Molecular Characterization of Pol Gene of HIV in Pakistan

By
Adnan Zeb

Registration no: 02272111006

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

2023

DRSML QAU

Molecular Characterization of Pol Gene of HIV in Pakistan

DRSML QAU

By

Adnan Zeb

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.

بسم الله الرحمن الرحيم

Begin with the Name of Allah the most Merciful and Mighty

DRSN

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 **"Molecular Characterization of Pol Gene of HIV in Pakistan"** submitted by **Mr. Adnan Zeb** 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. Shahid Mahmood Baig

Chairman

Pakistan Science Foundation

1-Constitution Avenue, G-5/2, Islamabad

Chairperson ____________________________

Dr. Javaria Qazi

Associate Professor

Department of Biotechnology

Quaid-i-Azam University Islamabad

Dated ________19-05-2023___________________

DRSML QAU

DECLARATION OF ORIGINALITY

I hereby declare that the work ***"Molecular Characterization of Pol Gene of HIV in Pakistan"*** 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:

Adnan Zeb

Date: __________

DRSML QAU

I dedicated this work, especially to my brothers "Hukam Zeb & Saran Zeb", my Parents, and my charming family

Thank you for your being

My best friends,

My comfort when things go wrong,

The ones who encourage me, believe me,

Change my "weak" to strong,

And for filling my heart with such a pride

I feel like the happiest person alive

With you all by my side

Adnan Zeb

Table of Contents

ACKNOWLEDGEMENT i
List of Abbreviations ii
List of Figures iii
List of Tables iv
Abstract v
CHAPTER 1 ....
1 Introduction 1
Aims and Objectives of Current Study 7
CHAPTER 2 ....
2 Literature review 8
2.1 Human Immunodeficiency Virus (HIV) 8
2.2 History of Human Immunodeficiency Virus (HIV) 8
2.3 Classification of HIV 10
2.4 Epidemiology of HIV-1 12
2.4.1 Africa 12
2.4.2 Europe 12
2.4.3 North and South America 13
2.4.4 Australia 13
2.4.5 Asia 13
2.5 Prevalence of HIV in Pakistan 14
2.6 Morphology of Human Immunodeficiency Virus (HIV) 15
2.6.1 General structure 15
2.6.2 Envelope 16
2.6.3 Matrix 16
2.6.4 Core 16
2.7 Genome organization of HIV 16
2.8 Life Cycle of HIV 19
2.8.1 Viral entry 19
2.8.2 Post viral entry step 19
2.8.3 Reverse Transcription 20
2.8.4 Nucleus import 20
2.8.5 Integration 21
2.8.6 Gene Expression 21
2.8.7 Viral Particle Production 22

DRSML QAU

2.8.8 Budding....22
2.8.9 Maturation....23
2.9 Transmission....23
2.10 Pathogenesis....24
2.11 HIV variability....26
2.12 Treatment of HIV....26
2.12.1 Antiretroviral therapies (ART)....26
2.12.2 Medicinal plants using for HIV treatment....31
2.12.3 HIV Treatment by Gene editing tools....32
2.13 Vaccine development....34
CHAPTER 3....
3 Material and Methods....36
3.1 Sample Collection....36
3.2 Inclusion and Exclusion Criteria....36
3.3 Primers designing for *pol* gene of HIV-1....36
3.4 RNA extraction and cDNA synthesis....37
3.5 PCR amplification of HIV *pol* gene....38
3.6 Preperation of TAE Buffer....39
3.7 PCR products detection on Agarose gel....40
3.8 Sequencing of HIV *pol* gene....41
3.9 Nucleotide Blast and Phylogenetic Analysis....41
3.10 Translation of nucleotide sequences....42
3.11 Alignment *pol* protein of HIV....42
3.12 Prediction of Secondary structures....42
3.13 Protein Structure Modeling and Visualization....42
3.14 Location of isolated Proteins in Reference Protein....42
3.15 Identification of Conserved regions....43
CHAPTER 4....
4 Results....44
4.1 Results of PCR amplified products through Gel electrophoresis....44
4.2 Nucleotide Sequencing results....45
4.3 Phylogenetic analysis....45
4.4 Amino acids substitutions in HIV-1 *pol* protein....48
4.5 Secondary structure prediction....52
4.6 Modeling and Visualization of Protein....57

DRSML QAU

4.7 Protein isolates Location in Reference Protein 59
4.8 Conserved Region Identification 60
CHAPTER 5 ....
5 Discussion 61
Conclusion 67
REFERENCES 68
PUBLICATIONS 86
ORIGNALITY REPORT 87
DRSML QAU

ACKNOWLEDGEMENT

The first and premier gratitude is to the **ALMIGHTY ALLAH** whose generosity and uncountable blessings enabled me to complete my research.

I am grateful to our Chairperson, **Prof. Dr. Javaria Qazi** for encouragement, cooperation, and affectionate treatment during my study. I am thankful to my dear 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 am very grateful to **Higher Education Commission (HEC)** for their support. This research work is financially supported by **National Research Program for University (NARPU)**.

I am very obliged to my seniors **Sidra Rahaman, Lubna Rahman, Akmal Zubair, Syed Ahsan Shahid, Sajid Ali Shah and Tazeen, Dawood Ghafoor** for their technical and professional support, sharp guidance, encouragement, and valuable suggestions during my research studies. Additionally, I would like to express my deep feelings of respect to my lab fellows, Asim Saeed, Hanbal Ahmad Khan, Asiya Nawaz, Iffat Irshad, Ahsan Saeed, Naveera Khan, Hassan Ayaz, Saad Ullah and Huma Gull for their aspiring guidance and valuable suggestions during my laboratory experiments.

I wish to express my sincere words of thanks to all my classmates and friends for their attention, unceasing encouragement, friendly advice, and continuous support during my study. Last but not the least. This thesis owes a lot to my beloved parents, brothers and all my family members. Without their kind support, constant words of encouragement, inspiration, suggestion, guideline in 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.

Adnan Zeb.

DRSML QAU

List of Abbreviations

%	Percentage
˚C	Centigrade
µl/µg	Microliter/Microgram
HAART	Highly Active Anti-Retroviral Therapy
ZFN	Zinc finger nucleases
HIV	Human Immunodeficiency Virus
Aa	Amino acid
INSTIs	Integrase strand transfer inhibitors
TALENs	Transcription activator-like effector nucleases
PR	*Protease*
PIs	Protease inhibitors
IN	Integrase
ml	Milliliter
bp	Base pair
NRTIs	Nucleoside Reverse-Transcriptase Inhibitors
kDa	Killo Dalton
KP	Khyber Pakhtunkhwa
3D	Three Dimensional
Min	Minute
PCR	Polymerase Chain Reaction
NNRTIs	Non-nucleoside reverse transcriptase inhibitors
RNA	Ribonucleic Acid
DNA	Deoxyribonucleic acid
Sec	Seconds
AIDS	Acquired Immuno-Deficiency Syndrome
PIC	pre-integration complex
RT	Reverse Transcriptase
WHO	World Health Organization
cDNA	Complementary DNA
RT-PCR	Real Time- Polymerase Chain Reaction
ELISA	Enzyme Linked Immunosorbent Assay
CRFs	Circulating Recombinant Forms

List of Figures

Figure 1: Graphical demonstration of HIV types and subtypes classification 11
Figure 2: Morphology of Human Immunodeficiency Virus 18
Figure 3: Demonstration of genomic organization of HIV 18
Figure 4: Summary of Viral Replication Cycle 23
Figure 5. Graphical representation of overall research work 35
Figure 6: Optimized conditions of HIV complementary DNA (cDNA) synthesis 37
Figure 7: Representation of Optimized conditions of Polymerase Chain Reaction (PCR) 39
Figure 8: Gel purification procedure of the amplified PCR product 41
Figure 9: Gel electrophoresis and visualization of amplified products 44
Figure 10: Graphical irepresentation of sequencing in the form of chromatogram 45
Figure 11: Phylogenetic analysis of pol gene of studied HIV-1 in Pakistan 47
Figure 12: Representation of the alignment of amino acid sequences 48
Figure 13: Secondary structure of the reference *pol* protein sequence of HIV-1 53
Figure 14: Secondary structure of *pol* protein of QAU-AZ1 isolate 54
Figure 15: Secondary structure of pol protein of QAU-AZ2 isolate 55
Figure 16: Secondary structure of pol protein of QAU-AZ3 isolate 56
Figure 17: The 3D structure representation of HIV-1 pol reference protein and isolate QAU-AZ1 57
Figure 18: The 3D structure representation of HIV-1 *pol* reference protein and isolate QAU-AZ2 58
Figure 19: The3D structure representation of HIV-1 pol reference protein and isolate QAU-AZ3 58
Figure 20. Representation of our isolates in 3D and Liner structure of Reverse Transcriptase protein 59
Figure 21. Representation of Conserved region in 3D structure of our isolated protein QAU-AZ1 60

DRSML QAU

List of Tables

Table 1: Demonstration of primer name, oligonucleotide sequence, and product size of amplified product of pol gene of HIV 36

Table 2: Different reagents used in cDNA synthesis 38

Table 3: Different reagents use in PCR along with its volume 38

Table 4: Reagents use in gel electrophoresis 40

Table 5: Amino acid substitutions in the HIV-1 isolates (QAU-AZ1) proteins compared with reference sequence of HIV-1 pol protein (AHF27452). 49

Table 6: Amino acid substitutions in the HIV-1 isolates (QAU-AZ2) protein compared with reference sequence of HIV-1 pol protein (AHF27452) 50

Table 7: Amino acid substitutions in the HIV-1 isolates (QAU-AZ3) proteins compared with reference sequence of HIV-1 pol protein (AHF27452) 51

DRSML QAU

Abstract

Human Immunodeficiency Virus (HIV) is a single stranded RNA virus belong to family *Retroviridae* and genus Lentivirus. HIV is the major causative agent of a lethal condition called Acquired-Immunodeficiency-Syndrome (AIDS). World Health Organization (WHO) reported 38 million HIV infected people globally. By the end of 2021, 1.7 million new HIV cases and 0.69 million deaths were reported. Pakistan is among those countries where HIV prevalence is increasing rapidly and a remarkable increase of 57% new HIV cases have been reported in the last decade. In current study, blood samples were taken from high-risk behavior individuals from different cities of Pakistan. The HIV positive samples were confirmed by Real Time-Polymerase Chain Reaction. The Real Time-PCR based positive samples were further processed for *pol* gene amplification followed by nucleotide sequencing through Sanger method. Phylogenic analysis was performed through Maximum likelihood method using Hasegawa-Kishino-Yano + Gamma with Invariant-sites (HKY+G+I) model. Moreover, amino acid substitutions were also identified. The amplified *pol* gene sequence of our isolates QAU-AZ1 and QAU-AZ3 clustered with a previously reported sequence from China (Accession no. MN908916), while isolate QAU-AZ2 clustered with a Pakistani sequence (Accession no. JQ011662). Moreover, amino acid substitutions were identified in all three isolates. The positions of these substitutions were analyzed with reference sequence (GenBank Accession No. AHF27452). The 2D and 3D structure of all isolates and reference protein revealed no change in structure of protein as the substituted amino acids are involved in function of proteins. Further studies are required to investigate the *pol* gene of the viral isolates from Pakistani patients to identify possible drug resistance associated with amino acid substitutions.

Keywords: HIV, *pol* gene, *pol* protein, Phylogenetic analysis, Mutational analysis.

DRSML QAU

CHAPTER 1

INTRODUCTION

DRSML QAU

1 Introduction

In 2019, 1.7 million individuals were infected by HIV, with 0.69 million deaths worldwide (Freiberg et al., 2021). World health organization (WHO) by the end of 2021, reported that HIV had infected approximately 38 million people globally (Tappuni, 2020; Tariq et al., 2022). Pakistan one of the few countries of Asia, in which the incidence of HIV is increasing significantly every year since 1987. Currently, the HIV prevalence rate is less than 0.7% among general population of Pakistan (Haq et al., 2020). In Pakistan, the HIV epidemic is highly heterogeneous with a large diversity in the transference 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 reported HIV cases till January 5th, 2020 (Haq et al., 2020). All four provinces, including the capital city Islamabad, Azad Kashmir, and Gilgit Baltistan, report more than 1,000 new HIV cases each month. This proves categorically that the HIV epidemic is currently being disseminated to the general population by homosexuals, heterosexuals, injecting drug users and transgender sex workers. In Pakistan, 9,773 persons tested positive for HIV during the last 10 months of 2022, seriously undermining efforts to prevent and control the disease and amply illuminating the HIV spread from vulnerable populations to the general public (Ahmed, 2022). The rising number of HIV infections has become a threat to the Pakistani and worldwide communities. As a result, quick diagnostic services, public awareness campaigns about this fatal illness, and the creation of treatment facilities are required throughout the country.

Human immunodeficiency virus (HIV) is a single stranded RNA virus, belongs to genus Lentivirus, subfamily Orthoretrovirinae and Family Retroviridae (Ali et al., 2017; Luciw, 1996). It is divided into two types, HIV-1 and HIV-2 which are based on its genetic characteristics and its viral antigen differences (Gao et al., 1999). Immunodeficiency viruses that affect non-human primates I.e.., Simian immunodeficiency virus, or SIV, is also belong to genus Lentivirus. HIV-1 and HIV-2 are evolved from Central African chimpanzees (SIVcpz) and West African sooty mangabeys (SIVsm) respectively, as non-human primate immunodeficiency viruses (Virus, 2016). The most common form among the two types is HIV-1 which infected millions of people globally. Although, HIV-2 infections is limited and reported in some areas of Western and Central Africa and also been recorded in some other countries including United States (Peruski et al., 2020).

DR. SML QAU

The mature HIV virus is spherical in shape, about 100-120nm in diameter and made up of a lipid bilayer membrane enclosing a dense truncated cone shaped nucleocapsid or core. Two identical copies of about 9.8 kb long positive sense, linear and single stranded RNA molecules are present inside the core (Zulfiqar et al., 2017); (Fanales-Belasio et. al, 2010). There are Nine genes present in HIV encodes for about fifteen viral proteins. The HIV genome composed of structural genes along with other regulatory/accessory genes. There are three structure genes present in HIV which are *gag, pol and env*. Gag gene of HIV encodes different types of structural proteins such as core proteins (p6, p7, p15, p24) and matrix protein (p17). The *pol* gene encoded for three different types of proteins or enzymes which are integrase, protease and reverse transcriptase. These protein play a crucial role in replication of HIV virus (Di Nunzio, 2018b). The *env* gene of HIV encoded two types of viral envelope glycoproteins named *gp41* and *gp120*, which serve a significant role in recognition of receptors present on cell surface (H. Gelderblom, Özel, & Pauli, 1989).

The regulatory or accessory genes of HIV play a major role in modulating the replication of virus. Among these genes *tat* and *rev* are essential regulatory genes which encode trans-activator (TAT) protein RNA splicing-regulator (Rev) protein. Both of these proteins act as RNA binding proteins (Zulfiqar et al., 2017). Besides these functions, Tat proteins are expressed very soon after the infection and it induces the expression of other genes. Rev proteins also allows the allocation of the processed mRNA and HIV genomic RNA from nucleus to the cytoplasm of the cell (Fanales-Belasio et al., 2010). The other accessory-genes negative *regulating factor (Nef), Viral Protein-U (VPU), Viral Protein-R (VPR), Viral Infectivity Factor (VIF),* and *Viral Protein X (VPX)* have multifunction. The Vif, Vpu, Vpr and Nef proteins helps in budding and pathogenies of HIV. These genes also have a major role in replication of virus. Vpx protein is present in HIV-2 instead of Vpu make it less pathogenic compare to HIV-1 (Vicenzi & Poli, 2013).

HIV infections primarily target vital immune cells such dendritic, macrophage, and T helper cells (CD4+ T cells) (Kanta, Unnati, & Ritu, 2011). When the immune cells become infected, they are destroyed through a variety of mechanisms, particularly in case of CD4+ T cells. The helper T-cells become susceptible to pyroptosis and uninfected bystander cells underwent apoptosis, direct viral infection induced cell death and CD8+ cytotoxic lymphocytes plays a destructive role in identification of infected CD4+ T-cells (Lindegaard et al., 2004). Low level of CD4+ T cells result in immunosuppression because of loss cell-mediated

DRSML QAU

immunity and make HIV patients more vulnerable to variety of opportunistic infections, ultimately leading to the development of AIDS (Laila et al., 2019). AIDS become a serious health problem despite the good progress in diagnosis, prevention, and treatment sector (Zak-Place & Stern, 2004).

Many HIV-positive individuals don't even exhibit any obvious signs and symptoms. According to recent research, majority of HIV-positive individuals (between 70% and 90%) practice flu-like symptoms within couple of weeks after infection (Kapila et al., 2016). There are three different phase of HIV infection such as acute phase, chronic phase/asymptomatic phase and AIDS. The early infection stage, called acute phase which develops within two-four weeks after HIV infection. During this stage HIV rapidly multiply and spreads into the body. The symptoms in this stages include headache, rashes, sore throat, vomiting, nausea, diarrhea, fatigue, fever, arthralgia and pharyngitis (Laila et al., 2019). The second phase of HIV infection is chronic phase or asymptomatic phase, or undetectable phase. During this phase HIV multiplied continuously at very low level. Individual in this stage don’t have any HIV like symptoms but can still spread the virus to healthy people. This stage, if last for 4-8 years without Antiretroviral Therapy (ART) may lead to AIDS (T. Li, 2022). Symptoms like lymphadenectasis, fatigue, vaginal candidiasis and oral thrush can also appear during this phase (Laila et al., 2019).

Afterwards, when the body immune system become weakened the infected individual may develop signs and symptoms like weight loss, prolonged diarrhea and fever, dyspnea, cough, sweats, fatigue and swelling of lymph nodes and groin region. It is the last stage of HIV called AIDS, in which the patient CD4+T-cell counts is less than 200/mL (Fanales-Belasio et al., 2010). If these conditions are left untreated, it may lead to other life threating disorders like cancer specifically Kaposi’s sarcoma and lymphomas and many other opportunistic infections like tuberculosis, pneumonia, candidiasis, salmonella infection, toxoplasmosis etc. (Yarchoan & Uldrick, 2018).

HIV infection is transmitted from person to person through numerous ways, but infected blood or blood components are the most common ways of transmission. It can also spread through sexual contact. However, the infection risk is lower in heterosexuals (male to female) as compared to homosexuals (male to male) (Gouw et. al, 2006). Additionally, HIV infection can be transmitted through contaminated needles or syringes, blood, semen, vaginal secretions, breast milk, pregnancy or childbirth. A cut or damaged tissue also allows HIV

DRSML QAU

transmission to the mucous membrane (Khan et al., 2010). Increasing numbers and low awareness of HIV or AIDS among the people lead to persistent HIV epidemics in Pakistan. In order to address such a dangerous health concern, the government must proactive the health policies. Majority of Pakistani population is safe from HIV but infectious diseases can quickly spread across the borders. HIV epidemics emerged a few times ago and infections are quickly spread to a peak rise, therefore regulatory authorities must establish strict rules and regulations to restrict the spread of AIDs.

There are several diagnostic assays are utilized for the identification of HIV infection. The most common laboratory tests used to diagnose HIV/AIDS are HIV antibody tests, HIV RNA nucleic acid assays (qualitative and quantitative), CD4+ T cell counts and HIV medication resistance testing. HIV antibody screening procedures including rapid testing, enzyme-linked immunosorbent assay (ELISA), immunochromatography, Inter-combs test (ICT), immune-fluorescence or chemiluminescent assays (Li, 2022). These techniques are very sensitive and allows HIV antibody detection within 1 or 2 weeks after HIV infection. Prior to HIV antibody detection, an antigenic protein named P24 become visible in HIV infected individual and can be detected. Most of the commercially available approaches for diagnosis of HIV are based on antigen-antibody complex principle which is followed in ELISA and ICT testing (Zulfiqar et al., 2017). ELISA and ICT are routine screening tests that can easily detect HIV in patients. Although ELISA is an expensive and cost-effective technique, but there are several limitations linked to ELISA based detection. The major issue that is associated with ELISA and ICT is false positivity (Shafiee et al., 2015). Other limitations associated with ELISA and ICT methods are the false negative results during initial HIV infection phase of in which antibodies are not in detectable range (Shafiee et al., 2015).

Therefore, numerous other confirmative assays like radio Immunoassay (RIA) and Real Time PCR are recommended to solve such problems. Some laboratories used Radio immunoassay (RIA) as a positive test in which viral radiolabel protein is combined with patient serum (Zulfiqar et al., 2017) (Shafiee et al., 2015). PCR is a very important tool which is employed for amplification and detection of nucleic acids. Nowadays, real-time techniques (PCR-based HIV RNA assays) are used for detection of HIV in which frequently detecting fluorescence are emanated during each cycle of PCR (Bustin & Mueller, 2005). HIV-1 RNA estimation is achieved during first phase of RT-PCR, as compared to conventional PCR or other technologies, and provide more reliable results. Owing to the newly developed assays,

DRSML QAU

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. Through this method detection and amplification of samples are done simultaneously (Patel, 2022). Drug resistance test is used to find out phenotype and genotype of virus. Upon comparison with phenotypic assays, genotypic assay is cheap, rapid and having great sensitivity to detect resistant and wild type viruses. Additionally, genotypic tests are employed globally prior to the start of ART and when the treatment strategy needs to be changed due to an inadequate reduction in viral load or virologic failure following medication (Li, 2022).

From the last 20 years, antiretroviral therapies (ART) have significantly improved the treatment strategies for HIV (Williams, 2013). Before the concept of combination therapy, highly active anti-retroviral therapy (HAART), anti-viral medications were given as monotherapy. This therapy involves giving a mixture of at least three different medications. HAART therapy decrease death rate in patient, inhibit viral replication and also aid to increase the life span of individual but cannot eliminate HIV infection (Persaud et al., 2013). Till now more than Thirty (30) antiviral medications for HIV protection have been identified and licensed (March 2015). These medications are classified as Integrase inhibitors (INSTIs), Protease inhibitors (PIs), nucleoside reverse-transcriptase inhibitors (NRTIs), Non-nucleoside reverse transcriptase inhibitors (NNRTIs), Entry inhibitors (fusion inhibitors and co-receptor antagonists) (Zulfiqar et al., 2017). In order to treat Human immunodeficiency virus scientists also target CCR5, a coreceptor of HIV, by different gene editing tools like CRISPR-Cas9, Transcription activator-like effector nucleases (TALENs) and zinc finger nuclease (ZFN) (Khalili et al.,2015).

HIV vaccine development is particularly very slow due to the extraordinary ability of HIV to evade immunological pressure and number of strain variations. The main focus of early research was to develop such a vaccine that have ability to produce neutralizing antibodies (Mascola et al., 1996). Researchers throughout the world conducted a lot of research on the safety and efficacy of different vaccines. The vaccines produced *in vitro* develops antibodies against HIV strain but failed to produce broadly neutralizing antibody against different HIV variants. Due to the mutating nature of HIV it is challenging and hard to make vaccine for HIV (Zulfiqar et al., 2017).

DRSML QAU

In order 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 eliminated 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.

The *pol* gene of HIV encodes protease (PR), integrase (IN) and reverse transcriptase (RT). Protease is also known as p12, integrase as p32, reverse transcriptase as p51, and RNase H as p15 or reverse transcriptase + RNase H together as p66 (Blood, 2016). Protease play essential role in maturation of virions into infectious form and also help in cleavage of large Gag-Pol precursor proteins into its components. Reverse transcriptase function is to make copy of DNA from viral RNA while Integrase protein helps in the integration of viral DNA into infected cell genomic DNA to produce provirus (Swanson & Malim, 2008). Through a ribosomal-translation frameshift, these proteins are also produced from the same mRNA as the Gag proteins. The precursor polyprotein with a molecular weight of 160 kDa should be cleaved which is important for viability (Bennett et al., 2019). These proteins are not translated as separate unit but a large segment of polyprotein named Pr160gag-pol is formed and these poly proteins are then used as precursor for nascent virions assembly. Monomeric subunits of these enzymes did not show any activity. They can only show its enzymatic activity after oligomerization or dimerization (Babe et al.,1995).

DRSML QAU

When PR is activated, the Pol proteins RT and IN are produced although until now the PR-mediated proteolytic activity is not clear that whether the processing of these proteins takes place, either in cis, in trans, or in combination of both (Pettit et al., 2004). HIV integrase protein is active in homodimer form, although relatively high order of homo-oligomers may contribute to various enzymatic activities. Mature infectious virions comprise RT in heterodimeric form which is made from p66 subunit and p51 subunit. Molecular weight of these subunits is 66 kDa and 51 kDa respectively. The gene responsible for reverse transcriptase only encode large subunit 66 kDa (p66). During the process virion maturation proteolytic cleavage occur between F440/Y441 amino-acid residues of p66 subunit and a small subunit p51 is form (Chattopadhyay et al., 1992). Both of these subunits have same amino acid sequences, but they are different from each other in folding context which lead to asymmetric heterodimeric structure. The p66 domain of reverse transcriptase enzyme in the form of heterodimer is responsible for RNAase

and DNA polymerase activity. The function of p51 subunit of reverse transcriptase is not well understood but it might have main role in structure of virus. However, the RT heterodimers (p66/p51) subunits is necessary for the replication virus whereas RT homodimers (p66/p66) have DNA polymerase and RNase H activities (Abram & Parniak, 2005).

The present study was aimed to investigate the HIV infection in high-risk behavior individuals who were referred for HIV testing. Furthermore, to validate HIV positive (ICT-based) samples by highly specific RNA quantification RT-PCR and to investigate the *pol* gene of HIV isolates for nucleotide sequencing, phylogenetic analysis, and mutational study in the Pakistani population.

Aims and Objectives of Current Study

1. To perform nucleotide sequencing and phylogenetic analysis of *pol* gene of HIV.
2. To study amino acids substitution of *pol* protein structure models from viral isolates circulating in Pakistani patients.

DRSML QAU

CHAPTER 2

LITERATURE REVIEW

DRSML QAU

2 Literature review

2.1 Human Immunodeficiency Virus (HIV)

Human immunodeficiency virus (HIV) is single stranded RNA virus or retrovirus, related to genus *Lentivirus*, subfamily *Orthoretrovirinae* and Family *Retroviridae*. These viruses are called retroviruses because they have the capability to reverse the normal genetic information flow. All biological organism have DNA as their genetic material, which transmits genetic information to RNA, which in turn interprets this information into the exact assembly of proteins, made up of amino acids (Lusso, 2006). Retroviruses use host's DNA and their own RNA genome to make new viral DNA. HIV attacked on immune system and cause the most dangerous and life threating disorder called acquired immune deficiency syndrome (AIDS). HIV is classified into 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.

The basic structure of HIV-1 and HIV-2 are same, however there are some differences in organization of their genome which make HIV-1 more fatal than HIV-2 (Freed, 1998). HIV infection can be transmitted through sexual, perinatal or parenteral routs. A series of intracellular events is initiated after initial attachment of HIV to host immune-cells lead to the production of a wide number of new virus, cause the death of infected cells and eventually led to the destruction of immune system (Honda & Oka, 2006). The helper T-cells or CD4+ cells play a significant function in immune system and are mainly targeted by HIV. When a person is infected with HIV, after months to years the virus kills all CD4 cells and make the immunity of patient incapable of defending itself against illnesses and tumors (Ifeanyi et al., 2017). When an individual is immunocompromised various opportunistic infections like Hepatitis B and C virus, sexually transmitted disease and a verity of co-infections are able to develop ultimately lead to death. Normally, these infections do not cause any serious or fatal health issues (Zahra et al., 2021).

2.2 History of Human Immunodeficiency Virus (HIV)

In Central and Eastern Africa, this disease persisted for very long time. In 1959 the serological evidence of HIV was found in human samples that were stored 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

DRSML QAU

of meat. Moreover, HIV-2 is 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 5 (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). Afterword similar cases were 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-Sinoussi et al., 1983). Robert Gallo and colleagues in May 1984, found enough evidence to persuade the scientific and medical communities that was a novel virus, later it was renamed as human T lymphotropic virus type III (HTLV-III). This new virus was considered as the etiologic agent for the AIDS (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 patient (Levy et al., 1984). The entire genome sequencing of ARV, LAV and HTLV-III viruses was done in February 1985, which defined that they were all the decedents of the similar virus (Ratner et al., 1985).

In year 1986, the name human immuno-deficiency 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 (Clavel et al., 1986), which showed AIDS like clinical symptoms but there were no antibodies detected against HIV which becomes the finding of a another virus named as human immune-deficiency virus type 2 (HIV-2). The first identified/old HIV was named again as human immune-deficiency virus type 1 (HIV-1). In 2008, Nobel prize in Medicine was awarded to Francoise Barré-Sinoussi

DRSML QAU

and Luc Montagnier (Pasteur Institute) for the discovery of HIV-1 virus. HIV-1 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 demise for U.S. men ranging from age 25 to 44 and that year 2,332 cases reported in San Francisco (Padamsee, 2020).

2.3 Classification of HIV

HIV is classified into HIV-1 and HIV-2 whereases HIV-1 spread all over the world. Relatively HIV-2 less common and mostly prevalent in West Africa and is seldom detected elsewhere. HIV-1 strains can be categorized into M (major), O (outlier), and two recently identified N and P groups. These four (4) groups might reflect four different simian immunodeficiency virus invasions into humans (Chalmet et al., 2010). The only confined HIV-1 group to central and western Africa is Group O. In 1998, N group of HIV-1 was discovered which is rarely found in Cameroon. A novel strain of virus closely related to gorilla Simian-Immunodeficiency Virus was found in a Cameroonian woman in 2009 which was chosen as group P of HIV-1. Globally more than ninety percent (90%) infections are caused by subtype M of HIV-1 virus (Rachinger et al., 2011).

Based on genetic variation Group M of HIV-1 is classified further into nine (9) subtypes/clades including subtype A, B, C, D, F, G, H, J and K (Fox et al., 2010) as shown in Figure 1. Inside the infected cell of individual, a hybrid form of virus termed as circulating-recombinant forms (CRFs) is generated after the collision and combination of genetic material of different viruses from different subtypes (Gross et al., 2004). For example, when subtype A is genetically combine with subtype B it make a new strain CRF A/B (Chalmet et al., 2010). It has been reported that there are more than 70 distinct epidemiologically stable CRF exist and new CRF are likely to emerge. Statistical analysis shows that 1 out of every 400 newly developed virus is recombinant virus (McCutchan, 2000). HIV strains classification into subtypes/clades and CRFs is a very complicated and the classifications are always susceptible to change after the discovery of new information. Instead of using the terms A and F, some scientists refer to the subtypes A1, A2, A3, F1 and F2, while others consider the former to be sub-subtypes (Ifeanyi et al., 2017).

DRSML QAU

The geographical distribution of different subtypes of HIV-1 includes: Subtype A is found in Sub-Saharan Africa in Central Africa. The Subtype B infected areas include Japan, Brazil, Caribbean, South America, India, Thailand, United States and Europe. Subtype C is prevalent in South Africa, India and Brazil while Subtype D is prevalent in sub-Saharan Africa and Central Africa. The Subtype E is found in Southeast Asia, Central African Republic and Thailand. The prevalence of Subtype F occurs in Democratic Republic of Congo (Zaire), Romania and Brazil. Similarly, Subtype G is prevalent in Central Africa, Gabon, Democratic Republic of Congo (Zaire), Russia and Thailand while Subtype H is predominated in Gabon, Democratic Republic of Congo (Zaire), Central Africa and Russia. Subtype I is present in Cyprus while Subtype O is present in Gabon and Cameroon. These subtypes are unequally dispersed all around the world. Currently, more than half of new HIV infections worldwide are caused by subtype C. The majority of subtypes are found in Africa; however, subtype B is less common (Ifeanyi et al., 2017). HIV-2 is usually less pathogenic usually than HIV-1 and so far been categorized into ten (10) distinct groups that are A, B, C, D, E, F1, F2, G, H, and I. Among these group A and group B are prevalent while different lineages have been hypothesized from group A of HIV-2. Moreover, no known HIV-2 subtypes have been discovered yet (Visseaux et al., 2021).

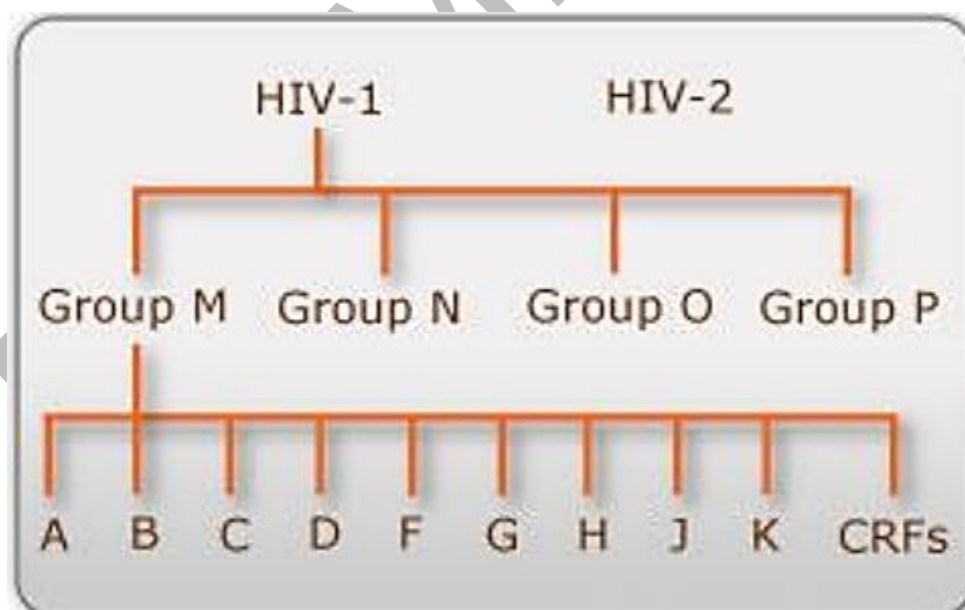

Figure 1: Graphical demonstration of HIV types and subtypes classification

HIV is categorized into HIV-1 and HIV-2. HIV-1 is further Grouped into M, N, O and P. Subtype M is further classified into A, B, C, D, F, G, H, J, K and CRFs (Fox et al., 2010).

QAU

2.4 Epidemiology of HIV-1

2.4.1 Africa

HIV-1 was initially originated from Chimpanzees in West of Central Africa and became one of the largest global public health issues 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 and AIDS. annually 0.73 million new HIV-1 infections occur because of 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 per day in infected patient. Consequently, the diagnoses, care, treatment and interventions against 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 predominant in Central and Western Africa, while in Southern and Eastern Africa, subtype C has predominated (Giovanetti et al., 2020).

2.4.2 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 rare 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. 10 cases most of them were injecting drug users and majority of those infected heterosexually are women (Hamers & Downs, 2003).

DRSML QAU

2.4.3 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).

2.4.4 Australia

In Australia, the diagnosis number of new HIV cases has been increased about 26 percent and the value of inhabitants has reached to 5.4/100,000 in 2012. HIV is mainly transmitted through sexual interaction among men. Subtype M:B of HIV-1 is mostly common in Australia. However, several other subtypes of HIV-1 are also identified which is mainly transferred from Asia. Lower pathogenicity and a defect in Nef region of subtype B have been identified in the beginning of epidemic (Churchill et al.,2007).

2.4.5 Asia

In Asia, HIV was reported quite late compared to other 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, Brown, & Phanuphak, 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 (3) groups for HIV prevalence: The first (1st) 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, Sittitrai, Vanichseni, & Thisyakorn, 1994). The second (2nd) group comprises those areas which were in transition and growing markedly in the 5 years including China, Nepal, Vietnam and Indonesia. The third (3rd) 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

DRSML QAU

determined mainly by sex workers, injecting drugs, 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).

2.5 Prevalence of HIV in Pakistan

It was estimated that approximately 150,000 children and adults 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). Seven outbreaks of HIV occur in Pakistan during last two decades (from 2000-2019). Out of these 7 outbreaks 3 occur in Sindh while 4 outbreaks occur in Punjab (Rabold et al., 2021). 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 (A. A. 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). In Pakistan HIV-1 subtype A is the most prevalent subtype (Rai et al., 2010).

In 2019, a high prevalence of more than 13% HIV cases have been reported from Kot Imrana, in 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) percent, 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 largest number of HIV positive cases (Ali et al., 2017).

DRSML QAU

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 seroprevalence of HIV. The results revealed that out of 2,062 individuals, 329 (16%) were HIV positive. Among all the positive in decidual, 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%), Chiniot (17%) ,D.G Khan (Dera Ghazi Khan) (21%), and Rawalpindi (19%) (Karim et al., 2020).

During the last ten months of 2022, 9,773 individuals in Pakistani were tested positive for HIV. The transference of HIV from infected populations to general public is clearly demonstrated, casting major doubt on HIV preventive and control efforts. According to the news, Punjab leads the way in new HIV infections with 6,106 positive tests, Sindh has 2,097 new cases while Khyber Pakhtunkhwa has 815 new HIV cases reported from January to t October 2022. Similarly, in 2022 from January-October, about 496 new cases of HIV were reported from Islamabad Capital Territory (ICT) whereas 316 new HIV cases were recorded from Balochistan (Ahmed, 2022). Looking, to all these numbers, which is highly threating to Pakistani as well as international community, therefore a rapid diagnostic services, public education efforts and the establishment of treatment facilities is required in all around the country.

2.6 Morphology of Human Immunodeficiency Virus (HIV)

2.6.1 General structure

By using negative staining of electron microscopy, HIV-1 and HIV-2 were thoroughly studied. A massive number of 130-200nm particles comprising 130nm long and 30-70 nm wide core of HIV-2 has been observed in culture. Morphologically the core is pear or conical shaped as shown in Figure 2. 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, SIV have projection like structure while HIV-2 comprises a knob like structure. The HIV virus is consists of envelope, inner membrane, and bullet like structure called core comprises genetic material (Chrystie & Almeida, 1988).

DRSML

2.6.2 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 play a vital role in attachment, adsorption, and entry of HIV into the host cell. The envelope of HIV is composed of 72 knobs comprise of transmembrane gp41 (TM) and surface gb120 (SU) trimer proteins (H. R. Gelderblom, 1991). It has 4.8nm in diameter head, 3.2nm long and 1.0nm wide stalk. The length of HIV varies with respect to the observable width of the envelope (Goto et al.,1998).

2.6.3 Matrix

The envelop of HIV cover a symmetrical exterior capsid membrane called Matrix contain matrix protein (p17 or MA). Matrix basically has ground equivalent substances which supports envelope and the core. In 1984, Bouillant and Becker described the matrix as an electron-lucent layer or region (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).

DRSML QAU

2.6.4 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 has 2 dimensional shapes which are bar like, triangular and sectorial (H. Gelderblom et al., 1988). Protein p24 is present in the core of HIV1. The core of HIV may be made of about 380 units, even though the size of these cores may differ (Ehrlich, Liu, Scarlata, Chu, & Carter, 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). Inside the capsid/core two identical viral gRNA molecules are present. The core also contains viral enzymes including integrase(IN) and reverse transcriptase (RT)/RNase H. During the maturation of virions, oligopeptides produced after release from the cell during the proteolytic processing of the pre-cursor proteins (p55, p160) are also found in viral particles (Luciw, 1996).

2.7 Genome organization of HIV

The human immunodeficiency virus has approximately 9.8 kb genomic length and comprises of two duplicates of positive sense, unspliced, 5′ capped and 3′ -polyadenylated single stranded RNA molecules (Rajarapu, 2014). They have nine overlapping open reading

frames (ORF) which codes for about fifteen viral proteins as shown in Figure 3. The HIV is made up of accessory, regulatory, and structural genes. All retroviruses have the structural genes gag, *pol*, and *env*, which code for structural proteins. Matrix proteins p17, core proteins p24, p7, p6, and p15 (gag) are all produced by three separate precursors by post translational cleavage. The *pol* gene HIV encode for key regulatory enzymes such reverse transcriptase (RT), integrase (IN) and protease (PR) (Di Nunzio, 2018a). Envelope glycoproteins gp41 and gp120 are present on the surface of virus which interact with the receptors of the cell (H. Gelderblom et al., 1989). Important regulatory genes include TAT and rev. The TAT gene produces the TAT protein very soon after infection and encourages the production of other HIV-related genes. The export of suitably processed mRNA and genomic RNA from the nucleus to the cytoplasm is made possible by the rev protein, which is produced by the rev gene. The additional accessory genes include Nef, VPU, VPR, VIF, and VPX. The viral cDNA is made easier to enter into the host nucleus, where it completes the integration of the virus into the host genome, due to the VPR proteins. The VPU proteins and other small proteins then interact to release the viral particle. By targeting the antiviral activity of the human enzyme "APOBEC3G" or A3G for cellular degradation and ubiquitination, the VIF protein improves the infectiousness of viral offspring (Rücker et al., 2004). The negative regulatory protein (NEF) performs a variety of tasks, including suppressing CD4 receptor expression on the surface of infected T-cells, improving viral replication, increasing virions release, and controlling the cellular signal transduction of the host cell (Lusic & Siliciano, 2017).

DRSML QAU

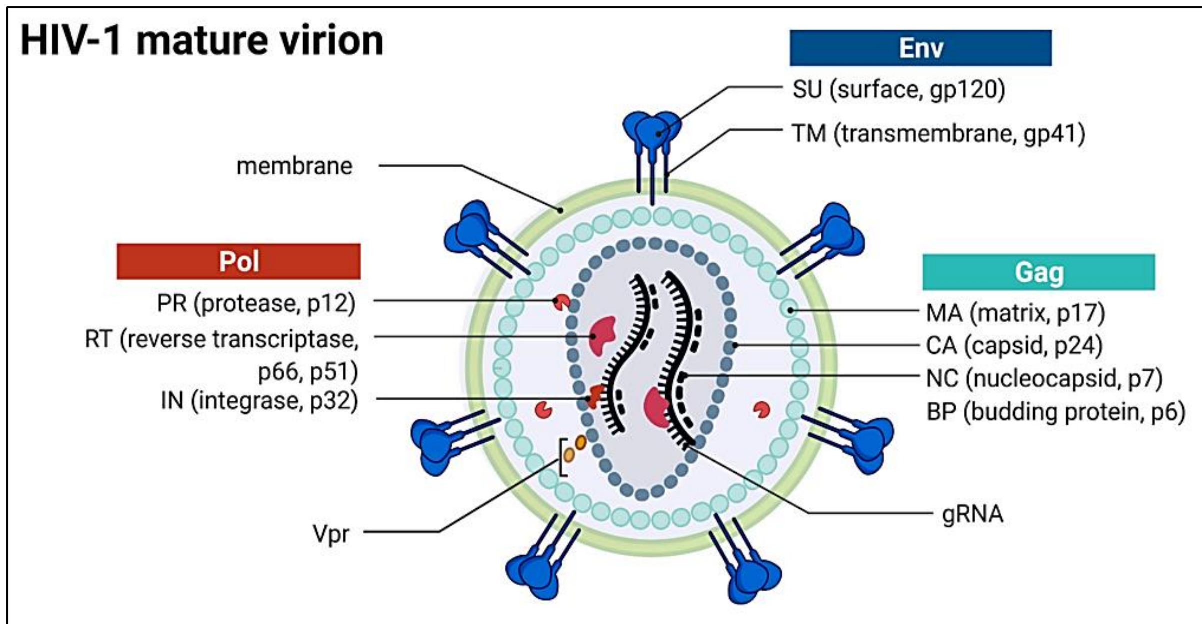

Figure 2: Morphology of Human Immunodeficiency Virus

The dark blue color represented konb imbedded in membrane (light green color) of HIV which enclosed Matrix (light blue color). A bullet like shape (dark gray color) inside the matrix represent capsid, red and yellow color represent viral proteins. Inside the capsid RNA is represented in dark color and other proteins are represented in dark red color (van Heuvel, Schatz, Rosengarten, & Stitz, 2022).

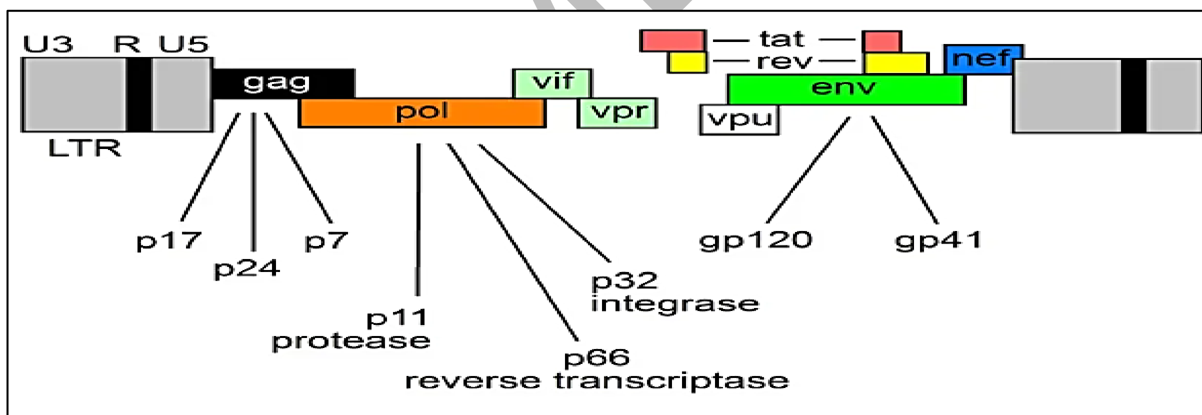

Figure 3: Demonstration of genomic organization of HIV

Light gray color represent LTR, Black color represent gag, Orange color represent pol, light green color represent vif and vpr, white color represent vpu, light red and yellow color represent tat and rev, green color represent env while blue color represent nef gene of HIV-1 (Hoffmann et al., 2003).

2.8 Life Cycle of HIV

The HIV replication cycle employs host cells, which is why they are referred to as intracellular parasites. HIV-1 has a complicated life cycle that may be generally separated into two (2) phases: Early replication and late replication phase. The early phase of HIV starts when the virion attaches to the surface of the cell and ends when proviral DNA is integrated in the host genome. After integration proviral transcription starts and late replication begins which is terminated when fully infectious virions are released. Extremely activated CD4+ T-cells have a short HIV life-cycle of one (1) to two (2) days, and both virally infected and non-infected recipient cells die as a result of programmed death (Kirchhoff, 2013).

2.8.1 Viral entry

Infected people often only have a 20-30 minutes half-life for cell-free HIV virions. Therefore, the virus needs to find a fresh target cell quickly and infect it. Macrophages, dendritic cells, and T-helper cells play a significant role in controlling the humoral and cellular response to any infection. HIV also targets immature dendritic cells, macrophages, and other T-cell subsets, however these cell types do not participate in virus replication (Kirchhoff, 2013). CD4 act as primary receptor for HIV-1 which is found on the surface of T lymphocytes, monocytes, dendritic cells, and macrophages (Deeks et al., 2015). 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 which is gp120 (glycoprotein 120) and gp41 (glycoprotein 41) (Freed, 2001). After HIV binds to CD4 cells, Env trimer conformational changes occur, allowing gp120 to engage with co-receptors CCR5 (R5) or CXCR4 (X4). Distinct T cell subsets express CXCR4 and CCR5 in different ways. CXCR4 is found on both naive and memory T-cells, but CCR5 is substantially expressed on only memory T lymphocytes. CCR5 is expressed by both dendritic cells and macrophages (Deeks et al., 2015). This conformational changes in gp41 brought about by the host cell attachment to the HIV promote membrane fusion between the host cell membrane and the lipid bilayer of HIV, allowing the viral core to enter the cytoplasm. (Zhou et al., 2004).

2.8.2 Post viral entry step

Early research claimed that uncoating can happen right away after viral penetration (Freed, 2001). The integrity of the capsid, according to more recent study, is crucial for HIV-1 infection and may last for several hours (Kirchhoff, 2013). Upon fusion to the plasma

DRSML QAU

membrane of host cell through viral envelope, the virus core may enter the host cytoplasm. The core (structure left over after the lipid bilayer is gone) is transformed into the reverse transcription complex (RTC), followed by pre-integration complex, during the "uncoating" process (PIC). While CA (capsid protein) appears to be lost throughout these phases, some of MA (matrix protein), NC (nucleocapsid protein), *pol* encoded enzymes RT and IN, as well as the auxiliary protein Vpr, stay attached (Freed, 2001). Additionally, it seems like uncoating is intimately related to reverse transcription and the translation from RTCs to PIC, which can incorporate viral DNA into host genome. Nevertheless, the precise moment of uncoating process is not well understood and is the topic of in-depth research (Arhel, 2010).

2.8.3 Reverse Transcription

The genome of HIV comprises two (2) nucleocapsid-protected positive (+) sense single stranded RNA molecules. Reverse transcription occur after getting entry of virus to the cytoplasm of host and both single stranded of RNAs are converted into dsDNA (Kirchhoff, 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).

2.8.4 Nucleus import

The viral genome (RNA, then DNA) stays connected to the high molecular weight RTC during reverse transcription. Eventually, as part of PIC, the viral DNA is taken into the nucleus. The PIC also contains the *pol*-encoded enzyme IN, which naturally has to be in the nucleus at the time of integration. In some study it is reported that this complex also comprises the Gag NC protein and auxiliary protein Vpr which are thought to contribute in nuclear import and also help in integrase activity (Gallay et al., 1997). Due to their size, HIV-1 PICs cannot passively diffuse through the nuclear pore; instead, they must be actively carried through it (Zennou et al., 2000). Nuclear entry was first thought to be mediated by the Vpr, viral matrix protein, and integrase. Recent studies suggest that the capsid protein may be an important component in this process and that none of these proteins are required for infection of non-dividing cells, however the underlying mechanisms are still unclear (Kirchhoff, 2013).

DRSML QAU

2.8.5 Integration

The viral genome is inserted into the host cell for successful infection and the production of its genes after being produced as linear dsDNA and transported across the nuclear membrane. The viral DNA integration into host DNA occur with the help of integrase enzyme (Anderson & Maldarelli, 2018). The integrated virus or viral genome in the host genome is known as the provirus, and it continues to infect the host cell for the rest of its life. The viral genome that has been incorporated into the host functions much like a biological gene. The host DNA and this proviral DNA work together to replicate the virus. The proviral DNA is multiplied or new cells are infected for the viral infection to spread (Chiu & Davies, 2004). Given that mutants frequently fail to transmit infections, integration is a crucial stage in retroviral replication. The integrated provirus in some cells might remain dormant or silent for years, which can be a significant barrier to virus eradication. The immune system is unable to identify it in its dormant form, making it impossible for it to be eliminated. The C-terminal domain, the central core, and the N-terminal with a zinc finger are the three domains of the integrase enzyme (Goodsell, 2015).

2.8.6 Gene Expression

The provirus after integration provides a template for viral RNAs, which code for all structural, regulatory and accessory proteins of HIV. Utilizing the polymerase enzyme of the host's cellular machinery, provial DNA act as a template viral for mRNAs and genomic RNA transcription (Wu & Marsh, 2003). At 5' LTR U3 region is located which comprises several cis-acting components necessary for the synthesis of RNA and initiates the transcription process of the provirus. The cellular transcription factor greatly influences how the virus expresses its genes (NFAT and NF-kB) (Gaynor, 1992). Due to the ineffective viral transcripts, the transcriptional output is initially minimal, and the viral trans-activator protein TAT is needed for successful expression of viral gene. The trans-acting response (TAR), of the R region in the 5' LTR is where the TAT attaches to speed up transcription. Large quantities of the three different types of viral RNAs are produced as a result. Rev, TAT, and Nef are all encoded by small, spliced mRNAs. VIF, VPU, VPR, and Env proteins are coded by partially spliced mRNAs, and Gag, *Pol*, and Gag polyprotein precursors are produced by unspliced mRNAs (Gatignol, 2007).

DRSML QAU

The TAT proteins play a crucial part in the viral transcription of RNA elongation, the Rev protein transports the viral RNA that has not been spliced to the cytoplasm, and the Nef protein performs a variety of tasks, one of which is to make the infected cell invisible to the host immune system. The TAT and Rev enable the synthesis of full-length mRNAs that code for the production of the Gag and Gag precursors, which are then processed for the synthesis of important structural and enzymatic proteins (Jeang et al., 1999).

2.8.7 Viral Particle Production

Gag and Gag-Pol precursors can multimerize as a result of interactions with Gag proteins. The two precursors are also found in lipid rafts in the inner leaflet of the plasma membrane, where the matrix domain of the two precursors is N-terminally myristoylated. (Ganser-Pornillos et al., 2008). They are drawn to these building platforms by the interactions between the matrix protein and the viral Env glycoproteins. Finally, two copies of the viral genome are attracted to this complex by the stem loops packaging signal of the viral RNA interacting with the zinc fingers present in the Gag NC domain. The viral Vif protein, among other biological components, is attracted to the sites of virion assembly and incorporated into viral particles where it blocks the restriction factor APOBEC3G. As viral proteins and RNA accumulate in the plasma membrane, the membrane first becomes coiled, and a membrane-coated spherical particle subsequently forms (Kirchhoff, 2013).

2.8.8 Budding

Budding is a process through which infected cells release their progeny viruses. This makes possible for newly produced HIV to pinch off and circulate. The assembly of many viral proteins precedes the start of the HIV virus development. The Gag precursor polyprotein is the major actor in the viral assembly. This step involves the end domain of the Gag (p6 part) and the cellular Tsg101 protein. These proteins contain elements that direct the viral particles to the cell membrane, support Gag-Gag interactions, encourage cell budding, and enclose the HIV genome for association with the Env viral glycoproteins (Von Schwedler et al., 2003).

DRSML QAU

2.8.9 Maturation

The viruses are visually distinguished by a thick layer of Gag and Gag-polyprotein precursors once they are liberated from the host plasma membrane in immature (non-infectious state). After the virus has finished budding, its proteases quickly come into action, cleaving the precursors of Gag and Gag-Pol before maturing their ultimate forms. As a result, the virus morphology changes to resemble a doughnut (Hikichi & Freed, 2021).

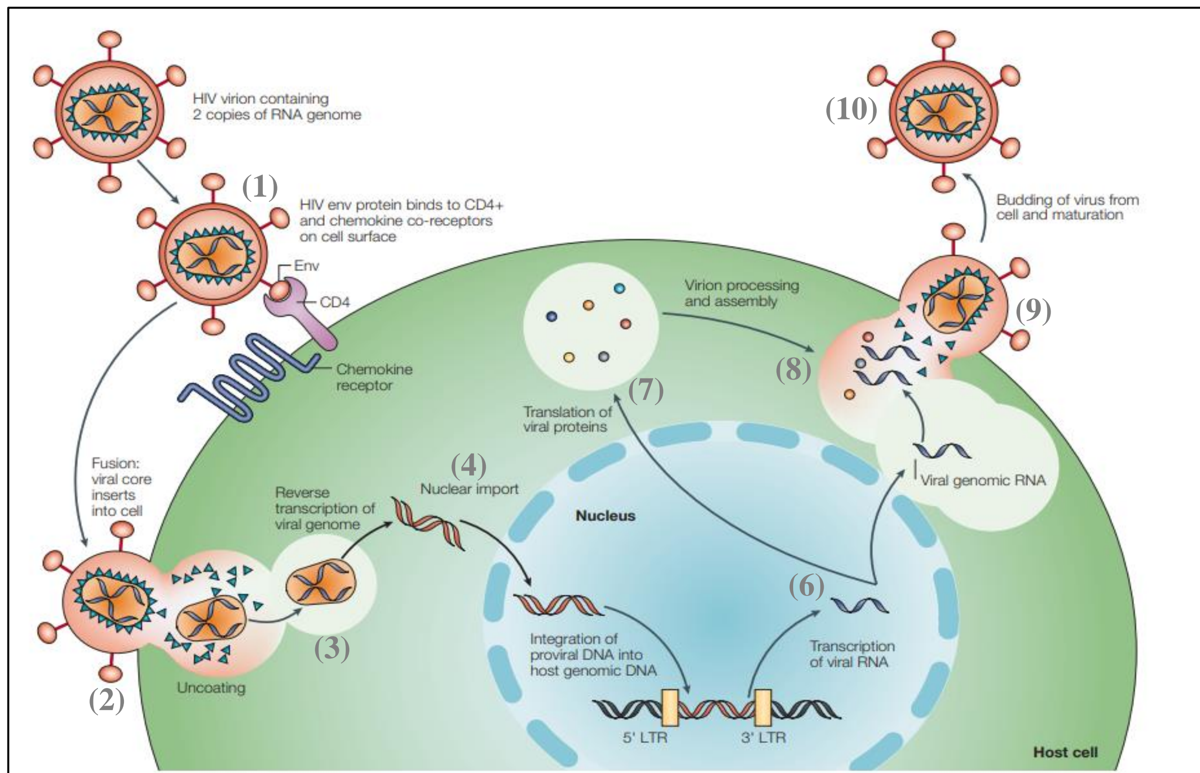

Figure 4: **Summary of Viral Replication Cycle**

The different steps involved in life cycle of HIV includes 1) Attachment, 2) insertion, 3) Reverse transcription, 4) Nuclear import, 5) Integration, 6) Transcription, 7) Translation, 8) Packing and Assembly, 9) Budding and 10) Maturation (Rambaut, Posada, Crandall, & Holmes, 2004)

DR

2.9 Transmission

HIV transmission is influenced by the host biological characteristics, the quantity and concentration of the virus in the bodily fluids that are exposed, and the degree to which the host is immunologically and cellularly vulnerable. A single virus can cause the initial HIV infection, which means that it can provide AIDS with a window of opportunity. HIV can spread by intravenous drug use, blood transfusion, sexual contact with contaminated blood, mother-to-

child transmission, organ transplantation, and sexual interaction (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).

2.10 Pathogenesis

The two major factors that promote 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).

HIV inhibits the immune system by two main mechanisms, the first of which is the destruction of CD4 and T cells, which ultimately results in immuno-deficiency. To enter the host cell, the virus uses its surface proteins, gp41 and gp120, to bind to CD4+ T cells. The chemokine receptor, a crucial co-receptor for HIV-1, is the second pathway through which HIV enters cells (Goto et al., 1998). There are two main chemokine receptors that can be used by

DRSML QAU

virus to enter into the host cell. These receptors are CCR5 and CXCR4 or Fusin (Speck et al., 1998).

The three main groups of HIV strains can be subdivided based on cellular tropism. These are known as R5 viruses with a non-syncytium inducing phenotype that infect PBMC (peripheral blood mononuclear cells), T lymphocytes, macrophages, and monocytes through CCR5 and not T-cell lines. These are present throughout all infection stages (Naif, 2013). The second is T-cell line tropism (T-tropic), also known as X4 viruses, which employ CXCR4 as their primary co-receptor to infect T lymphocytes and T-cell lines but are unable to infect monocytes and macrophages, delaying the course of AIDS. People that lack CCR5 expression are susceptible to X4 viruses but are really resistant to R5 viral infection (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). Other chemokine co-receptors, such as CCR2b and CCR3, can also serve as sites of attachment (primary or secondary) for HIV-1 and HIV-2 isolates, but they frequently are not engaged in infection. Additionally, in the brain, vagina, and intestine regions, Galactosyl ceramide (GalC), in addition to CD4, might potentially be a significant binding site for HIV-1 infection (Levy, 2009). Additionally, when HIV forms a complex with antibodies, complement receptors and Fc domains on T cells, macrophages, and other immune cells are easily accessed (Coffin & Swanstrom, 2013).

DRSML QAU

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). The immune system begins to deregulate as an acute HIV infection develops into AIDS, making the person more vulnerable to opportunistic infections. The increase in infected macrophages and CD4+T cells complements this. Although HAART (highly active antiretroviral therapy) is well known for reducing opportunistic infections when immune reconstitution takes place, neither HAART is widely accessible nor completely effective (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).

2.11 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 degrade the RNA strand of hybrids of DNA/RNA and after each replication round there is one substitution per HIV genome. Reverse transcriptase exhibits a nucleotide substitution rate of roughly 4-10 per nucleotide each replication cycle.

2) It has been claimed that the HIV-1 replication mechanism is extremely quick, with an infected person producing 109 virions on average every day.

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 (Fanales-Belasio et al., 2010).

2.12 Treatment of HIV

2.12.1 Antiretroviral therapies (ART)

Up to March 2015, more than 30 antiretroviral medications had been approved for the treatment of HIV infection. These medication are divided into different classes as given below (Zulfiqar et al., 2017).

DRSML QAU

2.12.1.1 Nucleoside reverse transcriptase inhibitors (NRTIs)

NRTIs were the first class of drugs approved by the FDA (Arts & Hazuda, 2012). Nucleoside reverse transcriptase inhibitors (NRTIs), also called nucleoside analogues, work by inhibiting the RT enzyme of HIV (or "nukes"). NRTIs compete with natural cellular nucleosides and act as substitute substrate (Zulfiqar et al., 2017). The absence of a OH group at 3′ end of the deoxyribose sugar in these nucleoside analogues hinders phosphodiester coupling with incoming 5′ nucleoside triphosphates and NRTI, which will ultimately result in the termination of the developing viral DNA chain (Arts & Hazuda, 2012). These analogues are regarded as pro-drugs because they undergo cellular kinase-mediated phosphorylation after endocytosis to transform into active metabolites. There are Eight FDA approved NRTIs drugs are available up to date in which include; zalcitabine (Hivid, ddC), stavudine (Zerit, d4T), zidovudine (Retrovir, AZT), abacavir (Ziagen, ABC), didanosine (Videx, ddI), emtricitabine (Emtriva, FTC), lamivudine (Epivir, 3TC), and a nucleotide RT 3′ inhibitor, Tenofovir disoprovil-fumarate (Viread, TDF) (Tang & Shafer, 2012). Some long-lasting side effect are also linked with these NRTIs including myelotoxicity, Polyneuropathy, lactate acidosis and pancreatitis etc. (Persaud et al., 2013)

2.12.1.2 Non-Nucleoside Reverse Transcriptase Inhibitors

The first NNRTI inhibitors that target the reverse transcriptase (RT) enzyme and nucleoside analogue were first described in 1990. NNRTI binding to RT and the creation of a hydrophobic pocket at the active site result in the suppression of HIV RT enzyme. Reduced polymerase activity and alterations in the spatial configuration of the substrate-binding region occur when NNRTIs bind to the RT enzyme (de Béthune, 2010). The NNRTI pocket enables the construction of highly selective inhibitors with minimal side effects and lower toxicity. Unlike active site or dNTP binding site, hydrophobic pocket does not need to be conserved in order to carry out enzymatic activity. due to the presence of resistance mutation sites, rapid mutations development and nonoverlapping inhibition mechanisms in NNRTIs, clinicians are encouraged to use this medication in conjunction with NRTIs (Das & Arnold, 2013).

The disruption of enzyme and inhibitor interaction is caused by NNRTI mutations which confer resistance in three different ways: NNRTIs binding pocket prevent the entry of inhibitor due K103N mutation, NNRTIs become less specific toward inhibitors due to Y188L mutation that alter the conformation and size of binding pocket, affecting the interactions

DRSML QAU

among residues of enzyme and inhibitor in the binding pocket due to Y181C mutation (Asahchop, Wainberg, Sloan, & Tremblay, 2012). The FDA approved NNRTI drugs to treat HIV infection include Efavirenz (EFV,Sustiva®), nevirapine (NEV, Viramune®), Rilpivirine (RPV, Edurant®), delavirdine (DLV, Rescriptor®) and etravirine (ETR, Intelence®) (de Béthune, 2010).

2.12.1.3 Integrase Inhibitors

The recent HIV-1 enzyme effectively targeted for therapeutic research is integrase enzyme. These inhibitors change how the enzyme interacts with divalent (Mg2+) cations and integrates correctly with the host genome by hooking up near to the enzyme's active site. (Arts & Hazuda, 2012). The strand transfer and 3′ processing steps of the integration process are carried out in tandem by the integrase enzyme and Mg2+ or Mn2+ divalent ions, which are supplied by the amino acids D64, D116, and E152 (Wainberg, Mesplède, & Quashie, 2012). When reverse transcription occur a double stranded DNA is formed which is cleaved by integrase enzyme at 3’ end in conserved dinucleotides lead to the formation of overhangs on both side genome. This process is known as 3' processing reaction. In strand transfer reaction, the confined integrase enzyme integrates the double-stranded viral DNA into the host genome at 3' overhangs before being transferred to the nucleus. Although the integrase enzyme catalyzes both of these reactions, the effective drugs used to treat HIV are those that can prevent the strand-transfer reaction. These drugs are known as integrase-strand transfer inhibitors (INSTIs), and they are attached to the enzyme that is already binding viral DNA (Ananworanich & Robb, 2014). INSTIs is made up of two parts: a hydrophobic group and a metal-binding pharmacophore. Magnesium from the active site is hidden by the metal-binding pharmacophore, and the hydrophobic group in the compound interacts with viral DNA and enzyme DNA. Raltegravir and MK-0518 were approved by the FDA in 2007, while GS-9137 and Elvitegravir inhibitors are still undergoing clinical testing (Arts & Hazuda, 2012). In close proximity to the trio of amino acids that aid in coordination with necessary magnesium cofactors, in the active site of the integrase enzyme, are mutations that confer resistance to INSTIs. These changes have a detrimental effect on viral replication and enzymatic activity (Hare et al., 2010).

DRSML QAU

2.12.1.4 Protease Inhibitors

The HIV-1 protease enzyme is the target of a different family of inhibitors known as protease inhibitors (PIs). The protease enzyme is crucial for viral maturation as well as for accelerating the digestion of precursor gag-pol polyprotein (Ananworanich & Robb, 2014). In the beginning, due to minute size and key role of protease in the life cycle of HIV-1, it was assumed that it would seldom undergo mutations. However, the protease gene showed significant flexibility, with more than twenty (20) amino acid substitutions and forty-nine (49) codon polymorphism was detected, both of these is associated to resistance (Arts & Hazuda, 2012). The primary resistance mutation in most PIs occurs in the enzyme's active area, causing amino acid changes that have a detrimental effect on the viral replicative fitness. Changes in the eight (8) primary cleavage sites, in addition to alterations in the protease gene, result in resistance to PIs (Zulfiqar et al., 2017). Ten (10) protease inhibitors (PIs) have been approved by FDA include: darunavir (TMC114, Prezista), atazanavir (ATZ, Reyataz), amprenavir (APV, Agenerase), fosamprenavir (Lexiva), lopinavir (LPV), indinavir (IDV, Crixivan), tipranavir (TPV, Aptivus), saquinavir (SQV, Fortovase/Invirase), nelfinavir (NFV, Viracept) and ritonavir (RTV, Norvir) (Zulfiqar et al., 2017).

2.12.1.5 Entry Inhibitors

The HIV-1 gp120 protein binds to the CD4-cell surface receptor, allowing HIV-1 to enter the host cell. After interaction with CD4, gp120 contains a structural element that is exposed and can interact with one of the two co-receptors on the host. This enables the gp41 transmembrane subunit to integrate into the cell membrane, resulting in the fusion of the cell membrane with the viral particle. Numerous medications have been developed to stop HIV-1 from entering cells. These medications include ibalizumab, enfuvirtide, and maraviroc (MVC) (Z. Li et al., 2013; Tilton & Doms, 2010). Entry inhibitors are further divided into different classes on the basis their ability to interrupt several stages of the HIV-1 life cycle (Arts & Hazuda, 2012).

Fusion Inhibitors

HIV-1 fuses to a host cell's cell membrane when two homologous regions of the viral protein gp41 come into contact. Heterologous proteins called fusion inhibitors imitate one of these domains in the viral protein, preventing it from interacting intramolecularly with other molecules. Alpha helical peptides imitate the leucine zipper domain and have potent antiviral

DRSML QAU

action (Arts & Hazuda, 2012). The only licenced fusion inhibitor that prevents the creation of gp41 hairpins is enfuvirtide. Two complementary sections of gp41 fold into one another to generate hair pins, shrinking the protein and fusing the membranes of the host and the viral cell (Ananworanich & Robb, 2014).

Attachment Inhibitors

The interaction between the gp120 and CD4 that is inhibited by attachment inhibitors stops the envelop from connecting with co-receptors (CXCR4 and CCR5), limiting the fusing of the viral and cellular membranes. These inhibitors are more effective against CXCR4 and CCR5 receptors of tropical strains than other chemokine receptors. BMS48804, TNX-355, and PRO542 are attachment inhibitors that have received FDA approval (Volberding, Sande, Greene, Lange, & Gallant, 2008).

Small Molecule CCR5 Antagonists

Other class of fusion inhibitor that successfully blocks HIV-1 entrance into the host cell is CCR5 antagonists. C-C chemokine receptor type-5 (CCR5) is expressed on the surface of variety of cells, including dendritic cells, macrophages, memory T-cells, astrocytes, fibroblasts, vascular smooth muscle, etc. The CCR5 receptor protein measures 40.6 kDa and comprises three hundred and fifty two (352) amino acids (Barmania & Pepper, 2013; Samson et al., 1996). The most crucial HIV co-receptor, CCR5 helps the HIV virus enter cells and spread from one cell to another (Lopalco, 2010). They work allosterically by preventing HIV-1 gp120 from attaching to host cell CCR5 (Tang & Shafer, 2012). When receptor antagonists bind to the hydrophobic pocket of CCR5, it undergoes conformational change which makes it challenging to identify by the viral protein gp120 (Arts & Hazuda, 2012). The major function of CCR5 replication of HIV and its genetic diversity in HIV resistance and illness control motivate researcher to investigate it receptor to treat HIV. Three CCR5 antagonists, Aplaviroc, MVC, and VCV, are currently available and can prevent HIV transmission in humans (Alkhatib, 2009; Contento et al., 2008).

2.12.1.6 Shortcomings of Recent Anti-Retroviral Therapy

The risky side effects of anti-HIV medications, their toxicity and detrimental impact on life quality, drug resistance limiting the of future treatment options of HIV, interactions of drugs, high costs, the limitations of current existing regimens, and treatment failure are all reasons to argue against early treatment in asymptomatic patients (Sendagire et al., 2009).

DRSML QAU

It is yet unknown when to start anti-HIV therapy. HIV-specific antiviral drugs also have an impact on host cells. Along with the HIV genome, they may also destroy the nuclear material of host cell. The deficiency of cellular DNA polymerase is primarily responsible for nucleoside reverse transcriptase inhibitor toxicity. Two very harmful and dose-dependent adverse effects include anaemia and neutropenia (Ramana, Anand, Sethuraman, & Krishnan, 2014).

Moreover, no vaccine against HIV-1 is available yet, these drugs are used in combination to prevent HIV in different stages of replication. HAART, or highly active antiretroviral therapy, is the standard of care and calls for the administration of three medicines from two classes of antiretroviral agents (McManus et al., 2019). The selection of strains resistant to one or more medication classes, as well as existing treatment associated toxicity might compromise efforts to target distinct stages of the viral reproduction cycle (Kaur, Sharma, Gupta, Ntie-Kang, & Kumar, 2020). 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 is 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).

2.12.2 Medicinal plants using for HIV treatment

Natural resources, such medicinal plants, are still chosen over synthetic remedies for the treatment of infectious and noninfectious diseases. It is often believed that medicinal herbs are used to treat HIV/AIDS since they have negligible or no side effects (Kaur et al., 2020). Because of the potential supply of antioxidants and nutraceutical components, medicinal herbs not only inhibit viral particle multiplication but also serve as immune stimulants and immunomodulators. *Rheum palmatum L., Trigonostema xyphophylloides, Vatica astrotricha, Hypoxias pelargonium, Rheum officinale, Sutherlandia frutescens, Sidoides hemerocallidea, Vernonia amygdalina,* etc. are just a few examples of widely used medicinal plants which shows to have substantial anti-HIV potential. To examine the therapeutic impact of medicinal plants, additional research is required to identify rigorous effectiveness and safety issues through undertaking large-scale clinical trials (Laila et al., 2019). RTI, PRI, ITI, Antioxidants, and Immunomodulators are some of the HIV-PR inhibitors that have been discovered to have activity (Kaur et al., 2020).

DRSML QAU

The primary chemical compounds which are responsible for various therapeutic effects are known as secondary metabolites. They frequently come from primary metabolites like as proteins, carbs, and so on. Multiple plant-based secondary metabolites, including polyphenolic chemicals, coumarins, flavonoids, alkaloids, and terpenoids have been identified and recognized as medicinal agents with several biological properties such as anti-HIV, anti-inflammatory, analgesics, anticancer activities (Tanahashi, 2017).

2.12.3 HIV Treatment by Gene editing tools

In year 1996, genetic variation in CCR5 was revealed in which normal protein was mutated by removal of 32 base pairs in range of 794-825 nucleotides (Dean et al., 1996). A mutant allele with 215 amino acids was created as a result of this deletion, which caused frame shift mutation in the protein and the insertion of seven (7) other amino acids. Numerous studies have provided evidence that this genetic variant is advantageous because it confers HIV protection on those who are homozygous for it (Liu et al., 1996). Based on this strategy, Hütter and his colleagues carried out a stem-cell bone marrow transplantation experiment on HIV-positive patients in 2009. These diseased people were used as transplant donors because they carried the homozygous CR5delta-32 mutation. Results of the bone marrow transplant showed that the HIV-positive person turned HIV negative (Hütter et al., 2009). Nonetheless, such bone marrow transplants faced a lot complications because finding a matched homozygous donor for the mutation is extremely unusual. As a result, researchers focused targeting the HIV chemoreceptor CCR5 utilizing a variety of gene editing tools, including as transcription activator like effector nucleases (TALENs), zinc finger nucleases (ZFN), and sharper cas-9 approaches, in order to overcome these challenges and treat HIV (Zulfiqar et al., 2017).

2.12.3.1 Zinc Finger Nucleases

Zinc finger nuclease (ZNF), an engineered protein with a zinc finger domain that perform gene editing and make a double stranded break in the desired DNA after binding in a specific region of DNA (Khalili et al., 2015). In 2014, Tebas and his coworkers performed experimental trials in which Zinc finger nuclease (ZNF) was employed for the disruption of CCR5 receptor in autologous CD4+ T cells. The trials were seemed successful when modified CD4+ cells administered into infected individuals. The individual gained resistance to HIV and the level of HIV is also drop. However, the reduction of HIV levels was slower than expected,

DRSML QAU

reveal that it can decrease disease progression but is not a everlasting treatment for HIV infection (Tebas et al., 2014).

2.12.3.2 Transcription Activator-Like Effector nucleases (TALENs)

TALENs, also known as transcription activator-like effector nucleases, are a very practical and efficient method used to destroy the CCR5 gene and stop HIV infection. Less cytotoxicity is caused by TALENs than by zinc finger nuclease (ZFN). These nucleases are created by joining a DNA binding domain made of 33–35 conserved amino acids derived from TALE with a non-specific nuclease domain (FokI) (Joung & Sander, 2013). Mock along with his coworkers used TALENs technique to knock out CCR5 gene in experimental trial. The result revealed the protection of altered CCR5 T-cells against R-5 tropic HIV (Joung & Sander, 2013). In contrast to ZNF technique which recognized three nucleotides per domain, TALENs technique identified one nucleotide per domain. So, TALENs is have more specificity than ZNF technique (Holkers et al., 2013). The fact that this technique can encourage genomic instability by inducing rearrangements in target cells, however, is a serious drawback. Consequently, more efficient processes are required for the use of TALENs in therapies (Hockemeyer et al., 2011).

2.12.3.3 CRISPR/CAS9

Antiretroviral therapy is used to suppress HIV-1 replication; however, the virus can hide in other sections of the host genome and resurface if the pharmaceutical therapy is discontinued. A potent genome editing method known as CRISPR/Cas9 has been developed by researchers. It can remove the viral genome from an infected individual by modifying and eventually suppressing the production of the HIV-1 provirus genome (Ebina et al., 2013). The powerful genome editing technique CRISPR-Cas9, also known as clustered regularly interspersed palindromic repeats and CRISPR associated protein 9, was developed in prokaryotes and makes use of the two essential biological components designed single-guided RNA and Cas9 protein. Cas9 is directed to cleave a complementary target sequence known as the (PAM) (Protospacer adjacent motif) by single-guided RNA (sgRNA), which Cas9 can detect (Mao et al., 2016).

Several methods, including pre-integrated proviral dsDNA blockage, CCR5 co-receptor disruption, making the latent provirus the target of cleavage or the excision of the viral genome by selectively targeting HIV-1's LTR sections, and altering viral genes that aid in viral

DRSML QAU

assembly and budding, can be used to eradicate HIV-1 using CRISPR/Cas9. The reactivation of a latent provirus with the aid of CRISPER technology suggests that antiretroviral medications may be used to treat HIV infection. HIV-1 can be cured using CRISPR/Cas9-based targeted editing of the CCR5 receptor. CRISPR/Cas9 provides on-target specificity in addition to being simple to employ and requires a single guided RNA (sgRNA) to designate the cleavage target site (Zulfiqar et al., 2017).

2.13 Vaccine development

The development of a preventative vaccination as well as improvement of ART therapy both are challenging due to very high variability of HIV. A powerful and long-lasting cellular and humoral immune response, including the induction of cross-clade neutralizing NAbs, must be induced by an effective HIV vaccine to support future global immunization campaigns. Thousands of healthy humans voluntarily contributed in testing of 30 candidate vaccines in 1987 (WHO). Still, it is probable that combining many vaccination platforms will be necessary to produce an effective polyvalent vaccine after three decades of completing clinical studies. Novel methods such as mRNA, viral vectored vaccines and HIV-derived virus like particle (VLP) based employing a number of different donor viruses will probably be included in this technology (van Heuvel et al., 2022).

By understanding how to the virus interacts with the human immune system has aided in the creation of number of vaccination platforms. Despite promising preliminary findings from vaccination studies, no effective vaccine against HIV-1 has been created yet. However, new promising vaccination platforms are under investigation. But it is still very difficult to develop an effective vaccine against HIV/AIDS due to the complexity of the process.

DRSML QAU

CHAPTER 3
MATERAL AND METHOD

DRSML QAU

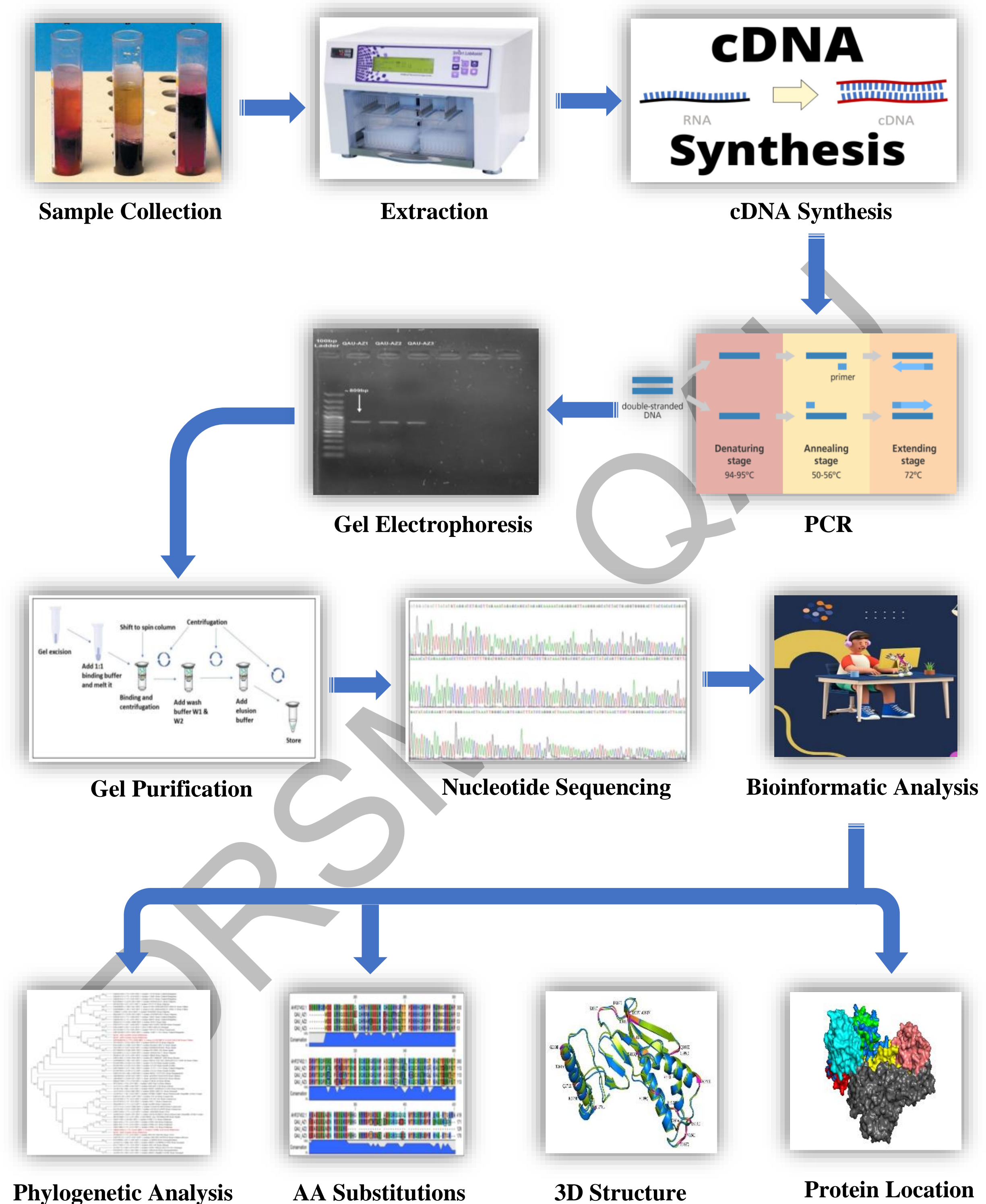

Figure 5. **Graphical representation of overall research work**

The figure shows the overall methodology of our study from sample collection to bioinformatic analysis.

3 Material and Methods

3.1 Sample Collection

Serum from HIV positive individuals was obtained from hospitals and HIV/AIDS centers in Islamabad, Lahore, Peshawar, and Kohat. A total of 20 HIV-positive individual serums were collected, of which 3 were adult females, 13 were adult males, and 4 were children under the age of 15 (1 male and 3 females).

3.2 Inclusion and Exclusion Criteria

HIV positive samples confirmed by PCR were included. Patients with chronic HIV infection, some naïve and some on different antiviral therapies. While Those patients who were negative for HIV were excluded from the study.

3.3 Primers designing for *pol* gene of HIV-1

Primers were designed for *pol* gene of HIV using online tool, Primer3 (http://bioinfo.ut.ee/primer3/) and were further validated using NCBI primer designing tool (Primer Blast) (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Primers were then modified according to the recently reported nucleotide sequences. For this purpose, HIV-1 sequences reported from Pakistan (particularly Accession no. KX232594; isolate HIV-1_PK001) was considered as a reference sequences. Moreover, *in-silico* PCR was also performed with the help of SnapGene (version 1.3.1) software, to confirm the product length and specificity of primers. Name, oligonucleotide sequence, and product size of the primers having maximum *in-silico* specificity are listed in Table 1.

Table 1: Demonstration of primer name, oligonucleotide sequence, and product size of amplified product of *pol* gene of HIV

S. No.	Primer name	Oligonucleotide seq.	PCR Product Size
1	PolF2592	TGCTTCCACAGGGATGGAAA	~ 809bp
2	PolR3393	TGACAAACTCCCATTCAGGA	

DRSML QAU

3.4 RNA extraction and cDNA synthesis

RNA extraction from HIV samples was performed by using automatic extractor TANBead Nucleic Acid Extractor, Model no. SLA-32 (Taiwan). The extracted product was immediately processed for cDNA synthesis or were stored immediately at -20ºC before processing it for cDNA synthesis. cDNA Synthesis Kit (RevertAid, ThermoScientific) was used for cDNA synthesis. The components for synthesis of cDNA 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). The total volume of reaction for was 20 µl as shown in Table 2. The reaction mixture was incubated for 5min at 25 ºC, followed by 42ºC for 60 min, 45 ºC for 30 minutes and finally reaction was completed at 75ºC for 5 mins. The reagents used in different quantities for cDNA synthesis are listed in Table 2. The optimized conditions for cDNA synthesis are given below in Figure 6.

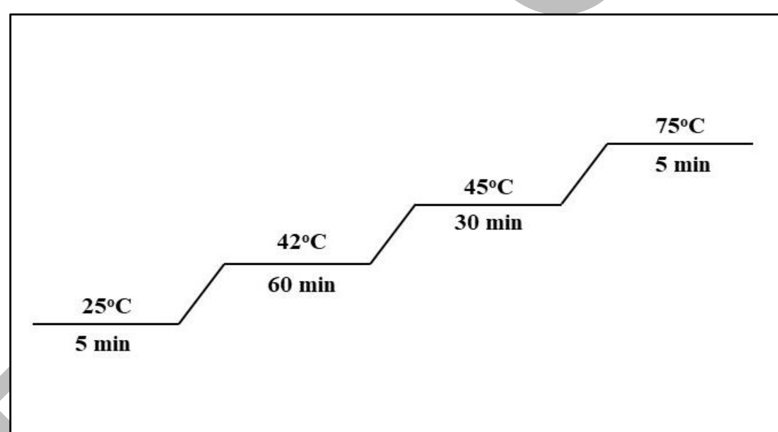

Figure 6: Optimized conditions of HIV complementary DNA (cDNA) synthesis

The optimized condition for cDNA synthesis was 25 ºC for 5 mins, 42ºC for 60 mins, 45 ºC for 30 mins and 75ºC for 5 mins.

DR QAU

Table 2: Different reagents used in cDNA synthesis

S. No.	Reagents	Quantity
1	5x reaction buffer	4µL
2	Random hexamer primers/ Reverse Primer	2µL
3	10 mM dNTP Mix	2µL
4	Template RNA	8µL
5	ddH2O	2µL
6	RiboLock (20U/µL)	1µL
7	RevertAid RT enzyme	1µL
	Total Volume	**20 µl**

3.5 PCR amplification of HIV *pol* gene

HIV *pol* gene was amplified through PCR and PCR Master Mix (2X Phusion High-Fidelity, ThermoScientific) was used. The reaction mixture of PCR contained 2X Phusion High Fidelity Master Mix 10µl, *pol* Forward primer (10µM) 1.5µl, Pol Reverse primer (10µM) 1.5µl, cDNA 6µl and double distilled water 1µl which makes a total of 20µl reaction as shown in Table 3. PCR cyclic condition were given as: initial denaturation at 94°C for 5mins followed by 35 cycles at 94°C for 10 seconds, 58°C for 30 seconds, 72°C for 15 seconds, and the final extension of PCR was at 72°C for 10 min. The hold temperature was 4°C for infinity. Figure 6 shows the optimized condition of PCR. The primers specifically targeted and amplified approximately 809 nucleotides long fragment. The amplified product was confirmed through gel electrophoresis as represented in Figure 7.

Table 3: Different reagents use in PCR along with its volume

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

DRSML QAU

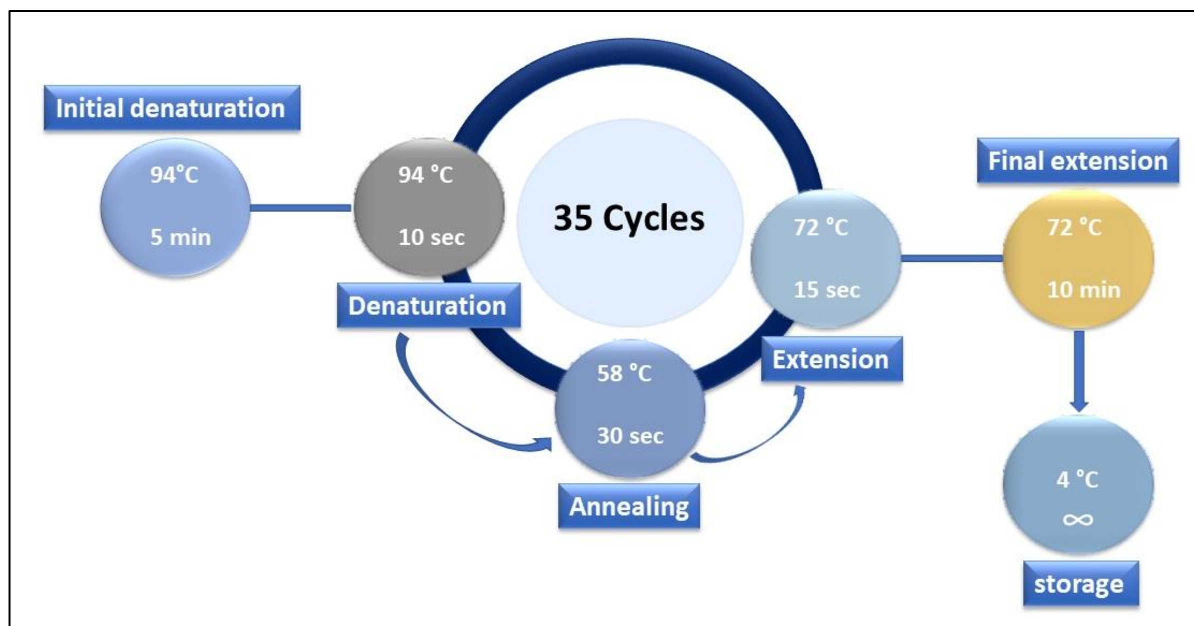

Figure 7: Representation of Optimized conditions of Polymerase Chain Reaction (PCR)

The condition for PCR includes denaturation in initial stage at 94°C for 5 mins, denaturation, annealing and extension in 2nd stage at 94°C for 10 sec, 58°C for 30 sec and 72°C for 15 sec. Final extension occurred at 72 in 10 min. Storage temperature was 4°C for infinite time.

3.6 Preperation of TAE Buffer

TAE is abbreviated as Tris Acetate EDTA (Ethylenediaminetetraacetic-acid), a running buffer which is employed to separate DNA in agarose gel during the process of gel electrophoresis. Reagents like Tris base, glacial acetic acid (Boric acid is used instead of glacial acetic acid in case of TBE buffer), EDTA and distal water (ddH_2O) are used in preparation of TAE buffer. For the preparation of 10X TAE buffer in 1 litter, 48.4g of Tris base was added to 800ml ddH_2O. Then 11.4 ml and 3.7g of glacial acetic acid and EDTA was added and mixed thoroughly. After that 200ml more ddH_2O was added to the mixture for making the final volume of 1000ml or 1 litre. Then 10ml of the prepared 10x buffer was added to 90ml of double distal water (ddH_2O) to make 100ml of 1x TAE buffer.

3.7 PCR products detection on Agarose gel

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 in conical flask. 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 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. Different components used in gel electrophoresis process are listed in Table 4. The amplified fragment (excised from the gel) was purified using Gene JET Gel Extraction Kit (ThermoScientific) as demonstrated in Figure 8. The total 30µl eluted products were further processed for nucleotide sequencing.

Table 4: Reagents use in gel electrophoresis

S. No.	Components of (1.5% gel)	Amount
1	1 X TAE buffer	40ml
2	Agarose	0.6g
3	Ethidium bromide solution	1.5µl

DRSML QAU

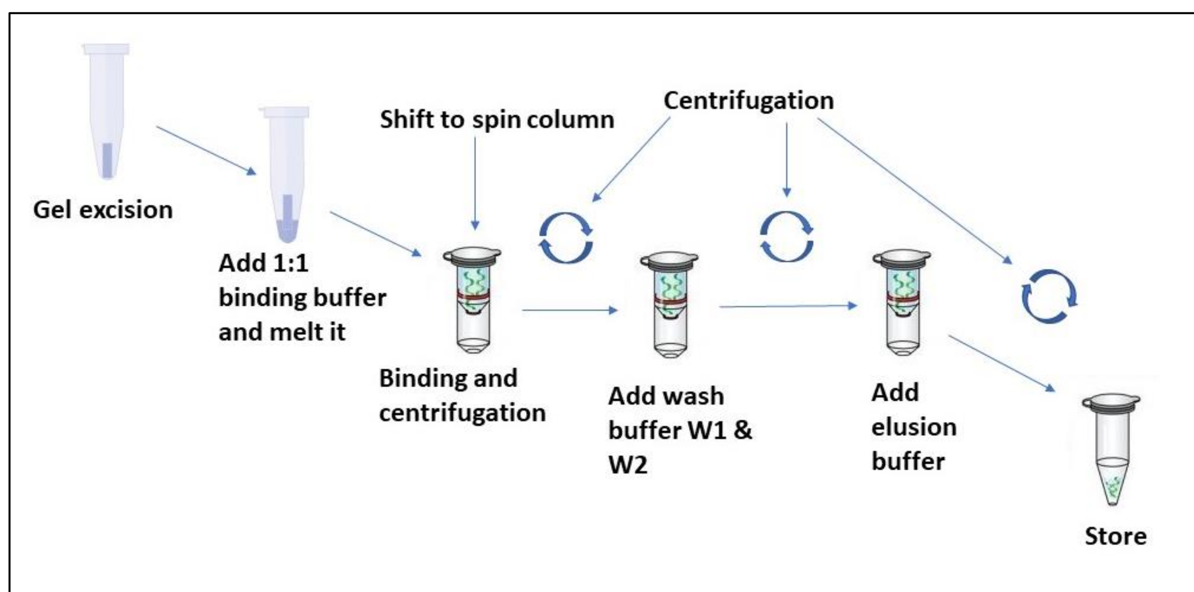

Figure 8: Gel purification procedure of the amplified PCR product

Steps: Gel excision, Adding Binding Buffer, centrifugation, Adding Wash buffer W1 & W2, adding elusion buffer and Storage.

DRSML QAU

3.8 Sequencing of HIV *pol* gene

Sanger sequencing of the gel purified PCR product was done with *pol* specific Forword primer. A sequence alignment tool called BioEdit was employed to visualized the nucleotide sequencing results. The nucleotide sequences of our isolates were obtained in "FASTA" format and further analysis was performed.

3.9 Nucleotide Blast and Phylogenetic Analysis

Nucleotide Basic local alignment search tool (BLASTn) (https://blast.ncbi.nlm.nih.gov/) was employed for homology analysis of the query sequence (QAU-AZ1, QAU-AZ2 and QAU-AZ3) with most closely related known HIV-1 from Pakistan and other regions of the world present in NCBI database. All the available nucleotide sequences of *pol* gene of HIV-1 were downloaded from the GenBank NCBI. The saved sequences were then aligned using MAFFT online server (https://mafft.cbrc.jp/alignment/server/index.html) (Katoh, Rozewicki, & Yamada, 2019). Phylogenetic analysis was performed using the isolate sequences and 60 other retrieved sequences Pol-HIV-1 from NCBI. The evolutionary history was concluded by means of Maximum likelihood Method using Hasegawa Kishino Yano + Gamma with Invariant sites

(HKY+G+I) model (Abecasis et al., 2018). All the evolutionary history analyses were conducted by using MEGA 11 package (Tamura et al., 2021).

3.10 Translation of nucleotide sequences

Nucleotides sequences of *pol* gene 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).

3.11 Alignment *pol* protein of HIV

The nucleotide sequences of our isolates were then translated into amino acid sequences through online server ExPASy (https://web.expasy.org/translate/). The alignment of amino acid sequences was executed using CLC (Yamaguchi, 2018). The isolates protein sequences were aligned with a WHO recommended protein sequence of HIV-1 *pol* protein (Accession no. AHF27452). Amino acid substitutions were then analyzed for positions that have been described in the previous literature.

3.12 Prediction of Secondary structures

To predict secondary structure of HIV-pol samples and reference sequence an online self-optimized prediction server SOPMA was used. SOPMA has been described to improve the success rate in the prediction of secondary structure of protein sequence (Geourjon and Deleage, 1995).

3.13 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 3D models of our isolates were obtained in "PDB" format and visualize on PyMOL software. Structure comparison of our isolates with reference protein was also performed by PyMOL (Yang et al., 2012).

3.14 Location of isolated Proteins in Reference Protein

The location of our isolated proteins was identified using online HIV sequence locator tool (https://www.hiv.lanl.gov/content/sequence/LOCATE/locate.html) and was further confirmed in Reference protein (Harrison et al., 2022) using PyMole software.

DRSML QAU

3.15 Identification of Conserved regions

Our isolated protein sequences were blast in NCBI protein blast and the closely related sequences was retrieved. After alignment the Conserved regions were identified using BioEdit software. The conserved regions were then visualized using PyMole software.

DRSML QAU

CHAPTER 4

RESULTS

DRSMLQAU

4 Results

In present study, total of 20 HIV suspected samples were collected from hospitals and HIV/AIDS centers of Islamabad, Lahore, Peshawar, and Kohat and confirmed through PCR. 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), 13 adult males and 3 adult females. Out of the total positive samples 5 samples were collected from Islamabad, 8 from Peshawar, 3 from Kohat and 4 from Lahore, Punjab.

4.1 Results of PCR amplified products through Gel electrophoresis.

Pol gene of HIV-1 was amplified by PCR by using a set of gene specific forward and reverse primers. After the PCR reaction the amplified products were run on a 1.5% agarose gel for the confirmation. The samples were labeled as QAU-AZ1, QAU-AZ2. and QAU-AZ3, respectively. The agarose gel shows amplified fragment of approximately 809 bp of HIV-1 Pol gene for all samples when compared with the DNA 100bp ladder as shown in Figure 9. Each band from gel was cut and stored for purification.

DR QAU

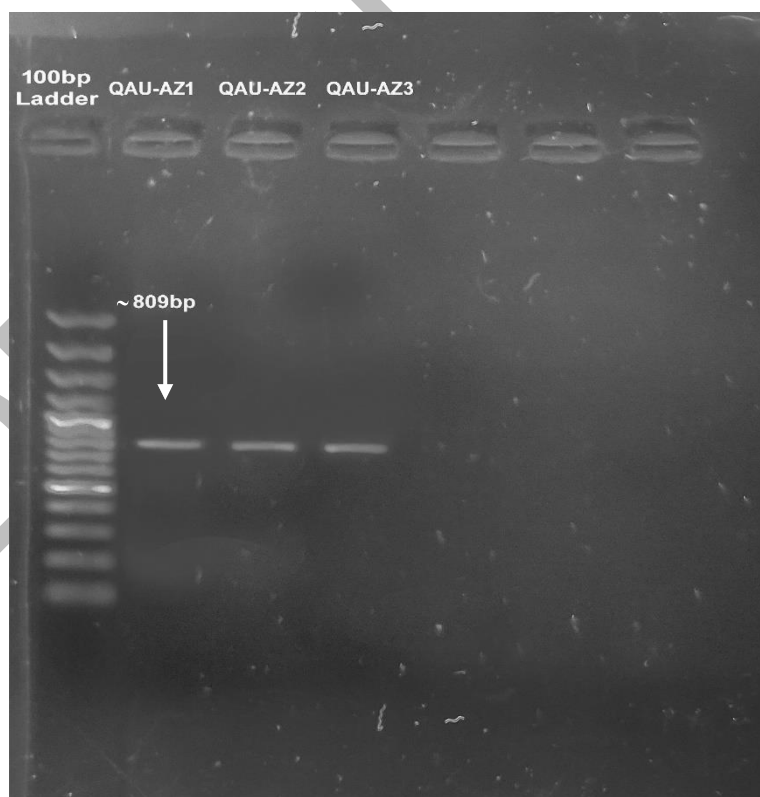

Figure 9: Gel electrophoresis and visualization of amplified products

Lane 1 represent 100bp ladder while Lane 2, Lane 3 and Lane 4 represent amplified PCR products of ~809bp.

4.2 Nucleotide Sequencing results

Each sample after amplification was purified from excised gel using gel extraction kit and three samples were further processed for sangers sequencing using *pol* forward primer (5'-TGCTTCCACAGGGATGGAAA-3'). Chromas software was employed for refining of sequences (Zajec, 1986). Three sequences of Pol HIV-1 were obtained in current study and were used further for post sequencing analysis through different bioinformatics tools. The chromatogram of nucleotide sequences is shown in Figure 10.

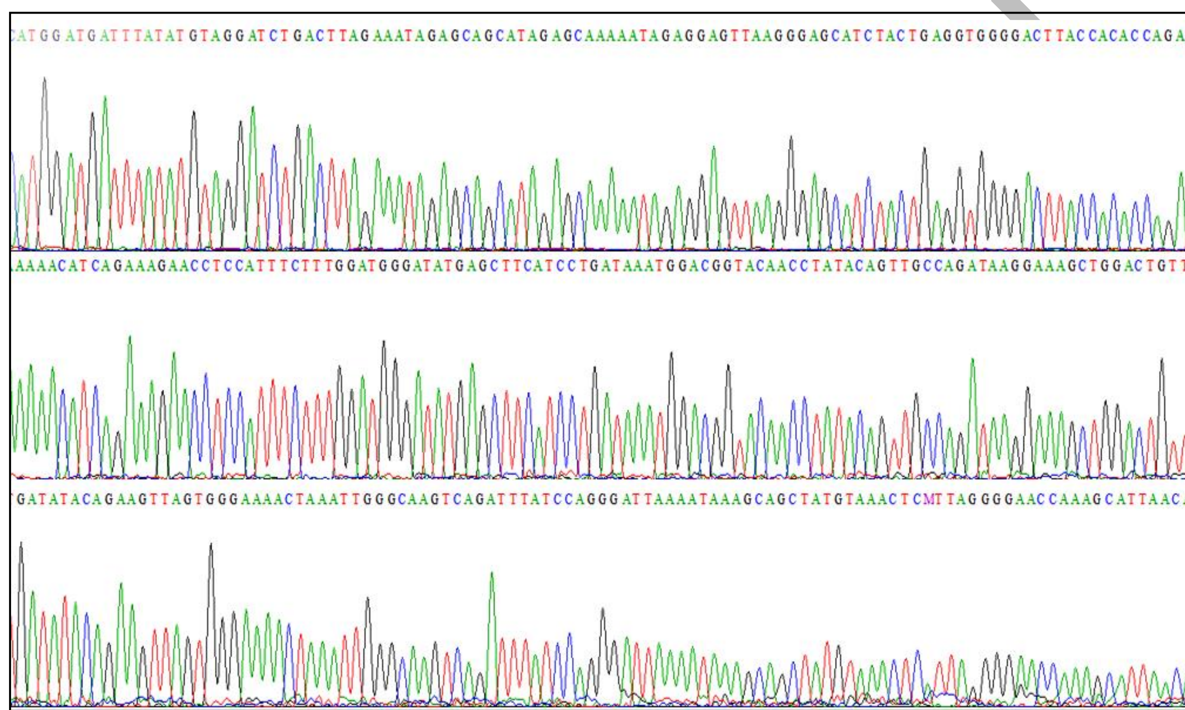

Figure 10: Graphical representation of sequencing in the form of chromatogram

Nucleotide sequencing result of QAU-AZ1 *pol* gene in the form of chromatogram. Nucleotides A, T, G and C are represented in Green, Red, Black, and Blue color respectively.

4.3 Phylogenetic analysis

Phylogenetics analysis was performed to understand the evolution and genetic variability of current isolates, based on similarities and differences with other virus sequences. Twenty-five (25) sequences for each isolate (QAU-AZ1, QAU-AZ2, QAU-AZ3) were retrieved from NCBI protein blast. After removal of homologous sequences by Notepad ++, only 60 sequences along with isolate sequences were aligned and subjected to MEGA 11 program for Phylogenetic analysis. Maximum Likelihood method was employed for phylogenetic analysis. Percentage of each clustered taxa in optimal trees are displayed next to every branch. In

Maximum Likelihood Method, Hasegawa Kishino Yano + Gamma with Invariant sites (HKY+G+I) model was used with 1000 boots strap replication to compute evolutionary history. Best fit model selection was performed using MEGA 11 software in which HKY+G+I was selected as best fit substitution model based on low Bayesian-Information-Criterion (BIC) value *i.e.* 10637.21756. The current study isolates QAU-AZ1 and QAU-AZ3 clustered with HIV-1 clone (Accession no: MN908916) reported from China, while QAU-AZ2 clustered with HIV-1 isolate 07PK-115 from Pakistan (Accession no. JQ011662). The study demonstrates that HIV-1 *pol* gene isolates clustered together with reference sequences from China and Pakistan while phylogenetically distinct from other region like Spain, Nigeria, Cameron, United Kingdom, Senegal, and USA *etc.* which depicts that they might have been originated from above-mentioned sequences recently and 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 *pol* gene are represented in black color as shown in the Figure 11.

DRSML QAU

Figure 11: Phylogenetic analysis of *pol* gene of studied HIV-1 in Pakistan

Red color represents our isolates (n=3) and black color represent reference sequences (n=60) reported from different countries of the word.

4.4 Amino acids substitutions in HIV-1 *pol* protein

The nucleotide sequences of all isolates were translated to amino acid sequences using online server ExPASy, followed by alignment of resultant amino acid sequences with reference sequences through CLC workbench (Version 8) as shown in the Figure 12. The *pol* reference protein sequence (Accession no. AHF27452) recommended by WHO as international reference sequence was retrieved from NCBI GenBank. Upon comparison with reference sequence, significant amino acid substitutions were found in all isolates. There are 23 amino acid substitutions found in pol protein of QAU-AZ1 isolates that are G260E, T264A, Q271E, K275R, F278L, V309Q, E312D, D314E, V340I, T350A, E355D, V356I, K361E, S386T, D388E, I390V, A391V, Q398E, G399D, K411I, -414L, T416Q, and K418N. Similarly, 16 amino acid substitutions such as T264A, I266V, Q271E, F278L, V309K, D314E, S315N, K339R, Q342H, L347I, T350A, E355D, V356I, P358T, L359M and K361E were found in QAU-AZ2 isolate. Moreover, in isolate QAU-AZ3, 17 amino acid substitutions were observed that are G260E, T264A, Q271E, K275R, F278L, V309Q, E312D, D314E P336A, V340I, L347I, E355D, V356I, K361E, S386T, I390V and G399D. Detailed information of amino acid substitutions of all isolates with reference sequence are given in Figure 12 and Table 5, 6 and 7.

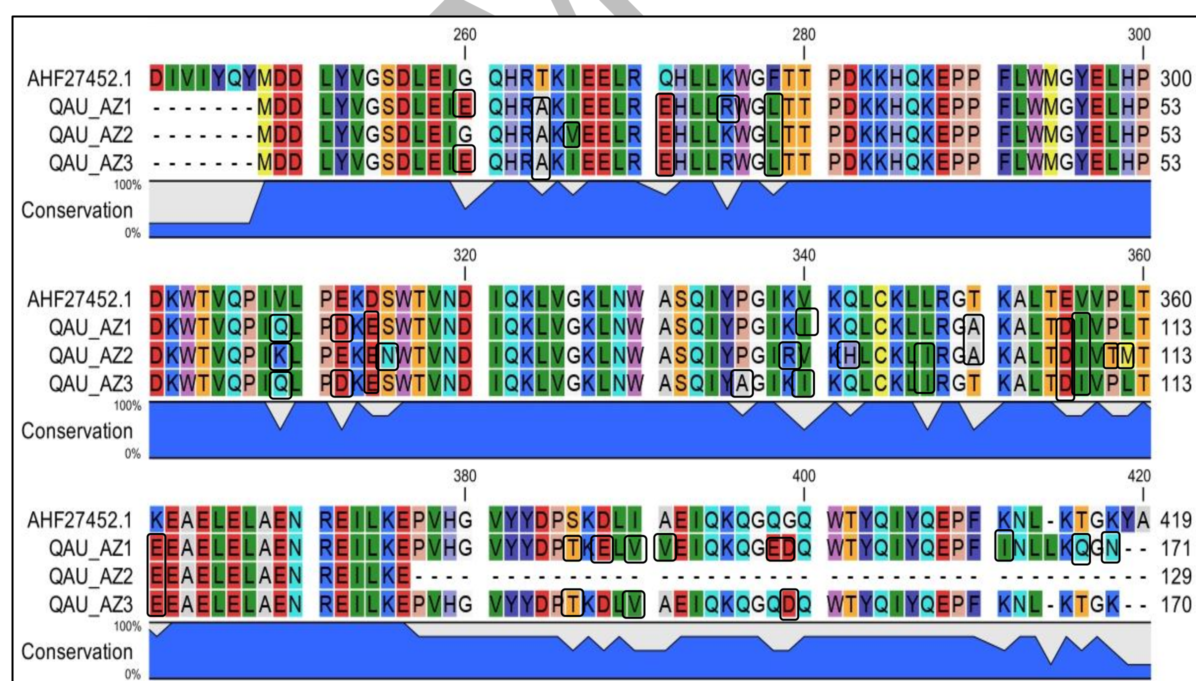

Figure 12: Representation of the alignment of amino acid sequences

QAU_AZ1, QAU_AZ2, QAU_AZ3, represents our isolates while AHF27452 represents reference sequence. The bule color represent conservation wile grove represent substitution.

Table 5: Amino acid substitutions in the HIV-1 isolates (QAU-AZ1) proteins compared with reference sequence of HIV-1 *pol* protein (AHF27452).

S. No.	Amino acid position*	Reference Protein	Charge	Polarity	Variant	Charge	Polarity
1	260	G (Glycine)	Neutral	Non-polar	E (Glutamic acid)	-ve	Polar
2	264	T (Threonine)	Neutral	Polar	A (Alanine)	Neutral	Non-polar
3	271	Q (Glutamine)	Neutral	Polar	E (Glutamic acid)	-ve	Polar
4	275	K (Lysine)	+ve	Polar	R (Arginine)	+ve	Polar
5	278	F (Phenylalanine)	Neutral	Non-polar	L (Leucine)	Neutral	Non-polar
6	309	V (Valine)	Neutral	Non-polar	Q (Glutamine)	Neutral	Polar
7	312	E (Glutamic acid)	-ve	Polar	D (Aspartic acid)	-ve	Polar
8	314	D (Aspartic acid)	-ve	Polar	E (Glutamic acid)	-ve	Polar
9	340	V (Valine)	Neutral	Non-polar	I (Isoleucine)	Neutral	Non-polar
10	350	T (Threonine)	Neutral	Polar	A (Alanine)	Neutral	Non-polar
11	355	E (Glutamic acid)	-ve	Polar	D (Aspartic acid)	-ve	Polar
12	356	V (Valine)	Neutral	Non-polar	I (Isoleucine)	Neutral	Non-polar
13	361	K (Lysine)	+ve	Polar	E (Glutamic acid)	-ve	Polar
14	386	S (Serine)	Neutral	Polar	T (Threonine)	Neutral	Polar
15	388	D	-ve	Polar	E	-ve	Polar

DRSML QAU

		(Aspartic acid)			(Glutamic acid)		
16	390	I (Isoleucine)	Neutral	Non-polar	V (Valine)	Neutral	Non-polar
17	391	A (Alanine)	Neutral	Non-polar	V (Valine)	Neutral	Non-polar
18	398	Q (Glutamine)	Neutral	Polar	E (Glutamic acid)	-ve	Polar
19	399	G (Glycine)	Neutral	Non-polar	D (Aspartic acid)	-ve	polar
20	411	K (Lysine)	+ve	Polar	I (Isoleucine)	Neutral	Non-polar
21	414	-			L (Leucine)	Neutral	Non-polar
22	416	T (Threonine)	Neutral	Polar	Q (Glutamine)	Neutral	Polar
23	418	K (Lysine)	+ve	Polar	N (Asparagine)	Neutral	Polar

* The Amino acid position was based on reference protein (Accession No. AHF27452).

Table 6: Amino acid substitutions in the HIV-1 isolates (QAU-AZ2) protein compared with reference sequence of HIV-1 *pol* protein (AHF27452)

S. No.	Amino acid position*	Reference Protein	Charge	Polarity	Variant	Charge	Polarity
1	264	T (Threonine)	Neutral	Polar	A (Alanine)	Neutral	Non-polar
2	266	I (Isoleucine)	Neutral	Non-polar	V (Valine)	Neutral	Non-polar
3	271	Q (Glutamine)	Neutral	Polar	E (Glutamic acid)	-ve	Polar
4	278	F (Phenylalanine)	Neutral	Non-polar	L (Leucine)	Neutral	Non-polar
5	309	V (Valine)	Neutral	Non-polar	K (Lysine)	+ve	Polar
6	314	D (Aspartic acid)	-ve	Polar	E (Glutamic acid)	-ve	Polar
7	315	S (Serine)	Neutral	Polar	N (Asparagine)	Neutral	Polar
8	339	K (Lysine)	+ve	Polar	R (Arginine)	+ve	Polar
9	342	Q (Glutamine)	Neutral	Polar	H (Histidine)	+ve	Polar

DRSML QAU

10	347	L (Leucine)	Neutral	Non-polar	I (Isoleucine)	Neutral	Non-polar
11	350	T (Threonine)	Neutral	Polar	A (Alanine)	Neutral	Non-polar
12	355	E (Glutamic acid)	-ve	Polar	D (Aspartic acid)	-ve	Polar
13	356	V (Valine)	Neutral	Non-polar	I (Isoleucine)	Neutral	Non-polar
14	358	P (Proline)	Neutral	Non-polar	T (Threonine)	Neutral	Polar
15	359	L (Leucine)	Neutral	Non-polar	M (Methionine)	Neutral	Non-polar
16	361	K (Lysine)	+ve	Polar	E (Glutamic acid)	-ve	Polar

* The Amino acid position was based on reference protein (Accession No. AHF27452).

Table 7: Amino acid substitutions in the HIV-1 isolates (QAU-AZ3) proteins compared with reference sequence of HIV-1 *pol* protein (AHF27452)

S. No.	Amino acid position*	Reference Protein	Charge	Polarity	Variant	Charge	Polarity
1	260	G (Glycine)	Neutral	Non-polar	E (Glutamic acid)	-ve	Polar
2	264	T (Threonine)	Neutral	Polar	A (Alanine)	Neutral	Non-polar
3	271	Q (Glutamine)	Neutral	Polar	E (Glutamic acid)	-ve	Polar
4	275	K (Lysine)	+ve	Polar	R (Arginine)	+ve	Polar
5	278	F (Phenylalanine)	Neutral	Non-polar	L (Leucine)	Neutral	Non-polar
6	309	V (Valine)	Neutral	Non-polar	Q (Glutamine)	Neutral	Polar
7	312	E (Glutamic acid)	-ve	Polar	D (Aspartic acid)	-ve	Polar
8	314	D (Aspartic acid)	-ve	Polar	E (Glutamic acid)	-ve	Polar
9	336	P (Proline)	Neutral	Non-polar	A (Alanine)	Neutral	Non-polar
10	340	V (Valine)	Neutral	Non-polar	I (Isoleucine)	Neutral	Non-polar

DRSML QAU

11	347	L (Leucine)	Neutral	Non-polar	I (Isoleucine)	Neutral	Non-polar
12	355	E (Glutamic acid)	-ve	Polar	D (Aspartic acid)	-ve	Polar
13	356	V (Valine)	Neutral	Non-polar	I (Isoleucine)	Neutral	Non-polar
14	361	K (Lysine)	+ve	Polar	E (Glutamic acid)	-ve	Polar
15	386	S (Serine)	Neutral	Polar	T (Threonine)	Neutral	Polar
16	390	I (Isoleucine)	Neutral	Non-polar	V (Valine)	Neutral	Non-polar
17	399	G (Glycine)	Neutral	Non-polar	D (Aspartic acid)	-ve	Polar

* The Amino acid position was based on reference protein (Accession No. AHF27452).

4.5 Secondary structure prediction

The secondary structure of HIV-1 *pol* protein was predicted via online server SOPMA (Geourjon, 1995). The server used bioinformatics approaches that attempt to determine the local secondary structures of proteins based only on their amino acid sequence. The secondary structure of Reference sequence comprises 45.29% Alpha helix, 15.88% Extended strand, 8.24% Beta turn and 30.59% Random coil as shown in Figure 13. HIV-1 isolate QAU-AZ1 has an alpha helix of 46.78%, extended strand of 14.04%, Beta turn of 7.6% and Random coils of 31.58% while in HIV-1 isolate QAU-AZ2 has an alpha helix of 48.06%, extended strand of 12.40%, Beta turn of 6.98% and Random coils of 52.56% as shown in Figure 14 and 16. Similarly, isolate QAU-AZ3 has alpha helix of 44.71%, extended strand of 13.53%, Beta turn of 8.24% and Random coils of 33.53% as shown in Figure 16. It is supposed that conserved regions are existing which maintained the structure of the Protein.

DRSML QAU

```
         10        20        30        40        50        60        70
          |         |         |         |         |         |         |
MDDLYVGSDLEIGQHRTKIEELRQHLLKWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDSWT
hceeeecccccchhhhhhhhhhhhhhhhhttcccccccccccccceeeeeeecttcceeeeeccccccccc
VNDIQKLVGKLNWASQIYPGIKVKQLCKLLRGTKALTEVVPLTKEAELELAENREILKEPVHGVYYDPSK
hhhhhhhhhhhhhhhhhcttcchhhhhhhhttccccchhhhhhhhhhhhhhhhhhhhhhhhtteeccttth
DLIAEIQKQGQGQWTYQIYQEPFKNLKTGK
hhhhhhhhccccceeeeeehtcccceeetc

Sequence length :   170

SOPMA :
   Alpha helix      (Hh) :    77 is  45.29%
   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  15.88%
   Beta turn        (Tt) :    14 is   8.24%
   Bend region      (Ss) :     0 is   0.00%
   Random coil      (Cc) :    52 is  30.59%
   Ambiguous states (?)  :     0 is   0.00%
   Other states          :     0 is   0.00%
```

20 40 60 80 100 120 140 160

Helix
Sheet
Turn
Coil

20 40 60 80 100 120 140 160

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

Figure 13: Secondary structure of the reference *pol* protein sequence of HIV-1

Different motifs and parts of *pol i.e.* Alpha helix, Extended sheet, Beta turn and Random coils are represented in Blue, Red, Green and Pink respectively along with their percentage and number.

```
        10        20        30        40        50        60        70
         |         |         |         |         |         |         |
MDDLYVGSDLEIEQHRAKIEELREHLLRWGLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIQLPDKESWT
hceeeeccccchhhhhhhhhhhhhhhhhttcccccccccccccceeeeeeccttcceeeeecccccccce
VNDIQKLVGKLNWASQIYPGIKIKQLCKLLRGAKALTDIVPLTEEAELELAENREILKEPVHGVYYDPTK
hhhhhhhhhhhhhhhhccttcchhhhhhhhttccchhhhhhhhhhhhhhhhhhhhhhhhhcttccttccth
ELVVEIQKQGEDQWTYQIYQEPFINLLKQGN
hhhhhhhhcccccceeeeeecccceh hhhttc

Sequence length :   171

SOPMA :
   Alpha helix      (Hh) :    80 is  46.78%
   3₁₀ helix        (Gg) :     0 is   0.00%
   Pi helix         (Ii) :     0 is   0.00%
   Beta bridge      (Bb) :     0 is   0.00%
   Extended strand  (Ee) :    24 is  14.04%
   Beta turn        (Tt) :    14 is   8.19%
   Bend region      (Ss) :     0 is   0.00%
   Random coil      (Cc) :    53 is  30.99%
   Ambiguous states (?)  :     0 is   0.00%
   Other states          :     0 is   0.00%
```

20 40 60 80 100 120 140 160

Helix
Sheet
Turn
Coil

20 40 60 80 100 120 140 160

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

Figure 14: Secondary structure of *pol* protein of QAU-AZ1 isolate

Different motifs and parts of *pol i.e.* Alpha helix, Extended sheet, Beta turn and Random coils are represented in Blue, Red, Green and Pink respectively along with their percentage and number.

```
        10        20        30        40        50        60        70
         |         |         |         |         |         |         |
MDDLYVGSDLEIGQHRAKVEELREHLLKWGLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIKLPEKENWT
hceeeecccccccchhhhhhhhhhhhhhttcccccccccccccceeeeeeccttcceeeeeecccccccc
VNDIQKLVGKLNWASQIYPGIRVKHLCKLIRGAKALTDIVTMTEEAELELAENREILKE
hhhhhhhhhhhhhhhtccttcchhhhhhhhttccchhhhhhhhhhhhhhhhhhhhhhhh

Sequence length :   129

SOPMA :
   Alpha helix       (Hh) :    62 is  48.06%
   3_10 helix         (Gg) :     0 is   0.00%
   Pi helix          (Ii) :     0 is   0.00%
   Beta bridge       (Bb) :     0 is   0.00%
   Extended strand   (Ee) :    16 is  12.40%
   Beta turn         (Tt) :     9 is   6.98%
   Bend region       (Ss) :     0 is   0.00%
   Random coil       (Cc) :    42 is  32.56%
   Ambiguous states  (?)  :     0 is   0.00%
   Other states           :     0 is   0.00%

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

Figure 15: Secondary structure of *pol* protein of QAU-AZ2 isolate

Different motifs and parts of *pol i.e.* Alpha helix, Extended sheet, Beta turn and Random coils are represented in Blue, Red, Green and Pink respectively along with their percentage and number.

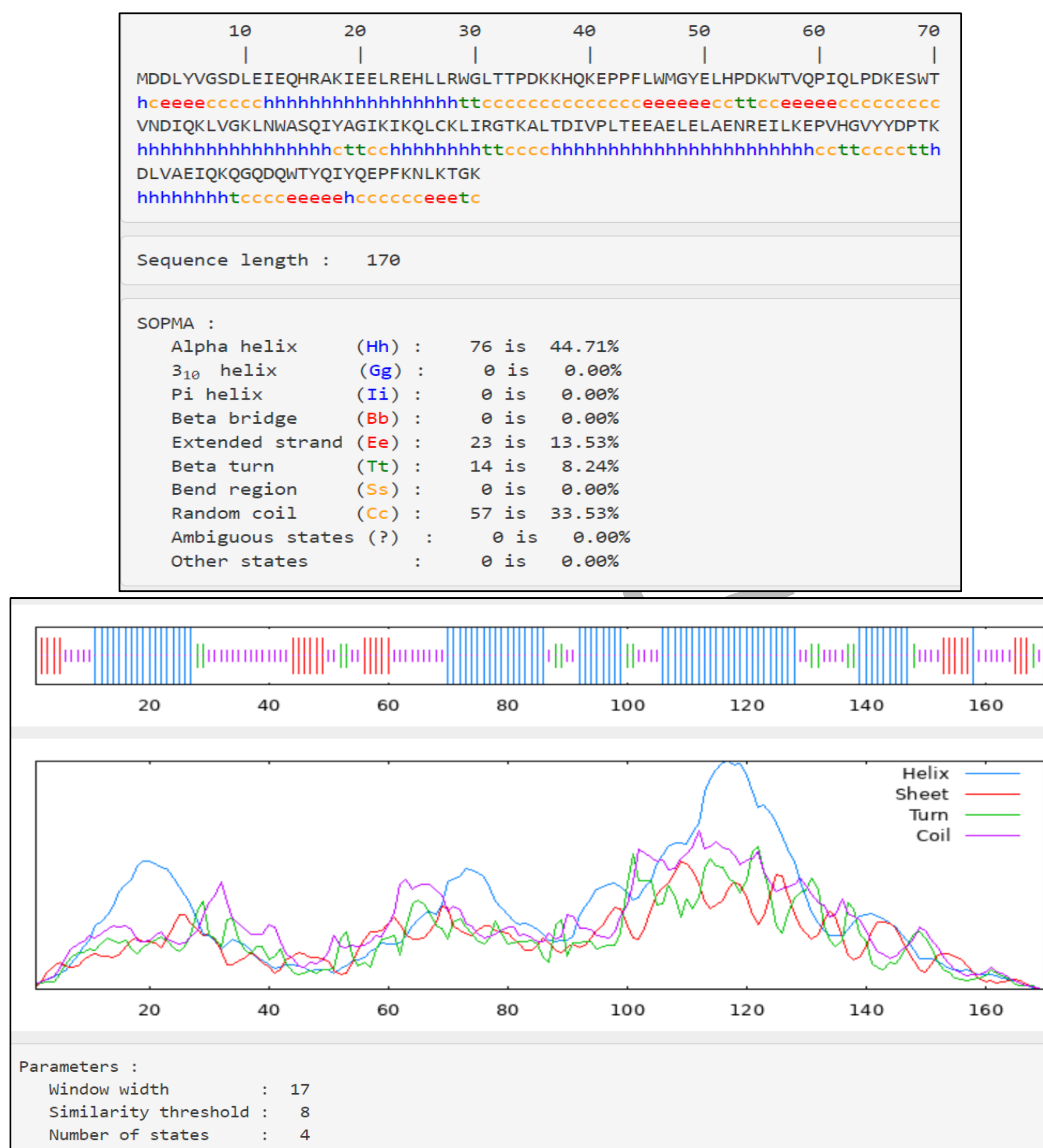

Figure 16: Secondary structure of *pol* protein of QAU-AZ3 isolate

Different motifs and parts of *pol i.e.* Alpha helix, Extended sheet, Beta turn and Random coils are represented in Blue, Red, Green and Pink respectively along with their percentage and Number.

4.6 Modeling and Visualization of Protein

The tertiary structure (3D) of protein sequences for three isolates QAU-AZ1, QAU-AZ2 and QAU-AZ3 were generated by Robetta server. Structural comparison of the isolates proteins with the reference protein sequence was made possible through visualization of protein models. The assessment of substitutions was made easier and clear through structure predication. Each variant protein structure was superimposed with reference structure via PyMOL software. The substitutions are highlighted in amino acid sequences of 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 as shown in Figure 17, 18, and 19.

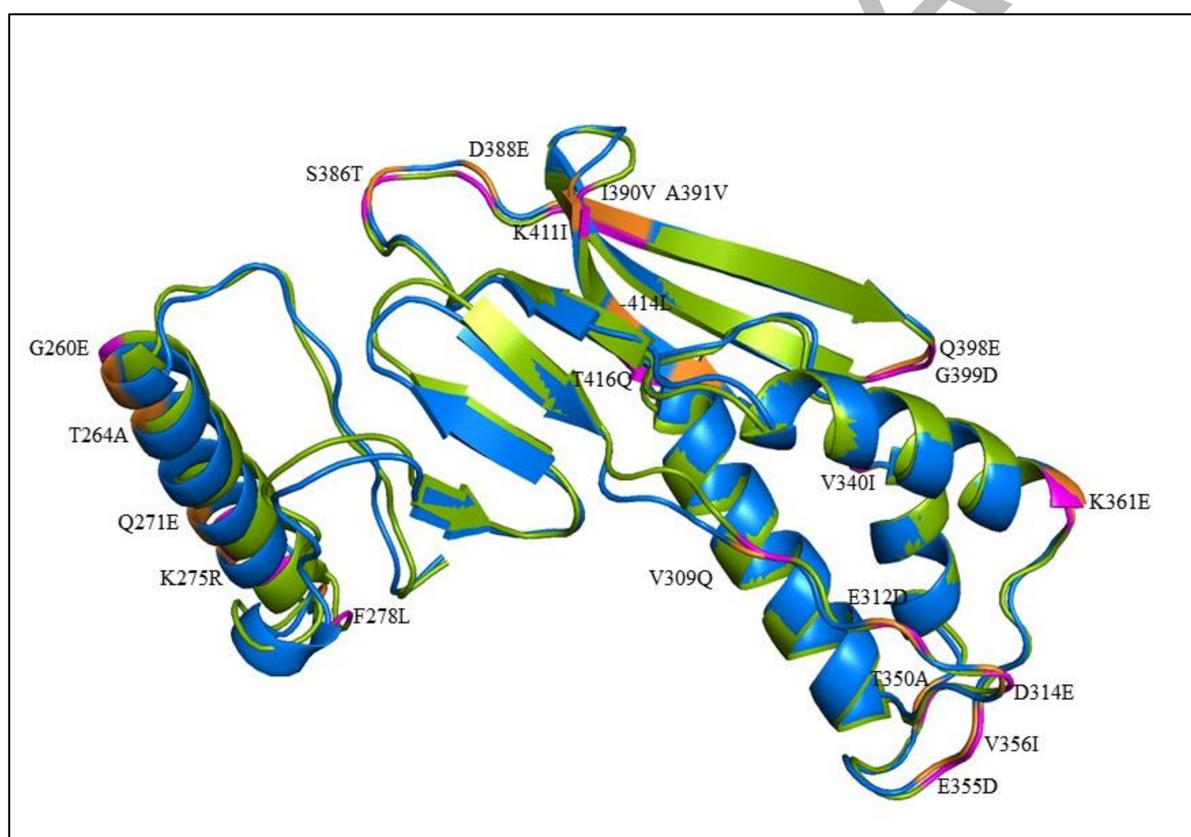

Figure 17: The 3D structure representation of HIV-1 *pol* reference protein and isolate QAU-AZ1

Splitpea Green color represent reference sequence with amino substitution in Magenta color while Marine Blue color represent isolate QAU-AZ1 with amino substitution in orange color.

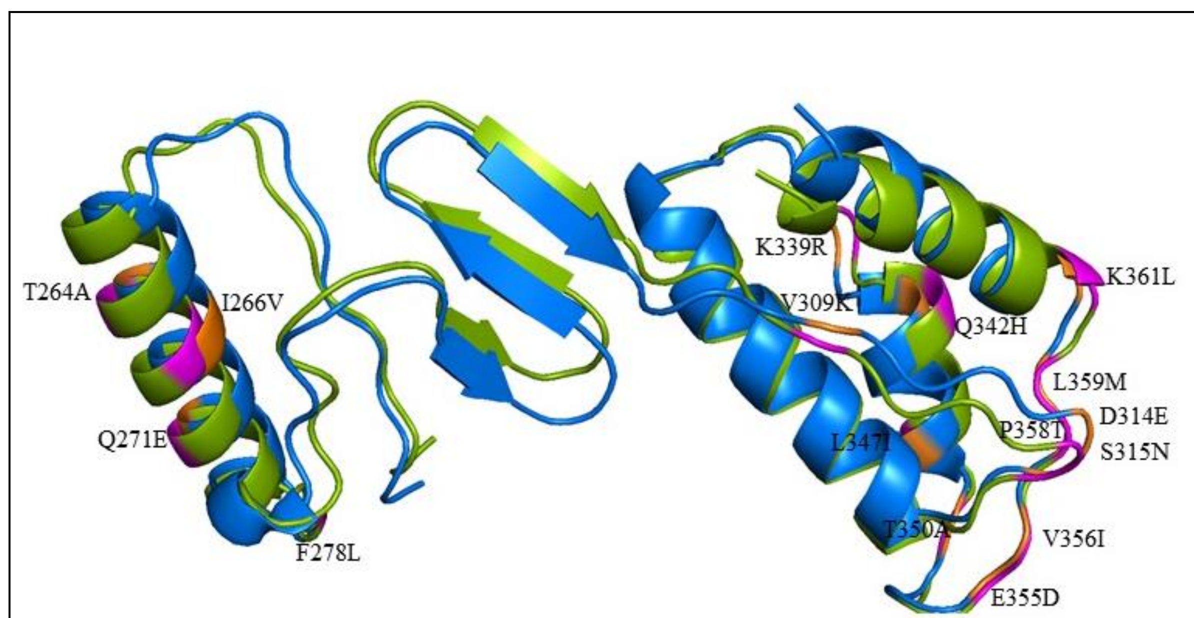

Figure 18: The 3D structure representation of HIV-1 *pol* reference protein and isolate QAU-AZ2

Splitpea Green color represent reference sequence with amino substitution in Magenta color while Marine Blue color represent isolate QAU-AZ2 with amino substitution in orange color.

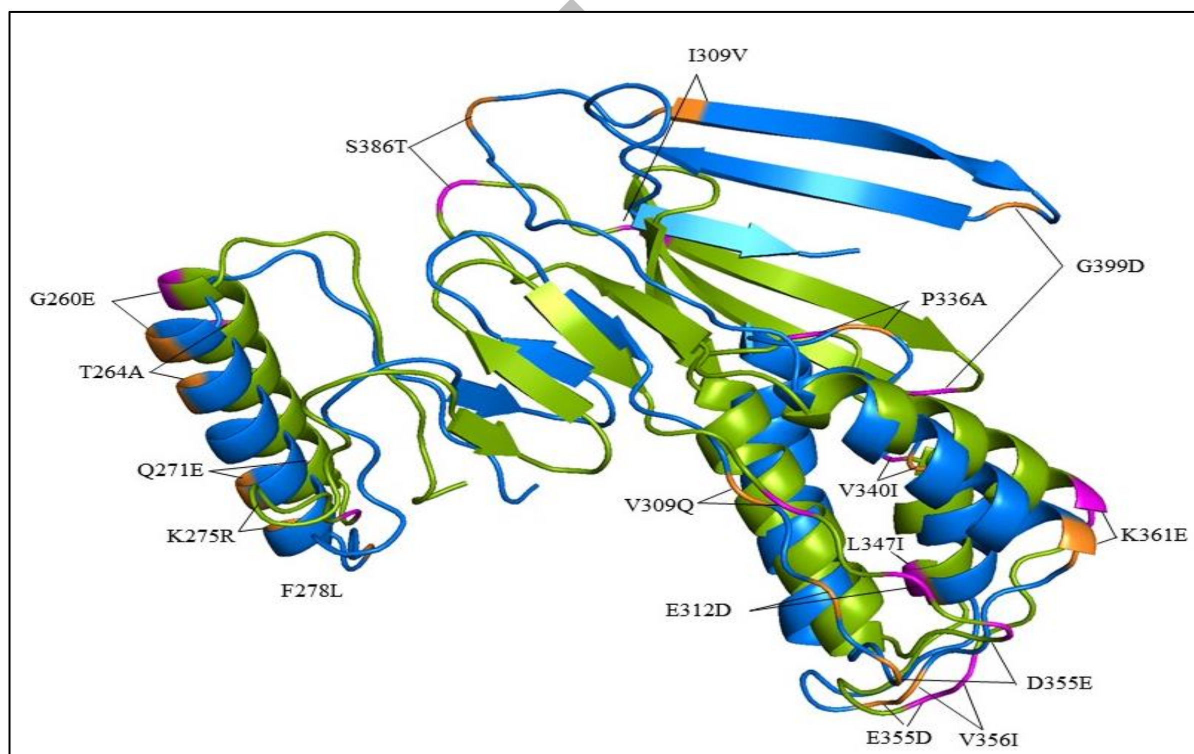

Figure 19: The3D structure representation of HIV-1 *pol* reference protein and isolate QAU-AZ3

Splitpea Green color represent reference sequence with amino substitution in Magenta color while Marine Blue color represent isolate QAU-AZ3 with amino substitution in orange color.

4.7 Protein isolates Location in Reference Protein

It was confirmed from online HIV Sequence Locator tool that all our isolate proteins are located in Reverse Transcriptase Protein of *pol* gene of HIV1. It was further confirmed by aligning our isolated proteins 3D structure with reference Reverse Transcriptase Protein (PDB ID: 1DLO) using PyMole software. After the alignment it was confirmed that all our isolates are located in Palm and Connection region of RTp66 subunit of HIV-1 reverse transcriptase enzyme. The locations of our isolates are represented in Figure 20.

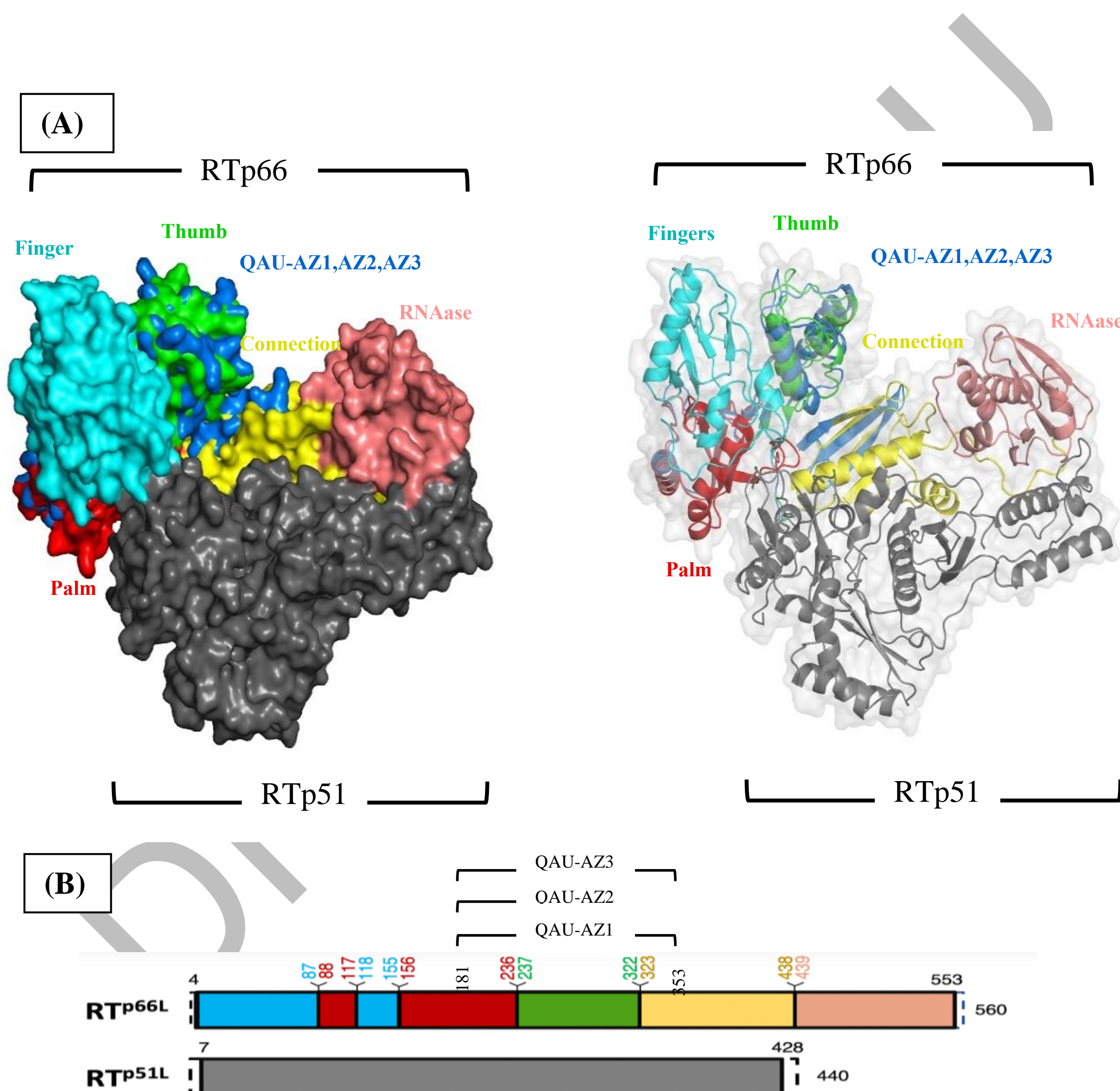

Figure 20. Representation of our isolates in 3D and Liner structure of Reverse Transcriptase protein

The 3D structure of reverse transcriptase protein is represented in **(A)** while its Linear structure is represented in **(B)** (Harrison et al., 2022).

4.8 Conserved Region Identification

The isolated proteins were blast in NCBI blast and 100 protein sequences was retrieved for each isolate. After the alignment of retrieved sequences with our isolates, the conserve region was identified using Bioedit software. Only two conserved regions were identified in our isolates and reference protein i-e. 32 TTPDKKHQKEPPFLWMGYELHPDKWTVQPI 61 and 68 SWTVNDIQKLVGKLNWASQIY 88. These conserved regions were marked in 3D structure using PhyMole as shown in Figure 21.

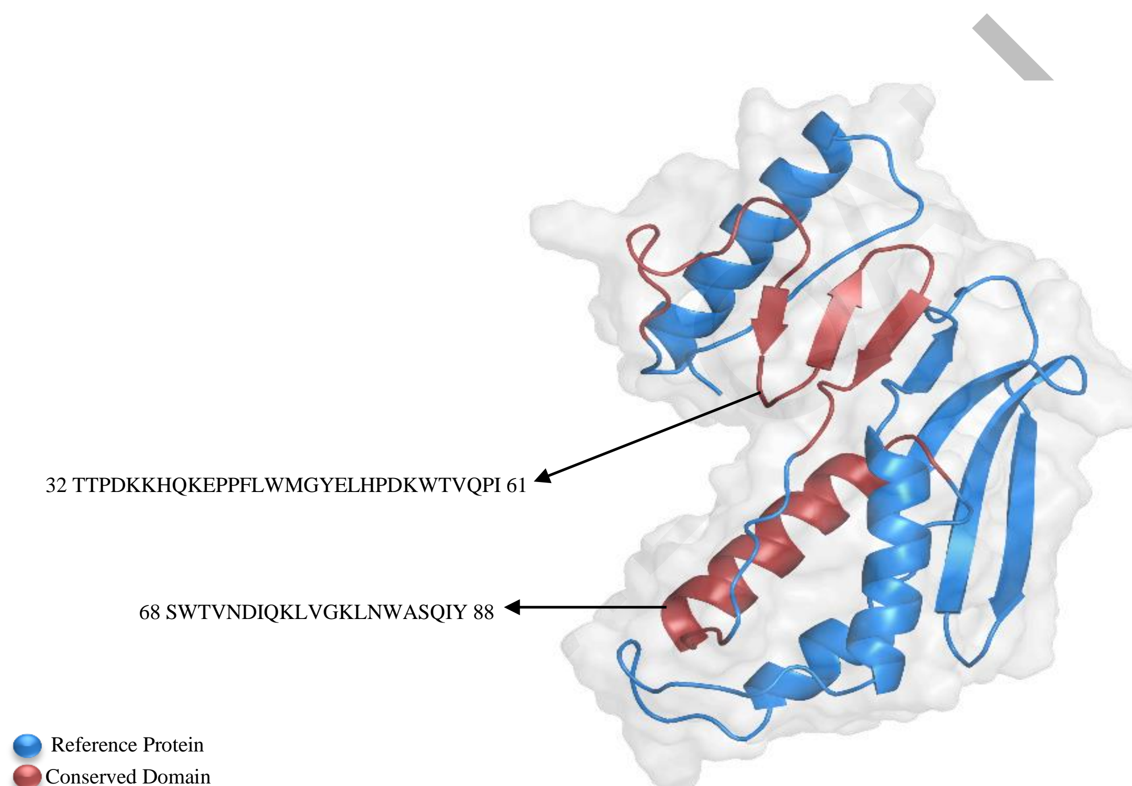

Figure 21. Representation of Conserved region in 3D structure of our isolated protein QAU-AZ1

Bule color represent our isolate while Dark Red color represent conserved regions.

DR

CHAPTER 5

DISCUSSION

DRSMLQAU

5 Discussion

HIV-1 infects around 38 million people worldwide and is considered a major threat for global health. Pakistan has become more prone to HIV-1 spread over the last 10 years and a noticeable increase of 57 percent new HIV-1 infections (Tariq et al., 2022). Therefore, a significant attention is required to control the spread of HIV. The 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 et al., 2010). HIV-1 is known for enormous amount of genetic variation which is associated with recombination of different strains of HIV-1, error during replication and evolution influenced by immunological selection (Tariq et al., 2022). The present study was performed from July 2022 to February 2023 at Infectious Disease and Molecular Pathology Laboratory (IDMPL), Department of Biotechnology, Quaid-i-Azam University Islamabad. A total of 20 HIV positive patients were included in this study, ranging from eight years old to 45 years. 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 RT-PCR for confirmation. 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, the positive samples were processed to assess HIV-Pol gene and its mutations through conventional PCR amplification and nucleotide sequencing.

In current study, gene specific primers were used to amplify HIV-1 *pol* gene of approximately 809bp. The findings of this study showed similarity with the previously reported study by Yaqub et al., 2021. *Pol* gene encoded different types of enzyme like protease (PR), reverse transcriptase (RT) and integrase (IN) which play a crucial role in viral replication (Yaqub et al., 2021). Protease (PR) protein helps in cleave of *Gag-Pol* precursor protein that allow the development of mature HIV virions which may infect CD4 cells (Meher, Vaishnavi, Kumar, Patel, & Kaushik, 2019). Mature HIV virion Reverse transcriptase (RT) is composed of heterodimer referred as *P51* and *P66*. *P66* act as DNA polymerase as well as RNAase H while *P51* provide structural support to *P66* (Sánchez-Murcia et al., 2020). Similarly, integrase is associated with virion maturation and reverse transcription. It is integrated into capsid encased viral cores together with reverse transcriptase and viral RNA (Engelman & Kvaratskhelia, 2022). The HIV-1 pol gene is a crucial marker for precisely defining HIV-1

DRSML QAU

genetic variation (Yaqub et al., 2021). HIV Pol 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 directions for the treatment strategies. The evolutionary genetics and hierarchal relationship data was provided. Nucleotide sequences of the current study were assessed for phylogenetic and evolutionary relationships with other sequences of HIV-1 *pol* gene of Pakistan and other parts of the world. Phylogenetic trees of current sequences were constructed by MEGA 11 software, version 11 (Tamura et al., 2021). The evolutionary relationship of Pol gene was performed by Maximum Likelihood Method with Hasegawa Kishino Yano + Gamma with Invariant sites (HKY+G+I) model. Similar method for phylogenetic analysis was also employed by (Kassaye et al., 2016). The query sequences (isolates QAU-AZ1 and QAU-AZ3) clustered with HIV-1 clone reported in China (Accession no: MN908916) while QAU-AZ2 isolate clustered with HIV-1 isolate 07PK-115 reported in Pakistan (Accession no. JQ011662). All Pakistani isolates makes a separate cluster and are phylogenetically distinct from isolates of other regions of the world. Further studies are needed to find out the subtypes and circulating recombinant fragments (CRFs) to understand the evolutionary relationship of Pakistani isolates with other sequences of the world.

Globally and locally, HIV is distributed according to a very varied molecular and genetic distribution. HIV-1 subtype-A is the most prevalent subtype in Pakistan (Shah et al., 2011). Several research has been conducted earlier which shown the prevalence of this phenomena in Pakistan (Khan et al., 2018). Nucleotide substitutions in the wild-type codon leads to antiviral drug resistance in HIV (Shafer, Kantor, & Gonzales, 2000). Another study was performed by Keulen et. al 1996 on codon alteration in reverse transcriptase and investigated drug resistance to RT enzyme. This concluded that mutation in nucleotide pattern of *pol* plays an important role in drug resistance (Keulen, Boucher, & Berkhout, 1996). Usually, drug resistance associated mutations occur in more than one codon. Production of molecular targeted drugs can specifically suppress HIV transmission and replication due to modern therapeutic strategies (Wahid, 2019; Wensing et al., 2019).

It is hard to treat HIV because to its genetic material persistently sustains variation in the form of substitutions or deletions (Ueda, Witaningrum, Khairunisa, Kotaki, & Kameoka, 2019). A global plan has been adopted by the World Health Organization (WHO) to eradicate

DRSML QAU

this illness through reducing the number of new HIV cases, death, providing accessible medicines, and raising awareness. To successfully address this issue, insights into the worldwide distribution, clinical data, and genetic diversity are needed (Kagan et al., 2019).

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-AZ1, QAU-AZ2 and QAU-AZ3) with WHO recommended Pol reference protein sequence (Accession no. AHF27452), several amino acid substitutions were observed as demonstrated in Figure 16, Figure 17, and Figure 18. There are 23 amino acid substitution in QAU-AZ1, 16 in QAU-AZ2 and 17 in QAU-AZ3. Details of the corresponding substitutions are as follows:

Glycine (G) is non-polar, hydrophobic, and neutral amino acid at position 260 in reference sequence which is substituted by hydrophilic Glutamic acid (E) which is polar -ve charge amino acid in both isolates QAU-AZ1 and QAU-AZ3. Glycine is also replaced by Aspartic acid (D) polar, negative charged and hydrophobic amino acid at position 399 in QAU-AZ1 and QAU-AZ2. Glycine has hydrogen atom in its side chain instead of carbon atom allow conformational flexibility of glycine and play a role in protein structures. Glutamic acid and Aspartic acid (D) are acidic in nature and is commonly present in the active or binding regions of proteins. They may occasionally play a function in the catalytic site of proteins like proteases or lipases (Betts & Russell, 2003).

Threonine (T) is hydrophobic, neutral, and polar amino acid at position 264 in reference protein which substituted by Alanine (A) which is a non-polar, hydrophobic, and neutral amino acid in all isolates QAU-AZ1, QAU-AZ2 and QAU-AZ3. At position 350 Threonine is substituted by Alanine in QAU-AZ1 and QAU-AZ2. Threonine at position 416 changed into Glutamine (Q) in QAU-AZ1. Threonine (T) occurs in protein surfaces or within interior part of the protein. They are present in protein functional sites, which is fairly reactive and form hydrogen bonding with different polar substrates (Frankel & Young, 1998). Alanine is rarely directly involved in the function of proteins because of the presence of non-reactive side chains but alanine can play an important role in the specificity or recognition of the substrate, especially in the interaction with non-reactive atoms (Kempf et al., 2001).

Isoleucine (I) is replaced by Valine (V) at position 266 in QAU-AZ2 isolate while remain conserve in other isolates. Isoleucine is also replaced by Valine (V) at position 390 in

DRSML QAU

QAU-AZ1 and QAU-AZ2. Valine (V) is replaced at position 340 of reference protein by Isoleucine (I) in QAU-AZ3 isolate. Valine at position 356 of reference sequence is substituted by Isoleucine (I). Both amino acids 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).

In all three isolates at position 271 the polar hydrophilic Glutamine (Q) amino acid is replaced by hydrophilic Glutamic acid (E) in all isolate. Glutamine is also replaced by Glutamic acid (E) at position 398 in QAU-1 and QAU-3. Glutamine is neutral amino acid while Glutamic acid (E) is negative charge amino acid. Both are polar in nature and are typically found in the active or binding regions of proteins. The interactions with other polar or charged atoms are because of the presence of polar side chain (Betts & Russell, 2003). The polar hydrophilic Glutamine (Q) amino acid at 342 position is replaced by polar Histidine (H) hydrophobic. Glutamine is neutral amino acid while Histidine positive charge amino acid. Both of these amino acids are involved in active site of protein (Betts & Russell, 2003).

Lysine (K) in reference sequence at position 275 substituted by Arginine (R) at the same position in isolates QAU-AZ1 and QAU-AZ2. Lysine (K) all three isolates at position 361 is replaced hydrophilic Glutamic acid (E) in all isolates. Lysine is substituted by Isoleucine (I) and Asparagine (N) at 411 and 418 position in QAU-AZ1. Both K and R amino acids are polar, positive charged and hydrophilic in nature in involved in active and binding site of protein with other polar side groups (Borders Jr et al., 1994; Stites, 1997). The substitute amino acids both share similar properties, so they did not have an impact on their regulatory functions.

Aspartic acid (D) is polar, negative charged and hydrophobic amino acid which is substituted at position 388 by hydrophilic Glutamic acid (E) in QAU-AZ1. Both amino acids are involved in active site of protein due to similar properties.

Phenylalanine (F) at position 278 in reference sequence is substituted by Leucine (L) in all isolates. Both are hydrophobic in nature, non-polar and neutral amino acid. Phenylalanine and Leucine amino acid side chain is largely non-responsive and infrequently involved in protein function, it plays a role in substrate identification (Betts & Russell, 2003).

Valine (V) is hydrophobic, non-polar and neutral amino acid 309 in reference sequence is substituted by hydrophilic, neutral, and polar amino acid Glutamine (Q) in QAU-AZ1 and

DRSML QAU

QAU-AZ-3 isolates. Valine (V) is substituted by Lysine (K) which is polar, positive charged and hydrophilic in nature in QAU-AZ2 isolate. It is already mentioned that Valine has lower reactivity and are rarely involved in protein functions while Glutamine and lysine is involved in active site regions of protein and interact with other polar atoms (Palanisamy et al., 2017).

At position 314 at reference protein a polar hydrophilic Glutamic acid (E) is replaced by polar hydrophilic Aspartic Acid (D) amino acid in QAU-AZ1 and QAU-AZ3 isolates. Similarly Glutamic acid (E) at 355 position is replaced by Aspartic Acid (D) in QAU-AZ1, QAU-AZ2 and QAU-AZ3. Both Glutamic acid and Aspartic acid are acidic in nature. Glutamic acid and Aspartates both are participated in the binding sites of proteins (Borders Jr et al., 1994; Stites, 1997).

The polar hydrophilic serine (S) amino acid at position 315 is substituted by the polar hydrophilic asparagine amino acid in QAU-AZ2 while remain conserved in other isolates. Both Serine and asparagine are neutral polar amino acids. Serine at position 386 is substituted by hydrophobic Threonine (T) in QAU-1 and QAU-3 which is involve in protein functions. Serine is commonly found in the functional centers of proteins. The serine amino acids had the ability to bind to a series of polar substances due to the presence of a reactive hydroxyl group. The Asparagine amino acid is commonly found in the binding or active sites of protein. The polar side chain of the asparagine is important for the interaction with other charged or polar atoms (Simon et al., 2002).

Proline (P), a neutral, non-polar and hydrophobic amino acid at position 336 of the reference sequence is substituted to a hydrophobic, non-polar, neutral amino acid Alanine (A) and at position 358 it is changed to Threonine (T) which is hydrophobic, neutral, and polar amino acid in QAU-AZ2 isolate. Alanine (A) is hydrophobic, non-polar and neutral amino acid which is replaced with Valine at position 390 in QAU-AZ1 isolates. Proline being less reactive is often involved in protein active or binding site while Alanine, Threonine and Valine involved in protein functions (Betts & Russell, 2003).

The polar hydrophilic Glutamine (Q) amino acid at position 342 is replaced by polar hydrophobic Histidine (H) in QAU-AZ2. Glutamine is neutral amino acid while Histidine positive charge amino acid. Similarly, Glutamine (Q) amino acid at position 398 is replace by hydrophilic -ve charge amino acid Glutamic acid (E) in QAU-AZ1. These amino acids are involved in active site of protein (Russell, 1998).

DRSML QAU

Leucine (L) in reference sequence at position 347 is replaced by Isoleucine (I) in QAU-AZ2 and QAU-AZ3. Subsequently, Leucine is substituted by Methionine (M) at position 359 in QAU-AZ2. All these amino acids are hydrophobic, non-polar, neutral in nature which are non-reactive and rarely involved in the protein function (Yu, Huang, & Wang, 2017). Methionine shares same properties as of Leucine. Like other hydrophobic amino acids, its role is in recognition or binding of hydrophobic ligands (Salzwedel et al., 1999).

The identified amino acid substitutions are mostly involved in protein binding or active site and have vital role in protein function. Therefore, almost no changes were observed in 3D structure of protein. The query sequences of all isolates in this study were located between palm and connection region in RTp66 subunit of reverse transcriptase protein which was confirmed in recently reported 3D structure of reverse transcriptase protein (Harrison et al., 2022). Moreover, in current study two conserved regions 32 TTPDKKHQKEPPFLWMGYELHPDKWTVQPI 61 and 68 SWTVNDIQKLVGKLNWASQIY 88 were also identified. Furthermore, each isolate is needed to pass out through different bioinformatic tools to check the activity of recently available drugs.

DRSML QAU

Conclusion

The World Health Organization has estimated 39 million new HIV cases globally. HIV prevalence in Pakistan is still rising and currently seems unstoppable. Understanding nucleotide sequence diversity and the resulting amino acid substitutions is crucial for the complete eradication of HIV. The HIV pol gene was amplified and sequenced in the current study, its phylogenetic analysis conceals that the current sequences (isolate QAU-AZ1, QAU-AZ2 and QAU-AZ3) grouped with reference sequences from China (MN908916) and Pakistan (JQ011662) respectively. This study also reports a number of amino acid substitutions in structural and functional domains of *pol* protein. These amino acid substitutions have no effect on 3D structure of protein. More in depth studies are required to find the association of these amino acid substitutions with resistance against antiretroviral drugs. It is further suggested to study the viral isolates from Pakistani patients for possible drug resistant mutations to prescribe specific antiretroviral therapies.

DRSML QAU

REFERENCES

DRSML QAU

REFERENCES

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.

Abram, M. E., & Parniak, M. A. (2005). Virion instability of human immunodeficiency virus type 1 reverse transcriptase (RT) mutated in the protease cleavage site between RT p51 and the RT RNase H domain. *Journal of virology, 79*(18), 11952-11961.

Ahmed, F. J. (2022). Scenario of HIV infection in Pakistan. *Pakistan Journal of Health Sciences*.

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.

Alkhatib, G. (2009). The biology of CCR5 and CXCR4. *Current Opinion in HIV and AIDS, 4*(2), 96.

Ananworanich, J., & Robb, M. L. (2014). The transient HIV remission in the Mississippi baby: why is this good news? *Journal of the International AIDS Society, 17*(1).

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.

Arhel, N. (2010). Revisiting HIV-1 uncoating. *Retrovirology, 7*(1), 1-10.

Arts, E. J., & Hazuda, D. J. (2012). HIV-1 antiretroviral drug therapy. *Cold Spring Harbor perspectives in medicine, 2*(4), a007161.

Asahchop, E. L., Wainberg, M. A., Sloan, R. D., & Tremblay, C. L. (2012). Antiviral drug resistance and the need for development of new HIV-1 reverse transcriptase inhibitors. *Antimicrobial agents and chemotherapy, 56*(10), 5000-5008.

DRSML QAU

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.

Babe, L. M., Rose, J., & Craik, C. S. (1995). Trans-dominant inhibitory human immunodeficiency virus type 1 protease monomers prevent protease activation and virion maturation. *Proceedings of the National Academy of Sciences, 92*(22), 10069-10073.

Barmania, F., & Pepper, M. S. (2013). CC chemokine receptor type five (CCR5): an emerging target for the control of HIV infection. *Applied & translational genomics, 2*, 3-16.

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.

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.

Bennett, J. E., Dolin, R., & Blaser, M. J. (2019). *Mandell, douglas, and bennett's principles and practice of infectious diseases E-book*: Elsevier Health Sciences.

Betts, M. J., & Russell, R. B. (2003). Amino acid properties and consequences of substitutions. *Bioinformatics for geneticists*, 289-316.

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.

Blood, G. A. C. (2016). Human immunodeficiency virus (HIV). *Transfusion Medicine and Hemotherapy, 43*(3), 203.

DRSML QAU

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.

Chalmet, K., Staelens, D., Blot, S., Dinakis, S., Pelgrom, J., Plum, J., . . . Verhofstede, C. (2010). Epidemiological study of phylogenetic transmission clusters in a local HIV-1 epidemic reveals distinct differences between subtype B and non-B infections. *BMC infectious diseases, 10*(1), 1-9.

Chattopadhyay, D., Evans, D., Deibel Jr, M., Vosters, A., Eckenrode, F., Einspahr, H., . . . Heinrikson, R. t. (1992). Purification and characterization of heterodimeric human immunodeficiency virus type 1 (HIV-1) reverse transcriptase produced by in vitro processing of p66 with recombinant HIV-1 protease. *Journal of Biological Chemistry, 267*(20), 14227-14232.

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.

Churchill, M. J., Chiavaroli, L., Wesselingh, S. L., & Gorry, P. R. (2007). Persistence of attenuated HIV-1 rev alleles in an epidemiologically linked cohort of long-term survivors infected with nef-deleted virus. *Retrovirology, 4*(1), 1-10.

DRSML QAU

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.

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.

Contento, R. L., Molon, B., Boularan, C., Pozzan, T., Manes, S., Marullo, S., & Viola, A. (2008). CXCR4–CCR5: a couple modulating T cell functions. *Proceedings of the National Academy of Sciences, 105*(29), 10101-10106.

Das, K., & Arnold, E. (2013). HIV-1 reverse transcriptase and antiviral drug resistance. Part 1. *Current opinion in virology, 3*(2), 111-118.

de Béthune, M.-P. (2010). Non-nucleoside reverse transcriptase inhibitors (NNRTIs), their discovery, development, and use in the treatment of HIV-1 infection: a review of the last 20 years (1989–2009). *Antiviral research, 85*(1), 75-90.

Dean, M., Carrington, M., Winkler, C., Huttley, G. A., Smith, M. W., Allikmets, R., . . . Gomperts, E. (1996). Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. *Science, 273*(5283), 1856-1862.

Deeks, S. G., Overbaugh, J., Phillips, A., & Buchbinder, S. (2015). HIV infection. *Nature reviews Disease primers, 1*(1), 1-22.

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

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

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

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

DRSML QAU

Ebina, H., Misawa, N., Kanemura, Y., & Koyanagi, Y. (2013). Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus. *Scientific reports, 3*(1), 1-7.

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.

Engelman, A. N., & Kvaratskhelia, M. (2022). Multimodal functionalities of HIV-1 integrase. *Viruses, 14*(5), 926.

Fanales-Belasio, E., Raimondo, M., Suligoi, B., & Buttò, S. (2010). HIV virology and pathogenetic mechanisms of infection: a brief overview. *Annali dell'Istituto superiore di sanita, 46*, 5-14.

Fox, J., Castro, H., Kaye, S., McClure, M., Weber, J. N., Fidler, S., & Resistance, U. C. G. o. H. D. (2010). Epidemiology of non-B clade forms of HIV-1 in men who have sex with men in the UK. *Aids, 24*(15), 2397-2401.

Frankel, A. D., & Young, J. A. (1998). HIV-1: fifteen proteins and an RNA. *Annual review of biochemistry, 67*(1), 1-25.

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.

Gallay, P., Hope, T., Chin, D., & Trono, D. (1997). HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway. *Proceedings of the National Academy of Sciences, 94*(18), 9825-9830.

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.

Ganser-Pornillos, B. K., Yeager, M., & Sundquist, W. I. (2008). The structural biology of HIV assembly. *Current opinion in structural biology, 18*(2), 203-217.

DRSML QAU

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.

Gelderblom, H. R. (1991). Assembly and morphology of HIV: potential effect of structure on viral function. *Aids, 5*(6), 617-638.

Giovanetti, M., Ciccozzi, M., Parolin, C., & Borsetti, A. (2020). Molecular epidemiology of HIV-1 in African countries: a comprehensive overview. *Pathogens, 9*(12), 1072.

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.

Gouws, E., White, P. J., Stover, J., & Brown, T. (2006). Short term estimates of adult HIV incidence by mode of transmission: Kenya and Thailand as examples. *Sexually transmitted infections, 82*(suppl 3), iii51-iii55.

Gross, K. L., Porco, T. C., & Grant, R. M. (2004). HIV-1 superinfection and viral diversity. *Aids, 18*(11), 1513-1520.

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

DRSML QAU

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.

Hare, S., Vos, A. M., Clayton, R. F., Thuring, J. W., Cummings, M. D., & Cherepanov, P. (2010). Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance. *Proceedings of the National Academy of Sciences, 107*(46), 20057-20062.

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

Hockemeyer, D., Wang, H., Kiani, S., Lai, C. S., Gao, Q., Cassady, J. P., . . . Miller, J. C. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. *Nature biotechnology, 29*(8), 731-734.

Hoffmann, C., Rockstroh, J., & Kamps, B. S. (2003). *HIV medicine*: Flying Publisher Paris, Cagliari, Wuppertal, Sevilla.

Holkers, M., Maggio, I., Liu, J., Janssen, J. M., Miselli, F., Mussolino, C., . . . Goncalves, M. A. (2013). Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic acids research, 41*(5), e63-e63.

Honda, M., & Oka, S. (2006). Current therapy for human immunodeficiency virus infection and acquired immunodeficiency syndrome. *International journal of hematology, 84*(1), 18-22.

Hütter, G., Nowak, D., Mossner, M., Ganepola, S., Müßig, A., Allers, K., . . . Blau, O. (2009). Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *New England Journal of Medicine, 360*(7), 692-698.

Ifeanyi, O. E., Amilo Grace, I., Uzoma, O. G., & Nnatuanya Isaac, N. (2017). Human Immunodeficiency Virus infection and cardiovascular. *Int. J. Curr. Res. Med. Sci, 3*(9), 9-37.

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.

DRSML QAU

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.

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.

Joung, J. K., & Sander, J. D. (2013). TALENs: a widely applicable technology for targeted genome editing. *Nature reviews Molecular cell biology, 14*(1), 49-55.

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.

Kanta, V., Unnati, S., & Ritu, M. (2011). A review on aids and herbal remedies. *International Journal of Research in Ayurveda & Pharmacy, 2*(6), 1709-1713.

Kapila, A., Chaudhary, S., Sharma, R., Vashist, H., Sisodia, S., & Gupta, A. (2016). A review on: Hiv aids. *Indian Journal of Pharmaceutical and Biological Research, 4*(3), 69-73.

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.

Kassaye, S. G., Grossman, Z., Balamane, M., Johnston-White, B., Liu, C., Kumar, P., . . . Dewar, R. (2016). Transmitted HIV drug resistance is high and longstanding in metropolitan Washington, DC. *Clinical infectious diseases, 63*(6), 836-843.

DRSML QAU

Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in bioinformatics, 20*(4), 1160-1166.

Kaur, R., Sharma, P., Gupta, G. K., Ntie-Kang, F., & Kumar, D. (2020). Structure-activity-relationship and mechanistic insights for anti-HIV natural products. *Molecules, 25*(9), 2070.

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.

Keulen, W., Boucher, C., & Berkhout, B. (1996). Nucleotide substitution patterns can predict the requirements for drug-resistance of HIV-1 proteins. *Antiviral research, 31*(1-2), 45-57.

Khalili, K., Kaminski, R., Gordon, J., Cosentino, L., & Hu, W. (2015). Genome editing strategies: potential tools for eradicating HIV-1/AIDS. *Journal of neurovirology, 21*(3), 310-321.

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*, 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.

Laila, U., Akram, M., Shariati, M. A., Hashmi, A. M., Akhtar, N., Tahir, I. M., . . . Akhter, N. (2019). Role of medicinal plants in HIV/AIDS therapy. *Clinical and Experimental Pharmacology and Physiology, 46*(12), 1063-1073.

DRSML QAU

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.

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.

Li, T. (2022). Chinese Guidelines for the Diagnosis and Treatment of HIV/AIDS (2021 Edition). *Infectious Diseases & Immunity*.

Li, Z., Zhou, N., Sun, Y., Ray, N., Lataillade, M., Hanna, G. J., & Krystal, M. (2013). Activity of the HIV-1 attachment inhibitor BMS-626529, the active component of the prodrug BMS-663068, against CD4-independent viruses and HIV-1 envelopes resistant to other entry inhibitors. *Antimicrobial agents and chemotherapy, 57*(9), 4172-4180.

Lindegaard, B., Hansen, A.-B. E., Pilegaard, H., Keller, P., Gerstoft, J., & Pedersen, B. K. (2004). Adipose tissue expression of IL-18 and HIV-associated lipodystrophy. *Aids, 18*(14), 1956-1958.

Liu, R., Paxton, W. A., Choe, S., Ceradini, D., Martin, S. R., Horuk, R., . . . Landau, N. R. (1996). Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell, 86*(3), 367-377.

Lopalco, L. (2010). CCR5: from natural resistance to a new anti-HIV strategy. *Viruses, 2*(2), 574-600.

Luciw, P. A. (1996). Human immunodeficiency viruses and their replication. *Fields virology, 2*, 1881-1952.

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

Lusso, P. (2006). HIV and the chemokine system: 10 years later. *The EMBO journal, 25*(3), 447-456.

DRSML QAU

Mao, X.-Y., Dai, J.-X., Zhou, H.-H., Liu, Z.-Q., & Jin, W.-L. (2016). Brain tumor modeling using the CRISPR/Cas9 system: state of the art and view to the future. *Oncotarget, 7*(22), 33461.

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.

Mascola, J. R., Snyder, S. W., Weislow, O. S., Belay, S. M., Belshe, R. B., Schwartz, D. H., . . . Gorse, G. J. (1996). Immunization with envelope subunit vaccine products elicits neutralizing antibodies against laboratory-adapted but not primary isolates of human immunodeficiency virus type 1. *Journal of Infectious Diseases, 173*(2), 340-348.

McCutchan, F. E. (2000). Understanding the genetic diversity of HIV-1. *AIDS (Supplement), 14*(3), S31-S44.

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.

Meher, B. R., Vaishnavi, M., Kumar, M. V. S., Patel, S., & Kaushik, S. (2019). Mutation Effects on 3D-Structural Reorganization Using HIV-1 Protease as a Case Study.

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*(S1), 26-30.

DRSML QAU

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. (2022). Advances in testing for infectious diseases—looking back and projecting forward. *Clinical chemistry, 68*(1), 10-15.

Persaud, D., Gay, H., Ziemniak, C., Chen, Y. H., Piatak Jr, M., Chun, T.-W., . . . Luzuriaga, K. (2013). Absence of detectable HIV-1 viremia after treatment cessation in an infant. *New England Journal of Medicine, 369*(19), 1828-1835.

Peruski, A. H., Wesolowski, L. G., Delaney, K. P., Chavez, P. R., Owen, S. M., Granade, T. C., . . . Brooks, J. T. (2020). Trends in HIV-2 diagnoses and use of the HIV-1/HIV-2 differentiation test—United States, 2010–2017. *Morbidity and Mortality Weekly Report, 69*(3), 63.

Pettit, S. C., Everitt, L. E., Choudhury, S., Dunn, B. M., & Kaplan, A. H. (2004). Initial cleavage of the human immunodeficiency virus type 1 GagPol precursor by its activated protease occurs by an intramolecular mechanism. *Journal of virology, 78*(16), 8477-8485.

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., Ali, H., Fernandez, D., Knuth, M., Schenkel, K., Asghar, R. J., . . . Morgan, O. (2021). Systematic review of reported HIV outbreaks, Pakistan, 2000–2019. *Emerging infectious diseases, 27*(4), 1039.

DRSML QAU

Rachinger, A., Manyenga, P., Burger, J. A., Derks van de Ven, T. L., Stolte, I. G., Prins, M., . . . Schuitemaker, H. (2011). Low incidence of HIV-1 superinfection even after episodes of unsafe sexual behavior of homosexual men in the Amsterdam Cohort Studies on HIV Infection and AIDS. *Journal of Infectious Diseases, 203*(11), 1621-1628.

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.

Ramana, L. N., Anand, A. R., Sethuraman, S., & Krishnan, U. M. (2014). Targeting strategies for delivery of anti-HIV drugs. *Journal of controlled release, 192*, 271-283.

Rambaut, A., Posada, D., Crandall, K. A., & Holmes, E. C. (2004). The causes and consequences of HIV evolution. *Nature Reviews Genetics, 5*(1), 52-61.

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.

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

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.

DRSML QAU

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.

Samson, M., Libert, F., Doranz, B. J., Rucker, J., Liesnard, C., Farber, C.-M., . . . Forceille, C. (1996). Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature, 382*(6593), 722-725.

Sánchez-Murcia, P. A., de Castro, S., García-Aparicio, C., Jiménez, M. A., Corona, A., Tramontano, E., . . . Gago, F. (2020). Peptides mimicking the β7/β8 loop of HIV-1 reverse transcriptase p51 as "hotspot-targeted" dimerization inhibitors. *ACS Medicinal Chemistry Letters, 11*(5), 811-817.

Sendagire, H., Easterbrook, P. J., Nankya, I., Arts, E., Thomas, D., & Reynolds, S. J. (2009). The challenge of HIV-1 antiretroviral resistance in Africa in the era of HAART. *AIDS reviews, 11*(2), 59.

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.

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.

DRSML-QAU

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.

Swanson, C. M., & Malim, M. H. (2008). SnapShot: HIV-1 proteins. *Cell, 133*(4), 742-742. e741.

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Molecular biology and evolution, 38*(7), 3022-3027.

Tanahashi, T. (2017). Diversity of secondary metabolites from some medicinal plants and cultivated lichen mycobionts. *Yakugaku Zasshi: Journal of the Pharmaceutical Society of Japan, 137*(12), 1443-1482.

Tang, M. W., & Shafer, R. W. (2012). HIV-1 antiretroviral resistance. *Drugs, 72*(9), e1-e25.

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

Tariq, U., Nazziwa, J., Sasinovich, S., Shah, S. A., Naeem, S., Abidi, S. H., & Esbjörnsson, J. (2022). Phylogenetic Characterization of HIV-1 Sub-Subtype A1 in Karachi, Pakistan. *Viruses, 14*(10), 2307.

Tebas, P., Stein, D., Tang, W. W., Frank, I., Wang, S. Q., Lee, G., . . . Nichol, G. (2014). Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *New England Journal of Medicine, 370*(10), 901-910.

DRSML QAU

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.

Tilton, J. C., & Doms, R. W. (2010). Entry inhibitors in the treatment of HIV-1 infection. *Antiviral research, 85*(1), 91-100.

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.

van Heuvel, Y., Schatz, S., Rosengarten, J. F., & Stitz, J. (2022). Infectious RNA: Human Immunodeficiency Virus (HIV) Biology, Therapeutic Intervention, and the Quest for a Vaccine. *Toxins, 14*(2), 138.

Vicenzi, E., & Poli, G. (2013). Novel factors interfering with human immunodeficiency virus-type 1 replication in vivo and in vitro. *Tissue Antigens,* 81(2), 61-71.

Virus, H. I. (2016). Transfus Med Hemother German Advisory Committee Blood (Arbeitskreis Blut). Subgroup 'Assessment of Pathogens Transmissible by Blood, 43(3), 203-222.

Visseaux, B., Bertine, M., Le Hingrat, Q., Ferré, V., Charpentier, C., Collin, F., . . . Descamps, D. (2021). HIV-2 diversity displays two clades within group A with distinct geographical distribution and evolution. *Virus evolution,* 7(1), veab024.

Volberding, P., Sande, M. A., Greene, W. C., Lange, J. M., & Gallant, J. E. (2008). Global hiv/aids medicine: *Elsevier Health Sciences.*

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.

Wainberg, M. A., Mesplède, T., & Quashie, P. K. (2012). The development of novel HIV integrase inhibitors and the problem of drug resistance. *Current opinion in virology*, 2(5), 656-662.

DRSML QAU

Wang, Z., Pan, Q., Gendron, P., Zhu, W., Guo, F., Cen, S., . . . Liang, C. (2016). CRISPR/Cas9-derived mutations both inhibit HIV-1 replication and accelerate viral escape. *Cell reports,* 15(3), 481-489.

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.

Williams, I. G. (2013). Treatment strategies for HIV infection. *Medicine,* 41(8), 470-473.

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.

Yaqub, S., Yaqub, T., Shabbir, M. Z., Nadeem, A., Shahid, M. F., Tahir, Z., & Mukhtar, N. (2021). Prevalence, Molecular Epidemiology and Drug Resistance Pattern of Human Immunodeficiency Virus Type 1 among Injecting Drug Users in Lahore, Pakistan. *Pakistan Journal of Zoology,* 53(2), 563.

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.

DRSML QAU

Zennou, V., Petit, C., Guetard, D., Nerhbass, U., Montagnier, L., & Charneau, P. (2000). HIV-1 genome nuclear import is mediated by a central DNA flap. *Cell,* 101(2), 173-185.

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.

Zulfiqar, H. F., Javed, A., Afroze, B., Ali, Q., Akbar, K., Nadeem, T., . . . Husnain, T. (2017). HIV diagnosis and treatment through advanced technologies. *Frontiers in public health*, 5, 32.

DRSML QAU

PUBLICATIONS

1. **Zeb, A**., Ali, S. S., Azad, A. K., Safdar, M., Anwar, Z., Suleman, M., ... & Wei, D. Q. (2021). Genome-wide screening of vaccine targets prioritization and reverse vaccinology aided design of peptides vaccine to enforce humoral immune response against Campylobacter jejuni. *Computers in Biology and Medicine*, *133*, 104412.
2. Khan, S., Rizwan, M., **Zeb, A**., Eldeen, M. A., Hassan, S., Ur Rehman, A., ... & Ullah, A. (2022). Identification of a Potential Vaccine against Treponema pallidum Using Subtractive Proteomics and Reverse-Vaccinology Approaches. *Vaccines*, *11*(1), 72.
3. Shah, S. A. A., Qureshi, N. A., Qureshi, M. Z., Alhewairini, S. S., Saleem, A., & **Zeb, A**. (2023). Characterization and bioactivities of M. arvensis, V. officinalis and P. glabrum: In-silico modeling of V. officinalis as a potential drug source. *Saudi Journal of Biological Sciences*, *30*(6), 103646.
4. Ghafoor, D., **Zeb, A**., Ali, S. S., Ali, M., Akbar, F., Ud Din, Z., ... & Khan, W. (2023). Immunoinformatic based designing of potential immunogenic novel mRNA and peptide-based prophylactic vaccines against H5N1 and H7N9 avian influenza viruses. *Journal of Biomolecular Structure and Dynamics*, 1-18.

DRSML QAU

ORIGNALITY REPORT

adnan thesis

ORIGINALITY REPORT

15%	10%	11%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

No.	Source	%
1	Sana Ullah, Muhammad Ali, Asmat Shaheen, Fatima Zia et al. "Sofosbuvir Resistance-associated Substitutions in the Palm Domain of HCV-NS5B RNA Dependent RNA Polymerase; Study of two Sofosbuvir non-responders", International Journal of Infectious Diseases, 2021 Publication	1%
2	www.frontiersin.org Internet Source	1%
3	"Encyclopedia of AIDS", Springer Science and Business Media LLC, 2018 Publication	1%
4	discovery.ucl.ac.uk Internet Source	1%
5	www.researchgate.net Internet Source	1%
6	Submitted to University of Wolverhampton Student Paper	<1%