Development of liquid chromatography mass spectrometric fingerprinting based method for HIV-1 integrase detection

By

Tikedzani Geraldine Vele

A dissertation submitted in fulfilment of the requirements for the degree of Master of Science in the subject of Biochemistry

At
The University of Venda
Department of Biochemistry and Microbiology

Supervisor: Dr N.E Madala
Co-supervisor: Mr L.M Mathomu

Submitted on

April 2021

Table of Contents

Declaration

I, **Tikedzani Geraldine Vele**, declare that this thesis submitted to the University of Venda for the Master of Science degree in Biochemistry under the School of Mathematical and Natural Sciences has not been submitted to any other University. This is my work with the exception of referenced material which has been properly cited in text and acknowledged.

Signature: ______________ **Date: 28/04/2021**

Dedication

This thesis is dedicated to my mother Vele Mashudu, my brother Vele Ipfi Nigel and the Netshiombo family.

Acknowledgements

I thank the Lord almighty for giving me strength to complete this study.

To my supervisor Dr Ntakadzeni E. Madala, thank you for pushing me out of my comfort zones to learn various techniques, thank you for always being present when help is needed, and taking us along on the path to success. Your knowledge and commitment has helped me reach a massive milestone in my life. Thank you for seeing a potential in me.

To my co-supervisor Mr Lutendo Mathomu, thank you for moulding the scientist you knew was there. I am eternally grateful for your mentorship.

I would also like to express my sincere thanks and gratitude to Dr Tshifhiwa Tambani, for providing necessary support, guidance and motivation.

I thank Prof Pascal Bessong of the HIV/AIDS Research Unit, at the University of Venda for the construct and the lab materials offered when needed.

I wish to further extend my acknowledgement to the following institutions for funding this project;

- The Department of Science and Innovation (DSI) and the Council for Scientific and Industrial Research (CSIR) of South Africa
- The University of Venda Research and Publication Committee (RPC)

I would like to extend my heartfelt gratitude to my friend Kgabo O. Phadu, thank you for your tremendous emotional support that helped me complete this research.

I am extremely grateful for my family and friends for their prayers, support, encouragement and their love that they showed throught out this whole project.

Lastly, I would like to thank my lab mates and the department of Biochemistry for their moral support and motivation throughout this study.

Preface

This thesis is comprised of four chapters. The outlines for each chapter are provided below.

Chapter 1: This is a general introduction encompassing all the background to the suggested broad aim and specific objectives as well as its identification by LC-MS. It also highlights the broad problem statement and spells out the study hypothesis.

Chapter 2: This chapter present the methods for recombinant production of IN protein and identification by LC-MS.

Chapter 3: This chapter reports on the purification of IN protein and the identification, detection of HIV-1 integrase by theoretical peptide fingerprinting, ESI based MS method and Multiple Reaction Monitoring (MRM).

Chapter 4: This chapter covers conclusion remarks.

Abstract

Human Immunodeficiency virus/ acquired immunodeficiency syndrome (HIV/AIDS) is a relatively modern disease that has caused extensive morbidity, mortality and suffering worldwide. HIV/AIDS is associated with different enzymes that are responsible for viral replication and one of such enzyme is the HIV-1 integrase. Integrase is responsible for the incorporation of viral DNA into the host cell, the enzyme is emerging as a novel target for intervention by chemotherapeutics, thus making the scientific investigation of HIV-1 integrase an important field of research. Currently, Enzyme-Linked Immunosorbet Assay (ELISA) based method are used for diagnosis of HIV infections across the world. However, other methods which are capable of detecting HIV-1 proteins with high sensitivity and specificity are imperative. Mass Spectrometry (MS) is known to offer unprecedented sensitivity and specificity during protein identification and, as such, its use for medical use is of greater interest. Therefore, this study sought to develop an MS method which can be applied for detection and characterization of HIV-1 integrase. The main objective of the current study was to identify and detect HIV-1 integrase using Multiple Reaction Monitoring (MRM). Detection of HIV-1 integrase using MRM will provide a method which may serve as a basis for diagnosis HIV/AIDS infection. Using Nested PCR and Sanger sequencing, pET15b and integrase construct were successfully confirmed. To this end, recombinant HIV-1 integrase was expressed in *E. coli* BL21 (DE3) cells. The recombinant HIV-1 integrase was successfully purified using nickel-affinity chromatography. Using electrospray ionization (ESI) based MS method, it was observed that integrase produces peptides masses that matches with theoretical masses, however some of the peptide masses differs due to multiple charge state. In addition, using precursor ion 600.84 for MRM, it was noted that the precursor ion produce three product ions acquired during ionization in the ESI chamber. Furthermore, the peak intensity of the product ions were noted to increase as the collision energy increases. These findings make the precursor ion 600.84 and its product ions a fingerprint that can be used as a tool integrase detection. As a proof of concept, the current method can be further developed by means of incorporating other MS ions originating from HIV-1 proteins, thereby increasing the likelihood of a future diagnosis.

List of symbols and abbreviations

Abbreviations of units	Symbol Interpretation
A600	Absorbance at 600 nanometres
AmpR	Ampicillin
bp	Base pair
CE	Collision energy
CID	Collision induced dissociation
CCD	Catalytic core domain
CXCR	Chemokine receptors
CTD	C-terminal domain
DDA	Data dependent acquisition
ºC	Degree Celsius
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotides
ELISA	Enzyme linked immunosorbert assay
ESI	Electronspray ionization
g	Gram
HIV-1	Human immunodeficiency virus 1
HIV/AIDS	Human immunodeficiency virus, acquired immunodeficiency syndrome
IEC	Ion exchange chromatography
IN	Integrase
IPTG	Isopropyl β-d-1-thiogalactopyranoside
kDa	Kilodalton
LC- MS	Liquid chromatography mass spectrometry
LC-Q-TOF	Liquid chromatography quadrupole time of flight
LEDGF	Lens epithelium-derived growth factor
LTR	Long terminal repeats

M	Molar
MgCl	Magnesium chloride
µl	Microlitre
µg	Microgram
µg/mL	Microgram per millilitre
mg/mL	Milligram per millilitre
mM	Millimolar
mL	Millilitre
mg	Milligram
MRM	Multiple reaction monitoring
MS/MS	Tandem mass spectrometry
MS	Mass spectrometry
m/z	Mass/charge
NaCl	Sodium chloride
NaH_2PO_4	Monosodium phosphate
ng	Nanogram
NCBI	National center for biotechnology information
Ni-NTA	Nickel-charged nitriloacetic acid
NTD	N-terminal domain
PMF	Peptide mass fingerprinting
%	Percent
PCR	Polymerase chain reaction
PIC	Pre-integration complex
PMSF	Phenylmethylsulfonyl fluoride
Q1	First quadrupole
Q-TOF	Quadrupole time-of flight
RNA	Ribonucleic acid

RT	Reverse transcriptase
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
STF	Strand transfer reaction
v/v	Volume per volume
w/v	Weight per volume

List of figures

List of tables

Chapter 1: Rationale and literature review

1.1 Background and motivation

Since its inception in 1981, HIV/AIDS has been a severe life-threatening disease known to cause millions of deaths. With Africa being the most affected region, wherein two-thirds of new infections are found. In Sub-Saharan Africa, HIV/AIDS is the leading cause of death and the fourth largest killer worldwide (WHO, 2016).

In 2012, HIV/AIDS has accounted for 70% of newly infected individuals in Sub Saharan Africa. HIV/AIDS figures illustrate its devastating consequences with both Eastern and Southern Africa responsible for 6% of infection of the global population, making up 52% of all people living with HIV/AIDS. In 2015, there were approximately 36.9 million people living with HIV/AIDS and 2.1 million newly infected with HIV/AIDS worldwide (UNAIDS, 2016).

HIV-1 integrase is an important enzyme that is involved in the replication of HIV in the host cell. The role of integrase is the development of a functional provirus that incorporates the viral DNA into the genetic material of the cell. It is an immunogenic and retained human immunodeficiency virus protein. Integrase has p31 antigen, that is known to help improve the specificity of techniques for HIV detection.

The diagnosis of HIV-1 is crucial for detecting and monitoring the virus. HIV/AIDS diagnosis has been related to different uses. Blood screening has helped protect countless individuals from HIV infection since 1985 (Branson, 2007). In addition, HIV-1 tests are also used for epidemiological monitoring, providing community-based health authorities with information on the level of infection, enabling them to prioritize the population for vaccine administration and care (Branson, 2007).

Currently, there are screening tests available that are used wolrdwide. Enzyme linked Immunosorbent Assay (ELISA), the nucleic acid test (contains DNA and RNA nucleotides), and the antigen-antibody test are included in these experiments, such tests, with different mechanism modes, are considered to diagnose HIV.

ELISA is known to detect blood antibodies. The nucleic acid test diagnoses both RNA (ribonucleic acid) and DNA (deoxyribonucleic acid) and, ultimately, the antibody/antigen test diagnoses both antigen and antibody (Butto *et al.,* 2010). While these techniques have been improved to unparalleled levels of sensitivity and specificity over the past few years, the identification of new variants has emphasized the sensitivity limits of existing assays (Parekh *et al.,* 2019). For the identification and quantification of protein, adequate sensitivity and multiplexing capacity are therefore urgently required.

A technique that would enhance the sensitivity of HIV detection must be established. Peptide mass fingerprinting is a well recognized method used to classify proteins isolated by Liquid Chromatography Quadrupole Time Of Flight (LC-QTOF) (Singh *et al.,* 2019). Apart from peptide mass fingerprinting, it is another tool can be used to identify highly sensitive proteins. In addition, detection of protein identified by peptide mass fingerprinting, multiple reaction monitoring (MRM), which is a tandem MS technique, can be used as HIV detection tool. However, no studies have been done to detect HIV proteins using peptide mass fingerprinting and multiple reaction monitoring, so this research will further explore the development of a tool for detecting HIV-1 integrase.

The research questions are summarized as follows:

a) Will trypsin enzyme digest HIV-1 integrase?
b) How will the purity of HIV-1 integrase affect identification of peptides?
c) How is HIV-1 integrase detected using peptide mass fingerprinting?
d) How does MRM detect HIV-1 integrase?

This study will develop a liquid chromatography mass fingerprinting method for HIV-1 integrase detection. This will be achieved through by looking at expression, purification of integrase (IN) as well as identifying the IN protein by peptide mass fingerprinting and multiple reaction monitoring.

1.2 Aim and objectives of the study

The aim of this study is to develop an LC-QTOF-MS based MRM technique for detection of HIV-1 integrase protein for subsequent HIV diagnosis.

The objectives of the study are:

1.2.1. To confirm the HIV-1 integrase gene/pET15b plasmid construct.

1.2.2. To express HIV-1 integrase in BL21 (DE3) cells and purify recombinant HIV-1 integrase protein using nickel affinity chromatography.

1.2.3. To develop a multiple reaction monitoring (MRM) method through LC-QTOF-MS for generation of HIV-1 integrase peptide fragment.

1.3 Literature review

1.3.1 HIV/AIDS infection

Human immunodeficiency virus (HIV) infection is accountable for a worldwide pandemic, with 36.7 million people infected with the virus. In 2016, UNAIDS reported that HIV was highly serious in Eastern and Southern Africa, with 19.4 million people living with HIV in 2016, making Southern Africa to accumulate for 64% of all new HIV infections. However, with increased efforts to tackle the pandemic, the number of new infections decreased from 300 000 to 160 000 in 2016, a 47% reduction over the 2010-2016 period. A steady decrease in HIV-related deaths was also recorded, with approximately 1 million people dying in 2016, as compared to 1.9 million in 2005. This solid decline is likely due to improvement in access to the treatment of antiretroviral drugs (UNAIDS, 2016).

Two major types of HIV virus have been identified as, HIV-1 and HIV-2, wherein HIV-1 is the predominant type. HIV-1 has excessive heterogeneity that is instigated through numerous factors including the error prone proof reading ability of reverse transcriptase (RT) (Op de Coul *et al.,* 2001), the excessive turnover rate of HIV *in vivo* (Ho *et al.,* 1995), selective immune pressures in the host (Michael, 1999) and recombination during the replication cycle (Temin, 1993).

In addition, HIV-1 is subdivided into four main classes because of such effects; M (Major), O (Outlier), N (Non-M and Non-O), and P (Pending the identification of further human cases) (Buonaguro *et al.,* 2007). The zoonotic transmission of the simian immunodeficiency virus emerged in these four classes (SIV). Of the four groups, M is the largest group globally observed and responsible for the prevailing HIV pandemic (Sharp & Hahn, 2011; Hemelaar, 2012).

Group M is similarly subdivided from A-K into 11 subtypes (Soto-Rifo *et al.,* 2011). In Europe, America, Japan and Australia, subtype B is prevalent, while subtype C is prevalent in Southern Africa, Eastern Africa, India, Nepal and China (Hellmund & Lever, 2016). With these nations accounting for most of the infections, subtype C is the most extensively spread subtype in the world (Buonaguro *et al.,* 2007). Buonaguro et al (2007) stated that groups N, O and P are less common in many Western and Central African countries, where group O accounts for approximately 5% of infections, and is very rare outside these regions. HIV infection is no longer fatal, but due to advances in antiretroviral drug treatment, it is now considered a chronic and manageable disease (WHO, 2021).

1.3.1.1 HIV virion and genomic organization

Three open reading frames (ORF) *gag*, *pol* and *env* are included in the HIV-1 genome; these three ORFs encode nine genes that encode at least 16 proteins that are necessary for the virus to function and survive (Arrildt *et al.,* 2012). The genomic organization and the resultant protein for each gene are depicted in Figure 1.1.

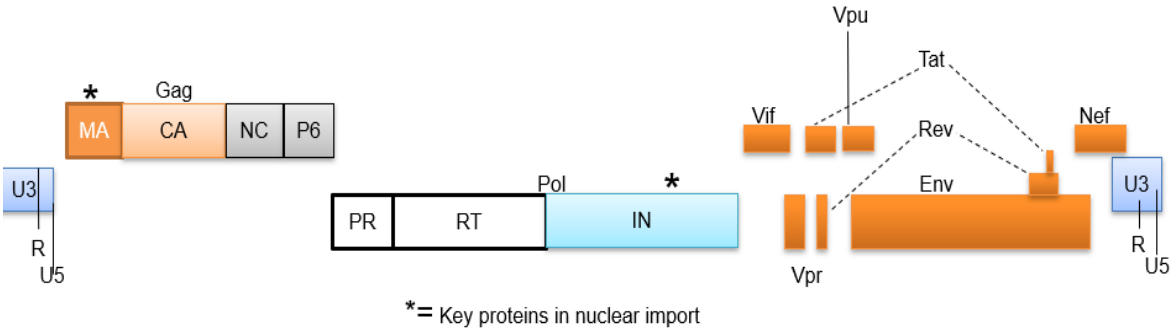


Figure 1. 1. The representation of the genome of the HIV-1 virus. There are 9 essential genes that encode for 16 proteins and are flanked by identical long terminal repeats. Gag, polymerase and envelope glycoproteins are encoded by *gag, pol* and *env* structural genes. The matrix (MA, dark orange), capsid (CA, light orange), nucleocapsid (NC, grey), P6, protease (PR, white), reverse transcriptase (RT) and integrase (IN, blue) are all formed by the cleavage of the gag-pol polyprotein. *Tat* and *rev* regulatory genes encode transcription transactivator and rev (dark orange) regulatory factor. Accessory genes such as vif,vpr, vpu

and nef (orange) encode viral infectivity factor, viral protein R, negative factor and viral protein U. Adopted from Costin, 2007.

As shown in Figure 1.2, the subsequent proteins assemble to form different portions of the virion. The *gag* gene translates into structural proteins (McGovern *et al.,* 2002), while the *pol* gene is responsible for generating the enzymes important to the process of replication (Klein *et al.,* 2010), and the *env* gene forms the glycoprotein protective coat (Figure 1.2; Simon *et al.,* 2006). Two plus strands of RNA each containing a copy of the nine genes make up the virion. The capsid, made up of p24, encloses the RNA and is surrounded by the p17 matrix (Figure 1.2; Hossein & Mac Gabhhan, 2012) . The matrix is enclosed by a host derived lipid bilayer where the glycoproteins gp120 and gp41 are embedded. The virion contains most of the components needed to replicate in a host cell with the aid of host cofactors (Figure 1.2; Hossein & Mac Gabhhan, 2012).

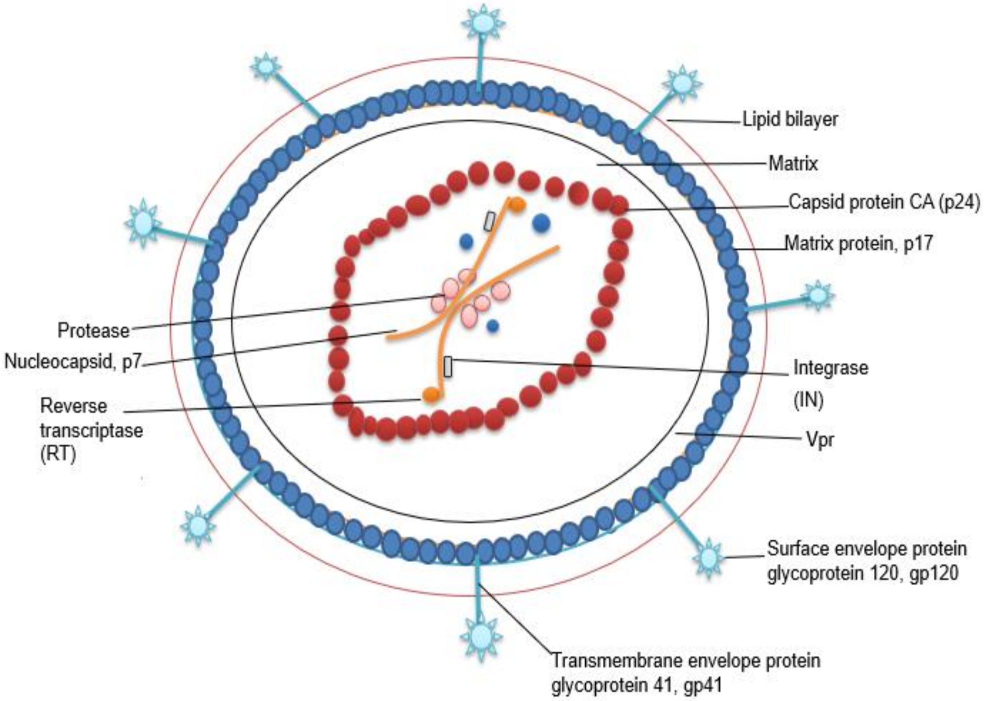


Figure 1. 2*.* Demonstrative figure of an HIV virion displaying structural and functional proteins. Capsid protein (p24), matrix protein (p17), nucleocapsid (p7), reverse transcriptase, integrase and protease, as

well as envelope glycoproteins gp120 and gp41, are all found in the virion. Adapted from Hosseini and Mac Gabhann, 2012.

1.3.1.2 HIV-1 Lifecycle

The HIV-1 lifecycle consists of seven phases: binding, viral fusion (entry), reverse transcription, integration, viral replication and eventually the assembly and budding of viral particles (Figure 1.3). CD4 cell has two major co-receptors on its surface known as chemokine receptors (CCR5 and CXCR4), which have been identified as the main co-receptors for T-cell line tropics and macrophages of HIV-1 isolates (Wang *et al.,* 2017). These receptors are known to allow passage of the virus into the host cell (Chen, 2019). When the HIV-1 trimeric gp120 binds to the CD4 receptor, the chain of events begins, inducing a conformational shift in gp120 that allows binding to co-receptors of chemokine; CXCR4 or CCR5 (Figure 1.3; Chan & Kim, 1998). The binding of the co-receptor allows viral entry and fusion (Chen, 2019). The viral envelope fuses with the membrane of the host cell, allowing the viral capsid to enter the cytoplasm of the host cell (Wyatt & Sudroski, 1998). The ribonuclease (RNA) genome, reverse transcriptase (RT), RNaseH and integrase are found in the viral capsid. The capsid is disassembled and reverse transcriptase (RT) convert the RNA genome into cDNA (Figure 1.3; Zheng *et al.,* 2005). The cDNA is then assembled into the complex of pre-integration (PIC). Viral cDNA, RT, IN, Vpr, capsid, matrix and host proteins are included in the PIC, which contains, among other proteins, lens epithelial derived growth factor (LEDGF/p75). The PIC contains all that is required for nuclear transport and viral cDNA integration into the host DNA (Panwar & Singh, 2021). Once integration has occurred, the integrated viral genome is transcribed and translated by host cellular enzymes (polymerase, nuclease and ligase), resulting in a newly synthesized viral RNA genome and polyproteins. The polyproteins Gag/GagPol and Envelope are split to form functional proteins by viral or host proteases, respectively. HIV-1 integration will be discussed in more detail in section 1.3.2.

All the components assemble near the cell surface and an immature virion emerges or bud from the infected cell surface (Figure 1.3; Butsch & Boris-Lawrie, 2002). The cleavage of the GagPol polyprotein to form a mature virion is completed by viral protease. The new viral particles can then invade neighboring cells and infect them (Rojas & Park, 2019).

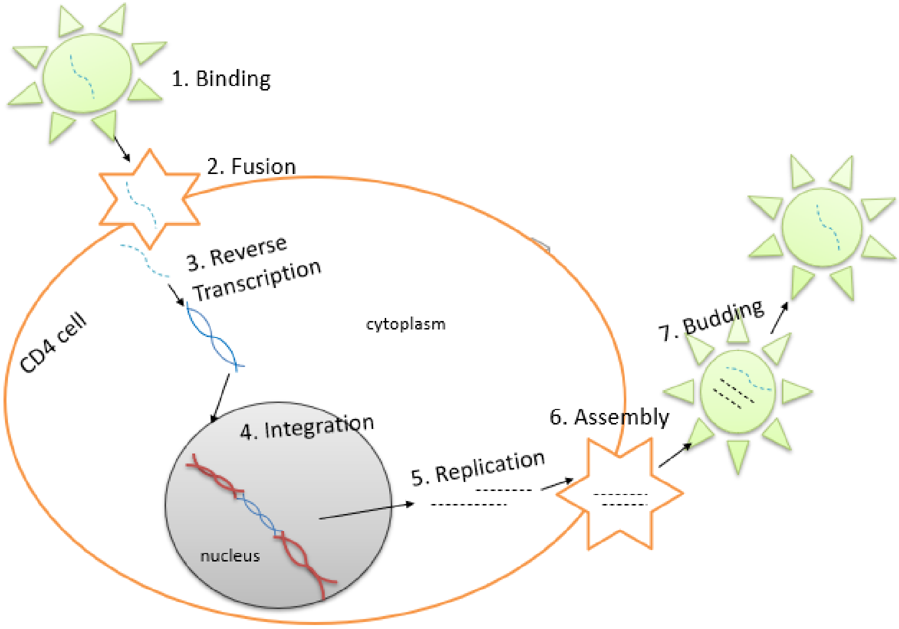


Figure 1. 3. Lifecycle of HIV-1 showing the individual steps of entry, reverse transcription, integration, transcription, translation, assembly and budding. Adapted from Rambaut *et al.,* 2004.

1.3.2 Biological mechanism of integrase

Integrase (IN) is a 288 amino acids protein that consists of three functional domains (Figure 1.4), the N-Terminal Domain (NTD), Catalytic Core Domain (CCD); and the C-Terminal Domain (CTD); and each of the integrase domains contain recognizable functional motifs (Wang *et al.,* 2001; Karki *et al.,* 2005; Santo, 2014). For instance, the N-terminal domain (residues 1-49) contains two histidine residues (His 12 and His 16) and two cysteine residues (Cys40 and Cys43), all of which are definitely conserved and form an HH-CC zinc-finger motif (Figure 1.4; Schweitzer *et al.,* 2013) that chelates one zinc atom per integrase monomer (Park *et al.,* 2019). The zinc atom serve as stabilization to the NTD folding and is pivotal to integrase activity (Elliot *et al.,* 2020). The CCD consists of three conserved negatively charged amino acids, ASP64, ASP116 and Glu152 (also known as DDE motif), which coordinate divalent metal ions crucial for the strand transfer (STF) reaction and comprise β sheet and alpha helices bound by flexible loop (Figure 1.4; Wang *et al.,* 2001).

The flexible loops allow conformational modifications needed for the 3’ end viral DNA processing and the stand transfer reaction, which are key steps of the integration reaction (Esposito & Craigie, 1999). Wang *et al* (2001) reported that substitution of any of the residues inside the DDE motif significantly inhibits integrase activity. Of the three domains, the C-terminal domain is the least conserved and has structural homologous with SH3 DNA binding domains and non-specifically binds DNA (Karki *et al.,* 2005; Quashie *et al.,* 2015). Overall, coordinated action between the three domains is, however, necessary to catalyze the 3’-end processing and STF reactions effectively (Elliot *et al.,* 2020).

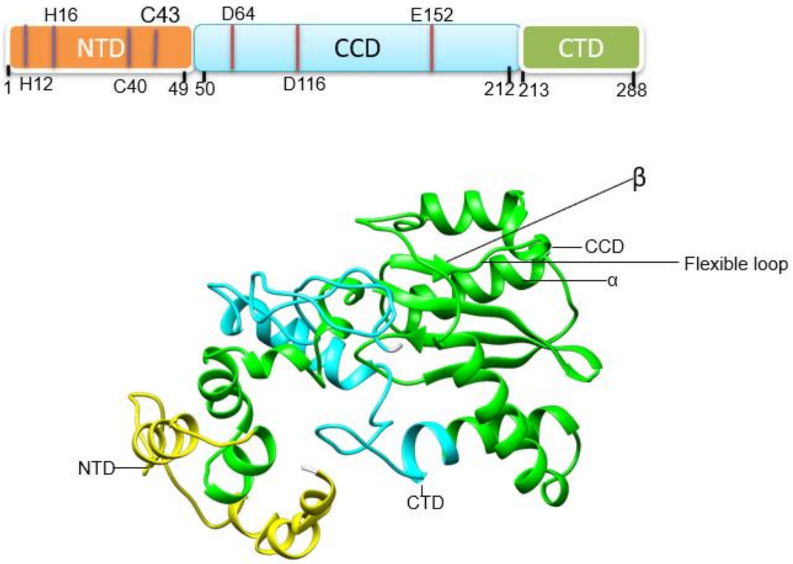


Figure 1. 4. Schematic annotation of three dimensional ribbon structure of HIV-1 integrase and its domains. The 3D model of HIV-1 integrase showing three structural domains. The N-terminal domain (NTD) represented in yellow consist of the HHCC motif (1-49), the Catalytic Core Domain (CCD) represented by green consist of the DDE triad (50-212) and the C-Terminal Domain (CTD) represented by cyan blue (213-288).

After completion of reverse transcription, the role of integrase during HIV-1 replication begins in the cytoplasm (Santo, 2014). Integration of viral DNA into the host genome confirms persistent infection with HIV-1, leading to both active and latent viral reservoirs (Delelis *et al.,* 2008). Integrase cleaves GT (guanine-thymine) dinucleotides to a conserved CA (cysteine-adenine) dinucleotide at both 3' ends of the viral DNA with the aid of using nucleophilic attack (Delelis *et al.,* 2008; Santo, 2014). The 3'-EP reaction creates the reactive 3-OH ends and integrase remains bound to long terminal repeats (LTR) ends, forming the complex preintegration (PIC) (Craigie, 2001).

Once the PIC is targeted at the host genomic DNA, the strand transfer reaction occurs within the nucleus (Craigie, 2012). Bound to the integrase, viral DNA binds to the host genomic DNA and uses the 3'-OH ends of the viral DNA produced during the 3'-EP reaction to conduct nucleophilic attack reactions on the host genomic DNA phosphodiester bonds (Li *et al.,* 2006). Both ends of the viral DNA remain similar to one another and the 3'-OH ends of viral DNA are ligated to the resulting 5'-phosphate ends of the host genomic DNA, with a 5-base pair stagger separating the ligation points in the genomic DNA (Figure 1.5; Brin *et al.,* 2000). The two nucleotides at the 5'end of viral DNA form a "flap" and are then trimmed; gap filling from the 3'-end of the host genomic DNA, completes integration (Schweitzer *et al.,* 2013).

The completion of the STF (strand transfer reaction) produces a functional integrated proviral DNA that forms the transcription template for new viral RNAs needed for the translation of viral proteins and enzymes and the transcription of new full-length viral RNAs needed for packaging into new virons (Jozwik *et al.,* 2020). Any integrated virus can create thousands of new viruses, amplifying the initial infection and the CD4^{+} T-cell death being the ultimate cost (Delelis *et al.,* 2008).

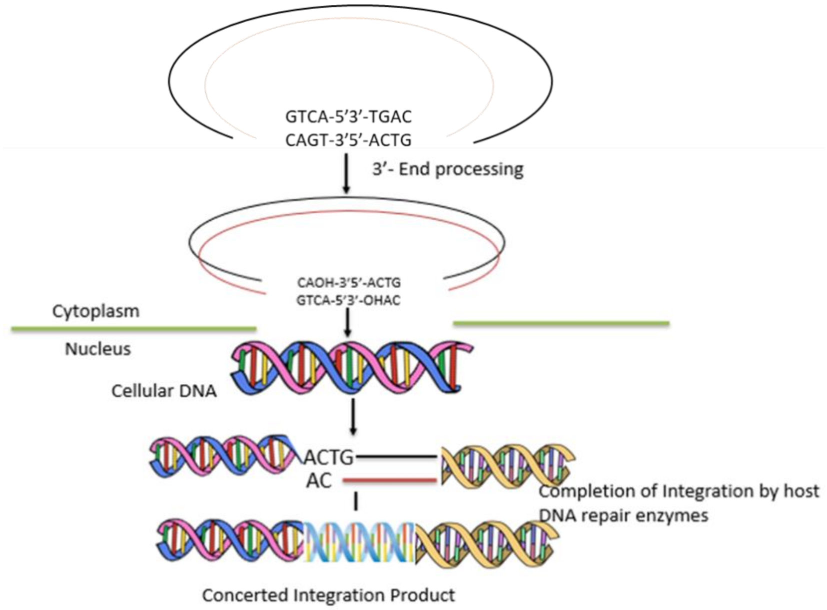


Figure 1. 5. Inforgraphical display of the biological mechanism essential for proviral DNA integration in *vivo*. The first step during the integration process is the binding of an integrase on the long terminal repeats to catalyze the 3' end processing, resulting in endonucleotically cleavage of dinucleotides. Inside the nucleas, the cleaved DNA serve as a substrate for integration onto the host DNA (Genome). The cleaved DNA attacks the host DNA's phosphodiester bond via a process known as strand transfer reaction. Cellular enzymes complete the integration by repairing the resulting integration intermediate. Adopted from Santo, 2014.

1.3.3 Current HIV-1 detection techniques

HIV-1 infection has been extensively diagnosed through a variety of approaches that specifically detect antibody, antigen, and DNA/RNA (Butto *et al.,* 2010). Herein, this section present a detailed assessments on available technologies for the detection and monitoring of HIV infection and to introduce more modern technologies that could provide significant improvements in diagnostics, surveillance, blood screening and disease monitoring (Butto *et al.,* 2010).

1.3.3.1 Enzyme-Linked ImmunoSorbent Assay

One of the methods used to detect antibodies in HIV infected patient's blood is an enzyme-linked immunosorbent assay (UNAIDS, 2006). Virtually 100% of HIV-1 infected individuals are found with HIV-specific antibodies, and HIV infection can be detected by testing the presence of those HIV-1 specific antibodies (IgM, IgG) (Armstrong & Taege, 2007). It was stated that the sensitivity of this method depends on the concentration of the virus (Gurtler *et al.,* 1998; Tehrani *et al.,* 2015).

1.3.3.2 Nucleic acid diagnostic technique

Nucleic acid tests (NAT) are commercial tests that can identify HIV nucleic acid (either RNA or proviral DNA) (Ochodo *et al.,* 2018). NAT can supplement antibody test for HIV diagnosis in special cases such as suspected acute infection, when antibodies are nevertheless undetectable and in the case of newborn HIV infected mothers (Huang *et al.,* 2017). As a prognostic marker for antiretroviral therapy and estimation of infection, the quantitative detection of HIV RNA in plasma is used ('viral load' test) (Jani *et al.,* 2014). The nucleic acid test uses HIV-1 subtype B infection primers thus detect other HIV-1 subtypes with low sensitivity through polymerase chain reaction (PCR) (Fiebig *et al.,* 2003).

1.3.3.3 Antigen diagnostic technique

HIV infection may also be diagnosed by the detection of virus components, preferably virus-specific antibodies. However, HIV antigen (p24 antigen) can only be detected intermittently in plasma throughout the asymptomatic period (James *et al.,* 2014). The detection of viral nucleic acid can be achieved by different laboratory techniques which can determine either the proviral cDNA in leukocytes or the viral RNA in the cell-free compartment (Lange *et al.,* 1986). These strategies have the advantage that the assay, once developed, can be practical to a large number of samples with a high degree of robustness and low cost, such tests are sensitive with low detection limits.

Although a diverse approach to unprecedented levels of sensitivity and specificity has been refined over the last few years, the identification of new variants has highlighted the sensitivity limits of existing assays (Fiebig *et al.,* 2003).

Huttenhain *et al* (2009) reported that the highly sensitive and specific ELISA allows the detection of low-concentration proteins ranging from ng/ml to pg/ml in plasma. Various immunoassays have been developed for the measurement of multiple proteins in a single assay (Gomez *et al.,* 2010). However, commercially available immunoassays are very costly and ELISA multiplexing is likely to be limited due to cross-reactivity of the antibodies (Guan, 2007). Accessibility of specific antibodies in contrast to innovative candidate proteins and time required for the development of new ELISA assays creates additional tailbacks in the biomarker pipeline (Hanly *et al.,* 2010). High-satisfactory sensitivity and multiplexing functionality are therefore urgently needed for identifying and detecting proteins.

1.3.4 Liquid chromatography mass spectrometry protein diagnostic technique

Proteomics is an emerging technology that detect and quantify protein variance expressions in a different states, and study protein-protein interaction characteristics. Depending on time and cellular status, the complex existence of protein expression, protein interactions, and protein modification requires measurement (Lin *et al.,* 2003). These types of studies require a number of measurements and therefore it is important to identify high-performance proteins.

Liquid chromatography (LC-MS) is an analytical technology that combines chromatography with MS to improve resolution by separating large amounts of analytes in the LC column (Khadir & Tiss, 2013). Mass Spectrometry (MS) measures mass to charging ratio *(m/z)* of charged particles (ions). Although the mass spectrometers are different, all use electrical or magnetic fields to control the movement of ions generated by the analyte of interest and determine their *m/z* (Chen & Pramanik, 2008). The fundamental components of a mass spectrometer include the ion source, the mass analyser, the detector and the vacuum and data systems. The ion source is used to ionize sample components injected in an MS system by electron beam, UV beams, laser beams, or corona discharge ionized components (Chen & Pramanik, 2008; Khadir & Tiss, 2013).

LC-MS is multidimensional and may be used to detect large molecular weights of the molecule after trypsin digestion (Ho *et al.,* 2003). Trypsin is a serine protease that cleaves proteins into peptides with an average size of 700-1500 daltons, the ideal range for MS and is regarded as the golden standard for protein digestion into peptides (Laskay *et al.,* 2013). Moreover, it is highly specific and cut arginine and lysine residues on the carboxyl side, therefore making the C-terminal peptides of arginine and lysine charged, thus detectable by MS (Laskay *et al.,* 2013). The key peptide/protein ionization methods currently used are ESI and MALDI methods and have been related to high-throughput sample preparation techniques (Lin *et al.,* 2003).

Electrospray ionization (ESI) is a technique used in mass spectrometry to produce ions using an electrospray in which a high voltage is applied to a liquid to produce an aerosol (Loo *et al.,* 1989; Pitt, 2009). It is particularly helpful in the development of ions from macromolecules as it overcomes the tendency of these ionizing molecules to fragment. ESI varies from other ionization processes (e.g. MALDI) because it can generate multiple-charged ions (Nadler *et al.,* 2017) and expand the mass of the analyzer efficiently to fit the kDa-MDa magnitude levels found in protein and its related polypeptide fragments (Lin *et al.,* 2003).

During electrospray ionization, the ion source converts and fragments the neutral sample molecules into gas phase ions that are sent to the mass analyser (Ho *et al.,* 2003; Nadler *et al.,* 2017). While the mass analyzer uses electrical and magnetic fields to sort ions by mass, the detector measures and amplifies the ion current to calculate the abundance of each mass-resolved ion (Loo *et al.,* 1989). In order to generate a mass spectrum that the human eye can easily recognize, the data system records, processes, stores and displays data on a computer.

In general, ESI process produce ions that are multiply charged. When ESI is operated at a positive ion mode, ions are formed as a result of protonation, thus the ionic species detected are not the true molecular ions (formed by the loss or gain of an electron) but rather protonated or deprotonated molecules (Banerjee & Mazumdar, 2012), which are mainly referred to as precursor or pseudo-molecular ions. Therefore, the mass spectra displays ions which are either protonated nor depronated and some can either carry one of multiple charges depending on the number of proton added or subtracted, respectively (Figure 1.6). Apart from proton addition (M + H)+ or abstraction (M – H)-, other ion species such as sodium ion cation (M + Na)+ are also prominent in ESI based mass spectrometry (Fenn *et al.,* 1989; Liuni & Wilson, 2011). Chen and Pramanik (2008) reported that large macromolecules can have multiple charge states, resulting in a distinct charge state envelope. Unlike some other ionization sources, all of these are even-electron ion species: electrons (alone) are not added or removed. Analytes are used in electrochemical processes, causing shifts in the corresponding mass spectrum peaks (Banerjee & Mazumdar, 2012).

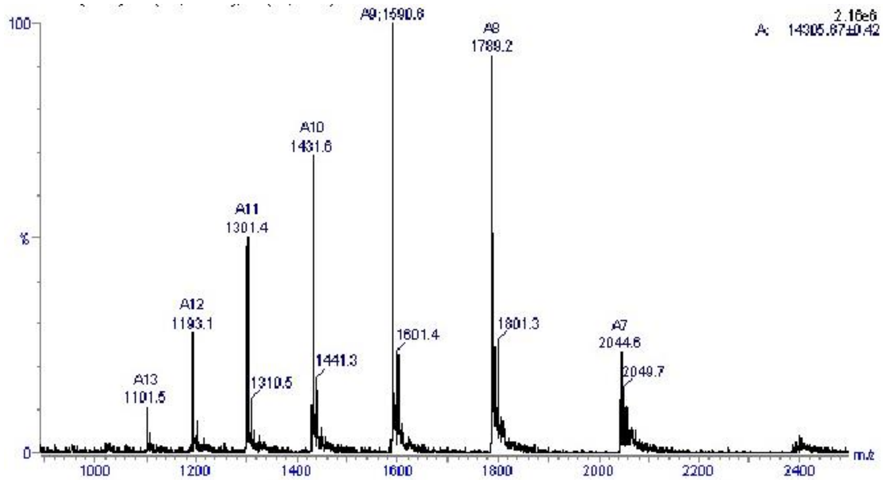


Figure 1. 6. A typical ESI mass spectrum of hen egg white lysozyme. Adopted from Mano and Goto, 2003

Among the many different kinds of mass analyzers, ESI is used with quadrupole-time of flight (Q-TOF) in both MS and tandem MS (MS/MS) modes. Quadrupole-TOFs are tools that combine a quadrupole mass analyzer for ion selection in tandem mass spectrometer mode with TOF analyzers to record high-resolution *m/z* values (Ho *et al.,* 2003; Pitt & James, 2009). In a collision cell, ions are disassociated between the two mass analyzers (Lin *et al.,* 2003). The values of *m/z* are chosen by their TOF in this tool and all others are deflected from the flight path (Chen & Pramanik, 2008). The ions then enter the collision cell and are exposed to high-energy collisions with inert gas, such as helium or argon, which cause the ions to fragment (Lin *et al.,* 2003). This rapid ion fragmentation is referred to as collision-induced dissociation (CID). For proteomic studies, all of the above mentioned mass spectrometers are efficient and distinguish on the basis of mass range, mass accuracy, sensitivity, resolution and cost.

1.3.4.1 Peptide mass fingerprinting approach for MS-based protein identification

Peptide mass fingerprinting (PMF) is gaining momentum in the identification of proteins analysis by LC-Q-TOF-MS (Hamza *et al.,* 2020). One of the key technologies driving growth of proteomics is the detection of mass peptides and therefore, the widespread use of PMF is chosen due to its simple and effective process for the detection of peptides with a specific mass and it is one of the key technologies driving the growth of proteomics (Hamza *et al.,* 2020).

A protein is a set of amino acids arranged in a specific sequence to produce a specific activity or property. Although some proteins share a high degree of homology (sequence similarity) with other proteins, some, if not many, protein sequences are unique (Jim, 2007; Jakubke & Sewald, 2008). If the protein could be cut predictably, the size of the pieces should form a fingerprint for that specific protein (Hamza *et al.,* 2020). Furthermore, if a protein sequence could be cut in silicon, the fingerprint produced can be used to identify such protein (Hamza *et al.,* 2020).

A protein must first be cleaved with a proteolytic enzyme (trypsin), this is achieved for various reasons, some proteins are large and therefore need to be digested into smaller peptides (Figure 1.7), in particular due to the fact that large endogenous proteins produce highly complex mass spectra that interfere with the analyte ions used for quantification, additionally proteins are difficult to manage (Henzel *et al.,* 2003). Secondly, mass spectroscopy is used to predict the actual mass of the peptides. This step offers the catalog of the peak inventory of the identified peptides (Henzel *et al.,* 2003). Thirdly, the obtained peaks are analyzed by comparing the masses of peptides which are found on an online databases (https://bio.tools/MASCOT). HIV-1 integrase has some synthetic peptides identifiable by LC-MS (Pala *et al.,* 2020). Additionally, some of the peptides are known to be involved in the inhibition process of HIV-1 IN. For instance, peptide DQAEHLK (147-175) which is known to be found on the catalytic core domain has been reported that IN inhibitors binds to the peptide for further IN inhibition process. It was further reported that it inhibit the strand transfer and 3'-processing activities (Sourgen *et al.,* 1996). Pala *et al* (2020) reported some peptides ions that could form a fingerprint that can be widely used for detection of IN, therefore diagnosing HIV/AIDS.

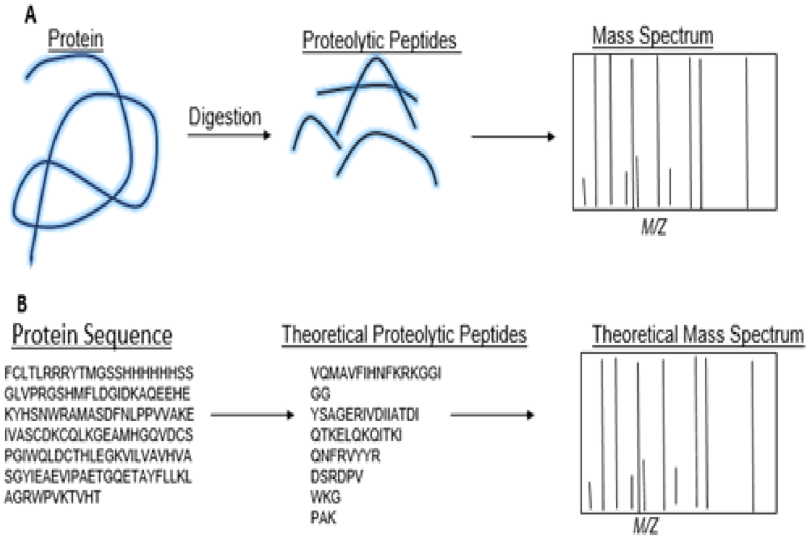


Figure 1. 7. Schematic representation of various steps followed during peptide mass fingerprinting. Adopted from Graves and Haystead, 2002.

Peptide mass fingerprinting has been widely used for the diagnosis of various diseases. For example, Kitamura *et al* (2017) used PMF for early diagnosis of Alzheimer's disease. Patient samples were analyzed using two-dimensional differential gel electrophoresis (2D-DIGE) combined with matrix-assisted laser desorption ionization time of the flight tandem (MALDI-TOF) mass spectrometry followed by peptide mass fingerprinting. It was reported that three down-regulated proteins have been identified as candidate for early diagnosis of Alzheimer's disease (Kitamura *et al.,* 2017).

Agranoff *et al* (2006) identified the diagnosis of tuberculosis biomarkers by proteomic fingerprinting. It was reported that 20 peaks were obtained with 90% accuracy, where two peptides (serum amyloid A protein and transthyretin) were identified (Agranoff *et al.,* 2006). The diagnostic accuracy obtained was 94%. Furthermore, the identified potential biomarkers for tuberculosis possess a biological connection with the disease and could be used to develop new diagnostic tests (Agranoff *et al.,* 2006).

Key advantages of using PMF is that it does not rely on protein sequencing for protein identification (Liebler, 2002). After PMF, protein has to be detected using MRM.

1.3.5 Multiple reaction monitoring technique

Multiple reaction monitoring is an extremely specific and sensitive label-free approach for the recognition of targeted peptides (Ong and Mann, 2005). It refers to the tandem MS scan mode coupled with triple quadrupole or hybrid quadrupole/trap MS instrumentation (Sherwood *et al.,* 2009), both types of instrumentation work very similarly in terms of their ion cycle and the singular release of *m/z* at any one time over the entire *m/z* range and are therefore capable of selecting predefined ions for analysis (Ong and Mann, 2005).

It is important to design transitions for the targeted protein before starting MRM, and this involves the selection of peptides (also known as precursor ions) and their corresponding product ions (Figure 1.8; Ong and Mann, 2005). Usually two or four transitions are selected for the target peptide and multiple peptides for each protein (Feldberg *et al.,* 2019). Based on the pre-designed transition lists, the primary quadrupole of the MS will be able to select and transmit the precursor ions to the second quadrupole for further fragmentation (Anderson and Hunter, 2006). The resulting product ions will then be transmitted to the third quadrupole (Figure 1.8), which detects only product ions with selected predefined *m/z* ions (Liebler & Zimmerman, 2013).

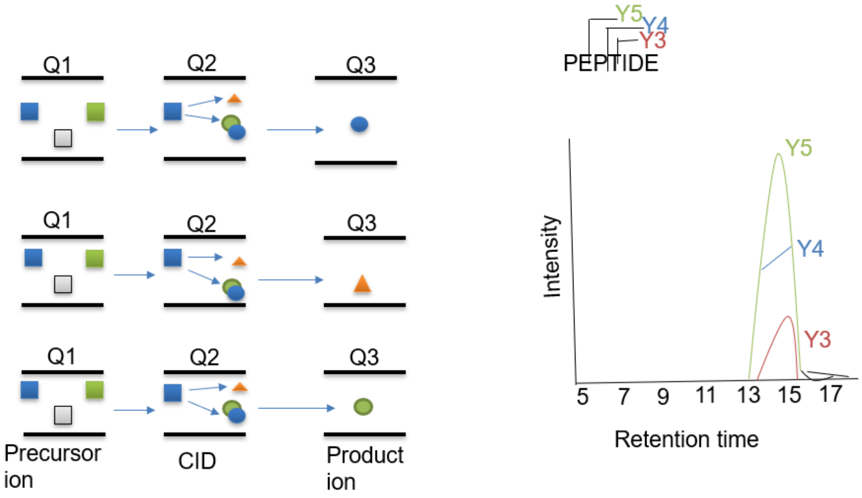


Figure 1. 8. Schematic representation of mass spectrometer triple-quadrupole operated in MRM mode. Source-generated ions are separated in first quadrupole (Q1), the targeted peptide then pass into the collision induced dissociation (CID) cell, where it is fragmented and allowed through the third quadrupole (Q3) into the detector. Adopted from Domon and Aebersold, 2006.

The specificity of the MRM is ensured by the two selection steps. At least three product ions become a unique signature in combination with the precursor ion for the MRM analysis to be successful and can then specifically target the peptide/protein for both relative and absolute quantification (Anderson and Hunter, 2006; Kuzyk *et al.,* 2011). The ability to specifically quantify target peptides allows MRM to be a valuable technique. Compared to the widely used immunoassay technology that is based on antigen-antibody interaction, the MRM method has two advantages: It does not require the use of antibody, thus eliminating the time needed to develop the antibody (Kuzyk *et al.,* 2009) and it is capable of instantly quantifying up to 40 peptides (Parker *et al.,* 2010).

Considering the clinical value of MS-based assays, direct contrasts are regularly made with ELISA, which is considered to be the golden technique for protein quantification in clinical samples. Aspects of ELISAs, such as "time to first result" (1–2 hours) and the ability to quantify 96 or 384 samples in parallel due to their plate-based microtiter format, are currently difficult to match with MS-based protein assays (Kingsmore, 2006). MRM protein assays can be regarded superior than ELISA based methods because they are able to target multiple proteins at a time, thus multiplexed protein assays. Because of the impact of multiplexed assays on genomics, there is a greater interest in multiplexed protein quantification in individual clinical samples (Kingsmore, 2006). Furthermore, compared to the time that it takes to develop antibodies for ELISA method, development of MRM takes relatively short time (Kingsmore, 2006). Finally. MRM-based strategies are superior because of samples throughput, where dosens of samples can be anlaysed for several biomarker evaluation in a relatively shorter time (Zheng *et al.,* 2007).

Chapter 2 : Materials and Methods

2.1. Materials

The plasmid constructs used are described (table 2.1). The reagents used are listed in (Appendix B).

Table 2. 1. *E. coli* strains and plasmid constructs used for expression of recombinant protein

Competent cells	Description
E. coli BL21(DE3)	*FhuA2 [Ion] ompT (ʎ DE3) [dcm] ΔhadS* ʎ DE3= ʎ sBamHIo ΔEcoRI-B int (lacl PlacUV5 T7 gene1) i21 Δnin5
Plasmid construct	Description
pET15b/IN	pET15b encoding IN, AmpR

2.2. Conformation of plasmid constructs

Nested PCR was performed to amplify a 864bp fragment of HIV-1 integrase (IN) gene. External primers shown in table 2.2 were used for first round reaction and internal primers were used for second round to amplify 864 bp fragment (Table 2.2). The final volume of PCR reaction was 20 µl, which contained master mix (10X PCR buffer, 1.5 mM of MgCl, 0.1 mM dNTPs, 0.25 U/µl Taq polymerase, PCR grade water or nuclease-free water, 0.1 mM primers) and DNA template. The thermal cycling conditions (Table 2.3) were set on the thermocycler for both first and second round, additionally the reaction involve 35 cycles for both rounds. The PCR products (5 µl) were confirmed on a 1 % agarose stained gel with 100bp marker used for size confirmation. The ethidium stained gel was visualized under Spectroline® UV transilluminator (Lasec, made in USA) at 300 amperes, 80 volts for 45 minutes. Plasmid and gene construct was also sent to Inqaba for sanger sequening (Appendix A. 3) and blast of the obtained DNA sequence was conducted on NCBI.

Table 2. 2. Primer sequences and thermal cycling conditions used to amplify 864bp fragment of HIV-1 integrase gene

Targeted region	**Primer specification**	**Primer sequences**	
Integrase gene	TTS TTZ	1st round	5' GTGAATCAGAGTTAGTCAACC 3' 5' GACTCCCTGACCCAAATGCC 3'
	MTTS MTTZ	2nd round	5'CTCGAGCCATGGCATTTCTAGATGGAATAGATAAGGC 3' 5'GGATCCTAGCTAATCTTCATCCTGTCTACC 3'

Table 2. 3. Thermal cycling conditions

Conditions	**Temperature**	**Time**
Initial	95 ºC	5 minutes
Denaturation	95 ºC	1 minute
Annealing	62 ºC	1 minute
Extension	72 ºC	2 minutes
Final extension	72 ºC	10 minutes

2.3. Preparation of *E. coli* BL21(DE3) competent cells

E .coli BL21 (DE3) cells were prepared following growth at 37 ºC on 2 x YT broth containing1.6 % w/v Tryptone, 1 % w/v yeast extract and 0.5 % w/v NaCl. Before use, the broth was autoclaved at 121 ºC for 15 minutes for sterilization and complete solubilization of components. A single colony of the bacteria was picked and inoculated into 5 ml 2 x

YT broth, where it grew overnight at 37 ºC. Overnight culture was transferred to fresh 2 x YT broth and grown to an OD_{600}. = 0.5, followed by centrifugation of the culture at 6000 rpm for 10 minutes at 4ºC. The supernatant was discarded and the pellet was treated with 0.1 M $MgCI_2$, followed by incubation on ice for 30 minutes. The suspension was centrifuged for 10 minutes at 6000 rpm, 4ºC, and the supernatant was discarded. The obtained pellet was resuspended in 0.1 M $CaCI_2$, followed by incubation on ice for 4 hours. After incubation, the suspension was followed by centrifugation and the latter supernatant discarded. The resulting cells were stored in a solution containg 0.1 M $CaCI_2$ and 30 % glycerol. Several aliquotes (200 µl per tube) were stored at -80ºC before use.

2.4 Transformation of plasmid into competent cells.

A volume of 2 µl of integrase/pET15b was added into 100 µl of BL21 (DE3) cells. *E. coli* BL21 (DE3) cells was incubated on ice for 30 minutes, followed by heat shock at 42 ºC for 45 seconds and placed immediately on ice for 10 minutes. Thereafter, a volume of 900 µl of 2 x YT broth was added, followed by incubation at 37 ºC for an hour with gentle stirring. The transformed BL21 (DE3) cells were transferred into LB agar plates containing ampicillin, followed by overnight incubation at 37 ºC.

2.5 Agarose gel electrophoresis for DNA

For agarose gel preparation, 1 g of agarose was dissolved in 1x TAE buffer (40 mM, 20 mM acetic acid and 1 mM EDTA) by heating with constant agitation. Agarose gel was cooled, prior to the addition of ethidium bromide (0.5 µg/ml), the gel was allowed to polymerize at room temperature for 15-30 minutes. The gel was then placed on electrophoresis chamber and covered with 1x TAE buffer. A volume of 4 µl of 10x DNA loading buffer (0.25% bromophenol blue + 30% glycerol) was added to 20 µl of the plasmid DNA sample followed by loading the samples into the wells. Electrophoresis was conducted at 100 volts for one hour. The gel was then visualized using UV light (GeneGenius Bioimaging System (Syngene), USA).

2.6. Recombinant expression of pET15b-IN constructs

Chemically prepared *E. coli* BL21 (DE3) competent cells (Appendix A.1) were transformed with pET15b-integrase gene construct. A colony was inoculated into 25 mL of 2 x YT media (1.6 % tryptone, 1 % yeast extract and 0.5 % NaCl) supplemented with 100 µg/mL ampicillin and incubated overnight with continuously shaking at 37 °C. The overnight culture was diluted 10 times into a freshly prepared 2 x YT media supplemented with 100 µg/mL ampicillin and incubated to mid exponential growth at OD600= 0.4-0.6. The expression of the constructs was induced with 1 mM Isopropyl β-d-1-thiogalactopyranoside (IPTG) and samples were collected every hour for 3 hours. The cells were harvested at 3 hours by spinning at 6000 x g for 30 min at 4 °C. The pellets were resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 1 mM PMSF and 1 mg/ml lysozyme)and frozen at -80 °C. Samples were prepared and analyzed using Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis (SDS-PAGE).

2.7. SDS PAGE analysis for integrase protein

Integrase protein was treated by boiling in SDS sample buffer (0.25% Bromophenol blue (R250), 2% SDS, 10 % glycerol (v/v), 100 mM Tris, and 1 % β-mercaptoethanol) at 95 ºC for 5 minutes and resolved using 12 % acrylamide resolving gel prepared in table A.1. The gel was transferred into electrophoresis tank, followed by addition of electrophoresis buffer (25 mM Tris, pH 8.3 250 mM glycine and 0.1% (w/v) SDS). Prestained protein ladder (ThermoFisher Scientific, USA) was loaded on the wells, followed by loading SDS boiled samples onto the wells. Electrophoresis was performed at 150 volts, 120 current for an hour, using Bio-Rad Mini protein electrophoresis system (Biorad, U.S.A).

2.8 Coomassie Staining

The gel was removed from the electrophoresis chamber and placed in a container with enough 0.5 % Coomassie Blue G-250 (prepared in 50% methanol and 10% acetic acid), the gel was stained for 30 minutes. The stain was discarded, followed by rinsing the gel with MilliQ water. The gel was destain with 40 % HPLC grade methanol and 10 % acetic acid, replacing the solution every 15-20 minutes until band appear.

2.9 Solubility study of recombinant pET15b-IN constructs

To determine the solubility of IN, cell pellets from -80 °C were allowed to thaw on ice. The cells were resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole). Phenylmethylsulfonyl fluoride (PMSF) was added to the final concentration of 1 mM and lysozyme added to the final concentration of 1 mg/ml, and stirred on ice for 30 minutes. A sample containing a whole cell lysate was collected and the remaining cell lysate was sonicated at amplitude setting of 30 of 5 cycles with 15 seconds pulse and 30 seconds pause, after each cycle. A sonicated sample was collected and the remaing cell lysate was centrifuged at 10000 xg for 30 minutes at 4 ºC. After centrifugation, the supernatant and pellet sample were collected. The whole cell lysate, sonicated, supernatant and pellet samples were prepared and analyzed using SDS-PAGE.

2.10 Purification of recombinant pET15b-IN constructs

A colony of pET15b, BL21 (DE3) cells was inoculated into 50 ml of 2 x YT broth supplemented with 50 µL of ampicillin and incubated overnight with shaking at 37 ºC. The overnight culture was transferred into 450 ml of fresh 2 x YT broth supplemented with 450 µL of ampicillin and incubated to mid experimental growth. An amount of 1 mM IPTG was used once the OD_{600} reaches 0,4 and culture was incubated for 5 hours. After 5 hours the culture was collected and harvested at 6000 xg for 15 minutes at 4 ºC, and the pellets were stored at -80 ºC.

The recombinant integrase protein was purified under non-denaturing and denaturing conditions. The pellet which were stored at -80 ºC was thawed on ice and resuspended into 5 ml lysis buffer (50 mM NaH_2PO_4, 30 mM NaCl and 10 mM imidazole). PMSF was added to the final concentration of 1 mM and lysozyme added to the final concentration of 1 mg/ml and stirred for 30 minutes on ice. The cell lysates were sonicated at amplitude setting of 30 of 5 cycles with 15 seconds pulse and 30 seconds pause, after each cycle. The cell lysates were centrifuged at 12000 g for 30 minutes at 4 ºC. After centrifugation, the supernatant was added to HisPur™ Nickel-charged nitriloacetic acid (Ni-NTA), immobilized metal affinity chromatography column (IMAC) to allow integrase in the soluble fraction to bind with 2 ml nickel beads at 4 ºC for 1 hour. The HisPur™ Ni-NTA IMAC column was washed with 20 ml of wash buffers of varying imidazole concentration (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM and 50 mM imidazole, pH 8.0). The bound protein was eluted using 10 ml of elution buffer of varying imidazole concentration (50 mM NaH_2PO_4, 300 mM NaCl and 250 mM and 500 mM imidazole).

2.11 Determination of protein concentration using Bradford assay

Integrase concentration was determined by Bradford’s method. Bovine serum albumin (BSA) standards were prepared in 0.15 M NaCl at concentrations ranging from 0 to 1 mg/ml. Bradfords reagent 200 µl (Sigma Aldrich, USA) was added to 10 µl of integrase and the reaction was incubated in the dark for five minutes at room temperature. Using a SpectraMax M3, absorbance was measured at 595 nm (Molecular devices, USA). The recombinant protein was similarly treated, and the concentration was calculated by extrapolation from the standard curve, shown in Appenix A. 2. The readings were taken in triplicate and the average was calculated.

2.12 Protein identification by peptide mass fingerprinting, theoretical and manual procedures

2.12.1 Theoretical study, retrieval of HIV-1 integrase genome

The whole genome sequence of novel integrase was obtained through (https://www.ncbi.nlm.nih.gov/) with accession no: HQ207727.1, 864 bp in length and described as HIV-1 isolate 5799 from Canada integrase (pol) gene.

2.12.2 Translation of HIV-1 integrase sequence

The obtained sequence of HIV-1 integrase was translated into protein for the identification. The translation was done by EMBL EMBOSS TRANSEQ (http://www.ebi.ac.uk). It translates the nucleotide sequence into an amino acid sequence. It gives three forward and three reverse frames.

2.12.3 In-silico protein digestion and mass calculation

Integrase sequence was digested by protein prospector and peptide mass obtained was calculated from an Online Bioinformatics tool (http://web.expasy.org/peptide_mass/). Wilkins et al (1997) reported that the peptide mass tool is designed to aid peptide-mapping experiments, culminating in the analysis of peptide-mass fingerprinting (PMF) and mass spectrometry data. Trypsin enzyme was used to cut the integrase sequence in order to estimate the number of peptide mass and the protein mass to charge ration (Gundry *et al.,* 2010). Peptides's peak list was arranged by *m/z* ratio of integrase protein from the protein mass calculations.

2.12.4 Protein identification by LC-ESI-Q-TOF-MS

For manual integrase digestion, 40 µL of integrase protein (20 µg) was mixed with 10 µL of 1X trypsin buffer (50 mM Tris-HCl, 20 mM $CaCl_2$, pH 8) and 7 µL Milli-Q water, followed by addition of 2 µL trypsin enzyme (20 µg). The trypsin reaction was set with a ratio of substrate: protein, which was 1:40. The protein and trypsin sample was incubated at 37 ºC for overnight (18-24h). After 18 hours, 200 µL of acetonitrile was added to the sample and stored at -20 ºC prior to MS use. An amount of 200 µL of the digested sample was added to the vials, then placed to the LC tray.

The trypsin reaction products (10 µL) were injected on an LC-QTOF-MS, model LC-MS 9030 instrument. For separation, a Shim Pack Velox C18 column (100 mm × 2.1 mm with particle size of 2.7 µm) (Shimadzu, Kyoto, Japan) thermostated at 40 °C was used. The mobile phase consistent of a binary solvent system consisting of solvent A: 0.1 % formic acid in water and solvent B: 0.1 % formic acid in acetonitrile (UHPLC grade, Romil SpS, Cambridge, UK) with a total flow rate of 0.4 mL/min. The chromatographic separation was achieved using a 30 min multiple gradient run consisting of the following steps: 5 % B for 3 min, followed by a short gradient to 40 % B over 12 min and the conditions (40 % B) were kept constant for 3 min, followed by another gradient to 90 % B in 3 min, another isocratic hold at 90 % for 3 min was also done and finally, the initial conditions (5 % B) were re-established in 3 min and a the column was re-equilibrated for a next run at 5 % B for 3 min.

For optimial MS detection, the following parameters were set as follows: ESI positive ionization mode; interface voltage of 4.0 kV; nebulizer gas flow at 3 L/min; heating gas flow at 10 L/min; heat block temperature at 400 °C; CDL temperature at 250 °C; detector voltage at 1.70 kV; TOF tube temperature at 42 °C. Sodium iodide (NaI) was used as calibrant to ensure high accurate mass (below 1 ppm). Data was acquired with a mass range of 100–2000 Da. For tandem MS (MS/MS) experiments, argon gas was used as collision gas for collision induced dissociation approach. A typical MS^E approach were collison energy spread (between 10 and 40 eV) and were used to generate framents for MRM method. For multiple reaction monitoring, selected precursor ions and their associated fragment/product ions were selected and appropriate collision energies were used. LabSolution software was used for data acquisition and post run analysis.

Chapter 3: Results and Discussion

3. 1 Confirmation of pET15b-IN plasmid integrity

The pET15b/IN gene construct was verified by nested PCR (Figure 3.1), which make use of internal primers that amplify the targeted amplicon with high precision. Negative controls (PCR reaction without DNA template) was also anlaysed and did not produce the desired band of interest (Figire 3.1, lanes 1 and 2). As seen below, the PCR reaction containing the pET15b/IN construct as a template produced a desired band of interest at 864 bp (Figure 3.1, lane 3).

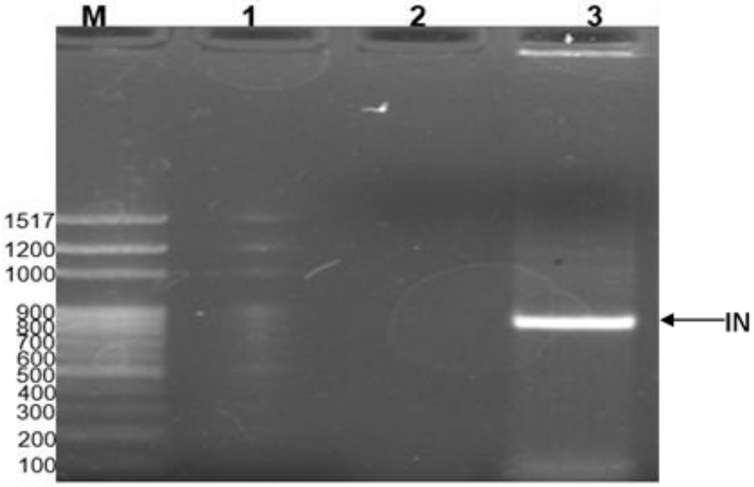


Figure 3. 1. A representative 1% agarose gel electrophoretogram showing nested PCR products amplified from pET15b/ integrase construct. A 1kb DNA size marker (New England Biolabs) is indicated as M. Lanes 1 and 2 are negative control that contain the master mix without the DNA template and lane 3 is the amplicon generated from pET15b/IN construct used as template. The arrow indicates the approximately 864 bp PCR amplicon (Integrase gene).

3. 2 Expression of the pET15b-IN

Recombinant IN was expressed in *E. coli* BL21 (DE3) cells at 37 °C. The expression was assessed using SDS-PAGE (Figure 3.2 A). The SDS-PAGE results show a distinct IN band before and after IPTG induction, which Is shown by the presence of a band at 32 kDa. The SDS-PAGE shows that IN is expressed prior to induction by IPTG, phenomenon described as leaky expression (Caumont *et al.,* 1996). The induction study shows a similar IN yield from the first hour to the fifth hour post-induction (Figure 3.2 A). To counteract the leaky expression, IN truncation and to also improve the expression yield, the growth temperature was reduced to 20 °C (Figure 3.2 B). Heterologous expression of IN at 20 °C resulted in truncation prior to and after induction (Figure 3.2 B). However, higher expression yield was observed at 3 and 4 hours post induction (Figure 3.2 B).

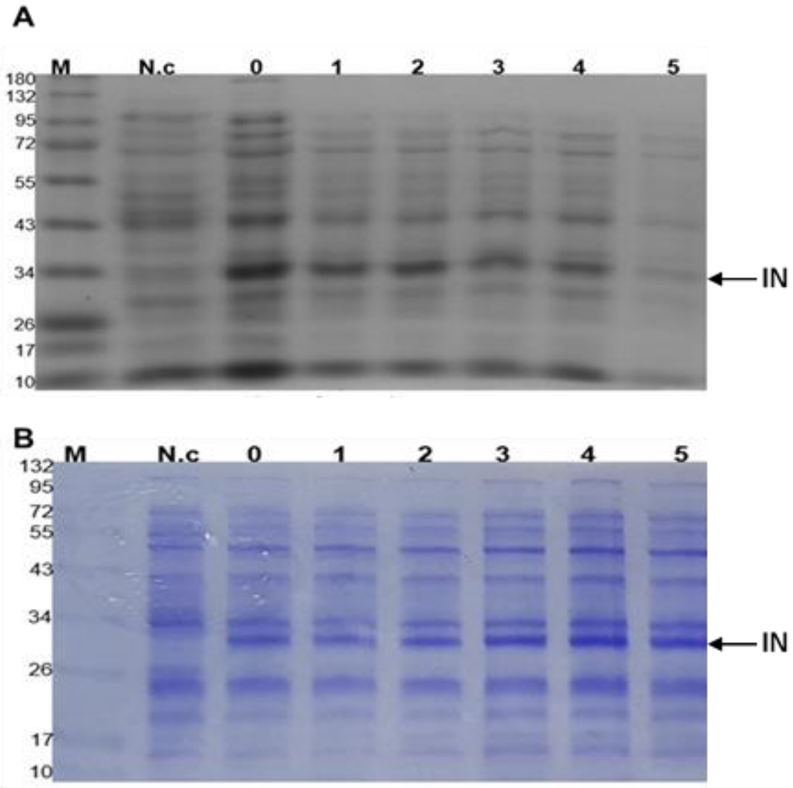


Figure 3. 2. Representative Sodium Dodecyl Sulfate Poly-Acrylamide Gel Electrophorectogram (SDS-PAGE) showing heterogeneous expression of integrase in *E. coli* BL21 (DE3) cells at 37 ºC (A) and 20 ºC (B). Samples are as follows; lane M is prestained protein marker, N.c-negative control which contain untransformed cells; 0-uninduced plasmid/ gene construct; 1-5 hours post induction samples.

3. 3 Solubility study of pET15b-IN

The solubility of recombinant pET15b-IN was assessed in the presence and absence of detergent. IN was not soluble in lysis buffer (in the absence of detergent), depicted by a 32 kDa band in the pellet fraction (Figure 3.3 A, Lane P), a band was observed on a sample after sonication, however a very faint band at 32 kDa was observed in the soluble fraction (Figure 3.3 A) which suggest IN may be expressed in inclusion bodies or possible hydrophobic interaction with the membrane making it insoluble (Jenkins *et al.,* 1996). The IN preparation was then treated with detergent (urea and Triton X-100) in order to solubilize the protein. The use of high concentration of urea (8 M) has been shown to completely denature proteins by disrupting the existing secondary structures (Singh *et al.,* 2015), the same has been reported when using 1 % of the nonionic detergent Triton X-100 (Noda *et al.,* 2017). Hereine, the detergent showed no significant difference in the solubilization of IN. The solubilization with 1 % Triton X-100 known to unfold proteins (Noda *et al.,* 2017) was partial (Figure 3.3 B), this is shown by the presence of a 32 kDa band in the insoluble fraction, however, there was still significant amount of protein observed in the soluble fraction (Figure 3.3 B). Highest concentration of 8 M urea known to unfold proteins (Singh *et al.,* 2015) also showed no significant improvement, this is shown by a pronounced 32 kDa band in the insoluble fraction and a faint band in the soluble fraction (Figure 3.3 B).

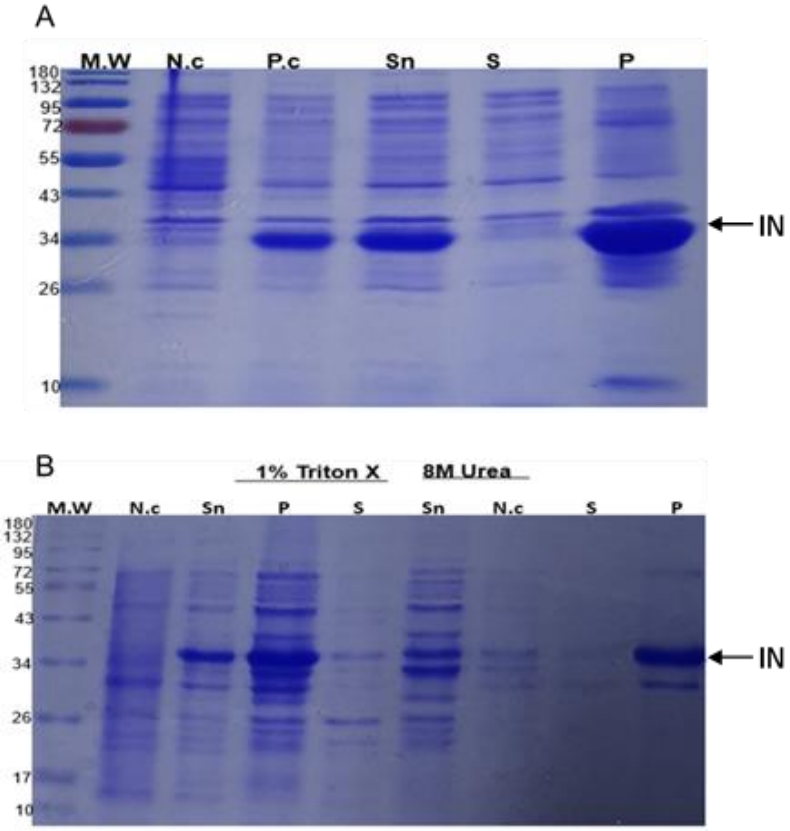


Figure 3. 3. Representative Sodium Dodecyl Sulfate Poly-Acrylamide Gel Electrophorectogram showing solubility study of IN. IN was solubilized in the absence (A) and presence of detergents, 1% Triton X and 8M urea (B). Lane M.W-prestained protein marker, N.c-negative control which is untransformed cell lysate, P.c- positive control which is whole cell lysate, Sn-sonication cells after sonication, S-supernatant which is soluble fraction, P-pellet which is insoluble fraction.

3. 4 Purification of recombinant IN

The recombinant IN was purified under non-detergent conditions using Nickel-affinity chromatography. After lowering the growth temperature from 37 ºC to 20 ºC, it was observed that IN was soluble, this was observed by detecting IN in the whole lysate as a 32 kDa species (Figure 3.4). This was also observed in the soluble fraction to show that IN was successfully solubilized (Figure 3.4). There was some protein observed in the flow through, which suggests that IN did not fully bind to the Nickel beads (Figure 3.4, lane

FT). Moreover, IN was absent from the wash sample (Figure 3.4, lane W1 and W2). It can be said that the purification of the desired protein was not entirely successful, owing to the faint band observed during final elution (Figure 3.4, lane E1 to E6). To obtain a pure protein with high concentration, several modifications to the purification protocol were made.

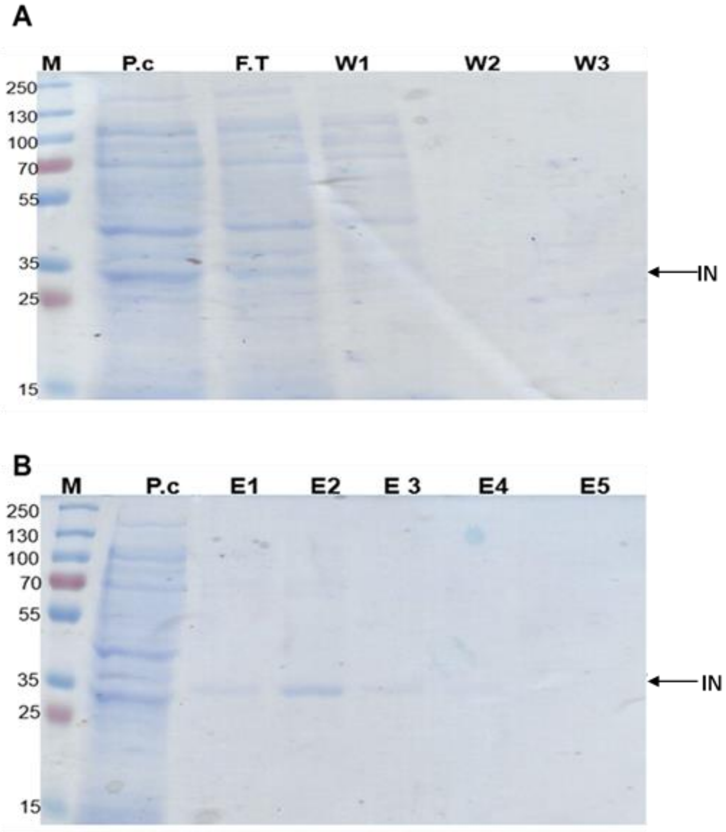


Figure 3. 4. Representative Sodium Dodecyl Sulfate Poly-Acrylamide Gel Electrophorectogram (SDS-PAGE) showing purification of recombinant IN, expressed at 20 ºC. Lane, M – Protein marker, P.c-positive control which contains a whole cell lysate, F.T-flow through, W1-W3 are wash samples, E1-E5 are elutions samples.

To maximize the purification of IN, experimental improvements were made which included the use of Triton X-100 (Figure 3.5). This was an attempt to increase the concentration of IN and to counteract the contamination which could have resulted to truncation. The protein protocol was further optimized by conducting the purification process under denaturing conditions using 1% triton X-100. As seen below, IN was observed in the supernatant as a 32 kDa band (Figure 3.5). Furthermore, a faint band was observed at flow through which may be due to IN not successfully binding to the nickel beads. Thus, no IN was further detected on the wash samples (Figure 3.5). Prior to purification, IN was eluted with varying imidazole concentrations (250 mM and 500 mM). A band at approximately 32 kDa was observed at elution 2 (Figure 3.5 B), however quantification resulted in lower concentration. As a result of low concentration observed, IN was purified using ion exchange chromatography.

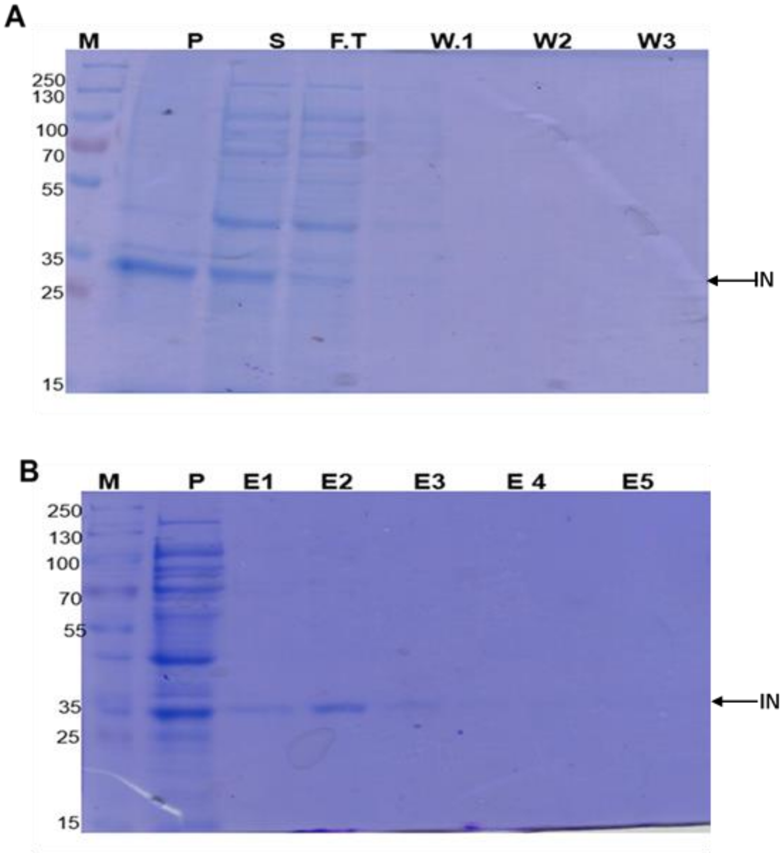


Figure 3. 5. Representative Sodium Dodecyl Sulfate Poly-Acrylamide Gel Electrophorectogram (SDS-PAGE) showing **p**urification optimization of recombinant pET15b-IN construct**.** Purification was optimized using 1% Triton X-100. Samples were loaded as follows; Lane, M– Protein marker, P-pellet, S-supernatant, F.T-flow through, W1-W3 are wash samples, E1-E3 are elutions samples with 250 mM imidazole, E4-E5 are elution samples with 500 mM imidazole.

In order to increase the concentration of IN, ion exchange chromatography (IEC) was attampted however with very little success (Figure 3.6). Furthermore, there was notable breakdown of IN detected in all samples. As a result of the breakdown, IN was purified under denaturing conditions (8 M urea; Figure 3.7).

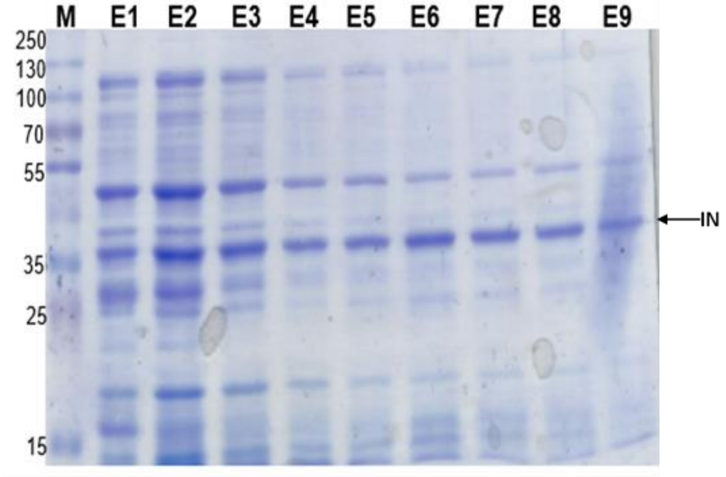


Figure 3. 6. Representative Sodium Dodecyl Sulfate Poly-Acrylamide Gel Electrophorectogram (SDS-PAGE) showing ion exchange chromatography purification optimization of recombinant pET15b-IN. Lane M-represents molecular weight markers (in kDa); Lanes E: represent the elution fraction.

Interestingly, using the lysate treated with 8M urea lysis buffer in order to solubilize the protein, IN was successfully purified (Figure 3.7). However, some of the protein did not fully bind to the beads as indicated by a 32 kDa band in the flow through (Figure 3.7). Additionally, some amount of protein was lost in the washes, which shows that increasing the imidazole concentration affected the binding of IN to the beads, though expected.

Selective displacement (Varying imidazole concentrations) were used for eluting integrase protein, and as seen on figure 3.7 B, the eluted proteins did not show any breakdown or other proteins except for IN, thus successfully purifying integrase with 8 M urea detergent (Figure 3.7 B).

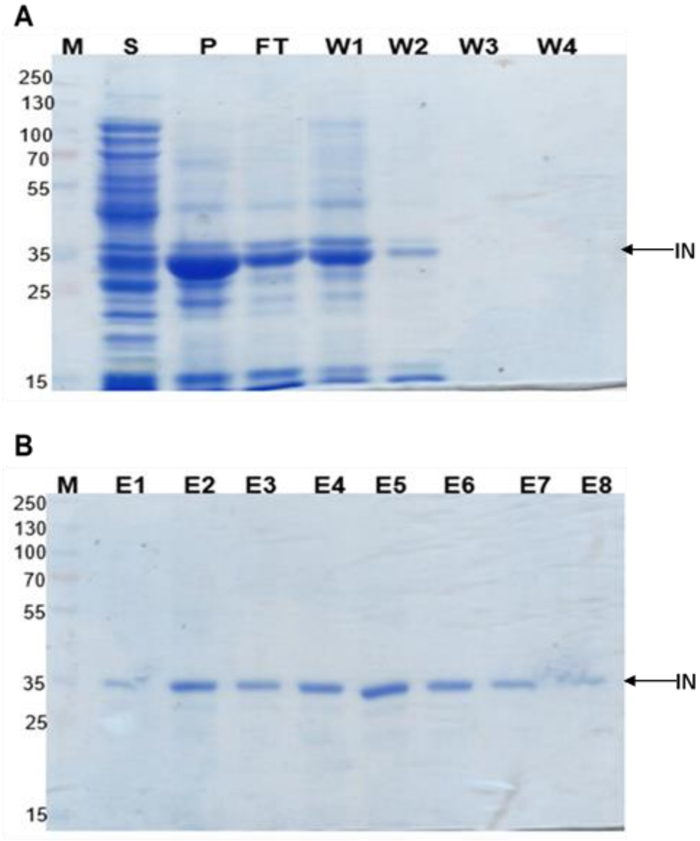
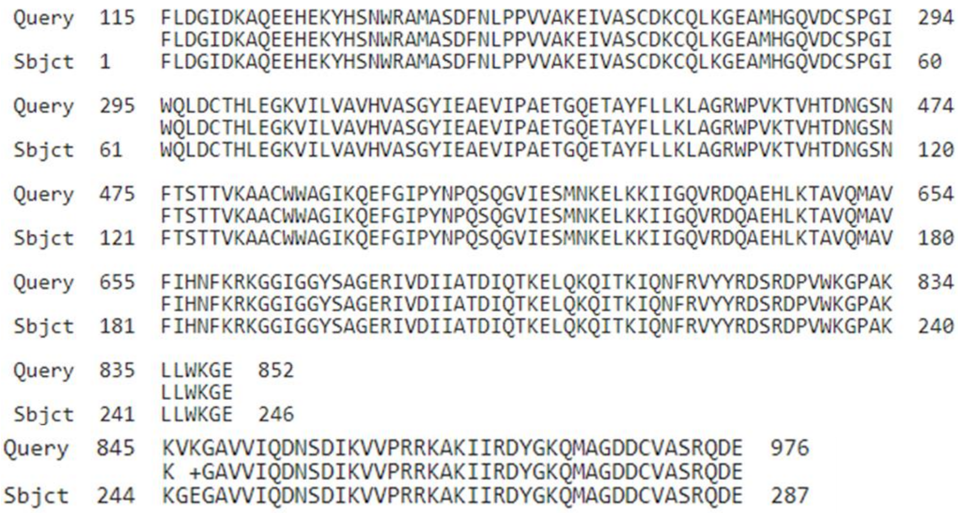


Figure 3. 7. Representative Sodium Dodecyl Sulfate Poly-Acrylamide Gel Electrophorectogram (SDS-PAGE) showing purification of IN using Ni-affinity chromatography. Purification was optimized using 8M urea. Samples were loaded as follows; Lanes, M – Protein marker.Lane S-supernatant; P-pelltet, FT-flow through; W1-W4 are wash samples, Lanes E1-E8 are elutions samples.

3. 5 Translation of HIV-1 integrase DNA sequence (sanger sequence) to protein sequence.

The obtained DNA sequence from sanger sequencing was translated into protein by NCBI translator. The protein sequence obtained was described as integrase, partial (human immunodeficiency virus), with 288 amino acid in length as indicated by NCBI (NCBI accession number: AAC37875.1). The sequence shows the amino acid composition with no modifications.



```
Query  115  FLDGIDKAQEEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLKGEAMHGQVDCSPGI  294
            FLDGIDKAQEEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLKGEAMHGQVDCSPGI
Sbjct  1    FLDGIDKAQEEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLKGEAMHGQVDCSPGI  60

Query  295  WQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKTVHTDNGSN  474
            WQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKTVHTDNGSN
Sbjct  61   WQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKTVHTDNGSN  120

Query  475  FTSTTVKAACWWAGIKQEFGIPYNPQSQGVIESMNKELKKIIGQVRDQAEHLKTAVQMAV  654
            FTSTTVKAACWWAGIKQEFGIPYNPQSQGVIESMNKELKKIIGQVRDQAEHLKTAVQMAV
Sbjct  121  FTSTTVKAACWWAGIKQEFGIPYNPQSQGVIESMNKELKKIIGQVRDQAEHLKTAVQMAV  180

Query  655  FIHNFKRKGGIGGYSAGERIVDIIATDIQTKELQKQITKIQNFRVYYRDSRDPVWKGPAK  834
            FIHNFKRKGGIGGYSAGERIVDIIATDIQTKELQKQITKIQNFRVYYRDSRDPVWKGPAK
Sbjct  181  FIHNFKRKGGIGGYSAGERIVDIIATDIQTKELQKQITKIQNFRVYYRDSRDPVWKGPAK  240

Query  835  LLWKGE  852
            LLWKGE
Sbjct  241  LLWKGE  246
Query  845  KVKGAVVIQDNSDIKVVPRRKAKIIRDYGKQMAGDDCVASRQDE  976
            K +GAVVIQDNSDIKVVPRRKAKIIRDYGKQMAGDDCVASRQDE
Sbjct  244  KGEGAVVIQDNSDIKVVPRRKAKIIRDYGKQMAGDDCVASRQDE  287
```

Figure 3. 8. Sequence alignment of translated HIV-1 integrase protein sequence, with a total of 288 amino acids. Query represent the protein of interest and Subject represent a protein within a database.

3. 6 Protein identification by *In silico* IN digestion using trypsin

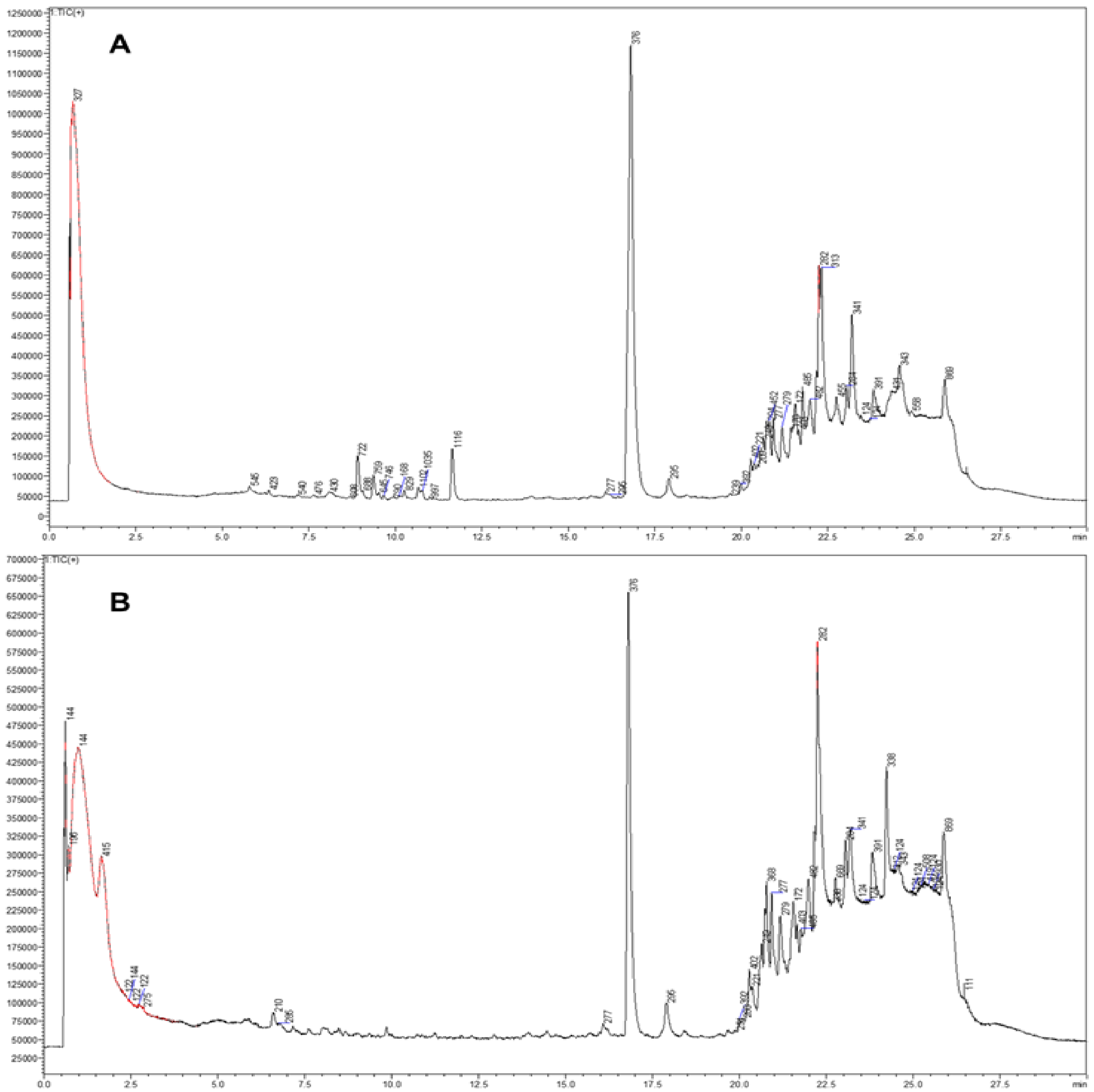
Table 3.1. Theoretical trypsin digest of the targeted protein. The protein was found to have the following properties :Theoretical pI: 9.03, Mw (average mass):31919.43 / Mw (monoisotopic mass): 31899.19].

MASS	POSITION	#MC	PEPTIDE SEQUENCE
2266.0808	168-187	0	QEFGIPYNPQSQGVIESMNK
2164.0002	9-27	0	YTMGSSHHHHHHSSGLVPR
1708.8136	143-158	0	TVHTDNGSNFTSTTVK
1505.7933	205-217	0	TAVQMAVFIHNFK
1459.7613	52-65	0	AMASDFNLPPVVAK
1329.7624	231-242	0	IVDIIATDIQTK
1219.5775	28-38	0	GSHMFLDGIDK
1023.4854	220-230	0	GGIGGYSAGER

1005.4974	159-167	0	AACWWAGIK
870.3952	39-45	0	AQEEHEK
864.4131	66-73	0	EIVASCDK
862.3954	46-51	0	YHSNWR
840.4210	198-204	0	DQAEHLK
752.4123	1-6	0	FCLTLR
685.4355	192-197	0	IIGQVR
677.3729	251-255	0	IQNFR
644.3402	263-267	0	DPVWK
621.2951	279-284	0	GSSNTR
600.3140	256-259	0	VYYR
559.3602	272-275	0	LLWK
529.3133	139-142	0	WPVK
517.2980	243-246	0	ELQK
491.2646	74-77	0	CQLK
489.3031	247-250	0	QITK
416.2616	135-138	0	LAGR
389.2394	188-190	0	ELK
377.1779	260-262	0	DSR
372.2241	268-271	0	GPAK
361.1830	276-278	0	GER
175.1189	7-7	0	R
175.1189	8-8	0	R
175.1189	218-218	0	R
147.1128	191-191	0	K
147.1128	219-219	0	K

3. 7 Protein analysis by LC-ESI-Q-TOF-MS though peptide mass fingerprinting approach

Liquid chromatography hyphernated to an Electron spray ionization quadrupole time of flight mass spectrometry was used to identify IN through a process called peptide mass fingerprinting. Two control were ran, one which did not contain trypsin enzyme (Figure 3.9 A) and the other which did not contain a template protein (Figure 3.9 B). The type of mass spectrometry used herein is unable to read large protein and, as such, an endopeptidase enzyme (trypsin to be specific) was used to reduce the size of the analyte. The results show that the negative controls (Figure 3.9 A and B) resulted in few trypsin fragment peaks. As seen below, the trypsin digest of IN protein resulted in multiple fragments as compared to the two controls (Figure 3. 9). The acquisition mass range was set between 100 – 2000 Da and higher masses resulted in an abrupt lost of sensitivity. In order to select possible ion to carry more experiments on, the obtained product ions were compared to theoretical ions (Table 3. 1).


A
1:TIC(+)
1250000
1200000
1150000
1100000
1050000
1000000
950000
900000
850000
800000
750000
700000
650000
600000
550000
500000
450000
400000
350000
300000
250000
200000
150000
100000
50000
0
327
376
722
1116
295
282
313
341
343
869
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0
27.5
min
B
1:TIC(+)
700000
675000
650000
625000
600000
575000
550000
525000
500000
475000
450000
425000
400000
375000
350000
325000
300000
275000
250000
225000
200000
175000
150000
125000
100000
75000
50000
25000
144
144
415
210
277
376
295
282
338
341
391
124
368
277
172
869
111
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0
27.5
min

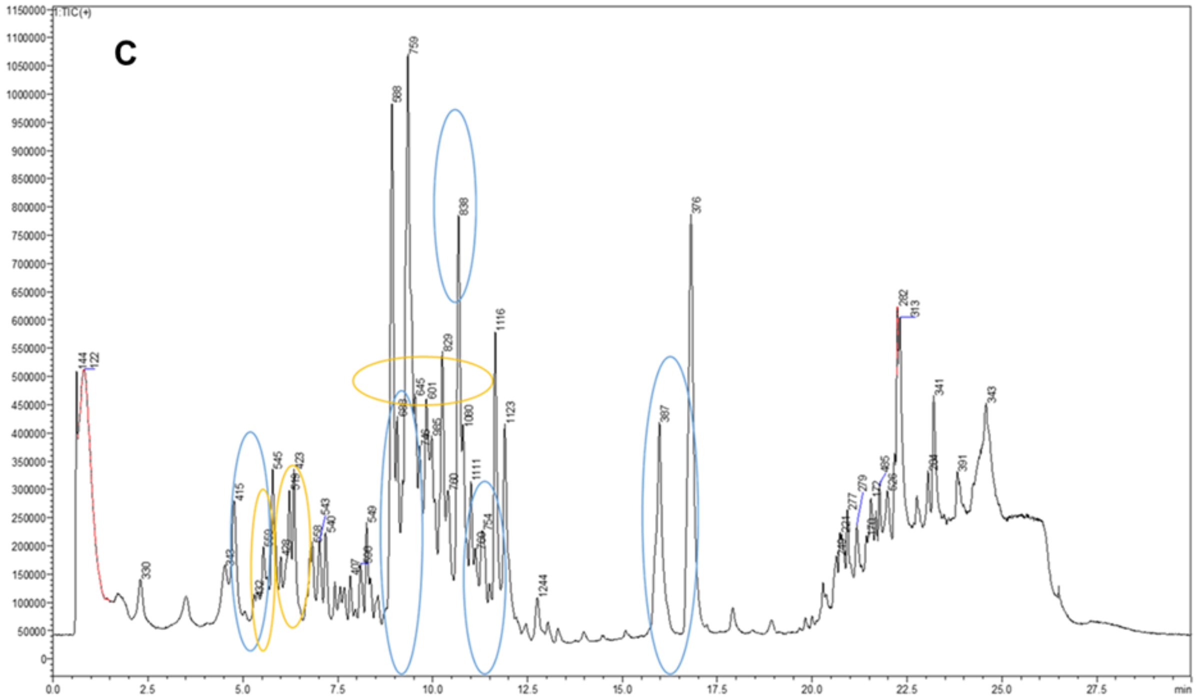


Figure 3. 9. Representative LC-MS, total ion chromatogram (TIC) showing separation of fragments in non-trypsin control (A), non-protein control (B) and trypsin-digested protein preparation (C). Some peaks obtained shows identical peptide mass with theoretical mass range (shown in orange), some differs with theoretical mass due to charge gained (shown in blue).

3. 8 MS/MS (MS²) validation for further fragmentation of precursor ion 600.84

The 600.84 precursor ion was selected for MS² fragmentation, because it was one of the ions seen in both the theoretical ions and the ions obtained through experimental MS results. This is assumed to be peptide VYYR with peptide mass of 600.84 (Table 3.1) and it was further subjected to MS² to generate products ions, unique to its composition. The 600.84 precursor ion was fragmented with the following parameters: *m/z* range 100 to 1000 Da, collision energy 25 eV, 35 eV, and 45 eV, with 15 eV spread, interface voltage of 4 kV. Product ions were efficiently generated with varying collision energy (CE) 10 to 40 eV (Figure 3.11 A), 20 to 50 eV (Figure 3.11 B) and 30 to 60 eV (Figure 3.11 C).

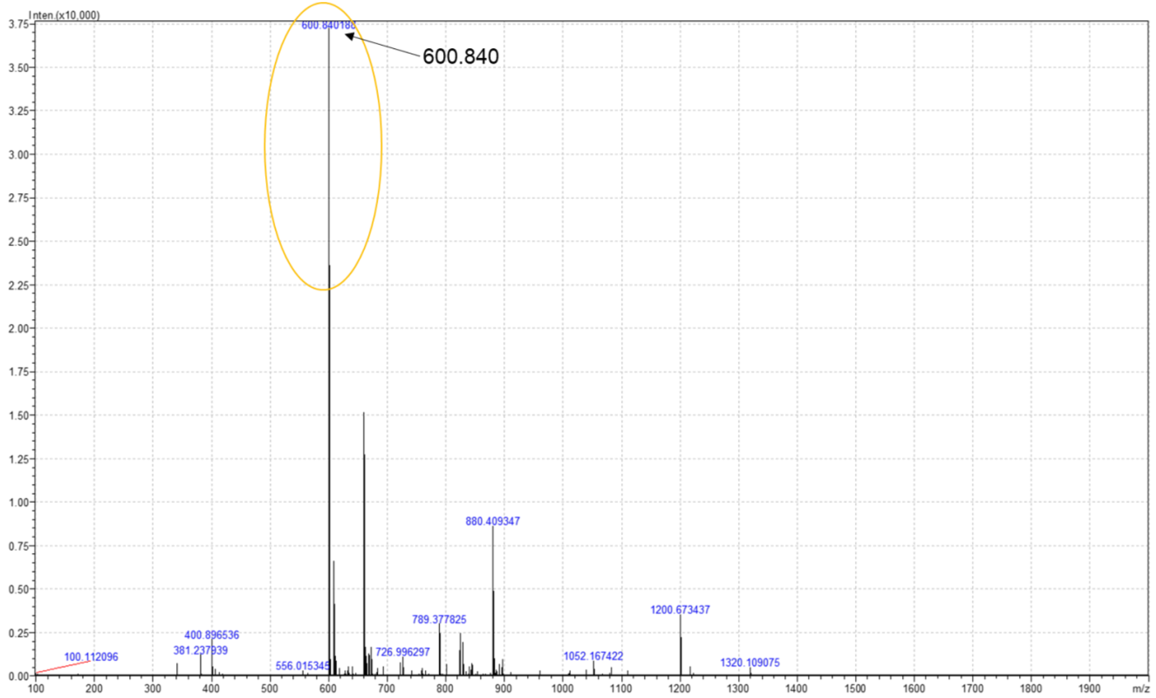


Figure 3. 10. MS² spectrum showing detection of one of the expected fragments of integrase enzyme at *m/z* **600.84.**.

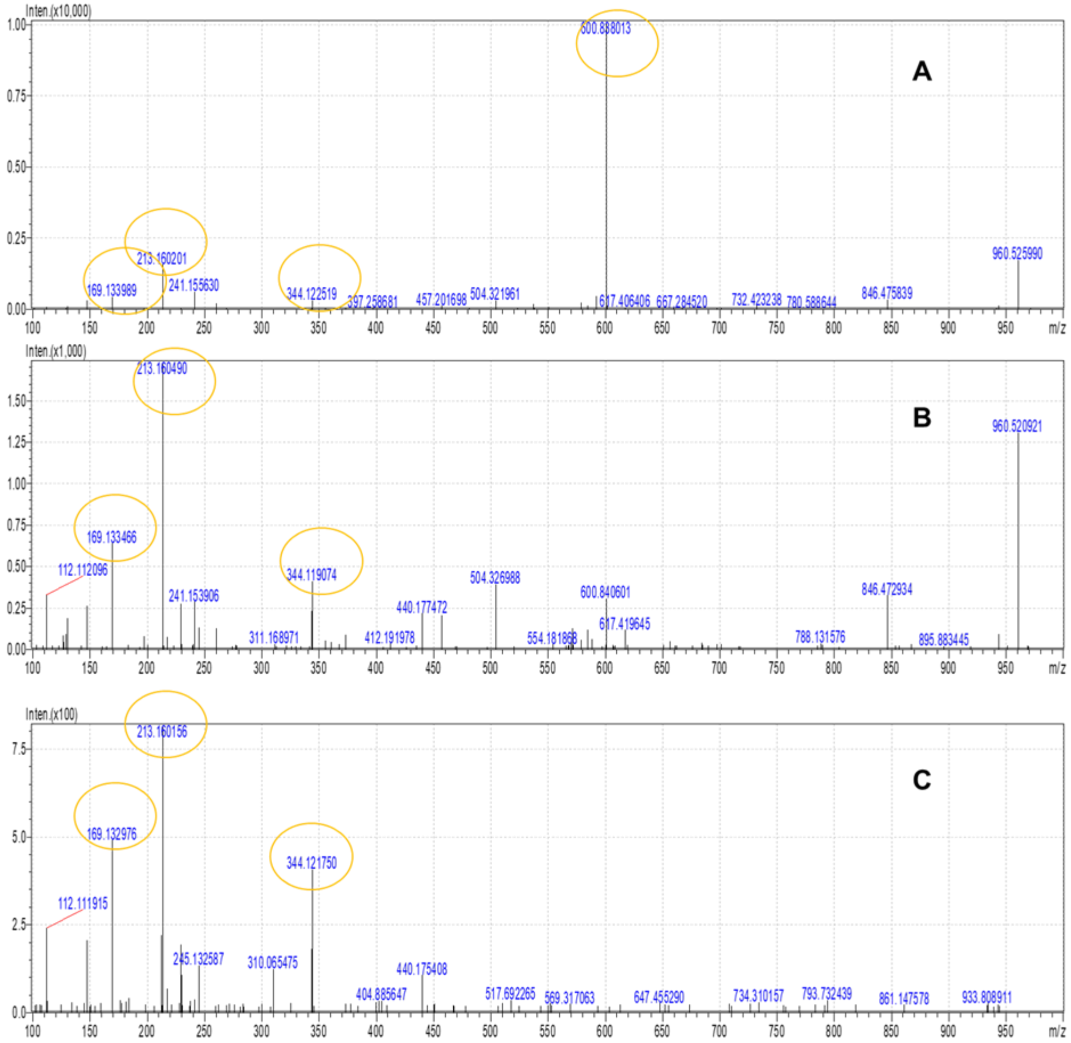


Figure 3. 11 . MS/MS spectrum transition of precursor ion 600.84 with a collision energy ramp of 10 to 40 eV (A), 20 to 50 eV (B) and 30 to 60 eV (C). The transition shows the product ion (600.840) and three product ion highlighted in orange 169.13; 213.16 and 344.11.

3. 9 Multiple reaction monitoring transitions for precursor ion 600.84

The selected 600.84 precursor ion was further subjected to MRM tandem mass spectrometry (Figure 3.12). This method was developed to produce unique signals which are associated with the product ions which were seen in MS^2 anlaysis (Figure. 3.11). MRM method are kown to be very specific and their use can be done in combination with other forms of tandem mass spectrometry to increase sensitivity (Masike & Madala, 2018). Herein, the selected precursor ion and its product ions were used to develop a very specific MRM method against a product which was deemed to be unique for integrase enzyme and signals associated with this ion were positively detected as seen below (Figure. 3.12). It is always advised that multiple ions are targeted in order to generate a more robust and reliable method of detection. For instance, in the current study, other ions which were generated as product of trypsin digest of IN were also selected for downstream MS^2 and their unfragmented mass spectra are show below (Figure 3.13). These ions showed typical MS^1 profile of proteins as shown elsewhere (example in Figure 1.6). As expected these peptides produced multiple charged species and this was found to pose an undisputed analytical challenge because selection of an ion for downstream fragmentation pattern was found to be challenging and, as such, no MRM method for these ions could be generated. Possible remedy to overcome this challenge is further discussed in section 3.10.

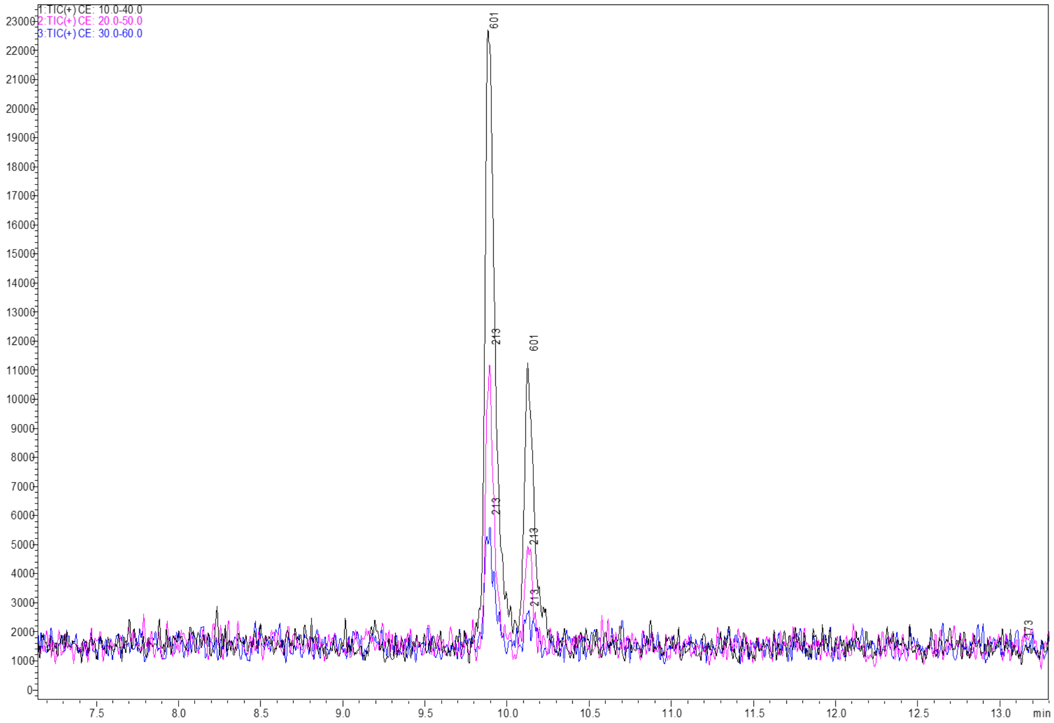
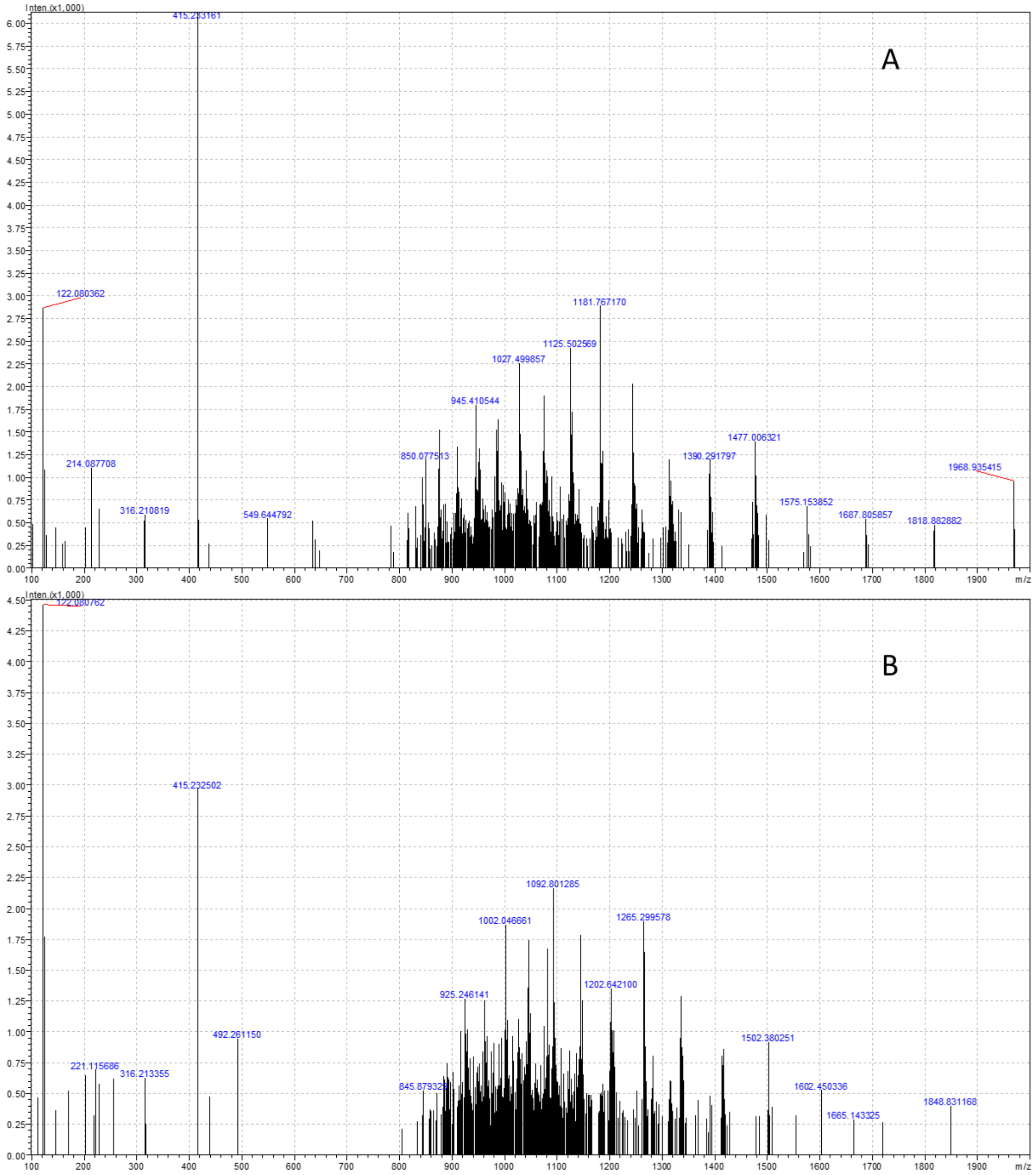


Figure 3. 12. MRM spectrum detection of a fragment ion at m/z 600.84, showing three product ion varying collision energy.



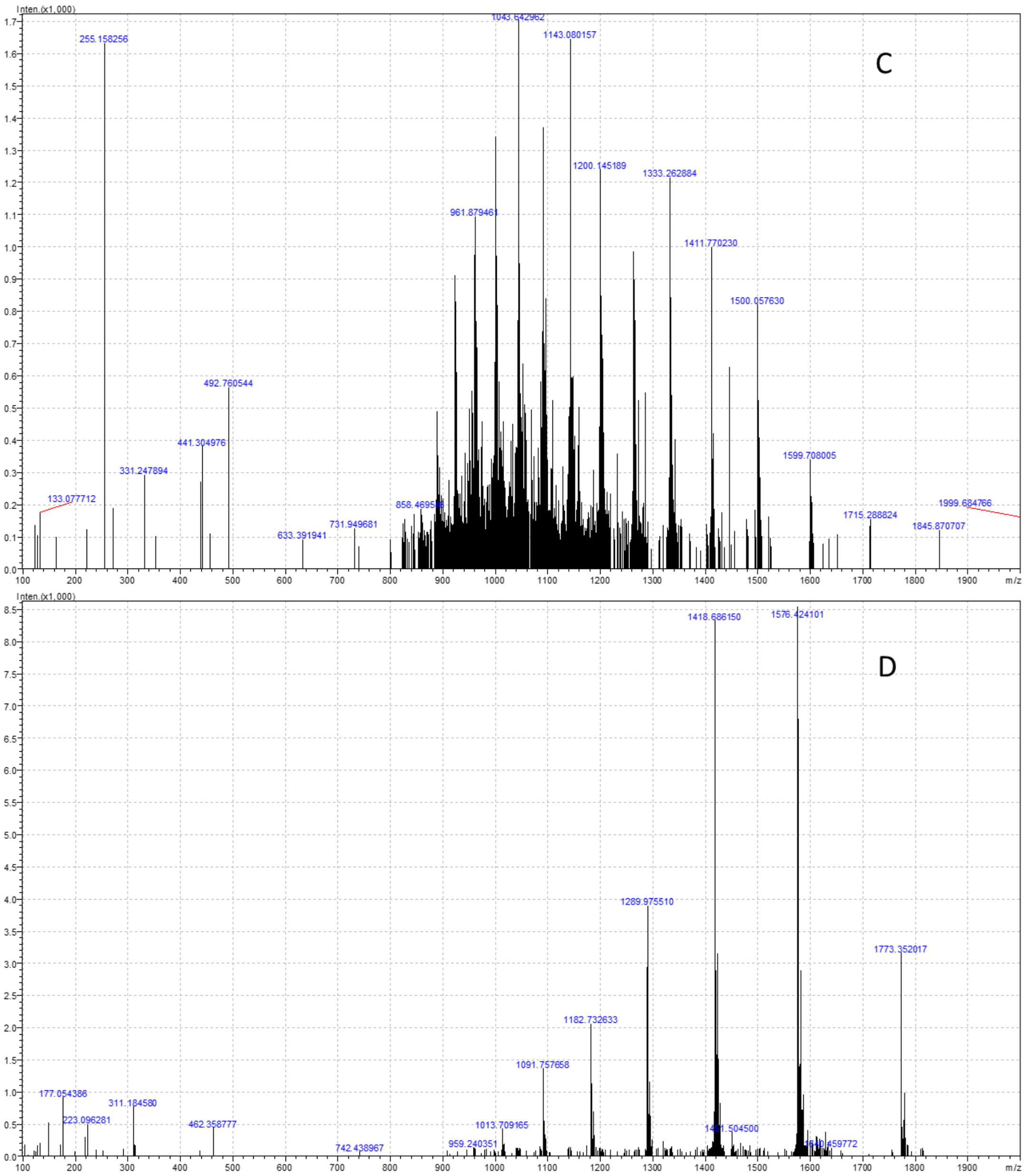


Figure 3. 13 B. Representative MS spectra of fragment ions associated with IN protein showing multiple charged species due to ESI ionization mode use herein.

3.10 Discussion

The aim of the study was to develop MS method to detect HIV associated protein (integrase enzyme) through the use of mass spectrometry. The pET15b/integrase gene construct was confirmed through nested PCR and sanger sequencing. Nested PCR for the amplification of an approximately 864bp fragment encompassing the HIV-1 integrase was successfully established. Agarose gel electrophoresis was used to visualize the resulting DNA fragments. Sanger sequening was also done to confirm the pET15b/IN gene construct (Appendix B.3), the identity of the sequence obtained was searched on NCBI and it was confirmed that it is HIV-1 integrase from Canada. The blasted DNA sequence further prove that an amplicon is identical to the target

Integrase was successfully expressed and purified through series of chromatographic optimization. Recombinant IN protein was shown to resolve at 32 kDa, the recombinant IN that correlate to the size (32 kDa) was creported by (Caumont *et al.,* 1996; Lataillade & Kozal, 2006). Various strategies were used to improve expression and solubility of IN. This includes lowering the growth temperature during the expression stage and use of various detergents to improve the solubility of the recombinant protein. The expression of IN was improved when transformed cells were grown at 20 ºC, this is because lowering the temperature slow down the physiological and metabolic process of the cells such as transcription, translation and refolding state which allows proper folding of the protein. Although there was an improvement on the expression levels, the solubility of IN remained challenging. Furthermore, the observed partial solubility might indicate that IN has some hydrophobic properties suggesting a possible functional activity along the cell membrane (Figure 3.7; (Rawlings, 2016). Oh and Shin, (1996) further stated that integrase is less soluble because it binds to DNA of the host very strongly.

Integrase was expressed in its insoluble form, this suggest the possible formation of inclusion bodies (Tehrani *et al.,* 2015; Zhao *et al.,* 2009). The use of 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) detergent to solubilize IN was found to be effective (Oh and Shin, 1996), since it is insoluble in its native form.

It was therefore assumed that the problems associated with solubility of IN was due to IN not being mutated which makes it difficult to solubilize even in the presence of detergents. A mutuation by systematic replacement of hydrophobic residues on a single amino acid (F185K) of the integrase was found to enhance the solubility (Jenkins *et al.,*1996). However, this was not an option in the current study, since the aim was to develop a method to detect this protein at its native state, hence changes will definitely affect this method of detection.

The leaky expression of IN in *E. coli* BL21 (DE3) cells was also observed. Leaky expression is a phenomenon where expression of a protein happens without induction of its promoter (Briand *et al.,* 2016). Thus, in the current study, expression happened prior induction by the IPTG, this phenomenon may indicate weakly regulated promoter (Figure 3.4).

Though various approaches failed to purify integrase, the use of 8 M urea successfully purified IN using Nickel-affinity chromatography (Figure 3.7). Urea in the lysis buffer makes their DNA binding loosen and subsequent sonication releases the integrase from the DNA. The protein was refolded by again lowering the urea concentration (to at least 2 M). It was found that the activity of integrase is completely restored by lowering urea concentrations (Balakrishnan *et al.,* 1996).

This study further employed LC-QTOF-MS for identification of IN protein. It was vital to ascertain the effectiveness of trypsin. The sequence of IN was translated into amino acids to theoritically generate possible fragemnts which were expected after trypsin digest. The *in silico*-based peptide digestions released series of possible framents as shown in Table 3.1. The obtained peptides were used as a control with the MS obtained peptides.

Trypsin was used as it exhibits superior cleavage specificity. It specifically hydrolyses peptide bonds only when carbonyl group is followed by arginine or lysine amino acid residue. However, cleavage will not occur if proline is on the carboxyl side or C-terminal of lysine or arginine. This is of extreme importance since in the process of searching peptide fragmentation spectra against sequence databases, potentially matching peptide sequences should confirm trypsin specificity (Ma *et al.,* 2014; Fleurbaaij *et al.,* 2015).

LC-Q-TOF-MS analyses of the trypsin digest reaction was used to produce a fingerprint of IN that can be widely used for its identification. For tandem MS (MS/MS experiments), Argon gas was used as the collision gas to fragment IN. Figure 3.9 A and B illustrates a representative negative control of the digestion reaction and all these two control showed very little fragments ions when compared to the experimental trypsin digestion reaction containing all the reaction components. No false positive was observed in the negative controls which emphasizes the reliability of the LC-MS technique (Figure 3.9 A and B).

Figure 3.9 C illustrates the experimental sample of interest, which is the digested IN. The success of this digestion was established by comparing the outcome of the reaction with those of theoretical digest, and ions which were common were selected for further experiments. Four peptides which had the same *m/z* as theoretical enzymatic were identified according to accurate mass and amino acid sequence using LC-MS/MS analysis (highlighted in orange; Figure 3.9 C). Theoretically, a single unique tryptic peptide would be sufficient for IN identification. Nevertheless, detecting several peptides ensures unambiguous identification (Duracova *et al.,* 2018). Product ions at *m/z* 559.27, 600.94, 644.132 and 1505.68 were the few which shared same masses with some product generated theoretically and their identity were also shown previously (Pala *et al.,* 2020). Therefore, these are fragmentations ions that can be used as a signature of IN (Figure 3.9 C). As expected , some of the fragmentation ions were found to differ with theoretical mass and those were not selected for further experiments. Other technical experimental settings such as collision energy are known to affect the results of peptide mass fingerprinting. Ho *et al* (2003) reported that collision energy can generate additional ion types due to side chain cleavage. Pitt (2009) also reported that ESI can generate multiple charge ion by addition of a proton to the analyte ($M+H^{+}$) when the ion source is

operated in positive mode. This may results in an envelope of ions with different charge states, thus producing multiple-charged ions, effectively extending the mass range. These findings correlate with our recent findings (Figure 3.9 C) of fragmented ions that differs with the theoretical peptide masses.

The selected ions were further analysed through MS² analyses. In this approach, fragmentation of peptide ion at *m/z* 600.84 (precursor ion) was further subjected to collison induced dissociation for further fragmentation. The transition of ions from the precursor to product ion is highly specific and as such can be exploited for identification of IN as demonstrated herein. The use of mass spectrometry for identification of proteins was shown to provide high degree of selectivity (Ma *et al.,* 2014; Feldberg *et al.,* 2020). Through MS/MS, the precursor ion at *m/z* 600.54 further fragmented to produce product at *m/z* 213.16, 169.13 and 344.12 product ions

According to European criteria, two MRM transitions are sufficient for protein identification (Feldberg *et al.,* 2020). Therefore, the ion at *m/z* 600.84 was further analysed through MRM approach since it was the only peptide ion detected by MS² and the only peptide ion with same mass as theoretical mass. Figure 3.12 demonstrates MRM chromatogram for an ion at *m/z* 600.84, with three specific MRM transitions (Figure 3.11 A,B and C). It was noted that two peaks which resulted in similar fragmentation pattern were detected, they also eluted very close to each other suggesting that they are closely related (Figure 3.12). The transitions of the ion at *m/z* 600.84 were found to happen at different collision energies.. Herein, three collision energy 25, 35 and 45 were used to generate multiple product ions which were further used to create the MRM method. The detection of IN was enabled based on the expected retention time and intensity of product ions of the three MRM transitions (Figure 3.11 A,B and C). It was also observed that an increase in CE, resulted in an increase of the peak intensity of the product ions.

Figure 13 depicts some of the multiply charged trypsin digest products from which the MS^2 couldn't be generated. Because of the multiple charge gained, it was discovered that it was impossible to generate MS/MS data because no MS/MS results were obtained when each of the ion clusters was chosen as a precursor. Figure 1.6 depicts cluster ions as a result of multiple charges, which corresponds to these findings (Mano & Goto, 2003). Furthermore, Data Dependent Acquisition (DDA) was used to automatically select the precursor, but no MS/MS results were obtained even with the automated approach. According to Blackburn *et al* (2010), it is unlikely that a peptide will be fragmented at the apex of its chromatographic peak, which may affect the interpretation of the MS2. Kuster *et al* (2005) and Michalski *et al* (2011) went on to say that DDA performance decreases as sample complexity increases because the semi-stochastic selection of precursor ions exacerbates certain identification limitations.

3.11 Data analysis workflows

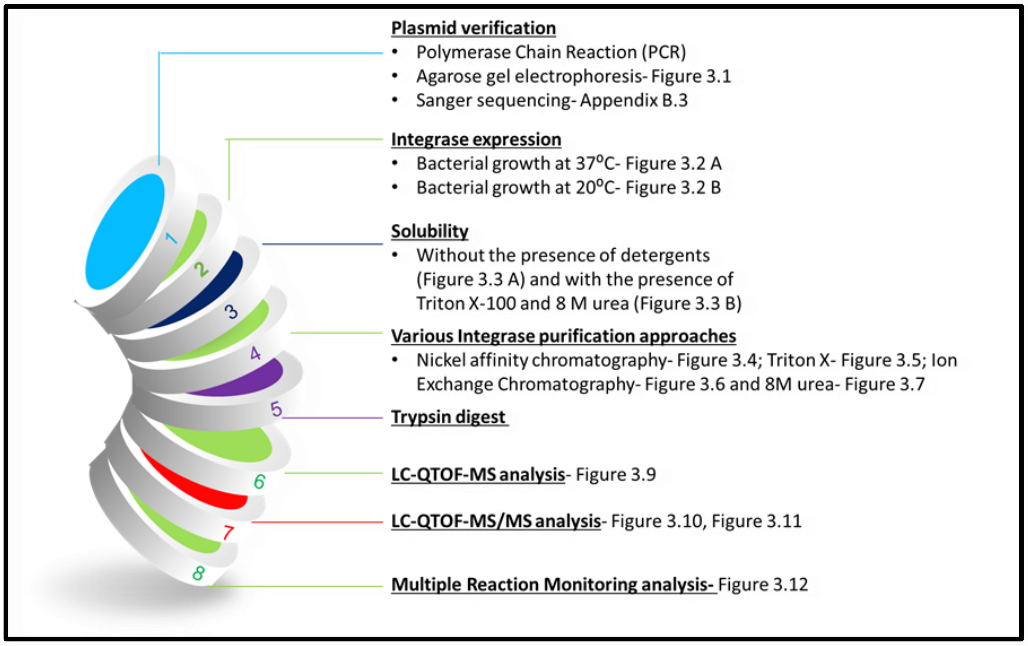

1
2
3
4
5
6
7
8
Plasmid verification
• Polymerase Chain Reaction (PCR)
• Agarose gel electrophoresis- Figure 3.1
• Sanger sequencing- Appendix B.3
Integrase expression
• Bacterial growth at 37°C- Figure 3.2 A
• Bacterial growth at 20°C- Figure 3.2 B
Solubility
• Without the presence of detergents (Figure 3.3 A) and with the presence of Triton X-100 and 8 M urea (Figure 3.3 B)
Various Integrase purification approaches
• Nickel affinity chromatography- Figure 3.4; Triton X- Figure 3.5; Ion Exchange Chromatography- Figure 3.6 and 8M urea- Figure 3.7
Trypsin digest
LC-QTOF-MS analysis- Figure 3.9
LC-QTOF-MS/MS analysis- Figure 3.10, Figure 3.11
Multiple Reaction Monitoring analysis- Figure 3.12

Figure 3.14. Inforgraphical display summerising the workflow followed during this study and the major findings thereof.

Chapter 4: References

Agranoff, D., Fernandez-Reyes, D., Papadopoulos, M.C, Rojas, S.A, Herbster, M, Loosemore, A, Tarelli, E. and Sheldon, J., 2006. Identification of diagnostic markers for tuberculosis by proteomic fingerprinting of serum. *Lancet,* 368: 1012-1021.

Anderson, L. and Hunter, C.L. 2006. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Molecular Cell Proteomics,* 6: 2212-2229.

Armstrong, W.S. and Taege, A.J., 2007. HIV screening for all: the new standard of care. *Cleve Clinical Journal of Medicine,* 74, 297-301.

Arrildt, K.T., Joseph, S.B. and Swanstrom, R., 2013. The HIV-1 Env protein: A coat of many colors. *Current HIV/AIDS Reports,* 9, 52-63.

Balakrishnan, A., Zastrow, D., Jonsson, C.B., 1996. Catalytic activities of the human T-cell leukemia virus type II integrase. *Virology*, 219, 77-86.

Banerjee, S. and Mazumdar, S., 2012. Electrospray ionization mass spectrometry: A technique to access the information beyond the molecular weight of the analyte. *International Journal of Analytical Chemistry.*

Blackburn, K., Mbeunkui, F., Mitra, S.K., Mentzel, T., 2010. Improving protein and proteome coverage through data-independent multiplexed peptide fragmentation. *Journal of Proteome Research,* 9, 3621-3637.

Branson, B.M., 2007. State of the ART for diagnosisof HIV Infection. Diagnosis of HIV infection. *Clinical Infectious Diseases,* 45, 221-225.

Briand, L., Marcion, G., Neiers, F., 2016. A self-inducible heterologous protein expression system in *Escherichia coli. Scientific Reports*, 6, 33037.

Brin, E.Y.J., Skalka, A.M. and Leis, J., 2000. Modelling the late stepsin HIV-1 retroviral integrase-catalyzed DNA integration. *Journal in Biological Chemistry*, 275, 39287-39295.

Buonaguro, L., Tornesello, M.L. and Buonaguro, F.M., 2007. Human Immunodeficiency Virus Type 1 Subtype Distribution in the Worldwide Epidemic: Pathogenetic and Therapeutic Implications. *Journal of Virology,* 81, 10209–10219.

Butsch, M. and Boris-Lawrie, K., 2002. Desting of unspliced retroviral RNA: Ribosome and/or virion?. *Journal of Virology,* 76, 3089-94.

Butto, S., Suligoi, B., Fanales-Belasio, E. and Raimondo, M., 2010. Laboratory diagnostics for HIV infection. *Diagnostic Tools for HIV Infection,* 46, 24-33.

Caumont, A.B., Jamieson, G.A., Pichuantes, S, Nguyen, A.T., Litvak, S. and Dupont, C.H., 1996. Expression of functional HIV-1 integrase in the yeast Saccharomyces cerevisiae leads to the emergence of a lethal phenotype: potential use for inhibitor screening. *Current Genetics*, 29, 503–510.

Chan, D.C. and Kim, P.S., 1998. HIV entry and its inhibition. *Cell,* 93, 681-4.

Chen, B., 2019. Molecular mechanism of HIV-1 entry. *Trends in Microbiology*, 27, 878-891.

Chen, G.D. and Pramanik, B.N., 2008. LC-MS for protein characterization: current capabilities and future trends. *Expert reviews. Proteomics,* 5, 435-444.

Costin, J.M., 2007. Cytopathic mechanisms of HIV-1. *Virology Journal,* 4, 100.

Craigie, R., 2001. HIV integrase, a brief overview from chemistry to therapeutics. *Journal of Biological Chemistry,* 276, 23213-23216.

Craigie, R., 2012. The molecular biology of HIV integrase. *Future Virology,* 7, 679-686.

Delelis, O., Carayon, K., Mouscadet, J.F., 2008. Integrase and integration: biochemical activities of HIV-1 integrase, *Retrovirology,* 5. 114.

Domon, B., Aebersold, R., 2006. Mass spectrometry and protein analysis. *Science,* 312, 212-217.

Duracova, M., Kilmentova, J., Fucikova, A. and Dresler, J., 2018. Proteomic methods of detection and quantification of protein toxins. *Toxins,* 10, 99, 1-30.

Elliot, J.L., Eschbach, J.E., Koneru, P.C., Li, W., Chaver, M.P., Townsend, D., Lawson, D.Q., Engelman, A.N., Kvaratskhelia, M., Kutluay, S.B., 2020. Integrase-RNA interactions underscore the critical role of integrase in HIV-1 virion morphogenesis. *Microbiology and Infectious Disease*, 9. 1-28.

Esposito, D. and Craigie, R., 1999. HIV integrase structure and fuction. *Advances in Virus Research,* 52, 319-324.

Feldberg, L., Schuster, O., Elhanany, E., Laskar, O., Yitzhaki, S. and Gura, S., 2019. Rapid and sensitive identification of ricin in environmental samples based on lactamyl agarose beads using LC-MS/MS (MRM). *Journal of Mass Spectrometry*, 55, e4482.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. and Whitehouse, C.M., 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science,* 246, 64-71.

Fiebig, E.W., Wright, D.J. and Rawai, B.D., 2003. Dynamics of HIV-1 viremia and antibody seroconversion in plasma donors: Implications of diagnosis and staging of primary HIV-1 infection. *AIDS,* 17, 1871-1879.

Fleurbaaij, F., van Leeuwen, H.C., Klychnikov, O.I., Kuijper, E.J., Hensbergen, P.J., 2015. Mass Spectrometry in Clinical Microbiology and Infectious Diseases. *Chromatographia*, 78, 379–389.

Gomez, E., Jespersen, D.J., Harring, J.A. and Binnicker, M.J., 2010. Evaluation of the Bio-Rad BioPlex 2200 syphilis multiplex flow immunoassay for the detection of IgM- and IgG-class antitreponemal antibodies. *Clinical Vaccine Immunology,* 17, 966–8.

Graves, P.R. and Haystead, T.A.J., 2002. Molecular biologist's guide to proteomics. *Microbiology and molecular biology reviews,* 66, 39-63.

Guan, M., 2007. Frequency, causes and new challenges of indeterminate results in western blot confirmatory testing for antibodies to human immunodeficiency virus. *Clinical Vaccine Immunology,* 14, 649-659.

Gundry, R.L., White, M.Y., Murray, C.I., Kane, L.A., Fu, Q., Stanley, B.A., Van Eyk, J.E., 2010. Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. *Current Protocols in Molecular Biology,* 25

Gurtler, L.A, Muhlbacher, U., Michl, H., Hofmann, G.G., Paggi, V., Bossi, R., Thorstebsson, R.G., Villaescusa, A., Eiras, J.H., Hernandez, W. and Melchior, F., 1998. Reduction of the diagnostic window with a new combined p24 antigen and human immunodeficiency virus antibody screening assay. *Journal of Virology Methods,* 75, 27-38.

Hanly, J.G., Su, L., Farewell, V. and Fritzler, M.J., 2010. Comparison between multiplex assays for autoantibody detection in systemic lupus erythematosus. *Journal of Immunology Methods,* 358, 75–80.

Hamza, M., Ali, A., Khan, S., Ahmed, S., 2020. nCOV-19 peptides mass fingerprinting identification, binding and blocking of inhibitors flavonoids and anthraquinone of Moringa oleifera and hydroxychloroquine. *Journal of Biomolecular Structure & Dynamics*, 39, 1-11.

Hellmund, C. and Lever, A.M.L., 2016. Coordination of genomic RNA packaging with viral assembly in HIV-1. *Viruses,* 8, 192.

Hemelaar, J., 2012. The origin and diversity of the HIV-1 pandemic. *Trends in Molecular Medicine*, 18, 182–192.

Henzel, W.J., Watanabe, C. and Stults, J.Y.J., 2003. Protein identification: The origins of peptide mass fingerprinting. *Journal of the American Society for Mass Spectrometry,* 14, 931-942.

Ho, C.S., Lam, C.W.K., Chan, M.H.M., Cheung, R.C.K., Law, L.W., Lit, L.C.W., Ng, K.F., Suen, M.W.M. and Tai, H.L., 2003. Electronspray ionization mass spectrometry: Principles and clinical applications. *The Clinical Biochemist Reviews,* 24, 3-12.

Ho, C.S., Lam, K.C., Chan, M.H., Cheung, R.C.K., Law, L.K., Lil, L.C.W., Ng, K.F., Suen, M. and Tai, H.L., 2003. Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. *The Clinical biochemist. Reviews / Australian Association of Clinical Biochemists,* 24, 3-1.

Ho, D.D., Neumann, A.U., Perelson, A.S., Chen, W., Leonard, J.M. and Markowitz, M., 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature,* 373, 123–126.

Hosseini, I. and Mac Gabhann, F., 2012. Multi-Scale Modeling of HIV Infection in vitro and APOBEC3G-Based Anti-Retroviral Therapy. *PLoS Compututer Biology,* 8.

Huang, W., Wei, W., Shi, X.T., Jiang, T., 2017. The analysis of the detection performance of nucleic acid testing and ELISA fort HIV, HBV and HCB. *Frontiers in Laboratory Medicine,* 1, 200-202.

Huttenhain, R., Malmstrom, J., Picotti, P. and Aebersold, R., 2009. Perspectives of targeted mass spectrometry for protein biomarker verification. *Current Opinion in Chemical Biology,* 13, 518–25.

Jakubke, H. and Sewald, N., 2008. Amino acids. Peptide from A to Z: A Concise Encyclopedia. *Germany: Wiley-VCH,* 20.

James, D.A., Meintjes, G. and Brown, G.D., 2014. A neglected epidemic: fungal infections in HIV/AIDS. *Trends in Microbiology,* 22, 120-127.

Jani, I.V., Megii, B., Mabunda, N., Vubil, A., Sitve, N.S., Tobaiwa, O. and Quevedo, J.I., 2014. Accurate early infant HIV diagnosis in primary health clinics using a point of care nucleic acid test. *Journal of Acquired Immune Deficiency Syndrome*, 67, e1-e4.

Jenkins, T.M., Engelma, A., Ghirlando, R. and Craigie, R., 1996. A soluble active mutant of HIV-1 integrase-involvement of both the core and carboxyl-terminal domains in multimerization. *Journal of Biological Chemistry,* 271, 7712-7718.

Jim, C., 2007. An introduction to amino acids. *Chemguide.*

Jozwik, I.K., Passos, O. D., Lyumkis, D., 2020. Structural Biology of HIV integrase strand transfer inhibitors, *Trends in Pharmacological Sciences*, 41, 611-626.

Karki, R.G., Tang, Y., Burke, T.R., Nicklaus, M.C., 2005. Model of full length HIV-1 integrase complexed with viral DNA as template for anti-HIV drug design. *Journal of Computer-Aided Molecular Design,* 18, 739-60.

Khadir, A. and Tiss, A., 2013. Proteomics Approaches towards Early Detection and Diagnosis of Cancer. *Journal of Carcinogenesis & Mutagenesis. Cancer Diagnosis, Treatment and Therapy,* S14.

Kingsmore S.F., 2006. Multiplexed protein measurement: technologies and applications of protein and antibody arrays. *Natural Review Drug Discovery,* 5, 310–320.

Kitamura, Y., Usami, R., Ichihana, S., Kida, H., Satoh, M. and Tocnimoto, H., 2017. Plasma protein profillinf gor potential biomarkerts in the early diagnosis of Alzheimer's disease. *Neurological Research,* 39, 231-238.

Klein, J., Bjorkman, P.J. and Rall, G.F., 2010. Few and far between: How HIV may be evading antibody avidity. *PLoS Pathogens,* 6, e1000908.

Kuster, B., Schirle, M., Mallick, P., Aebersold, R., 2005. Scoring proteomes with proteotypic peptide probes. *Natural Reviews of Molecular Cell Biology,* 6, 577-583.

Kuzyk, M.A., Smith, D., Yang, J., Cross, T.J., Jackson, A.M., Hardie, D.B., Anderson, N.L., Borchers, C.H., 2009. Multiple reaction monitoting-based, multiplexed, absolute quantitation of 45 proteins in human plasma. *Molecular Cell Proteomics,* 8, 1860-77.

Kuzyk, M.A., Parker, C.E. and Borchers, C.H., 2011. Development of MRM based assays for plasma proteins. In: Backvall H, ed. *Methods in Molecular Biology.* New York, NY: Human Press.

Lange, J.M., Paul, D.A., Huisman, H.G., Wolf, F., Berg, H., Coutonho, R.A., Danner, S.A., Noorda, J. and Goudsmit, J., 1986. Persistent HIV antigenanemia and decline of HIV core antibodies associated with trabsition to AIDS. *British Medical Journal*, 293, 1469-1462.

Lataillade, M.P.H. and Kozal, M.D., 2006. The Hunt for HIV-1 Integrase Inhibitors. *AIDS Patient Care and STDs*, 20, 489-501.

Laskay, U.A., Lobas, A.A., Srzentic, K., Gorshkov, M.V., Tsybin, Y.O., 2013. Proteome digestion specificity analysis for rational design of extended bottom-up and middle-down prpteomics experiments. *Journal of Proteome,* 12, 5558-5569.

Li, M., Mizuuchi, M., Burke, T.R. and Craigie, R., 2006. Retroviral DNA integration: reaction pathway and critical intermediates. *EMBO Journal,* 25, 1295-1304.

Liebler, D.C., 2002. Introduction to proteomics: Tools for the new biology (1st ed.). Totowa, NJ: Humana Press.

Liebler, D.C., and Zimmerman, L.J., 2013. Targeted quantitation of proteins by mass spectrometry. *Biochemistry,* 52, 3797-3806.

Lin, D., Tabb, D.L. and Yates, J.R., 2003 Large-scale protein identification using mass spectrometry. *Biochimica et Biophysica*, 1646, 1-10.

Liuni, P. and Wilsom, D., 2011. Understanding and optimizing electrospray ionization techniques for proteomic analysis. *Expert Review of Proteomics,* 8, 197-209.

Loo, J.A., Udseth, H.R. and Smith, R.D., 1989. Peptide and protein-analysis by electrospray ionization mass-spectrometry and capillary electrophoresis mass-spectrometry. *Analytical Biochemistry,* 179, 404-412.

Ma, X., Tang, J. and Li, C., 2014. Identification and quantification of ricin in biomedical samples by magnetic immunocapture enrichment and liquid chromatography electrospray ionization tandem mass spectrometry. *Analytical and Bioanalytical Chemistry*, 406, 5147-5155.

Mano, N. and Goto, J., 2003. Biomedical and biological mass spectrometry. *Analytical Sciences,* 19, 3-14.

Masike, K and Madala, N.E., 2018. Synchronized survey scan approach allows for efficient discrimination of isomeric and isobaric compounds during LC-MS analyses. *Journal of Analytical Methods in Chemistry*.

McGovern, S.L., Caselli, E., Griyorieff, N. and Shoichet, B.K., 2002. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *Journal of Medical Chemistry,* 45, 1712-22.

Michael, N.L., 1999. Host genetic influences on HIV-1 pathogenesis. *Current Opinion in Immunology,* 11, 466–474.

Michalski, A., Cox, J., Mann, M., 2011. More than 100 000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *Journal of Proteome Research,* 10, 1785-1793.

Nadler, W.M., Waidelich, D., Kerner, A., Hanke, S., Berg, R., Trumpp, A. and Rosli, C., 2017. MALDI verus ESI: The impact of the ion source on peptide identification*. Journal of proteome research,* 16, 1207-1215.

Noda, A.A., Fleitas, O., Rodriguez, I., Beltran J.F., Falcon, R., Almaguer, T. and Samaha, T.H., 2017. Triton X-100 Vs. Triton X-114: Isolation of Outer Membrane Proteins from Leptospira Spp. *International Journal of Veterinary Science & Technology*, 1, 007-012.

Ochodo, E.A., Kakourou, A., Mallet, S. and Deeks, J.J., 2018. Point of care tests detecting HIV nucleic acids for diagnosis of HIV infection in infants and children aged 18 months or less. *The Cochrane Database of systematic reviews,* 11, cd013207.

Oh, J.W. and Shin, C.G., 1996. Purification and characterization of the human immunodeficiency virus type 1 integrase expressed in *Escherichia coli*. *Molecules and Cells,* 6, 96–100

Ong, S. and Mann, M., 2005. Mass spectrometry-based proteomics turns quatitative. *Natural Chemistry Biology,* 1, 252-262.

Op de Coul, E.L., Prins, M., Cornelissen, M., van der Schoot, A., Boufassa, F., Brettle, R.P., Hernández-Aguado, L., Schiffer, V., McMenamin, J., Rezza, G., Robertson, R., Zangerle, R., Goudsmit, J., Coutinho, R.A. and Lukashov, V.V., European and Italian Seroconverter Studies 2001. Using phylogenetic analysis to trace HIV-1 migration among western European injecting drug users seroconverting from 1984 to 1997. *AIDS London. England,* 15, 257–266.

Pala, N., Esposito, F., Tramontano, E., Singh, P.K., Sanna, V., Carcelli, M., Haigh, L.D., Satta, S., Sechi, M., 2020. Development of a raltegravir-based photoaffinity-labeled probe for human immunodeficiency virus-1 integrase capture. *Medicinal Chemistry Society,* 11, 1986-1992.

Panwar, U. and Singh, S., K., 2021. In Silico virtual screening of potent inhibitor to hamper the interaction between HIV-1 integrase and LEDGF/p75 interaction using E-pharmacophore modeling, molecular docking, and dynamics simulations. *Computational Biology and Chemistry,* 93,107509.

Park, J.N., Yun, J.N., Shi, Y., Han, J., Li, X., Jin, Z., Kim, T., Park, J., Park, S., Liu, H. and Lee, W., 2019. Non-cryogenic structure and dynamics of HIV-1 integrase catalytic core domain by X-ray free electron lasers. *International Journal of Molecular Sciences,* 20, 1943.

Parekh, B.S., Ou, C.Y., Fonjungo, P.N., Kalou, M.B., Rottimghaus, E., Puren, A., Alexander, H., Cox, M.H., Nkengasong, J.N., (2019). Diagnosis of Human Immunodeficiency virus infection. *Clinical Microbiology Reviews*, 32, 1-55.

Parker, C.E., Pearson, T.W., Anderson, N.L. and Borchers, C.H., 2010. Mass spectrometry based clinical proteomics- a review and prospective. *Analyst,* 135, 1830-1838.

Pitt, J., 2009. Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *Clinical Biochemistry Review,* 30, 19-34.

Quashie, P.K., Han, Y.S., Hassounah, S., Mesplede, T. and Wainberg, M.A., 2015. Structural studies of the HIV-1 integrase protein: compound screening and structural characterization of a DNA-binding inhibitor. *PLoS One,* 10: e0128310.

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

Rawlings, A.E., 2016. Membrane proteins: always an insoluble protein?. *Biochemistry Society Transactions,* 44, 790-795.

Rojas, V.K., and Park, W., 2019. Role of the ubiquitin proteasome system (UPS) in the HIV-1 life cycle. *International Journal of Molecular Sciences,* 20, 2984.

Santo, R., 2014. Inhibiting the HIV integration process: past, present and the future. *Journal of Medicinal Chemistry,* 57, 539-566.

Schweitzer, C.J., Jagadish, T., Haverland, N., Ciborowski, T., Belshan, M., 2013. Proteomic analysis of early HIV-1 nucleoprotein complexes. *Journal of Proteome Research*, 12, 559-72.

Sharp, P.M. and Hahn, B.H., 2011. Origins of HIV and the AIDS pandemic. Cold Spring Harb. *Perspectives in Medicine,* 1, a006841.

Sherwood, C.A., Eastham, A., Lee, L.W., Peterson, A., Eng, J.K., Shteynberg, D., Mendoza, L., Deutsch, E.W., Risler, J., Tasman, N., 2009. A software application for spectral library-based MRM transitions list assembly. *Journal of Proteome Research,* 8, 4396-4405.

Simon, V., Ho, D.D. and Karim, Q.A., 2006. HIV/AIDS epidemiology, pathogenesis, prevention, and treatment. *Lancet,* 368, 489–504.

Singh, S.K., Goel, G., Rathore, A.S., (2019). A novel approach for protein identification from complex cell proteome using modified peptide mass fingerprinting algorithm. *Electrophoresis*, 40, 3062-3073.

Singh, R., Hassan M.I., Islam A. and Ahmad, F., 2015. Cooperative Unfolding of Residual Structure in Heat Denatured Proteins by Urea and Guanidinium Chloride. *PLoS One,* 10. e0128740.

Soto-Rifo, R., Limousin, T., Rubilar, P.S., Ricci, E.P., Decimo, D., Moncorge, O., Trabaud, M.A., Andre, P., Cimarelli, A. and Ohlmann, T., 2011. Different effects of the TAR structure on HIV-1 and HIV-2 genomic RNA translation. *Nucleic Acids Research,* 40, 2653-2667.

Sourgen, F., Maroun, R.G., Frere, V. and Bouziane, M., 1996. A synthetic peptide from the human immunodeficiency virus type-1 integrase exhibits coiled-coil properties and interferes with the in vitro integration activity of the enzyme. *European Journal of Biochemistry,* 240, 765-773.

Tehrani, Z.R., Azadmanesh, K., Mostafavi, E., Soori, S., Azizi, M. and Khabiri, A., 2015. Development of an Integrase-based ELISA for Specific Diagnosis of Individuals Infected with HIV. *Journal of Virological Methods.*

Temin, H.M., 1993. Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. *Proceedings of the National Academy of Sciences*, 90, 6900–6903.

UNAIDS., 2006. AIDS epidemic update: special report on HIV/AIDS : December 2006.

UNAIDS., 2016. United Nations Programme on HIV/AIDS report on the global AIDS epidemic. United Nations.

UNAIDS fact sheet - Latest statistics on the status of the AIDS epidemic - UNAIDS_FactSheet_en.pdf [WWW Document], n.d. URL http://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf (accessed 11.13.17).

Wang, J.Y., Ling, H., Yang, W., Craigie, R., 2001. Structure of a two-domain fragment of HIV-1 integrase: Implications for domain organization in the intact protein. *European Molecular Biology Organization,* 20, 7333-7343.

Wang, Z., Shang, H. and Jiang, Y., 2017. Chemokine and chemokine receptors: Accomplices for human immunodeficiency virus infection and latency. *Frontiers in Immunology,* 8, 1274.

Wilkins, M.R., Lindskog, I., Gasteiger, E., 1997. Detailed peptide characterization using PEPTIMASS- a World-Wide-Web accessible tool. *Electrophoresis,* 18, 403-408.

World Health Organization., 2016. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection, Recommendations for a public health approach. Second ed. Geneva: World Health Organisation, Accessed 4 May 2017.

World Health Organization., 2021. Global progress report on HIV, viral hepatitis and sexually transmitted infections, Accountability for the global health sector stratergies 2016-2021: actions for impact.

Wyatt, R. and Sodroski, J., 1998. The HIV-1 envelope glycoproteins: fusogens, antigens and immunogens. *Science,* 280, 1884-8.

Zhao, X.H., He, X.W., Li, W.M., Wang, J.H., Yang, L.S., Peng, Y.P., Liu, X.Y., 2009. Synthesis of HIV-1 integrase gene and its high level expression in *Escherichia coli. Progress in Modern Biomedicine,* 9, 4024-4026.

Zheng, J.S., Lange, V., Ossola, R., Eckhardt, K., Krek, W., Aebersold, R. and Domon, R., 2007. High Sensitivity Detection of Plasma Proteins by Multiple Reaction Monitoring of N-Glycosites. *Molecular & Cellular Proteomics,* 6, 1809–1817.

Zheng, Y.H., Loosin, N. and Peterlin, B.M., 2005. Newly identified host factors modulate HIV replication. *Immunology Letters,* 97, 225-34.

Chapter 5: Conclusion and future perspectives

The main aim of the current study was to develop a proof of concept idea of using mass spectrometry for medical diagnosis. Here, a detailed approach of using QTOF-MS was conducted using recombinant HIV-1 integrase as a target. This gene was placed under the control of the pET15b expression vector which was later successfully transformed in *E. coli* BL21 (DE3) cells. To confirm the identity of the target protein/ gene, PCR in combination with sanger sequencing was used and the online identity search revealed this protein to be an HIV-1 IN subtype B, which was isolated from a Canadian HIV-1 positive patient.

The recombinant protein was successfully expressed in *E.coli* cells using a low growth temperature of around 20 ºC as shown by several electrophoretograms presented herein. This protein was found to resolve at a size of 32 kDa, which is consistent with the size reported in literature. Furthermore, solubility studies revealed this protein to be highly insoluble, however, the use of detergent slightly improved its solubility. For purification, both the insoluble and soluble fraction was used and the protein was successfully purified using Nickel (Ni) affinity chromatography. To get a pure protein, various purification steps were conducted until a very pure protein was achieved, which was devoid of other impurities, thus other endogenous *E. coli* proteins.

Before LC-MS, *in silico* trypsin digest of the sequence retrieved from NCBI was conducted. This generated fingerprinting was essential during the optimization of the MS parameters such as ionization mode and mass range coverage of the instrument. The LC-MS analyses reveled the presence of multiple trypsin digestion products of IN protein. Furthermore, fragment ions with mass similarities to those achieved through *in silico* digest were selected for downstream processing. A fragmnet with precursor ion around *m/z* 600 was selected and further subjected to MS/MS and MRM analyses. This ion was successfully detected using tandem MS^2 method and MRM. It can be said that its fragmentation conditions which was achieved through differential collision energies can be used for further development of a robust method aimed at detection of IN from various samples. Other ions which are symptomatic of IN were also detected in the current study, however, their downstream processing through MS^2 were found to be impossible due to charge multiplicity, a common occurrence during electrospray ionization mode. A very robust form of MS/MS approach known as data dependent acquisition (DDA) approach was also used to automatically select the precursor ions was also conducted without much improvement on the MS^2 analyses.

The results of the current study has confirmed that MS is a feasible technique for protein based diagnosis of diseases. Herein, a more detailed approach using recombinant protein associated with HIV infection is presented. The shortfall of the current study was on the ion selection for building a more robust method which at least target multiple ions associated with a specific proteins, especially when ESI mode is used since it results in multiple charged ion species. To circumvent this problem, a newly developed MS approached know as data independent acquisition (DIA) approach will be used for ion selection which will aid in more ions being selected for MRM method development. Lastly, even though the current study was only a proof of concept, it will be ideal to use it on real samples isolated from HIV infected individuals.

Appendix A: Supplementary data

A.1. Colony PCR conformation of transformation of pET15b/IN constructs into BL21(DE3) cells.

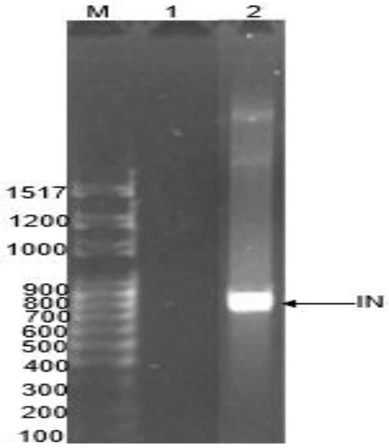


Figure A.1. pET15b/IN and BL21(DE3) cells colony PCR. Lane M: DNA molecular weight marker; Lane 1 is the negative control-untransformed cells, lane 2 is the transformed cells.

A.2. Determination of protein concentration using Bradford assay

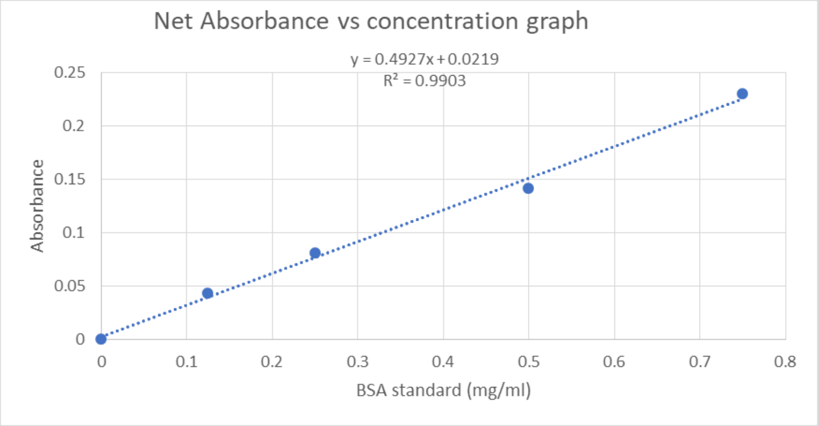


Figure A.2. BSA standard curve generated from 10 mg/mL BSA stock. Absorbance of different BSA dilution read at 595 nm was plotted against respective concentration. Linear regression and the R2 value were generated using Excel.

A 3.Sanger sequencing

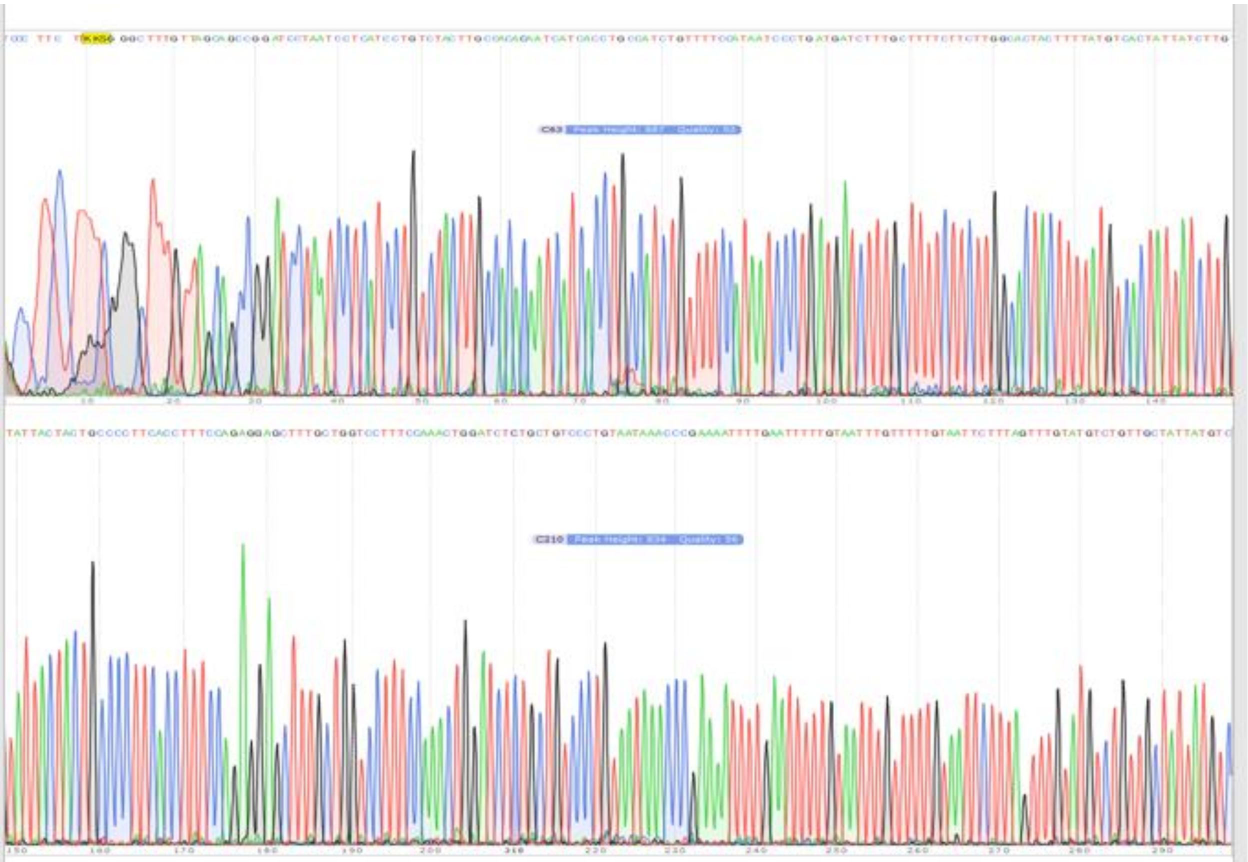


Figure A.3 (A). Confirmation of IN gene using DNA sequencing. Sequencing was performed using the Big Dye method. A T7 reverse primer was used.

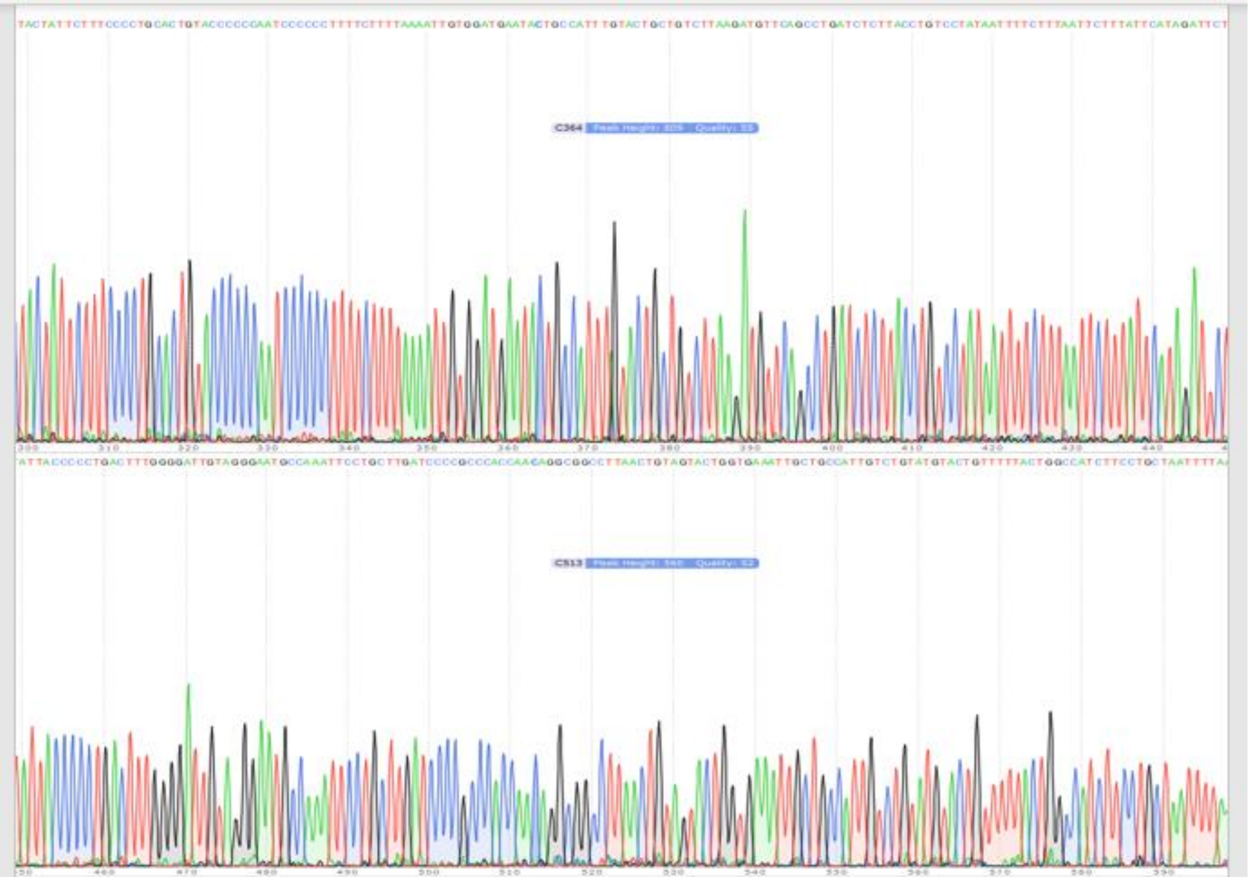


Figure A.2 (B). Confirmation of IN gene using DNA sequencing. Sequencing was performed using the Big Dye method. A T7 reverse primer was used.

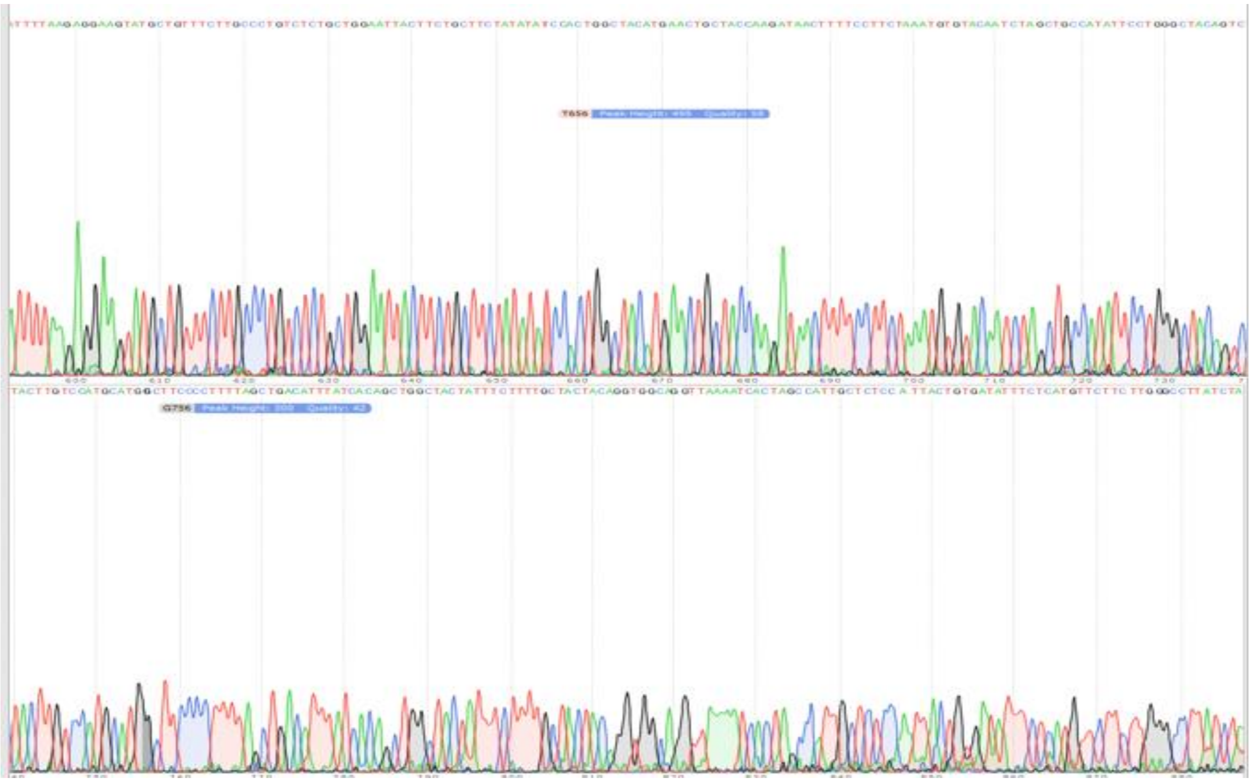


Figure A.3 (C). Confirmation of IN gene using DNA sequencing. Sequencing was performed using the Big Dye method. A T7 reverse primer was used.

Appendix B: **List of Reagents**

Reagent	Supplier
Acetic acid	Merck, Germany
Agarose whitehead scientific	South Africa
Ammonium persulphate	Merck, Germany
Ampicillin	Sigma, U.S.A
Bovine serum albumin	Sigma, U.S.A
Bromophenol blue	Sigma, U.S.A
Calcium chloride	Merck, Germany
Chloramphenicol	Sigma, U.S.A
Coomasie brilliant blue R250	Merck, Germany
Diethithreitol	Sigma, U.S.A
DreamTaq master mix	Thermo Scientific, U.S.A
Ethidium bromide	Sigma, U.S.A
Glacial acetic acid	Merck, Germany
Glycerol	Merck, Germany
Glycine	Merck, Germany
Imidazole	Sigma, U.S.A
Isopropyl-1-thio-D-galacopyranoside	Sigma, U.S.A
Lysozyme	Merck, Germany
Magnesium chloride	Merck, Germany
Methanol	Merck, Germany
Ni-NTA resin	Thermo Scientific, U.S.A
PagerRuler Prestained Protein Ladder	Thermo Scientific, U.S.A
Potassium chloride	Merck, Germany
Sodium chloride	Merck, Germany
Sodium dodecyl sulphate	Merck, Germany
Sodium hydroxide	Merck, Germany
TEMED	Sigma, U.S.A
Tris	Merck, Germany
Tryptone Merck	Merck, Germany
Tween 20	Merck, Germany
Urea	Melford, UK

Yeast extract powder	Merck, Germany
β-mercaptoethanol	Sigma, U.S.A