. *Retrovirology, 15*(1), 1-22.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., . . . Gasteiger, E. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic acids research, 40*(W1), W597-W603.

Avanzi, V. M., Vicente, B. A., Beloto, N. C. P., Gomes-da-Silva, M. M., Ribeiro, C. E. L., Tuon, F. F., . . . Raboni, S. M. (2017). Profile of HIV subtypes in HIV/HBV-and HIV/HCV-coinfected patients in Southern Brazil. *Revista da Sociedade Brasileira de Medicina Tropical, 50*, 470-477.

Azim, T., Chowdhury, E. I., Reza, M., Faruque, M. O., Ahmed, G., Khan, R., . . . Strathdee, S. A. (2008). Prevalence of infections, HIV risk behaviors and factors associated with HIV infection among male injecting drug users attending a needle/syringe exchange program in Dhaka, Bangladesh. *Substance use & misuse, 43*(14), 2124-2144.

Barré-Sinoussi, F., Chermann, J.-C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., . . . Rouzioux, C. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science, 220*(4599), 868-871.

Bell, N. M., & Lever, A. M. (2013). HIV Gag polyprotein: processing and early viral particle assembly. *Trends in microbiology, 21*(3), 136-144.

Beloukas, A., Psarris, A., Giannelou, P., Kostaki, E., Hatzakis, A., & Paraskevis, D. (2016). Molecular epidemiology of HIV-1 infection in Europe: An overview. *Infection, Genetics and Evolution, 46*, 180-189.

Betts, M. J., & Russell, R. B. (2003). Amino acid properties and consequences of substitutions. *Bioinformatics for geneticists, 317*, 289.

Beyrer, C., Razak, M. H., Lisam, K., Chen, J., Lui, W., & Yu, X.-F. (2000). Overland heroin trafficking routes and HIV-1 spread in south and south-east Asia. *Aids, 14*(1), 75-83.

Borders Jr, C., Broadwater, J. A., Bekeny, P. A., Salmon, J. E., Lee, A. S., Eldridge, A. M., & Pett, V. B. (1994). A structural role for arginine in proteins: multiple hydrogen bonds to backbone carbonyl oxygens. *Protein Science, 3*(4), 541-548.

Bour, S., Geleziunas, R., & Wainberg, M. A. (1995). The human immunodeficiency virus type 1 (HIV-1) CD4 receptor and its central role in promotion of HIV-1 infection. *Microbiological reviews, 59*(1), 63-93.

Brown, T., Sittitrai, W., Vanichseni, S., & Thisyakorn, U. (1994). The recent epidemiology of HIV and AIDS in Thailand. *AIDS (London, England), 8*, S131-S141.

Bustin, S. A., & Mueller, R. (2005). Real-time reverse transcription PCR (qRT-PCR) and its potential use in clinical diagnosis. *Clinical Science, 109*(4), 365-379.

Chiu, T. K., & Davies, D. R. (2004). Structure and function of HIV-1 integrase. *Current topics in medicinal chemistry, 4*(9), 965-977.

Chrystie, I., & Almeida, J. D. (1988). The morphology of human immunodeficiency virus (HIV) by negative staining. *Journal of medical virology, 25*(3), 281-288.

Chu, M., Zhang, W., Zhang, X., Jiang, W., Huan, X., Meng, X., . . . Tian, T. (2017). HIV-1 CRF01_AE strain is associated with faster HIV/AIDS progression in Jiangsu Province, China. *Scientific reports, 7*(1), 1-8.

Clavel, F., Guétard, D., Brun-Vézinet, F., Chamaret, S., Rey, M.-A., Santos-Ferreira, M. O., . . . Rouzioux, C. (1986). Isolation of a new human retrovirus from West African patients with AIDS. *Science, 233*(4761), 343-346.

Clavel, F., & Mammano, F. (2010). Role of Gag in HIV resistance to protease inhibitors. *Viruses, 2*(7), 1411-1426.

Coffin, J., & Swanstrom, R. (2013). HIV pathogenesis: dynamics and genetics of viral populations and infected cells. *Cold Spring Harbor perspectives in medicine, 3*(1), a012526.

Cohen, M. S., Gay, C. L., Busch, M. P., & Hecht, F. M. (2010). The detection of acute HIV infection. *The Journal of infectious diseases, 202*(Supplement_2), S270-S277.

Coren, L. V., Thomas, J. A., Chertova, E., Sowder, R. C., Gagliardi, T. D., Gorelick, R. J., & Ott, D. E. (2007). Mutational analysis of the C-terminal gag cleavage sites in human immunodeficiency virus type 1. *Journal of virology, 81*(18), 10047-10054.

Dam, E., Quercia, R., Glass, B., Descamps, D., Launay, O., Duval, X., . . . Group, A. S. (2009). Gag mutations strongly contribute to HIV-1 resistance to protease inhibitors in highly drug-experienced patients besides compensating for fitness loss. *PLoS pathogens, 5*(3), e1000345.

Derdeyn, C. A., & Silvestri, G. (2005). Viral and host factors in the pathogenesis of HIV infection. *Current opinion in immunology, 17*(4), 366-373.

Dore, G. J., Kaldor, J. M., Ungchusak, K., & Mertens, T. E. (1996). 9.2 Epidemiology of HIV and AIDS in the Asia-Pacific region. *Medical journal of Australia, 165*(9), 494-498.

Ehrlich, L. S., Liu, T., Scarlata, S., Chu, B., & Carter, C. A. (2001). HIV-1 capsid protein forms spherical (immature-like) and tubular (mature-like) particles in vitro: structure switching by pH-induced conformational changes. *Biophysical Journal, 81*(1), 586-594.

Farid-ul-Hasnain, S., Johansson, E., & Krantz, G. (2009). What do young adults know about the HIV/AIDS epidemic? Findings from a population based study in Karachi, Pakistan. *BMC Infectious Diseases, 9*(1), 1-11.

Frankel, A. D., & Young, J. A. (1998). HIV-1: fifteen proteins and an RNA. *Annual review of biochemistry, 67*, 1.

Freed, E. O. (1998). HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology, 251*(1), 1-15.

Freed, E. O. (2001). HIV-1 replication. *Somatic cell and molecular genetics, 26*(1), 13-33.

Freiberg, M. S., Duncan, M. S., Alcorn, C., Chang, C. C. H., Kundu, S., Mumpuni, A., . . . Vittinghoff, E. (2021). HIV Infection and the Risk of World Health Organization–Defined Sudden Cardiac Death. *Journal of the American Heart Association, 10*(18), e021268.

Gallo, R. C., Salahuddin, S. Z., Popovic, M., Shearer, G. M., Kaplan, M., Haynes, B. F., . . . Safai, B. (1984). Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science, 224*(4648), 500-503.

Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F., . . . Shaw, G. M. (1999). Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature, 397*(6718), 436-441.

Gatignol, A. (2007). Transcription of HIV: Tat and cellular chromatin. *Advances in pharmacology, 55*, 137-159.

Gaynor, R. (1992). Cellular transcription factors involved in the regulation of HIV-1 gene expression. *Aids, 6*(4), 347-364.

Gelderblom, H., Özel, M., Hausmann, E., Winkel, T., Pauli, G., & Koch, M. (1988). Fine structure of human immunodeficiency virus (HIV), immunolocalization of structural proteins and virus-cell relation. *Micron and Microscopica Acta, 19*(1), 41-60.

Gelderblom, H., Özel, M., & Pauli, G. (1989). Morphogenesis and morphology of HIV structure-function relations. *Archives of virology, 106*(1), 1-13.

Geourjon, C., & Deleage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics, 11*(6), 681-684.

Giovanetti, M., Ciccozzi, M., Parolin, C., & Borsetti, A. (2020). Molecular Epidemiology of HIV-1 in African Countries: A Comprehensive Overview. *Pathogens, 9*(12), 1072.

González, S. A., & Affranchino, J. L. (1998). Substitution of leucine 8 in the simian immunodeficiency virus matrix protein impairs particle formation without affecting N-myristylation of the Gag precursor. *Virology, 240*(1), 27-35.

Goodsell, D. S. (2015). Illustrations of the HIV life cycle. *The Future of HIV-1 Therapeutics*, 243-252.

Goto, T., Nakai, M., & Ikuta, K. (1998). The life-cycle of human immunodeficiency virus type 1. *Micron, 29*(2-3), 123-138.

Hamers, F. F., & Downs, A. M. (2003). HIV in central and eastern Europe. *The Lancet, 361*(9362), 1035-1044.

Haq, I., Ullah, R., Din, M., Ahmad, S., Anwar, F., Ali, M., & Khan, H. U. (2020). Unrecognized HIV infection in asymptomatic volunteer blood donors at district Peshawar, Khyber Pakhtunkhwa, Pakistan. *New Microbes and New Infections, 35*, 100685.

Harden, V. A. (2009). Acquired Immune Deficiency Syndrome (AIDS): History and Time Line. *Essentials of Public Health Biology: A Guide for the Study of Pathophysiology, 4*, 321.

Hikichi, Y., & Freed, E. O. (2021). Maturation of HIV-1. *Science, 373*(6555), 621-622.

Indriati, D. W., Kotaki, T., Khairunisa, S. Q., Witaningrum, A. M., Matondang, M. Q., Ueda, S., . . . Kameoka, M. (2018). Appearance of drug resistance mutations among the dominant HIV-1 subtype, CRF01_AE in Maumere, Indonesia. *Current HIV Research, 16*(2), 158-166.

Jacobson, J. M., Colman, N., Ostrow, N. A., Simson, R. W., Tomesch, D., Marlin, L., . . . Prince, A. M. (1993). Passive immunotherapy in the treatment of advanced human immunodeficiency virus infection. *Journal of Infectious Diseases, 168*(2), 298-305.

Jaffar, S., Grant, A. D., Whitworth, J., Smith, P. G., & Whittle, H. (2004). The natural history of HIV-1 and HIV-2 infections in adults in Africa: a literature review. *Bulletin of the World Health Organization, 82*, 462-469.

Jeang, K.-T., Xiao, H., & Rich, E. A. (1999). Multifaceted activities of the HIV-1 transactivator of transcription, Tat. *Journal of Biological Chemistry, 274*(41), 28837-28840.

Jenkins, L. M. M., Hara, T., Durell, S. R., Hayashi, R., Inman, J. K., Piquemal, J.-P., . . . Appella, E. (2007). Specificity of acyl transfer from 2-mercaptobenzamide thioesters to the HIV-1 nucleocapsid protein. *Journal of the American Chemical Society, 129*(36), 11067-11078.

Jenness, S. M., Sharma, A., Goodreau, S. M., Rosenberg, E. S., Weiss, K. M., Hoover, K. W., . . . Sullivan, P. (2017). Individual HIV risk versus population impact of risk compensation after HIV preexposure prophylaxis initiation among men who have sex with men. *PloS one, 12*(1), e0169484.

Jin, M., Hui, H., Robertson, D., Müller, M., Barré-Sinoussi, F., Hirsch, V., . . . Hahn, B. (1994). Mosaic genome structure of simian immunodeficiency virus from west African green monkeys. *The EMBO journal, 13*(12), 2935-2947.

Kagan, R. M., Dunn, K. J., Snell, G. P., Nettles, R. E., & Kaufman, H. W. (2019). Trends in HIV-1 drug resistance mutations from a US reference laboratory from 2006 to 2017. *AIDS Research and Human Retroviruses, 35*(8), 698-709.

Karim, A. M., Karim, M. A., Shinwari, Z. K., Lee, J. H., & Lee, S. H. (2020). Prevalence of Human Immunodeficiency Virus Infection in Rural Pakistan. *Iranian Journal of Public Health, 49*(12), 2421.

Kempf, D., King, M., Brun, S., Sylte, J., Marsh, T., Stewart, K., . . . Park, A. (2001). Association of the nucleocapsid-p1 (NC-p1) Gag cleavage site alanine to valine mutation with mutational patterns in HIV protease producing protease inhibitor resistance. *Antiviral Therapy, 6*, 59-59.

Khan, A. A., Khan, A., & Bokhari, A. (2010). The HIV epidemic in Pakistan. *J Pak Med Assoc, 60*(4), 300-307.

Khan, S., Zahid, M., Qureshi, M. A., Mughal, M. N., & Ujjan, I. D. (2018). HIV-1 genetic diversity, geographical linkages and antiretroviral drug resistance among individuals from Pakistan. *Archives of virology, 163*(1), 33-40.

Khawaja, Z. A., Gibney, L., Ahmed, A. J., & Vermund, S. H. (1997). HIV/AIDS and its risk factors in Pakistan. *Aids, 11*(7), 843-848.

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

Krakauer, D. C., & Nowak, M. (1999). T–cell induced pathogenesis in HIV: bystander effects and latent infection. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 266*(1423), 1069-1075.

Lackner, A., Lederman, M. M., & Rodriguez, B. (2012). HIV pathogenesis: the host. *Cold Spring Harbor perspectives in medicine, 2*(9), a007005.

Lemey, P., Pybus, O. G., Wang, B., Saksena, N. K., Salemi, M., & Vandamme, A.-M. (2003). Tracing the origin and history of the HIV-2 epidemic. *Proceedings of the National Academy of Sciences, 100*(11), 6588-6592.

Levy, J. A. (2009). HIV pathogenesis: 25 years of progress and persistent challenges. *Aids, 23*(2), 147-160. doi:10.1097/QAD.0b013e3283217f9f

Levy, J. A., Hoffman, A. D., Kramer, S. M., Landis, J. A., Shimabukuro, J. M., & Oshiro, L. S. (1984). Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science, 225*(4664), 840-842.

Lusic, M., & Siliciano, R. F. (2017). Nuclear landscape of HIV-1 infection and integration. *Nature Reviews Microbiology, 15*(2), 69-82.

Lv, Z., Chu, Y., & Wang, Y. (2015). HIV protease inhibitors: a review of molecular selectivity and toxicity. *Hiv/aids (Auckland, NZ), 7*, 95.

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., . . . Gonzales, N. R. (2010). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic acids research, 39*(suppl_1), D225-D229.

Marx, P., Munn, R., & Joy, K. (1988). Computer emulation of thin section electron microscopy predicts an envelope-associated icosadeltahedral capsid for human immunodeficiency virus. *Laboratory investigation; a journal of technical methods and pathology, 58*(1), 112-118.

McManus, W. R., Bale, M. J., Spindler, J., Wiegand, A., Musick, A., Patro, S. C., . . . Cyktor, J. C. (2019). HIV-1 in lymph nodes is maintained by cellular proliferation during antiretroviral therapy. *The Journal of clinical investigation, 129*(11), 4629-4642.

Merson, M. H. (1993). Slowing the spread of HIV: agenda for the 1990s. *Science, 260*(5112), 1266-1268.

Mertens, T. E., & Low-Beer, D. (1996). HIV and AIDS: where is the epidemic going? *Bulletin of the World Health Organization, 74*(2), 121.

Mir, F., Mahmood, F., Siddiqui, A. R., Baqi, S., Abidi, S. H., Kazi, A. M., . . . Soofi, S. B. (2020). HIV infection predominantly affecting children in Sindh, Pakistan, 2019: a cross-sectional study of an outbreak. *The Lancet Infectious Diseases, 20*(3), 362-370.

Montoya, C. J., Velilla, P. A., Chougnet, C., Landay, A. L., & Rugeles, M. T. (2006). Increased IFN-γ production by NK and CD3+/CD56+ cells in sexually HIV-1-exposed but uninfected individuals. *Clinical immunology, 120*(2), 138-146.

Naif, H. M. (2013). Pathogenesis of HIV infection. *Infectious Disease Reports, 5*(11), e6.

Nunzio, F. D. (2018). Nuclear pore complexes, genome organization and HIV-1 infection *Nuclear Pore Complexes in Genome Organization, Function and Maintenance* (pp. 183-199): Springer.

Oster, A. M., Switzer, W. M., Hernandez, A. L., Saduvala, N., Wertheim, J. O., Nwangwu-Ike, N., . . . Hall, H. I. (2017). Increasing HIV-1 subtype diversity in seven states, United States, 2006–2013. *Annals of epidemiology, 27*(4), 244-251. e241.

Padamsee, T. J. (2020). Fighting an epidemic in political context: thirty-five years of HIV/AIDS policy making in the United States. *Social History of Medicine, 33*(3), 1001-1028.

Palanisamy, N., Osman, N., Ohnona, F., Xu, H.-T., Brenner, B., Mesplède, T., & Wainberg, M. A. (2017). Does antiretroviral treatment change HIV-1 codon usage patterns in its genes: a preliminary bioinformatics study. *AIDS Research and Therapy, 14*(1), 1-10.

Papadopulos-Eleopulos, E., Turner, V. F., Papadimitriou, J., Page, B., Causer, D., Alfonso, H., . . . Fiala, C. (2004). A critique of the Montagnier evidence for the HIV/AIDS hypothesis. *Medical hypotheses, 63*(4), 597-601.

Patel, R. (2021). Advances in Testing for Infectious Diseases—Looking Back and Projecting Forward. *Clinical Chemistry, 68*(1), 10-15. doi:10.1093/clinchem/hvAb110

Peeling, R. W., Mabey, D., Herring, A., & Hook, E. W. (2006). Why do we need quality-assured diagnostic tests for sexually transmitted infections? *Nature Reviews Microbiology, 4*(12), 909-921.

Pillay, D., Geretti, A. M., & Weiss, R. A. (2009). Human immunodeficiency viruses. *Principles and practice of clinical virology*(Ed. 6), 897-938.

Quinn, T. C. (2008). HIV epidemiology and the effects of antiviral therapy on long-term consequences. *AIDS (London, England), 22*(Suppl 3), S7.

Rabold, E. M., Shaikh, S. A., Schenkel, K., Baig, M. A., Asghar, R. J., Liban, A., . . . Ali, H. (2021). Notes from the Field: Pediatric HIV Outbreak in Ratodero, Pakistan—April 2019–April 2020. *Morbidity and Mortality Weekly Report, 70*(42), 1489.

Rai, M. A., Nerurkar, V. R., Khoja, S., Khan, S., Yanagihara, R., Rehman, A., . . . Ali, S. H. (2010). Evidence for a" Founder Effect" among HIV-infected injection drug users (IDUs) in Pakistan. *BMC Infectious Diseases, 10*(1), 1-5.

Rajarapu, G. (2014). Genes and Genome of HIV-1. *Journal of Phylogenetics & Evolutionary Biology, 2*(1), 1000126.

Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., . . . Baumeister, K. (1985). Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature, 313*(6000), 277-284.

Raza, A., Ahmadian, A., Rafiq, M., Salahshour, S., Naveed, M., Ferrara, M., & Soori, A. H. (2020). Modeling the effect of delay strategy on transmission dynamics of HIV/AIDS disease. *Advances in Difference Equations, 2020*(1), 1-13.

Richardson, M., & Page, I. (2018). Role of serological tests in the diagnosis of mold infections. *Current fungal infection reports, 12*(3), 127-136.

Rothenberg, R. B., Scarlett, M., del Rio, C., Reznik, D., & O'Daniels, C. (1998). Oral transmission of HIV. *Aids, 12*(16), 2095-2105.

Rotheram-Borus, M. J., Davis, E., & Rezai, R. (2018). Stopping the rise of HIV among adolescents globally. *Current opinion in pediatrics, 30*(1), 131.

Royce, R. A., Sena, A., Cates Jr, W., & Cohen, M. S. (1997). Sexual transmission of HIV. *New England Journal of Medicine, 336*(15), 1072-1078.

Rücker, E., Grivel, J.-C., Münch, J., Kirchhoff, F., & Margolis, L. (2004). Vpr and Vpu are important for efficient human immunodeficiency virus type 1 replication and CD4+ T-cell depletion in human lymphoid tissue ex vivo. *Journal of virology, 78*(22), 12689-12693.

Russell, R. B. (1998). Detection of protein three-dimensional side-chain patterns: new examples of convergent evolution. *Journal of molecular biology, 279*(5), 1211-1227.

Ruxrungtham, K., Brown, T., & Phanuphak, P. (2004). Hiv/aids in asia. *The Lancet, 364*(9428), 69-82.

Saez-Cirion, A., & Manel, N. (2018). Immune responses to retroviruses. *Annu Rev Immunol, 36*(1), 193-220.

Salzwedel, K., West, J. T., & Hunter, E. (1999). A conserved tryptophan-rich motif in the membrane-proximal region of the human immunodeficiency virus type 1 gp41 ectodomain is important for Env-mediated fusion and virus infectivity. *Journal of virology, 73*(3), 2469-2480.

Santoro, M. M., & Perno, C. F. (2013). HIV-1 genetic variability and clinical implications. *International Scholarly Research Notices, 2013*.

Shafer, R. W., Kantor, R., & Gonzales, M. J. (2000). The genetic basis of HIV-1 resistance to reverse transcriptase and protease inhibitors. *AIDS reviews, 2*(4), 211.

Shafiee, H., Wang, S., Inci, F., Toy, M., Henrich, T. J., Kuritzkes, D. R., & Demirci, U. (2015). Emerging technologies for point-of-care management of HIV infection. *Annual review of medicine, 66*, 387-405.

Shah, S., Xing, H., Altaf, A., Chen, B., Liao, L., Jia, Y., . . . Shao, Y. (2011). Antiretroviral drug resistance mutations among treated and treatment-naive patients in Pakistan: diversity of the HIV type 1 pol gene in Pakistan. *AIDS Research and Human Retroviruses, 27*(12), 1277-1282.

Shimizu, N., Tanaka, A., Oue, A., Mori, T., Apichartpiyakul, C., & Hoshino, H. (2008). A short amino acid sequence containing tyrosine in the N-terminal region of G protein-coupled receptors is critical for their potential use as co-receptors for human and simian immunodeficiency viruses. *Journal of general virology, 89*(12), 3126-3136.

Sigal, A., & Baltimore, D. (2012). As good as it gets? The problem of HIV persistence despite antiretroviral drugs. *Cell host & microbe, 12*(2), 132-138.

Simon, V., Vanderhoeven, J., Hurley, A., Ramratnam, B., Louie, M., Dawson, K., . . . Markowitz, M. (2002). Evolving patterns of HIV-1 resistance to antiretroviral agents in newly infected individuals. *Aids, 16*(11), 1511-1519.

Speck, R. F., Penn, M. L., Wimmer, J. r., Esser, U., Hague, B. F., Kindt, T. J., . . . Goldsmith, M. A. (1998). Rabbit cells expressing human CD4 and human CCR5 are highly permissive for human immunodeficiency virus type 1 infection. *Journal of virology, 72*(7), 5728-5734.

Staszewski, S., Miller, V., Sabin, C., Schlecht, C., Gute, P., Stamm, S., . . . Hill, A. (1999). Determinants of sustainable CD4 lymphocyte count increases in response to antiretroviral therapy. *Aids, 13*(8), 951-956.

Stevenson, M. (2003). HIV-1 pathogenesis. *Nature medicine, 9*(7), 853-860.

Stites, W. E. (1997). Protein− protein interactions: interface structure, binding thermodynamics, and mutational analysis. *Chemical reviews, 97*(5), 1233-1250.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol, 30*(12), 2725-2729. doi:10.1093/molbev/mst197

Tantillo, C., Ding, J., Jacobo-Molina, A., Nanni, R. G., Boyer, P. L., Hughes, S. H., . . . Arnold, E. (1994). Locations of anti-AIDS drug binding sites and resistance mutations in the three-dimensional structure of HIV-1 reverse transcriptase: implications for mechanisms of drug inhibition and resistance. *Journal of molecular biology, 243*(3), 369-387.

Tappuni, A. R. (2020). The global changing pattern of the oral manifestations of HIV. *Oral Diseases, 26*, 22-27.

Thomson, M. M., Villahermosa, M. L., Vázquez-de-Parga, E., Cuevas, M. T., Delgado, E., Manjón, N., . . . Carrillo, M. G. (2000). Widespread circulation of a B/F intersubtype recombinant form among HIV-1-infected individuals in Buenos Aires, Argentina. *Aids, 14*(7), 897.

Ueda, S., Witaningrum, A. M., Khairunisa, S. Q., Kotaki, T., & Kameoka, M. (2019). Genetic diversity and drug resistance of HIV-1 circulating in North Sulawesi, Indonesia. *AIDS Research and Human Retroviruses, 35*(4), 407-413.

Von Schwedler, U. K., Stuchell, M., Müller, B., Ward, D. M., Chung, H.-Y., Morita, E., . . . Cimbora, D. M. (2003). The protein network of HIV budding. *Cell, 114*(6), 701-713.

Wahid, B. (2019). An update on the severe outbreak of HIV in Kot Imrana, Pakistan. *The Lancet Infectious Diseases, 19*(3), 241.

Waksman, G., & Sansom, C. (2005). Introduction: Proteomics and protein-protein interactions: Biology, chemistry, bioinformatics, and drug design *Proteomics and Protein-Protein Interactions* (pp. 1-18): Springer.

Wazir, M. S., Mehmood, S., Ahmed, A., & Jadoon, H. R. (2008). Awareness among barbers about health hazards associated with their profession. *J Ayub Med Coll Abbottabad, 20*(2), 35-38.

Wensing, A. M., Calvez, V., Ceccherini-Silberstein, F., Charpentier, C., Günthard, H. F., Paredes, R., . . . Richman, D. D. (2019). 2019 update of the drug resistance mutations in HIV-1. *Topics in antiviral medicine, 27*(3), 111.

Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Church, D. M., . . . Helmberg, W. (2005). Database resources of the national center for biotechnology information. *Nucleic acids research, 33*(suppl_1), D39-D45.

Wodarz, D., & Nowak, M. A. (2002). Mathematical models of HIV pathogenesis and treatment. *BioEssays, 24*(12), 1178-1187.

Wu, Y., & Marsh, J. W. (2003). Gene transcription in HIV infection. *Microbes and infection, 5*(11), 1023-1027.

Yamaguchi, J., Olivo, A., Laeyendecker, O., Forberg, K., Ndembi, N., Mbanya, D., . . . Rodgers, M. A. (2018). Universal target capture of HIV sequences from NGS libraries. *Frontiers in microbiology, 9*, 2150.

Yang, Z., Lasker, K., Schneidman-Duhovny, D., Webb, B., Huang, C. C., Pettersen, E. F., . . . Ferrin, T. E. (2012). UCSF Chimera, MODELLER, and IMP: an integrated modeling system. *Journal of structural biology, 179*(3), 269-278.

Yarchoan, R., & Uldrick, T. S. (2018). HIV-associated cancers and related diseases. *New England Journal of Medicine, 378*(11), 1029-1041.

Yeni, P. (2006). Update on HAART in HIV. *Journal of Hepatology, 44*, S100-S103.

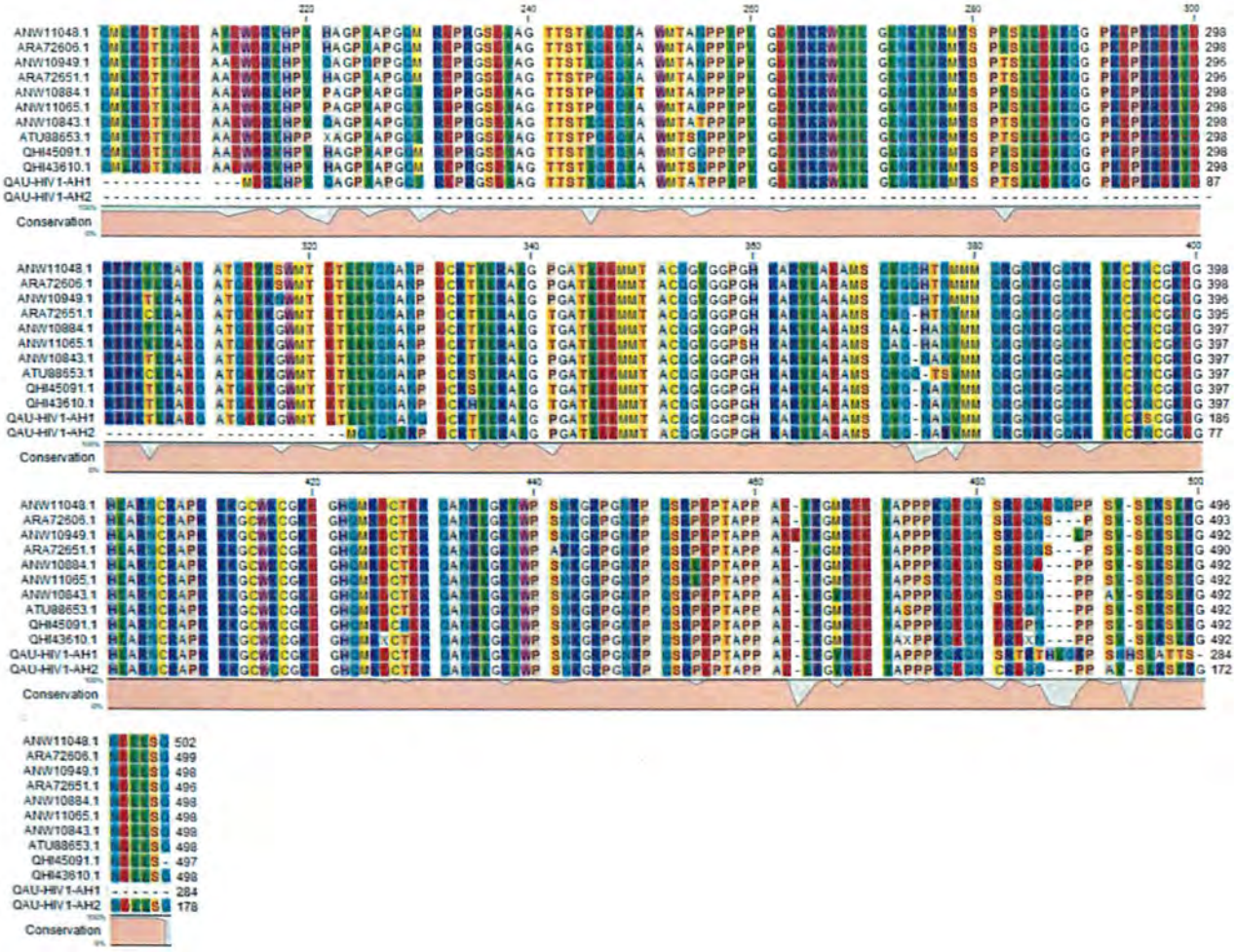
Yu, F.-H., Huang, K.-J., & Wang, C.-T. (2017). C-terminal HIV-1 transframe p6* tetrapeptide blocks enhanced Gag cleavage incurred by leucine zipper replacement of a deleted p6* domain. *Journal of virology, 91*(10), e00103-00117.

Zahra, A., Saleem, M. A., Javed, H., Khan, M. A. U., Naveed, M., & Shakoori, A. R. (2021). Single Nucleotide polymorphism induces a positive selection pressure at gag-pol sites in human immunodeficiency virus favoring drug resistance mutations. *Pakistan J. Zool, 53*(5), 1703-1714.

Zak-Place, J., & Stern, M. (2004). Health belief factors and dispositional optimism as predictors of STD and HIV preventive behavior. *Journal of American college health, 52*(5), 229-236.

Zhou, J., Yuan, X., Dismuke, D., Forshey, B. M., Lundquist, C., Lee, K.-H., . . . Chen, C. H. (2004). Small-molecule inhibition of human immunodeficiency virus type 1 replication by specific targeting of the final step of virion maturation. *Journal of virology, 78*(2), 922-929.

APPENDIX



Comparison of Query Sequences (QAU-HIV1-AH1 and QAU-HIV1-AH2) with reported amino acid sequences from Kenya, Uganda, USA, and Pakistan showing amino acid substitutions and mutations.