Preparation of HIV-1 founder virus envelope glycoprotein core crystals, and computational characterisation of covalent and non-covalent complexes it forms with 2dCD4

Dominic Sevenster

A dissertation submitted to the Faculty of Health Sciences,

University of the Witwatersrand,

In fulfilment of the requirements for the degree of

Master of Science in Medicine

Johannesburg, April 2021

Declaration

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

FACULTY OF HEALTH SCIENCES

PLAGIARISM DECLARATION TO BE SIGNED BY ALL HIGHER DEGREE STUDENTS

SENATE PLAGIARISM POLICY: APPENDIX ONE

I Dominic Sevenster (Student number: 814679) am a student registered for the degree of MSc Med Molecular Medicine in the academic year 2021.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.
- I have included as an appendix a report from "Turnitin" (or other approved plagiarism detection) software indicating the level of plagiarism in my research document.

Signature: _____ Date: 1November 2021

Abstract

The envelope glycoprotein (Env) is responsible for human immunodeficiency virus type 1 (HIV-1) entry into CD4+ T cells through the binding of CD4 and represents an ideal target for therapeutics as well as vaccine development strategies. During transmission, only a limited number of founder viruses (FV) are capable of establishing infection in a new host, thus an enhanced understanding of the FV Env structure is critical in vaccine immunogen development. Previous findings showed that immunizations of small animals with a novel Env–2dCD4^{S60C} complex generated broadly neutralizing antibodies against clinically relevant HIV-1. The mechanism behind this enhanced immune response is attributed to the targeted disulphide bond between the gp120 component of Env and 2dCD4^{S60C} but has not yet been elucidated. This study aimed to optimize the crystallization conditions of a modified unliganded gp120 core FV sequence (gp120$_{FVC}$ExC), as well as compare gp120$_{FVC}$ExC to a similarly modified gp120 core subtype C (gp120Core$_e$) sequence. This allowed for a better understanding of the structure and conformational dynamics of gp120 FV protein compared to a non-founder virus variant. MD analyses were also performed on the gp120$_{FVC}$ExC–2dCD4^{S60C} and gp120$_{FVC}$ExC – 2dCD4^{WT} complexes to identify potential structural changes that may be responsible for the enhanced immunogenicity produced by the S60C mutation in the Env – 2dCD4^{S60C} immunogen.

The gp120 FV sequence was designed and modified to allow for unliganded crystallization through the truncation of the V1, V2, and V3 loops. The modified gp120 was expressed in modified GnTi-deficient cells, purified using lectin affinity chromatography, and deglycosylated with Endoglycosidase H to enhance homogeneity. The purified and homogenate protein was successfully crystallized using a Morpheus Index kit in an under-oil diffusion method, validating the chosen modifications for crystallization. Homology models of gp120$_{FVC}$ExC, gp120Core$_e$, 2dCD4^{WT}, 2dCD4^{S60C}, and the gp120$_{FVC}$ExC – 2dCD4$^{S60C/WT}$ complexes were generated using the SWISS-MODEL platform. Generated homology models were used for 50 ns MD simulations on the Maestro platform at a 1 ns timestep to investigate structural and dynamic differences between the proteins. Further analyses of the MD simulated proteins were performed on the Schrodinger platform. Comparative MD analyses of gp120$_{FVC}$ExC and gp120Core$_e$ showed marked differences in the variable loop regions (V1-V5) as well as overall less fluctuations of gp120$_{FVC}$ExC residues.

The V1 and V4 loops show increased flexibility while the V3 loop exhibited a drastic decrease of approximately 4 Å in fluctuations in $gp120_{FVC}ExC$ when compared to $gp120Core_e$. MD simulations of the $gp120_{FVC}ExC$ – $2dCD4^{S60C}$ complex shows a decrease in fluctuation near the N-terminus at residues 38-43 and marked increase of 2 Å in fluctuations of residues in the V3 loop comparative to $gp120_{FVC}ExC$ – $2dCD4^{WT}$. These investigations show distinct structural and conformational differences between the $gp120_{FVC}ExC$ and the $gp120Core_e$ protein that may be characteristic of a gp120 FV sequence. Additionally, the differences in fluctuations observed in the comparative simulation of $gp120_{FVC}ExC$ – $2dCD4^{S60C/WT}$ complexes of residues 38-43 (reduced in $gp120_{FVC}ExC$ – $2dCD4^{S60C}$) and residues 214-224 (increased in $gp120_{FVC}ExC$ – $2dCD4^{S60C}$) provided insight and an initiation point for further investigation into the enhanced immunogenicity of the novel Env–$2dCD4^{S60C}$ vaccine immunogens.

Table of Contents

List of figures

List of abbreviations

AIDS	: Acquired Immunodeficiency Syndrome
ARV	: Antiretroviral
BCA	: Bicinchonic Acid Assay
BSA	: Bovine Serum Albumin
bNAb	: Broadly neutralising antibody
CD4bs	: CD4 binding site
cDNA	: Complementary DNA
DMSO	: Dimethyl Sulfoxide
Endo H	: Endoglycosidase H
Env	: HIV-1 Envelope glycoprotein
GlcNAc	: N-acetylglucosamine subunit
GnTi	: N-acetylglucosaminyltransferase
gp120	: Envelope glycoprotein 120
gp41	: Envelope glycoprotein 41
$gp120Core_e$	: Envelope glycoprotein 120 Extended Core
$gp120_{FV}$	: Envelope glycoprotein 120 Founder Virus
$gp120_{FVC}ExC$	: Envelope glycoprotein 120 Founder Virus Extended Core
HIV-1	: Human Immunodeficiency Virus type-1
HMW	: High Molecular Weight
kDa	: Kilodalton
LB	: Lysogeny Broth
LMW	: Low Molecular Weight
MD	: Molecular Dynamics
MHC II	: Major Histocompatibility Complex II
MM	: Molecular Modelling
MMP	: Methyl alpha-D-manno-pyranoside

NAb	: Neutralising antibody
NMR	: Nuclear Magnetic Resonance
PBS	: Phosphate Buffered Saline
PBST	: PBS with Tween20
PEI	: Polyethylenimine
PEG	: Polyethylene glycol
PTM	: Post Translational Modification
RMSD	: Root Mean Square Deviation
RMSF	: Root Mean Square Fluctuation
S60C	: Serine – 60 – Cysteine substitution
SDS-PAGE	: Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	: Size Exclusion Chromatography
SSE	: Secondary Structural Elements
TEMED	: Tetramethylethylenediamine
WT	: Wild Type

1.Introduction

1.1. HIV-1 disease prevalence and burden

Human Immunodeficiency Virus type-1 (HIV-1), the causative agent of Acquired Immunodeficiency Syndrome (AIDS), has resulted in a global pandemic since its discovery in the 1980's (Gottlieb et al., 1981; Klatzmann and Gluckman, 1986; Popovic et al., 1984). The disease progression of HIV-1 infection to AIDS without antiretroviral (ARV) intervention leads to the marked fatality, due to a compromised immune system in the infected individual. The estimates of global HIV-1 infections were approximately 38 million individuals, with the highest numbers in Eastern and Southern Africa with 20.7 million individuals, during 2019/2020 (Unaids, 2020). The continual and improved development, as well as increased access to ARV treatment has drastically improved the quality of life, disease outcomes, and life expectancy. This is easily observed by the steady reduction in AIDS related deaths from 1.7 million in 2005 to 690 000 in 2020 (Unaids, 2020). However, the access to ARV treatment is neither universal nor sufficient to address the disease burden, as observed by only 51.2 - 87.7 % of HIV-1 infected individuals having adequate access to ARV treatment (Unaids, 2020). This is despite the increase of 2 million individuals accessing ARV treatment from 2018 to 2019 (Unaids, 2020). Although the prevalence, outcomes, and treatment for HIV-1 point to improved outcomes for infected individuals; the need for an alternative approach to the HIV-1 pandemic is needed since transmissions are still ongoing. This need for an improved response is particularly stark in countries with poor socio-economic outcomes such as Africa.

An alluring approach to dealing with the HIV-1 pandemic is the development and implementation of a prophylactic vaccine. This would greatly limit, if not prevent viral transmission and reduce the continual burden of ever-increasing resources required to combat the disease, increasing accessibility to those already afflicted by HIV-1. Resource-limited countries and settings that already exhibit limited access to adequate treatment, would benefit immensely from such a preventive strategy.

1.2. Structure of HIV-1

HIV-1 is classified in the *Retroviridae* family, and has the characteristic of being positive stranded RNA viruses that transcribe their RNA into complementary DNA (cDNA) to continue the viral replication cycle (Hu and Hughes, 2012). The two positive single stranded RNA molecules of HIV-1 encode for 9 genes (*gag, pol, env, tat, rev, nef, vif, vpu, and vpr*) that result in at least 16 mature proteins used in the life cycle of HIV-1 (Goto et al., 1998; Murray et al., 2011; Watts et al., 2009). HIV-1 consists of a conical capsid that is comprised of protein p24, which itself is encapsulated by protein p17 (Figure 1.1; Murray et al., 2011). This conical capsid houses the two positive RNA stands as well as proteins integral for viral replication and processing such as integrase, protease, RNase H and reverse transcriptase (Figure 1.1). The conical capsid and the surrounding p17 matrix are encapsulated in a host-derived lipid membrane that is acquired when the virus buds out of infected host cells (Checkley et al., 2011). The host derived membrane contains protruding envelope glycoprotein (Env) heterotrimers that are the main target of the host immune response (Pantophlet and Burton, 2006). The protruding Env heterotrimers are comprised of the extracellular gp120 and gp41 transmembrane protein components (Figure 1.1), bound through noncovalent bonds (Binley et al., 2000). This Env is responsible for viral transmission and host cell entry primarily through the binding of gp120 to the host CD4 proteins on $CD4^+$ T cells (Rizzuto et al., 1998). Following CD4 binding, host cell entry is further mediated by the binding of primarily the V3 loop of gp120 to co-receptors CCR5 or CXCR4 (Hoffman, and Doms, 1999).

Pertinently, the HIV-1 Env is the target for neutralizing antibodies (NAb) generated by the host immune system (Pantophlet and Burton, 2006). These NAb responses would be required for the efficacy of a potential prophylactic vaccine.

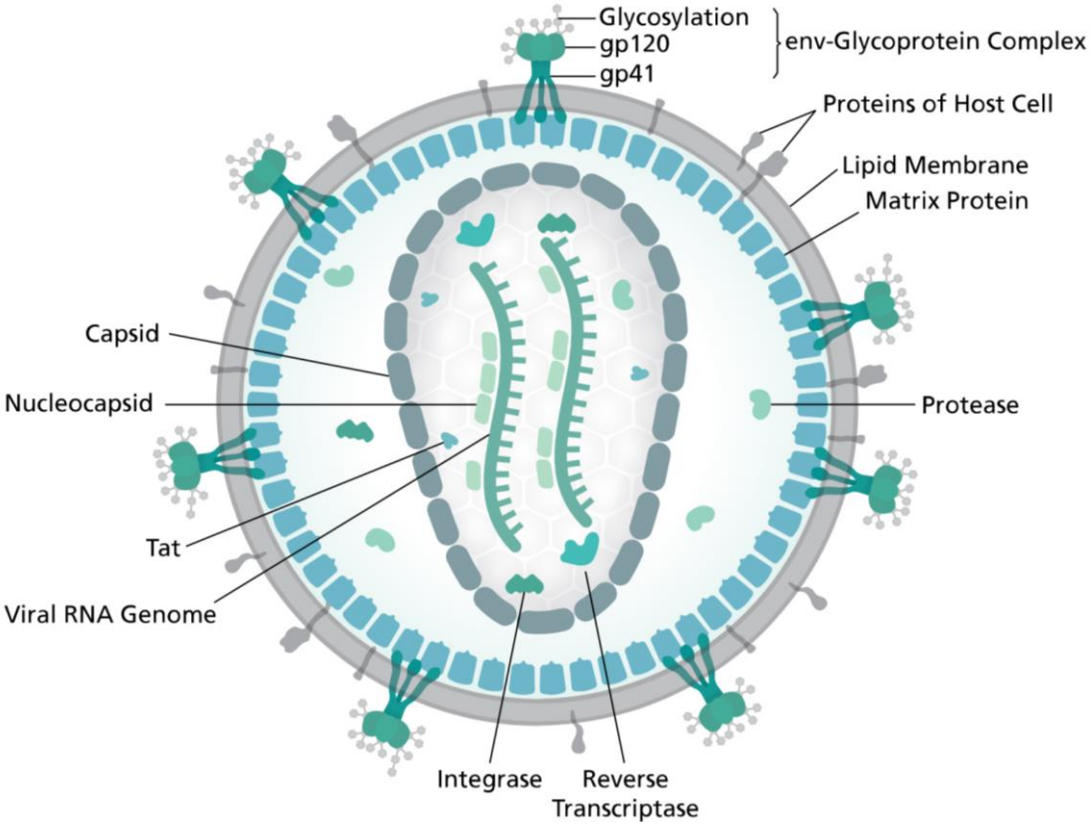

Figure 1.1: A structural representation of the HIV-1 virion with associated structures and proteins. The env-glycoprotein complex (glycans, gp120, gp41) that bind to the host receptor CD4 are situated on the lipid membrane of the virus. The viral RNA genome along with proteins required for replication are encapsulated within a matrix protein that is subsequently covered by a lipid membrane. Image obtained from https://commons.wikimedia.org/wiki/File:HI-virion-structure_en.svg.

1.3. HIV-1 life cycle

The HIV-1 viral and infection replication cycle (Figure 1.2) begins with the binding of the gp120 component to the CD4 glycoprotein on T lymphocytes and other immune cells that express CD4 (Figure 1.2, 1) (Wilen et al., 2012). The binding of gp120 to CD4 initiates structural changes in the HIV-1 Env that allow for the binding of gp120 to chemokine co-receptors CCR5 or CXCR4 (Berger et al., 1999). The tropism of the co-receptor binding is determined by the V3 region of gp120 and appears integral to co-receptor binding (Hoffman, and Doms, 1999). The binding of gp120 to CD4 and the co-receptors induces conformational changes that leads to the exposure of

gp41 of the Env molecule that is essential to host-virus membrane fusion (Weiss, 2003). Exposure of the fusion peptide on gp41 and subsequent formation of a six-helix bundle results in the fusion of the host and viral membranes (Figure 1.2, 2) (Melikyan et al., 2000).

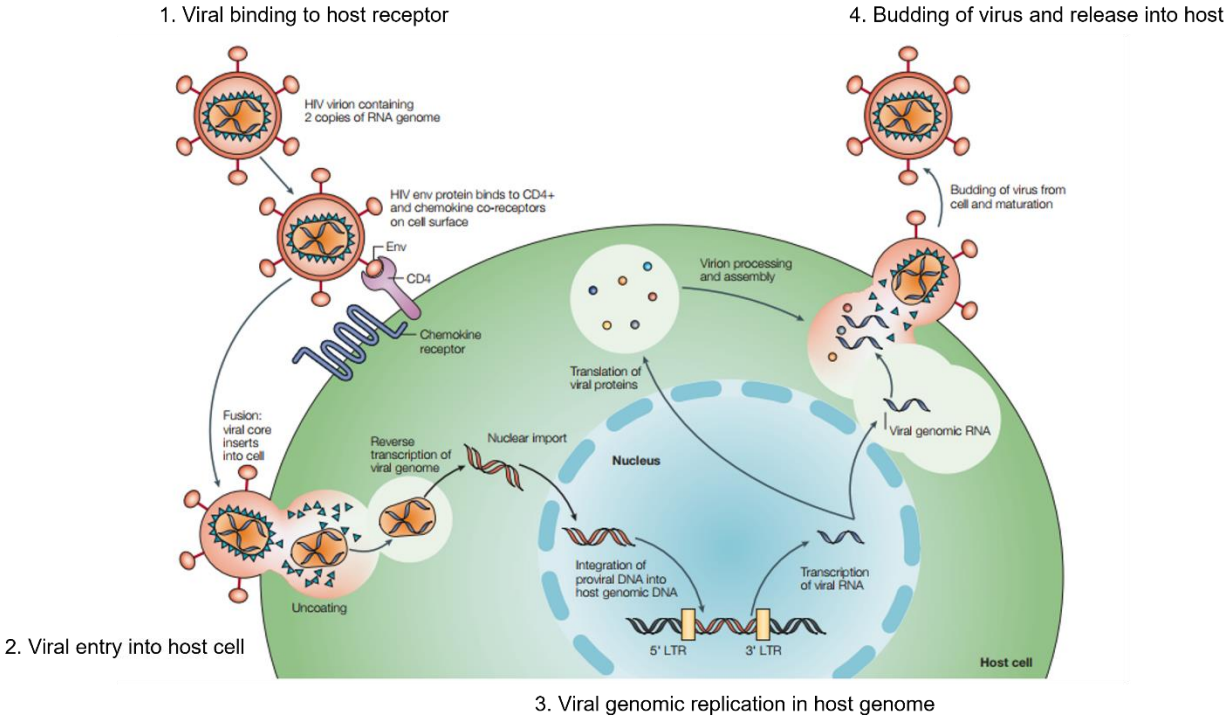

Figure 1.2: Diagrammatic representation of the HIV-1 infection and replication cycle in a host CD4+ cell. HIV-1 initiates binding with the host through env proteins that bind to CD4 on host cells and subsequently to a chemokine co-receptor (1) to initiate membrane fusion (2). This leads to reverse transcription of viral RNA and entry into host genomic DNA where it will be replicated through host machinery (3). The final step in the viral replication cycle is the budding and subsequent release of the virus into the host system (4) for subsequent rounds of infection. Image adapted from (Rambaut et al., 2004).

Once the HIV-1 virus has fused with the host membrane, the conical p24 capsid enters the host cell and uncoats to release the viral RNA strands and other enzymes required for the replication of the virus (Figure 1.2, 2). The viral RNA strands are then transcribed into cDNA through the assistance of the reverse transcriptase enzyme (Bebenek et al., 1993). This is a highly error prone process that leads to the multitude of mutations observed in the HIV-1 genome and subsequent

generation of quasi-species that are responsible for viral escape from ARV treatment as well as the host immune system (Bebenek et al., 1993). Integrase forms a complex with the generated viral cDNA to form a pre-integration complex (PIC) with the assistance of the viral Vpr protein and other host proteins (Vandegraaff and Engelman, 2007). The PIC enters the host nucleus where the cDNA is permanently integrated into the host genome by integrase– allowing for continual expression of the viral genes and resulting in the establishment of a latent HIV-1 infection (Figure 1.2, 3) (Vandegraaff and Engelman, 2007). The integrated viral cDNA is expressed, and the resulting mRNA is exported to the cytoplasm to produce the 16 viral protein required to further the life cycle of the virus (Figure 1.3, 3) (Li and De Clercq, 2016). Precursor viral proteins are expressed and cleaved to form functional viral proteins that can be assembled into a mature HIV-1 viral particle (Figure 1.2, 4). The gp160 molecule which undergoes post-translational glycosylation, a precursor to gp120 and gp41 are cleaved by host furin to generate the constituent protein products (Hallenberger et al., 1992). These gp120 and gp41 molecules remain bound by a noncovalent interaction. Other proteins such as the Gag-Pol polyproteins are cleaved by the viral protease. Mature viral proteins assemble under the host cell membrane in cholesterol rich lipid rafts, in which the gp120-gp41 proteins are embedded, to utilise the host membrane as the viral membrane (Figure 1.2, 4) (Ganser-Pornillos et al., 2012). An immature viral particle buds from the host membrane to circulate through the host (Figure 1.2, 4). The budding is facilitated by the viral Gag protein which interacts with the host machinery to export the viral particle (Morita et al., 2011). The final stage in the budding of the viral particle is through the cleavage of the Gag-Pol polypeptide responsible for the budding of the viral particle (Pettit et al., 2004). Following budding, the viral protease completes the cleavage of all the viral proteins to form a mature virion capable of another round of infection. The only exposed viral proteins that protrude from the host derived membrane are the gp120 and parts of the gp41 heterotrimers and are thus critical targets for neutralizing antibodies as well as a potential prophylactic vaccine.

1.4. The structure of gp120 and the role in CD4 binding

The investigation of monomeric gp120 through crystallographic studies, as well as through Molecular Dynamic (MD) simulations has provided great insight into the structure, dynamics, and

structural changes to key binding partners such as CD4 (Kwong et al., 1998; Pandey et al., 2019). Many of the crystallographic studies have been performed largely on modified gp120 molecules that have truncated variable loops and have undergone deglycosylation; important modifications necessary to crystallise gp120 for structural investigation through X-ray diffraction (Kwong et al., 1999). Studies analysing the structure of monomeric gp120 have identified an inner and an outer domain of the protein, denoted by their relation to the internal and external portions of the protein, where the outer domain is exposed to the surrounding environment (Figure 1.3). The inner and outer domains are connected by a four stranded β-sheet that bridges the two domains (Figure 1.3) (Kwong et al., 1998). Amino acid sequence analysis of virus isolates have shown that the gp120 consists of five variable (V1-V5) regions and five constant regions (C1-C5) (Kwong et al., 1998). Investigation into the variable region have yet to determine the conformations and flexibility of these loops, partially due to the immense sequence variation that can exist in these regions. The CD4 binding loop is a small depression that is essential in the binding to domain 1 of CD4 and subsequent viral entry (Kwong et al., 1998). This loop forms a small depression at the junction of the inner and outer domains, between the V2 and V3 loops, above the bridging sheet (Yoon et al., 2010).

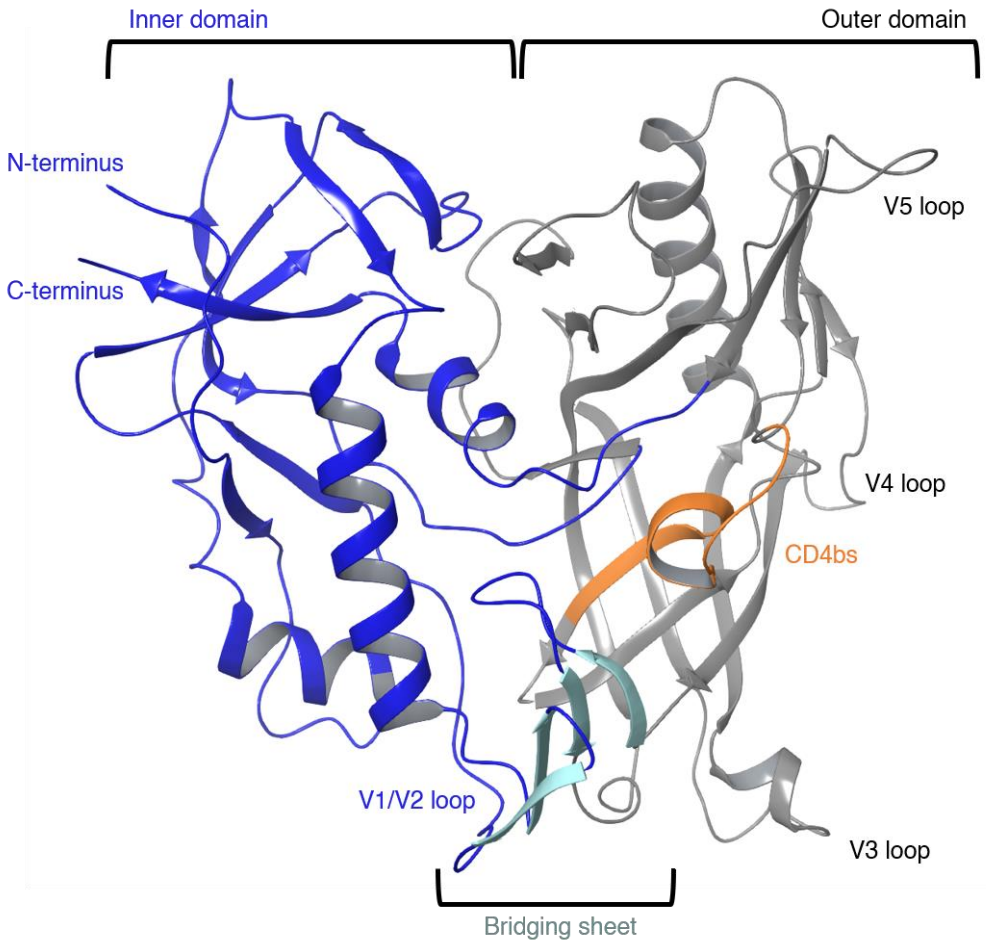

Figure 1.3: The core structure of gp120 depicting key structural elements of the gp120 protein. The gp120 is divided into an inner domain (blue) and outer domain (grey) based on the orientation in the Env trimeric spike. The orientation of the protein is such that the V1/V2 loop, bridging sheet, and V3 loop are distal from the viral membrane. This bridging sheet (light blue) between the two domains consists of four β-sheets that bridge the inner and outer domains. The CD4 binding site and visible loop are depicted in orange. This is the primary interaction site of gp120 with CD4. Variable loops are labelled for reference. The gp120 was generated from the published gp120 structure from *Kwon et al,* (2012) (PDB ID: 3TGR).

The gp120 appears to be extremely flexible and undergoes large conformational changes upon CD4 binding. It primarily takes either a "closed" conformation that is more tightly packed, or an "open" conformation characterized with outwardly rotated subunits (Liu et al., 2008). The structural investigations of gp120 indicate that it retains a more "closed" conformation when bound to antibodies that inhibit structural rearrangement such as the monoclonal antibody (MAb) VRC01 that binds the CD4 binding site of gp120 (Li et al., 2011), and a more "open" conformation when bound to CD4 or antibodies that recognise epitopes exposed upon CD4 binding. However, the only

successful account of crystallography on an unliganded gp120 structure indicates that it takes a similar "open" orientation to CD4 bound gp120 (Kwon et al., 2012). However, this could be due to the modifications employed to enable the unliganded crystallization of gp120. A further characteristic of HIV-1 Env, in particular the gp120 component is the extensive glycosylation of these proteins (Figure 1.4). N-linked glycans comprise over 50 % of the weight of the HIV-1 Env and has been referred to as a "glycan shield" due to its role in shielding neutralizing epitopes from the host immune response. These are a hurdle in both the development of vaccines, as well as determining the structure of the gp120 component due to the innate flexibility of the glycans producing conformational heterogeneity.

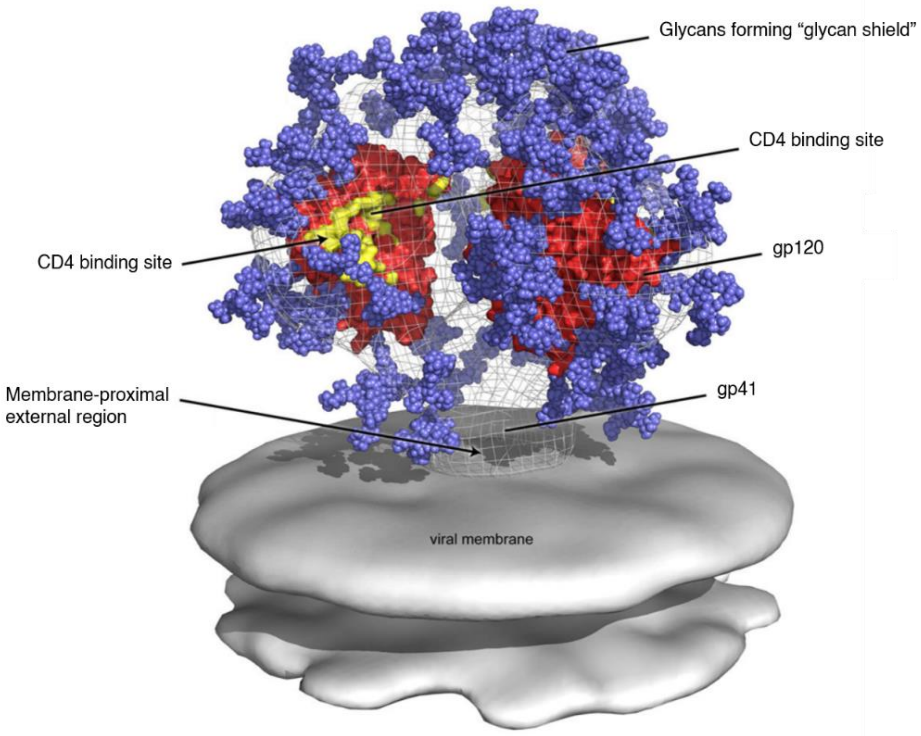

Figure 1.4: Diagram depicting the extensive glycosylation of the HIV-1 Env and the effect on the accessibility of the gp120 and gp41 components of the Env. Image adapted from (Burton et al., 2012).

The primary binding partner in HIV-1 infection is the CD4 receptor on host $CD4^+$ T cells that binds to gp120. CD4 is expressed on $CD4^+$ cells and primarily on T lymphocytes, that are

responsible for an integral role in the host adaptive immune response. It has also been shown to be expressed on other cells in the immune response such as monocytes, macrophages, and B lymphocytes to name a few (Kaleem et al., 2001; Kazazi et al., 1989; Zhen et al., 2014). Apart from the contextual receptor function for HIV-1, CD4 interacts with major histocompatibility complexes type II (MHC II) to stabilise T-cell receptor interactions (König et al., 1992; Li et al., 2013). These MHC II molecules are present on antigen presenting cells and induce signalling to T-cells.

The CD4 molecule is a 55 kDa type 1 integral membrane protein that consists of four extracellular domains. It consists of an extracellular region, a transmembrane portion, and a cytoplasmic tail. The first two domains (those integral to the interaction with HIV-1 gp120; 2dCD4) have been crystallised and characterised and reside in the extracellular domain of the protein (Kwong et al., 1998).

The gp120-CD4 interaction has been a large focus in the investigation of HIV-1 and has led to the production of many structural depictions and insights into the interaction. It has been well established that the gp120 molecule undergoes extensive conformational changes upon binding to CD4. The domain 1 of CD4 binds to the CD4 binding loop that resides above the bridging sheet between the inner and outer domains of gp120 (Figure 1.4). The portion of domain 1 in CD4 that contacts gp120 spans over residues 25-64 and forms a loop structure that protrudes from this domain (Kwong et al., 1998; Sharma et al., 2005). The regions of gp120 that contact CD4 are far more extensive and are spread over six regions of gp120, owed to its highly complex conformational structure (Sharma et al., 2005). The primary interactions are between Phe-43 and residues Glu-370 and Trp-427 of gp120 (Moebius et al., 1992). The second primary contact is the Arg-59 residue that bonds electrostatically to residue Asp-368 in gp120 (Moebius et al., 1992). MCH II interactions with CD4 are encompassed over the same region of residues that are integral to the gp120 interaction – where Phe-43 and Arg-59 of CD4 play equally integral roles (J. Wang et al., 2001).

The binding of gp120 to CD4 exposes the bridging sheet between the inner and outer domains of gp120 and stabilises this conformation. Cryo-EM of trimeric HIV-1 gp120 bound to CD4 indicates that the V1 and V2 loop undergo conformational rearrangement upon binding (Wang et al., 2016). It has also been observed that the V1, V2, and V3 loops prevent gp120 from adopting the more

"open" CD4 structure, as the deletion of these regions in an unliganded gp120 crystal structure shows that the more open conformation is adopted (Kwon et al., 2012). There is also evidence to suggest that the V3 loop orients closer to the co-receptor as this loop determines viral tropism, as well as the deletion of key residues on the tip of the loop prevents viral entry into host cells (Bagnarelli et al., 2003, p. 3; Xiang et al., 2010). MD simulations analysing the trajectories of CD4 bound gp120 compared to unliganded gp120 corroborate the large-scale reorientation of the V1, V2, and V3 loops (Li et al., 2020) as well as an increase in the mobility and conformational flexibility of these regions.

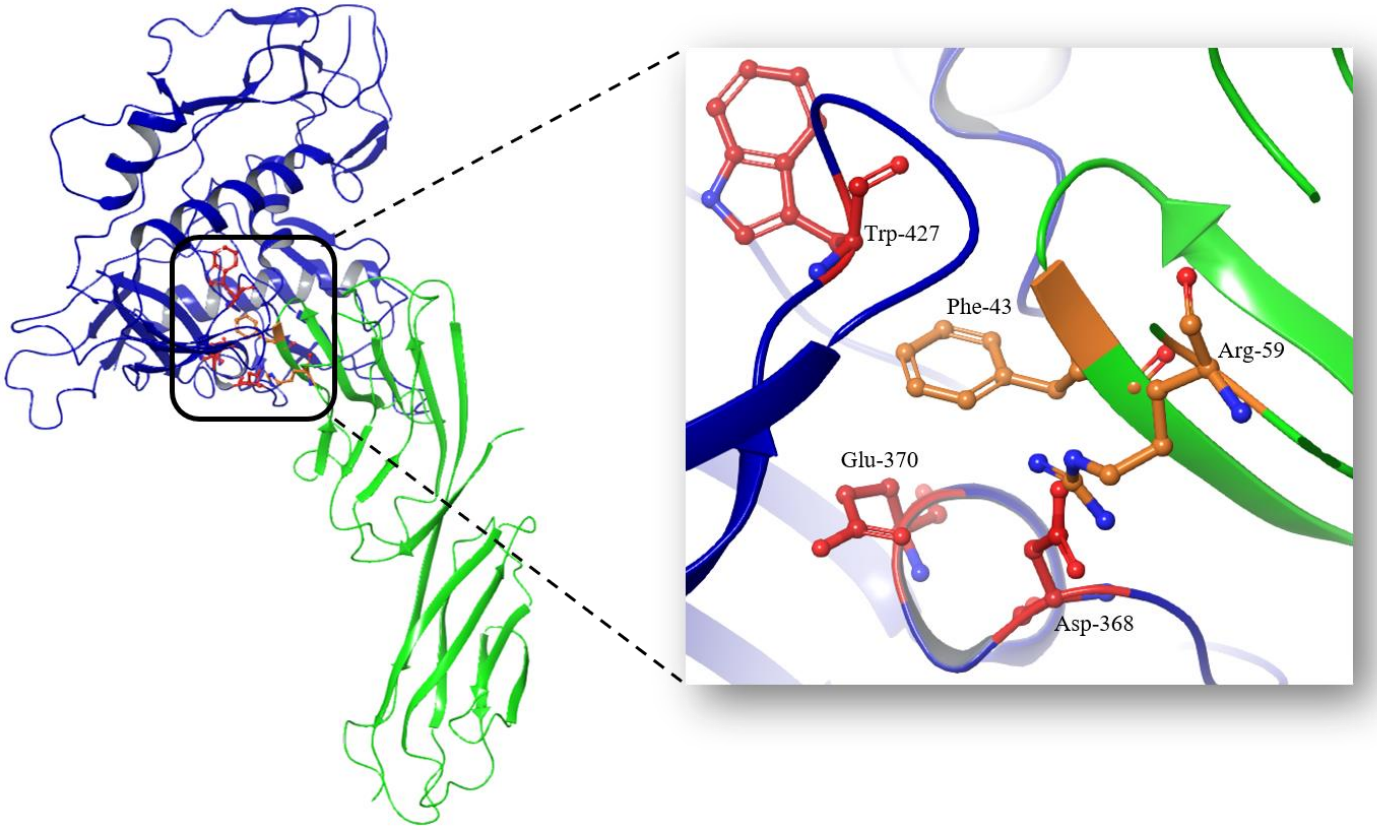

Figure 1.5: Depiction of gp120 (blue) binding to the first domain of two domain CD4 (green) with the contacts between. Key residues involved in binding are depicted for gp120 (red) and CD4 (orange). Arg-59 binds electrostatically to Asp-368 while the Phe-43 has hydrophobic interactions between Trp-427 and Glu-370 of gp120. These form the primary interactions between the gp120 and CD4 molecule as CD4 binds into the binding pocket of gp120. The image was generated using PDB: 6L1Y.

1.5. Prophylactic vaccines as a strategy for HIV-1 prevention

The principle of a prophylactic vaccine is the introduction of an immunogen to elicit an appropriate immune response, either cellular or humoral (antibody) immune responses that aim to target as wide a range of naturally occurring genetic variants of a pathogen as possible (Lelièvre and Lévy, 2016). For chronic infections such as HIV-1, cellular immune responses impact on viral replication and viral set point following infection, whereas appropriate antibody responses can neutralize virus thereby preventing transmission.

Antibodies would need to target highly conserved regions of the HIV-1 proteins that are less susceptible to immune escape, allowing them to neutralize a broad range of HIV-1 strains and prevent infection from the wide array of virus quasispecies that could be transmitted (Lelièvre and Lévy, 2016). Broadly neutralizing antibodies (bNAbs) develop naturally in HIV-1 infected individuals, but in only approximately 20% of these individuals (NISC Comparative Sequencing Program et al., 2013). These naturally occurring bNAbs exhibit varying degrees of breadth and potency and are susceptible to viral escape (Walker et al., 2011). An efficacious prophylactic vaccine against HIV would thus need to elicit a bNAb response, preferably with increased breadth and potency than the naturally occurring bNAbs.

In the context of HIV-1, vaccines that stimulate appropriate cellular immune responses, e.g., HIV specific CD8+ cytotoxic T lymphocyte responses, have proven ineffective in numerous human clinical trials (Reviewed in: (Koup and Douek, 2011). The use of monomeric gp120 to stimulate an antibody response has also failed in human clinical trials (Rerks-Ngarm et al., 2009); rgp120 (HIV Vaccine Study Group, 2005). Investigations of potentially trimeric gp120 mutagens have provided limited success in inducing bNAb responses and are able to display quaternary epitopes unavailable to monomeric gp120 (Jones et al., 2017). Based on evolving structural information of the Env, the understanding of its dynamic, trimeric nature and how the antibody response matures in infected individuals who develop broadly neutralizing antibodies (bNAbs), novel immunogen designs and vaccination regimens are being investigated.

1.5.1. The characteristics and use of a transmitted Founder Virus sequence for epitope targeting.

An HIV-1 founder virus (FV) is the virus quasispecies capable of establishing an infection in a new host, and an estimated 80% of heterosexual infections are established by a single FV (Keele et al., 2008). These are attractive targets for a prophylactic vaccine and preventing the establishment of an HIV-1 infection. Investigations of these founder viruses has shown that the gp120 components exhibit unique characteristics compared to viruses active during a chronic HIV-1 infection. Structurally FV gp120 has been observed to have shorter variable loops (V1-V4) and less genetic diversity amongst founder viruses of the same subtype than viruses involved in chronic infections (Go et al., 2011; Rademeyer et al., 2007). Studies have also shown that the glycosylation profile of gp120 FV for subtype B and C exhibit more similarities to each other, than to chronic viruses of the same subtype (Go et al., 2011). Further it has been determined that gp120 subtype C FV exhibit fewer N-linked glycosylation sites and consequently less glycosylation (Rademeyer et al., 2007). These characteristics, along with potentially undiscovered traits may provide insight into the infectious ability of FV variants, as well as potential strategies to combat infection.

There have been numerous attempts at generating consensus sequences of these viral sequences to identify common characteristics among the viral strains (Keele et al., 2008; Killick et al., 2014). The rationale for using such a gp120 founder virus consensus sequence in an immune strategy is to identify conserved characteristics in these viral proteins to direct an immune response to. This would allow for antibodies to potentially neutralise numerous species of HIV-1 that can establish an infection, a necessity for a prophylactic vaccine strategy. This study utilised an HIV-1 subtype C FV consensus sequence ($gp120_{FV}$) that was created using bioinformatic determination analysis of 1,894 subtype C founder viruses (Killick et al., 2014). HIV-1 subtype C is the predominant circulating subtype in South Africa (Bbosa et al., 2019; Musyoki et al., 2012).

1.6. Generation of a potent bNAb response through a novel $gp120_{FV}$-2dCD4^{S60C} complex

Various strategies have been employed in recent years that focus on the introduction of various Env-based immunogens to elicit the production of bNAbs against HIV-1. These include the

introduction of gp120 Env molecules (Georgiev et al., 2013), monomeric gp120–2dCD4 complexes (Kang et al., 1994) and modified soluble HIV-1 SOSIP trimers as possible immunogens (Torrents de la Peña et al., 2018). However, these have shown limited success in the generation of bNAbs following immunization of small animal models, non-human primates, and in human clinical trials in some instances (Kang et al., 1994; Rerks-Ngarm et al., 2009; Schwartz et al., 2016; Yang et al., 2004). An additional strategy that has been utilized is the stabilization of the gp120–CD4 complex, a method that has shown significant success (Cerutti *et al.*, 2010; Thompson *et al.*, 2015).

Our laboratory has focused on testing the immunogenicity of a novel gp120–CD4 complex in small animals (Killick et al., 2015) and non-human primates (unpublished data). The immunogen design is based on the use of an HIV-1 subtype C FV sequence ($gp120_{FVC}$). It is important to use a consensus sequence to account for the large range of genetic diversity and to expose conserved sequences and structures throughout those founder viruses. The second modification is a mutation in 2dCD4 through the substitution of serine at location 60 on the CD4 α-helix 1 with cysteine to create $2dCD4^{S60C}$ (Cerutti et al., 2010b). This substitution attempts to create a targeted disulphide bond between gp120 and CD4 to stabilize the complex. This is achieved by utilizing an existing disulphide bond found on gp120 between C126-C196. The introduction of the S60C mutation shifts the disulphide bond resulting in its formation between C60 on CD4 and C126 on gp120 (Cerutti et al., 2010b) as depicted in Figure 1.6.

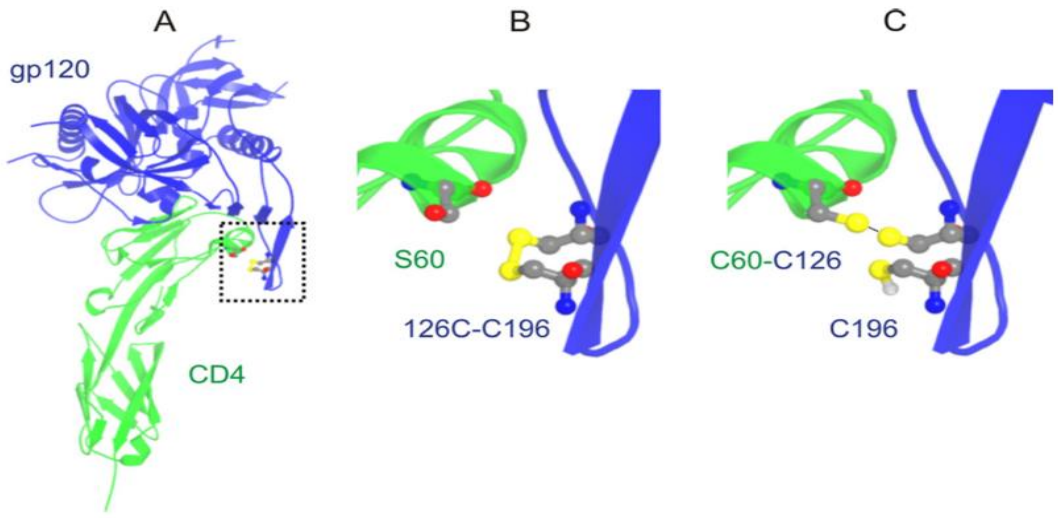

Figure 1.6: Molecular model depicting the binding interaction of the gp120(core)- 2dCD4^{S60C} complex. Model B indicates the S60C modification on CD4 (green) and the native Cys126–Cys196 disulphide bond on gp120 (blue). The novel disulphide bond after targeted disulphide exchange is present on Model C between C60 on CD4 and C126 on gp120 (adapted from (Cerutti et al., 2010b)**.**

This 2dCD4^{S60C} protein has been shown to bind gp120 at a much higher affinity than native 2dCD4 (2dCD4^{WT}), and ultimately confers an increased immunogenic effect (Cerutti *et al*., 2010; Killick *et al*., 2015). This immunogenic effect induces a bNAb response capable of generating a neutralizing response against tier 1, and clinically relevant tier 2 and 3 pseudovirus strains in small mammals (rabbits) (Killick *et al*., 2015) and non-human primates (unpublished data).

The mechanisms responsible for the enhanced immunogenicity of the gp120$_{FVC}$ExC – 2dCD4^{S60C} are not fully understood but could include the following (but not limited to): exposure of novel epitopes induced by the disulphide bond, pronounced exposure of well-known conserved epitopes, and/or increased stability or decreased dynamics of epitope conformations that allow for a bNAb response to be generated. It is critical to start understanding the mechanisms involved in the enhanced immunogenicity, and to start to determine whether structural changes induced by the disulphide bond of the S60C complex compared to the wild type CD4 complex contribute to its enhanced immunogenicity.

1.7. Experimental and theoretical determination of HIV-1 Env protein structure

1.7.1. Protein crystallization and modifications of gp120 for crystal formation

The gold standard for protein structure determination is protein crystallization and X-ray crystallography due to the ability to generate high resolution structures. Crystallization of proteins is notoriously difficult and requires high purity and homogeneity to be successful. As such, multiple techniques have been established to increase the probability of successful crystallization of the HIV-1 Env. A common approach in crystallization is the use of antibodies and ligands to stabilize the structures, reducing protein conformational flexibility to improve the probability of crystal formation (Kwong *et al*., 1998). The crystallization of HIV-1 Env has also been optimized using various techniques. The first is the usage of the monomeric form of Env gp120, as the native trimeric Env form does not readily create crystals due to its size, complexity, and dynamic nature. This is also evident as the gp41 component is a transmembrane protein and is extremely challenging to crystallize. Monomeric gp120 nonetheless contains the CD4 binding site and as such still provides the necessary information regarding the interactions between gp120 and CD4 (Kwong *et al*., 1999), and is thus useful and relevant in elucidating novel Env-2dCD4^{S60C} interactions. The second is the use of sequence adaptations that remove mobile native loop structures that inhibit crystallization, and generate a non-native loop to aid in stabilizing and improving crystal formation, known as gp120Core$_e$ and gp120Core$_{min}$ (Kwon et al., 2012). The final modification is the deglycosylation of gp120 to create homogeneity and reduce the dynamic nature of the glycoprotein by removing glycan groups that have been shown to inhibit crystallization (Kwong *et al*., 1998).

1.7.2. Molecular Dynamic Simulations

The study of protein structures, and by extension their function has revolutionized and enhanced our understanding of protein biochemistry. This progress has been due most notably to the advent of structural analytical techniques such as X-ray crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy that have been implemented since the 1950s. However, these techniques only provide static images of proteins which are inherently dynamic in nature. These images have been

invaluable to study proteins but are limited in the resolution they can capture and the protein dynamics they can investigate. These limitations have in part led to the development of techniques that can more effectively investigate the dynamics of proteins. One such technique is an *in silico* computational method known as Molecular Dynamic (MD) simulation.

MD simulation is a form of Molecular Modelling (MM) that is used mostly computationally to model the behaviour of atoms in a system over a set time period (McCammon and Harvey, 1988). Briefly, the atoms are assigned both coordinates to indicate their position, and velocities calculated from the Newtonian laws of motion. These are used to calculate the incremental change in position of atoms in the system over discrete time steps (McCammon and Harvey, 1988). For protein simulations this technique requires an experimentally derived protein structure (such as those from X-ray crystallography or NMR spectroscopy) to act as a template for simulating the protein dynamics (Ji et al., 2012; Shahlaei et al., 2011). The calculations that govern these systems are relatively simplistic as they are based of the Newtonian laws of motion and require the use of mechanical analogues to represent the atoms and their interactions (Allen, 1989; Hollingsworth and Dror, 2018). For example, springs are used to model the bonds of the system and atoms are represented by singular point charges (Allen, 1989).

MD simulation was first performed on proteins in the 1970's on a small bovine pancreatic trypsin inhibitor which had a well resolved crystal structure (McCammon et al., 1977). The first simulation was performed on the singular protein in a vacuum, at an extremely short picosecond simulation length (McCammon et al., 1977). Despite the computational limitations of the time, MD simulation has sustained as a technique and since covered an immense amount of ground in terms of the simulations that are possible and routinely performed. These improvements have been brought on largely by the increase in widely accessible computational power and improvements made in the implementation of the software to do more with less. The field has progressed from simulating systems of a mere hundreds of atoms to routinely simulating system of 50 000 – 100 000 atoms that allow for the investigation of systems with larger complexity that are more biologically relevant than ever before (Hospital et al., 2015). It is also possible to perform longer simulations with more incremental time-steps that can simulate at biologically relevant timescales. These improvements have led to improved investigation of enzymatic activity (Spiwok et al.,

2007), membrane protein complexes (Kandt et al., 2007), and other conformational changes associated with protein function (Lee et al., 2015).

There are a few variations of MD simulations that are routinely performed that will be discussed briefly. First, the nature of the calculations used divide MD simulations into classical MD simulations that use solely classical Newtonian laws of motion to calculate the forces of the atoms in the system (Boldon et al., 2015), and Quantum MD simulations (QMDS) that include the addition of Schrodinger's uncertainty principles (Boldon et al., 2015). Quantum MD has materialized in recent years due to the enormous increase in computational power available. The key benefit to this technique is the ability to simulate the breaking and formation of covalent bonds by including the uncertainty principles that are also known to contribute to the behaviour and properties of these interactions, but that are not accounted for in traditional Newtonian MD simulation (Hollingsworth and Dror, 2018). While this additional bond simulation in QMDS allows for more accurate simulations for protein interactions it is computationally expensive. This makes employing QMDS for simulating reasonable biological scales unfeasible using commonly accessible computing power and is thus only performed in a small number of specialized facilitates with access to supercomputers. Classical MD is far more accessible and common for MD simulations and is thus more applicable to the current experimental landscape and is discussed in further detail in section 1.7.2.1.

The second way of dividing these simulations is through the detail at an atomic level (referred to as resolution) that is modelled – dividing them into atomistic and coarse-grained models. These models differ solely on the level of detail and resolution that is simulated. Atomistic models simulate the individual atoms of the system and their interactions (Boldon et al., 2015). In contrast, coarse models treat atoms as approximate groups to minimize the number of atoms simulated. The decision behind which model to utilize is based on the complexity of the system, the computational power available, and the required length of the simulation. Systems that require a longer simulation time (longer biological processes), have more atoms due to large structures, or require more time steps to evaluate intermediate states will increase the computational and time cost of running a simulation, often making it unfeasible to use an atomistic model. However, wherever possible the atomistic model is used due to the increased accuracy and resolution of the model being superior.

1.7.2.1. Classic MD simulation and Force-Fields

MD simulations use both classical Newtonian mechanics and so-called "force-fields" to calculate the positions of the atoms and interactions of the simulation. The force-field used is the primary contributor to the calculations and represents the parameters used to determine the potential energy of the system (Guvench and MacKerell, 2008). Parameters are the characteristics used to define the system – i.e., they define the characteristics of the atoms and the interactions between them (Guvench and MacKerell, 2008). This potential energy calculation (*Eq. 1*) is a mathematical expression of the internal energy of the system, which is equal to the sum of all the potential and kinetic energies of the system. The potential energy calculations include the energy of the bonded and non-bonded interactions, represented by the covalent and long-range van der Waals and electrostatic forces. However, as mentioned, due to the nature of the equations and calculations used in classical MD simulations, they are unable to calculate the formation and breakage of covalent bonds (Hollingsworth and Dror, 2018). It is important to note that as atoms are represented as fixed charge particles, their charge remains constant throughout the simulation instead of fluctuating based on their environment (Riniker, 2018).

$E_{total} = E_{bonded} + E_{nonbonded}$: which comprises the following:

$E_{bonded} = E_{bond} + E_{angle} + E_{dihedral}$

$E_{nonbonded} = E_{electrostatic} + E_{van\ der\ Waals}$

Eq. 1: Potential energy function used to calculate the potential energy of the simulation environment.

The force-field and the calculations approximate the mechanics that describe the atoms and their interactions. The atoms are modelled as fixed-charged spheres that represent a single point where the spheres are allowed to overlap to represent atomic interaction (Riniker, 2018). These atomic interactions are calculated using the criteria for the so called "Lennard-Jones" potential of charged particles (Riniker, 2018). Springs act as the analogues for the bonds, angles, and the torsion of the angles (González, 2011). In all dynamic simulations that include kinetics, the atomic velocities are calculated using the Newtonian laws of motion (F = m × a) to determine the movement and

sequential position of the atoms in the system. The positions of the atoms are represented as coordinates that change based on the time-steps of the simulation - these are used to generate the dynamic representation of the simulation. These time-step snapshots are parsed together to create a "movie" of the atomic movement.

The parameters of the atoms vary between the individual force-fields that are used and are either calculated, experimentally derived, or both (González, 2011). This makes force-fields unique to the atoms they were validated on. The parameters of the force-fields often apply to every individual atom of the system, whereas coarse model force-fields will cluster the atoms into groups and characterize them as a unit (Nielsen et al., 2004). This reduces the computational requirements at the cost of accuracy. The experimental validation of the parameters for macromolecular structures such as proteins are often determined by smaller organic molecules that are easier to parameterize (Guvench and MacKerell, 2008). A great deal of atoms have been parameterized but is by no means exhaustive – should a simulation require parameters that have not yet been established they will have to be determined prior to running the simulation as opposed to using an existing force-field. In protein MD simulations there are numerous established and validated force-fields to choose from that perform similarly, apart from a few context specific applications. These differ based on the mechanisms, calculations, and validations that have been used to establish the parameters of the atoms that will be simulated. These fields have been expanded, validated, and improved over the years to increase their accuracy.

1.7.2.2. Preparation of the simulation

To perform MD simulations, both the structure and the environment in which the simulation is performed need to be prepared. This preparation follows a standard procedure that varies in its specifics based on the nature of the experiment and the software used to run the simulation. Most suites have integrated tools to prepare the structures and environment for simulation that has greatly increased the accessibility of performing MD simulations.

Homology models are frequently generated as the first step in the course of performing MD simulations. These models fulfil a dual purpose in the preparation of the proteins for MD

simulations. The first is the ability to generate a model of a protein sequence for which the experimental structure has not yet been determined (Ji et al., 2012; Muhammed and Aki-Yalcin, 2019). The second function is the correction of artefacts of crystallization and correcting for certain residues that have not been appropriately resolved through experimental methods such as X-ray crystallography (Muhammed and Aki-Yalcin, 2019). These homology models are generated by constructing a model based on a protein sequence while using a homologous protein template. The initial structure that is derived from experimental techniques, such as NMR and X-ray crystallography, acts as a base for the generated homology model. These structures act as a base template in conjunction with the desired sequence to be modelled to generate a model free from artefacts of crystallization.

To run an MD simulation, both the protein and the simulation environment need to be prepared. The procedure for these preparations follows a standard workflow. This preparation starts with the PDB file of the structure or homology model. This file is modified to remove any artifacts of crystallization such as the non-native crystal contacts that occur between the protein molecules in the crystal, set the correct orientation of the bonds, and introduce the hydrogen atoms that are not detectable during techniques such as X-ray crystallography (Muhammed and Aki-Yalcin, 2019). Missing side chains and residues are also added to complete the structure. Disulphide bonds are also detected and modelled based on the topography of the molecule. There are various recommendations around the act of removing water molecules from the crystallization, and this seems to be either software or experimentally specific. However, it is recommended that external water molecules are removed, but ones internal to the protein or in contact with the protein should be left in the file as they add valuable conformational information to the protein. The protein structure can then be optimized by correcting the orientation of the added hydrogen bonds and optimizing the existing bond orientations of the protein. Once this preparation has been completed the simulation environment can be generated and the molecule added. The system consists initially of boundaries around the protein referred to as the unit cell – these boundaries are a box of sorts and can take various shapes based on what is required or best suits the shape of the modelled protein (Gajula et al., 2016). When deciding on a shape and size of the unit cell, it is important that it is large enough to accommodate for fluctuations of the protein to prevent it from leaving the simulation boundary. It is important to choose a unit cell that is also not too large as the system will increase in complexity due to the added water molecules and increase computational cost and

increase the time required to run the simulation. The unit-cell around the protein is then solvated – this is the introduction of water molecules or solvent that the protein will be simulated in, which is traditionally a mimic of its physiological environment (Kandt et al., 2007). Following this, the system needs to be neutralized by adding counter ions to either reach a neutral charge or a desired charge for the experiment (Gajula et al., 2016). Finally, the structure is relaxed through an energy minimization step. This removes any steric hindrance that may have been introduced in the preparation of the file and aims to reduce the overall energy landscape of the protein (Kandt et al., 2007). There are numerous ways to perform energy minimization which differ from software to software such as the steepest descent method that, like the other methods attempts to achieve the nearest local energy minimum (Gajula et al., 2016). The last step prior to running the simulation is the selection of a force-field. Depending on the software used this could already be introduced in the preparation phase (Gajula et al., 2016).

There are multiple software options available to choose from when deciding to run MD simulations. Many of these are open source and free to the public – like GROMACS (Van Der Spoel et al., 2005), CHARMM (Brooks et al., 2009), and AMBER (Case et al., 2005). There are also many proprietary software suites that can provide an easier user experience with a less steep learning curve, of which the Schrodinger package is one of the most popular examples. The calculations used by the simulation software are similar and many can accommodate most of the popular force-fields available. This makes the choice of software interchangeable, particularly in terms of simulation accuracy. However, the software is often optimized to utilize certain force-fields and operate on certain hardware configurations (CPU or GPU based).

1.7.2.3. Applications and Limitations of MD simulations

1.7.2.3.1. Applications

As an analytical technique MD simulation has provided unparallel capability to investigate and analyse the dynamics of molecular structures at an atomic level. It has been particularly revolutionary where proteins are concerned by directing further experimental approaches such as the investigation of structure modifications in binding assays and creating representations of proteins that may not be experimentally possible – such as high temperature experiments or those

that happen in an extremely short timeframe (Settanni and Fersht, 2008). This technique allows for the observation of transition states as proteins move between conformations (Lee et al., 2015). These simulations have been used quite extensively to tackle problems such as protein folding prediction and *ab initio* prediction of protein structure (Cheung and Yu, 2018; Kuhlman and Bradley, 2019). This has led to improvements in the capabilities of predicting the quaternary structure of a protein based on the amino acid sequence. This is particularly important in the field of rational protein design by allowing us to create proteins with certain conformations to produce or enhance a therapeutic effect such as drugs for cancer and rheumatoid arthritis (Singh et al., 2019) as well as rational drug design as potential therapeutics (Do et al., 2018). MD simulation has also been implemented to investigate the biological activity of proteins in term of enzyme catalysis (Spiwok et al., 2007) and protein-protein interactions (Gupta et al., 2016).

1.7.2.3.2. Limitations

There are limitations intrinsic to the nature of running a computational model. The most pertinent is the computational limitation imposed by the hardware utilized to run these simulations, creating not only an accessibility issue of gaining access to sophisticated hardware but also limiting the complexity of the simulations. While there are institutions that have supercomputer capabilities, even these are limited in the complexity and length of simulations they can run. Importantly the length of the individual timesteps in the simulation are determined by the fastest moving atom in the system (Tuckerman and Berne, 1991) – that is to say that the length of the timesteps are limited to a certain time which may increase the number of steps in a simulation. The amount of timesteps for a simulation that is of biological relevance may cause it to surpass the computational capabilities of the hardware.

The nature of calculating the atomic forces using the Newtonian laws of motion requires the integration of the calculations. The act of mathematical integration of these algorithms introduces numerical errors in the calculations due to inherent uncertainty in these calculations (Wang et al., 2001). These errors compound over the length of the simulation, decreasing the accuracy (Kim, 2014).

1.7.4. MD simulation and its use in the investigation of HIV-1

1.7.4.1. HIV-1 gp120–CD4 complex MD Simulation for conformation investigation

Through traditional imaging techniques such as X-ray crystallography it has long since been established that gp120 undergoes distinct conformational changes upon CD4 binding (Li et al., 2020). However, the dynamics, intermediary conformations, and implications of the conformational changes have remained elusive with more to uncover. Using MD simulation for the comparison modelling of gp120 and the gp120–CD4 complex the understanding of the conformational dynamics of these proteins and their interactions with proteins have been expanded. It has been found that CD4 binding reduces conformational flexibility of gp120, increases stability, and prevents it from moving out of its ligand bound conformation (Li et al., 2020). As a poignant example of how these models can translate into therapeutic strategies the authors of Li et al., (2020) propose restraining the loops responsible from preventing reorientation through small molecules to prevent conformational changes upon CD4 binding based on results from MD analysis. Many of the models are used to investigate potential therapeutic mechanisms to treat HIV-1 infections. Another such study utilized mutated CD4 proteins in complex with gp120 to determine potential drug locations and identify characteristics that these ligands would require to be effective (Chong Teoh et al., 2011). There are numerous studies identifying highly conserved or accessible regions of gp120 as potential sites for entry inhibitor binding – preventing viral entry into host cells (Pandey et al., 2019). MD simulation utility becomes astonishingly apparent in its ease to investigate characteristics of proteins that would be extremely challenging using traditional techniques like X-ray crystallography. This utility becomes apparent for proteins as those in HIV-1 through the investigation of the glycans on gp120. Gp120 is a highly glycosylated protein for which the glycans are often removed prior to crystallization to increase the probability of successful crystal formation. It has also been established that these glycans play a large role in the binding of broadly neutralizing antibodies (bNAbs) – they are often necessary for binding and mutations can lead to resistance to bNAbs (Townsley et al., 2016). Models have been used to elucidate the function of various surface glycans and potentially direct therapeutic strategies. Some glycans have been shown to enhance gp120-CD4 complex formation and gp120-antibody associations (Zhang et al., 2017). All these models and strategies culminate in increased knowledge of protein and structural function – bringing new avenues to light for potential

treatment. Pertinently there have been many MD simulation studies that have investigated the structural changes that occur with the binding of gp120 to CD4. These changes have found that primarily the V1, V2, and V3 regions are involved in the reorientation that gp120 undergoes when binding CD4. Primarily the dynamics of the gp120 reduce when bound to CD4 as the V1V2 region is prevented from moving towards the core of gp120. The stabilization of V1V2 prevents reorientation and subsequent movement out of the CD4 bound state (Kwon et al., 2012b).

1.7.4.2. MD Simulation investigation of therapeutic intervention strategies of HIV-1

For drug design and therapeutic investigation MD simulation has proven invaluable for understanding mechanisms of resistance to existing therapeutics, the investigation of novel drugs, and to increase the understanding of existing therapeutic functions. An avenue for investigation is the treatment through drugs such as ARV's. Modelling has been used to not only identify sites in which these drugs can bind (Badaya and Sasidhar, 2020) but also how mutations can drastically increase resistance to existing ARV therapies (Chitongo et al., 2020; Moonsamy et al., 2016; Nizami et al., 2016).

Investigating bNAbs for HIV-1 treatment and prevention is another strategy that is actively being studied through MD analysis, both in terms of investigating the mutations that confer resistance, as well as identifying the conserved regions that would make effective targets to achieve sufficient breadth and potency. Insights have also been provided into the effect of mutations into the sequence of g120 and the resulting changes in conformation of the variable loop structures that result in altered steric hindrance to the exposure of immunogenic epitopes (Hollingsworth et al., 2018; Ovchinnikov et al., 2018). Investigation of the glycan shield through modifications and the resulting exposure of epitopes have been investigated as well in an attempt to identify epitopes that may lead to the exposure of novel epitopes to elicit host antibody responses (Mengistu et al., 2017). The investigation of drugs, small particle inhibitors, and the effect of bNAbs against gp120 have all been investigated through MD simulations (Zhang et al., 2012). The investigation of bNAbs and potential neutralizing epitopes have been distinctly alluring due to the cross clade neutralization of HIV-1 and as such identifying sequences or structures to elicit this response

without an HIV-1 infection has been posited as a strategy for prophylactic HIV-1 vaccines (Alberto et al., 2020).

Overall, the implementation of MD simulations in the analysis of a core gp120 FV will provide critical insights into the structural and conformational dynamics of the protein, as compared to a similarly modified HIV-1 subtype C gp120 ($gp120Core_e$). Furthermore, MD simulations can contribute to the understanding of the potential mechanisms and critical residues responsible for the enhanced immunogenicity of Env–$2dC4^{S60C}$ complexes. Furthermore, results from the MD simulation can provide supplementing information to X-ray crystallography structures of various Env and Env-$2dCD4^{S60C}$ complexes (based on FV sequences). Thus, it is critical to design an HIV-1 subtype C Core gp120 construct based on FV sequences, with modifications to allow for crystallization, and to perform MD simulations on this $2dCD4^{S60C}$ liganded and unliganded protein and appropriate controls.

1.8. Aims and Objectives

The overall aim of the study was to crystallize and computationally characterize recombinant HIV-1 subtype C core envelope glycoproteins based on a consensus transmitted FV sequence ($gp120_{FVC}ExC$) and a modified clinical isolate sequence ($gp120Core_e$), to gain insights into structural dynamics of transmitted founder viruses, as well as mechanisms involved in the enhanced immunogenicity of a $gp120_{FVC}ExC$ – $2dCD4^{S60C}$ vaccine immunogen resulting from the introduced putative disulphide bond.

The objectives of the study were:

- To design a recombinant modified $gp120_{FVC}$ extended core ($gp120_{FVC}ExC$) protein, express it in mammalian cell lines, and purify it to homogeneity.
- To determine the optimal conditions that allow for the crystallization of a deglycosylated $gp120_{FVC}ExC$ protein.
- To conduct MD simulations to compare the $gp120_{FVC}ExC$, $gp120Core_e$, $2dCD4^{WT}$, $2dCD4^{S60C}$, $gp120_{FVC}ExC–2dCD4^{WT}$ and $gp120_{FVC}ExC–2dCD4^{S60C}$ complexes to identify novel structural characteristics and dynamics associated with transmitted founder viruses, as well as the enhanced immunogenicity of $gp120_{FVC}ExC–2dCD4^{S60C}$ complexes.

2. Methods

2.1.Reagents and sequences used in this study

The amino acid sequences used to generate the expression constructs, and in the MD simulation study are shown below:

2dCD4^{WT}: The sequence was generated from the first two amino-terminal domains of human CD4 and used in the generation of homology models in conjunction with the X-ray diffraction structure of 2dCD4 (PDB ID: 2NY4; Figure 2.1).

2dCD4^{WT}

MKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNL
KIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQL
ELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKLE

Figure 2.1: Sequence of 2dCD4^{WT} generated from the first two amino-terminal domains of human CD4.

2dCD4^{S60C}: The sequence for 2dCD4^{WT} was used with a Serine to Cysteine substitution at position 60. The sequence was used to generate a homology model in conjunction with the X-ray diffraction structure of 2dCD4 (PDB ID: 2NY4; Figure 2.2).

2dCD4^{S60C}

MKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRCLWDQGNFPLIIKNL
KIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQL
ELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKLEHHHHHH

Figure 2.2: Sequence of 2dCD4^{S60C} generated from the first two amino-terminal domains of human CD4 with a substitution of Serine to Cysteine at position 60 (indicated in red)

gp120Core$_{min}$ and gp120Core$_{e}$: These sequences were obtained from *Kwon et al*, (2012) where the modifications of the V1/V2, and V3 loops were used to modify our gp120$_{FVC}$ sequence.

gp120$_{FVC}$: HIV-1 gp120 founder virus consensus sequence was obtained from Killick et al., (2014)

gp120$_{FVC}$ExC: This sequence was generated by adapting the V1/V2, and V3 loops of the gp120$_{FVC}$ExC sequence (Killick et al., 2014). The V1/V2 -PLCVGAGCNT- sequence adaptation was

taken from gp120Core$_{min}$ (Kwon et al., 2012), while the V3 -GGSGSG- sequence modification was taken from gp120Core$_e$. This ultimately resulted in our gp120$_{FVC}$ExC sequence, depicted in Figure 2.3 below.

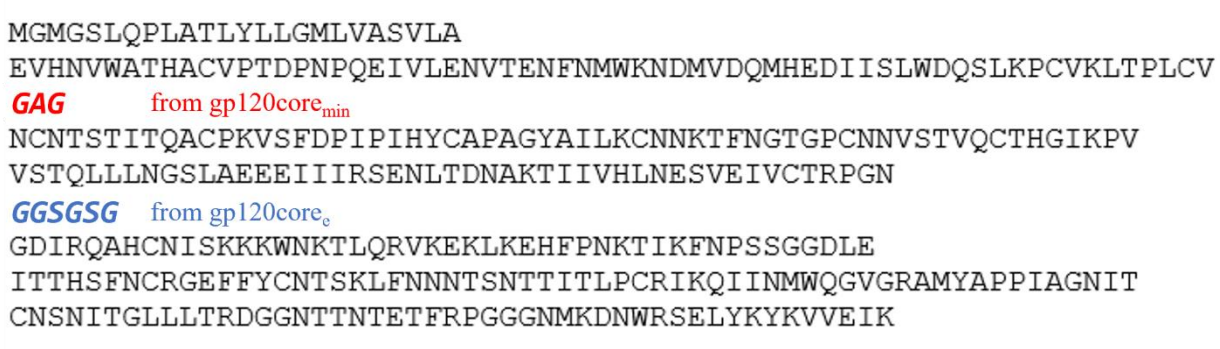

Figure 2.3: Sequence of gp120$_{FVC}$ExC generated from gp120$_{FVC}$ (Killick et al., 2014). The GAG modification was adapted from gp120Core$_{min}$ as indicated by the -GAG- sequence (red). The V3 modification was adapted from gp120Core$_e$ and is indicated with the -GGSGSG- region (blue).

The pDSgp120$_{FVC}$ExC mammalian expression plasmid was generated by codon optimizing (humanizing) the gp120$_{FVC}$ExC sequence and synthesizing by GeneArt and cloning it into the pcDNA 3.3 – TOPO plasmid (Invitrogen). The plasmid contains a leader peptide sequence for protein extracellular export of the protein and Geneticin resistance for selection of stably transfected mammalian cells.

Escherichia (E.) coli DH5α bacterial cells were obtained from Invitrogen, Life Technologies to allow for the transformation, propagation and purification of the pDSgp120$_{FVC}$ExC plasmid, and subsequent transfection of the purified recombinant plasmid into mammalian cells for gp120$_{FVC}$ExC protein expression.

The HEK293S GnTi$^-$ mammalian cells were obtained from ThermoFischer Scientific and were used for the expression of gp120$_{FVC}$ExC as the lack of N-acetylglucosaminyltransferase I prevents the formation of complex glycans, allowing for more uniform deglycosylation.

2.2.Design, recombinant protein expression and purification of gp120$_{FVC}$ExC

2.2.1. Design of a modified gp120$_{FVC}$ExC sequence

The crystallization of unliganded gp120 has proven to be notoriously difficult (Kwong et al., 1998). However, *Kwon et al*. (2012) has successfully utilized two primary modifications that allow for the crystallization of unliganded gp120 that they have coined gp120Core$_e$. These modifications have adopted into the FV sequence to increase the probability of crystallizing the FV sequence. The two modifications employed by *Kwon et al.* are aimed at reducing heterogeneity and substituting the V1, V2, and V3 regions. While *Kwon et al.* replaced the V1 and V2 regions with a -GG- sequence, the V1/V2 sequence from gp120Core$_{min}$ was adopted, allowing for the inclusion of residues from the V1 region necessary for the formation of the putative disulphide bond in our immunogenic complex. The HIV-1 subtype C FV sequence described by Killick et al., (2014) was used as the base for the gp120$_{FVC}$ExC construct. The V1, V2 region consists of a -PLCVGAGCNT- sequence, which was obtained from gp120Core$_{min}$ and the V3 region was modified identically to gp120Core$_e$ with the introduction of -GGSGSG- to replace a portion of the V3 region.

The pDSgp120$_{FVC}$ExC plasmid was used to transform chemically competent *E. coli* DH5α cells at a ratio of 200 µg DNA to 20 µl of DH5α cells and left on ice for 30 min. The cells were heat shocked at 42°C for 42 seconds and placed on ice for 2 minutes. The cells were spread on 1.5% agar plates supplemented with 100 µg/ml Ampicillin (Roche) and left to incubate overnight at 37°C. A single selected colony was inoculated into 10 ml of Lysogeny Broth containing ampicillin (1 g Tryptone, 0.5 g Yeast Extract, 1 g NaCl, 100 µg/ml ampicillin) and left for 8 hours at 37°C. The inoculated media was transferred to 100 ml autoclaved Lysogeny Broth containing ampicillin and left overnight at 37°C at 250 orbitals per minute.

The Qiagen Maxiprep DNA extraction kit and protocol were used to extract the pDSgp120$_{FVC}$ExC plasmid. In brief, the culture media was centrifuged for 15 minutes at 6000 × g at 4°C and the supernatant removed. The bacterial pellet was resuspended in cooled buffer P1 and buffer P2 was added to the resuspension and left to incubate at room temperature for 5 minutes. Cooled buffer P3 was added to the solution, mixed by inversion, and left to incubate on ice for 20 minutes. The solution was centrifuged at 14 000 × g for 30 min at 4°C. Qiagen-tip 500 column was equilibrated

with buffer QBT and allowed to gravity drain, after which the supernatant was added to the column and allowed to flow through. The column was washed twice with buffer QC prior to the DNA being eluted off the column with buffer QF. Once eluted the DNA was precipitated using isopropanol and centrifuged at 14 000 × g for 30 min at 4°C. The supernatant was poured off and resuspended in 70 % ethanol prior to another round of centrifugation at 14 000 × g for 5 min at 4°C. The supernatant was removed, and the pellet was left to air dry prior to being resuspended in RNAse free water. The DNA concentrations were measured using a Nanodrop™ (ThermoFischer) at a 260/280 nm ratio and stored at -20°C.

2.2.2. Expression of $gp120_{FVC}ExC$

The expression of $gp120_{FVC}ExC$ was performed in a HEK293S GnTi$^-$ deficient cell line to prevent the formation of complex glycans, allowing for homogenous downstream deglycosylation. The generation of a stably transfected cell line was performed to allow for continuous expression of the $gp120_{FVC}ExC$ protein. Cell stock of stably transfected HEK293S GnTi$^-$ cells were already available in our laboratory. To generate these cells in brief, T25 flat bottom culture flasks were seeded with HEK293S GnTi$^-$ cells with 1×10^6 cells and 5 ml of Ham's F-12 complete media (10% FCS). Cells were transfected for flasks that exhibited approximately 70-80% confluency. The transfection solution was generated by mixing 5 ml Opti-MEM (Gibco, ThermoFischer) with 4 ug pDS$gp120_{FVC}ExC$ plasmid DNA and 12 µl PEI Max (1 mg/ml; Polysciences) per flask and vortexed discontinuously for 10 seconds and left to incubate at room temperature for 15 minutes. On the day of transfection, the culture media was removed, and the cells were incubated overnight with the transfection solution. The following morning the transfection solution was aspirated from the cells and 5 ml of Ham's F-12 complete media (10% FCS) was added to the cells. For selection 100 µg/ml of Geneticin (Gibco, ThermoFischer) was added to each flask. Cells that survived selection and expression validated by dot blot as described in section 2.2.6 below were upscaled and frozen for future stocks. Frozen cell stocks were maintained by freezing cells in 90% FCS (Gibco, Life Technologies) and 10% DMSO (Pierce™, ThermoFischer Scientific) at -20°C. Recombinant protein expression was confirmed by dot blot as described in section 2.2.6 below.

Stably transfected cells were rapidly thawed and resuspended in Ham's F-12 complete media (10% FCS, 50 µg/ml Geneticin, and 25 µg/ml Gentamycin) (ThermoFischer Scientific) before being centrifuged at 800 × g for 3 minutes. Cells were resuspended in the media and added to a T25 flat bottom culture flask and incubated at 37°C with 5% CO_2. Once the cells reached 90% confluency the cells were seeded into 5 × T75 flasks prior to being seeded into roller bottles once 90% confluency was reached in the T75 flasks. Cells in the roller bottles were maintained with Ham's F-12 media (5% FCS, 25 µg/ml Geneticin, and 25 µg/ml Gentamycin) (ThermoFischer Scientific). The roller bottles were rotated at 0.3 rpm at 37°C and 5% CO_2. Media containing $gp120_{FVC}$ExC was harvested twice a week until a total volume of 2 L was reached, after which protein expression levels decline. The harvested media was centrifuged at 14 000 × g for 5 minutes and filtered using a 0.22-micron filter to remove cells and debris from the media and frozen at -4°C until needed.

2.2.3. Purification of $gp120_{FVC}$ExC through lectin affinity chromatography

Lectin affinity chromatography was used to bind the native mannose glycans on the surface of the expressed $gp120_{FVC}$ExC and purify the recombinant protein. Two millilitres of lectin-agarose (Sigma Aldrich) was suspended in 1xPBS (pH 7.5) (ThermoFischer Scientific) and equilibrated with 100 × column volumes (CV) of 1xPBS. Filtered culture media containing $gp120_{FVC}$ExC was thawed and run through the column. The column was washed to reduce non-specific binding with 100 × CV of PBS containing first 0.5 M NaCl, followed by 1 M NaCl. Elution of $gp120_{FVC}$ExC was achieved by using 25 ml of 1 M methyl alpha-D-manno-pyranoside (MMP) (Sigma Aldrich). The column was first left to equilibrate for 5 min with 10 ml of the MMP solution before collecting this volume of the resulting eluate, after which the additional 15 ml MMP solution was added and passed through the column to remove the remaining $gp120_{FVC}$ExC. Once the protein had been eluted the column was regenerated using 100 × CV 1.5 M NaCl and stored in PBS with 0.02% Sodium Azide. Recombinant protein purification was confirmed by SDS-PAGE and Western blot analyses, as described in sections 2.2.7 and 2.2.8 below.

Eluted protein was concentrated to 1 mg/ml using a concentrating centrifuge tube with a 10 kDa size cut-off (Centricon®) filtered using a 0.22-micron filter, and protein expression and

concentration was determined using a Bicinchonic Acid assay (BCA assay) (ThermoFischer Scientific), as per manufacturer's instructions, and stored at 4°C.

2.2.4. Deglycosylation of $gp120_{FVC}ExC$

Deglycosylating $gp120_{FVC}ExC$ is an integral step in the process of crystallization of glycosylated proteins as it increases the probability of successful crystallization. This is achieved using Endoglycosidase H (Endo H) (Sigma Aldrich) to cleave between two N-acetylglucosamine (GlcNAc) subunits proximal to the asparagine residue (Figure 2.4). This truncates the glycans on the surface by leaving a remaining GlcNAc molecule. Glycans are inherently flexible in nature, and this flexibility creates heterogeneity and flexibility that impairs crystallization – which requires a high degree of homogeneity to create an identical lattice to make up the crystal structure. Endo H thus truncates the glycans to remove flexibility of the glycans and creates a more homogenous protein sample. The lectin affinity purified $gp120_{FVC}ExC$ was buffer exchanged through centrifugation into PBS (pH 5.5) for optimal enzymatic activity, using a 10 kDa Centricon®. The protein was concentrated to 1 mg/ml prior to the addition of Endo H. Optimized amount of Endo H was determined to be 2 units per 100 µg of purified $gp120_{FVC}ExC$ when incubated for 48 hours. The $gp120_{FVC}ExC$ – Endo H mixture was incubated overnight at 37°C after which the solution was filtered using a 0.22-micron filter to remove precipitated protein in order to prevent further aggregation of the sample. It was then left to incubate for a further 24 hours at room temperature. The deglycosylated protein was subsequently buffer exchanged through centrifugation into PBS (pH 7.5) using a 10 kDa Centricon® and concentrated to 1 mg/ml and filtered using a 0.22-micron filter. The deglycosylated $gp120_{FVC}ExC$ was stored at 4°C.

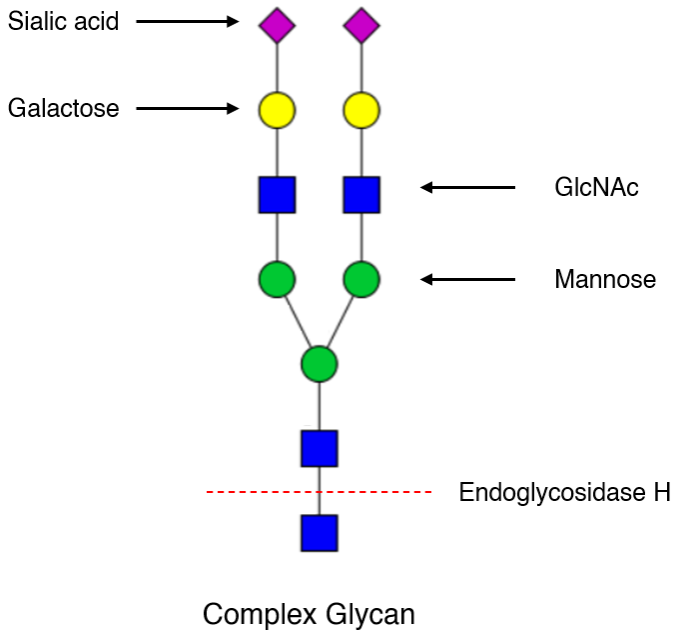

Figure 2.4: Representation of a complex N-linked glycan. Expression through GnTi$^-$ cells prevents the formation of the structure beyond the mannose residues. The cleavage site by Endoglycosidase H is indicated with a red dash, with cleavage resulting in a single GlcNAc residue remaining. Image adapted from https://en.wikipedia.org/wiki/File:Types_of_glycans.svg.

2.2.5. Separation of deglycosylated gp120$_{FVC}$ExC species

Two distinct species of gp120$_{FVC}$ExC were generated during the deglycosylation process, likely due to variability in the initial glycosylation process of the protein. This generated a species of a higher molecular weight (HMW) and one of a lower molecular weight (LMW). For the purpose of crystallization, species need to be separated to ensure homogeneity. To achieve the separation a Superdex HiLoad 75 pg (Mw range of 3 000 – 70 000 Da) (GE Lifescience) was used. The column was equilibrated with the crystallization buffer (2.5 mM Tris-HCl, pH 7.5, 350 mM NaCl, 0.02% NaN_3). The protein was loaded at a volume of 1 ml and a concentration of 2 mg/ml and passed through the column at a flow rate of 0.8 ml/min. Fractions were collected at a volume of 0.5 ml to prevent extensive overlap of the species in the fractions as they were not completely resolved on the column. Fractions were evaluated by SDS-PAGE, as described in section 2.2.7 below.

2.3.Dot Blot

To perform a Dot blot, a nitrocellulose membrane (BioRad) was coated with the target protein. Three microliters of each protein sample were added to the same location three times, and samples were allowed to dry between applications. The membrane was blocked for an hour using 3% BSA (Roche) in PBST (0.02% Tween 20 in PBS). It was then incubated with anti-gp120 rabbit sera (HPRU) diluted 1:2000 in PBST for an hour before being washed in PBST 3 times for 5 minutes. An anti-rabbit linked reporter IgG antibody (GE Healthcare UK Limited, ECL™) was added to the membrane and left for an hour prior to being washed 3 times for 5 minutes in PBST. Pierce ECL Western blotting substrate (Thermo Scientific) was added to the membrane and left to develop for 15 minutes prior to visualization on the Gel Doc imaging system set to Chemi-High Sensitivity.

2.4.SDS-PAGE

SDS-PAGE consists of two vertical layers of gel that allow for the separation of proteins based on molecular weight. The protein separates in a vertical fashion with the lower molecular weight resolving faster and appearing lower on the gel. The Mini-PROTEAN® TGX™ BioRad system was utilized to prepare and run the gels. The lower portion of the gel (separating gel) allows for the separation and resolution of the proteins. The top layer (stacking gel) allows for the loading of protein and ensures that the proteins enter the separating gel at the same time. The separating gel (10%) was prepared by mixing 3.3 ml of monomer solution (30% (w/v) acrylamide (Sigma-Aldrich), 2.7% (w/v) bis-acrylamide (Sigma-Aldrich)), 2.5 ml separating buffer (1.5 M Tris-HCl, pH 8.8), 10 µl TEMED (Tetramethylethylenediamine) (ThermoFischer Scientific), 100 µl SDS (Calibiochem) (10% (w/v), 4.1 ml dH_2O), and 32 µl of ammonium persulfate (Sigma-Aldrich) (10% (w/v). The gel was then poured and covered in isopropanol to protect the gel from oxygen that inhibits the polymerization of the gel. The gel was left to polymerize for an hour.

The stacking gel (4%) was prepared by mixing 1.3 ml monomer solution (30% (w/v) acrylamide (Sigma-Aldrich), 2.7% (w/v) bis-acrylamide (Sigma-Aldrich)), 2.5 ml stacking buffer (0.5 M Tris-HCl, pH 6.8,), 10 µl TEMED (ThermoFischer Scientific), 100 µl SDS (Calibiochem) (10% (w/v)),

6.1 ml dH_2O, and 100 µl ammonium persulfate (Sigma-Aldrich) (10% (w/v)). The stacking gel was poured over the separating gel once the isopropanol had been removed completely. A comb was then added into the stacking gel solution to form the wells and the gel was left to polymerize for another hour.

Protein samples were prepared using a 1:1 ratio of Sample Buffer (125 mM Tris, 4% SDS, 20% Glycerol (Saarchem, Merck), 2 mg bromophenol blue (Saarchem, Merck), 10% 2-Mercaptoethanol, pH 6.8) to protein. Samples were heated at 95°C for 5 minutes to aid the denaturation of the protein that occurs from the SDS and pulse centrifuged to collect any sample droplets prior to being added to the wells. Gels were run in Running Buffer (2.5 mM Tris 19.2 mM Glycine, 0.01% SDS) at 180 V for 1 hour on the Proteon BioRad system. They were subsequently stained using R-250 Coomassie blue stain (5% Methanol, 10 % Acetic acid, 40% d H_2O, 0.025% Coomassie blue R-250 (Sigma-Aldrich)) for an hour and destained using destaining solution (10% Acetic acid) until bands were visible. Gels were imaged using the Chemidoc XRS™ GelDoc (BioRad) imaging system and QuantityOne™ imaging software.

2.5.Western Blot

The protein separated in the SDS-PAGE gels were transferred to a nitrocellulose membrane (BioRad). Filter papers and the nitrocellulose membrane were equilibrated in transfer buffer (25 mM Tris-HCl (pH 7.6), 192 mM Glycine, 20% Methanol, 0.03% Sodium Dodecyl Sulphate) for 5 – 10 minutes using a semi-dry procedure on a Trans-Blot SD Semi-Dry Transfer Cell (BioRad) system. The protein was transferred for 20 minutes at 20 V. After protein transfer, the gel was stained with using R-250 Coomassie blue stain (5% Methanol, 10 % Acetic acid, 40% d H_2O, 0.025% Coomassie blue R-250 (Sigma-Aldrich)) to determine the relative amount of protein present as a gauge of transfer efficacy.

The membrane was blocked using Bovine Serum Albumin (BSA) to prevent non-specific antibody binding. This was performed using 3% BSA (Roche) in PBST (PBS with 0.02% Tween20), which was placed on the membrane for an hour. The membrane was washed 3 times for 5 minutes using PBST. A primary anti-gp120 rabbit serum (HPRU) directed at the protein of interest was diluted

1:2000 in PBS and placed on the membrane for 1 hour, prior to washing as described. A secondary anti-rabbit linked reporter IgG antibody (GE Healthcare UK Limited, ECL™) that binds the conserved region of the IgG primary antibody was placed on the membrane for an hour and washed as previously described. The membrane was developed using Pierce ECL™ Western blotting substrate (Thermo Scientific) for 15 minutes before being imaged using a GelDoc XRS™ set to the "Chemi-High Sensitivity" function that detects the fluorescence from the secondary antibody.

2.6.Crystal Trials – screening and optimization

Of the two species produced by the Endo-H deglycosylation of $gp120_{FVC}$ExC the Low Molecular Weight (LMW) species (section 2.2.4 and 2.2.5) was selected for further studies. This was chosen under the assumption that the LMW species had undergone more extensive deglycosylation that will reduce conformational flexibility and increase the probability of crystallization. The LMW $gp120_{FVC}$ExC was pooled from the selected SEC fractions and concentrated immediately prior to crystallization to ensure minimal degradation and precipitation of the protein. All crystallization trials were performed on the automated Oryx platform, using a Hamptons Index kit to test 96 common crystallization conditions. The protein was concentrated to 4 mg/ml and crystal trials were conducted using a sitting drop method. The drops were generated at a protein to mother liquor ratio of 1:1, 1:2, and 1:3 for each of the 96 conditions. The plate was left at 20°C and periodically analysed for crystal formation. No crystals viable for upscaling were produced. However, the formation of needles was observed that were subsequently used as a seed stock for further rounds of crystal trials.

The seed stock was generated from the existing protein needles. The needles were harvested using a loop and placed into 10 µl of mother liquor in which the needles were grown before being broken with a glass wand. A glass bead was added to the mixture containing the protein crystal fragments and centrifuged discontinuously for 5 seconds to further fragment the needles. The seed stock solution containing the needles was diluted into the various mother liquor solutions of the Hamptons Index kit at a ratio of 1:10 seed stock to mother liquor. The soluble deglycosylated

gp120$_{FVC}$ExC protein was subsequently added to the various mother liquor solutions containing the seed stock at a concentration of 1 mg/ml and a ratio of 1:1.

Additional trials were run using another prominent kit (Morpheus) to determine potential conditions for crystallization. The Morpheus kit also has 96 common conditions but uses an under-oil method for crystallization. This also allows for the use of smaller amounts of protein per trial, allowing for a wider variety of protein concentrations to be attempted. Three plates were set up with ratios of 1:1, 1:2, and 1:3 of protein to mother liquor. The protein concentration was kept at 4 mg/ml. The plates were left at 20°C and periodically analysed for crystal growth. Two conditions indicated crystal growth that can be upscaled to increase the size of the crystals for X-ray diffraction.

2.7.Generation of Homology Models and Molecular Dynamic (MD) Simulations

To prepare the proteins for MD simulations, homology models were generated of the following proteins: gp120Core$_e$, gp120$_{FVC}$ExC, 2dCD4^{WT}, and 2dCD4^{S60C}. Homology models utilize the sequence from the proteins of interest (gp120$_{FVC}$ExC and 2dCD4$^{S60C)}$ and existing X-ray crystallography models (gp120Core$_e$ and 2dCD4^{WT}) to generate a model for where there is a larger than 75% sequence similarity. The unliganded gp120$_{FVC}$ExC model was generated using our gp120$_{FVC}$ExC sequence and the crystal structure of gp120Core$_e$ (PDB ID: 3TGR) to generate a protein model of our sequence. The 2dCD4^{S60C} and 2dCD4^{WT} homology model was generated using the X-ray structure of wild type (WT) 2dCD4 (PDB ID: 2NY4). These models were generated using the SWISS-MODEL platform (www.swissmodel.expasy.org) providing our own protein templates that closely resemble our proteins of interest. Generation of the gp120$_{FVC}$ExC – 2dCD4$^{S60C/WT}$ complexes was performed on the PyMOL platform (www.pymol.org) by fitting the constituent proteins (gp120 and 2dCD4 $^{S60C/WT}$) over existing crystal structures of gp120-2dCD4 complexes (PDB ID: 3LQA) with the accompanying antibody removed before it was used as a template. The fitted homology models were exported to retain the structure of the proteins and their orientation. Once these model of the gp120$_{FVC}$ExC – 2dCD4 ^{S60} was generated, a disulphide bond was introduced in the gp120$_{FVC}$ExC – 2dCD4^{S60C} using PyMOL by linking the chains of the relevant residues, and the length of the bond was generated at 1.5 Angstroms (Å).

The molecular dynamic (MD) simulations, as well as the preparation of the proteins were performed on the Schrodinger Maestro 12.2 platform. The proteins were prepared using a standard procedure using the OPLS2005 force field. The homology models underwent optimization to rotate bonds to more favourable energy orientations in the absence of crystallization artefacts. Due to the nature of homology model generation, water molecules were removed and as such needed to be reintroduced to the structures. The protein was placed in a cubic unit cell in which the protein was cantered leaving 1nm between the protein and the edges of the box. The unit cell was filled with solvent and neutralized to a net charge with the addition of the necessary amount of Sodium (Na^+) and Chloride (Cl^-) ions. The system was relaxed using a step wise energy minimization procedure to the nearest local minimum. The final simulation was run at 300K for 50 ns at 1 ps intervals for a total of 1000 frames.

All other analyses of the MD simulation proteins were performed on the Maestro platform to analyse the Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), and Secondary Structural Elements (SSE's) of the proteins that underwent simulation. Protein simulations were visualized on the Maestro platform.

The Root Mean Square Deviation (RMSD) calculates the global fluctuation of the entire protein structure per frame for the entire simulation. This provides insight into the mobility of the entire structure on a frame-by-frame basis. It is important to note that the overall trend is the result of interest as the frames between the two structures are not comparable due to the different initiation points of each simulation post relaxation of the protein. However, it also acts as a good measure for the quality of the simulation and is often employed as a quality control step, as an extremely large fluctuation in the RMSD usually indicates that the molecule has moved beyond the bounds of the water-box containing the protein. This boundary is considered to be between 2.5-3 Å and all the simulations we performed fell within the acceptable bounds.

The Root Mean Square Fluctuation (RMSF) of the residues in an MD simulation determines the amount of movement for each residue compared to the reference frame (frame 0) over the entire period of the simulation, which in our case was 50 ns. This provides insight into the changes in structure, dynamics, and subsequent interatomic interactions between the compared structures. A strong reduction in the fluctuation of a region can also indicate the truncation of a region.

To aid in determining the protein residues and regions that exhibit significant difference between the gp120 structures in the MD simulation, a strategy was employed to reveal significant differences to be investigated. This was achieved by calculating the percentage difference in the RMSF of the individual residues between the two proteins using the following formula:

$$\left(1-\left(\frac{RMSF\ of\ gp120Core_e\ residue}{RMSF\ of\ gp120_{FVC}ExC\ residue}\right)\right)\times 100\ = Percentage\ differnece\ in\ RMSF\ of\ the\ residues$$

This ultimately allowed us to investigate individual residues and regions that display significant differences in fluctuations when the structures are compared. A 30% cut-off was used to attempt to isolate the residues that experienced the most significant change between the two structures for investigation. The 30% cut-off was chosen based on the results from the MD analysis. This allowed for the isolation of regions in the structures that comprise more than single residues and exhibit the largest fluctuations that are more likely to be relevant in the analysis.

3. Results

3.1.Design of a modified $gp120_{FVC}ExC$ sequence

To increase the probability of crystallizing the expressed $gp120_{FVC}ExC$, the $gp120_{FVC}$ sequence was modified by the introduction of substitutions in the V1/V2 region (as found in $gp120Core_{min}$) and V3 region (as found in the $gp120Core_e$), while the V1 and V2 included an additional nine residues compared to $gp120Core_e$ (Figure 3.1). The additional nine residues included in the V1/V2 region are integral to the formation of the putative disulphide bond between Cysteine 113 of gp120 (that resides in the V1 loop) and the Cysteine residue at position 60 of $2dCD4^{S60C}$ of the immunogenic $gp120_{FVC}ExC$ – $2dCD4^{S60C}$ complex. The inclusion of these resides allow for the validation of our ability to crystallize the $gp120_{FVC}ExC$ in isolation, as well as an eventual component in the $gp120_{FVC}ExC$ – $2dCD4^{WT/S60C}$ complexes.

Apart from the adopted modifications performed on the $gp120_{FVC}ExC$ sequence, there were other sequence differences between the $gp120_{FVC}ExC$ and the $gp120Core_e$ (PDB ID: 3TGR) sequence, both of which are HIV-1 subtype C derived sequences. Several amino acids throughout the sequence, with certain regions containing large sequence differences were noted. The sequence between the V3 region and the CD4 binding loop show numerous differences that may be characteristic of the FV sequence and may impact the functionality of the V3 and CD4 binding loops (Figure 3.1). The other notable regions are the V4 region of $gp120_{FVC}ExC$ that has an 11 amino acid deletion, as well as a three amino acid residue insertion in the V5 region (Figure 3.1).

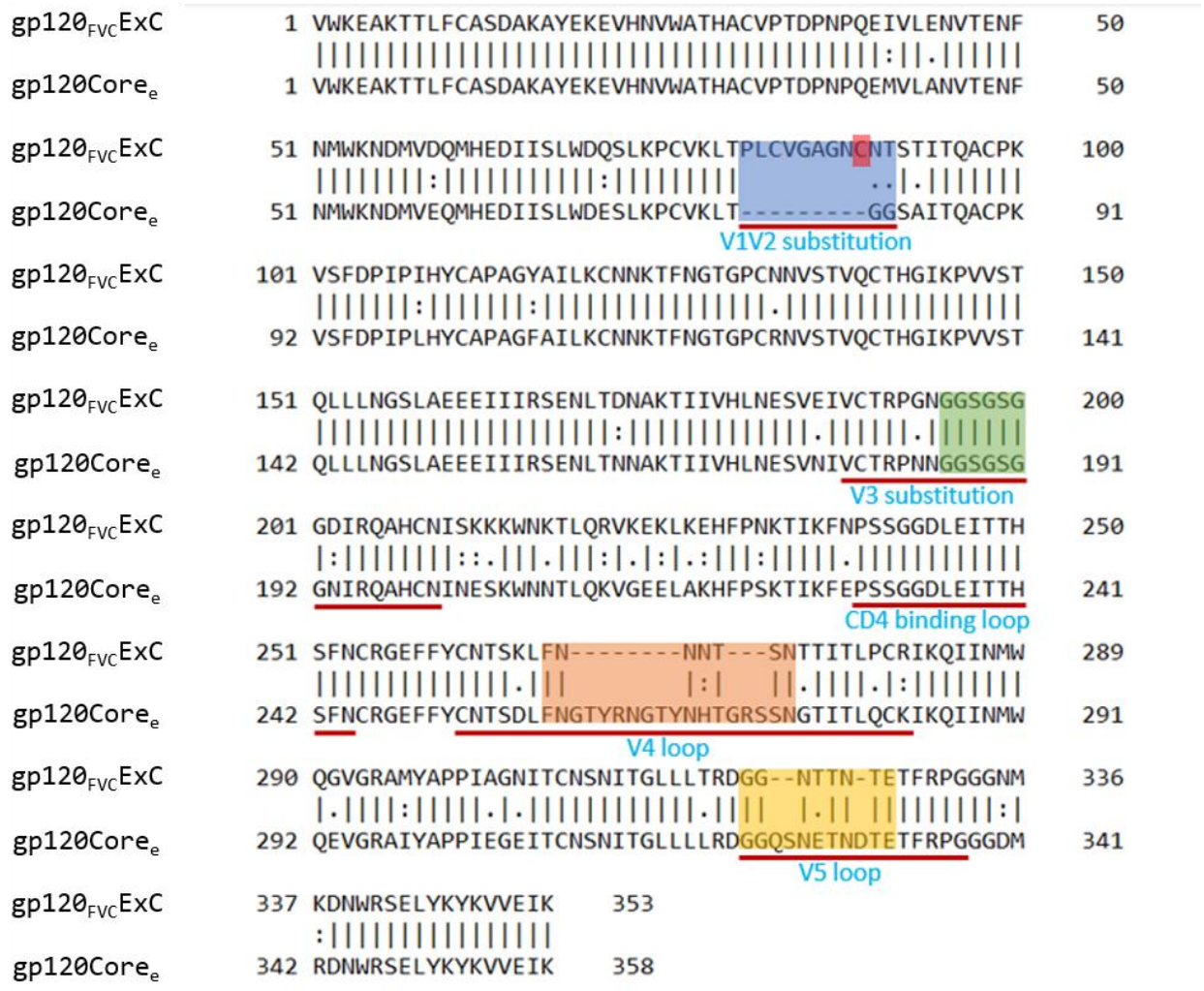

Figure 3.1: Amino acid sequence comparison of the novel gp120$_{FVC}$ExC consensus sequence compared to the gp120Core$_e$ (3TGR). Functional regions of the gp120Core$_e$ and gp120$_{FVC}$ExC proteins are indicated by red bars. Changes in the sequences are indicated by the blue, orange, and yellow blocks. The green block indicates the substitution of the V3 loop. The red highlighted block indicates the residue involved in the putative disulphide bond in gp120$_{FVC}$ExC-2dCD4^{S60C} complexes.

3.2.Successful gp120$_{FVC}$ExC expression in mammalian cell lines

Dot blot analysis confirmed the presence of secreted gp120$_{FVC}$ExC from GnTi- deficient HEK293S cells in the tissue culture media (Figure 3.2). This step was performed once the cells

reached confluency in the roller bottles to ensure an adequately detectable amount of recombinant protein was being expressed and secreted into the supernatant. The expression was confirmed by the positive control of previously purified gp120 and a negative control of fresh culture media (Figure 3.2).

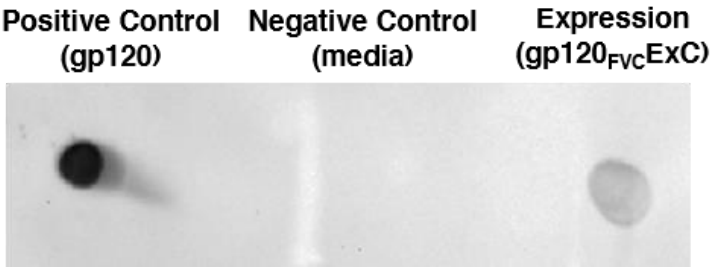

Figure 3.2: Dot blot confirming the expression of $gp120_{FVC}ExC$ following detection using anti-gp120 rabbit serum. Positive and negative controls were included, where the positive control is a known gp120 sample and the negative control is untransfected cell culture media.

3.3.Effective purification of $gp120_{FVC}ExC$ through lectin affinity chromatography

A purification profile of $gp120_{FVC}ExC$ was generated to determine the efficiency of the lectin affinity chromatography purification, and to identify if subsequent purifications were necessary to ensure optimal yield of purified protein from the harvested culture supernatant. Two purifications were performed in succession on the same batch of harvested culture supernatant and analysed through SDS-PAGE and western blot (Figures 3.3 and 3.4).

The initial purification showed distinct bands visible in the media and flow-through samples that corresponds to the size of $gp120_{FVC}ExC$ at approximately 58 kDa (Figure 3.3). However, as indicated by the corresponding western blot (Figure 3.4) this is not $gp120_{FVC}ExC$ but rather a protein native to the expression media as the bands in the western blot are smaller and less diffuse. These proteins are likely FCS protein bands from the cell media. These proteins have a size of 66.5 kDa which on a 10% SDS-PAGE gel would overlap with the $gp120_{FVC}ExC$ band at 58 kDa. Importantly the presence of protein in the flow-through sample visible on the western blot is indicative of the need for an additional purification (Figure 3.4). A faint band is visible at ~58 kDa

in the first wash sample (Figure 3.3), however, the corresponding band on the western blot shows little to no detection of gp120$_{FVC}$ExC (Figure 3.4). The eluted and concentrated bands of the first purification show proteins at ~58 kDa as expected for gp120$_{FVC}$ExC (Figure 3.3). These bands in the eluted and concentrated samples on the corresponding western blot (Figure 3.4) indicate that they represent gp120$_{FVC}$ExC with minimal contamination as the bands corresponds to those of the SDS-PAGE in intensity and size. The second purification is similar to that of the first with a markedly less intense band in the concentrated sample of gp120$_{FVC}$ExC observed on the SDS-PAGE (Figure 3.3). However, the corresponding western blot shows a similar profile in the second purification to that of the first (Figure 3.4). It should be noted that the protein yield after a third round of purification is drastically less than that of the second round of purifications (results not shown) and as such only two rounds of purifications were performed per batch of gp120$_{FVC}$ExC containing culture supernatant.

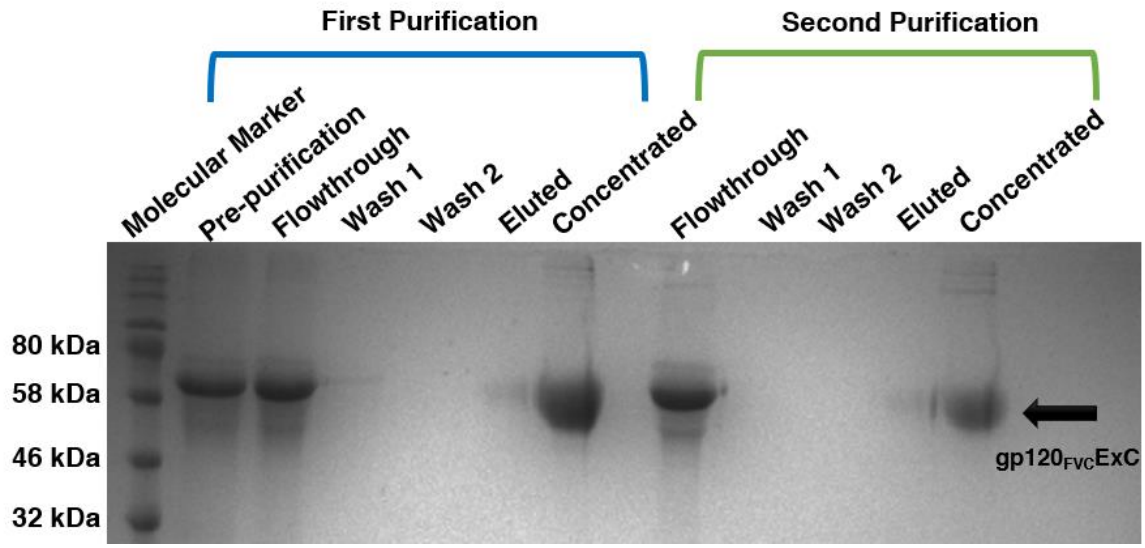

Figure 3.3: A 10% SDS-PAGE gel indicates that the lectin purification steps used are adequately purifying gp120$_{FVC}$ExC while maintaining a high protein yield. While some protein is visible in the initial wash step (Wash 1, blue), protein is only observed in the elution steps. Further, two rounds of purification are necessary to ensure optimal purification. The Molecular Weight marker (P7712, New England Biolabs® Inc.) is indicated on the left, in kilodaltons (kDa). The arrow points to gp120$_{FVC}$ExC.

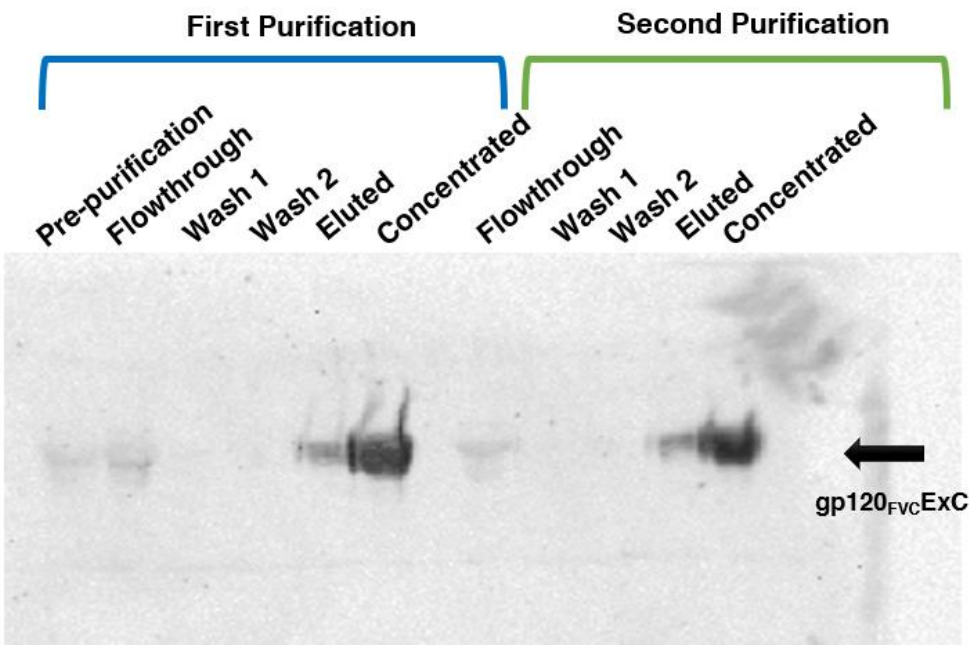

Figure 3.4: Western blot analysis confirming the presence of gp120$_{FVC}$ExC in the elution and concentrated samples of the various gp120$_{FVC}$ExC purification steps, as well as confirming the presence of a contaminating non gp120$_{FVC}$ExC protein of a similar size to gp120$_{FVC}$ExC in the expression media. The arrow points to gp120$_{FVC}$ExC.

Therefore, gp120$_{FVC}$ExC was successfully purified using lectin affinity chromatography, but more than one purification per batch of media was required to maximize protein yields based on the purification resin and column sizes used.

3.4.Cleavage of gp120$_{FVC}$ExC N-linked glycans using Endo H

Deglycosylation of gp120$_{FVC}$ExC was performed using Endoglycosidase H (Endo H) to cleave the N-linked glycans, reducing the overall glycosylation of the recombinant protein. A reduction in the size of gp120$_{FVC}$ExC from 58 kDa to approximately 32-46 kDa indicates successful deglycosylation (Figure 3.5). A band is visible at approximately 28 kDa in the deglycosylated samples that corresponds to the molecular weight of Endo H (Figure 3.5). Two distinct bands are also visible in the deglycosylated protein samples that indicates successful deglycosylation that

results in two distinct deglycosylated gp120$_{FVC}$ExC species. The formation of the two deglycosylated species is likely due to the variable native glycosylation of gp120$_{FVC}$ExC during expression and subsequent post-translational modifications, as noted by the large disperse band presented in the glycosylated gp120$_{FVC}$ExC lane (Figure 3.5). Successful deglycosylation of gp120$_{FVC}$ExC was achieved using 2 units of Endo H per 100 µg of gp120$_{FVC}$ExC over 48 hours at a pH of 5.5.

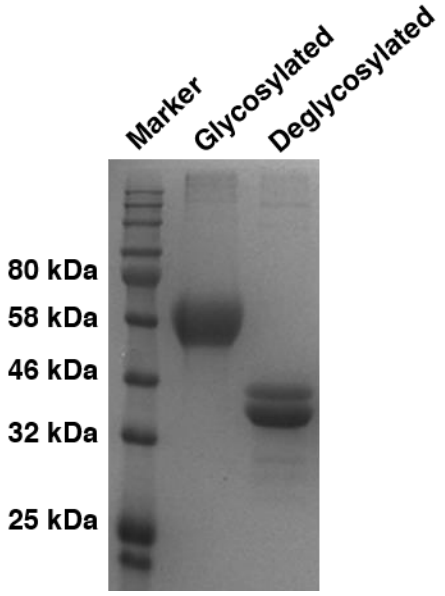

Figure 3.5: A 10% SDS-PAGE gel indicates that deglycosylation of gp120$_{FVC}$ExC using Endo H generates two distinct species between 32- and 46 kDa, a significant shift compared to Glycosylated gp120$_{FVC}$ExC at 58 kDa. The Molecular Weight marker (P7712, New England Biolabs® Inc.) is indicated on the left, in kilodaltons (kDa).

3.5.Separation of deglycosylated gp120$_{FVC}$ExC species using Size exclusion chromatography

Separation of the two deglycosylated gp120$_{FVC}$ExC species was performed with a HiLoad Superdex 75pg column, as indicated by the two overlapping peaks (Figure 3.6). Fractions were collected for analysis through SDS-PAGE to select samples that contain distinctly one species of gp120$_{FVC}$ExC, ensuring homogeneity for crystallization (Figure 3.7). Incomplete separation of the two species was achieved through SEC as observed by the overlapping peaks of the chromatogram (Figure 3.6) and the SDS-PAGE of the collected fractions shows fractions containing bands

corresponding to both $gp120_{FVC}$ExC species (Figure 3.7). The larger species of $gp120_{FVC}$ExC is represented by the first peak on the chromatogram, while the second peak represents the smaller molecular weight species (Figure 3.6) as shown by the SDS-PAGE of the collected fractions of the corresponding peaks (Figure 3.7). The second peak representing the smaller molecular weight species of $gp120_{FVC}$ExC has a higher UV absorbance reading than that of the higher molecular weight, indicating a larger amount of the lower molecular weight $gp120_{FVC}$ExC species present (Figure 3.6). The SDS-PAGE of fractions 15-23 show limited contamination of the low molecular weight species of $gp120_{FVC}$ExC (Figure 3.7) as indicated by a singular band of protein. This lower molecular weight species was selected and pooled for further crystallization studies due to the likely more complete deglycosylation and subsequent lower conformational dynamics of these proteins. Fractions 15 – 23 were selected and pooled for crystallization based on the purity of the low molecular weight species of protein (Figure 3.7). This aims to ensure as much homogeneity of the protein present to improve the probability of crystal formation.

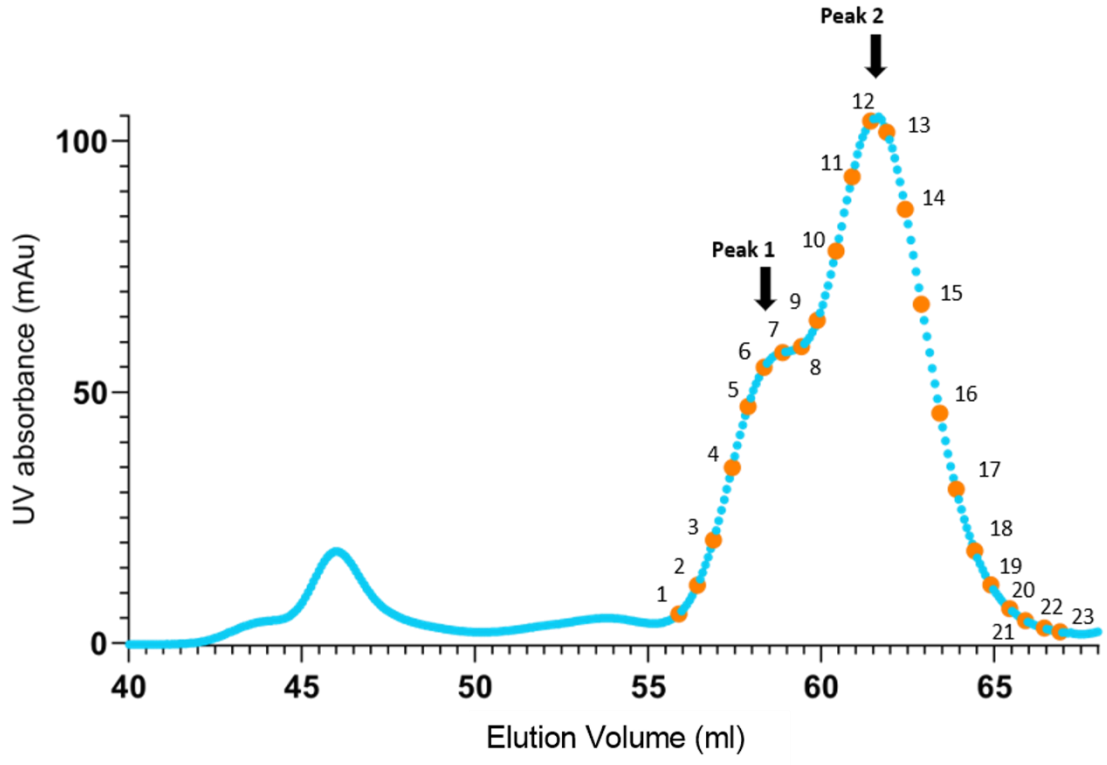

Figure 3.6: Chromatogram showing the separation and purification of the two deglycosylated gp120$_{FVC}$ExC species following SEC. The overlaps of the first and second peak indicate non-discrete separation of the proteins. Orange dots represent the collection fractions from the purification. The protein was run through a Hi-Load Superdex 75pg column.

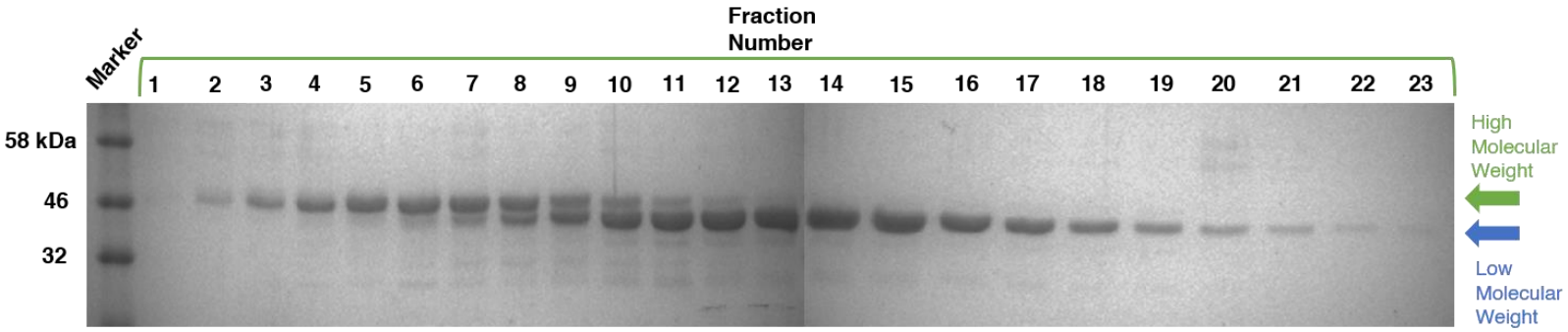

Figure 3.7: Separation of the two deglycosylated Low Molecular Weight (LMW, blue arrow) and High Molecular Weight (HMW, green arrow) gp120$_{FVC}$ExC protein species achieved through SEC. Samples were run on a 10% SDS-PAGE gel under reducing conditions. The Molecular Weight marker (P7712, New England Biolabs® Inc.) is indicated on the left, in kilodaltons (kDa).

3.6.Conditions in the Morpheus crystallization kit produces $gp120_{FVC}$ExC protein crystals

Two kits were used for the crystallization of $gp120_{FVC}$ExC. The Hamptons Index kit, using a sitting drop vapour diffusion setup to assess 96 common conditions for crystallization resulted primarily in precipitation, clear drops, and few amorphous crystals that are unlikely to be protein in nature. However, one condition (44, from Hamptons Index Kit HR2-144) (0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3 350) produced protein needles when trialed with a protein concentration of 4 mg/ml at a ratio of 1:1 with mother liquor (Figure 3.8). These crystals were too small in size to both isolate and perform X-ray diffraction on. Instead these crystals were harvested and used to produce a seed stock to further seed a Hamptons Index kit to increase the size and probability of crystal formation through the presence of protein fragments to act as nuclei. This did not yield any results and predominately led to precipitation despite the reduction in starting protein concentration to 1 mg/ml.

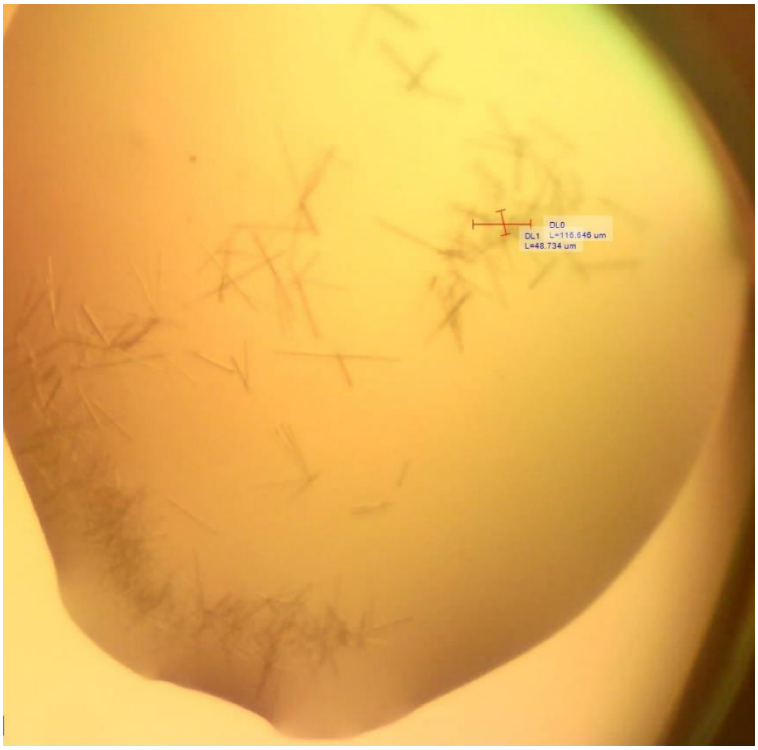

Figure 3.8: The formation of protein needles using a Hamptons Index kit. Condition 44 Hamptons Index Kit HR2-144 (0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3 350) with a protein concentration of 4 mg/ml and a ratio of protein to mother liquor of 1:1 produced needle shaped protein crystals.

Subsquently both the Hamptons Index kit and the Morpheus kit were used in an under-oil set-up. An under-oil set up allows for different results to standard vapor diffusion as the rate at which the protein concentrates vary between the methods, as well as allowing certain proteins to crystallize (although the mechanism for this is not quite understood). The Hamptons Index kit showed no results of value. The Morpheus kit had two promising conditions (0.06 M (0.3 M Calcium chloride dihydrate, 0.3 M Magnesium chloride hexahydrate); 0.1 M Tris, BICINE pH 8.5; 30% (40% v/v Ethylene glycol, 20% w/v PEG 8000; and 0.06 M (0.3 M Calcium chloride dihydrate, 0.3 M Magnesium Chloride hexahydrate); 0.1 M Tris, BICINE pH 8.5; 37.5% (25% v/v MPD, 25% w/v PEG 1000, 25% w/v PEG 3350) that produced multiple protein crystals that exhibited birefringence with results from the first condition shown in Figure 3.9. These crystals were produced both at a ratio of 1:2 and 1:3 protein to mother liquor, respectively. These conditions that produced protein crystals were noted, to be upscaled to produce larger crystals viable for future X-ray diffraction studies of $gp120_{FVC}$ExC.

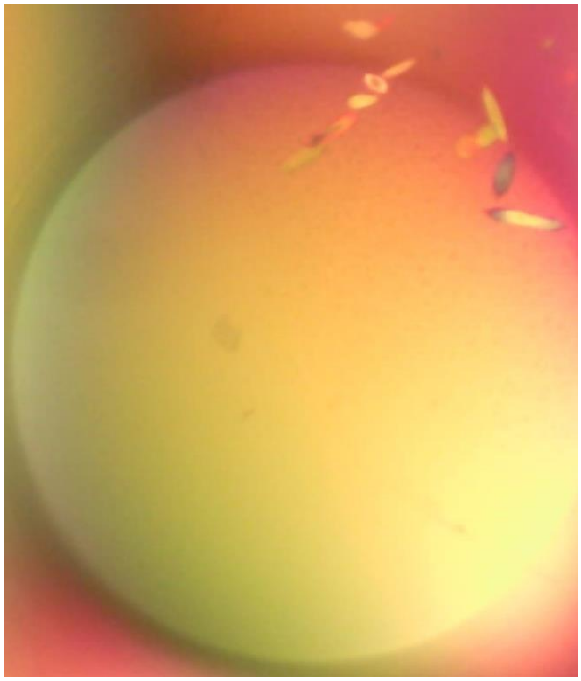

Figure 3.9: Protein crystal formation of deglycosylated $gp120_{FVC}$ExC.using a Morpheus kit in an under-oil setup. These crystals were produced at a protein concentration of 4 mg/ml and a ratio of 1:2 protein to mother liquor for condition (0.06 M (0.3 M Calcium chloride dihydrate, 0.3 M Magnesium chloride hexahydrate); 0.1 M Tris, BICINE pH 8.5; 30% (40% v/v Ethylene glycol, 20% w/v PEG 8000) of the kit.

3.7.MD simulations

3.7.1. Novel $gp120_{FVC}ExC$ sequence has reduced conformational dynamics compared to $gp120Core_e$ sequence in MD simulations

Homology models of both the $gp120_{FVC}ExC$ and $gp120Core_e$ proteins were successfully generated to obtain valuable data and comparisons between the structural and conformational dynamics of the proteins in MD simulations. The Root Mean Square Deviation (RMSD) for $gp120_{FVC}ExC$ and $gp120Core_e$ is shown in Figure 3.10. The RMSD calculated the global fluctuation of the entire protein structure per frame for the entire simulation. This provides insight into the mobility of the entire structure on a frame-by-frame basis. It is important to note that the overall trend is the result of interest as the frames between the two structures are not comparable due to the different initiation points of each simulation post relaxation of the protein. However, it also acts as a good measure for the quality of the simulation and is often employed as a quality control step, as an extremely large fluctuation in the RMSD usually indicates that the molecule has moved beyond the bounds of the water-box containing the protein. This boundary is considered to be between 2.5-3 Å and all the simulations performed for $gp120_{FVC}ExC$ and $gp120Core_e$ fell within the acceptable bounds.

The RMSD of $gp120_{FVC}ExC$ is overall lower than that of $gp120Core_e$ (Figure 3.10). This is particularly apparent from frame 500 onward once the structures have reached equilibration. The average RMSD for the $gp120_{FVC}ExC$ is 0.128 Å smaller than that of $gp120Core_e$, further suggesting that the conformational flexibility of $gp120_{FVC}ExC$ may be lower than that of the $gp120Core_e$ structure.

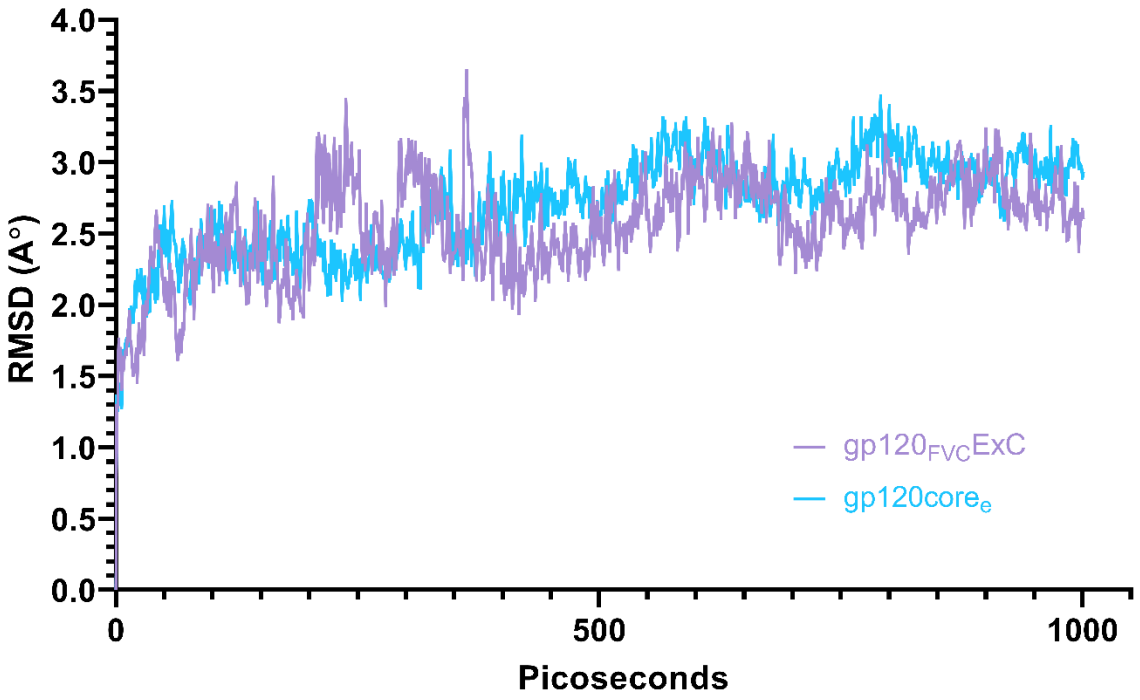

Figure 3.10: The RMSD of $gp120_{FVC}ExC$ (purple) and gp120 3TGR (blue) of the averaged alpha carbons, showing minimal change between the structures once equilibrated (from approximately 500 picoseconds onwards). However, the RMSD of $gp120_{FVC}ExC$ remains lower than that of $gp120Core_e$.

The RMSF shows a strong reduction in the fluctuation of the regions between the variable loops of the structures, possibly indicating truncations (Figure 3.11). The aligned RMSF profiles of the structures shows gross differences between the two structures due to sequence variation and fluctuations of the residues (Figure 3.11).

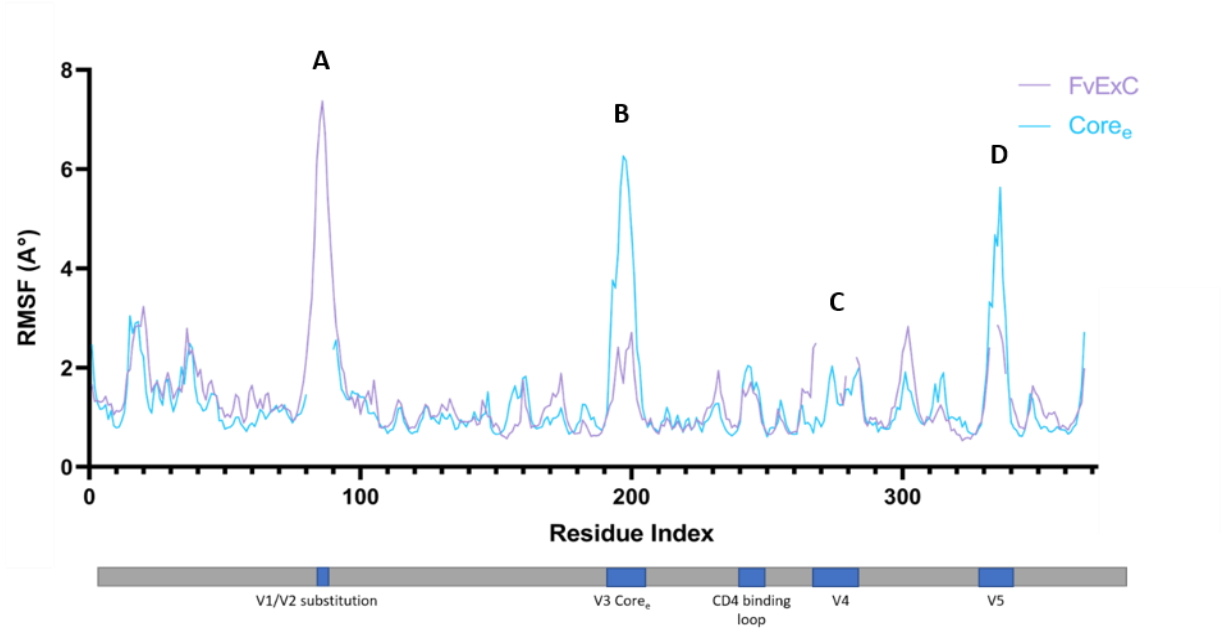

Figure 3.11: The Root Mean Square Fluctuation (RMSF) of gp120$_{FVC}$ExC and gp120Core$_e$ depicts the conformational changes made to the gp120$_{FVC}$ExC (purple) structure compared to a gp120Core$_e$ (blue) to enhance the probability of crystallization. This is most evident in the large fluctuations seen in the residues over time of the Variable loop regions (V1-V5) that are indicated with letters A-D (V1 and V2 regions overlap in A). The grey bar indicates the structure of the Env protein, with key functional residues indicated in blue.

The percentage difference in the RMSF of the individual residues between the two proteins was calculated (Figure 3.12). While many regions showed fluctuation above the 30% cut-off, they are primarily single residues and were subsequently omitted. However, four regions that indicated significant differences, as well as observable structural variation were identified for further analysis (Figure 3.12, Figure 3.13). These regions comprise the V1, V2, V3, V4, and V5 regions (V1 and V2 overlap, highlighted section A in Figure 3.12 and Figure 3.13). It should be noted that despite not meeting the cut-off criteria for analysis, the CD4 binding region, as a functional region of interest, exhibits some structural variations that are worthy of discussion. Of these residues in the CD4 binding loop, residues 241, 242, 243, and 253 have reduced fluctuations for gp120$_{FVC}$ExC compared to gp120Core$_e$, while residue 248 of this region for gp120$_{FVC}$ExC has more flexibility.

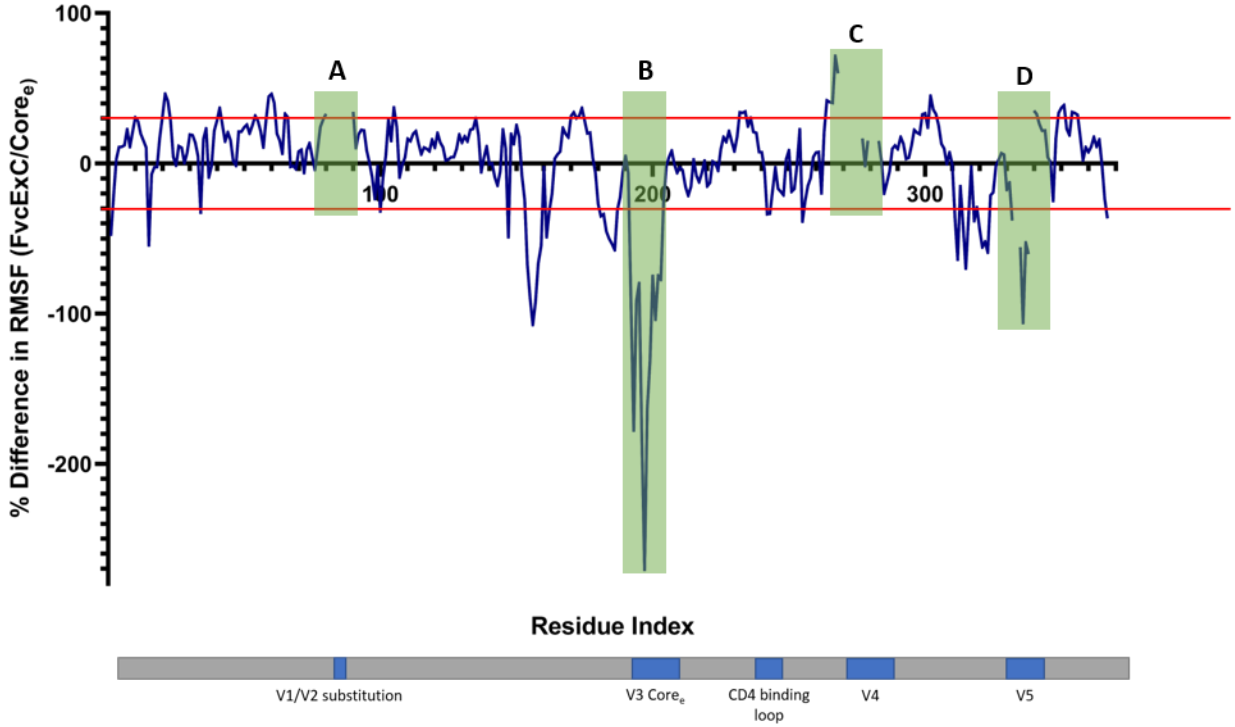

Figure 3.12: The percentage difference in the RMSF for the individual residues between gp120$_{FVC}$ExC and gp120Core$_e$ shows distinct differences between the structures – notably in the V3, CD4 binding loop, and V5 regions. A positive percentage difference indicates that gp120Core$_e$ residues are more conformationally flexible. Red bars indicate the 30% cut-off metric by which residues that experience significant fluctuations was determined. These regions are indicated with letters A-D and highlighted with green squares. The grey bar indicates the structure of the Env protein, with key functional residues indicated in blue.

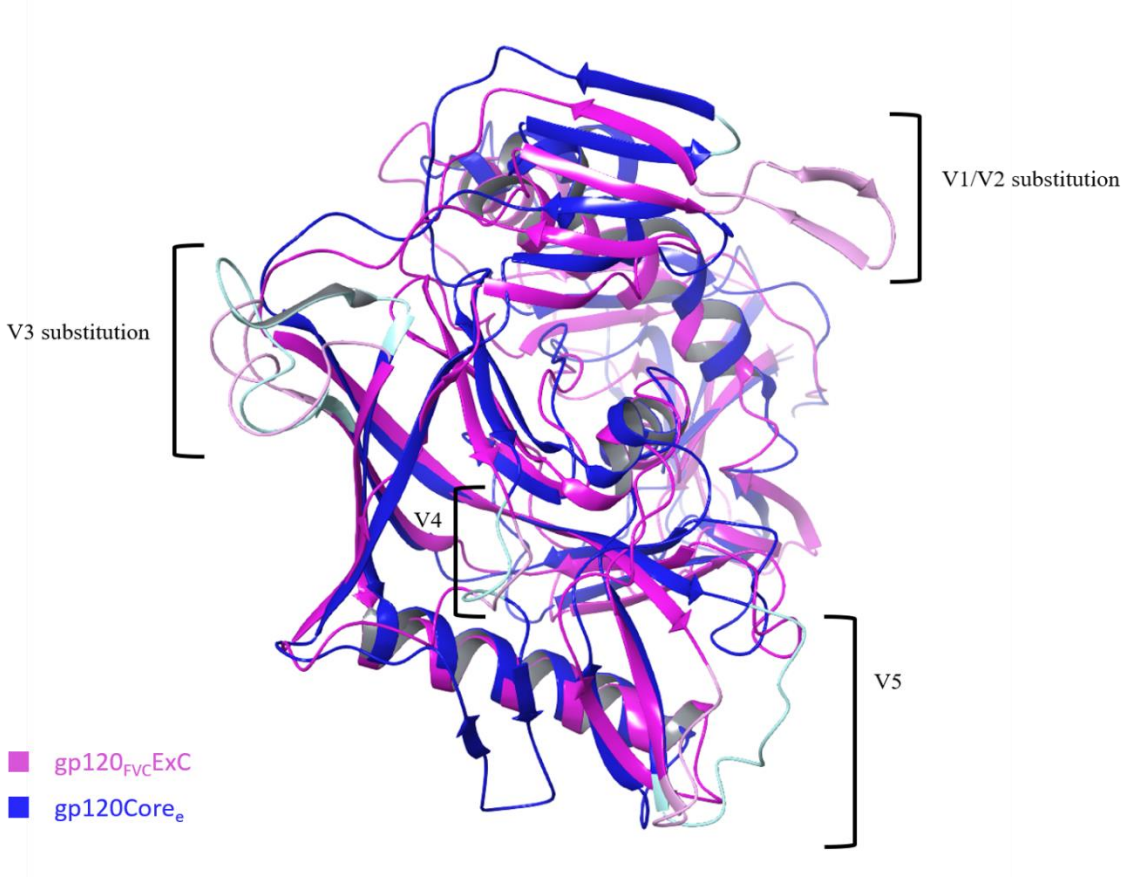

Figure 3.13: Ribbon representation of our $gp120_{FVC}ExC$ homology model (purple) superimposed over $gp120Core_e$ (blue) (PDB: 3TGR). The V1/V2 and V3 modifications that differentiate gp120 from $gp120Core_e$ and $gp120_{FVC}ExC$ structure to aid in unliganded crystallisation are highlighted. Regions investigated in the analysis are labelled.

The four regions (V1/V2, V3, V4, and V5) determined by the percentage difference to be significant were further inspected through MD simulations (Figures 3.13 to 3.17). These images are merely single representative snapshots of the MD simulations, and are used only to illustrate the differences, but do not account for the full dynamic fluctuations observed.

3.7.2. Longer V1V2 region of $gp120_{FVC}ExC$ compared to $gp120Core_e$ leads to increased fluctuation in the residues of this region

The residues adopted from $gp120_{min}$ in $gp120_{FVC}ExC$ compared to $gp120Core_e$ required for the putative disulphide bond to form between residue 113 of $gp120_{FVC}ExC$ and residue 60 of 2dCD4

increase the fluctuations of the residues in the V1V2 substitution region. This is apparent when looking at the RMSF of gp120$_{FVC}$ExC compared to gp120Core$_e$ (Figure 3.11). However, due to the disparity of gp120$_{FVC}$ExC containing the additional residues necessary for the putative disulphide bond, calculating the percentage difference is not possible. It was included as the difference is easily visually observed. The visual representation of the fluctuation (Figure 3.14) shows a larger loop structure in gp120$_{FVC}$ExC when compared to the loop found on gp120Core$_e$ (Figure 3.14) leading to a larger conformationally flexible loop and increased fluctuation of the flanking residues. A single snapshot of the MD simulation to better illustrate the differences between gp120$_{FVC}$ExC and gp120Core$_e$ show the extent of the loop formed by the additional residues and the conformational dynamics in the V1V2 loop structure.

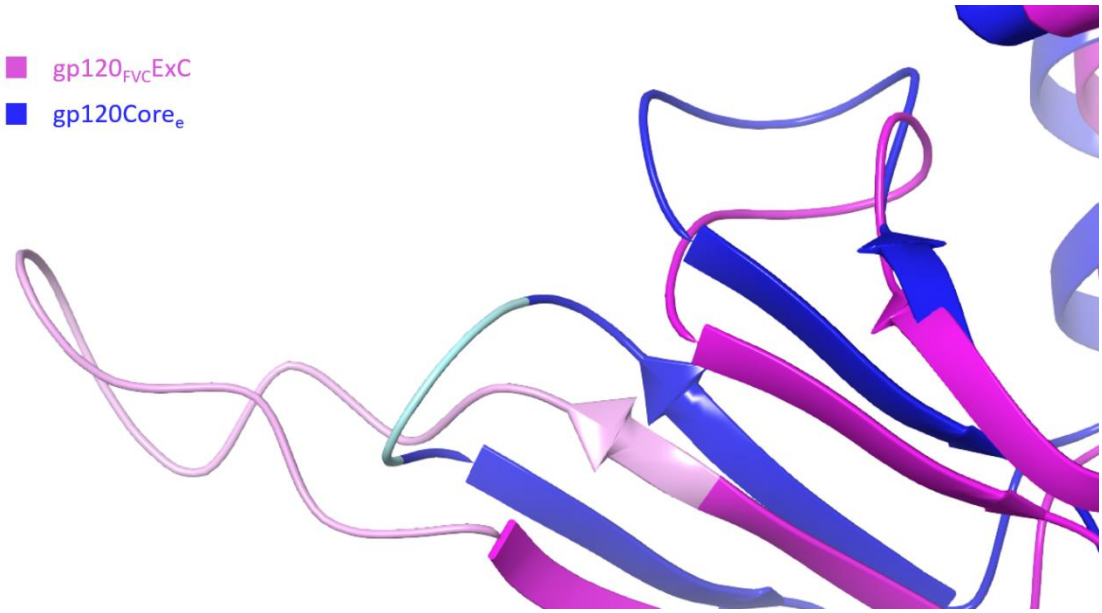

Figure 3.14: Visual representation of the V1/V2 substitution region, key residues with a difference in fluctuation are highlighted for gp120Core$_e$ (light blue) and gp120$_{FVC}$ExC (light purple). The additional residues in gp120$_{FVC}$ExC form a larger more conformationally flexible loop and to a lesser extent the flanking residues.

3.7.3. gp120$_{FVC}$ExC shows reduced fluctuation in residues that comprise the V3 region

The largest difference in fluctuation is observed around the V3 loop region where gp120Core$_e$ has much larger fluctuation of the residues in this region compared to gp120$_{FVC}$ExC (Figure 3.11,

Figure 3.12), despite the proteins' sequence similarity in this region. This region of fluctuation is for both the residues before the V3 region (166-187) and the residues that form the modified V3 region (190-205). The MD visual representation of the fluctuation shows that while gp120$_{FVC}$ExC takes a more coiled conformation, the corresponding residues of gp120Core$_e$ remain in a more open and flexible conformation (Figure 3.15). This appears to be in part due to the Asn to Gly-217 and Asn to Asp-266 mutations in gp120$_{FVC}$ExC compared to gp120Core$_e$. This allows for the formation of more hydrogen bonds in gp120Core$_e$ that result in coiling of the structure. Mostly all but the outermost residues appear to exhibit differences in fluctuations in the V3 region.

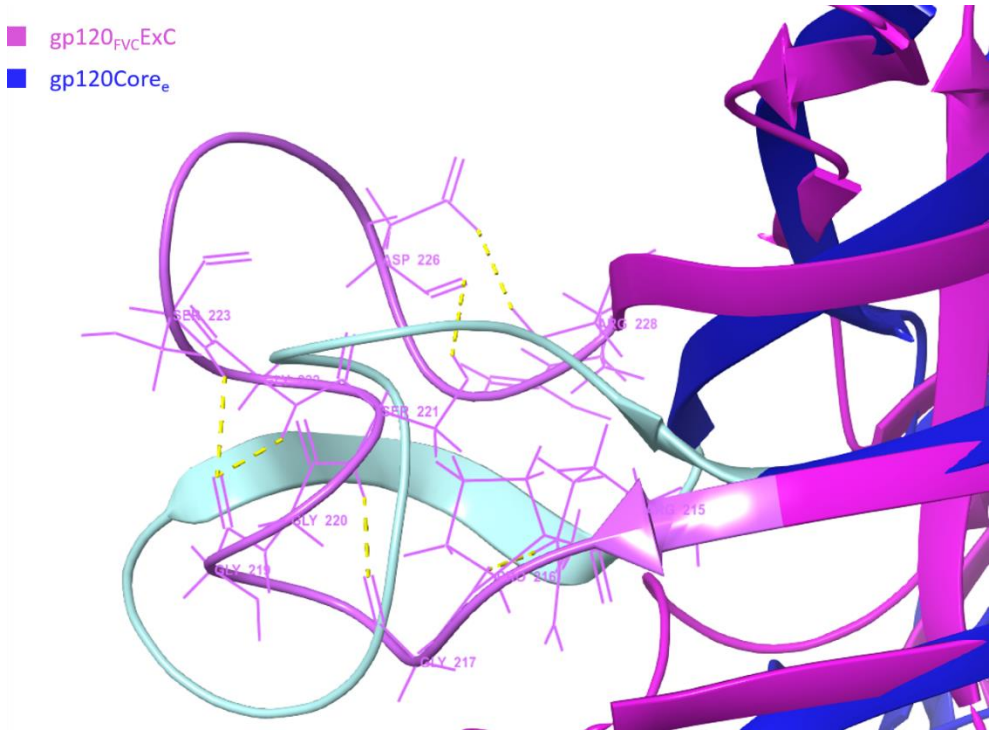

Figure 3.15: Visual representation of the V3 substitution region of gp120$_{FVC}$ExC (light purple) and gp120Core$_e$ (light blue), with key residues with a difference in fluctuations highlighted. The key residues of gp120Core$_e$ remain in an open configuration while the corresponding residues of gp120$_{FVC}$ExC persist in a coiled conformation. Yellow dashes indicate formed hydrogen bonds in gp120$_{FVC}$ExC that are responsible for the fluctuation. Key residues Asp 226 and Gly 217 in gp120$_{FVC}$ExC are absent in gp120Core$_e$.

3.7.4. Additional residues in the V4 region of gp120Core$_e$ compared to gp120$_{FVC}$ExC stabilize the V4 loop region of gp120Core$_e$

The V4 loop region exhibits sequence differences between the gp120$_{FVC}$ExC and gp120Core$_e$ proteins in that gp120Core$_e$ contains more residues than gp120$_{FVC}$ExC, potentially due to gp120$_{FVC}$ExC being a FV consensus sequence which has been found to have shorter V1-V4 loops (Rademeyer et al., 2007). Of all the regions that experience significant differences in RMSF between the gp120$_{FVC}$ExC and gp120Core$_e$ proteins, the V4 region is the only region of gp120$_{FVC}$ExC (263–266) that appears to experience more fluctuation than that of gp120Core$_e$ (Figure 3.16). This is potentially due to the additional residues present in gp120Core$_e$ that flank the residues in this region, increasing the stability of this region (Figure 3.16). When visualized through MD simulation the residues that exhibit fluctuation in the V4 loop of gp120$_{FVC}$ExC have a more open conformation that fluctuates more readily than the corresponding residues of gp120Core$_e$, that remain in a coiled configuration for larger portions of the simulation. The coiled configuration appears to be primarily due to the presence of a hydrogen bond between Ser-255 and Leu-257 that exists in gp120Core$_e$ but not gp120$_{FVC}$ExC. This could be due to an Asp to Lys mutation in gp120$_{FVC}$ExC that lies between these residues. It should also be noted that the remaining residues that comprise the V4 loop in gp120Core$_e$ that are absent in gp120$_{FVC}$ExC contain β-sheets that have hydrogen bonds to further stabilize the adjacent region and potentially the region in which the significant difference was found between the gp120 proteins.

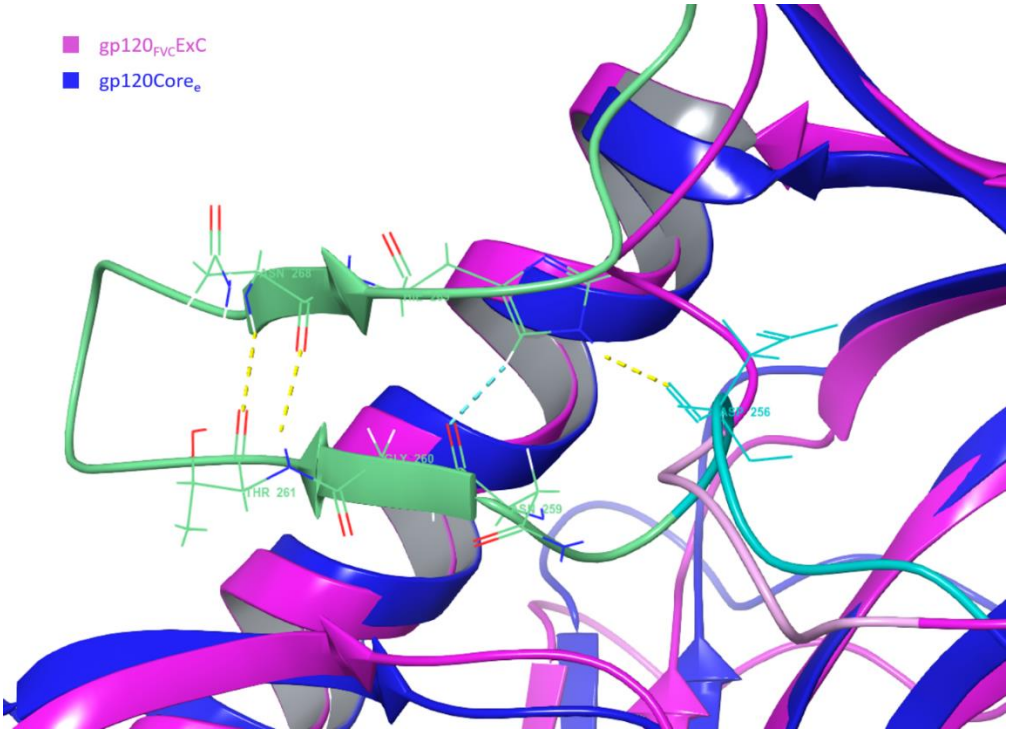

Figure 3.16: Visual representation of the V4 loop shows the more coiled conformation of gp120Core$_e$ (light blue) compared to the more open and mobile conformation of gp120$_{FVC}$ExC (light purple), leading to increased fluctuations of the gp120$_{FVC}$ExC residues. The residues for which both gp120$_{FVC}$ExC and gp120Core$_e$ overlap, appear to be stabilized by a hydrogen bond in gp120Core$_e$ (yellow dash) between Asp 256 and His 269. The remaining portion of the V4 regions are shown for gp120Core$_e$ (green) and gp120$_{FVC}$ExC (dark purple). Further stabilization is likely due to hydrogen bonds present between the β-sheets in gp120Core$_e$.

3.7.5. V5 loop region of gp120Core$_e$ has increased mobility due to a larger V5 region compared to gp120$_{FVC}$ExC

Unlike the V4 loop where the addition of residues appears to stabilize a region, the presence of additional residues in the V5 for gp120Core$_e$ appear to increase the mobility of this region compared to gp120$_{FVC}$ExC (Figure 3.17). The identified residues of the V5 region of gp120Core$_e$ vacillate between a coiled and uncoiled structure throughout the simulation. This change in coiling appears to be a result of the additional residues in gp120Core$_e$ producing more hydrogen bonds

between the residues of the V5 region that cause the structure to coil and uncoil as these bonds form and break throughout the simulation. This appears to be the reason behind the increased fluctuations of these residues in gp120Core$_e$ as the corresponding residues in gp120$_{FVC}$ExC remain in a more open but stable conformation (Figure 3.17).

These results indicate that the gp120$_{FVC}$ExC structure is potentially overall less dynamic than gp120Core$_e$ in the variable regions of the protein. However, there are many small regions (Figure 3.12) with few residues found throughout gp120$_{FVC}$ExC that appear to undergo larger degrees of fluctuations outside of these key variable regions (V1, V2, V3, V5). Ultimately, gp120$_{FVC}$ExC appears to be more stable with key regions experiencing less conformational fluctuations and change in dynamics of these residues, except for the V4 loop region.

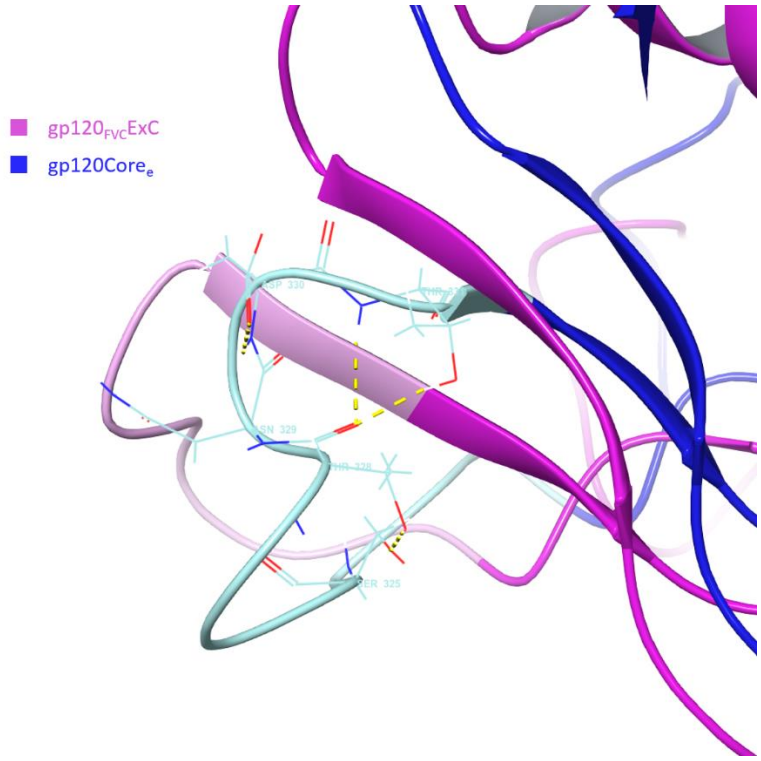

Figure 3.17: Visual representation of the V5 loop shows a coiling and uncoiling of the key residues in gp120Core$_e$ (blue), which appears in contrast to the open but less mobile residues of gp120$_{FVC}$ExC (purple), increasing the fluctuation of the gp120Core$_e$ residues.

3.7.6. Introduction of the S60C mutation minimally decreases conformational dynamics of unbound 2dCD4

The addition of the S60C mutation to 2dCD4 appears to slightly reduce the conformational dynamics of the 2dCD4 protein compared to 2dCD4^{WT} by 0.2 Å (Figure. 3.18). The average RMSD difference is 0.2035 Å larger for 2dCD4^{WT} compared to 2dCD4^{S60C} suggesting a reduction in conformational mobility induced by the S60C mutation in 2dCD4.

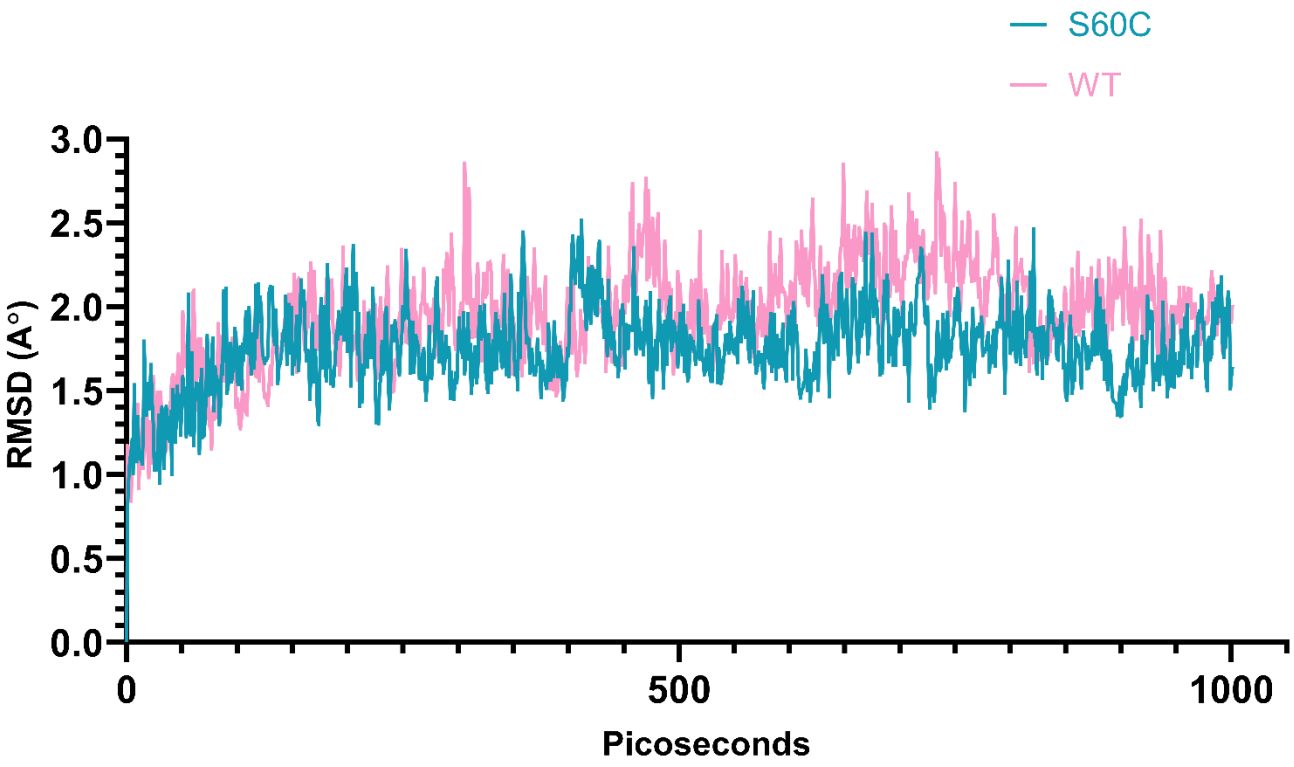

Figure 3.18: RMSD of 2dCD4^{S60C} (blue) and 2dCD4^{WT} (pink) alpha carbons averaged for the residues of the proteins per frame of the simulation. A global reduction of RMSD is observed in the 2dCD4^{S60C} structure, particularly as the simulation progresses and equilibrates.

Comparing the RMSF of the 2dCD4^{WT} and 2dCD4^{S60C} structures shows a slight overall reduction of approximately 0.2 Å in the mobility of 2dCD4^{S60C} when the Serine to Cysteine mutation is introduced (Figure 3.19) and the same trend in reduction of mobility is observed in the RMSD analysis (Figure 3.18). Discerning further variations becomes precarious due to the scale of

changes observed. However, there is a visible reduction in mobility of the S60C mutation residue (Figure 3.19).

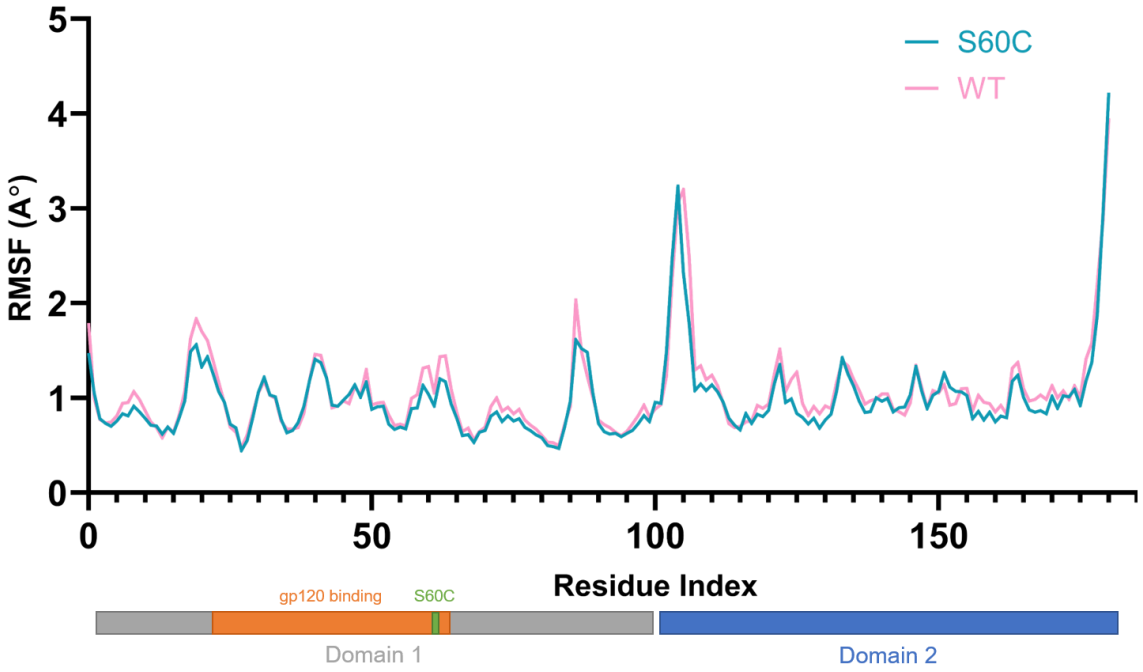

Figure 3.19: The RMSF of 2dCD4^{S60C} (purple) and 2dCD4^{WT} (blue) residues over a 50 ns MD simulation shows a decrease in mobility of the modified 2dCD4^{S60C} compared to 2dCD4^{WT}. The domain bar depicts CD4 domain 1 (grey) containing the gp120 binding region (orange) and the location of the S60C mutation (green). CD4 domain 2 is depicted in blue. A reduction is also noted at the location of the Cysteine substitution for 2dCD4^{S60C} (green).

The percentage difference of the ratio in RMSF (30% cut-off) identified 2 regions to investigate that show large scale differences in fluctuation. It should be noted that a reduction (negative) percentage indicates less mobility of the 2dCD4^{S60C} structure when compared to the 2dCD4^{WT} structure (Figure 3.20). This indicates that all the regions of significant difference experience a reduction in dynamics. However, a limited number of residues exhibit fluctuations and as such visualizing the structure in the MD simulations provides limited insight. The regions that experience significant differences are those in domain 2 of the 2dCD4 structure, with residues

Asp-105, Thr-106 and Ser-124, Ser-125 experiencing the largest change, as a reduction in fluctuation for 2dCD4^{S60C} compared to 2dCD4^{WT}.

Analysis of the residues outside of the 30% difference threshold provided further insight into the changes and potential structural effects of the S60C mutation. Residues Ser-19, Gln-20, and Lys-21 reside in the gp120 binding domain of CD4 which is less mobile in the 2dCD4^{S60C} variant (Figure 3.20). In the gp120 binding domain Arg-59, Cys/Ser-60, Trp-62, and Asp-63 also have reduced mobility of these residues (Figure 3.20). Importantly the S60C mutation has reduced mobility and potentially stabilized the key residue involved in the disulphide bond in the complex. There exists a key overlap between the gp120 binding domain (residues 25-64) and that of the residues involved in binding MHC II molecules (25-85). Of this overlap Arg-59, Cys/Ser-60, Trp-62, and Asp-63 residues experience a significant reduction in mobility for the S60C variant, as with residues in the gp120 binding domain. Overall, however, the 2dCD4^{S60C} structures shows limited difference in structure of the protein in its entirety compared to 2dCD4^{WT} as is expected due to a single conserved point mutation in the structure.

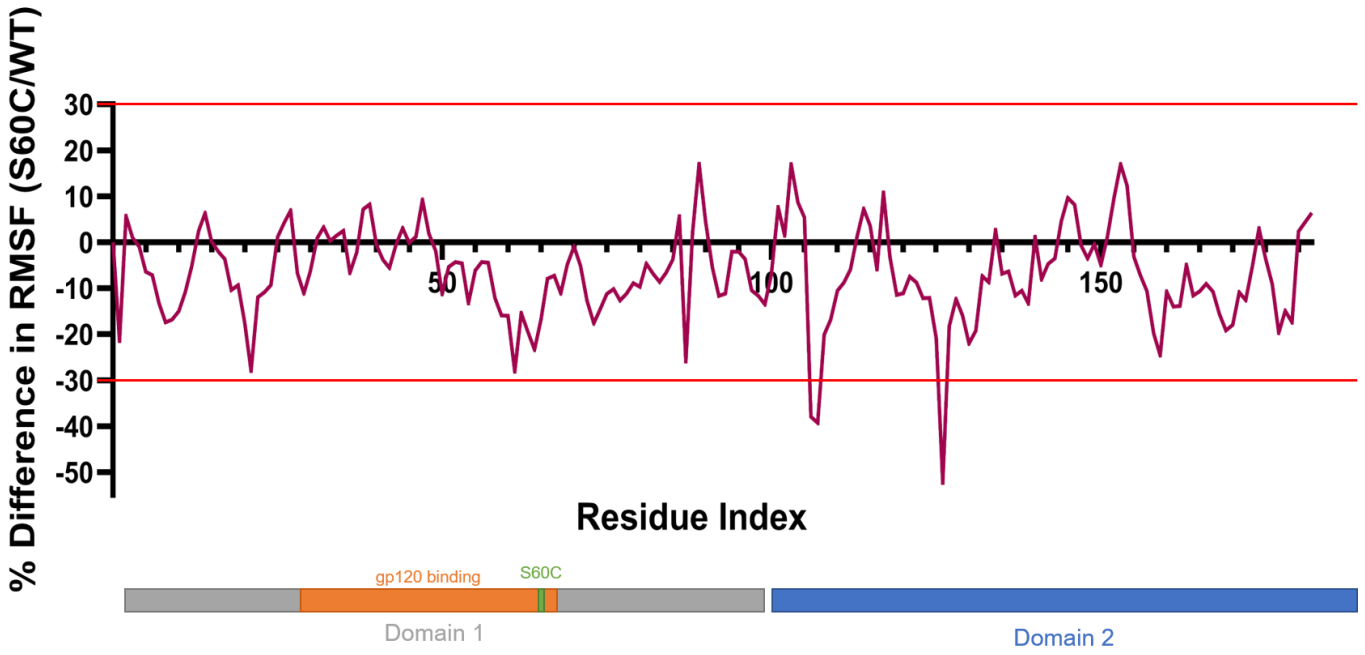

Figure 3.20: The percentage difference in RMSF between the 2dCD4^{S60C} and 2dCD4^{WT} residues show structural variation between specific residues of which the S60C mutation (green) causes a reduction of flexibility at this residue compared to 2dCD4^{WT}. Notable changes in a reduction of conformational flexibility is also observed in domain 2 for 2dCD4^{S60C}. A positive percentage difference indicates that there is more flexibility of residues for 2dCD4^{S60C} than that of 2dCD4^{WT}. A negative percentage difference indicates that 2dCD4^{WT} residues are more conformationally flexible. The domain bar depicts domain 1 (grey) containing the gp120 binding region (orange) and the location of the S60C mutation (green). Domain 2 is depicted in blue.

3.7.7. The 2dCD4 S60C mutation increases the conformational dynamics of gp120$_{FVC}$ExC – 2dCD4 complex

The RMSF for the gp120$_{FVC}$ExC – 2dCD4$^{S60C/WT}$ complexes were determined and compared to ascertain the residues or regions that experienced a significant difference in fluctuation due to the putative disulphide bond introduced with the S60C mutation of 2dCD4 in the gp120$_{FVC}$ExC – 2dCD4^{S60C} complex (Figure 3.21). This provides some insight into the nature of the difference in the dynamics of the two complexes, such that residue 214-224, gp120$_{FVC}$ExC – 2dCD4^{S60C} exhibits larger fluctuation of these residues compared to those of the gp120$_{FVC}$ExC – 2dCD4^{WT} complex

(Figure 3.21 A). This region along with the overall trend may indicate a higher overall conformational flexibility resulted from the introduction of the putative disulphide bond between Cys-113 of $gp120_{FVC}ExC$ and Cys-60 of $2dCD4^{S60C}$.

The percentage difference of the RMSF of the residues (cut-off of 30%) highlighted regions that experience significant change in fluctuation as a result of the putative disulphide bond (Figure 3.21 B). Two regions that experience large fluctuations and consist of more than 3 residues were identified and reside around the N-terminus (38-43) and the V3 region (214-224). Other regions that experience differences below the cut-off or consisted of few residues but fell within functional regions included residue Cys-113 in $gp120_{FVC}ExC$, which shows a reduction in fluctuation due to the introduction of the putative disulphide bond in the $gp120_{FVC}ExC – 2dCD4^{S60C}$ complex (Figure 3.21 B). The residues of the V4 and V5 loops of the S60C complex also display a reduction in residue fluctuation compared to those of $gp120_{FVC}ExC – 2dCD4^{WT}$. The 2dCD4 component of the complexes experience less change in residue fluctuation than those of the $gp120_{FVC}ExC$ component likely due to the putative disulphide bond. However, some variation is observed in the gp120 binding domain as well as a large reduction in mobility of domain 2 in the S60C 2dCD4 component of the complex. The overall trend of the RMSD and the RMSF shows a larger conformational flexibility and mobility of residues with the introduction of the putative disulphide bond between $gp120_{FVC}ExC$ and $2dCD4^{S60C}$ proteins.

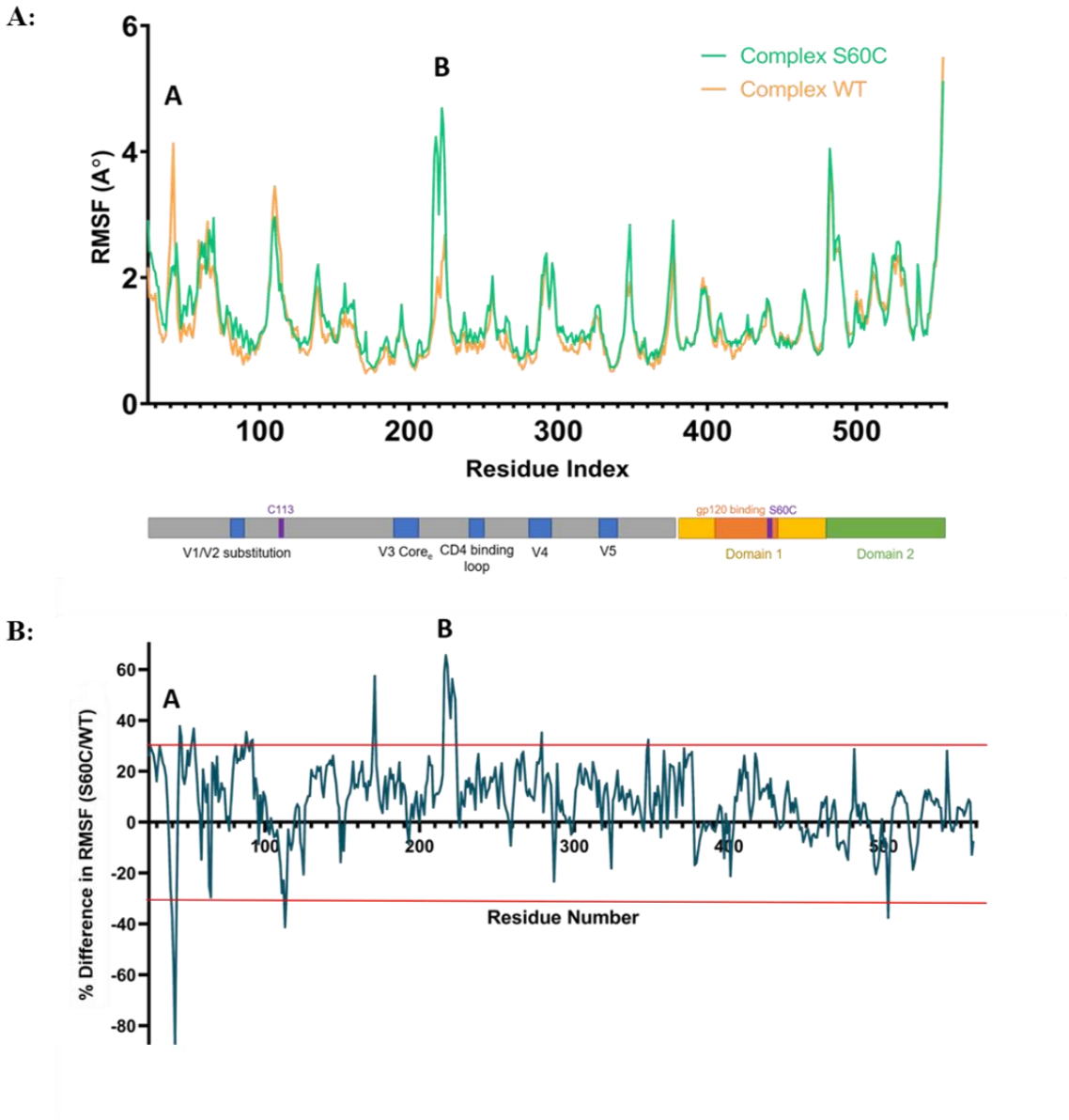

Figure 3.21. A: The RMSF of the residues in the $gp120_{FVC}$ExC-2dCD4^{S60C} complex (green) indicates a large change in conformational flexibility around residue 200 after the V3 loop and between the V3 region and the V4 loop, as well as compared to $gp120_{FVC}$ExC – 2dCD4^{WT} (orange). A reduction is also observed around residue 40 prior to the V1/V2 substitution region. Importantly a reduction is observed at the C113 region – which forms part of the introduced putative disulphide bond. Globally there are minimal overall differences between the two structures. However, $gp120_{FVC}$ExC-2dCD4^{WT} displays less flexibility than $gp120_{FVC}$ExC-2dCD4^{S60C}. **B:** The percentage difference in RMSF between $gp120_{FVC}$ExC-2dCD4^{WT} and $gp120_{FVC}$ExC-2dCD4^{S60C} shows many regions of structural variation between the two structures. Notably around C113, the region that forms part of the putative disulphide bond in conjunction with the S60C

mutation is less dynamic. Additionally, the region that resides between the V3 and the CD4 binding loop of $gp120_{FVC}ExC$ is more dynamic. Variation is also observed in the gp120 binding domain of $2dCD4^{S60C}$ which is more dynamic than that of $2dCD4^{WT}$.

Visual representation of the two regions determined by the percentage difference in the RMSF of the residues for $gp120_{FVC}ExC$ - $2dCD4^{WT}$ and $gp120_{FVC}ExC$ - $2dCD4^{S60C}$ were obtained from the MD simulations to better illustrate the changes in dynamics of these regions (Figures 3.24 and 3.25). These images are merely single snapshots of the MD simulations, and are used to illustrate the differences, but do not account for all the fluctuations observed. The image below shows the regions that were investigated in the context of the aligned protein structures (Figure 3.22).

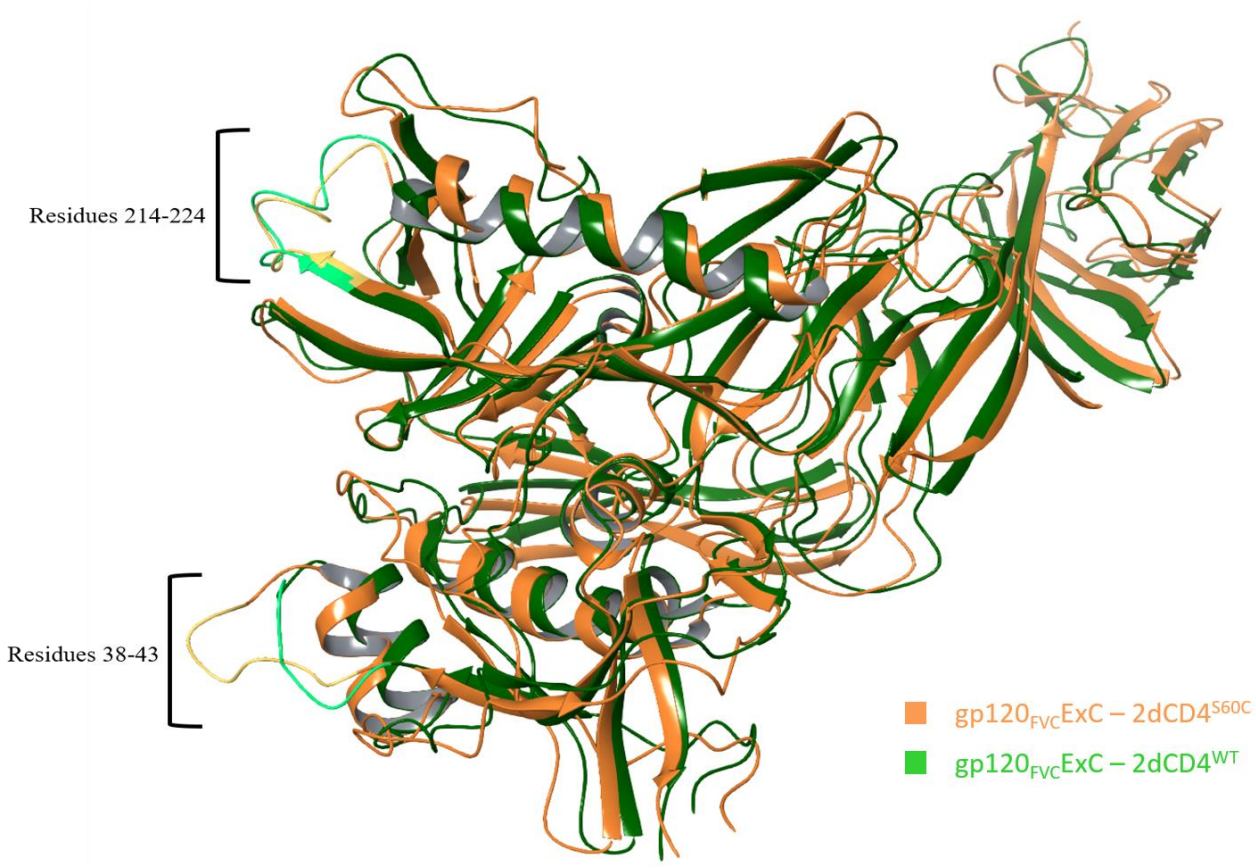

Figure 3.22: Superimposed $gp120_{FVC}ExC$ – $2dCD4^{WT\ (green)}$ and $S60C\ (orange)$ complexes with the regions visually investigated highlighted and labelled.

3.7.8. Residues 38-43 of gp120$_{FVC}$ExC – 2dCD4^{S60C} show a large reduction in fluctuation compared to the WT variant of the complex

Residues 38 – 43 of the gp120$_{FVC}$ExC - 2dCD4^{S60C} complex show a drastic 85% (Figure 3.21 B) reduction in fluctuation compared to the corresponding residues of gp120$_{FVC}$ExC - 2dCD4^{WT}. Figure 3.23 shows that while gp120$_{FVC}$ExC - 2dCD4^{S60C} remains relatively immobile due to a constant hydrogen bond between Lys 40 and Asn 48, the loop of these residues in gp120$_{FVC}$ExC - 2dCD4^{WT} fluctuates more readily in the y-axis of the reference image and appears to make coil like structures more readily. This fluctuation spans the distance to the gp120$_{FVC}$ExC – 2dCD4^{S60C} loop and the depicted position of gp120$_{FVC}$ExC – 2dCD4^{WT}.

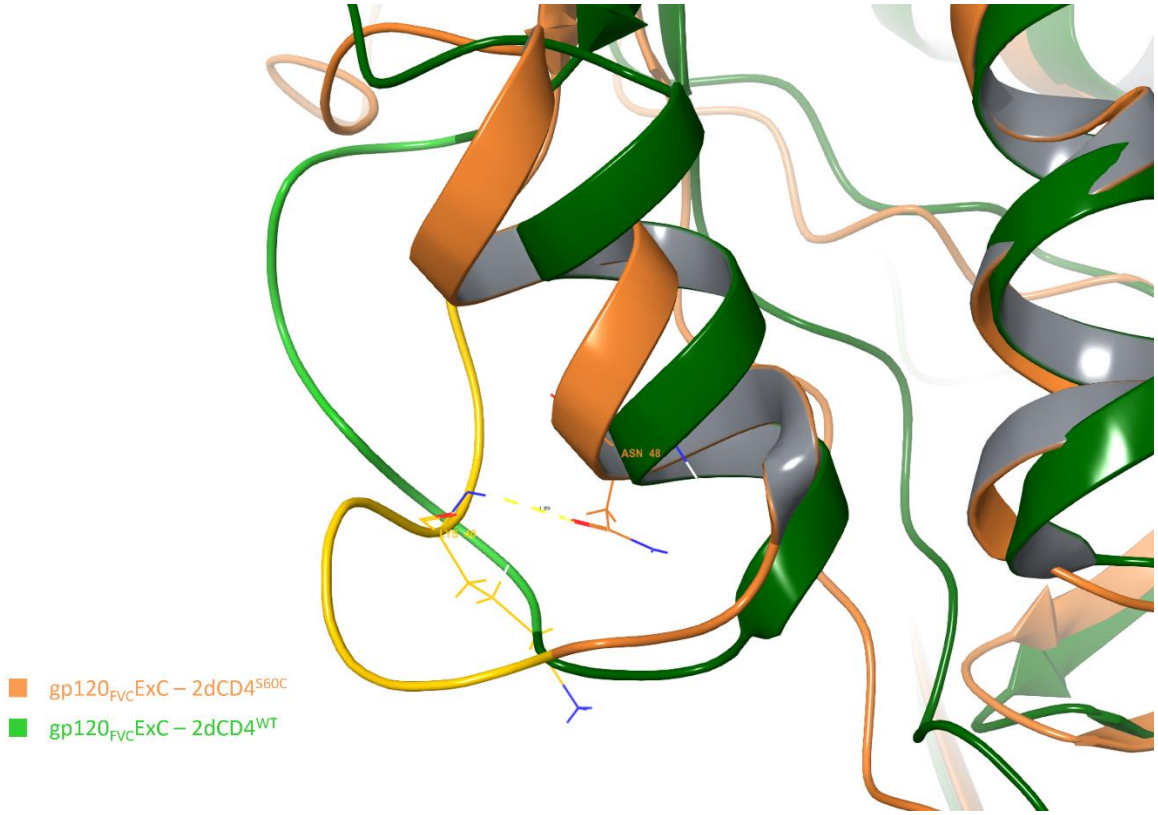

Figure 3.23: Visual representation of the residue 38-43 loop shows a movement along the y-axis of the image for the key residues in gp120$_{FVC}$ExC – 2dCD4^{WT} (green) while the loop of gp120$_{FVC}$ExC – 2dCD4^{S60C} (orange) remains in the depicted position with minor movements along the y-axis. The bond stabilizing the loop of gp120$_{FVC}$ExC – 2dCD4^{S60C} is indicated with a yellow dash between Lys 40 and Asn 48.

3.7.9. Residues 214-224 in the V3 loop of $gp120_{FVC}ExC - 2dCD4^{S60C}$ have increased mobility compared to $gp120_{FVC}ExC - 2dCD4^{WT}$

Residues 214-224 in the V3 loop show the largest increase in residue fluctuation of $gp120_{FVC}ExC - 2dCD4^{S60C}$ compared to the corresponding residues of $gp120_{FVC}ExC - 2dCD4^{WT}$ (Figure 3.24). This increase in fluctuations for this region is also the change that encompasses the largest number of adjacent residues. Visualizing these residues during MD simulation showed that the residues that make up this loop are highly mobile for $gp120_{FVC}ExC - 2dCD4^{S60C}$. These residues fluctuate frequently between a coiled and an open conformation in $gp120_{FVC}ExC - 2dCD4^{S60C}$. The corresponding residues in $gp120_{FVC}ExC - 2dCD4^{WT}$ often encompass an α-helix that stabilizes the residues and reduces the fluctuations of flanking residues that is present throughout large portions of the simulation. The coiling and uncoiling of the $gp120_{FVC}ExC - 2dCD4^{S60C}$ residues appear to be the reason behind the increased fluctuation of these residues.

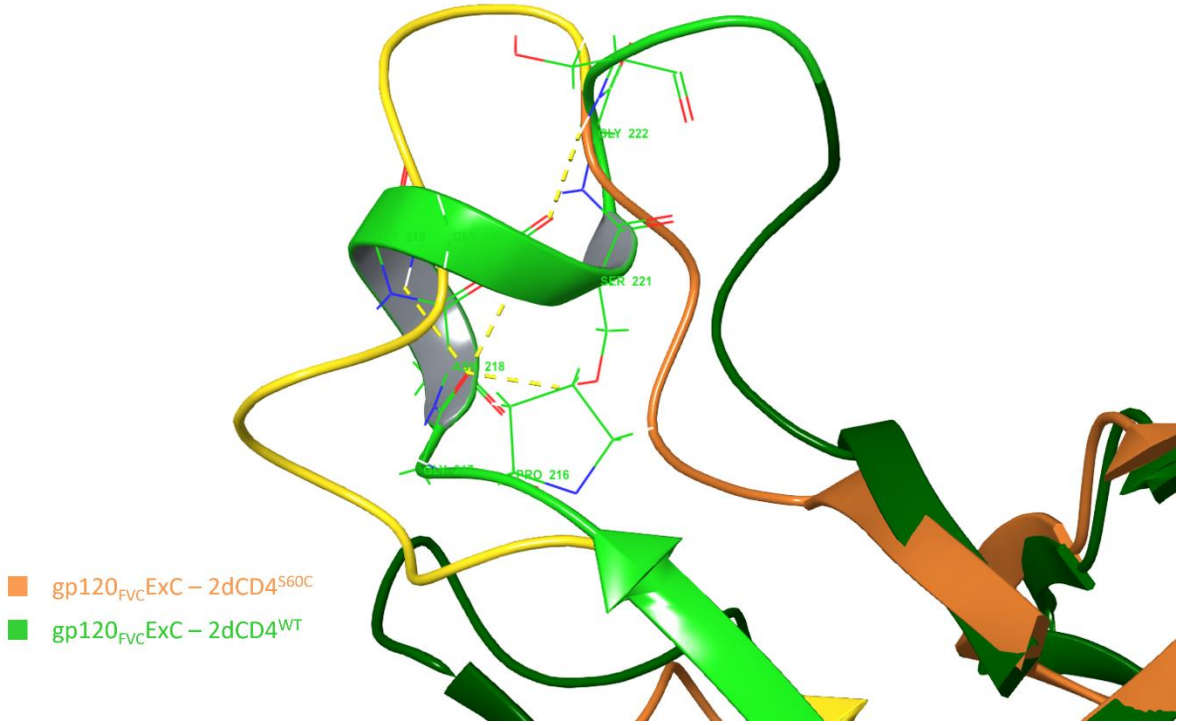

Figure 3.24: Visual representation of residues 214-224 of the V3 loop shows a frequent and persistent coiling and uncoiling in $gp120_{FVC}ExC - 2dCD4^{S60C}$ (orange) while the loop of $gp120_{FVC}ExC - 2dCD4^{WT}$ (green) remains in a predominantly α-helix conformation.

Analysis of the Secondary Structural Elements (SSE) over the period of the simulation corroborate the presence of the α-helix present in $gp120_{FVC}ExC - 2dCD4^{WT}$ that is absent in $gp120_{FVC}ExC - 2dCD4^{S60C}$ at residues 214-224 (Figure 3.25). This analysis of SSE's represents the secondary structures (α-helix, β-sheets) in the MD simulation that are present for over 70% of the simulation time. This causes an increase in mobility of these residues for $gp120_{FVC}ExC\text{-}2dCD4^{S60C}$ as observed in the RMSF (Figure 3.21 A). Interestingly, while the introduction of the S60C mutation in unliganded 2dCD4 appears to reduce the overall conformational flexibility, the putative disulphide bond introduced due to the mutation in the $gp120_{FVC}ExC\text{-}2dCD4^{S60C}$ complex increases the conformational dynamics compared to the $gp120_{FVC}ExC\text{-}2dCD4^{WT}$ complex.

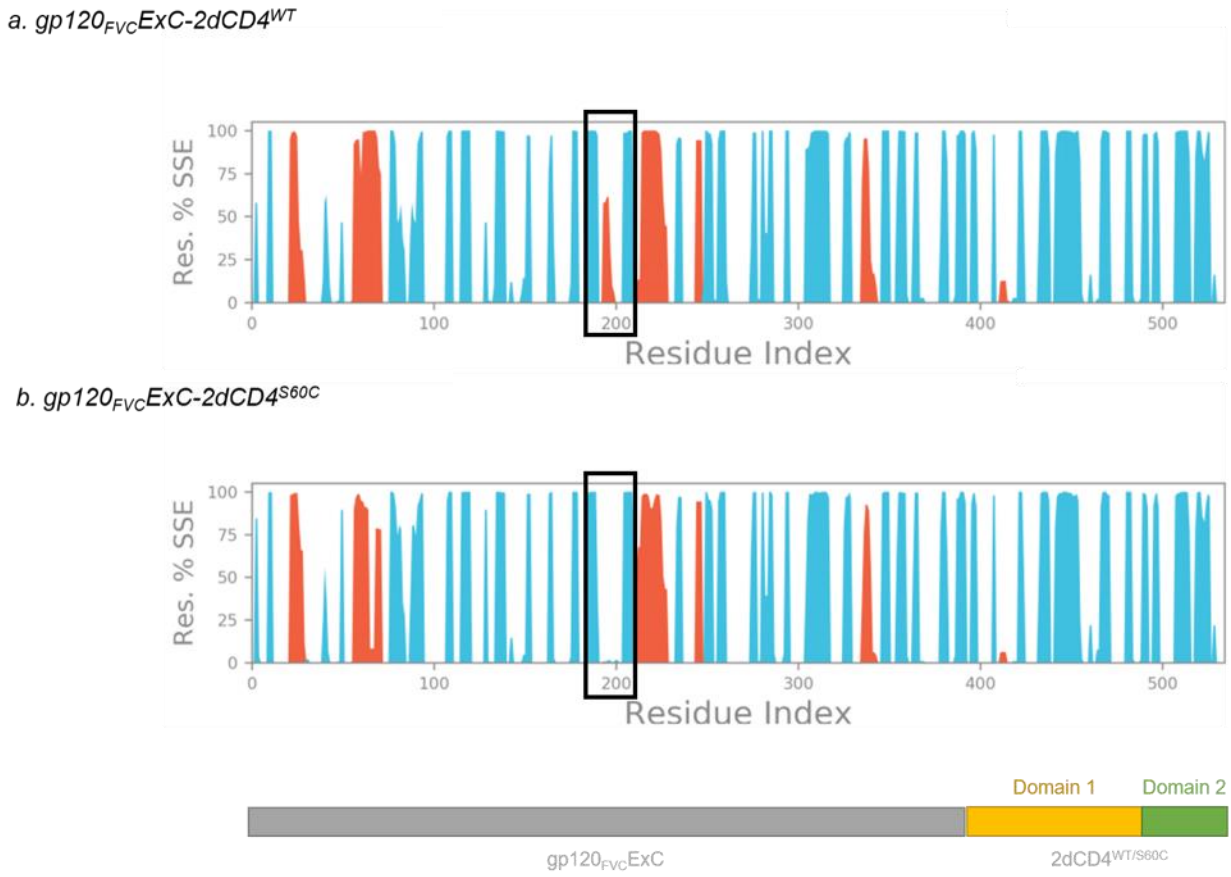

Figure 3.25: Secondary structural elements of $gp120_{FVC}ExC\text{-}2dCD4^{S60C/WT}$ complexes as determined by MD simulations of the Maestro (Schrodinger) platform. There is a reduction in an α-helix in the $gp120_{FVC}ExC\text{-}2dCD4^{S60C}$ complex at residues 190-200 (b) compared to the $gp120_{FVC}ExC\text{-}2dCD4^{WT}$ complex (a). Blue bars represent β-sheets, and red bars indicate α-helices. The domain bar below the structure represents the $gp120_{FVC}ExC$ (grey) component of the complex while domain 1 of 2dCD4 is indicated in yellow and domain 2 of 2dCD4 is indicated in green. The black square indicates the region where the α-helix change happens due to the introduction of the putative disulphide bond.

4. Discussion

It is well recognized that HIV-1 diversity presents a significant challenge to creating a global vaccine. In an attempt to address this problem, extensive characterization and comparisons of individual clinical virus, founder virus, consensus, and mosaic sequences have been undertaken and tested in various platforms. To this end, the current study described conditions for the successful crystallization of an HIV-1 subtype C FV consensus sequence, designated $gp120_{FVC}ExC$, as well as an MD simulation comparison of $gp120_{FVC}ExC$ to the well characterized $gp120Core_e$. Through these MD simulations, regions that experience a large difference in the fluctuations of the residues, which are primarily focussed around the V1, V2, V3, V4, and V5 variable loop regions of gp120 were identified. Additionally, differences observed in the CD4 binding loop were described, as a step towards better understanding the dynamics of transmitted founder viruses and the observed enhanced immunogenicity of a $gp120_{FVC}ExC$ – $2dCD4^{S60C}$ vaccine immunogen.

4.1.Sequence modifications increase the probability of $gp120_{FVC}ExC$ crystallization

The ability to successfully crystallize unliganded $gp120_{FVC}ExC$ was likely the direct result of the modifications adapted from *Kwon et al.* that modified the V1, V2, and V3 regions of gp120 to reduce conformational heterogeneity. The journey undertaken by *Kwon et al* started with the design of the $gp120Core_{min}$ sequence (Kwon et al., 2012). This sequence contained large reductions in the V1, V2, and V3 regions. The V1 and V2 regions were largely replaced with a -GAG- sequence while the V3 region was replaced with a -GG- sequence. They found that the $gp120Core_{min}$ was able to form protein crystals when bound to 2dCD4 but were unable to form crystals unliganded. They subsequently designed the first gp120 sequence that was able to generate unliganded crystals. This gp120 was coined $gp120Core_e$ due to an extended linker sequence in the V3 region -GGSGSG- they employed. The $gp120Core_e$ sequence also contains a shorter -GG- sequence in the V1 and V2 regions. The designed $gp120_{FVC}ExC$ sequence (Figure 3.1) incorporated motifs/domains from both $gp120Core_{min}$ and $gp120Core_e$. We adopted the V1 and V2 modifications from $gp120Core_{min}$ along with retaining residues that were omitted since the study end goal is to crystallize the $gp120_{FVC}ExC$ – $2dCD4^{S60C}$ complex. The reason for the

retention of the -GAG- substitution was due to the gp120Core$_{min}$ protein's ability to generate 2dCD4 liganded crystals. The additional residues that were retained in the gp120$_{FVC}$ExC V1 and V2 region is due to residue 113 being critical in the putative disulphide bond of the gp120$_{FVC}$ExC – 2dCD4^{S60C} complex. The V3 modification from gp120Core$_e$ was included in an attempt to increase the ability to generate unliganded gp120$_{FVC}$ExC crystals to compare to existing unliganded gp120 structures. It appears that the linker sequence modification of the V3 region in gp120Core$_e$ and gp120$_{FVC}$ExC are integral to the generation of unliganded protein crystals. Additionally, the gp120$_{FVC}$ExC V1 and V2 region is more mobile than observed for gp120Core$_e$, as determined by MD simulation analysis when observing the RMSF of these regions (Figure 3.11). While this increase in mobility may have reduced the ability to crystallize gp120$_{FVC}$ExC, it was in no way detrimental due to our ability to generate protein crystals and speaks to the importance of the V3 -GGSGSG- substitution. It is also possible that the lower RMSD observed for the gp120$_{FVC}$ExC compared to gp120Core$_e$, likely is due to a lower RMSF in the V3 and V5 regions, which may have mitigated some of the increased fluctuations introduced by the additional residues in the V1/V2 region (Figure 3.10 and 3.11). Reductions in conformational heterogeneity have been found to be an integral factor in increasing the probability of crystallizing proteins, particularly for conformationally flexible proteins like gp120 (Kwong et al., 1999). Collectively, these sequence modifications, particularly to the V3 region have likely been integral in the ability to produce unliganded gp120$_{FVC}$ExC crystals.

4.2. Post translational modifications and protein species homogeneity of gp120$_{FVC}$ExC further increase the probability of crystallization

The generation of protein crystals, even with a modified sequence is notoriously difficult (Kwong et al., 1998), and as such other modifications and methods are employed to increase conformational heterogeneity and purity of the intended protein. The methods employed in the current study were likely equally integral in the success of producing gp120$_{FVC}$ExC crystals.

As gp120 is a largely glycosylated protein, and critical epitopes are protected by a glycan shield (Berndsen et al., 2020), it naturally has a large amount of conformational heterogeneity due to the mobility and multitude of orientations of the glycans. The first step in successful crystallization

was to reduce the formation of N-linked glycans present on $gp120_{FVC}$ExC. This was achieved through the expression of $gp120_{FVC}$ExC in GnTi-deficient cells that lack N-acetylglucosaminyltransferase activity that prevents the formation of complex N-glycans (Reeves et al., 2002). These truncated glycans produced by the GnTi-deficient cells leave the mannose glycans on the surface of $gp120_{FVC}$ExC unaffected, allowing the subsequent purification of $gp120_{FVC}$ExC using lectin affinity chromatography. The lack of complex N-linked glycans importantly facilitate more homogenous deglycosylation through Endoglycosidase H (Endo H) by targeting now exposed N-linked glycans between two N-acetylglucosamine subunits (GlcNAc) proximal to the asparagine residue of the subunit (Freeze and Kranz, 2010). During the deglycosylation of $gp120_{FVC}$ExC, we observed two distinct species of deglycosylated $gp120_{FVC}$ExC between 32 and 46 kDa, which is an overall reduction of ~26 kDa in protein weight (Figure 3.5).

It is unclear why the deglycosylation of $gp120_{FVC}$ExC produced two distinct deglycosylated species instead of uniform deglycosylation. However, a few possibilities may account for this phenomenon. Variable post-translational modification (PTM) of the individual proteins during expression may account for subsequent variable cleavage by Endo H and is supported by the width and smearing observed of expressed $g120_{FVC}$ExC (Raska et al., 2010). The variable cleavage resulting from potentially variable PTM may result in altered accessibility of GlcNAc residues by Endo H between the proteins, either due to the presence of multiple cleavage sites or steric interference of neighbouring glycans preventing the access to cleavage sites. The order of the cleavage may be responsible for altered deglycosylation. Should Endo H molecules approach the glycan (Figure 2.4) and cleave certain sites, this may occlude subsequent sites from being cleaved either by accessibility or structural rearrangement. This could explain the equilibrium shift to a more complete deglycosylation as observed in Figure 3.7, where a subset of glycans are cleaved unfavourably preventing complete deglycosylation. A strategy employed in the deglycosylation of proteins is the denaturation of the protein to increase the accessibility of Endo H to GlcNAc cleavage sites (Krenkova et al., 2013), and as such it is possible that conformational differences in the gp120 protein structure may account for variable accessibility of Endo H residues. To investigate this phenomenon, it is possible to perform Mass spectrometry to investigate the nature of the glycans on our sequence for both the more and less deglycosylated species that may validate the proposed hypotheses.

The conditions for successful crystallization of $gp120_{FVC}ExC$ were found in an under-oil crystallization method using the Morpheus Index kit (Figure 3.9). Attempts at using the vapour diffusion method, as well as the Hampton Index kit to produce $gp120_{FVC}ExC$ crystals proved futile. The vapour diffusion method produced primarily needles, resulted in aggregation, or clear drops (Figure 3.8). The likely benefit of the under-oil method is the rate of evaporation of the mother liquor resulting in an altered rate of increased protein concentration, leading to the formation of protein crystals (Brumshtein et al., 2008). Additionally, attempts at producing $gp120_{FVC}ExC$ crystals using a seeding method for both the Hamptons and Morpheus kits were unsuccessful. It appears that an initial lower concentration of our protein, or the absence of a source of nucleation is more suited to the formation of $gp120_{FVC}ExC$ protein crystals, regardless of whether a vapour diffusion or under-oil method was used.

The two conditions found in the Morpheus kit using under-oil diffusion that were successful at producing protein crystals were successful over a wider range of protein concentrations. They produced numerous similarly shaped birefringent crystals that were determined to be protein crystals. While these crystals appear to be suited to X-ray diffraction, they could not be upscaled or tested using diffraction due to accessibility of resources and time constraints due to the COVID-19 pandemic. As such we are unable to state whether these conditions were ultimately successful at producing diffracting $gp120_{FVC}ExC$ crystals.

Ultimately the success in producing crystals of the unbound $gp120_{FVC}ExC$ (which has been notoriously troublesome) provides future avenues to further investigate the nature of a founder virus gp120 sequence as compared to a non-founder virus gp120 structure. This could provide valuable insights to relevant Env motifs/domains critical for successful HIV-1 transmission and establishment of productive infection in a new host. It also provides a measure of confidence in the future ability to crystallize and compare the $gp120_{FVC}ExC–2dCD4^{S60C}$ complex and supplement the $gp120_{FVC}ExC$ and $gp120_{FVC}ExC – 2dCD4^{S60C}$ structures with data produced in this study. This would identify critical epitopes responsible for the enhanced immunogenicity of these complexes in small animals and non-human primates.

4.3.MD investigation of gp120$_{FVC}$ExC compared to gp120Core$_e$ shows differences in the variable loop regions

Comparative MD simulations of gp120$_{FVC}$ExC and gp120Core$_e$ (PDB ID: 3TGR) provided insights into the structure of gp120$_{FVC}$ExC as well as conformational differences that exist in gp120$_{FVC}$ExC through the analysis of the residue fluctuations compared to gp120Core$_e$ (Figures 3.10 to 3.17). The bulk of the difference in fluctuations observed between gp120$_{FVC}$ExC and gp120Core$_e$ were observed primarily in the V1, V2, V3, V4, and V5 regions (Figure 3.12). Observations of the CD4 binding loop have also been included despite the differences not meeting our criterion for selection due to its functional relevance at establishing infection through the binding of CD4 (Kwong et al., 1998). The differences observed in the V1 and V2 regions of gp120$_{FVC}$ExC compared to gp120Core$_e$ have already been discussed above (Section 4.1). Due to the nature of these differences being introduced to aid crystallization, it does not allow for reasonable comparison between the structures from an *in vivo* standpoint. The focus of this discussion therefore surrounds the comparable V3, V4, and V5 regions of gp120$_{FVC}$ExC compared to gp120Core$_e$.

Of these regions that have been identified, the V3 region of gp120$_{FVC}$ExC experiences the largest difference in fluctuation compared to gp120Core$_e$ (Figure 3.12 and 3.15). This region has undergone modification in both gp120$_{FVC}$ExC and gp120Core$_e$ to substitute a large portion of the loop with a -GGSGSG- sequence. However, the gp120$_{FVC}$ExC sequence differs from the gp120Core$_e$ surrounding this substitution region with a point mutation proximal and distal to the substitution. Through analysis of the MD simulations, it was observed that these point mutations: Asn-217 to Gly and Asn-266 to Asp, allow for the formation of more hydrogen bonds that result in a more coiled structure compared to gp120Core$_e$ that retains a looser conformation (Figure 3.15). The differences observed for this region does not appear to have been detrimental in the crystallization of gp120$_{FVC}$ExC and the reduction in fluctuations may have allowed for more conformational heterogeneity. The possibility also exists that the differences between the amino acids of the protein sequences in the region that resides between the V3 loop and the CD4 binding loop may have contributed to reducing the fluctuations of the V3 region in gp120$_{FVC}$ExC compared to gp120Core$_e$ (Figure 3.1). However, no analysis has provided insight as to the differences that

may be induced by these residues. These two point mutations (Asn-217 to Gly and Asn-266 to Asp) that exist in $gp120_{FVC}ExC$ compared to $gp120Core_e$ along with differences in the unmodified loops of these sequence may affect the function of the V3 loop. The V3 loop has been shown to be integral to the infectivity of HIV-1 and neutralizing antibodies directed to this loop, particularly the tip of the loop that has been shown to be a potent neutralization epitope, and antibodies directed to this loop inhibit viral entry (Ivanoff et al., 1992; Yokoyama et al., 2016). This is primarily due to the V3 loop contacting the co-receptors (CCR5 and/or CXCR4) that are required for viral entry into the host cells (Ivanoff et al., 1992) and are a main determinant of viral tropism (Yokoyama et al., 2016). It has also been determined that mutations in the V3 loop, particularly in the regions that are more conserved, such as the tip of the V3 loop, can be deleterious to the function of HIV-1 viral entry (Ivanoff et al., 1992). Structural investigations have also found a synergy between the V1/V2 region and its masking of the V3 tip that prevents accessibility from neutralizing antibodies (Yokoyama et al., 2016). It is therefore possible that through innate sequence differences between the founder virus $gp120_{FVC}ExC$, as well as the potential increased conformational stability of the V3 region, that it allows for enhanced viral entry through more efficient binding to the host co-receptors, a key feature of HIV-1 founder viruses. The decreased fluctuations of the V3 region in $gp120_{FVC}ExC$ compared to $gp120Core_e$ may translate to the non-modified $gp120_{FVC}ExC$ structure that lacks the linker for crystallization. Should this be the case, it is possible that the V3 region may be more or less shielded by the V1/V2 loop, altering its immunogenicity and allowing for increased viral immune escape and establishment of infection.

The V4 region in $gp120Core_e$ contains additional resides compared to those of $gp120_{FVC}ExC$, resulting in a longer V4 loop structure (Figure 3.1 and 3.16). A direct comparison of corresponding aligned residues through structural analysis shows that $gp120Core_e$ is stabilized by hydrogen bonds that are not as prevalent in $gp120_{FVC}ExC$ (Figure 3.16). More interestingly, is that the larger V4 loop region of $gp120Core_e$ contains an anti-parallel β-sheet that appears to stabilize the region due to the hydrogen bonds that form between the sheets resulting in less fluctuations of the residues in this region compared to $gp120_{FVC}ExC$ (Figure 3.16). The functions of the V4 loop in gp120 has thus far not been properly characterized in regards to viral entry into host cells, however it is suggested that the V4 region affects the conformation of gp120 and may provide steric hindrance to prevent neutralization by antibodies (Frost et al., 2005; Wei et al., 2003). The V4 region has also been found to act as a neutralizing epitope during infection (Moore et al., 2008). Interestingly

the V4 region appears to be more resistant to modification than either V1, V2, or V3 (Nakane et al., 2015) and the replacement or deletion of the V4 region with a linker sequence of the same length appears to affect the binding of gp120 to CD4 and the CCR5 or CXCR4 co-receptors (Yuan et al., 2013). It has also been observed that the deletion of the V4 region negatively affects the display of gp120 on the cell surface (Yuan et al., 2013). The smaller V4 region of $gp120_{FVC}ExC$ may be beneficial to the display of Env on the host cells or may allow for a different conformation of gp120 that aids in the binding of either the receptor or co-receptor, particularly as it doesn't contain a highly stable β-sheet as observed in $gp120Core_e$. These are merely speculation to provide insight into the nature of a founder virus consensus sequence that is capable of establishing an HIV-1 viral infection and to act as a point of further investigation.

The V5 region of $gp120Core_e$ contains additional residues compared to $gp120_{FVC}ExC$ resulting in a longer more conformationally flexible loop (Figure 3.1 and 3.17). Comparative investigation through MD simulations appear to show no interactions between the residues that would account for the increased fluctuation observed in $gp120Core_e$ or the lesser fluctuations observed in $gp120_{FVC}ExC$. The fluctuation in residues appear to be primarily due to coiling and un-coiling of the $gp120Core_e$ V5 region through the formation and breakage of bonds with adjacent residues. As with the V4 region of gp120, the function of the V5 loop has not yet been fully determined but appears to play a role in the conformation of gp120 as well as potential steric hindrance to shield epitopes (Yuan et al., 2013). The V5 region also appears to be quite resistant to modifications, requiring a deletion or complete substitution to affect the functions of gp120 (Yuan et al., 2013). Consequently the V5 as well as the V4 region appear to undergo limited conformational change upon CD4 binding, and is therefore not likely involved in the enhancement of immunogenicity (Nakane et al., 2015). As the V5 loop consists of fewer residues and projects outward from the gp120 structure, its role in steric hindrance in CD4 binding and epitope exposure may not be functionally relevant, further than that of V4.

As mentioned, our criterion excluded the differences observed in the CD4 binding loop. However, it is pertinent to mention the changes observed in this region due to its functional relevance in a founder virus sequence and subsequent establishment of an HIV-1 infection, and as such subtle changes may be functionally relevant. The CD4 binding loop has been shown to be a highly conserved region among gp120 sequences (Kwong et al., 2002). It was observed that three

residues: Gly-267, The-272, and Asn-277 show a sizeable reduction in fluctuation in $gp120_{FVC}ExC$ compared to $gp120Core_e$. The stabilization of these residues, particularly with modifications observed in the V5 region of $gp120_{FVC}ExC$ may stabilize these residues into a conformation that leads to enhanced CD4 binding. However, the relevance of these residues would need to be further investigated to establish whether these are functionally relevant in a founder virus sequence and providing a transmission advantage for that particular variant. It should be noted that the binding functionality of $gp120_{FVC}ExC$ sequence has been validated with regards to 2dCD4 binding (Cerutti et al., 2010) and as such these point mutations are not deleterious to the binding ability of our $gp120_{FVC}ExC$ sequence.

4.4.$gp120_{FVC}ExC$ in the context of HIV-1 founder virus envelopes

As mentioned previously, HIV-1 founder viruses have been found to play an integral role in the establishment of an infection in the host, with an infection being established by a single virus in approximately 80% of cases (Abrahams et al., 2009). This reiterates their importance in the role of prophylactic vaccine design and the prevention of HIV-1 infection. Further investigation of these viruses has shown that founder viruses exhibit unique characteristics compared to viruses active during a chronic HIV-1 infection. These characteristics likely have an impact on the results from this study and will be integral for future investigations regarding $gp120_{FVC}ExC$.

An important consideration for any HIV-1 Env analysis is the importance of the glycosylation profile of the Env protein. In this study we have observed a variable deglycosylation pattern despite our efforts to reduce conformational heterogeneity through the use of $GnTi^-$ cells and Endo H deglycosylation. This is likely a pattern that will be observed in other HIV-1 Env founder viruses as the glycosylation profile of both subtype C and B viruses are more alike across subtypes than with chronic viruses of the same subtype (Go et al., 2011). This heterogeneity of glycosylation profile is particularly apparent in glycan site occupancy and the amount of complex glycans (Go et al., 2011) with subtype C founder viruses exhibiting fewer N-linked glycosylation sites (Rademeyer et al., 2007). This may have implications in improving the deglycosylation of $gp120_{FVC}ExC$ for further optimization regarding the homogeneity of the deglycosylated protein species and increasing protein yields for future upscaled crystallization. Unique sequence

characteristics have also been discovered through recent studies that are corroborated by our MD simulations as well as our sequence alignment (Figure 3.1, 3.16, and 3.17). It has been shown that there is a propensity for HIV-1 founder viruses to have shorter variable loops (V1-V4) (Rademeyer et al., 2007). This is observed in our sequence alignment (Figure 3.1) as well as our MD analysis (Figure 3.16). In the instance of the $gp120_{FVC}$ExC sequence the V5 loop is also longer than that of $gp120Core_e$, reinforcing existing findings (Figure 3.1 and 3.17). Furthermore, MD analysis on $gp120_{FVC}$ExC without the V1/V2, and V3 modifications will provide insight into the dynamics of the loops and potentially corroborate a lower RMSD for the structure which may allow for a more representative complete $gp120_{FVC}$ExC protein structure to be crystallized through more inclusion of the V1-V3 loops.

In summation, the overall trend of stability and less conformational flexibility of $gp120_{FVC}$ExC observed through the calculated RMSD (Figure 3.10), RMSF (Figure 3.11), and visually though MD simulations compared to $gp120Core_e$ appear to be due to large differences of fluctuations in core regions of the gp120 sequences. These differences may be functionally relevant to the nature of a founder virus sequence and warrant further investigation, particularly those of the variable loop regions. Other smaller differences in residue fluctuations were observed but proved difficult to quantify. Further investigation into the nature of $gp120_{FVC}$ExC may allow the current data to support and supplement those investigations and provide insight into differences that are determined to be functionally relevant to the mechanisms specific to founder viruses. However, the overall trend of increased stability of $gp120_{FVC}$ExC compared to $gp120Core_e$ may allow for more favourable conformations that lead to increased ability to establish infection into host cells.

4.5. The introduction of a Cysteine residue at position 60 reduces conformational dynamics of 2dCD4

Comparative MD analysis of $2dCD4^{WT}$ with $2dCD4^{S60C}$ containing the S60C mutation provided considerable insight into the changes in dynamics of 2dCD4 due to the mutation. Overall, the trend appears to be that the S60C mutation reduces the conformational dynamics of the protein as observed by the reduction in RMSD and RMSF of the $2dCD4^{S60C}$ protein (Figures 3.18 and 3.19), which may confer increased stability. This increase in stability as a result of reduced fluctuations

in the residues of 2dCD4^{S60C} compared to 2dCD4^{WT} may result in these residues adopting a conformation that confers altered function in the binding abilities of the protein. The introduction of the S60C mutation has been shown to allow 2dCD4^{S60C} to bind to gp120 at a higher affinity than 2dCD4^{WT} (Cerutti et al., 2010) and the observed reduction in conformational flexibility may allow for this enhanced binding to gp120 through beneficial conformations of 2dCD4^{S60C}. However, overall, it appears that the introduction of this residue does not greatly alter the structure and in the case of gp120 binding, the function.

Multiple changes in the fluctuation of residues have been observed in the comparative MD analysis of 2dCD4^{WT} compared to 2dCD4^{S60C}. The following residues: Ser-19, Gln-20, and Lys-21 that reside proximal to the gp120 binding region experience less fluctuation in 2dCD4^{S60C} than the corresponding residues in 2dCD4^{WT} (Figure 3.20). In addition, the gp120 binding region is comprised of residues that exhibit less fluctuation in 2dCD4^{S60C} compared to 2dCD4^{WT}, namely: Arg-59, Cys/Ser-60, Trp-62, and Asp-63. There is evidence that Cys/Ser-60 and Asp-63 may be directly involved in the binding of gp120 to CD4 (Fontenot et al., 2007; Kwong et al., 1998; Moebius et al., 1992), despite these residues are not the primary residues involved in the binding of gp120 (i.e. Phe-43 and Arg-59) (Fontenot et al., 2007; Kwong et al., 1998; Moebius et al., 1992). The reduction of these resides in 2dCD4^{S60C} when compared to 2dCD4^{WT} may aid in the binding of this region to gp120. Particularly, the reduction in fluctuations of Phe-43 and Arg-59 may stabilize them in a conformation conducive to binding gp120. As mentioned previously, the S60C mutation increases the binding affinity of gp120 to CD4 (Killick et al., 2015). This may be due to the nature of the residues in serine to cysteine substitution or the change in fluctuations of the neighbouring key residues (Arg-59) due to the S60C mutation.

The most significant differences observed for 2dCD4^{S60C} compared to 2dCD4^{WT} were in domain 2 of the protein structure. The Asp-105, Thr-106, Ser-124 and Ser-125 residues experienced the largest reduction of fluctuation in 2dCD4^{S60C} compared to 2dCD4^{WT}. Despite the fact that the binding region to gp120 is in domain 1 of 2dCD4, the binding of 2dCD4 to gp120 is dependent on the formation and breakage of an allosteric disulphide bond that resides in domain 2 of the protein between Cys-130 and Cys-159 (Matthias et al., 2002). The reduction of this allosteric Cys-130-Cys-159 bond results in conformational changes in domain 1 of CD4 that lead to the binding of gp120 (Owen et al., 2018). As such it has been found that certain mutations in domain 2 of 2dCD4

can inhibit binding to gp120 (Mizukami et al., 1988). While our prominent fluctuations do not overlap with the described mutations, it is possible that these fluctuations exert a conformational impact to gp120 binding.

Interestingly, the gp120 binding region of 2dCD4^{S60C} and the CD4 binding loop of gp120$_{FVC}$ExC that are used to generate the gp120$_{FVC}$ExC – 2dCD4^{S60C} complex show reduced mobility of these binding regions. This may confer increased stability that facilitates the formation of a stable gp120$_{FVC}$ExC – 2dCD4^{S60C} immunogen complex and warrants further investigation.

4.6. The putative disulphide bond in gp120$_{FVC}$ExC – 2dCD4^{S60C} increases conformational dynamics of the complex compared to gp120$_{FVC}$ExC – 2dCD4^{WT}

An integral part of this study was the investigation of the gp120$_{FVC}$ExC – 2dCD4^{S60C} complex that contains the putative disulphide bond integral to the Env – 2dCD4^{S60C} complex that produces potent, broadly neutralizing antibodies in a rabbit model capable of neutralizing clinically relevant (tier 2 and tier 3) HIV-1 pseudoviruses (Killick et al., 2015). The investigation of this complex was performed through MD simulations with the aim of supplementing future structural information of the crystal structure of gp120$_{FVC}$ExC – 2dCD4^{S60C}, as compared to a WT complex. The MD simulation data was interpreted together with results from our lab using Hydrogen-Deuterium (HD) exchange to determine the change in accessibility of residues to surrounding solvent due to the introduction of the putative disulphide bond, since epitope mapping of the rabbit serum proved inconclusive.

The introduction of the putative disulphide bond between residue Cys-113 and Cys-60 increases the overall conformational flexibility of the gp120$_{FVC}$ExC – 2dCD4^{S60C} complex when compared to the gp120$_{FVC}$ExC – 2dCD4^{WT} complex (Figure 3.21). This increased conformational flexibility suggests that the basis for enhanced immunogenicity of the covalently bound complexes may be due, in part, to potential exposure of novel epitopes or the reduction in steric hindrance that shield neutralizing epitopes. This increase in overall dynamics of the residues in the S60C complex is observed primarily in the gp120$_{FVC}$ExC portion of the complex and to a lesser extent the 2dCD4 component of the complex. The putative disulphide bond introduced by the S60C mutation appears

to stabilize the residues involved in the putative disulphide bond with a reduction in both Cys-113 ($gp120_{FVC}ExC$) and Cys-60 ($2dCD4^{S60C}$) residues in the $gp120_{FVC}ExC$ – $2dCD4^{S60C}$ complex (Figure 3.21).

Through the selected criterion we have determined two regions of $gp120_{FVC}ExC$ in the $gp120_{FVC}ExC$ – $2dCD4^{S60C}$ that experience drastic changes in fluctuation when observed through MD analysis, as compared to $gp120_{FVC}ExC$ – $2dCD4^{WT}$ (Figure 3.23). The first region is closer to the N-terminus of the $gp120_{FVC}ExC$ protein at residues 38-43 for which the residues experience an 85% decrease in fluctuation compared to the complex without the putative disulphide bond (Figure3.21 B). These residues form a loop structure that in the disulphide bond complex remains fairly stable compared to the complex without the bond. It should be noted that both complexes contain the same sequences and as such the disulphide bonds appears to produce structure wide changes in residue fluctuation. While no effective antibodies have been discovered that interact with residues 38-43, novel antibodies may potentially be directed to stabilized epitopes in this region. It is possible that the conformational stability in these residues in the S60C complex affect the exposure of neighbouring epitopes to the host immune system. In essence, it may be more stable in a more accessible conformation that increases the exposure of novel epitopes or reduces steric hindrance, allowing the exposure of neighbouring epitopes in the structure.

The second region that exhibits a drastic change in dynamics is situated in the beginning of the V3 loop with smaller fluctuations observed between the V3 loop and the CD4 binding loop (Figure 3.21) – both of which are highly immunogenic and integral functional regions of gp120 (Han et al., 2019; Pantophlet et al., 2008; Raja et al., 2003; Zhou et al., 2010). These residues, 214-224, of $gp120_{FVC}ExC$ – $2dCD4^{S60C}$ are much more mobile than those of the $gp120_{FVC}ExC$ – $2dCD4^{WT}$ complex. These residues of the disulphide linked complex fluctuate between a coiled and un-coiled conformation that may have implications for the exposure of novel epitopes (Figure 3.24), or the steric interference of the neighbouring epitopes. The coiling of these residues may allow for the accessibility of residues in the structure by increasing their exposure, as well surrounding epitopes by reducing the steric interference that may be caused by a more open loop structure. Interestingly, for residues 214-224 there is a more constant α-helix structure in the $gp120_{FVC}ExC$ – $2dCD4^{WT}$ complex (Figure 3.24). This α-helix structure, while not constant, is present for at least 70% of the MD simulation (Figure 3.25). This further strengthens the impression that the putative disulphide

bond creates conformational changes throughout the $gp120_{FVC}ExC - 2dCD4^{S60C}$ complex. This may also account for the increased exposure of potential epitopes on residues 214-224, as well as neighbouring regions in the V3 loop and CD4 binding regions that have been shown to be immunogenic (Han et al., 2019; Pantophlet et al., 2008; Raja et al., 2003; Zhou et al., 2010). It should be noted that because the V1, V2, and V3 regions have been modified to aid in the crystallization of both the $gp120_{FVC}ExC$ and the $gp120_{FVC}ExC - 2dCD4^{S60C}$ complex, it does present limitations in drawing conclusions through MD analysis. However, these differing dynamics may be present in an unmodified gp120 founder virus protein and create direction for further investigation.

It is also pertinent to touch on the increased fluctuations observed in the CD4 binding loop in the $gp120_{FVC}ExC - 2dCD4^{S60C}$ complex compared to the $gp120_{FVC}ExC - 2dCD4^{WT}$ complex due to its functional relevance in CD4 binding and immunogenic epitopes (Raja et al., 2003, p. 4; Wyatt and Sodroski, 1998). The CD4 binding site on gp120 is located in a depression of which half the residues of gp120 contact CD4 only through main chain atoms (Wyatt and Sodroski, 1998). The increased fluctuations of these residues in $gp120_{FVC}ExC - 2dCD4^{S60C}$ may allow for enhanced CD4 binding through the increased accessibility of these residues situated in the CD4 binding loop. It is also possible that the increased fluctuations in the residues of this region allow for increased epitope exposure, as these have been found to be an immunogenic region in gp120 that prevents CD4 binding (Raja et al., 2003).

In summation, it appears that the putative disulphide bond introduced through the S60C mutation in 2dCD4 increases the conformational flexibility of $gp120_{FVC}ExC$, particularly for residues 38-43 and residues 214-224 in the V3 loop of gp120. The resulting increased conformational flexibility may increase the exposure of novel epitopes, or decrease steric hindrance that may be responsible for the increased immunogenicity of $Env - 2dCD4^{S60C}$ complexes (Killick et al., 2015).

5. Conclusion and future direction

The need for an HIV-1 vaccine remains integral in the eradication of the HIV-1 pandemic despite the increased quality of life of those afflicted afforded by improved ARV treatments and intervention. It is to this end that HIV-1 Env founder viruses remain an attractive avenue as a potential prophylactic vaccine candidate. The use of an HIV-1 gp120 FV sequence (in conjunction with a 2dCD4 binding partner) aims to elicit bNAbs to conserved sites on gp120 that will neutralize and prevent infection from single HIV-1 FV variants that are responsible for establishing productive infection in 70-80% of transmission cases. To investigate the nature of a gp120 FV sequence, comparative MD simulation analysis was performed on $gp120_{FVC}ExC$ and compared to a non-FV variant $gp120Core_e$ to investigate potential characteristics unique to this generated variant, as well as validate methodology of sequence modification to produce protein crystals of $gp120_{FVC}ExC$ for X-ray diffraction studies. Further, we performed MD simulation analysis to investigate potential characteristics of the $gp120_{FVC}ExC$ – $2dCD4^{S60C}$ complex compared to $gp120_{FVC}ExC$ – $2dCD4^{WT}$ that may be responsible for the enhanced immunogenicity observed in Env – $2dCD4^{S60C}$ (Killick et al., 2015).

The sequence modifications employed in the truncation of the V1, V2, and V3 loops from *Kwon et al,* (2012) have been successfully implemented and validated in the generation of protein crystals of $gp120_{FVC}ExC$ and form a validation for future gp120 crystallography studies and eventual X-ray diffraction of $gp120_{FVC}ExC$. Marked differences were also observed in the variable loop regions (V1-V5) that align with previous findings (Rademeyer et al., 2007) and act as a point of further future investigation into the nature of a gp120 subtype C FV consensus sequence. MD simulations of the $gp120_{FVC}ExC$ – $2dCD4^{S60C}$ complex show a decrease in fluctuations near the N-terminus at residue 38-43 with a marked increase in fluctuation of the V3 region of the $gp120_{FVC}ExC$ component when compared to $gp120_{FVC}ExC$ – $2dCD4^{WT}$. The future direction is the generation of X-ray diffraction crystal structures of $gp120_{FVC}ExC$ to further investigate the structural characteristics of $gp120_{FVC}ExC$, as well as the crystallization and structure determination of $gp120_{FVC}ExC$ – $2dCD4^{S60C}$ for further investigation into the nature of the immunogenicity conferred by the putative disulphide bond introduced by the S60C mutation. The results from this

study create points of investigation for future structural studies regarding these recombinant proteins and may play a role in the increased understanding of Env FV prophylactic vaccines.

6. References

Abrahams, M.-R., Anderson, J.A., Giorgi, E.E., Seoighe, C., Mlisana, K., Ping, L.-H., Athreya, G.S., Treurnicht, F.K., Keele, B.F., Wood, N., Salazar-Gonzalez, J.F., Bhattacharya, T., Chu, H., Hoffman, I., Galvin, S., Mapanje, C., Kazembe, P., Thebus, R., Fiscus, S., Hide, W., Cohen, M.S., Karim, S.A., Haynes, B.F., Shaw, G.M., Hahn, B.H., Korber, B.T., Swanstrom, R., Williamson, C., 2009. Quantitating the Multiplicity of Infection with Human Immunodeficiency Virus Type 1 Subtype C Reveals a Non-Poisson Distribution of Transmitted Variants. J. Virol. 83, 3556–3567.

Alberto, R.-F.R., Martiniano, B., Saúl, R.-H., Jazmín, G.-M., Mara, G.-S., Rubén, E.-P.A., Jonathan, F.-V.M., Vicente, M.-M.J., José, C.-B., 2020. In silico and in vivo studies of gp120-HIV-derived peptides in complex with G4-PAMAM dendrimers. RSC Adv. 10, 20414–20426.

Allen, L.C., 1989. Lewis-Langmuir atomic charges. J. Am. Chem. Soc. 111, 9115–9116.

Badaya, A., Sasidhar, Y.U., 2020. Inhibition of the activity of HIV-1 protease through antibody binding and mutations probed by molecular dynamics simulations. Sci. Rep. 10, 5501.

Bagnarelli, P., Fiorelli, L., Vecchi, M., Monachetti, A., Menzo, S., Clementi, M., 2003. Analysis of the functional relationship between V3 loop and gp120 context with regard to human immunodeficiency virus coreceptor usage using naturally selected sequences and different viral backbones. Virology 307, 328–340.

Bbosa, N., Kaleebu, P., Ssemwanga, D., 2019. HIV subtype diversity worldwide. Curr. Opin. HIV AIDS 14, 153–160.

Bebenek, K., Abbotts, J., Wilson, S.H., Kunkel, T.A., 1993. Error-prone polymerization by HIV-1 reverse transcriptase. Contribution of template-primer misalignment, miscoding, and termination probability to mutational hot spots. J. Biol. Chem. 268, 10324–10334.

Berger, E.A., Murphy, P.M., Farber, J.M., 1999. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. Annu. Rev. Immunol. 17, 657–700.

Berndsen, Z.T., Chakraborty, S., Wang, X., Cottrell, C.A., Torres, J.L., Diedrich, J.K., López, C.A., Yates, J.R., Gils, M.J. van, Paulson, J.C., Gnanakaran, S., Ward, A.B., 2020. Visualization of the HIV-1 Env glycan shield across scales. Proc. Natl. Acad. Sci. 117, 28014–28025.

Binley, J.M., Sanders, R.W., Clas, B., Schuelke, N., Master, A., Guo, Y., Kajumo, F., Anselma, D.J., Maddon, P.J., Olson, W.C., Moore, J.P., 2000. A Recombinant Human Immunodeficiency Virus Type 1 Envelope Glycoprotein Complex Stabilized by an Intermolecular Disulfide Bond between the gp120 and gp41 Subunits Is an Antigenic Mimic of the Trimeric Virion-Associated Structure. J. Virol. 74, 627–643.

Boldon, L., Laliberte, F., Liu, L., 2015. Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application. Nano Rev. 6, 25661.

Brooks, B.R., Brooks, C.L., MacKerell, A.D., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A.R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R.W., Post, C.B., Pu, J.Z., Schaefer, M., Tidor, B., Venable, R.M., Woodcock, H.L., Wu, X., Yang, W., York, D.M., Karplus, M., 2009. CHARMM: The Biomolecular Simulation Program. J. Comput. Chem. 30, 1545–1614.

Brumshtein, B., Greenblatt, H.M., Futerman, A.H., Silman, I., Sussman, J.L., 2008. Control of the rate of evaporation in protein crystallization by the 'microbatch under oil' method. J. Appl. Crystallogr. 41, 969–971.

Burton, D.R., Ahmed, R., Barouch, D.H., Butera, S.T., Crotty, S., Godzik, A., Kaufmann, D.E., McElrath, M.J., Nussenzweig, M.C., Pulendran, B., Scanlan, C.N., Schief, W.R., Silvestri, G., Streeck, H., Walker, B.D., Walker, L.M., Ward, A.B., Wilson, I.A., Wyatt, R., 2012. A Blueprint for HIV Vaccine Discovery. Cell Host Microbe 12, 396–407.

Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz, K.M., Onufriev, A., Simmerling, C., Wang, B., Woods, R.J., 2005. The Amber biomolecular simulation programs. J. Comput. Chem. 26, 1668–1688.

Cerutti, N., Mendelow, B.V., Napier, G.B., Papathanasopoulos, M.A., Killick, M., Khati, M., Stevens, W., Capovilla, A., 2010a. Stabilization of HIV-1 gp120-CD4 Receptor Complex through Targeted Interchain Disulfide Exchange. J. Biol. Chem. 285, 25743–25752.

Cerutti, N., Mendelow, B. V., Napier, G.B., Papathanasopoulos, M.A., Killick, M., Khati, M., Stevens, W., Capovilla, A., 2010b. Stabilization of HIV-1 gp120-CD4 receptor complex through targeted interchain disulfide exchange. J. Biol. Chem. 285, 25743–25752.

Checkley, M.A., Luttge, B.G., Freed, E.O., 2011. HIV-1 Envelope Glycoprotein Biosynthesis, Trafficking, and Incorporation. J. Mol. Biol. 410, 582–608.

Cheung, N.J., Yu, W., 2018. De novo protein structure prediction using ultra-fast molecular dynamics simulation. PLOS ONE 13, e0205819.

Chitongo, R., Obasa, A.E., Mikasi, S.G., Jacobs, G.B., Cloete, R., 2020. Molecular dynamic simulations to investigate the structural impact of known drug resistance mutations on HIV-1C Integrase-Dolutegravir binding. PloS One 15, e0223464.

Chong Teoh, T., Heidelberg, T., Rizman-Idid, M., 2011. Systematic protein-protein docking and molecular dynamics studies of HIV-1 gp120 and CD4: insights for new drug development. Daru J. Fac. Pharm. Tehran Univ. Med. Sci. 19, 469–475.

Do, P.-C., Lee, E.H., Le, L., 2018. Steered Molecular Dynamics Simulation in Rational Drug Design. J. Chem. Inf. Model. 58, 1473–1482.

Fontenot, D., Jones, J.K., Hossain, M.M., Nehete, P.N., Vela, E.M., Dwyer, V.A., Jagannadha Sastry, K., 2007. Critical role of Arg59 in the high-affinity gp120-binding region of CD4 for human immunodeficiency virus type 1 infection. Virology 363, 69–78.

Freeze, H.H., Kranz, C., 2010. Endoglycosidase and Glycoamidase Release of N-Linked Glycans. Curr. Protoc. Mol. Biol. Ed. Frederick M Ausubel Al 0 17.

Frost, S.D.W., Wrin, T., Smith, D.M., Kosakovsky Pond, S.L., Liu, Y., Paxinos, E., Chappey, C., Galovich, J., Beauchaine, J., Petropoulos, C.J., Little, S.J., Richman, D.D., 2005. Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection. Proc. Natl. Acad. Sci. U. S. A. 102, 18514–18519.

Gajula, M.P., Kumar, A., Ijaq, J., 2016. Protocol for Molecular Dynamics Simulations of Proteins. Bio-Protoc. 6, e2051–e2051.

Ganser-Pornillos, B.K., Yeager, M., Pornillos, O., 2012. Assembly and architecture of HIV. Adv. Exp. Med. Biol. 726, 441–465.

Georgiev, I.S., Joyce, M.G., Zhou, T., Kwong, P.D., 2013. Elicitation of HIV-1-neutralizing antibodies against the CD4-binding site. Curr. Opin. HIV AIDS 8, 382–392.

Go, E.P., Hewawasam, G., Liao, H.-X., Chen, H., Ping, L.-H., Anderson, J.A., Hua, D.C., Haynes, B.F., Desaire, H., 2011. Characterization of Glycosylation Profiles of HIV-1 Transmitted/Founder Envelopes by Mass Spectrometry. J. Virol. 85, 8270–8284.

González, M.A., 2011. Force fields and molecular dynamics simulations. Éc. Thématique Société Fr. Neutron. 12, 169–200.

Goto, T., Nakai, M., Ikuta, K., 1998. The life-cycle of human immunodeficiency virus type 1. Micron Oxf. Engl. 1993 29, 123–138.

Gottlieb, M.S., Schroff, R., Schanker, H.M., Weisman, J.D., Fan, P.T., Wolf, R.A., Saxon, A., 1981. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. N. Engl. J. Med. 305, 1425–1431.

Group, T. rgp120 H.V.S., 2005. Placebo-Controlled Phase 3 Trial of a Recombinant Glycoprotein 120 Vaccine to Prevent HIV-1 Infection. J. Infect. Dis. 191, 654–665.

Gupta, M., Chauhan, R., Prasad, Y., Wadhwa, G., Jain, C.K., 2016. Protein-protein interaction and molecular dynamics analysis for identification of novel inhibitors in Burkholderia cepacia GG4. Comput. Biol. Chem. 65, 80–90.

Guvench, O., MacKerell, A.D., 2008. Comparison of protein force fields for molecular dynamics simulations. Methods Mol. Biol. Clifton NJ 443, 63–88.

Hallenberger, S., Bosch, V., Angliker, H., Shaw, E., Klenk, H.D., Garten, W., 1992. Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. Nature 360, 358–361.

Han, Q., Jones, J.A., Nicely, N.I., Reed, R.K., Shen, X., Mansouri, K., Louder, M., Trama, A.M., Alam, S.M., Edwards, R.J., Bonsignori, M., Tomaras, G.D., Korber, B., Montefiori, D.C., Mascola, J.R., Seaman, M.S., Haynes, B.F., Saunders, K.O., 2019. Difficult-to-neutralize global HIV-1 isolates are neutralized by antibodies targeting open envelope conformations. Nat. Commun. 10, 2898.

Hoffman, T.L., Doms, R.W., 1999. HIV-1 envelope determinants for cell tropism and chemokine receptor use. Mol. Membr. Biol. 16, 57–65.

Hollingsworth, L.R., Brown, A.M., Gandour, R.D., Bevan, D.R., 2018. Computational study of HIV gp120 as a target for polyanionic entry inhibitors: Exploiting the V3 loop region. PLoS ONE 13.

Hollingsworth, S.A., Dror, R.O., 2018. Molecular Dynamics Simulation for All. Neuron 99, 1129–1143.

Hospital, A., Goñi, J.R., Orozco, M., Gelpí, J.L., 2015. Molecular dynamics simulations: advances and applications. Adv. Appl. Bioinforma. Chem. AABC 8, 37–47.

Hu, W.-S., Hughes, S.H., 2012. HIV-1 Reverse Transcription. Cold Spring Harb. Perspect. Med. 2.

Ivanoff, L.A., Dubay, J.W., Morris, J.F., Roberts, S.J., Gutshall, L., Sternberg, E.J., Hunter, E., Matthews, T.J., Petteway, S.R., 1992. V3 loop region of the HIV-1 gp120 envelope protein is essential for virus infectivity. Virology 187, 423–432.

Ji, X., Wang, W., Zheng, Y., Hao, J., Sun, M., 2012. Homology Modeling and Molecular Dynamics Simulation Studies of a Marine Alkaline Protease. Bioinforma. Biol. Insights 6, 255–263.

Jones, A.T., Chamcha, V., Kesavardhana, S., Shen, X., Beaumont, D., Das, R., Wyatt, L.S., LaBranche, C.C., Stanfield-Oakley, S., Ferrari, G., Montefiori, D.C., Moss, B., Tomaras, G.D., Varadarajan, R., Amara, R.R., 2017. A Trimeric HIV-1 Envelope gp120 Immunogen Induces Potent and Broad Anti-V1V2 Loop Antibodies against HIV-1 in Rabbits and Rhesus Macaques. J. Virol. 92

Kaleem, Z., White, G., Zutter, M.M., 2001. Aberrant Expression of T-Cell–Associated Antigens on B-Cell Non-Hodgkin Lymphomas. Am. J. Clin. Pathol. 115, 396–403.

Kandt, C., Ash, W.L., Tieleman, D.P., 2007. Setting up and running molecular dynamics simulations of membrane proteins. Methods San Diego Calif 41, 475–488.

Kang, C.Y., Hariharan, K., Nara, P.L., Sodroski, J., Moore, J.P., 1994. Immunization with a soluble CD4-gp120 complex preferentially induces neutralizing anti-human immunodeficiency virus type 1 antibodies directed to conformation-dependent epitopes of gp120. J. Virol. 68, 5854–62.

Kazazi, F., Mathijs, J.-M., Foley, P., Cunningham, A.L., 1989. Variations in CD4 Expression by Human Monocytes and Macrophages and Their Relationship to Infection with the Human Immunodeficiency Virus. J. Gen. Virol. 70, 2661–2672.

Keele, B.F., Giorgi, E.E., Salazar-Gonzalez, J.F., Decker, J.M., Pham, K.T., Salazar, M.G., Sun, C., Grayson, T., Wang, S., Li, H., Wei, X., Jiang, C., Kirchherr, J.L., Gao, F., Anderson, J.A., Ping, L.-H., Swanstrom, R., Tomaras, G.D., Blattner, W.A., Goepfert, P.A., Kilby, J.M., Saag, M.S., Delwart, E.L., Busch, M.P., Cohen, M.S., Montefiori, D.C., Haynes, B.F., Gaschen, B., Athreya, G.S., Lee,

H.Y., Wood, N., Seoighe, C., Perelson, A.S., Bhattacharya, T., Korber, B.T., Hahn, B.H., Shaw, G.M., 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc. Natl. Acad. Sci. 105, 7552–7557.

Killick, M., Capovilla, A., Papathanasopoulos, M., 2012. Immunogenicity of native and CD4 liganded monomeric and trimeric envelope glycoproteins based on HIV-1 Subtype C consensus Founder virus sequences. Retrovirology 9, P21.

Killick, M., Capovilla, A., Papathanasopoulos, M.A., 2014. Generation and Characterization of an HIV-1 Subtype C Transmitted and Early Founder Virus Consensus Sequence. AIDS Res. Hum. Retroviruses 30, 1001–1005.

Killick, M.A., Grant, M.L., Cerutti, N.M., Capovilla, A., Papathanasopoulos, M.A., 2015. Env–2dCD4S60C complexes act as super immunogens and elicit potent, broadly neutralizing antibodies against clinically relevant human immunodeficiency virus type 1 (HIV-1). Vaccine 33, 6298–6306.

Kim, S., 2014. Issues on the Choice of a Proper Time Step in Molecular Dynamics. Phys. Procedia 53, 60–62.

Klatzmann, D., Gluckman, J.C., 1986. HIV infection: facts and hypotheses. Immunol. Today 7, 291–296.

König, R., Huang, L.Y., Germain, R.N., 1992. MHC class II interaction with CD4 mediated by a region analogous to the MHC class I binding site for CD8. Nature 356, 796–798.

Koup, R.A., Douek, D.C., 2011. Vaccine Design for CD8 T Lymphocyte Responses. Cold Spring Harb. Perspect. Med. 1, a007252–a007252.

Krenkova, J., Szekrenyes, A., Keresztessy, Z., Foret, F., Guttman, A., 2013. Oriented immobilization of peptide-N-glycosidase F on a monolithic support for glycosylation analysis. J. Chromatogr. A 1322, 54–61.

Kuhlman, B., Bradley, P., 2019. Advances in protein structure prediction and design. Nat. Rev. Mol. Cell Biol. 20, 681–697.

Kwon, Y.D., Finzi, A., Wu, X., Dogo-Isonagie, C., Lee, L.K., Moore, L.R., Schmidt, S.D., Stuckey, J., Yang, Y., Zhou, T., Zhu, J., Vicic, D.A., Debnath, A.K., Shapiro, L., Bewley, C.A., Mascola, J.R., Sodroski, J.G., Kwong, P.D., 2012a. Unliganded HIV-1 gp120 core structures assume the CD4-bound conformation with regulation by quaternary interactions and variable loops. Proc. Natl. Acad. Sci. 109, 5663–5668.

Kwon, Y.D., Finzi, A., Wu, X., Dogo-Isonagie, C., Lee, L.K., Moore, L.R., Schmidt, S.D., Stuckey, J., Yang, Y., Zhou, T., Zhu, J., Vicic, D.A., Debnath, A.K., Shapiro, L., Bewley, C.A., Mascola, J.R., Sodroski, J.G., Kwong, P.D., 2012b. Unliganded HIV-1 gp120 core structures assume the CD4-bound conformation with regulation by quaternary interactions and variable loops. Proc. Natl. Acad. Sci. 109, 5663–5668.

Kwong, P.D., Doyle, M.L., Casper, D.J., Cicala, C., Leavitt, S.A., Majeed, S., Steenbeke, T.D., Venturi, M., Chaiken, I., Fung, M., Katinger, H., Parren, P.W.I.H., Robinson, J., Van Ryk, D., Wang, L., Burton, D.R., Freire, E., Wyatt, R., Sodroski, J., Hendrickson, W.A., Arthos, J., 2002. HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. Nature 420, 678–682.

Kwong, P.D., Wyatt, R., Desjardins, E., Robinson, J., Culp, J.S., Hellmig, B.D., Sweet, R.W., Sodroski, J., Hendrickson, W.A., 1999. Probability Analysis of Variational Crystallization and Its Application to gp120, The Exterior Envelope Glycoprotein of Type 1 Human Immunodeficiency Virus (HIV-1)*. J. Biol. Chem. 274, 4115–4123.

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

Lee, H.S., Qi, Y., Im, W., 2015. Effects of N -glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study. Sci. Rep. 5, 8926.
Lelièvre, J.-D., Lévy, Y., 2016. HIV-1 prophylactic vaccines: state of the art. J. Virus Erad. 2, 5–11.
Li, G., De Clercq, E., 2016. HIV Genome-Wide Protein Associations: a Review of 30 Years of Research. Microbiol. Mol. Biol. Rev. MMBR 80, 679–731.
Li, X., Teng, M., Reinherz, E.L., Wang, J., 2013. Strict Major Histocompatibility Complex Molecule Class-Specific Binding by Co-Receptors Enforces MHC-Restricted αβ TCR Recognition during T Lineage Subset Commitment. Front. Immunol. 4.
Li, Y., Deng, L., Liang, J., Dong, G.-H., Xia, Y.-L., Fu, Y.-X., Liu, S.-Q., 2020. Molecular dynamics simulations reveal distinct differences in conformational dynamics and thermodynamics between the unliganded and CD4-bound states of HIV-1 gp120. Phys. Chem. Chem. Phys. 22, 5548–5560.
Li, Y., O'Dell, S., Walker, L.M., Wu, X., Guenaga, J., Feng, Y., Schmidt, S.D., McKee, K., Louder, M.K., Ledgerwood, J.E., Graham, B.S., Haynes, B.F., Burton, D.R., Wyatt, R.T., Mascola, J.R., 2011. Mechanism of Neutralization by the Broadly Neutralizing HIV-1 Monoclonal Antibody VRC01 ▿ . J. Virol. 85, 8954–8967.
Liu, J., Bartesaghi, A., Borgnia, M.J., Sapiro, G., Subramaniam, S., 2008. Molecular architecture of native HIV-1 gp120 trimers. Nature 455, 109–113.
Matthias, L.J., Yam, P.T.W., Jiang, X.-M., Vandegraaff, N., Li, P., Poumbourios, P., Donoghue, N., Hogg, P.J., 2002. Disulfide exchange in domain 2 of CD4 is required for entry of HIV-1. Nat. Immunol. 3, 727–732.
McCammon, J.A., Gelin, B.R., Karplus, M., 1977. Dynamics of folded proteins. Nature 267, 585–590.
McCammon, J.A., Harvey, S.C., 1988. Dynamics of Proteins and Nucleic Acids. Cambridge University Press.
Melikyan, G.B., Markosyan, R.M., Hemmati, H., Delmedico, M.K., Lambert, D.M., Cohen, F.S., 2000. Evidence That the Transition of HIV-1 Gp41 into a Six-Helix Bundle, Not the Bundle Configuration, Induces Membrane Fusion. J. Cell Biol. 151, 413–424.
Mengistu, M., Tang, A., Foulke, J.S., Blanpied, T.A., Gonzalez, M.W., Spouge, J.L., Gallo, R.C., Lewis, G.K., DeVico, A.L., 2017. Patterns of conserved gp120 epitope presentation on attached HIV-1 virions. Proc. Natl. Acad. Sci. 114, E9893–E9902.
Mizukami, T., Fuerst, T.R., Berger, E.A., Moss, B., 1988. Binding region for human immunodeficiency virus (HIV) and epitopes for HIV-blocking monoclonal antibodies of the CD4 molecule defined by site-directed mutagenesis. Proc. Natl. Acad. Sci. 85, 9273–9277.
Moebius, U., Harrisoni, S.C., Reinherz, E.L., n.d. Human immunodeficiency virus gpl20 binding C'C" ridge of CD4 domain 1 is also involved in interaction with class II major histocompatibility complex molecules 5.
Moonsamy, S., Bhakat, S., Walker, R.C., Soliman, M.E.S., 2016. Single Active Site Mutation Causes Serious Resistance of HIV Reverse Transcriptase to Lamivudine: Insight from Multiple Molecular Dynamics Simulations. Cell Biochem. Biophys. 74, 35–48.
Moore, P.L., Gray, E.S., Choge, I.A., Ranchobe, N., Mlisana, K., Abdool Karim, S.S., Williamson, C., Morris, L., CAPRISA 002 Study Team, 2008. The c3-v4 region is a major target of autologous neutralizing antibodies in human immunodeficiency virus type 1 subtype C infection. J. Virol. 82, 1860–1869.
Morita, E., Sandrin, V., McCullough, J., Katsuyama, A., Baci Hamilton, I., Sundquist, W.I., 2011. ESCRT-III protein requirements for HIV-1 budding. Cell Host Microbe 9, 235–242.
Muhammed, M.T., Aki-Yalcin, E., 2019. Homology modeling in drug discovery: Overview, current applications, and future perspectives. Chem. Biol. Drug Des. 93, 12–20.

Murray, J.M., Kelleher, A.D., Cooper, D.A., 2011. Timing of the Components of the HIV Life Cycle in Productively Infected CD4+ T Cells in a Population of HIV-Infected Individuals. J. Virol. 85, 10798–10805.

Musyoki, A., Mothapo, K., Rakgole, J., Lukhwareni, A., Bessong, P., Selabe, G., Bredell, H., Williamson, C., Mphahlele, M.J., 2012. Genetic Characterization of HIV Before Widespread Testing of HIV Vaccine Candidates at a Clinical Trial Site in Pretoria, South Africa. AIDS Res. Hum. Retroviruses 28, 1131–1138.

Nakane, S., Iwamoto, A., Matsuda, Z., 2015. The V4 and V5 Variable Loops of HIV-1 Envelope Glycoprotein Are Tolerant to Insertion of Green Fluorescent Protein and Are Useful Targets for Labeling. J. Biol. Chem. 290, 15279–15291.

Nielsen, S.O., Lopez, C.F., Srinivas, G., Klein, M.L., 2004. Coarse grain models and the computer simulation of soft materials. J. Phys. Condens. Matter 16, R481–R512.

NISC Comparative Sequencing Program, Liao, H.-X., Lynch, R., Zhou, T., Gao, F., Alam, S.M., Boyd, S.D., Fire, A.Z., Roskin, K.M., Schramm, C.A., Zhang, Z., Zhu, J., Shapiro, L., Mullikin, J.C., Gnanakaran, S., Hraber, P., Wiehe, K., Kelsoe, G., Yang, G., Xia, S.-M., Montefiori, D.C., Parks, R., Lloyd, K.E., Scearce, R.M., Soderberg, K.A., Cohen, M., Kamanga, G., Louder, M.K., Tran, L.M., Chen, Y., Cai, F., Chen, S., Moquin, S., Du, X., Joyce, M.G., Srivatsan, S., Zhang, B., Zheng, A., Shaw, G.M., Hahn, B.H., Kepler, T.B., Korber, B.T.M., Kwong, P.D., Mascola, J.R., Haynes, B.F., 2013. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. Nature 496, 469–476.

Nizami, B., Sydow, D., Wolber, G., Honarparvar, B., 2016. Molecular insight on the binding of NNRTI to K103N mutated HIV-1 RT: molecular dynamics simulations and dynamic pharmacophore analysis. Mol. Biosyst. 12, 3385–3395. https://doi.org/10.1039/c6mb00428h

Ovchinnikov, V., Louveau, J.E., Barton, J.P., Karplus, M., Chakraborty, A.K., 2018. Role of framework mutations and antibody flexibility in the evolution of broadly neutralizing antibodies. eLife 7.

Owen, G.R., Le, D., Stoychev, S., Cerutti, N.M., Papathanasopoulos, M., 2018. Redox exchange of the disulfides of human two-domain CD4 regulates the conformational dynamics of each domain, providing insight into its mechanisms of control. Biochem. Biophys. Res. Commun. 497, 811–817.

Pandey, V., Tiwari, G., Mall, V.S., Tiwari, R.K., Ojha, R.P., 2019. Interaction between gp120 and ligand in HIV-1 env protein: Molecular dynamics simulations and binding free energy calculations. AIP Conf. Proc. 2142, 110031.

Pantophlet, R., Burton, D.R., 2006. GP120: Target for Neutralizing HIV-1 Antibodies 33.

Pantophlet, R., Wrin, T., CavaciniJ, L.A., Robinson, J.E., Burton, D.R., 2008. Neutralizing activity of antibodies to the V3 loop region of HIV-1 gp120 relative to their epitope fine specificity. Virology 381, 251–260.

Pettit, S.C., Everitt, L.E., Choudhury, S., Dunn, B.M., Kaplan, A.H., 2004. Initial Cleavage of the Human Immunodeficiency Virus Type 1 GagPol Precursor by Its Activated Protease Occurs by an Intramolecular Mechanism. J. Virol. 78, 8477–8485.

Popovic, M., Sarngadharan, M., Read, E., Gallo, R., 1984. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science 224, 497–500.

Rademeyer, C., Moore, P.L., Taylor, N., Martin, D.P., Choge, I.A., Gray, E.S., Sheppard, H.W., Gray, C., Morris, L., Williamson, C., 2007. Genetic characteristics of HIV-1 subtype C envelopes inducing cross-neutralizing antibodies. Virology 368, 172–181.

Raja, A., Venturi, M., Kwong, P., Sodroski, J., 2003. CD4 Binding Site Antibodies Inhibit Human Immunodeficiency Virus gp120 Envelope Glycoprotein Interaction with CCR5. J. Virol. 77, 713–718.

Rambaut, A., Posada, D., Crandall, K.A., Holmes, E.C., 2004. The causes and consequences of HIV evolution. Nat. Rev. Genet. 5, 52–61.

Raska, M., Takahashi, K., Czernekova, L., Zachova, K., Hall, S., Moldoveanu, Z., Elliott, M.C., Wilson, L., Brown, R., Jancova, D., Barnes, S., Vrbkova, J., Tomana, M., Smith, P.D., Mestecky, J., Renfrow, M.B., Novak, J., 2010. Glycosylation Patterns of HIV-1 gp120 Depend on the Type of Expressing Cells and Affect Antibody Recognition. J. Biol. Chem. 285, 20860–20869.

Reeves, P.J., Callewaert, N., Contreras, R., Khorana, H.G., 2002. Structure and function in rhodopsin: High-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. Proc. Natl. Acad. Sci. U. S. A. 99, 13419–13424.

Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., Kaewkungwal, J., Chiu, J., Paris, R., Premsri, N., Namwat, C., de Souza, M., Adams, E., Benenson, M., Gurunathan, S., Tartaglia, J., McNeil, J.G., Francis, D.P., Stablein, D., Birx, D.L., Chunsuttiwat, S., Khamboonruang, C., Thongcharoen, P., Robb, M.L., Michael, N.L., Kunasol, P., Kim, J.H., 2009. Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand. N. Engl. J. Med. 361, 2209–2220.

Riniker, S., 2018. Fixed-Charge Atomistic Force Fields for Molecular Dynamics Simulations in the Condensed Phase: An Overview. J. Chem. Inf. Model. 58, 565–578.

Rizzuto, C.D., Wyatt, R., Hernández-Ramos, N., Sun, Y., Kwong, P.D., Hendrickson, W.A., Sodroski, J., 1998. A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. Science 280, 1949–1953.

S, A., Matchado, M.S., Snijesh, V.P., Kumar, A., Singh, S., 2019. An insight into anti-arthritic property OF C25H34O7 for Rheumatoid arthritis using molecular modelling and molecular dynamics approach. Inform. Med. Unlocked 16, 100145. https://doi.org/10.1016/j.imu.2018.11.001

Schwartz, J.A., Prado, I., Misamore, J., Weiss, D., Francis, J., Pal, R., Huaman, M., Cristillo, A., Lewis, G.K., Gallo, R.C., DeVico, A.L., Fouts, T.R., 2016. An HIV gp120-CD4 Immunogen Does Not Elicit Autoimmune Antibody Responses in Cynomolgus Macaques. Clin. Vaccine Immunol. CVI 23, 618–27.

Settanni, G., Fersht, A.R., 2008. High temperature unfolding simulations of the TRPZ1 peptide. Biophys. J. 94, 4444–4453.

Shahlaei, M., Madadkar-Sobhani, A., Mahnam, K., Fassihi, A., Saghaie, L., Mansourian, M., 2011. Homology modeling of human CCR5 and analysis of its binding properties through molecular docking and molecular dynamics simulation. Biochim. Biophys. Acta BBA - Biomembr., Including the Special Section: Protein translocation across or insertion into membranes 1808, 802–817.

Sharma, D., Balamurali, M.M., Chakraborty, K., Kumaran, S., Jeganathan, S., Rashid, U., Ingallinella, P., Varadarajan, R., 2005. Protein minimization of the gp120 binding region of human CD4. Biochemistry 44, 16192–16202.

Spiwok, V., Lipovová, P., Skálová, T., Dusková, J., Dohnálek, J., Hasek, J., Russell, N.J., Králová, B., 2007. Cold-active enzymes studied by comparative molecular dynamics simulation. J. Mol. Model. 13, 485–497.

Thompson, J., Kumar, P., Yi, J., Bowder, D., Wood, C., Xiang, S.-H., 2015. Stabilization of Outer Domain of gp120 from HIV-1 Subtype C for Vaccine Immunogen Design. Procedia Vaccinol., Procedia of the 8th Vaccine & ISV Congress, Philadelphia, USA, 2015 9, 6–15.

Torrents de la Peña, A., de Taeye, S.W., Sliepen, K., LaBranche, C.C., Burger, J.A., Schermer, E.E., Montefiori, D.C., Moore, J.P., Klasse, P.J., Sanders, R.W., 2018. Immunogenicity in Rabbits of HIV-1 SOSIP Trimers from Clades A, B, and C, Given Individually, Sequentially, or in Combination. J. Virol. 92.

Townsley, S., Li, Y., Kozyrev, Y., Cleveland, B., Hu, S.-L., 2016. Conserved Role of an N-Linked Glycan on the Surface Antigen of Human Immunodeficiency Virus Type 1 Modulating Virus Sensitivity to Broadly Neutralizing Antibodies against the Receptor and Coreceptor Binding Sites. J. Virol. 90, 829–841.

Tuckerman, M.E., Berne, B.J., 1991. Stochastic molecular dynamics in systems with multiple time scales and memory friction. J. Chem. Phys. 95, 4389–4396.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., Berendsen, H.J.C., 2005. GROMACS: fast, flexible, and free. J. Comput. Chem. 26, 1701–1718.

Vandegraaff, N., Engelman, A., 2007. Molecular mechanisms of HIV integration and therapeutic intervention. Expert Rev. Mol. Med. 9, 1–19.

Walker, L.M., Huber, M., Doores, K.J., Falkowska, E., Pejchal, R., Julien, J.-P., Wang, S.-K., Ramos, A., Chan-Hui, P.-Y., Moyle, M., Mitcham, J.L., Hammond, P.W., Olsen, O.A., Phung, P., Fling, S., Wong, C.-H., Phogat, S., Wrin, T., Simek, M.D., Principal Investigators, P.G., Koff, W.C., Wilson, I.A., Burton, D.R., Poignard, P., 2011. Broad neutralization coverage of HIV by multiple highly potent antibodies. Nature 477, 466–470.

Wang, H., Cohen, A.A., Galimidi, R.P., Gristick, H.B., Jensen, G.J., Bjorkman, P.J., 2016. Cryo-EM structure of a CD4-bound open HIV-1 envelope trimer reveals structural rearrangements of the gp120 V1V2 loop. Proc. Natl. Acad. Sci. 113, E7151–E7158.

Wang, J., Meijers, R., Xiong, Y., Liu, J., Sakihama, T., Zhang, R., Joachimiak, A., Reinherz, E.L., 2001. Crystal structure of the human CD4 N-terminal two-domain fragment complexed to a class II MHC molecule. Proc. Natl. Acad. Sci. 98, 10799–10804.

Wang, W., Donini, O., Reyes, C.M., Kollman, P.A., 2001. Biomolecular Simulations: Recent Developments in Force Fields, Simulations of Enzyme Catalysis, Protein-Ligand, Protein-Protein, and Protein-Nucleic Acid Noncovalent Interactions. Annu. Rev. Biophys. Biomol. Struct. 30, 211–243.

Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W., Swanstrom, R., Burch, C.L., Weeks, K.M., 2009. Architecture and Secondary Structure of an Entire HIV-1 RNA Genome. Nature 460, 711–716. h

Wei, X., Decker, J.M., Wang, S., Hui, H., Kappes, J.C., Wu, X., Salazar-Gonzalez, J.F., Salazar, M.G., Kilby, J.M., Saag, M.S., Komarova, N.L., Nowak, M.A., Hahn, B.H., Kwong, P.D., Shaw, G.M., 2003. Antibody neutralization and escape by HIV-1. Nature 422, 307–312.

Weiss, C.D., 2003. HIV-1 gp41: mediator of fusion and target for inhibition. AIDS Rev. 5, 214–221.

Wilen, C.B., Tilton, J.C., Doms, R.W., 2012. HIV: cell binding and entry. Cold Spring Harb. Perspect. Med. 2.

Wyatt, R., Sodroski, J., 1998. The HIV-1 Envelope Glycoproteins: Fusogens, Antigens, and Immunogens. Science 280, 1884–1888.

Xiang, S.-H., Finzi, A., Pacheco, B., Alexander, K., Yuan, W., Rizzuto, C., Huang, C.-C., Kwong, P.D., Sodroski, J., 2010. A V3 Loop-Dependent gp120 Element Disrupted by CD4 Binding Stabilizes the Human Immunodeficiency Virus Envelope Glycoprotein Trimer. J. Virol. 84, 3147–3161.

Yang, X., Tomov, V., Kurteva, S., Wang, L., Ren, X., Gorny, M.K., Zolla-Pazner, S., Sodroski, J., 2004. Characterization of the outer domain of the gp120 glycoprotein from human immunodeficiency virus type 1. J. Virol. 78, 12975–86. https://doi.org/10.1128/JVI.78.23.12975-12986.2004

Yokoyama, M., Nomaguchi, M., Doi, N., Kanda, T., Adachi, A., Sato, H., 2016. In silico Analysis of HIV-1 Env-gp120 Reveals Structural Bases for Viral Adaptation in Growth-Restrictive Cells. Front. Microbiol. 7.

Yoon, V., Fridkis-Hareli, M., Munisamy, S., Lee, J., Anastasiades, D., Stevceva, L., 2010. The GP120 Molecule of HIV-1 and its Interaction with T Cells. Curr. Med. Chem. 17, 741–749.

Yuan, T., Li, J., Zhang, M.-Y., 2013. HIV-1 Envelope Glycoprotein Variable Loops Are Indispensable for Envelope Structural Integrity and Virus Entry. PLoS ONE 8, e69789.

Zhang, Y., Niu, Y., Tian, J., Liu, X., Yao, X., Liu, H., 2017. The glycan-mediated mechanism on the interactions of gp120 with CD4 and antibody: Insights from molecular dynamics simulation. Chem. Biol. Drug Des. 90, 1237–1246. https://doi.org/10.1111/cbdd.13045

Zhang, Y., Pan, D., Shen, Y., Jin, N., Liu, H., Yao, X., 2012. Understanding the molecular mechanism of the broad and potent neutralization of HIV-1 by antibody VRC01 from the perspective of molecular dynamics simulation and binding free energy calculations. J. Mol. Model. 18, 4517–4527.

Zhen, A., Krutzik, S.R., Levin, B.R., Kasparian, S., Zack, J.A., Kitchen, S.G., 2014. CD4 Ligation on Human Blood Monocytes Triggers Macrophage Differentiation and Enhances HIV Infection. J. Virol. 88, 9934–9946.

Zhou, T., Georgiev, I., Wu, X., Yang, Z.-Y., Dai, K., Finzi, A., Kwon, Y.D., Scheid, J., Shi, W., Xu, L., Yang, Y., Zhu, J., Nussenzweig, M.C., Sodroski, J., Shapiro, L., Nabel, G.J., Mascola, J.R., Kwong, P.D., 2010. Structural Basis for Broad and Potent Neutralization of HIV-1 by Antibody VRC01. Science 329, 811–817.

Ethics Waiver

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

FACULTY OF HEALTH SCIENCES

University of the Witwatersrand Student Ethics Declaration Form

(To be completed during the protocol assessor meeting)

Background

All Research conducted by a University of the Witwatersrand student, with human subjects or animals, requires approval by the Wits Human Research Ethics Committee or Animal Research Ethics Committee, respectively.

If research has been undertaken without the necessary ethics approvals, this is considered an ethics violation. This will be reported to the relevant structures, the data will have to be discarded, and in the case of students, they cannot use the data towards their degree.

To prevent any ethics violations, the ethics requirements for the proposed project will be discussed with you at the protocol assessment.

Declaration

Based on the current protocol assessment (and any proposed changes suggested by the assessor committee), we, the undersigned, understand that the proposed research requires:

1. Human Research Ethics clearance certificate	Yes	**No**
a. Covered under existing supervisor ethics	Yes	**No**
b. Requires a new HREC application	Yes	**No**
2. Animal Research Ethics clearance certificate	Yes	**No**
3. No Human or Animal Ethics Clearance	**Yes**	No
4. Unclear, will seek appropriate guidance from the HREC/AREC committees (whichever relevant)	Yes	**No**

Signatures

Supervisor/s: _____ Student: _____

_____

Date: 25/05/2021

11 March 2019/MP