Institutionen för mikrobiologi, tumör och cellbiologi (MTC)

Amino acid predictors of HIV-1 coreceptor use in different subtypes and their application to antiretroviral treatment

AKADEMISK AVHANDLING
som för avläggande av medicine doktorsexamen vid Karolinska Institutet offentligen försvaras i Hillarpsalen, Retzius väg 8, Karolinska Institutet Campus Solna

Fredagen den 31 januari 2014, kl. 09.30

av
Lotta Pramanik Sollerkvist
M.D.

Huvudhandledare (nuvarande):
Professor Gunilla Karlsson Hedestam
Karolinska Institutet
Institutionen för mikrobiologi, tumör och cellbiologi

Huvudhandledare (till 10/12/2012):
Professor Anneka Ehrnst †
Karolinska Institutet
Institutionen för mikrobiologi, tumör och cellbiologi

Bihandledare:
Dr. Peter Clevestig
Stockholm International Peace Research Institute
Arms Control and Non-proliferation Programme &
Karolinska Institutet
Institutionen för mikrobiologi, tumör och cellbiologi

Fakultetsopponent:
Professor Ronald Swanstrom
University of North Carolina at Chapel Hill
Department of Biophysics and Biochemistry

Betygsnämnd:
Professor Patrik Medstrand
Lunds Universitet
Institutionen för laboratoriemedicin i Malmö

Docent Annika Karlsson
Karolinska Institutet
Institutionen för laboratoriemedicin

Professor Anna Mia Ekström
Karolinska Institutet
Institutionen för folkhälsovetenskap

Stockholm 2014

From the DEPARTMENT OF MICROBIOLOGY, TUMOR AND CELL BIOLOGY
Karolinska Institutet, Stockholm, Sweden

AMINO ACID PREDICTORS OF HIV-1 CORECEPTOR USE IN DIFFERENT SUBTYPES AND THEIR APPLICATION TO ANTIRETROVIRAL TREATMENT

Lotta Pramanik Sollerkvist

Karolinska Institutet

Stockholm 2014

The cover illustration shows the structure of the HIV-1 virion created by Peter Clevestig.
All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet. Printed by US-AB.

ISBN 978-91-7549-450-0

In memory of
my inspiring teacher
and dearest friend
Anneka Ehrnst

ABSTRACT

The human immunodeficiency virus type 1 (HIV-1) group M is responsible for the HIV pandemic and has great genetic variability. The main subtypes and circulating recombinant forms belonging to this group differ in their worldwide distribution, their disease progression patterns, as well as in their coreceptor use phenotype distributions. The coreceptor use phenotypes of HIV-1 are based on which coreceptor the virus uses for cell entry, with the two main coreceptors being CCR5 and CXCR4, corresponding accordingly to the R5 and X4 phenotype. The R5 phenotype is found early in the infection while the X4 emerges over time and is associated with disease progression. The main determinant of coreceptor use in HIV-1 is the third variable region (V3) of the HIV-1 envelope glycoprotein 120, which is the docking surface protein that attaches itself to the primary CD4 receptor and a coreceptor during cell entry.

In **papers I** and **II**, observations were made regarding the role of the V3 glycan and V3 charge in coreceptor use, based on sequences belonging to different HIV-1 subtypes retrieved from the Los Alamos HIV Sequence database, which contains sequences submitted from infected individuals all over the world. The V3 glycan was shown in **paper I** to be strongly associated with CCR5 use while a net high charge acquired from different positions in the V3 was shown in **paper II** to be important for CXCR4 use. As a result, a model adjustable for different subtypes was created, referred to as the glycan-charge model, for distinguishing between the coreceptor use phenotypes based on their biological properties that can be deduced from the V3 amino acid sequence.

CCR5 inhibitors are a class of antiretroviral drugs that target the CCR5 coreceptor, thereby blocking the entry of R5 viruses into cells. However, prior to their administration it is important to verify that a patient does not harbour CXCR4-using variants, which could otherwise be selected for. Biological methods of coreceptor use determination are expensive and time-consuming. Hence, coreceptor use prediction algorithms, which can predict the coreceptor use from HIV-1 V3 sequences, could help to make CCR5 inhibitors more universally accessible, but their prediction accuracy needs to be improved.

Infections with HIV-1 subtype C, which is the dominating subtype worldwide and in sub-Saharan Africa, are usually associated with low CXCR4 use, but several studies have found an increased CXCR4 use among treatment failure patients. To investigate this further, 24 treatment failure patients infected with subtype C in Botswana were in **paper III** compared with 26 treatment-naïve patients with regard to coreceptor use, which was determined using the coreceptor use prediction algorithm Geno2pheno with a false positive cut-off rate of 10 % as well as the glycan-charge model on population sequences. Increased CXCR4 use was found in the treatment-experienced group, suggesting that treatment with the only CCR5 inhibitor in clinical use to date, maraviroc, would be less suitable in this group, which is of special significance since maraviroc is mainly used as a salvage therapy drug.

Finally, in **paper IV** all currently available coreceptor use prediction algorithms, including the glycan-charge model algorithm, were evaluated from a CCR5 inhibitor

treatment perspective in a uniquely suited testing material, which consisted of V3 sequences of the major HIV-1 subtypes retrieved from the Los Alamos HIV Sequence database. A rigorous scrutiny of the original source articles was performed to verify that the reported coreceptor use was determined biologically. The results showed that learning algorithms were found to perform well in all studied subtypes, along with subtype-specific complex rule algorithms.

In summary, **papers I** and **II** elucidated the biological properties of coreceptor use in different subtypes, which could be determined using the V3 amino acid sequence, **paper III** applied this knowledge to help investigate the increased CXCR4 use in subtype C infected treatment failure patients, while **paper IV** compared all current coreceptor use prediction algorithms, including the glycan-charge model based on the observations in **papers I** and **II**, and applied in **paper III**, from a CCR5 inhibitor treatment perspective.

LIST OF PUBLICATIONS

I. Clevestig P, **Pramanik L**, Leitner T, Ehrnst A.
CCR5 use of human immunodeficiency virus type 1 is associated closely with the gp120 V3 loop N-linked glycosylation site.
J Gen Virol. 2006 Mar;87(Pt 3):607-12.

II. **Pramanik L**, Fried U, Clevestig P, Ehrnst A.
Charged amino acid patterns of coreceptor use in the major subtypes of human immunodeficiency virus type 1.
J Gen Virol. 2011 Aug;92(Pt 8):1917-22.

III. **Pramanik Sollerkvist L**,* Gaseitsiwe S,* Mine M, Sebetso G, Mphoyakgosi T, Diphoko T, Essex M, Ehrnst A.
*These authors contributed equally to this work.
Increased CXCR4 use of HIV-1 subtype C identified by population sequencing in patients failing antiretroviral treatment compared with treatment-naïve patients in Botswana.
AIDS Res Hum Retroviruses. 2013 Nov 8. E-published ahead of print.

IV. **Pramanik Sollerkvist L**, Caridha R, Liotta M, Clevestig P, Ehrnst A.
Comparison of bioinformatic prediction models for coreceptor use of HIV-1 in different subtypes from a CCR5 inhibitor treatment perspective.
Manuscript.

“Learning never exhausts the mind”

“The noblest pleasure is the joy of understanding”

– Leonardo da Vinci

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AIDS	Acquired immunodeficiency syndrome
ART	Antiretroviral treatment
CA	Capsid protein
CCR3	C-C chemokine receptor type 3
CCR5	C-C chemokine receptor type 5
CRF01_AE	Circulating recombinant form 01_AE
CXCR4	C-X-C chemokine receptor type 4
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
ECL2	Extracellular loop 2
env	Envelope gene
FDA	U.S. Food and Drug Administration
gag	Group antigen gene
G2P	Geno2pheno [coreceptor] algorithm
GALT	Gut-associated lymphoid tissue
IN	Integrase
MA	Matrix protein
MIP-1α	Macrophage inflammatory protein-1α
mRNA	Messenger ribonucleic acid
M-tropic	Macrophage/monocyte-tropic phenotype
NC	Nucleocapsid proteins
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NNT	Asparagine-Asparagine-Threonine, a common glycosylation motif
NRTI	Nucleoside reverse transcriptase inhibitor
NSI	Non-syncytium inducing phenotype
HAART	Highly Active Antiretroviral Treatment
HIV-1	Human immunodeficiency virus type 1
PBMC	Peripheral blood mononuclear cells
PCR	Polymerase chain reaction
PI	Protease inhibitor
pol	Polymerase gene
PR	Protease
RANTES	Regulated on Activation Normal T cell Expressed and Secreted
RT	Reverse transcriptase
SDF-1	Stromal cell-derived factor 1
SI	Syncytium inducing phenotype
SIV	Simian immunodeficiency virus
tRNA	Transfer ribonucleic acid
T-tropic	T-cell-tropic phenotype
V3	3rd variable region of envelope glycoprotein 120
-viroc	Viral receptor occupancy

AMINO ACID IUPAC CODES

Code	Abbreviation	Amino Acid
A	Ala	Alanine
B	Asx	Aspartic acid or Asparagine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
X	Xaa	Any amino acid
Y	Tyr	Tyrosine
Z	Glx	Glutamine or Glutamic acid

1 BACKGROUND

1.1 HIV-1 BASICS

One of the greatest pandemics of our time is caused by the human immunodeficiency virus (HIV), which is the etiological agent of acquired immunodeficiency syndrome (AIDS), a condition where the patient's immune system deteriorates slowly, eventually allowing the onset of opportunistic infections, which ultimately lead to premature death if no interventions are made.

1.1.1 Origins

HIV is closely related to the simian immunodeficiency virus (SIV), which primarily infects nonhuman primates, including chimpanzees (*Pan troglodytes troglodytes*), gorillas (*Gorilla gorilla*), and sooty mangabeys (*Cercocebus atys*), inducing a condition reminiscent of AIDS. It is believed that HIV is of simian origin, and is the result of multiple cross-species transmissions of SIV from nonhuman primates to humans [1, 2]. Four different cross-species transmission events of SIV from chimpanzees and gorillas gave rise to HIV type 1 (HIV-1) [1-4], while the less pathogenic HIV type 2 (HIV-2), which is rarely found outside of West Africa [5], was the result of eight different cross-species transmissions of SIV to humans from sooty mangabeys [1, 2, 6, 7]. The most likely mode of transmission in these instances was due to contact with blood or other bodily fluids of nonhuman primates infected with SIV during the hunt or the handling of primate bushmeat [1, 2, 8, 9], while bites or other wounds inflicted from pet nonhuman primates constitute another possible mode of transmission [2, 9]. Despite the occurrence of multiple cross-species transmission events of SIV, only one of these, the origin of the major HIV-1 subgroup, HIV-1 group M, accounts for the HIV-1 pandemic [1, 2, 9].

1.1.2 Pandemic

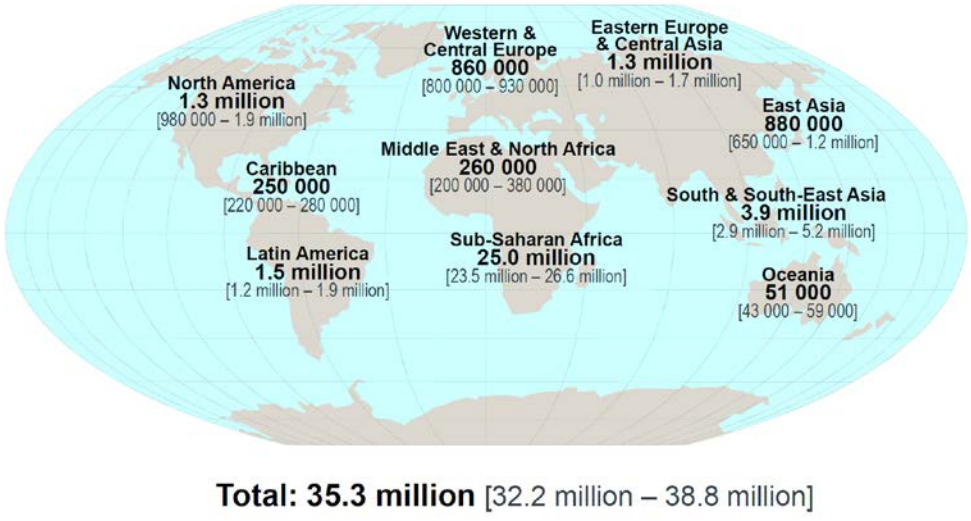

Figure 1. Adults and children estimated to be living with HIV in 2012. Credit: UNAIDS/ ONUSIDA 2013 [10].

The first emergence of HIV-1 group M has been estimated using phylogenetical and statistical analyses to have occurred in 1931 (confidence interval of 1915-1941) [11], or even earlier in 1908 (1884-1924) [12]. The most probable location of the early epidemic is considered to have been Léopoldville [1, 9], now known as Kinshasa, the capital city of the Democratic Republic of the Congo, where also the earliest samples of HIV-1 group M were found, dating from 1959 [13] and 1960 [12]. Kinshasa was at the time a growing community due to colonial administration and trade, and was the largest city in west central Africa where urban lifestyle was continually on the increase, which probably facilitated the spread of the epidemic [1, 9, 12, 13].

The AIDS epidemic was not brought to global attention until decades later in 1981, when opportunistic infections as well as uncommon malignancies started emerging in young male homosexuals in the US [14, 15]. Cases were also identified among intravenous drug users, Haitians, and hemophiliacs, and the Center for Disease Control named the new disease acquired immunodeficiency syndrome or AIDS, in 1982 [15]. In 1983, the cause of AIDS was identified to be a new human retrovirus [16, 17], at first known by the name human T-cell lymphotropic virus type III (HTLV-III), but subsequently renamed human immunodeficiency virus (HIV) in 1986 by the International Committee of the Taxonomy of Viruses [18]. The first blood test for the detection of HIV [19, 20] was approved for screening blood donors by the U.S Food and Drug Administration (FDA) in 1985 [15], while the first drug against HIV infection arrived in 1987 in the form of the cancer drug zidovudine (AZT), albeit the treatment was not very successful due to the quick emergence of drug-resistant variants [15]. The morbidity and mortality rates remained high until the turning point came in 1996 when the triple drug regimen combination therapy was introduced [21-23], known as Highly Active Antiretroviral Treatment (HAART), transforming AIDS from a fatal disease into a somewhat controllable condition [15]. Nevertheless, the antiretroviral treatment continued to be cumbersome and wrought with side effects for years to come until new and improved drug regimens were developed during the last decade [24-26]. This development has improved the quality of life and considerably increased the life expectancy of HIV-infected patients in high-income, industrialized countries [27] where the latest treatment is readily available as opposed to in low-income countries [28], where the majority of afflicted patients reside [10].

Currently, 35.3 million adults and children are estimated to be living with HIV with 25 million, the absolute majority, located in Sub-Saharan Africa (**Figure 1**) [10]. In 2012, 1.6 million people died of AIDS, while 2.3 million became newly infected [10], indicating the continual endurance of the pandemic, with different factors contributing to its sustainment, such as drug availability [28], lack of a vaccine [15, 29], socioeconomic conditions [30, 31], and risk-behaviour [32].

1.1.2.1 Botswana

Botswana is a sub-Saharan country where 17.6 % of the population aged above 18 months is infected with HIV, while the HIV prevalence among adults is 25 % and among pregnant women is as high as 30.4 % [33], rendering Botswana the country with the second-highest infection rate in the world [34]. As a result of the high HIV mortality and infection rates, the life expectancy in Botswana has radically decreased

[33]. In 2001, Botswana adopted the United Nations General Assembly Special Session on HIV and AIDS (UNGASS) Declaration of Commitment, leading among other things to the introduction of a national antiretroviral treatment program in 2002 [33]. Since then, the mortality rate has been reduced by 60 %, and 95 % receive antiretroviral treatment of those who are eligible for it [33].

1.1.3 Structure

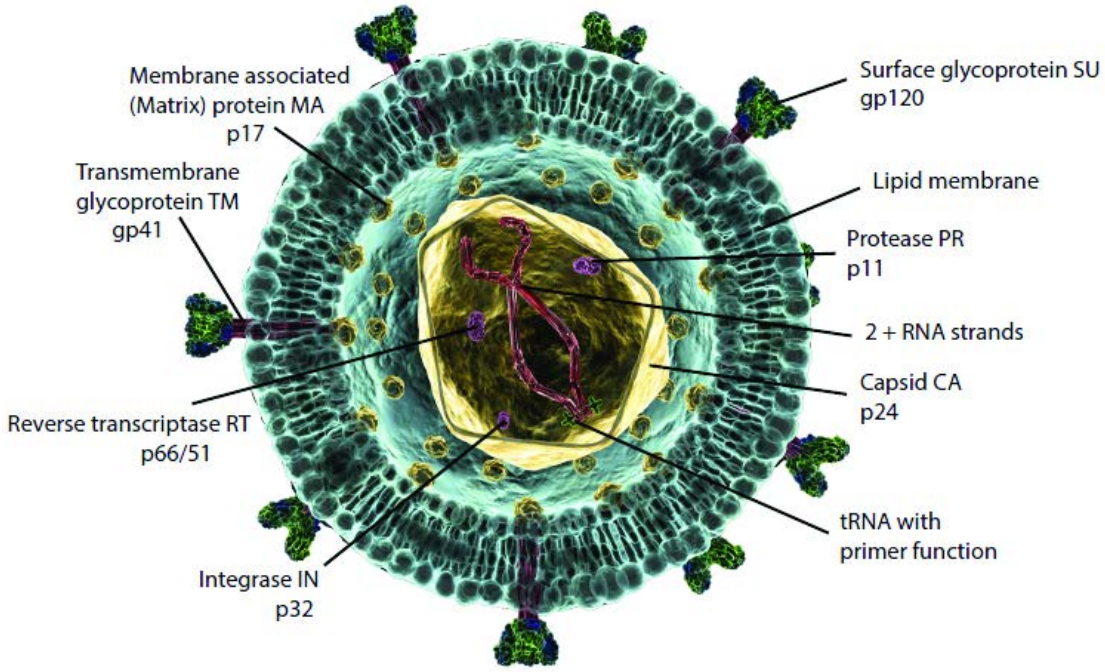

Figure 2. The structure of the HIV-1 virion. Adapted with permission from an illustration by Peter Clevestig.

HIV-1 is a *Lentivirus* that belongs to the *Retroviridae* family, which is distinguished by the presence of a reverse transcriptase enzyme. A mature virion (**Figure 2)** is roughly 145 nm (± 25 nm) in diameter [35] and consists of a host-derived lipid bilayer (the envelope) and a truncated cone-shaped capsid (p24). The components of the capsid are the enzymes reverse transcriptase (p66/p51), integrase (p32), and protease (p11), the nucleocapsid core proteins (p7, p9), tRNA molecules necessary for primer function during reverse transcription, and the RNA genome, which consists of two non-covalently-linked, positive sense RNA strands. The membrane associated matrix protein (p17) is found between the capsid and the envelope, while the trimeric spikes exposed on the surface of the virion consist of the docking glycoprotein (gp120) and the transmembrane fusion glycoprotein (gp41). Since the envelope is derived from the host cell membrane, other host specific proteins are also present on the surface, such as the major histocompatibility complex (MHC) [36] and intercellular adhesion molecule-1 (ICAM-1) [37].

1.1.4 Genome

As mentioned above, the HIV-1 genome is diploid, with two identical positive sense RNA strands, each approximately 9-10 kilobases long. The 5’ end of each strand has a methylated cap and the 3’end a poly-A tail, with long terminal repeat regions (LTRs) at both ends, which include promoter and enhancer sequences, as well as the primer

binding sequence (PBS), to which tRNA molecules associate that initiate reverse transcription. [38]

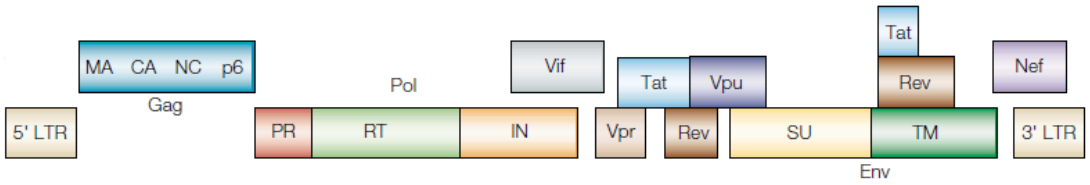

Figure 3. The HIV-1 genome. Adapted with permission from [39].

The genome has multiple open reading frames, encoding the three main structural polyproteins Gag (group antigen), Pol (polymerase), and Env (the envelope glycoproteins), but also a number of accessory and regulatory proteins (**Figure 3**). The *gag* gene encodes the matrix protein (MA), which is important for virion assembly and facilitates the transport of complementary DNA to the host cell's nucleus [40], the capsid protein (CA), and the nucleocapsid proteins (NC), which coat the viral RNA and are involved in its delivery during the virus assembly process [38]. The *pol* gene encodes the three enzymes, reverse trancriptase (RT), which transcribes the genomic RNA into complementary DNA (cDNA), integrase (IN), which integrates the cDNA with the host cell chromosomal DNA in the nucleus, and protease (PR), which completes the virion's maturation by cleaving the Gag-Pol polyprotein in the already assembled virion [38]. The *env* gene encodes the 160 kD glycoprotein gp160, which consists of the docking surface protein (SU) gp120 and the transmembrane (TM) fusion protein gp41, which stay non-covalently associated and form the trimeric spikes that are interspersed on the virion's surface [41]. Gp120 is involved in the cell attachment process through its binding to the primary receptor CD4 and subsequent to a conformational change in the spike structure, to one of several possible coreceptors. Coreceptor binding results in a second conformational change, which allows gp41 to mediate fusion with the cell membrane, enabling the virion to enter the cell [38].

The regulatory proteins of HIV-1 are Tat (Transactivator of transcription) and Rev (Regulator of virion expression). The function of Tat is to enhance the rate of transcription, while Rev facilitates the transport of spliced and full-length RNA from the cell nucleus to the cytoplasm, where translation and assembly of new virions take place [38].

HIV-1's accessory proteins comprise Nef (Negative factor), Vif (Virion infectivity factor), Vpr (Viral protein R), and Vpu (Viral protein U). Nef increases viral infectivity and is of importance for disease progression and viral spread. One of its main functions is to downregulate CD4 and MHC class I expression on the host cell surface, thereby evading detection by cytotoxic T cells [42].Vif is also important for infectivity. In particular, Vif induces the degradation of the antiviral factor APOBEC3G, thereby preventing APOBEC3G from incorporating itself into HIV-1 virions and interfering with the replication process [43]. Vpr has multiple functions, including a role in nuclear import of the pre-integration complex and cell growth arrest [38]. Vpu exhibits an antagonizing effect on tetherin, thereby preventing tetherin from inhibiting HIV-1 virion release from the cell [44]. Interestingly, Vpu is not encoded by HIV-2, which instead encodes Vpx (Viral protein X), which is absent in HIV-1. The function of Vpx

in HIV-2 involves the degradation of SAMHD1, which interferes with the reverse transcriptase step of lentiviruses [45].

1.1.5 Replication Cycle

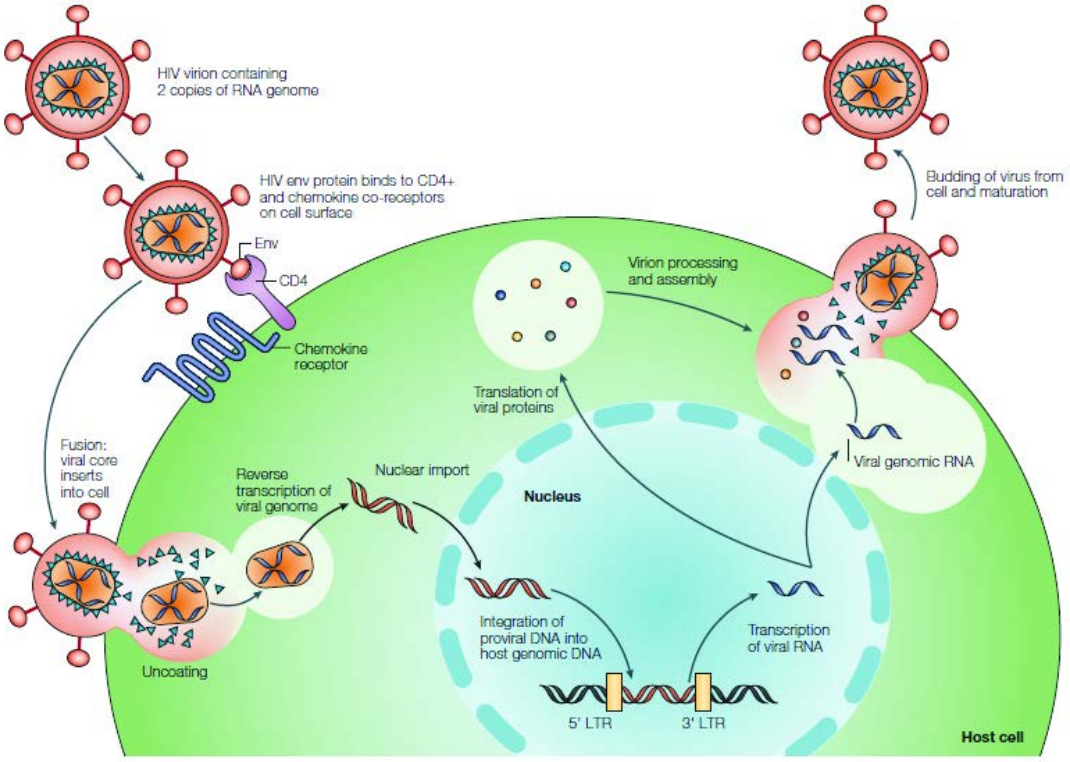

Figure 4. The HIV-1 replication cycle. Reproduced with permission from [46].

One HIV-1 replication cycle (**Figure 4**) takes on average approximately 1-2 days [47]. HIV-1's primary target cells are CD4 positive T-cell lymphocytes [16, 17, 48], as well as other CD4-expressing cells, such as macrophages, dendritic cells and microglia [48-51].

1.1.5.1 Attachment, binding and entry

The entry process of HIV-1 into a host cell can be broken down into four steps [52]. The first step is the binding of gp120 to CD4, the primary receptor [53, 54]. The exterior envelope glycoprotein gp120 forms heterodimers with the transmembrane-anchored fusion protein gp41 and together they constitute the heavily glycosylated trimeric spike on HIV-1's surface. One subunit of gp120 consists of five constant regions (C1-C5) interspersed with five variable regions (V1-V5), out of which four form flexible loop structures [52]. As a consequence of binding to CD4, a conformational change is induced in gp120 where V1/V2 and then V3 are rearranged, leading to the exposure of coreceptor binding determinants. These determinants consist mainly of the V3 loop [55, 56], but the bridging sheet, a four-stranded β sheet formed after gp120's binding to CD4, also plays a role in the coreceptor binding process [57-59]. This enables the second step, the interaction of gp120 with a coreceptor, usually the C-C chemokine receptor type 5 (CCR5) [60-62] or C-X-C chemokine receptor type 4 (CXCR4) [63, 64], allowing the exposure of the hydrophobic gp41 fusion peptide. The third step involves the "surfing" of the virus across the surface with the help of host cell machinery, in an attempt to find the optimal location for fusion with

the cell membrane [65]. Finally, the fourth step comprises the actual fusion. The gp41 fusion peptide is inserted into the cell membrane, thereby locking the viral and cell membranes together. This facilitates the fusion peptide in each gp41 subunit of the trimer to form a six-helix bundle, where the carboxy-terminal helical regions remain close to the viral membrane while the opposing amino-terminal helical regions stay close to the cell membrane [66, 67]. Consequently the six-helix bundle brings the membranes together, allowing the fusion of the virion with the cell membrane via the formation of a fusion pore, ensuing in the insertion of the viral core into the cytoplasm, where the process of uncoating starts [52].

1.1.5.2 Reverse transcription and integration

When the viral core has been released into the cytoplasm, the virion is uncoated, possibly with the aid of cyclophillin A, a cellular chaperon protein associated with the capsid [68]. Simultaneously, the reverse transciption complex is formed, containing MA, CA, NC, IN, Vpr, tRNA and the reverse transcriptase, RT [69]. Reverse transcription is initiated with the help of tRNA (Lys3) acting as a primer at the 5' end of the single-stranded RNA genome. The DNA polymerase function of the RT is responsible for creating a complimentary negative sense DNA strand (added to the 3' end of the tRNA molecule), while the RNase H function of the RT subsequently degrades the template RNA. The newly synthesized DNA strand is transferred to the 3' end of the viral RNA where the direct repeats section (R) is located, allowing for the continuation and completion of the negative sense DNA strand synthesis simultaneously with the RNA degradation. The RNA strand contains a purine rich sequence called the polypurine tract, which prevents the RNase H from degrading the RNA further, thereby leaving an RNA sequence to serve as a primer for the synthesis of the positive sense DNA strand. The positive sense DNA strand is synthesized until 18 nucleotides of the tRNA molecule are copied, allowing the RNase H to in turn degrade the tRNA molecule, leaving only an A ribonucleotide at the 5' end of the negative sense DNA strand. The positive sense DNA strand is transferred to the primer binding site region of the negative sense DNA strand, which will be complimentary with the 18 nucleotides of the positive strand, copied from the tRNA molecule. This allows the two DNA strands to anneal to each other, followed by the completion of the synthesis of the negative and positive strands, forming a linear double-stranded DNA of the full genome with some extra length at each end, comprised of U3-R-U5, where U3 was derived from the original RNA template's 3' end and U5 from its 5' end. These U3-R-U5 regions constitute the long terminal repeats (LTR). [70]

Late in the reverse transcription process, the reverse transcription complex is converted to a preintegration complex, which includes the integrase, IN and other viral proteins. The preintegration complex is actively imported to the cell nucleus, where IN removes two nucleotides from the 3' ends of the double-stranded DNA, allowing these ends to target phosphodiester bonds on opposite ends of the aimed section of the host cell's DNA, leading to the joining of the viral DNA's 3' ends to the host DNA's 5' ends. Subsequently, the unpaired two bases at the 5' ends of the viral DNA are removed, allowing the ligation of these 5' ends to the host DNA. IN catalyzes the processing of the 3' ends and the DNA transfer stage, but other events are likely catalyzed with the help of cellular enzymes. The integration of viral DNA can occur at multiple locations

in the host genome, and the integrated DNA is called a provirus, which forms the basis of retroviral infection. [71]

1.1.5.3 Transcription and translation

The integrated provirus can remain latent or be transcribed by the host cell's RNA polymerase II upon activation of the cell cycle, producing full-length mRNAs. The mRNAs are spliced in multiple ways [72-74], generating approximately 30 different transcripts. The early transcripts are multiply spliced to produce the regulatory proteins Tat and Rev, as well as the accessory protein Nef. Tat enhances the transcription process, Rev regulates the export of both spliced and unspliced mRNA to the cytoplasm, and Nef downregulates molecules on the cell surface such as CD4, which helps the virus to evade the immune system. The proteins Vpr, Vpu, Vif, and Env are produced from mRNAs that are singly spliced, while the unspliced mRNA transcript forms the Gag-Pol polyprotein. [75]

HIV-1 uses free polyribosomes in the cytoplasm to translate its mRNA into proteins. As mentioned above, HIV-1 uses alternative splicing of mRNA to generate different proteins to make the best use of its relatively small genome. Ribosomal frameshifting is another strategy utilized by HIV-1 for this purpose, which allows the production of Gag-Pro-Pol polyproteins even though the *gag* and *pol* genes are in different open reading frames that nonetheless overlap [76], thereby also maintaining the correct ratio between produced Gag polyproteins and Gag-Pro-Pol polyproteins necessary for the regulation of the replication cycle and for the correct amount of monomers of replication enzymes to be packaged during the assembly of new virion particles [77]. The Env glycoprotein precursor, gp160, is translated from a bicistrionic mRNA, followed by co-translational modification in the lumen of the endoplasmic reticulum, where N-linked oligomannosyl carbohydrate residues are added and trimmed [78]. The gp160 precursor glycoprotein is then transported to the Golgi complex where gp160 is cleaved into gp120 and gp41 by cellular proteases and the N-linked glycans are modified to generate complex type carbohydrates [78]. After these maturation steps, the glycoproteins are transported to the cell surface where they will become part of the envelopes of budding virus particles.

1.1.5.4 Assembly, release and maturation

The assembly of the virion takes place at the cell membrane, where the envelope trimeric spikes have already been incorporated, intermingled with host cell surface proteins. The different domains of the Gag polyprotein direct the assembly process: MA binds to the cell membrane and recruits the envelope glycoproteins, CA facilitates the necessary protein-protein interactions and produces the capsid, and NC seizes the viral genome and acts as a nucleic acid chaperone to tRNA during its annealing to the genome as a primer. The different viral components are gathered into spherical immature particles, where the Gag polyproteins are attached to the envelopes and radially projected toward the centers of the particles. As the budding takes place, PR is activated, resulting in the cleavage of the Gag polyproteins into MA, CA, NC, and p6, and the Pol section of Gag-Pro-Pol polyproteins into RT, IN, and PR, converting the produced viral particles into mature virions capable of infecting and replicating in new host cells. [79]

1.1.6 Pathogenesis

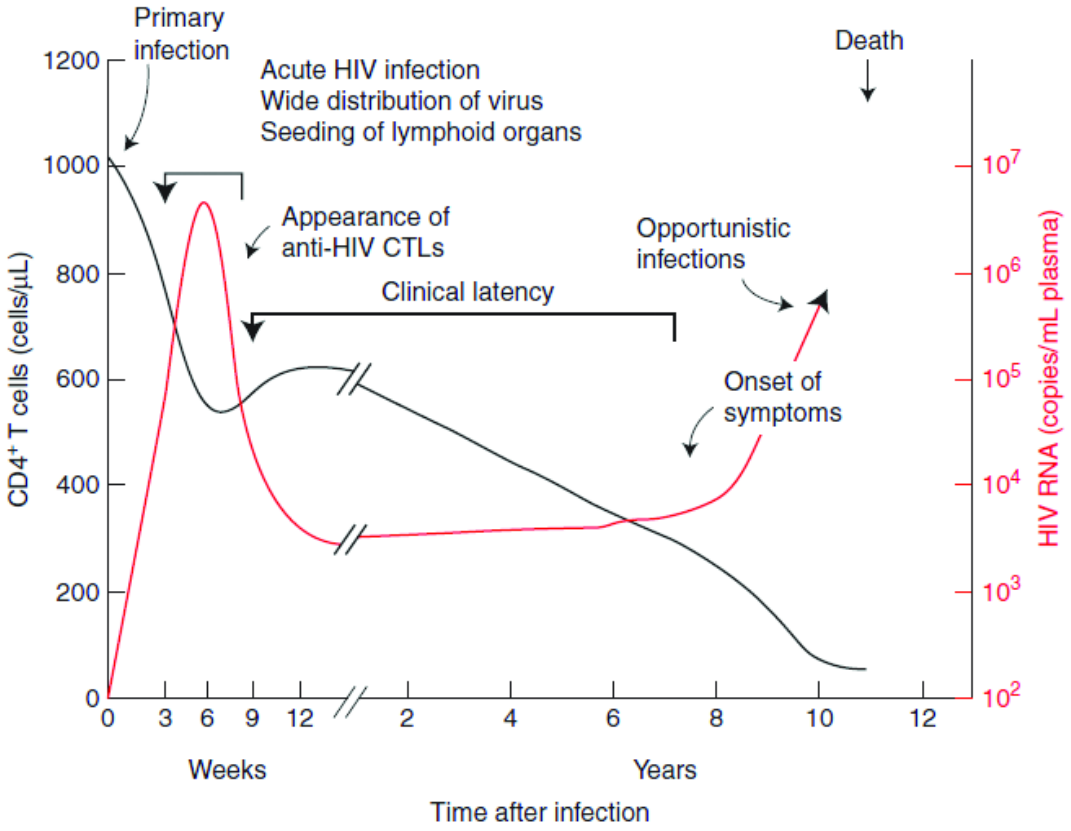

Figure 5. The natural time course of HIV-1 infection. Reproduced with permission from [47].

Exposure to HIV-1 at mucosal surfaces or through percutaneous inoculation can lead to infection [30]. The main modes of transmission are sexual, vertical, and by intravenous drug injections [30]. Blood transfusions are another possible source of transmission in countries where donated blood is not always screened properly for HIV.

HIV-1 infection can be divided into different phases (**Figure 5**). The first is the eclipse phase, which lasts 1-2 weeks, when the virus spreads from the site of infection to other organs and tissues, such as CD4+ T cells in draining lymph nodes and gut-associated lymphoid tissue (GALT), where it is able to replicate [80]. During this stage, the virus cannot be detected in the blood, and the patient is asymptomatic, since the immune system has not yet mounted a proper response. The next 2-4 weeks mark the acute phase, also called the primary infection. Some patients become symptomatic during this stage with “flu-like” symptoms, such as fever, throat ache and enlarged lymph nodes. The virus levels in the blood peak during this period up to approximately 10^7 RNA copies/mL plasma, perhaps as a result of the lack of an early immune response and the increase in available CD4+ T cells as the immune system begins to mount an antibody and CD8+ T cell response against HIV-1. At the end of the acute phase, the virus levels in the blood drop dramatically, roughly 100-fold, coupled with the transitory decline of the amount of CD4+ T cells measured in the blood. [47]

The acute phase is sometimes divided into Fiebig stages, which give an indication of the type of laboratory markers that can be used for diagnosis at the different time points

[81]. HIV-1 RNA is detectable in the blood by polymerase chain reaction (PCR) after about day 10 of the infection (Fiebig I), viral p24 antigen appears roughly after day 17 (Fiebig II), while HIV-specific antibodies can be detected by a recombinant protein-based enzyme-linked immunosorbant assay after approximately day 23 (Fiebig III) and by Western blot after day 26 (Fiebig IV) [81].

The acute phase is followed by the chronic phase, also known as "clinical latency", which can last from ≈ 1 to 20 years, during which the patient is asymptomatic, and often unaware of the infection. During this time, the viral RNA levels are mostly fixed at around 1-100 000 copies/mL blood, referred to as the "set point", but may also be gradually increasing. The CD4+ T cell count is close to normal levels at roughly 1000 cells/µL, but may begin to progressively decline. [47]

When CD4+ T cells have dropped to approximately 300-500 cells/µL, some symptoms of immunodeficiency may start to appear [82], but an HIV-infected patient is not considered to have AIDS until the CD4 count is below 200 cells/µL or if the patient has acquired an AIDS-defining illness, examples of which are the opportunistic infections *Pneumocystis jirovecii* pneumonia, esophageal candidiasis, and brain toxoplasmosis, and malignancies such as Kaposi's sarcoma and Burkitt's lymphoma [83-85]. During the AIDS stage, viremia steadily rises simultaneously with the decline of the CD4 count, since the immune system can no longer control the infection. Without antiretroviral treatment at this point, the mortality rate is above 95 % and the patient will most likely die before long [47].

1.2 HIV-1 GENETIC VARIABILITY

HIV-1 has a high genetic variability due to the error rate and the predisposition to retroviral recombination of the reverse transcriptase enzyme during replication, coupled with the high turnover of virions [86-89]. The single-step point mutation rate of HIV-1 is $\approx 3 \times 10^{-5}$ mutations per base per replication cycle [90], and about 10^{11} viral particles are likely generated daily [47].

1.2.1 Groups

HIV-1 can be further subdivided into groups (**Figure 6**), group M (major), N (non-M, non-O), O (outlier), and the more recently discovered group P [91]. All four groups are thought to be the result of separate cross-species transmission events [1], with group M and N originating from SIV in chimpanzees [3], group P from SIV in gorillas [91, 92], while the origins of group O are unknown, but a chimpanzee or gorilla SIV origin is likely [1, 2]. Only two cases have been found of Group P [91, 92] and 13 of group N [1, 93], all being of Cameroonian origin. Group O represents less than 1 % of all HIV infections and is mainly found in Cameroon, Nigeria, and Gabon [94, 95]. Group M on the other hand represents the vast majority of HIV-1 infections with millions of people infected, and is responsible for the global spread of the infection, existing in almost every country [1, 2].

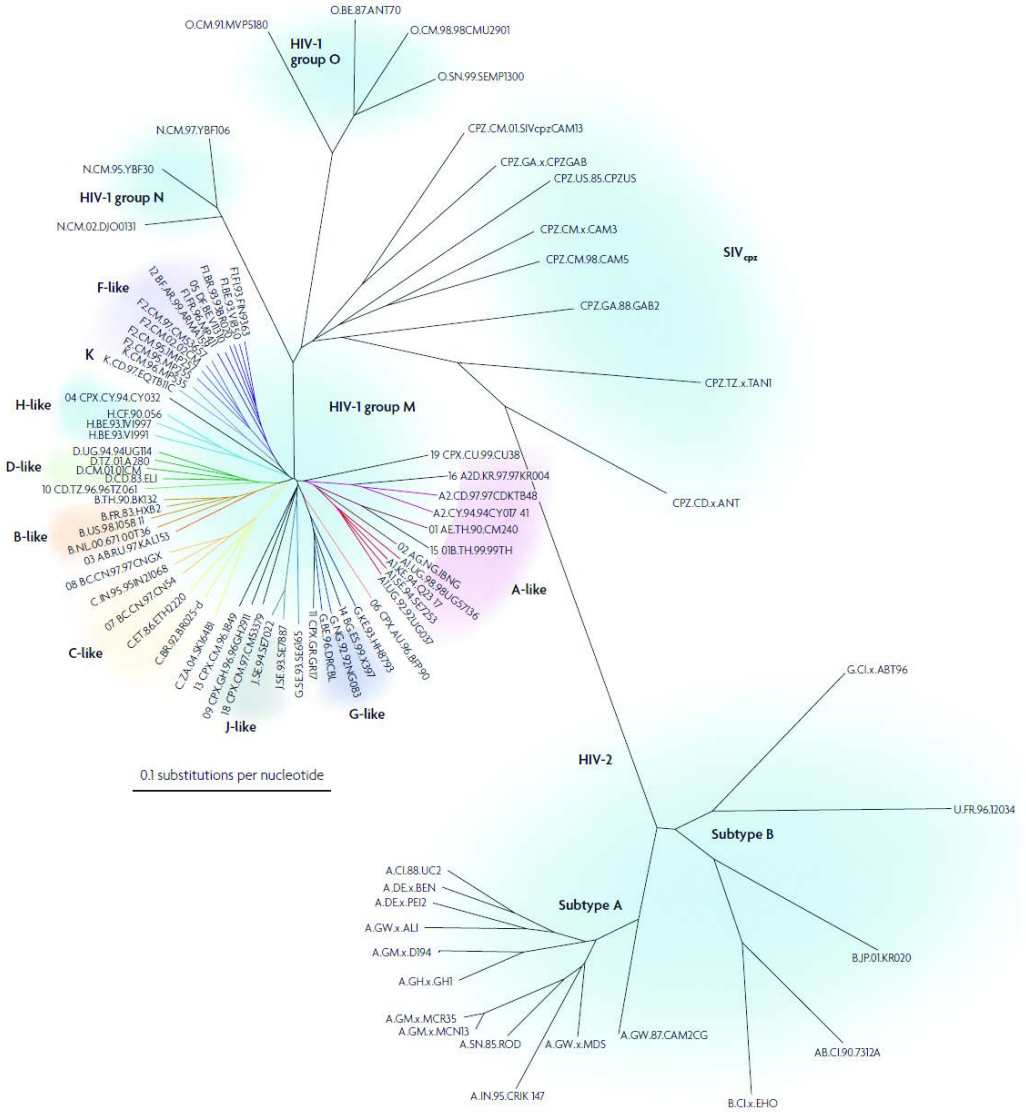

Figure 6. HIV-1 groups and their phylogenetic relationship with SIV and HIV-2. Reproduced with permission from [96].

1.2.2 Subtypes

HIV-1 group M is divided into subtypes based on their genetic similarity that are distributed differently geographically (**Figure 7**). There are nine subtypes designated with the letters A-D, F-H, J and K, as well as numerous recombinant forms. Subtypes E and I were during later analysis found to be recombinant, and were subsequently renamed accordingly to CRF01_AE and CRF04_cpx (or in some cases U for unclassified), respectively [97]. The amino acid level of variation within a subtype is between 8 and 17 %, but can be up to 30 %, while the level of variation between subtypes ranges from 17 to 35 %, but can be as high as 42 %, with the highest diversity observed in the *env* gene [98, 99]. Hence, subtypes B and D could due to their genetic similarity be potentially classified as the same subtype, but are not so owing to historical reasons [2].

Subtype C is the dominating subtype worldwide, responsible for more than 50 % of all HIV infections [96, 100]. Subtype C is responsible for nearly 100 % of all HIV infections in southern African countries such as South Africa, Zimbabwe,

Mozambique, Malawi, Swaziland and Botswana, but is also common in East Africa, as well as in India and Oceania [2]. Subtype B predominates in Australia, Western Europe and North America [101], and is perhaps hence the most studied subtype to date. It originated most likely in central Africa, but spread later through Haiti to MSM (men who have sex with men), intravenous drug user and hemophiliac populations in USA and Europe [102]. South America also has a large proportion of subtype B infections, along with subtype F infections. Subtype A is found among intravenous drug users in Eastern Europe, as well as in East African countries, such as Uganda, Kenya and Tanzania, where it co-circulates with subtypes C and D, together with various recombinant forms based on these three subtypes [2]. Subtype D is largely restricted to the East African countries Uganda, Kenya, South Sudan and Tanzania [103], as is subtype G to West Africa [96].

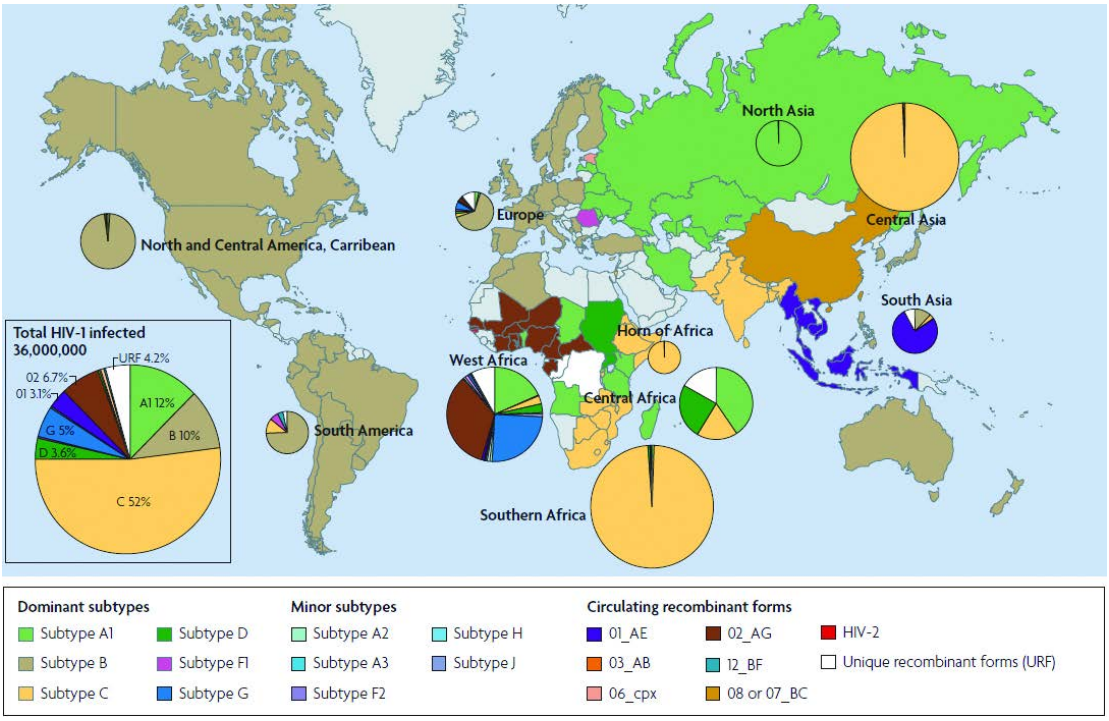

Figure 7. The geographical distribution of HIV-1 group M subtypes. Reproduced with permission from [96].

It has been suggested that the subtypes are not only differently distributed and genetically dissimilar, but also have varying transmission efficiency and differing disease progression [103]. Mother-to-child transmission in utero has been suggested to be preferential in subtype C compared to subtypes A and D [104, 105], perhaps due to increased vaginal shedding [105]. A study conducted in Senegal showed that females infected with subtype A were less prone to progress to AIDS compared to females infected with subtypes C, D and G [106], while a Swedish study on the other hand showed no difference in disease progression between patients infected with subtypes A, B, C and D [107]. Subtype D has nevertheless been associated with a faster disease progression compared to subtypes A, B, and/or C in several studies [108-112]. It has however been debated whether or not some of these observations are confounded by factors such as mode of transmission, host genetic factors, nutritional status and access to medical treatment [103].

1.2.3 Circulating Recombinant Forms

The creation of circulating recombinant forms (CRFs) is the result of co-infection in the same individual with different subtypes, allowing a recombination of their genomes [103]. For a virus to be classified as a CRF, it needs to be found in at least three epidemiologically unlinked individuals. Viruses that are found in fewer individuals are designated as unique recombinant forms (URFs). There are at present 58 existing CRFs and innumerable reports of URFs. CRFs are named according to a number and letter system, where the number stands for the order in which the CRF was discovered and the letters stand for the subtypes that the CRF is a recombination of. For instance, CRF01_AE was the first recombinant form to be discovered (hence the number 01), and is comprised of *gag* and *pol* genes from subtype A and an *env* gene from subtype E (as mentioned above, subtype E is no longer considered a separate subtype and all its sequences were consequently classified as CRF01_AE). If a CRF is a combination of three or more subtypes, the letter code in the name is replaced with "cpx", as is for instance the case with CRF04_cpx (previously subtype I). [2]

The two most common CRFs are CRF01_AE, which is found primarily in south-east Asia, and CRF02_AG, which predominates in West Africa causing up to 50-80 % of the HIV infections in that region. Among intravenous drug users, CRF03_AB is common in Eastern Europe and CRF07_BC and CRF08_BC in China, while CRF06_cpx possibly represents as much as 20-50 % of all HIV infections in the West African countries Togo, Burkina Faso, Niger and Nigeria. [2]

CRF01_AE has been associated with a faster disease progression compared to other CRFs and subtype B in Singapore [113], as well as in Thailand compared to subtype B infections in developed countries [114]. A Thai study on intravenous drug users found a higher probability of transmission of CRF01_AE compared to subtype B, but could not exclude that epidemiological, virological or host factors affected the results [115]. CRF02_AG has on the other hand not been associated with faster disease progression compared with other CRFs and subtypes in western and west-central Africa [116]. However, according to a more recent study conducted in Guinea-Bissau, patients infected with the A3/CRF02_AG (recombinant form between sub-subtype A3 and CRF02_AG) progressed faster to AIDS compared to patients infected with sub-subtype A3, suggesting that disease progression may vary among A-like subtypes/CRFs [117].

1.3 HIV-1 PHENOTYPE CLASSIFICATIONS

HIV-1 has since its discovery been classified into biological phenotypes in different ways. Even though these classifications overlap with each other, they are not synonymous [118], and ought therefore not to be used interchangeably.

1.3.1 Cell Tropism

In 1987, two genetically related but distinct HIV variants in the same patient were isolated from brain tissue and cerebrospinal fluid, respectively. Both viruses could infect peripheral blood mononuclear cells (PBMC), but only the virus from the brain tissue could infect macrophages and monocytes, indicating that genetic variation of HIV within a patient may lead to altered cell tropism of the patient's viruses [119].

Apart from *in vivo* observations, it could also be shown *in vitro* that HIV isolated from PBMCs could be co-cultivated with differentiated macrophages or phytohemagglutinin stimulated CD4+ T lymphoblasts [120]. HIV variants derived from monocytes could infect both macrophages and CD4+ T lymphoblasts, while viruses derived from lymphoblasts showed a sole preference for infecting T lymphoblasts [121]. Based on these observations as well as other similar ones, HIV was classified into M-tropic (macrophage- or monocyte-tropic) and T-tropic (T-cell-tropic) viruses, indicating that HIV viruses could either be dualtropic for macrophages and T cells or have mixed populations of M-tropic and T-tropic viruses coexisting simultaneously within the same individual [122].

1.3.2 Growth Kinetics

Another classification, based on growth kinetic properties, was introduced by a Swedish research group in 1989 [123]. Viruses derived from asymptomatic patients or patients with minor symptoms replicated slowly and inefficiently in PBMC cultures and could often not be passaged at all in CD4+ T-cell lines, while viruses from patients with severe AIDS replicated rapidly and efficiently both in T cell lines and PBMC cultures. Hence, the viruses were classified into slow/low and rapid/high, respectively, and it was observed that a shift from one to the other could occur in a patient over time. The slow/low viruses were exclusively M-tropic, while the rapid/high were both M- and T-tropic, based on the viruses' abilities to infect indicator monocytoid and T-lymphoid cell lines, respectively [123].

1.3.3 Syncytium Inducement

The slow/low and rapid/high viruses also differed from each other by displaying different cytopathogenic effects in cell culture [123]. The rapid/high dual T- and M-tropic viruses caused extensive syncytium formation in MT2 cell lines, which are derived from T cells, while the M-tropic slow/low viruses rarely did so. This lead to the classification of viruses into non-syncytium inducing (NSI) and syncytium inducing (SI), resulting in the MT2 cell culture test being used as a prognostic assay to determine the disease stage of the patient [124].

1.3.4 Coreceptor Use

That HIV-1 used the CD4 receptor for cell entry was found early on after the discovery of the virus [53, 54, 125], but even then it was acknowledged that some other factor was necessary to complete the entry process [126]. In 1996, it was established that HIV-1 uses chemokine receptors as coreceptors in addition to CD4 to enter cells (**Figure 8**), and these receptors could be either CCR5 [60-62] or CXCR4 [63, 64]. A new phenotype classification was consequently introduced in 1998 [118], where viruses using CCR5 were proposed to be called R5, while viruses using CXCR4 were to be called X4. Dualtropic viruses [127] were denoted R5X4, and later as new HIV-1 coreceptors were added to the list [127, 128], they received similar abbreviations, such as R3 for viruses using the C-C chemokine receptor type 3 (CCR3).
The new classification based on coreceptor use could more precisely describe the biological properties of a viral isolate [118]. R5 viruses would not be syncytium inducing (NSI), since MT2 cell lines lack CCR5 receptors, but they would be both M-

and T-tropic, since both macrophages and T-cells express CCR5, and most likely slow/low. X4 viruses on the other hand would be syncytium inducing (SI) due to the expression of CXCR4 on MT2 cell lines, T-tropic, and rapid/high. Dualtropic R5X4 viruses would combine properties from both R5 and X4, and be SI, M-tropic and T-tropic, as well as likely rapid/high.

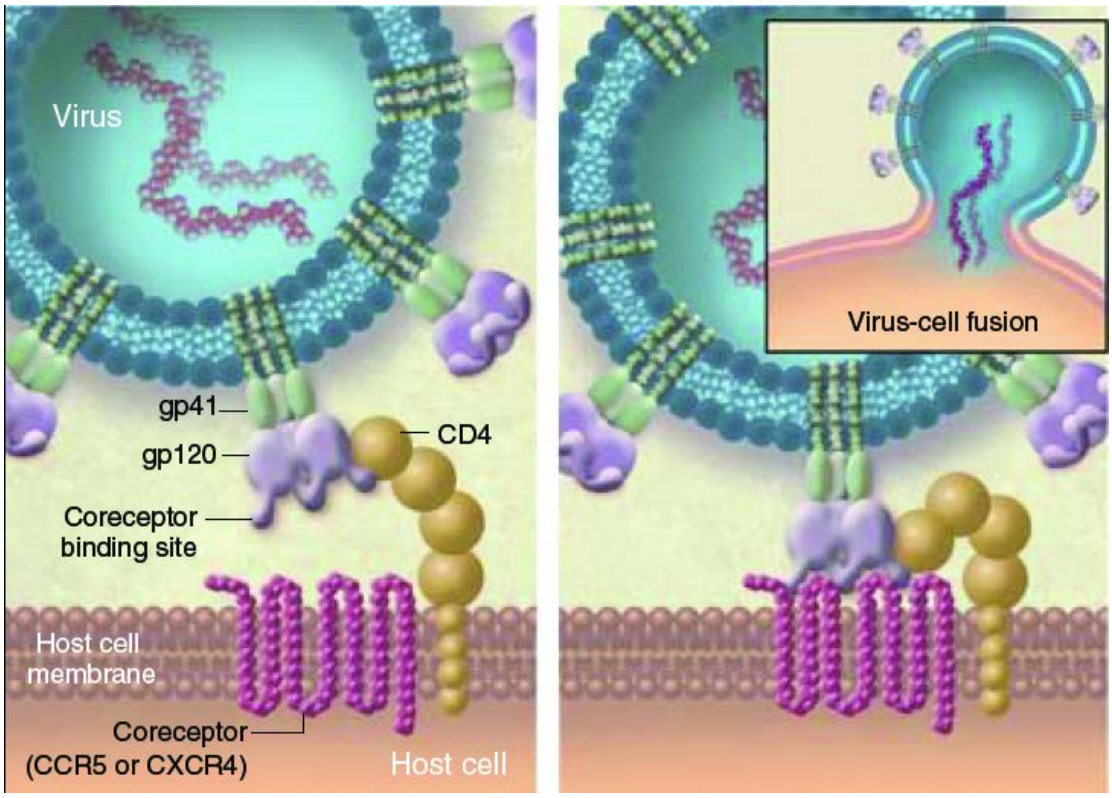

Figure 8. HIV-1 binding to the host cell, involving the primary receptor CD4 and the coreceptors CCR5 or CXCR4. Adapted with permission from [129]. Illustrator: Taina Litwak.

1.4 HIV-1 ENTRY CORECEPTORS

The main coreceptors of HIV-1 are considered to be CCR5 and CXCR4, but the virus may also utilize other receptors such as CCR1, CCR2b, CCR3, CCR4, CCR8, V28, BoB (GPR15), Bonzo (STRL33) and Apj [128, 130-133].

1.4.1 CCR5

The natural ligands of chemokine receptors are chemotactic cytokines (chemokines), which facilitate the chemotaxis of phagocytic cells and T cells to areas of inflammation. The chemokines that naturally bind the CCR5 receptor are the β-chemokines RANTES (Regulated on Activation Normal T cell Expressed and Secreted), MIP1α (Macrophage inflammatory protein-1α) and MIP1β [134, 135]. RANTES is involved in chemoattracting monocytes, memory T cells and eosinophils as well as in the inducement of histamine release by basophils [134]. MIP1α and MIP1β both mediate macrophage migration, but when it comes to other leukocytes, MIP1α primarily mediates migration of B lymphocytes, activated CD8+ T cells, NK cells and eosinophils, while MIP1β preferentially mediates migration of CD4+ T cells [136, 137]. All three chemokines are capable of blocking R5 virus infection *in vitro* [138], but *in vivo* they can limit replication, but not hinder disease progression [139].

CCR5 is substantially expressed on activated memory CD4+ T cells and GALT, as well as on macrophages, monocytes, microglia (resident macrophage type in the central nervous system), NK cells, dendritic cells and Langerhans cells (dendritic cells in mucosal epithelia), which also express lower levels of CD4, making these cells target cells for R5 viruses [140-147], which require both CD4 and CCR5 expression for cell entry.

The CCR5 receptor is a member of the G protein-coupled receptor superfamily. It is a 7-transmembrane protein, composed of four intracellular and three extracellular loops joined by hydrophobic membrane spanning domains, as well as an extracellular N-terminal and a cytoplasmic C-terminal domain [148]. A single potential N-linked glycosylation site is present in the third extracellular loop (ECL3) [134], but it is not posttranslationally modified [149], likely due to its proximity to the disulfide bridge formed between the N-terminal and ECL3. Unlike most other G protein-coupled receptors, no potential glycosylation sites are present in the N-terminal [134], which is important for ligand binding [148] along with ECL2 [150].

Viruses with the R5 phenotype are generally the first to be present at the start of the infection [131, 151-154], and have also been associated with transmission [30, 153-157]. These variants are mainly T-tropic [154, 155, 157] and require high amounts of CD4 on the host cell's surface [156, 157], as opposed to R5 viruses that develop later during the infection that have adapted to entering host cells with fewer CD4 molecules on the surface, such as macrophages [153, 158], which play a role in HIV-1's infection of the central nervous system [158, 159].

A deletion of 32 base pairs in the gene coding for the CCR5 receptor, designated as *CCR5*Δ32, causes a premature stop codon in ECL2 of CCR5, leading to the retention of the mutant CCR5 protein in the endoplasmic reticulum [160-162]. The allelic frequency of this mutation is about 10 % in Caucasians [161, 162], and homozygous individuals are strongly resistant to HIV infection [160, 161, 163], supporting the importance of R5 viruses in transmission [30]. However, homozygosity is not an absolute protection against infection [164-166], since a few homozygous individuals were inferentially alleged to have been infected with X4 viruses [164-166]. Heterozygous individuals are also protected to some extent from infection [161, 162] and disease progression [161, 163]. Even though homozygosity for *CCR5*Δ32 does not seem to result in any apparent detectable phenotypic consequences (apart from HIV infection protection) in the affected individual [160-162], it has been found that homozygotes are at an increased risk of symptomatic West Nile virus infection with fatal outcome [167]. On the other hand, homozygous individuals are to some extent resistant to acquiring non-HIV related non-Hodgkin lymphoma [168].

1.4.2 CXCR4

The natural ligand of the CXCR4 receptor is SDF-1 (Stromal cell-derived factor 1) [169, 170], initially named owing to it being a bone marrow stromal cell derived factor that enhances B cell proliferation [171]. It also acts as a chemoattractant to T lymphocytes and monocytes, but not neutrophils [172]. SDF-1 is capable of blocking

the infection of X4 viruses *in vitro* [169, 170], but seems not to have an inhibitory role *in vivo* in patients with advanced disease, suggesting that the virus adapts to escape antiviral chemokines [151].

The target cells of X4 viruses are primarily naïve CD4+ T cells [140, 141, 173, 174], but CXCR4 is also expressed on other leukocytes, including memory CD4+ T cells, GALT, Langerhans cells, and NK cells, which represent additional targets [140, 141, 145, 147, 174, 175]. In general, CXCR4 is expressed at various levels in nearly all lymphocyte subsets, as compared to the more restricted CCR5 [174]. Macrophages and monocytes also express CXCR4 [176, 177], but it remains hitherto unclear whether or not X4 viruses are capable of productively infecting macrophages [178].

The CXCR4 receptor is also a member of the G protein-coupled receptor superfamily, and is hence in many ways structurally similar to CCR5 [179]. Interestingly, the five acidic amino acid residues Asp-97$^{2.63}$, Asp-171$^{4.60}$, Asp-187 (ECL2), Asp-193$^{5.32}$, and Asp-262$^{6.58}$ in CXCR4 that are found to be important for ligand binding [180, 181], as well as infection with HIV [181], are replaced by the uncharged residues Tyr-89$^{2.63}$, Gly-163$^{4.60}$, Ser-179, Gln-188$^{5.32}$, and Asn-258$^{6.58}$ in CCR5 [179]. Also, the N-terminal of CXCR4 has nine acidic residues as opposed to only three in the N-terminal of CCR5 [179]. Moreover, CXCR4 contains two potential N-linked glycosylation sites, one in the N-terminal and one in ECL2, respectively, that are absent in CCR5 [182].

X4 viruses are believed to evolve from R5 viruses within a patient [126, 153], and generally start to appear with disease progression [126, 131, 151-153], accompanied by a rapid CD4 count decline [126, 131, 152]. If X4 viruses are the cause or the effect of disease progression remains unclear [126, 183]. However, X4 viruses are easier to neutralize compared to R5 viruses [183, 184], which could partially explain their appearance when the immune system has sufficiently deteriorated [153, 183].

1.4.3 Dual Tropism

Dualtropic viruses that utilize both CCR5 and CXCR4 represent an intermediate intra-patient evolutionary state that arises when R5 viruses acquire enough mutations in gp120 to allow the broadening and/or switch of coreceptor use [126, 185]. The properties of dualtropic viruses vary, depending on whether or not they use CCR5 or CXCR4 more efficiently [186, 187], sometimes denoted as Dual-R and Dual-X, respectively [186]. Dual-R viruses have a closer resemblance to R5 viruses genetically, while the same is true for Dual-X and X4 viruses [186, 187]. The terminology dual/mixed (DM) often appears with regard to dualtropic viruses, indicating that it cannot be distinguished whether the bulk of viral isolates contains truly dualtropic variants or whether a mixture of R5 and X4 and possibly also dualtropic viruses is present [183].

1.4.4 Other Coreceptors

As mentioned above, HIV-1 can utilize multiple other receptors as coreceptors *in vitro* apart from CCR5 and CXCR4, but their role *in vivo* is likely marginal [41, 132, 183], with the possible exception of CCR3, which may along with CCR5 play an important role in the infection of microglia in the central nervous system [128, 144]. CXCR4-

using viruses in advanced AIDS patients sometimes evolve the ability to use additional coreceptors such as V28, CCR8, and Apj [132], but HIV-1 does not normally utilize alternative coreceptors *in vivo* without also being able to use either CCR5 or CXCR4 [127, 132, 188].

1.4.5 Glycoprotein 120's Interaction with Entry Coreceptors

The gp120 V3 region was found early on to be important in HIV-1 tropism and syncytium-inducing capacity [189-194]. In most isolates, it is composed of 35 amino acids forming a loop structure with the help of a disulfide bridge created by the cysteine residues found at positions 1 and 35. The V3 loop tip consists of the GPG crown, which is fairly conserved and forms a β-turn, whereas the flanking regions of the crown are variable and together form the two strands of an antiparallel β-sheet [150, 195-197]. The crown plays an important role in the interaction of HIV-1 with the ECL2 of CCR5 [57, 179, 198] and ECL2 and ECL3 of CXCR4 [180, 199], while the N-terminal region of CCR5 interacts with the V3 base [41, 179, 198, 200] and the adjoining bridging sheet [57, 200].

The evolution from R5 viruses to X4 viruses involves various amino acid substitutions [190-193, 201-206], in particular those that confer a change in the net charge, where a higher net charge has been associated with CXCR4 use [190, 192, 193, 202, 203, 205, 206]. Hence, X4 viruses are also more heterogeneous in their V3 loop compared to R5 viruses [194, 205-207]. Furthermore, basic amino acid residues in positions 11, 24, 25 and 28 of the V3 loop have been coupled with CXCR4 use [190, 192, 201, 202, 205, 206]. The basic nature of the protruding V3 loop of X4 viruses has good electrostatic compatibility with the acidic residues in the binding pocket of CXCR4 [179-181], additionally supporting the importance of high V3 charge for CXCR4 use.

A potential N-linked glycosylation site, N301, is located in the V3 positions 6-8. The loss of this site has been associated with a coreceptor switch from CCR5 to CXCR4 [208-210], as well as with an increased sensitivity to neutralization [210, 211]. Moreover, when this site was removed using site-directed mutagenesis in a dualtropic isolate, it lost its CCR5-dependent fusion activity entirely, whereas 50 % of the wild-type CXCR4-dependent fusion activity was retained, emphasizing the importance of this glycosylation site for CCR5 use [212].

Other regions of gp120 can also affect coreceptor use [199, 204, 208, 212-216], in particular mutations and deletions in V1/V2 [199, 208, 214-216], while mutations in the fusion peptide gp41 may have an impact on coreceptor use too [217].

1.4.6 Biological Coreceptor Use Determination

In order to determine coreceptor use biologically *ex vivo*, patient-derived HIV-1 isolates are tested on specific cell lines such as U87 and GHOST, which apart from expressing CD4 are also transfected with CCR5, CXCR4, or other coreceptors of interest [62, 183, 218]. Another way to determine coreceptor use *ex vivo* is by infecting cell lines expressing both CCR5 and CXCR4 with patient-derived HIV-1 isolates while alternately adding coreceptor antagonists such as the CCR5-inhibiting TAK799 and the CXCR4-inhibiting AMD3100 [219, 220]. In more recent years, quite a few

recombinant phenotypic assays for determining coreceptor use have been developed, including the Trofile assay (Monogram Biosciences) [221]. Partial or full-length viral *env* is amplified from patient plasma and used to create pseudoviruses or infectious recombinant viruses. These are then tested for entry on indicator cell lines expressing CD4 along with either CCR5 or CXCR4. The advantage of these assays is that it is possible to distinguish between pure R5, pure X4 and dualtropic/mixed populations [183]. Also, the enhanced sensitivity Trofile assay can detect minor X4 populations that constitute as little as 0.3 % of the total quasispecies within a patient with 100 % sensitivity [222, 223].

1.4.7 Coreceptor Use and HIV-1 Subtypes

CXCR4-using variants appear in about half of the AIDS patients infected with subtype B [224, 225], but the prevalence in non-B HIV-1 subtypes varies. The subtype most closely related to subtype B, HIV-1 subtype D, is also associated with increased CXCR4 use [112, 186, 226], while several reports have suggested the same for CRF01_AE [227-229]. The acquisition of CXCR4 use seems on the other hand less likely in subtype C even at late clinical stages [207, 230-232], perhaps owing to the requirement of more mutations for a coreceptor switch [233]. Then again, later studies indicate that CXCR4 use in subtype C may not be as low as earlier reported [234-237], which could be attributed to an evolving subtype C epidemic [183, 234, 235]. In particular, the increase in CXCR4 use among treatment failure patients infected with subtype C [234, 236, 237], is intriguing.

1.5 ANTIRETROVIRAL TREATMENT

Antiretroviral drugs target different steps of HIV-1's replication cycle (**Figure 9**), and are consequently divided into drug classes based on the step they target.

1.5.1 Drug Classes

The drug classes that make up the core components of HAART, the triple drug combination given to HIV-1 infected patients in order to avoid the emergence of drug resistance, are nucleoside reverse transcriptase inhibitors (NRTI), non-nucleoside reverse transcriptase inhibitors (NNRTI) and protease inhibitors (PI). NRTIs target the reverse transcription step by imitating natural dNTPs (deoxynucleoside triphosphates) leading to their incorporation into viral DNA and thereby prematurely terminating the chain, while NNRTIs target the same step by allosterically inhibiting the reverse transcriptase enzyme. PIs, as implied by their name, block the maturation of HIV-1 virions by acting as competitive inhibitors to the protease enzyme. [238, 239]

First-line treatment usually consists of a combination of two NRTIs and one NNRTI or one PI, while other drug classes are generally introduced in second- and third-line treatments. If patient compliance desists, single mutations are enough to confer resistance to NRTIs and NNRTIs [240, 241], while PIs are more forgiving and require more mutations for resistance [242].

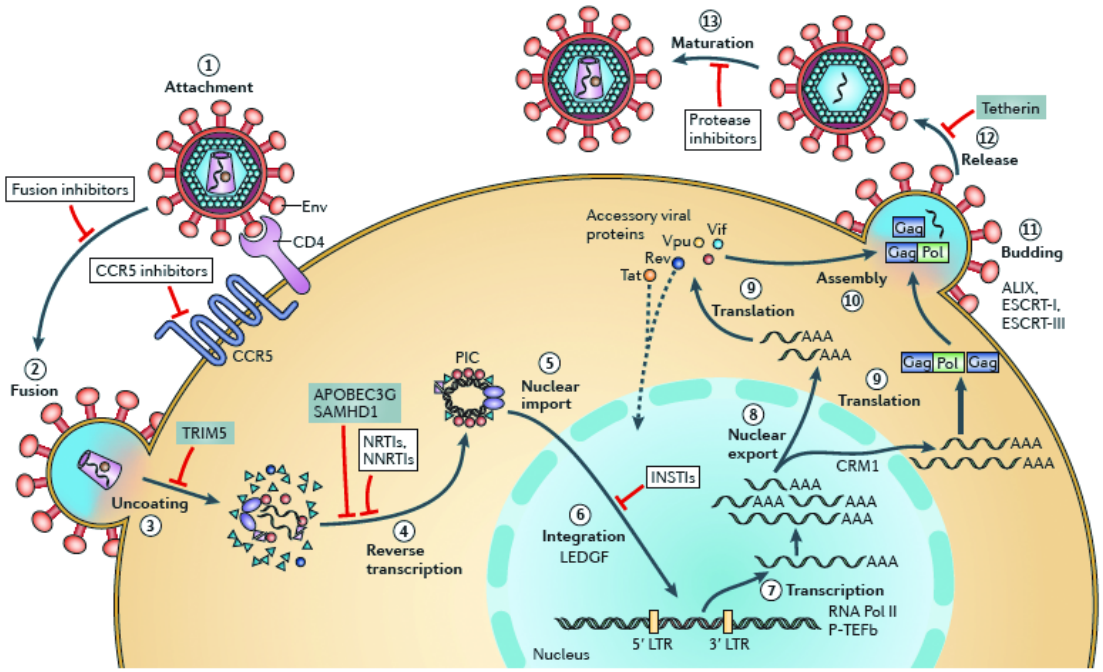

Figure 9. Antiviral drug classes and HIV-1's replication cycle. The antiviral drug classes are boxed in white while host proteins interfering with the viral life cycle are boxed in grey. Reproduced with permission from [238].

Newer drug classes target other steps of the viral replication cycle, such as the integration and the entry step. IN strand transfer inhibitors (INSTIs) block the viral DNA transfer activity of the integrase enzyme, thereby preventing integration of the viral DNA into the host's genome. Entry inhibitors can be divided into two subclasses: fusion inhibitors, which disrupt the six-helix bundle that brings together the viral and host membranes and in so doing prevents fusion, and coreceptor inhibitors. [238, 239]

1.5.2 Coreceptor Inhibitors

Coreceptor inhibitors are meant to block CCR5 or CXCR4, thereby precluding the binding of HIV-1. No CXCR4 inhibitor has been licensed for clinical use to date [243]. The synthetic small molecule CXCR4 inhibitors AMD3100 and the orally bioavailable AMD11070 were quite well tolerated, but did not show an antiviral effect in the form of a significant decrease in viral loads in clinical trials, although a reduction of the amount of X4 viruses in the patients was seen [244-246], indicating the potential usefulness of giving CXCR4 inhibitors in conjunction with CCR5 inhibitors [243, 246]. Further clinical development of AMD11070 has however been halted due to potential hepatotoxic effects observed in long-term animal studies [246].

Substantially more CCR5 inhibitors have been in development compared to CXCR4 inhibitors, but only one, maraviroc (**Figure 10**), has been licensed by FDA for clinical use [243, 247]. The existence of individuals lacking the CCR5 receptor due to the *CCR5*Δ32 deletion, but without any significant health impacts, suggested the safety of blocking CCR5 for clinical use. At first, modified natural ligands were tested, but owing to their poor oral bioavailability, the research focus shifted to generating small-molecule antagonists instead, which are given the suffix "viroc" in their names meaning "viral receptor occupancy" [243]. These antagonists allosterically alter

CCR5's conformation, resulting in the inability of HIV-1 to bind to the coreceptor [179, 243]. Some examples of developed CCR5 inhibitors apart from maraviroc are vicriviroc, which did not show significant efficacy in phase III clinical trials [248], aplaviroc, which was shown to be hepatotoxic in phase III clinical trials [249], and cenicriviroc, which is currently being tested in phase IIb clinical trials [243, 250].

1.5.3 Maraviroc

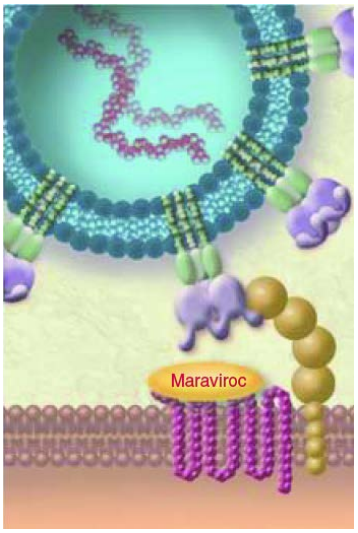

Figure 10. Schematic overview of how maraviroc blocks the entry process of HIV-1. Adapted with permission from [129]. Illustrator: Taina Litwak.

Maraviroc (Selzentry®) was approved by FDA for use in treatment-experienced adult HIV-1 infected patients in combination with other antiretroviral drugs in 2007 [247]. The license was extended to include treatment-naive patients in 2009 [243]. Maraviroc is administered on a twice daily dosing schedule and is well tolerated [129, 243].

Despite its approval for first-line treatment, maraviroc is mainly given as a second-line drug, due to several limiting factors. The twice daily dosing regimen prevents the co-formulation of maraviroc with other first-line drugs such as Truvada® and Epzicom®, which combine several drugs in one tablet administered only once per day. Reducing the amount of drugs given to patients daily has a significant impact on compliance, resulting in endeavours to limit first-line treatment to one-tablet-a-day dosing schedules. [243]

Another limiting factor is the requirement of an HIV-1 coreceptor use test to be performed before administering the drug, since the prior existence of CXCR4-using variants in the patient may lead to limited efficacy of the drug as well as drug resistance through the selection for these variants [251, 252]. In the U.S., biological determination of coreceptor use is recommended to be performed with the expensive and fairly time-consuming enhanced sensitivity Trofile® assay, which requires a minimum plasma HIV-1 RNA level of $\geq$ 1,000 copies/ml [221, 243, 253]. Genotypic assays, which attempt to predict coreceptor use from HIV-1 sequences, are however the preferred coreceptor use determining method in Europe, in particular those utilizing the Geno2Pheno algorithm with a false-positive rate of 10 % [254], which when tested retrospectively on material from clinical trials correlated well with the clinical outcome of maraviroc treatment [255].

Apart from drug resistance acquired through the development and/or selection for pre-existing CXCR4-using viruses [251, 252], resistance to maraviroc can also arise through the emergence of HIV-1 R5 variants associated with particular mutations in V3 that can utilize the new drug-bound CCR5 conformation [243, 256]. Interestingly, such variants do however retain full susceptibility to other CCR5-inhibitors [256].

1.5.4 Antiretroviral Treatment Program in Botswana

As mentioned earlier, the national antiretroviral treatment program in Botswana was introduced in 2002 [33]. At the time, the standard first-line treatment consisted of two NRTIs, zidovudine (AZT) and lamivudine (3TC), together with one NNRTI, either nevirapine (NVP) or efavirenz (EFV), while second-line treatment usually consisted of two other NRTIs, didanosine (ddI) and stavudine (d4t), and one PI, nelfinavir (NFV) [257]. The latest guidelines from 2012, however, recommend the following regimens: two NRTIs, tenofovir disoproxil (TDF) and emtricitabine (FTC), together with one NNRTI, EFV, often given as one tablet (Atripla®), as first-line treatment (alternatively, the Truvada® tablet can be used, which contains the same drugs, but replaces FTC with 3TC), and two NRTIs, AZT and 3TC, together with one PI combined with a booster, lopinavir/ritonavir (LPV/r), as second-line treatment [258]. Drug resistance testing is routinely performed only after the patient has failed second-line treatment [258].

1.6 CORECEPTOR USE PREDICTION MODELS

Coreceptor use prediction models, as the name implies, use patient-derived HIV-1 gp120 V3 sequences to predict the coreceptor use of the HIV-1 variants harboured within a patient. Most models do not distinguish between dualtropic and X4 variants, categorizing them together as CXCR4 users instead.

1.6.1 Simple Rule Algorithms

Algorithms based on simple rules appeared even before the discovery of HIV-1's coreceptors, and were used to distinguish between slow/low, M-tropic NSI and rapid/high, T-tropic SI variants, as they were classified at the time, and have since then been inferentially translated to coreceptor use phenotypes. The most prominent of these is the 11/25 rule, which classifies sequences as CXCR4-using if they contain a positively charged amino acid (arginine or lysine) in position 11 of the V3 loop and a neutral amino acid in position 25, or just a positively charged amino acid in position 25 [190]. Another simple rule algorithm derived from the observations of high V3 charge in CXCR4 use [190, 192, 193, 202, 203, 205, 206], classifies sequences with a total V3 charge of ≥ 5 as CXCR4-using.

1.6.2 Learning Algorithms

As the field of bioinformatics evolved, the use of learning algorithms became common for predicting coreceptor use. Learning algorithms are machine-learning methods that have been fed sequences with known coreceptor use, based on which scores are created. These scores represent how similar the tested sequences are to CXCR4-using or non-CXCR4-using sequences that were used to create the algorithm. The scores can be visualized on a scale where different false positive rate cut-off points may be chosen to denote what scores will be used to classify sequences as CXCR4-using. Examples of learning algorithms are SVM (support vector machine) [259], PSSM (position specific scoring matrix) [260], and Geno2Pheno [coreceptor] [261], which is available with different false positive rate cut-offs, where the 10 % is recommended by the European Consensus Group on clinical management of HIV-1 tropism testing [254].

1.6.3 Complex Rule Algorithms

Complex rule algorithms combine a set of rules into a decision tree. Several have become available recently that have been developed for specific subtypes. One of them is CoRSeqV3-C (**Figure 11**) [262], which is meant to be tested on subtype C sequences. Other examples are algorithms developed for subtype D [263] and CRF01_AE [264] by Raymond *et al.* that are based on 11/25 and net charge rules, as well as the mutations of the V3 N-linked glycosylation site.

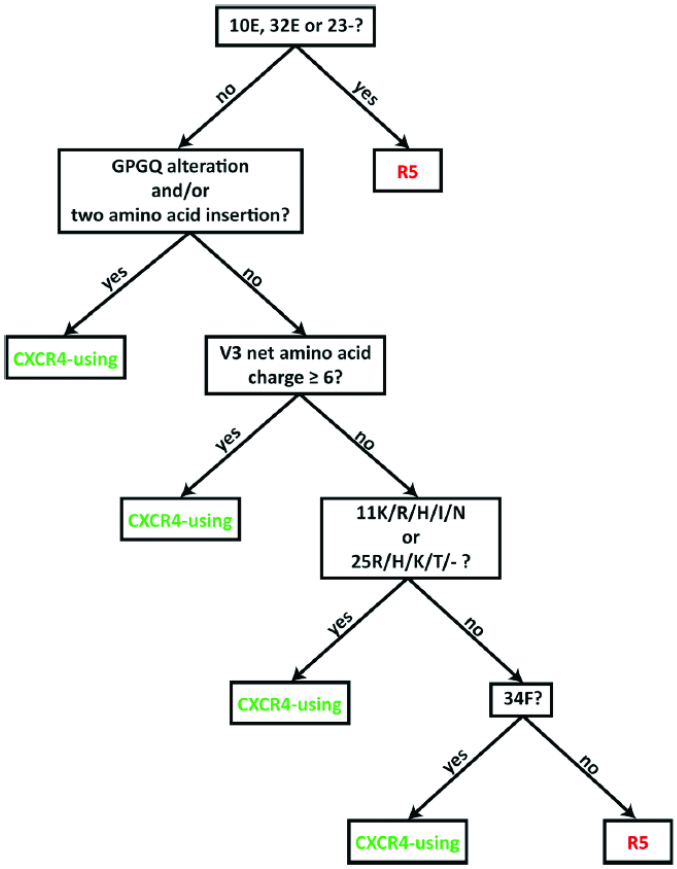

Figure 11. The decision tree of CoRSeqV3-C showing how CXCR4-using sequences are classified based on specific V3 amino acids and total charge. Reproduced with permission from [262].

2 AIMS

The general aim of this thesis was to investigate amino acid predictors of HIV-1 coreceptor use in different subtypes and apply the obtained knowledge to antiretroviral treatment.

The more specific aims of each paper were:

Paper I: To investigate in HIV-1 groups and subtypes if the occurrence of potential N-linked glycosylation sites in and around the gp120 V3 loop may be correlated with specific coreceptor use, in particular if N301 is associated with CCR5 use, using all available sequences from the Los Alamos HIV Sequence database and allowing only one sequence per patient.

Paper II: To characterize charged V3 amino acids in R5 and X4 sequences belonging to the major HIV-1 subtypes and describe the coreceptor switch from R5 to X4 from this perspective, using all available sequences from the Los Alamos HIV Sequence database and allowing only one sequence per patient.

Paper III: To investigate if there is an increase in CXCR4-using viruses in HIV-1 subtype C patients failing antiretroviral treatment in Botswana compared to treatment-naïve individuals using population sequencing, as well as to further assess the coreceptor use of minor HIV-1 quasispecies in treatment-experienced patients using single genome sequencing.

Paper IV: To compare the performance of current available coreceptor use prediction algorithms in major HIV-1 subtypes from a CCR5 inhibitor treatment perspective, as well as to determine if the "analytical" glycan-charge model, based on observations from **papers I** and **II**, can complement existing prediction algorithms, using all available sequences from the Los Alamos Sequence database and allowing one sequence per patient per phenotype.

3 MATERIALS AND METHODS

3.1 MATERIALS

3.1.1 Sequences from the Los Alamos HIV Sequence Database (Papers I, II and IV)

In **paper I**, the dataset consisted of 176 sequences retrieved from the Los Alamos HIV Sequence database. The sequences were from unique patients and spanned the gp120 V3 loop and flanking regions. The coreceptor use of each sequence was determined biologically and among the 176 sequences, 133 used CCR5, 29 used CXCR4 and 14 were dualtropic. Most of the sequences belonged to HIV-1 group M (n = 171), but group O (n = 4) and N (n = 1) sequences were also included. The group M sequences included subtypes A (n = 22), B (n = 62), C (n = 28), D (n = 12), G (n = 5) and U (n = 1), the circulating recombinant forms CRF01_AE (n = 20) and CRF02_AG (n = 11), as well as other recombinants, namely AC (n = 4), AD (n = 1), A1DGJ (n = 1), AU (n = 1), BC (n = 1) and BF1 (n = 2).

In **paper II**, the primary dataset consisted of 528 sequences retrieved from the Los Alamos HIV Sequence database, out of which 482 were R5 and 46 X4 sequences. The sequences were from unique patients, had biologically determined coreceptor use, spanned the gp120 V3 loop, and belonged to HIV-1 group M subtypes A (n = 48), B (n = 231), C (n = 180), D (n = 37) and the circulating recombinant form CRF01_AE (n = 32). A secondary dataset, which was used only for within phenotype analyses, consisted of an additional 12 R5 and 9 X4 sequences, resulting in a total of 494 R5 and 55 X4 sequences. Thus, in the secondary dataset, one patient could have a maximum of two sequences included: 1 R5 and 1 X4.

In **paper IV**, the dataset consisted of 1,273 sequences retrieved from the Los Alamos HIV Sequence database. One sequence per individual per phenotype was included. The sequences spanned the gp120 V3 loop and had biologically determined coreceptor use. The dataset contained 1,046 R5, 141 R5X4, and 86 X4 sequences belonging to HIV-1 group M subtypes A (n = 93), B (n = 607), C (n = 352), D (n = 84) and the circulating recombinant form CRF01_AE (n = 137).

3.1.2 Patient Samples from Botswana (Paper III)

In **paper III**, blood samples were collected from all adult patients experiencing second-line treatment failure in Botswana's national antiretroviral treatment (ART) program over a two-month period (May to June) in 2006. In all, twenty-four patients were identified, and both sexes were represented in the material. The patients were infected with HIV-1 subtype C and were on treatment for a minimum of one year (the exact duration of treatment was unavailable). Virological failure was determined as two consecutive viral load measures above 1,000 RNA copies/mL, which enabled drug resistance genotyping. Drug resistance profiles were available for twelve of the patients.

Twenty-six treatment-naïve patients infected with HIV-1 subtype C were selected as controls from a previously described cohort with samples collected in 2003 [265], rendering it likely that the treatment failure patients were started on treatment at a time point similar to the collection date of the treatment-naïve control samples, since the national ART program in Botswana was initiated only a year earlier [33, 257]. Furthermore, the probability of finding treatment-naïve patients with low CD4 counts would be greater in a cohort from an earlier time point, since the national guidelines for treatment commencement were different at the time. The inclusion criteria for the controls were an available sequence spanning the envelope gp120 V3 loop and a CD4 count of ≤ 250 cells/μL, which would have made these patients eligible for treatment in 2006, and hence increased the likelihood of the CD4 counts being similar in the patient and control groups.

3.2 METHODS

3.2.1 Selection of Sequences from the HIV Sequence Database

(Papers I, II and IV)

In **paper I**, the selection tool on the Los Alamos HIV Sequence database website (http://www.hiv.lanl.gov/) was used to collect and sort the material. Initially, 1,015 sequences spanning the gp120 V3 loop, comprising all HIV-1 groups, group M subtypes, sub-subtypes and circulating recombinant forms with a reported coreceptor use that were available in the database as of October 2004, were retrieved. Subsequently, since there were multiple sequences in the dataset belonging to the same individuals, one sequence per patient was selected randomly using lottery, without regard to the coreceptor use phenotypes present. Sequences lacking a patient identity code or listed with coreceptor use phenotypes other than R5, X4 or dualtropic R5X4, were discarded.

The selection process in **paper II** was similar to that of **paper I**, but differed in some aspects. Initially, all V3 sequences with reported exclusive CCR5 or CXCR4 use in HIV-1 group M subtypes A, B, C, D, and CRF01_AE as of June 16th 2009 were retrieved from the database, resulting in a total of 3,307 sequences. One sequence per patient was chosen using random number generators diminishing the number of sequences to 724. The original articles were then perused to confirm that the reported coreceptor use of the sequences was determined biologically, decreasing the amount of sequences further to 482. During this scrutiny, additional sequences were added from the original articles that were either not reported to the database at all or were not listed in the database's initial selection, resulting in a total of 528 sequences, which constituted the primary dataset. The secondary dataset allowed the use of one sequence per patient per phenotype. Consequently, the material was expanded to include additional sequences belonging to patients in the primary dataset that had sequences of both phenotypes. The additional sequences in the secondary dataset had sequences belonging to the other phenotype of these patients that was not selected in the primary dataset, resulting in a total of 549 sequences.

The dataset selection process in **paper IV** was identical to that of **paper II**, with the exception of also including exclusive dualtropic R5X4 sequences, and allowing one

sequence per patient per phenotype in the whole material. Initially, 7,878 sequences were retrieved, representing all V3 sequences with reported exclusive R5, X4 or dualtropic phenotype in HIV-1 group M subtypes A, B, C, D, and CRF01_AE as of October 9th 2013. After the completion of the rigorous selection process (**Figure 12**), 1,273 sequences remained, derived from 123 original articles.

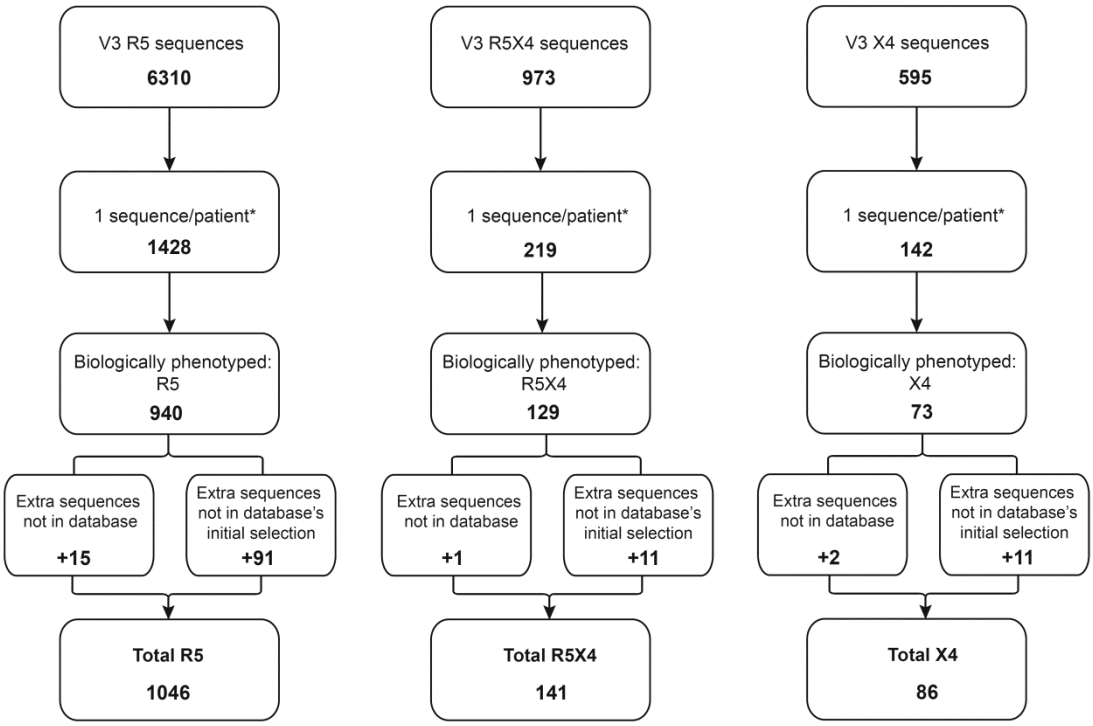

Figure 12. The sequence selection process in paper IV. *If available, 1 R5 and 1 R5X4 or X4 was allowed per patient. Reproduced from paper IV.

3.2.2 **Population Sequencing** (Paper III)

In **paper III**, population sequencing was performed on the samples obtained from the 24 patients failing treatment in order to obtain sequences for evaluating coreceptor use. DNA was extracted from 200 µL of buffy coat using the QIAamp Blood kit (Qiagen, Chatsworth, CA), eluted in 200 µL and stored at – 20ºC until ready for use. Sequences corresponding to the HIV-1 gp120 V3 region as well as flanking C2-C3 regions were amplified using two rounds of nested PCR consisting of 40 cycles each [266]. The primers used in the first round were JA167 and JA170, while JA168 and JA169 were used in the second round [267].

The PCRs were run in tubes and the results were visualized using gel electrophoresis, followed by the purification of the positive PCR products using the QIA quick PCR purification kit (QIAGEN, GmbH, Hilden, Germany). Ten ng/µL (Nanodrop) of the purified PCR product was used in a 20 µL Big Dye terminator 3.1 (Applied Biosystems, Foster City, CA) sequencing reaction, followed by the transfer of the tubes to Eurofins MWG Operon, Ebersberg, Germany for sequencing.

The sequences were then edited and assembled into overlapping fragments using the software program Sequencher (Gene Codes Corporation, Ann Arbor, MI). Each sequence was analyzed for ambiguities at the nucleic acid level for the inclusion into

phylogenetic trees. Additionally, majority rule was applied to the V3 region at the amino acid level for genotypic assay coreceptor use analysis.

Population sequences from the control patients were retrieved from the Los Alamos HIV Sequence database with the help of the accession numbers listed in Ndung'u *et al*. [265].

3.2.3 Single Genome Sequencing (Paper III)

In addition to population sequencing, single genome sequencing was also performed on the samples from the 24 patients failing treatment **in paper III**, in order to further assess the coreceptor use of minor HIV-1 quasispecies. The procedure was mostly the same as for population sequencing, but with several additional steps and alterations.

A limiting dilution PCR was performed prior to the regular nested PCR in order to determine the dilution that would most likely generate PCR products from single quasispecies. The limiting dilution PCR was run in four-fold dilutions of the sample DNA with four replicas of each dilution, starting at 1:8, in MicroAmp Optical 96-well reaction plates (AB Applied Biosystems, Singapore). Based on the Poisson theory of random distribution, a yield of approximately a third positive PCR reactions likely represents single proviruses [268-270]. Hence, a dilution was estimated from the positive PCR products in the limiting dilution PCR that would likely generate ≈ 25 % positive reactions. This dilution was then used to run regular PCR, also in MicroAmp Optical 96-well reaction plates, in the hope of amplifying single viral genomes.

Finally, a sequence would be regarded as a single V3 genome only if ambiguities were completely absent from V3. If no more than one ambiguity was present, the sequence would represent two single V3 genomes and both would be included for the coreceptor use determination and phylogenetic analysis. Ambiguities in the flanking regions were nevertheless allowed, meaning that each molecular clone would be V3 unique, but could consist of multiple variants sharing the same V3 nucleotide sequence. Sequences that contained more than one ambiguity in V3 were discarded.

3.2.4 Phylogenetic Analysis (Paper III)

Phylogenetic trees were created in **paper III** to illustrate that the generated sequences belonged to separate patients, as well as to evaluate within patients that harboured both R5 and CXCR4-using sequences how the coreceptor use phenotypes were related to each other.

All generated sequences were aligned using Clustal X version 1.81 (www.clustal.org/clustal2/) and manual adjustments to the alignment were then carried out in the Bio Edit Sequence Alignment Editor (www.mbio.ncsu.edu/BioEdit/ BioEdit.html). Maximum likelihood phylogenetic trees were generated with bootstrap values based on 1,000 data set replicates using Phylip version 3.69 (http://evolution. genetics.washington.edu/phylip.html), specifically with the "dnaml," "seqboot," and "consense" programs. Outgroup rooting was utilized on the phylogenetic tree containing all patient sequences (**Figure 13**) using the group O sequence MVP5180

with the accession number L20571. Midpoint rooting was utilized on the phylogenetic tree where the R5 and CXCR4-using sequences of specific patients were evaluated.

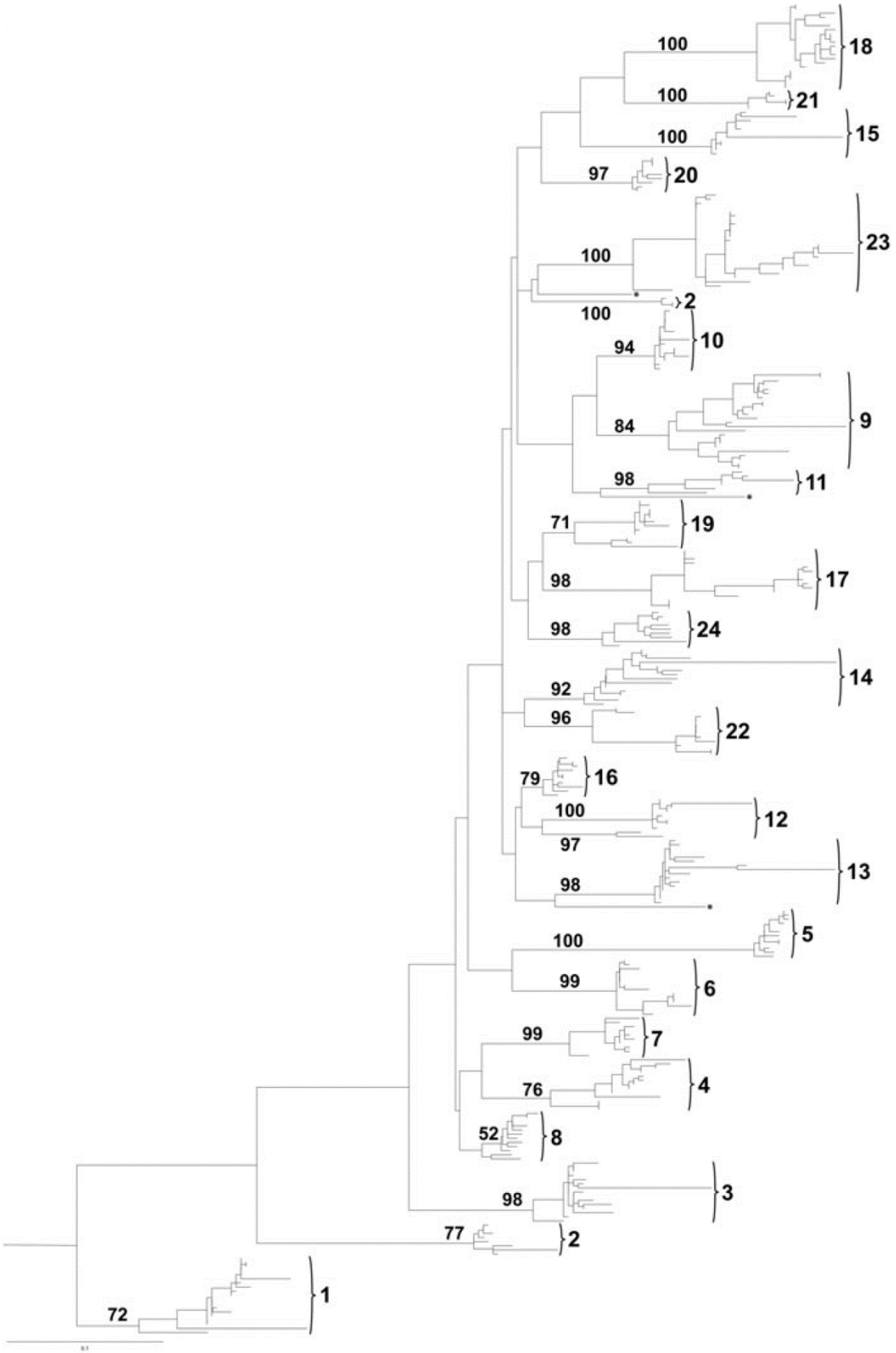

Figure 13. Example of phylogenetic analysis, illustrated using the phylogenetic tree of all patient sequences in paper III. The curly brackets indicate separate patients. Bootstrap values are given in percentages. Reproduced with permission from paper III.

3.2.5 V3 Charge and N-Linked Glycan Site Determination (Papers I to IV)

In **papers I-IV**, total charge of V3 sequences was calculated. The premises for these calculations are that the amino acids lysine (K) and arginine (R) each have a charge of +1, aspartic acid (D) and glutamic acid (E) each have a charge of -1, while histidine (H) has a charge of +0.1 under physiological conditions.

In **papers I**, **III** and **IV**, potential N-linked glycosylation sites, or sequon motifs, were identified in sequences. The sequons were governed by the amino acid order asparagine-X-threonine/serine-Y (N-X-S/T-Y) [271], where X can be any amino acid except proline (P) in the threonine (T) context [272, 273] and also not tryptophan (W), aspartic acid (D), or glutamine (Q) in a serine (S) context [274]. For the Y position, only proline would completely hinder the glycan addition through steric hindrance in both contexts [272, 273].

3.2.6 Coreceptor Use Phenotype Prediction (Papers III and IV)

In **papers III** and **IV**, available phenotype prediction algorithms were used to evaluate the coreceptor use of sequences. The Geno2pheno [coreceptor] learning algorithm [261] with a false positive rate cut-off of 10 % (http://coreceptor.bioinf.mpi-inf.mpg.de/index.php) was used as the primary coreceptor use determination method in **paper III**, since this is the recommended method by the European Consensus Group on clinical management of HIV-1 tropism testing [254].

Geno2pheno with different cut-offs was also one of the evaluated methods used in **paper IV**, along with the learning algorithms PSSM (http://indra.mullins.microbiol.washington.edu/webpssm/) [260] and SVM (http://genegames.org/wetcat/v3.html) [259], and the simple rule algorithms 11/25 [190] and the total charge rule (see section "1.6.1 Simple rule algorithms" for more detailed descriptions). Complex rule algorithms were also evaluated and consisted of C4.5, C4.5-8-12 and PART (http://genegames.org/wetcat/v3.html) [259], as well as the more recent subtype-specific CoRSeqV3-C (**Figure 11**) [262] and Raymond *et al.*'s prediction models for subtype D [263] and CRF01_AE [264] (see section "1.6.3 Complex rule algorithms" for more detailed descriptions).

3.2.7 Statistical Methods (Papers I to IV)

Various statistical methods were utilized in **papers I-IV**. In **paper I**, two-tailed Fisher's exact test was used to compare the significance of the presence or lack of specific glycosylation sites in conjunction with CCR5, CXCR4 or dualtropic use, and a one-tailed test in the analysis of total net charge between R5 and X4, and R5 and dualtropic sequences, respectively.

In **paper II**, the median test was used to compare the difference in charge between R5 and X4 strains in different subtypes. Also, the Wilcoxon signed rank test was used to compare the impact of positions 11 and 25 versus other positions in charge acquisition in the X4 phenotype in each subtype, and the unstructured mixed repeated effect's model was used for the same comparison, but in all subtypes pooled together.

In **paper III**, two-sided Fisher's exact test was applied when comparing the outcome between the treatment-naïve and treatment failure group, using the same method (population sequencing), while the two-sided exact sign test was applied when comparing different methods (population versus single genome sequencing) within the treatment failure group only. Also, two-sided Student's t-test was applied when comparing the genetic distances from the root of the phylogenetic tree between R5 and CXCR4-using strains.

In **paper IV**, sensitivity, specificity, positive predictive values and negative predictive values were calculated for all coreceptor use phenotype prediction algorithms from a CCR5 inhibitor usage perspective, i.e. sequences were classified into CXCR4-using and non-CXCR4-using, with detected CXCR4 use designated as a positive test.

3.2.8 Ethical Considerations (Paper III)

All patients enrolled in **paper III** provided informed consent. The study was approved by the ethical committee in Botswana with the registration numbers HRU-13/18/1 Vol II (5), HRU-13/18/1 Vol VI (37), and PPME-13/18/1 Vol I (89), as well as by the regional board of ethical vetting in Stockholm with the registration number 2008:2007/1496-31/3.

Papers I, **II** and **IV** used publically available sequences from the Los Alamos HIV Sequence database, and hence did not require ethical approvals.

4 RESULTS AND DISCUSSION

4.1 V3 GLYCAN AND CORECEPTOR USE (Paper I)

The main aim of **paper I** was to study the occurrence of potential N-linked glycosylation sites in and around the gp120 V3 loop in 176 individually unique sequences covering the majority of HIV-1 groups, subtypes, and recombinant forms, and investigate if any of these glycan sites could be related to coreceptor use, in particular N301 to CCR5 use. Eight potential N-linked glycosylation sites were examined (**Figure 14**), but only the fifth site, N301, which is located in the V3 loop, showed a correlation with coreceptor use, namely with CCR5 use as compared to exclusive CXCR4 use ($P = \leq 4.6 \times 10^{-12}$).

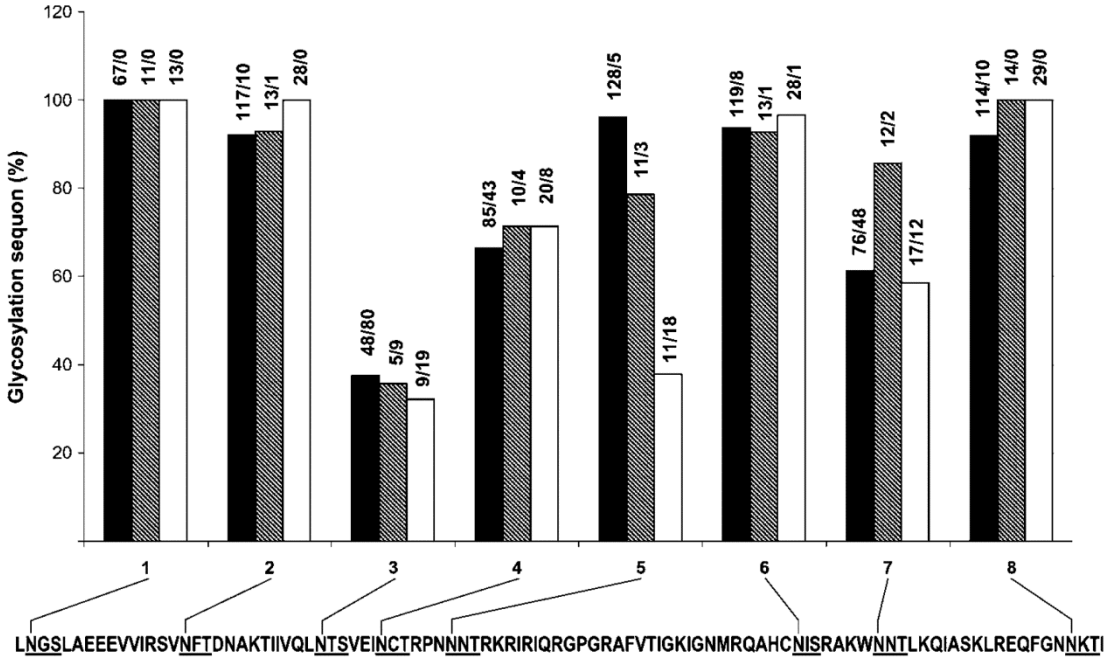

Figure 14. Eight potential N-linked glycosylation sites and their frequencies in the studied V3 sequences. The position of each glycosylation site is underlined on the HXB2R reference sequence, and the presented amino acids have the numbering 228-327 in gp160. The bars display in percentages in how many sequences each glycan site was present, respectively. The filled bars represent R5 sequences, the hatched dualtropic (R5X4), and the empty bars X4 sequences. The number of present/absent glycosylation motifs is given above each bar. Reproduced with permission from **paper I**.

As can be seen from **Figure 14**, only 5 out of 133 R5 sequences lacked the V3 glycosylation motif. Moreover, the particular NNT motif was conserved within the R5 sequences of the known HIV-1 group M subtypes in all sequences except one (120/121; 99.2 %). Interestingly, the 3 group M R5 sequences that lacked the V3 glycosylation site, had the amino acid motifs TNT, HNT, and NNA, respectively, thereby differing only by 1 nucleotide each from the NNT motif.

The loss of the V3 glycosylation site had in previous studies been associated with a coreceptor use phenotype switch from R5 to X4 [208-210]. The findings in **paper I** are in agreement with previous results, but emphasize the importance of the V3 glycan for CCR5 use in HIV-1 group M. A site-directed mutagenesis study targeting the V3 glycan in a dualtropic isolate likewise emphasized the importance of the V3 glycan for CCR5 use, while the effect on CXCR4 use was moderate [212]. Interestingly, when a revertant mutant virus acquired changes in the V3 loop resulting in an increase in net charge, it regained full use of CXCR4 [212]. Hence, the loss of the glycan in itself may not be enough to attain CXCR4 use, since additional changes in the form of the acquirement of positively charged amino acids to increase the total V3 charge seem necessary [41, 212, 260]. Furthermore, it has been observed that the presence or absence of the V3 glycan did not influence the replication ability of X4 viruses in permissive cells [275].

Intriguingly, a later structural study [57] of how the CCR5 N-terminal domain interacts with HIV-1 gp120, pointed out the importance of Asn302 (the middle amino acid position of the V3 glycosylation motif). Asn302 interacts with the sulfotyrosine Tyr14 on the N-terminal of CCR5, and the modification of a single nitrogen, leading to the transformation of Asn302 to Asp302, ablates recognition of CCR5 [57]. A similar modification of the nearby Asn300, which is not part of the V3 glycosylation motif, had little impact on the interaction with CCR5 [57]. This could explain the conservation of the particular NNT motif in the known HIV-1 group M subtypes found in **paper I**.

Of additional interest is that CCR5 lacks two potential N-linked glycosylation sites, one in the N-terminal and one in ECL2, respectively, that are present in CXCR4 [182]. Elimination of these glycans in CXCR4 broadens its coreceptor capacity to include several R5 strains, without affecting entrance of X4 strains [182].

Moreover, R5 viruses have been found to be less sensitive to neutralizing antibodies than co-existing CXCR4 viruses [183, 184]. Hence, yet another observation supporting the importance of the V3 glycan for CCR5 use is that the loss of this site results in increased sensitivity to neutralization [210, 211].

Consequently, the correlation of the V3 glycan with CCR5 use obtained in **paper I** in a material collected globally and spanning practically all HIV-1 groups and subtypes, strengthens the results obtained by previous and newer studies, by displaying their validity in a worldwide and subtype context.

4.2 V3 CHARGE AND CORECEPTOR USE (Papers I and II)

In **paper I**, the role of the V3 loop net charge in relation to coreceptor use phenotype was also evaluated, and as expected, the net charge of R5 viruses (median charge +3.0) was lower than that of X4 (median +5.9) (P = 8.8 x 10^{-12}) and dualtropic viruses (median +5.5) (P = 0.002). In addition, a breaking point at a charge of +4.2 was established, below which there were no X4 sequences and above which there were only 10/133 (7.5 %) R5 sequences. Hence, using net charge above +4.2 to distinguish the X4 phenotype had a sensitivity of a 100 % in the examined material, emphasizing the role

of high V3 net charge for X4 viruses established in previous studies [190, 192, 193, 202, 203, 205, 206].

The main aim of **paper II** was to characterize the underlying amino acid components that result in increased V3 charge in X4 sequences as opposed to R5 sequences in the major HIV-1 subtypes, and from this perspective to describe the coreceptor switch from R5 to X4. Dualtropic sequences were not included in this material in order to ascertain that the observed amino acid properties would pertain to either CCR5 or CXCR4 use.

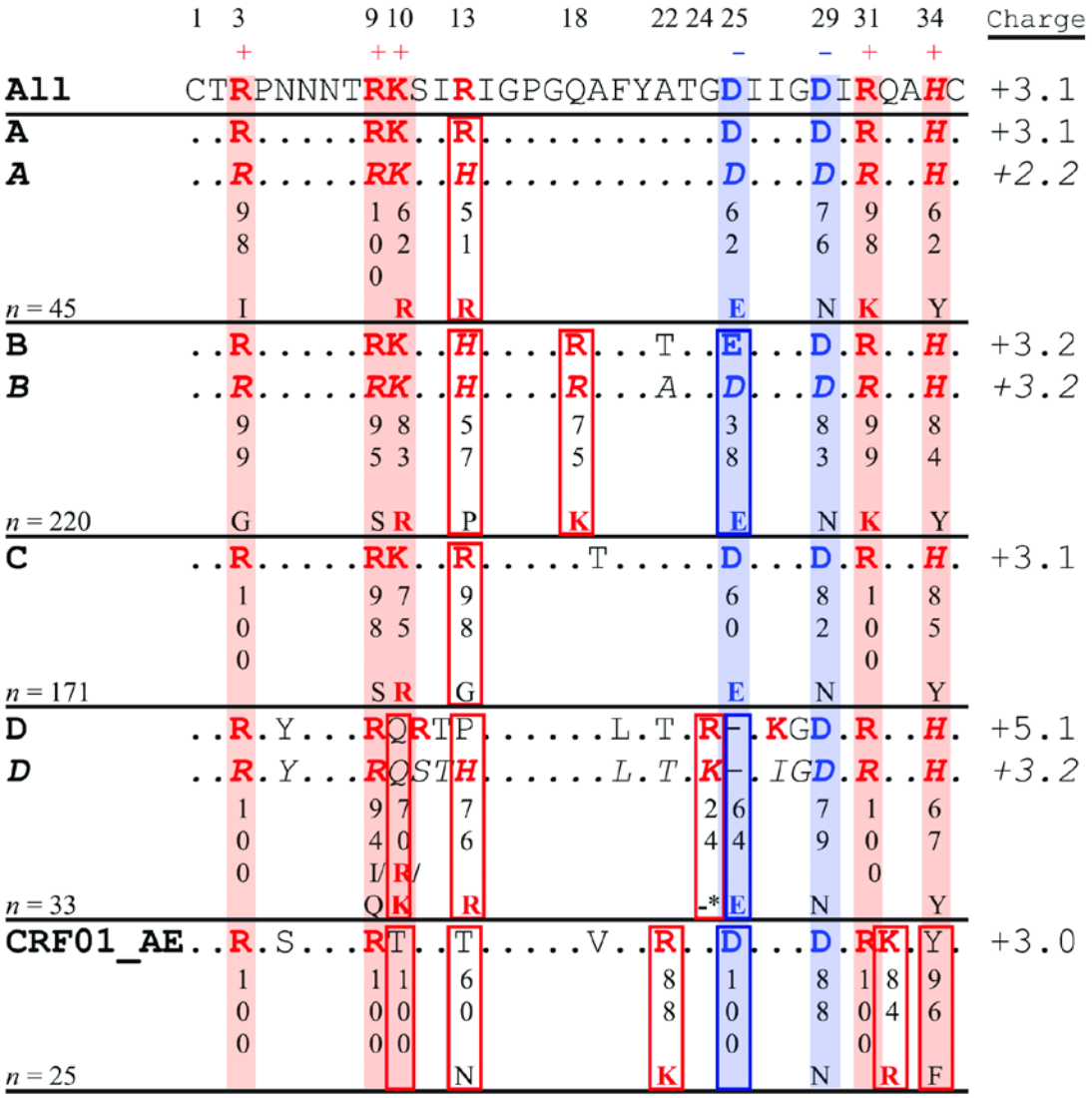

Figure 15. The charged amino acid distributions in the R5 phenotype in different HIV-1 subtypes. The V3 loop position numbering is indicated at the top. Consensus sequences for each subtype derived from the Los Alamos HIV Sequence database are displayed in black. When the consensus sequences based on the material differed from that of the database, they were additionally displayed in italics. The total charge of each consensus sequence is indicated in the charge column. Charged amino acids are bold, the charged amino acid positions conserved in all subtypes are shaded, and the subtype-specific ones are boxed, with positively charged being red and negatively charged being blue. The proportion of the most common amino acid is indicated with a percentage vertically below it, followed by the second most common amino acid. *n* = number of sequences; * = frequency of the second most common amino acid was equal to the first. Reproduced with permission from **paper II**.

Both generally conserved and subtype-specific positively and negatively charged amino acid positions were observed in the R5 phenotype (**Figure 15**). Interestingly, the same positions were discerned in the X4 sequences despite their heterogeneity. Nevertheless, the X4 sequences had a higher net charge than R5. This was attributed to the acquisition of positively charged amino acids other than the subtype-specific and generally conserved ones, and the loss of negatively charged amino acids.

The presence of positively charged amino acids at the V3 loop positions 11 and 25 have in multiple studies been found to be important markers of the X4 phenotype [190, 192, 201, 202, 205, 206]. In **paper II**, the proportion of acquired positively charged amino acids at positions 11 and 25 versus other positions that were not subtype-specific or generally conserved ones were calculated. For all five subtypes, only 41/161 (25 %) acquired positively charged amino acids were at positions 11/25, while 120/161 (75 %) were at other positions ($P < 0.0001$, Wilcoxon signed rank test). Moreover, less than half of the X4 sequences (26/55, 47 %) followed the 11/25 rule [190]. Besides, when all subtypes were pooled together the impact of other positions that were not subtype-specific or generally conserved ones on the overall gain in charge was greater than the impact of positions 11 and 25 ($P = 0.0001$, mixed repeated effect's model). Similarly, other positions had an equal or greater impact in individual sequences on overall charge increase compared with positions 11 and 25 in 73 % of the total X4 sequences.

Consequently, the observations in **paper II** of less position-dependent acquisition of positive charge provide a simple and plausible explanation for the evolution of X4 from R5, decreasing the fundamental role of the 11/25 positions. The importance of the basic character of the V3 loop in X4 viruses described in **paper II** is further strengthened by its electrostatic compatibility with the acidic residues in the binding pocket of CXCR4 observed in mutagenesis and structural studies [180, 181].

4.3 GLYCAN-CHARGE MODEL (Papers I to IV)

The V3 sequence properties associated with coreceptor use that were observed in **papers I** and **II** were used to construct a model for discriminating between different coreceptor use phenotypes, referred to as the glycan-charge model (**Table 1**).

According to the glycan-charge model, R5 sequences are characterized by the presence of the V3 glycosylation site N301 and low V3 charge, while X4 sequences are characterized by high V3 charge, regardless of the presence of the V3 glycan. Consequently, based on these premises the dualtropic sequences should combine the R5 property of having the glycan with the X4 property of high V3 charge.

As mentioned above, a charge cut-off of +4.2 below which there were no X4 sequences was noted in **paper I**. This cut-off was used in the glycan-charge model (**Table 1a**) to define the lower charge limit of the dualtropic group, while the higher charge limit was defined using the highest V3 charge (+6.1) among the dualtropic sequences in **paper I**. However, within these values, R5 and X4 sequences could be found too. Hence, this group is designated as mixed, since the model cannot distinguish between R5, R5X4

and X4 when the charge of the V3 sequences falls into this category and the glycan motif is concurrently present.

Table 1a. Glycan-charge model based on **paper I**.

	Envelope Glycoprotein 120 V3 Region Charge		
	< 4.2	**≥ 4.2 – 6.1 ≤**	**> 6.1**
Glycan N301 +	R5	R5, R5X4, X4	X4
Glycan N301 –	Undetermined	X4	X4

Table 1b. Subtype-specific versions of the glycan-charge model based on **paper II**.

	Envelope Glycoprotein 120 V3 Region Charge		
Subtype A	**< 5.3**	**≥ 5.3**	**≥ 5.3**
Subtype B	**< 3.3**	**≥ 3.3 – 7.1 ≤**	**> 7.1**
Subtype C	**< 5.0**	**≥ 5.0 – 5.1 ≤**	**> 5.1**
Subtype D	**< 5.0**	**≥ 5.0 – 6.2 ≤**	**> 6.2**
CRF01_AE	**< 4.0**	**≥ 4.0 – 5.0 ≤**	**> 5.0**
Glycan N301 +	R5	R5, R5X4, X4	X4
Glycan N301 –	Undetermined	X4	X4

The tables are adapted from **paper IV**.

The characterization of charged V3 amino acids in different coreceptor use phenotypes described in **paper II** allowed the modification of the glycan-charge model by the application of subtype-specific charge cut-offs. However, no dualtropic sequences were included in **paper II**, so the charge cut-offs of the mixed group were defined using the range between the X4 sequence with the lowest V3 charge and the R5 sequence with the highest V3 charge for each studied subtype and recombinant form. In the subtype A sequences there was no overlap between the charge of R5 and X4 sequences, and hence the mixed group could not be distinguished from the X4 group by using charge cut-offs.

When V3 sequences lacking the V3 glycan are encountered with a charge too low to be classified as CXCR4-using, they are classified as "undetermined," since they cannot be explained using the glycan-charge model. Such sequences are however rare, and in almost all cases their actual biological coreceptor use phenotype is R5. A theoretical explanation to the presence of such sequences could be that they were derived from plasma samples, where uninfectious or defective viruses also may circulate. Another explanation could be that sequences lacking the glycan might still be able to use CCR5 when the charge is too low for CXCR4 utilization, but that those viruses would probably be outcompeted *in vivo* by R5 viruses containing the V3 glycan, since those variants would be better at evading the immune system [210, 211], which could also account for why sequences belonging to the "undetermined" category are rarely encountered.

The glycan-charge model attempts to describe coreceptor use based on biological properties. In order for it to be of clinical use, however, it needed to be converted into an algorithm form that would categorize sequences into CXCR4-using and non-CXCR4-using, which is achieved by classifying the mixed group category as CXCR4-using, even though this entails the risk of at times generating false positive results, since the mixed group normally also may contain pure R5. As the glycan-charge model is based on biological properties, the algorithm derived from it was classified as an "analytical" one. This algorithm was evaluated along with other algorithms for coreceptor use prediction in **paper IV**. A subtype C-specific version of the glycan-charge model algorithm was also applied on sequences from treatment failure patients in **paper III**.

4.4 CORECEPTOR USE PHENOTYPE DISTRIBUTION IN DIFFERENT SUBTYPES (Papers I, II and IV)

In **papers I**, **II** and **IV**, sequences belonging to different HIV-1 subtypes with known biologically determined coreceptor use were retrieved from the Los Alamos HIV Sequence database, which contains sequences collected worldwide since the beginning of the 1990s. This opened the possibility to study the coreceptor use phenotype distribution among the major subtypes, leaving in mind, however, that the perceived distribution changes could partially be attributed to the choice of patients being included in research studies.

The sequences in **paper I** were a mix of HIV-1 groups, subtypes, sub-subtypes and recombinant forms. Unfortunately, however, within each category there were too few sequences for meaningful within-subtype analysis of coreceptor use. There was especially a lack of dualtropic and X4 sequences, although it can be noted that subtype D had more X4 sequences (n = 9) than R5 (n = 2), which would be in accordance with the observations of increased CXCR4 use in subtype D made in other studies [112, 186, 226].

Table 2. Subtype-specific distributions of the R5 and X4 phenotypes in **paper II**.

	Number of sequences	R5	X4	% R5
Subtype A	48	45	3	93.8
Subtype B	231	210	21	90.9
Subtype C	180	171	9	95.0
Subtype D	37	31	6	83.8
CRF01_AE	32	25	7	78.1
Total	**528**	**482**	**46**	**91.3**

The table was adapted with permission from the supplementary material in **paper II**.

In **paper II**, coreceptor use was studied in the major HIV-1 group M subtypes, consisting of subtypes A, B, C, D and the circulating recombinant form CRF01_AE (**Table 2**). Also here, the amount of X4 sequences was low, but subtype C displayed nevertheless a significantly higher proportion of the R5 phenotype ($P = 0.03$, two-tailed Fisher's exact test), while CRF01_AE displayed a significantly higher proportion of the X4 phenotype ($P = 0.015$) compared to the other subtypes included in the study. This is in accordance with earlier studies associating subtype C with low CXCR4 use [207, 230-232], as well as with both later and earlier reports associating CRF01_AE with higher CXCR4 use [227-229].

The same subtypes were studied in **paper IV**, but a lot more sequences could be included due to the increase of available sequences in the Los Alamos HIV Sequence database. However, since the criteria of sequence inclusion were one sequence per patient per phenotype, no coreceptor use distribution calculations would be valid, since there would be an overrepresentation of some individuals. Nevertheless, all subtypes had by now more reported R5 sequences than they had dualtropic or X4 sequences, and subtype A had an exceedingly low amount of reported X4 sequences ($n = 4$), which could either indicate the smaller amount of studies being conducted on this subtype or imply an actual indication of low CXCR4 use in this subtype.

4.5 CORECEPTOR USE IN TREATMENT FAILURE IN SUBTYPE C

(Paper III)

Despite the correlation of subtype C with low CXCR4 use observed in **paper II** and in other earlier studies [207, 230-232], there have more recently been quite a few reports of increased CXCR4 use in treatment failure patients infected with subtype C [234, 236, 237]. The main aim of **paper III** was to investigate if there is an increase in CXCR4-using viruses in HIV-1 subtype C patients failing antiretroviral treatment in Botswana compared to treatment-naïve individuals using population sequencing and genotypic assays. The increase of CXCR4-using sequences in treatment-experienced patients, especially in a subtype previously reported to have little CXCR4 use, is of special consequence with regard to treatment with CCR5 inhibitors, since the only CCR5 inhibitor in clinical use, maraviroc, is mostly used as a salvage therapy drug [243].

To assess the coreceptor use phenotypes obtained from population sequencing in the treatment-naïve and treatment failure patients, the Geno2pheno algorithm with a 10 % false positive rate cut-off was used, which is currently the recommended algorithm for genotypic assays according to the European Consensus Group on clinical management of HIV-1 tropism testing [254]. In addition, the glycan-charge algorithm adjusted for subtype C based on **papers I** and **II**, was also used for phenotype prediction.

Of the 26 population sequences belonging to the treatment-naïve patients, 2 were predicted to be from viruses able to use CXCR4 by the Geno2pheno method, while the glycan-charge algorithm predicted all 26 to be R5 (**Table 3**). Among the 24 population sequences belonging to the treatment-experienced patients, 8 were predicted to be from CXCR4-using viruses by the Geno2Pheno method, while the glycan-charge algorithm predicted 5 sequences to be potential CXCR4 users (**Table 3**). The seemingly higher

proportion of CXCR4-using viruses in the treated individuals' populations (8/24, 33.3%) versus the treatment-naïve individuals' populations (2/26, 7.7 %) as predicted by Geno2pheno was statistically significant (P = 0.03). This was also the case for the predictions made using the glycan-charge algorithm: 5/24 (20.8 %) treated CXCR4 users versus 0/26 (9%) treatment-naïve users (P = 0.02).

Table 3. Prevalence of potential CXCR4-using HIV-1 subtype C sequences in treatment-naïve and treatment failure patients using the Geno2pheno and glycan-charge algorithms, respectively, for coreceptor use phenotype prediction.

		Geno2Pheno [coreceptor] (European recom. cutoff)			**Glycan-Charge Model* (subtype C)**		
		Potential CXCR4-using sequences			Potential CXCR4-using sequences		
Patient population	Sequencing protocol	No. of patients	Percent (%)	Statistical tests	No. of patients	Percent (%)	Statistical tests
Treatment-naïve	Population sequencing	2/26	7.7	} p = 0.03†	0/26	0	} p = 0.02†
Treatment failure	Population sequencing	8/24	33.3	} p = 0.5‡	5/24	20.8	} p = 0.001‡
Treatment failure	Single genome sequencing	10/24	41.7		16/24	66.7	

*The mixed group is regarded as CXCR4-using. †Two-tailed Fisher's exact test. ‡Two-tailed exact sign test. European recom. cut-off = recommendations from the European Consensus Group on clinical management of HIV-1 tropism testing (10 % false-positive rate). Adapted with permission from **paper III**.

Single genome sequencing was applied in an attempt to identify CXCR4 use in minor HIV-1 quasispecies in the treatment-experienced patients that may have not been detected using population sequencing. Two more patients were found to harbour CXCR4-using viruses using this sequencing method and the Geno2pheno algorithm compared with population sequencing results with the same algorithm, while correspondingly 11 more individuals were found to harbour CXCR4-using viruses using the glycan-charge algorithm (**Table 3**).

Consequently, a statistically significant increased frequency of CXCR4-using viruses was detected by genotypic tropism testing on population sequences in HIV-1 subtype C infected patients failing antiretroviral treatment compared with treatment-naïve patients, suggesting that CCR5 inhibitors may be less suitable as drugs in treatment-experienced as opposed to treatment-naïve patients. Additionally, an even higher frequency of CXCR4-using viruses was detected using single genome sequencing in treatment-experienced patients.

An increase in CXCR4 use in treatment failure patients has previously been ascribed to factors such as suboptimal treatment [234], specific drug resistance mutations [234, 276], or low CD4 counts [234, 236, 237]. However, in Botswana the treatment

adherence rates have been good [257], and no statistically significant correlation could be observed in **paper III** between coreceptor use and specific drug resistance mutations or drug classes, although it should be noted that drug resistance profiles were available for only half of the patients. With regard to low CD4 counts, an attempt was made to minimize this bias by including control patients from an earlier cohort with CD4 counts that would have warranted treatment according to the current national guidelines, and hence likely matched the CD4 counts of the treatment-experienced patients. However, consequently since the collection date was later for the treatment-experienced patients, the observed increased CXCR4 use could be the result of the evolving subtype C epidemic [183, 234, 235].

Another large recent study in Botswana, which focused on subtype C infected treatment-naïve women with low CD4 counts, showed that the prevalence of CXCR4-using virus was still low in this group [277] compared to prevalence in subtype B infected individuals. These results in conjunction with those obtained in **paper III** further support the usefulness of maraviroc as a first-line drug among subtype C patients, at least in Botswana.

A limitation of **paper III** was that patients failing first-line ART could not be included, since according to the national treatment guidelines, only patients failing second-line treatment are eligible for drug resistance genotyping. As the source of samples was the national drug resistance genotyping laboratory, only samples where genotyping had been performed for clinical use could consequently be accessed. Hence, the treatment-experienced patients included in **paper III** may have had more opportunity to acquire multiple drug resistance mutations, which has been linked to CXCR4 use [278], compared to treatment-experienced patients who fail first-line treatment.

Another limitation of **paper III** was the unavailability of material for creating single genomes from the treatment-naïve patients, since only population sequencing data from 2003 were available for the treatment-naïve patients and not their DNA samples. For this reason, a comparison between the treatment-naïve and the treatment-experienced patients could only be performed on a population sequencing level, ruling out the possibility to assess how the potential existence of minor CXCR4-using quasispecies in the treatment-naïve group would affect the comparison with the treatment-experienced group, where it was shown that more CXCR4 users were detected with single genome sequencing. On the other hand, the clinical impact of small proportions of CXCR4-using virus among a patient's population of viruses is not known. In **paper III**, the potential CXCR4-using strains in a single individual's viral population could be detected if they represented at least 7 % of the total quasispecies. If even smaller proportions of CXCR4 use in a patient's circulating population of viruses are of clinical importance, more sensitive methods would need to be used, such as pyrosequencing [279, 280], which however are more expensive and labour intensive, and hence more difficult to use in resource-limited settings.

4.6 CORECEPTOR USE PREDICTION MODELS AND CCR5 INHIBITOR TREATMENT (Papers III and IV)

In **paper III**, two different coreceptor use prediction models were used, namely the recommended Geno2pheno method and the glycan-charge algorithm adjusted for subtype C based on **papers I** and **II**. The predictions made by both algorithms were plausible even when they did not overlap, as demonstrated using phylogenetic analysis methods, implying that the glycan-charge model adjusted for subtype C might have potential for further development for clinical use purposes.

Since the introduction of CCR5 inhibitors in ART regimens, accurate coreceptor use prediction models for utilization in genotypic assays, in particular those that take into consideration subtype variations, have been actively sought after, since biological determination of coreceptor use is time-consuming and expensive, and hence acts as a limitation to the accessibility of CCR5 inhibitors. The main aim of **paper IV** was to compare the performance of current available coreceptor use prediction algorithms in major HIV-1 subtypes from a CCR5 inhibitor treatment perspective. The collected testing material in **paper IV** was uniquely suited for the purpose of testing these algorithms due to several aspects. It encompassed patient-derived sequences gathered globally that had been accumulated over roughly the latest 15 years of the epidemic up until present day. Sequences from even earlier time points were not included, since they were not classified according to the coreceptor use phenotype classification, which was not implemented until 1996 [118], thereby reducing possible misclassifications of coreceptor use due to inference from previous phenotype classifications. Furthermore, the material underwent a uniquely rigorous scrutiny with regard to the verification of the biological determination of the coreceptor use phenotypes, which was based on the original article sources as opposed to the information provided by the database due to the database information often being incomplete and at times incorrect.

The 24 evaluated algorithms included simple rule algorithms, learning rule algorithms, complex rule algorithms, and the “analytical” glycan-charge algorithm based on **papers I** and **II** and applied in **paper III**. Some of the algorithms were subtype-specific and were hence only evaluated on sequences from their respective subtype.

Learning algorithms generally performed well at predicting coreceptor use phenotypes in all studied subtypes. The algorithm that performed best on the whole material was the learning algorithm Geno2pheno (G2P): 2.5 % (false positive rate cut-off) with specificity and negative predictive value at 94 % or above, positive predictive value at 84 %, and sensitivity at 77 %, closely followed by G2P: 5 %, which performed similarly, but had higher sensitivity at 77 %, but lower positive predictive value at 70 %. The clinically recommended G2P: 10 % performed well as well, with sensitivity and specificity above 80 % and negative predictive value at 96 %, but had a low positive predictive value at 55 %, indicating a trade-off for higher sensitivity by allowing more false positives.

The rule algorithms, both the simple and complex, had consistently high specificities, ranging from 92 to 99 % in all subtypes, but were usually not very sensitive. An exception to that were the more recent subtype-specific complex algorithms, where the

subtype C specific CorSeqV3-C [262] outperformed all the studied methods with all values above 90 %, but as it was subtype-specific, its values were only applicable for that subtype. Similarly, the subtype-specific complex algorithms created by Raymond *et al.* for subtype D [263] and CRF01_AE [264] also had the best sensitivities among the remaining rule algorithms, with 90 % and 67 %, respectively.

Sensitivities were generally low for all algorithms tested on subtype A, ranging from 33 to 75 %, perhaps implying that subtype A X4 sequences may be more homogeneous and hence more similar to subtype A R5 sequences, rendering them more difficult to distinguish, which would increase the risk of generating false negatives. In contrast to this, the algorithms tested on subtype D were not very specific, since 9 out of 18 tested methods had values below 75 %, possibly implying that subtype D R5 sequences may be more heterogeneous and hence more similar to subtype D X4 sequences, making them more difficult to distinguish as well, which would increase the risk of generating false positives. A plausible explanation for these observations could be that viruses of different subtypes need to travel different evolutionary distances for their R5 viruses to evolve CXCR4 use [41]. The observations made in **paper IV** regarding subtypes A and D are of special interest, since these subtypes co-exist in the same regions of sub-Saharan Africa, but subtype D has been associated with a larger proportion of CXCR4-using viruses [112, 186, 226] and a faster disease progression compared to subtype A [108-111]. Of additional interest was that none of the tested algorithms performed particularly well on subtype B even though most of them were designed using subtype B sequences. A similar trend was noted for subtype D sequences, which are closely related to subtype B sequences [2, 281] and display quite a high level of heterogeneity.

The “analytical” glycan-charge algorithm had generally an average performance on all subtypes, excepting perhaps subtype C where it performed quite well. The glycan-charge model is well suited for descriptive and explanatory purposes of the biological properties of coreceptor use, but needs to be developed further in order to be useful as an algorithm designed for clinical use. The glycan-charge model had particular difficulties with classifying correctly the dualtropic sequences, which in fact was difficult for all of the evaluated algorithms. One possible reason for this was that some of the sequences that were denoted as dualtropic actually could be R5 or X4, since they might have been derived from a dual/mixed population. Another possible reason could be that some of the dualtropic sequences utilized CCR5 or CXCR4 more efficiently than the other coreceptor [186, 245], and hence the sequences would look more R5- or X4-like, respectively [186]. It now remains to elucidate the biological role of these dualtropics. Perhaps it is not crucial for prediction algorithms in clinical use to detect R5-like dualtropic sequences, since it is possible that their impact on drug resistance is negligible due to an inefficient use of CXCR4, and that maybe only the dualtropic sequences that are X4-like will not be blocked by CCR5 inhibitors. This is inferred from that X4-like dualtropic sequences were efficiently inhibited by CXCR4 inhibitors, while R5-like dualtropic sequences were not [245]. Hence, studies like **paper II**, where the amino acid predictors of coreceptor use are studied in pure R5 and X4 sequences, are crucial for the validation of the actual properties that pertain to CCR5 and CXCR4 use, respectively, that can then be applied to how dualtropic sequences are classified. Moreover, some dualtropic sequences have identical V3 loops with R5 sequences [186], indicating the potential role of regions outside of V3 in coreceptor use. How

important these regions are for the ability to acquire efficient CXCR4 use needs to however be evaluated further.

4.7 LIMITATIONS OF DATABASE MATERIAL (Papers I, II and IV)

The Los Alamos HIV Sequence database, from which material for **papers I**, **II** and **IV** was retrieved, is a tremendous and unique asset for the HIV-1 research field, since it enables large amounts of sequences representing the whole pandemic to be studied. However, it has a number of limitations, in particular regarding the reporting of coreceptor use, that need to be kept in mind during the retrieval and subsequent analysis of the database sequences.

There is no current golden standard regarding the biological determination of coreceptor use, which may lead to quality variations of the collected sequences, with there not always being a perfect correlation between given sequence and alleged coreceptor use phenotype, exemplified by sequences derived from plasma although the coreceptor use determination was performed using viral isolates in entry assays. Also, the coreceptor use of many older sequences is inferred from previous classifications.

There is a skewed distribution in the database with regard to R5 and X4 sequences, with there being a lot fewer dualtropic and X4 sequences submitted. The low amount of CXCR4-using sequences will influence the ability to reach statistical significance of findings when R5 and X4 sequences are compared with each other, and the validity of any observations made in R5 sequences will inevitably weigh greater, since they will be based on more sequences.

Even though the sequences by and large are representative of the pandemic, there may be a bias in the subtype and phenotype distributions regard to which patients and in which countries most studies are conducted.

In addition to all of the above, the information accompanying the sequences in the database may be incomplete or sometimes even incorrect. For instance, sequences listed as dualtropic are often actually dual/mixed, and some sequences with reported coreceptor use were in fact never evaluated for that. Consequently, it is of great importance that the information given in the database is verified with the original articles, like was done in **papers II** and **IV**, during the selection and retrieval of sequences from the database for analysis.

5 CONCLUDING REMARKS AND FUTURE PERSPECTIVES

To study the evolution of the coreceptor switch from R5 to X4 is important, since it may have clinical implications. It remains unknown if the acquirement of CXCR4 use by HIV-1 variants within a patient is the cause or the result of disease progression. Regardless, more knowledge is needed concerning the amino acid predictors of coreceptor use in HIV-1 sequences, since this could provide information about disease stage as well as be applied to antiretroviral treatment, in particular to the clinical use of CCR5 inhibitors.

The findings in **papers I** and **II** support and strengthen current knowledge regarding the role of the V3 glycan and charge in coreceptor use, as well as demonstrate the validity of this knowledge in a global and subtype context. This information can help improve as well as create new coreceptor use prediction algorithms utilized in genotypic tropism testing prior to the administration of CCR5 inhibitors, thereby improving their accessibility in resource-limited settings. The analytic approach of the glycan-charge coreceptor prediction model, based on the observations in **papers I** and **II**, that is firmly rooted in the biological properties of coreceptor use, can be of particular use in the improvement of coreceptor use prediction algorithms, since it elucidates the truly important components determining coreceptor use. Hence, elements derived from the glycan-charge model, such as the importance of the V3 glycan for CCR5 use, can be incorporated into complex rule prediction algorithms where they would weigh more heavily in the prediction process compared to other rules not derived from biological property observations.

The varying distribution of HIV-1 subtypes worldwide and their possible differences in disease progression and coreceptor use emphasize the need to develop coreceptor use prediction algorithms that are subtype-specific. Moreover, subtype-specific prediction algorithms were shown in **paper IV** to perform the best with regard to predicting coreceptor use from a CCR5 inhibitor treatment perspective. As new coreceptor prediction algorithms frequently emerge and methods to determine biological coreceptor use phenotype become more and more standardized, the continual evaluation of the performance of available algorithms, as was done in **paper IV**, becomes imperative.

The findings in **paper III** of increased CXCR4 use in treatment failure patients infected with subtype C compared with treatment-naïve patients could likewise have significant clinical implications, since these findings help to elucidate which patient group might be a poorer candidate for treatment with CCR5 inhibitors. Future studies are warranted that compare treatment outcomes and drug resistance development in patients treated with CCR5 inhibitors within salvage therapy regimens compared to within first-line treatment regimens. These studies would be able to address the pressing question regarding what proportions of CXCR4-using variants and which particular CXCR4-using variants within a patient are of actual clinical importance, which could help in the choice of sequence generating method as well as of phenotype prediction algorithm.

6 ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to the people that have aided and supported me during this long and tedious but rewarding journey we all call PhD education:

First of all, none of this would be possible without my late main supervisor, **Anneka Ehrnst**. Anneka, you were not only a great source of inspiration to me ever since I first entered your lab at the age of 17, you were an enthusiastic teacher and a generous person, who invested so much time in your PhD students, but above all, to me, you were my dearest friend. Words cannot express how much I miss you, but in my heart and in my mind you were guiding me all the way up until this moment, even though you are no longer with us. Thank you for preparing me so well and for the wonderful years in science that I got to spend working with you.

I am also infinitely grateful to my current main supervisor, **Gunilla Karlsson Hedestam**. Thank you, Nilla for welcoming me into your wonderful group and giving me such free reigns to complete all of my work. You have inspired and guided me more than you know, and I have really learned a lot from you during this past year. I look forward to our future collaborations together.

Peter Clevestig, my co-supervisor, you have been there from the very start to the end. I learned my very first lab skills from you, just pipetting with water, and I'll never forget the highlights of my early scientific days, which were composed of solving crossword puzzles with you during our coffee breaks.

My former lab group:

Rozina Caridha, my dear, dear friend. My lab life changed forever when you joined our group. Thank you for the immense support over the years through the highs and lows, and for being such an unselfish, wonderful being.

Sven Grützmeier, for all the inspiring clinical discussions over the years, which has helped form me in my role as a doctor. Thank you also for all the lovely dinners we've had with our little group at your place. It has really helped us keep our identity and remember our dear Anneka.

Simani Gaseitsiwe, for the wonderful collaboration on **paper III**, and for teaching me about the Botswanian culture.

My current lab group:
Marjon, **Gabriel**, **Monika**, **Ganesh**, **Martina**, **Paola**, **Marc**, **Faezzah**, **Saskia**, **Flavio**, **Anna**, and **Gerry**. Thank you for making me feel so welcome. You are all such great people and I'm glad I got to spend so much time with you this year.

Ingemar Ernberg, thank you for being such an inspirational and enthusiastic group leader. Your advice over the years have been invaluable.
Ingemar Ernberg's group: I've really enjoyed all the Friday coffee breaks with you! Thank you for helping me out during the times when I shared a lab with you.

To my mentor:
Hanna Brauner, you have been superb at supporting me over the years!

To my half-time committee members:
Jan Albert, **Patrik Medstrand**, and **Annelie Tjernlund**, for giving me the right guidance and feedback that has helped me reach the point where I am now.

Special thanks to **Francesca Chiodi** and **The-Hung Bui**, for always showing an interest and being supportive.

To my students in my former group:
Johanna, **Desmond**, and **Marcus** – thank you for helping me learn to teach and for being such enthusiastic learners yourselves.

To all of my co-authors not mentioned above:
Max Essex, **Madisa Mine**, **Thomas Leitner**, **Ulrik Fried**, **Gaseene Sebetso**, **Thongbotho Mphoyakgosi**, and **Thabo Diphoko**. Thank you for all your valuable comments and input in our joint collaborations.

To all the kind MTC people that have made my PhD years so much more fun:
Maryam, **Chaniya**, **Maria-Lisa**, **Lech**, **Jola**, **Stefano**, **Qin**, **Frank**, **Tatiana**, and many many more – you know who you are!

To all of my friends with special thanks to:
Camilla, for teaching me how to use ReadCube and letting me share and vent when times have been low in science.
Cecilia, for all the constant pep talks and the lovely pep card that helped me through the last half year of my PhD education.
Also thanks to **Vendela**, **Eyvind**, **Olga** and **Tomas**, **Shermineh**, **Nastya**, **Robert**, **Jonathan**, **Hannes**, **Anders** and **Theresia**, **Sofia** and **Hannes**, **Malin** and **Henrik**, **Resha**, **Fredrik**, **Piotr**, **Katri** and **Oleg** and their **girls**, **Irina** and **Hossein** with their **boys**, **Elena Grigorjevna**, **Lena** and **Patrik** with their **girls**, **Pia** and **Svante**, **Valera** and **Lilja**, and everyone else who's been there for me!

And then of course to my wonderful family:

My grandparents **Larissa** and **Slava**, who have provided with inspiration, food, support and love all throughout these times.

To my parents-in-law, **Frans** and **Eva**, as well as my brother-in-law, **David**, for welcoming me into their family and always being ready to help out when computer problems were nigh.

To my dear cousins, **Shakil** and **Sheila**, **Shohag** and **Prova**, **Moon**, **Mehrab** and of course the unforgettable darlings **Nusaib** and **Samrina**. Thank your for believing in me and always being there.

To my aunt **Larisa**, who always has the right word to say when I'm feeling low.

To my mother **Svetlana**, who has sat up through long nights in solidarity when I've been working hard, and has always pushed me to be better, which I'm very thankful for.

To my father **Aladdin**, my big scientific inspiration. When I hear the word scientist, I always think of you first, because you are the complete personification of what a scientist should be! Thank you for reviewing every part of this thesis and supporting me so much! I'm so proud to be walking in your footsteps.

To my beloved little brother **Alec**, you are the pride and joy of my life!! Thank you for helping me with the formatting of this thesis. You are truly the best brother one can have.

And last but not least, to the love of my life, **Daniel**, my husband. You've truly been there for me all the way from the very beginning, and you understand me like no one else. I'm immensely grateful for the life that we share together.

7 REFERENCES

1. Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med.* Sep 2011;1(1):a006841.

2. Peeters M, Jung M, Ayouba A. The origin and molecular epidemiology of HIV. *Expert Rev Anti Infect Ther.* Sep 2013;11(9):885-896.

3. Keele BF, Van Heuverswyn F, Li Y, et al. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science.* Jul 28 2006;313(5786):523-526.

4. Van Heuverswyn F, Li Y, Neel C, et al. Human immunodeficiency viruses: SIV infection in wild gorillas. *Nature.* Nov 9 2006;444(7116):164.

5. de Silva TI, Cotten M, Rowland-Jones SL. HIV-2: the forgotten AIDS virus. *Trends Microbiol.* Dec 2008;16(12):588-595.

6. Gao F, Yue L, White AT, et al. Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. *Nature.* Aug 6 1992;358(6386):495-499.

7. Chen Z, Telfier P, Gettie A, et al. Genetic characterization of new West African simian immunodeficiency virus SIVsm: geographic clustering of household-derived SIV strains with human immunodeficiency virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey troop. *J Virol.* Jun 1996;70(6):3617-3627.

8. Peeters M, Courgnaud V, Abela B, et al. Risk to human health from a plethora of simian immunodeficiency viruses in primate bushmeat. *Emerg Infect Dis.* May 2002;8(5):451-457.

9. Locatelli S, Peeters M. Cross-species transmission of simian retroviruses: how and why they could lead to the emergence of new diseases in the human population. *AIDS.* Mar 27 2012;26(6):659-673.

10. UNAIDS. 2013 Global Report Epidemiology Slides. 2013.

11. Korber B, Muldoon M, Theiler J, et al. Timing the ancestor of the HIV-1 pandemic strains. *Science.* Jun 9 2000;288(5472):1789-1796.

12. Worobey M, Gemmel M, Teuwen DE, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature.* Oct 2 2008;455(7213):661-664.

13. Zhu T, Korber BT, Nahmias AJ, Hooper E, Sharp PM, Ho DD. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature.* Feb 5 1998;391(6667):594-597.

14. Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California. *MMWR Morb Mortal Wkly Rep.* Jul 3 1981;30(25):305-308.

15. Greene WC. A history of AIDS: looking back to see ahead. *Eur J Immunol.* Nov 2007;37 Suppl 1:S94-102.

16. Barre-Sinoussi F, Chermann JC, Rey F, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science.* May 20 1983;220(4599):868-871.

17. Gallo RC, Salahuddin SZ, Popovic M, et al. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science.* May 4 1984;224(4648):500-503.

18. Coffin J, Haase A, Levy JA, et al. What to call the AIDS virus? *Nature.* May 1-7 1986;321(6065):10.

19. Safai B, Sarngadharan MG, Groopman JE, et al. Seroepidemiological studies of human T-lymphotropic retrovirus type III in acquired immunodeficiency syndrome. *Lancet.* Jun 30 1984;1(8392):1438-1440.

20. Sarngadharan MG, Popovic M, Bruch L, Schupbach J, Gallo RC. Antibodies reactive with human T-lymphotropic retroviruses (HTLV-III) in the serum of patients with AIDS. *Science.* May 4 1984;224(4648):506-508.
21. Collier AC, Coombs RW, Schoenfeld DA, et al. Treatment of human immunodeficiency virus infection with saquinavir, zidovudine, and zalcitabine. AIDS Clinical Trials Group. *N Engl J Med.* Apr 18 1996;334(16):1011-1017.
22. D'Aquila RT, Hughes MD, Johnson VA, et al. Nevirapine, zidovudine, and didanosine compared with zidovudine and didanosine in patients with HIV-1 infection. A randomized, double-blind, placebo-controlled trial. National Institute of Allergy and Infectious Diseases AIDS Clinical Trials Group Protocol 241 Investigators. *Ann Intern Med.* Jun 15 1996;124(12):1019-1030.
23. Staszewski S, Miller V, Rehmet S, et al. Virological and immunological analysis of a triple combination pilot study with loviride, lamivudine and zidovudine in HIV-1-infected patients. *AIDS.* May 1996;10(5):F1-7.
24. Gallant JE, DeJesus E, Arribas JR, et al. Tenofovir DF, emtricitabine, and efavirenz vs. zidovudine, lamivudine, and efavirenz for HIV. *N Engl J Med.* Jan 19 2006;354(3):251-260.
25. Robbins GK, De Gruttola V, Shafer RW, et al. Comparison of sequential three-drug regimens as initial therapy for HIV-1 infection. *N Engl J Med.* Dec 11 2003;349(24):2293-2303.
26. van Leth F, Phanuphak P, Ruxrungtham K, et al. Comparison of first-line antiretroviral therapy with regimens including nevirapine, efavirenz, or both drugs, plus stavudine and lamivudine: a randomised open-label trial, the 2NN Study. *Lancet.* Apr 17 2004;363(9417):1253-1263.
27. Life expectancy of individuals on combination antiretroviral therapy in high-income countries: a collaborative analysis of 14 cohort studies. *Lancet.* Jul 26 2008;372(9635):293-299.
28. Braitstein P, Brinkhof MW, Dabis F, et al. Mortality of HIV-1-infected patients in the first year of antiretroviral therapy: comparison between low-income and high-income countries. *Lancet.* Mar 11 2006;367(9513):817-824.
29. Barouch DH. Challenges in the development of an HIV-1 vaccine. *Nature.* Oct 2 2008;455(7213):613-619.
30. Shaw GM, Hunter E. HIV transmission. *Cold Spring Harb Perspect Med.* Nov 2012;2(11).
31. El-Sadr WM, Mayer KH, Hodder SL. AIDS in America--forgotten but not gone. *N Engl J Med.* Mar 18 2010;362(11):967-970.
32. Crepaz N, Marks G. Towards an understanding of sexual risk behavior in people living with HIV: a review of social, psychological, and medical findings. *AIDS.* Jan 25 2002;16(2):135-149.
33. National AIDS Coordinating Agency. Botswana 2012 Global AIDS Response Report. 2012. <http://www.unaids.org/en/dataanalysis/knowyourresponse/countryprogressreports/2012countries/ce_BW_Narrative_Report[1].pdf>
34. Kandala NB, Campbell EK, Rakgoasi SD, Madi-Segwagwe BC, Fako TT. The geography of HIV/AIDS prevalence rates in Botswana. *HIV AIDS (Auckl).* 2012;4:95-102.
35. Briggs JA, Wilk T, Welker R, Krausslich HG, Fuller SD. Structural organization of authentic, mature HIV-1 virions and cores. *EMBO J.* Apr 1 2003;22(7):1707-1715.
36. Meerloo T, Sheikh MA, Bloem AC, et al. Host cell membrane proteins on human immunodeficiency virus type 1 after in vitro infection of H9 cells and blood mononuclear cells. An immuno-electron microscopic study. *J Gen Virol.* Jan 1993;74 (Pt 1):129-135.

37. Rizzuto CD, Sodroski JG. Contribution of virion ICAM-1 to human immunodeficiency virus infectivity and sensitivity to neutralization. *J Virol.* Jun 1997;71(6):4847-4851.
38. Frankel AD, Young JA. HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem.* 1998;67:1-25.
39. Robinson HL. New hope for an AIDS vaccine. *Nat Rev Immunol.* Apr 2002;2(4):239-250.
40. Hearps AC, Jans DA. Regulating the functions of the HIV-1 matrix protein. *AIDS Res Hum Retroviruses.* Mar 2007;23(3):341-346.
41. Arrildt KT, Joseph SB, Swanstrom R. The HIV-1 env protein: a coat of many colors. *Curr HIV/AIDS Rep.* Mar 2012;9(1):52-63.
42. Fackler OT, Baur AS. Live and let die: Nef functions beyond HIV replication. *Immunity.* Apr 2002;16(4):493-497.
43. Marin M, Rose KM, Kozak SL, Kabat D. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nat Med.* Nov 2003;9(11):1398-1403.
44. Neil SJ, Zang T, Bieniasz PD. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature.* Jan 24 2008;451(7177):425-430.
45. Strebel K. HIV accessory proteins versus host restriction factors. *Curr Opin Virol.* Dec 2013;3(6):692-699.
46. Rambaut A, Posada D, Crandall KA, Holmes EC. The causes and consequences of HIV evolution. *Nat Rev Genet.* Jan 2004;5(1):52-61.
47. Coffin J, Swanstrom R. HIV pathogenesis: dynamics and genetics of viral populations and infected cells. *Cold Spring Harb Perspect Med.* Jan 2013;3(1):a012526.
48. Gartner S, Markovits P, Markovitz DM, Kaplan MH, Gallo RC, Popovic M. The role of mononuclear phagocytes in HTLV-III/LAV infection. *Science.* Jul 11 1986;233(4760):215-219.
49. Price RW, Brew B, Sidtis J, Rosenblum M, Scheck AC, Cleary P. The brain in AIDS: central nervous system HIV-1 infection and AIDS dementia complex. *Science.* Feb 5 1988;239(4840):586-592.
50. Watkins BA, Dorn HH, Kelly WB, et al. Specific tropism of HIV-1 for microglial cells in primary human brain cultures. *Science.* Aug 3 1990;249(4968):549-553.
51. Donaghy H, Gazzard B, Gotch F, Patterson S. Dysfunction and infection of freshly isolated blood myeloid and plasmacytoid dendritic cells in patients infected with HIV-1. *Blood.* Jun 1 2003;101(11):4505-4511.
52. Wilen CB, Tilton JC, Doms RW. HIV: cell binding and entry. *Cold Spring Harb Perspect Med.* 2012;2(8).
53. Maddon PJ, Dalgleish AG, McDougal JS, Clapham PR, Weiss RA, Axel R. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell.* Nov 7 1986;47(3):333-348.
54. McDougal JS, Kennedy MS, Sligh JM, Cort SP, Mawle A, Nicholson JK. Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule. *Science.* Jan 24 1986;231(4736):382-385.
55. Wu L, Gerard NP, Wyatt R, et al. CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. *Nature.* Nov 14 1996;384(6605):179-183.
56. Trkola A, Dragic T, Arthos J, et al. CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5. *Nature.* Nov 14 1996;384(6605):184-187.

57. Huang CC, Lam SN, Acharya P, et al. Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. *Science.* Sep 28 2007;317(5846):1930-1934.
58. Reeves JD, Miamidian JL, Biscone MJ, et al. Impact of mutations in the coreceptor binding site on human immunodeficiency virus type 1 fusion, infection, and entry inhibitor sensitivity. *J Virol.* May 2004;78(10):5476-5485.
59. Forsell MN, Dey B, Morner A, et al. B cell recognition of the conserved HIV-1 co-receptor binding site is altered by endogenous primate CD4. *PLoS Pathog.* 2008;4(10):e1000171.
60. Alkhatib G, Combadiere C, Broder CC, et al. CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science.* Jun 28 1996;272(5270):1955-1958.
61. Dragic T, Litwin V, Allaway GP, et al. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature.* Jun 20 1996;381(6584):667-673.
62. Deng H, Liu R, Ellmeier W, et al. Identification of a major co-receptor for primary isolates of HIV-1. *Nature.* Jun 20 1996;381(6584):661-666.
63. Feng Y, Broder CC, Kennedy PE, Berger EA. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science.* May 10 1996;272(5263):872-877.
64. Berson JF, Long D, Doranz BJ, Rucker J, Jirik FR, Doms RW. A seven-transmembrane domain receptor involved in fusion and entry of T-cell-tropic human immunodeficiency virus type 1 strains. *J Virol.* Sep 1996;70(9):6288-6295.
65. Lehmann MJ, Sherer NM, Marks CB, Pypaert M, Mothes W. Actin- and myosin-driven movement of viruses along filopodia precedes their entry into cells. *J Cell Biol.* Jul 18 2005;170(2):317-325.
66. Chan DC, Fass D, Berger JM, Kim PS. Core structure of gp41 from the HIV envelope glycoprotein. *Cell.* Apr 18 1997;89(2):263-273.
67. Weissenhorn W, Dessen A, Harrison SC, Skehel JJ, Wiley DC. Atomic structure of the ectodomain from HIV-1 gp41. *Nature.* May 22 1997;387(6631):426-430.
68. Gamble TR, Vajdos FF, Yoo S, et al. Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell.* Dec 27 1996;87(7):1285-1294.
69. Fassati A, Goff SP. Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1. *J Virol.* Apr 2001;75(8):3626-3635.
70. Hu WS, Hughes SH. HIV-1 reverse transcription. *Cold Spring Harb Perspect Med.* Oct 2012;2(10).
71. Craigie R, Bushman FD. HIV DNA integration. *Cold Spring Harb Perspect Med.* Jul 2012;2(7):a006890.
72. Schwartz S, Felber BK, Benko DM, Fenyo EM, Pavlakis GN. Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. *J Virol.* Jun 1990;64(6):2519-2529.
73. Schwartz S, Felber BK, Fenyo EM, Pavlakis GN. Env and Vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs. *J Virol.* Nov 1990;64(11):5448-5456.
74. Schwartz S, Felber BK, Pavlakis GN. Expression of human immunodeficiency virus type 1 vif and vpr mRNAs is Rev-dependent and regulated by splicing. *Virology.* Aug 1991;183(2):677-686.

75. Wu Y. HIV-1 gene expression: lessons from provirus and non-integrated DNA. *Retrovirology.* 2004;1:13.
76. Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature.* Jan 21 1988;331(6153):280-283.
77. Farabaugh PJ. Programmed translational frameshifting. *Microbiol Rev.* Mar 1996;60(1):103-134.
78. Stein BS, Engleman EG. Intracellular processing of the gp160 HIV-1 envelope precursor. Endoproteolytic cleavage occurs in a cis or medial compartment of the Golgi complex. *J Biol Chem.* Feb 15 1990;265(5):2640-2649.
79. Sundquist WI, Krausslich HG. HIV-1 assembly, budding, and maturation. *Cold Spring Harb Perspect Med.* Jul 2012;2(7):a006924.
80. Haase AT. Targeting early infection to prevent HIV-1 mucosal transmission. *Nature.* Mar 11 2010;464(7286):217-223.
81. Fiebig EW, Wright DJ, Rawal BD, et al. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. *AIDS.* Sep 5 2003;17(13):1871-1879.
82. Jung AC, Paauw DS. Diagnosing HIV-related disease: using the CD4 count as a guide. *J Gen Intern Med.* Feb 1998;13(2):131-136.
83. Park RA. European AIDS definition. *Lancet.* Mar 14 1992;339(8794):671.
84. Ancelle-Park R. Expanded European AIDS case definition. *Lancet.* Feb 13 1993;341(8842):441.
85. Revision of the CDC surveillance case definition for acquired immunodeficiency syndrome. Council of State and Territorial Epidemiologists; AIDS Program, Center for Infectious Diseases. *MMWR Morb Mortal Wkly Rep.* Aug 14 1987;36 Suppl 1:1S-15S.
86. Smyth RP, Davenport MP, Mak J. The origin of genetic diversity in HIV-1. *Virus Res.* Nov 2012;169(2):415-429.
87. Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature.* Jan 12 1995;373(6510):123-126.
88. Roberts JD, Bebenek K, Kunkel TA. The accuracy of reverse transcriptase from HIV-1. *Science.* Nov 25 1988;242(4882):1171-1173.
89. Hu WS, Temin HM. Retroviral recombination and reverse transcription. *Science.* Nov 30 1990;250(4985):1227-1233.
90. Mansky LM, Temin HM. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol.* Aug 1995;69(8):5087-5094.
91. Plantier JC, Leoz M, Dickerson JE, et al. A new human immunodeficiency virus derived from gorillas. *Nat Med.* Aug 2009;15(8):871-872.
92. Vallari A, Holzmayer V, Harris B, et al. Confirmation of putative HIV-1 group P in Cameroon. *J Virol.* Feb 2011;85(3):1403-1407.
93. Vallari A, Bodelle P, Ngansop C, et al. Four new HIV-1 group N isolates from Cameroon: Prevalence continues to be low. *AIDS Res Hum Retroviruses.* Jan 2010;26(1):109-115.
94. Mauclere P, Loussert-Ajaka I, Damond F, et al. Serological and virological characterization of HIV-1 group O infection in Cameroon. *AIDS.* Mar 15 1997;11(4):445-453.
95. Peeters M, Gueye A, Mboup S, et al. Geographical distribution of HIV-1 group O viruses in Africa. *AIDS.* Mar 15 1997;11(4):493-498.
96. Arien KK, Vanham G, Arts EJ. Is HIV-1 evolving to a less virulent form in humans? *Nat Rev Microbiol.* Feb 2007;5(2):141-151.

97. Robertson DL, Anderson JP, Bradac JA, et al. HIV-1 nomenclature proposal. *Science.* Apr 7 2000;288(5463):55-56.
98. Korber B, Gaschen B, Yusim K, Thakallapally R, Kesmir C, Detours V. Evolutionary and immunological implications of contemporary HIV-1 variation. *Br Med Bull.* 2001;58:19-42.
99. Hemelaar J. The origin and diversity of the HIV-1 pandemic. *Trends Mol Med.* Mar 2012;18(3):182-192.
100. Essex M. Human immunodeficiency viruses in the developing world. *Adv Virus Res.* 1999;53:71-88.
101. Hemelaar J, Gouws E, Ghys PD, Osmanov S. Global trends in molecular epidemiology of HIV-1 during 2000-2007. *AIDS.* Mar 13 2011;25(5):679-689.
102. Gilbert MT, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, Worobey M. The emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci U S A.* Nov 20 2007;104(47):18566-18570.
103. Taylor BS, Sobieszczyk ME, McCutchan FE, Hammer SM. The challenge of HIV-1 subtype diversity. *N Engl J Med.* Apr 10 2008;358(15):1590-1602.
104. Renjifo B, Gilbert P, Chaplin B, et al. Preferential in-utero transmission of HIV-1 subtype C as compared to HIV-1 subtype A or D. *AIDS.* Aug 20 2004;18(12):1629-1636.
105. John-Stewart GC, Nduati RW, Rousseau CM, et al. Subtype C Is associated with increased vaginal shedding of HIV-1. *J Infect Dis.* Aug 1 2005;192(3):492-496.
106. Kanki PJ, Hamel DJ, Sankale JL, et al. Human immunodeficiency virus type 1 subtypes differ in disease progression. *J Infect Dis.* Jan 1999;179(1):68-73.
107. Alaeus A, Lidman K, Bjorkman A, Giesecke J, Albert J. Similar rate of disease progression among individuals infected with HIV-1 genetic subtypes A-D. *AIDS.* May 28 1999;13(8):901-907.
108. Vasan A, Renjifo B, Hertzmark E, et al. Different rates of disease progression of HIV type 1 infection in Tanzania based on infecting subtype. *Clin Infect Dis.* Mar 15 2006;42(6):843-852.
109. Baeten JM, Chohan B, Lavreys L, et al. HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads. *J Infect Dis.* Apr 15 2007;195(8):1177-1180.
110. Kiwanuka N, Robb M, Laeyendecker O, et al. HIV-1 viral subtype differences in the rate of CD4+ T-cell decline among HIV seroincident antiretroviral naive persons in Rakai district, Uganda. *J Acquir Immune Defic Syndr.* Jun 2010;54(2):180-184.
111. Easterbrook PJ, Smith M, Mullen J, et al. Impact of HIV-1 viral subtype on disease progression and response to antiretroviral therapy. *J Int AIDS Soc.* 2010;13:4.
112. Kaleebu P, Nankya IL, Yirrell DL, et al. Relation between chemokine receptor use, disease stage, and HIV-1 subtypes A and D: results from a rural Ugandan cohort. *J Acquir Immune Defic Syndr.* May 1 2007;45(1):28-33.
113. Ng OT, Lin L, Laeyendecker O, et al. Increased rate of CD4+ T-cell decline and faster time to antiretroviral therapy in HIV-1 subtype CRF01_AE infected seroconverters in Singapore. *PLoS ONE.* 2011;6(1):e15738.
114. Costello C, Nelson KE, Suriyanon V, et al. HIV-1 subtype E progression among northern Thai couples: traditional and non-traditional predictors of survival. *Int J Epidemiol.* Jun 2005;34(3):577-584.
115. Hudgens MG, Longini IM, Jr., Vanichseni S, et al. Subtype-specific transmission probabilities for human immunodeficiency virus type 1 among

injecting drug users in Bangkok, Thailand. *Am J Epidemiol.* Jan 15 2002;155(2):159-168.

116. Laurent C, Bourgeois A, Faye MA, et al. No difference in clinical progression between patients infected with the predominant human immunodeficiency virus type 1 circulating recombinant form (CRF) 02_AG strain and patients not infected with CRF02_AG, in Western and West-Central Africa: a four-year prospective multicenter study. *J Infect Dis.* Aug 15 2002;186(4):486-492.

117. Palm AA, Esbjornsson J, Mansson F, et al. Faster Progression to AIDS and AIDS-Related Death Among Seroincident Individuals Infected With Recombinant HIV-1 A3/CRF02_AG Compared With Sub-subtype A3. *J Infect Dis.* Aug 30 2013.

118. Berger EA, Doms RW, Fenyo EM, et al. A new classification for HIV-1. *Nature.* Jan 15 1998;391(6664):240.

119. Koyanagi Y, Miles S, Mitsuyasu RT, Merrill JE, Vinters HV, Chen IS. Dual infection of the central nervous system by AIDS viruses with distinct cellular tropisms. *Science.* May 15 1987;236(4803):819-822.

120. Gendelman HE, Orenstein JM, Martin MA, et al. Efficient isolation and propagation of human immunodeficiency virus on recombinant colony-stimulating factor 1-treated monocytes. *J Exp Med.* Apr 1 1988;167(4):1428-1441.

121. Gendelman HE, Baca LM, Husayni H, et al. Macrophage-HIV interaction: viral isolation and target cell tropism. *AIDS.* Mar 1990;4(3):221-228.

122. Chesebro B, Nishio J, Perryman S, et al. Identification of human immunodeficiency virus envelope gene sequences influencing viral entry into CD4-positive HeLa cells, T-leukemia cells, and macrophages. *J Virol.* Nov 1991;65(11):5782-5789.

123. Fenyo EM, Albert J, Asjo B. Replicative capacity, cytopathic effect and cell tropism of HIV. *AIDS.* 1989;3 Suppl 1:S5-12.

124. Koot M, Vos AH, Keet RP, et al. HIV-1 biological phenotype in long-term infected individuals evaluated with an MT-2 cocultivation assay. *AIDS.* Jan 1992;6(1):49-54.

125. Dalgleish AG, Beverley PC, Clapham PR, Crawford DH, Greaves MF, Weiss RA. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature.* Dec 20-1985 Jan 2 1984;312(5996):763-767.

126. Verhofstede C, Nijhuis M, Vandekerckhove L. Correlation of coreceptor usage and disease progression. *Curr Opin HIV AIDS.* Sep 2012;7(5):432-439.

127. Doranz BJ, Rucker J, Yi Y, et al. A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors. *Cell.* Jun 28 1996;85(7):1149-1158.

128. Choe H, Farzan M, Sun Y, et al. The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell.* Jun 28 1996;85(7):1135-1148.

129. Yost R, Pasquale TR, Sahloff EG. Maraviroc: a coreceptor CCR5 antagonist for management of HIV infection. *Am J Health Syst Pharm.* Apr 15 2009;66(8):715-726.

130. Deng HK, Unutmaz D, KewalRamani VN, Littman DR. Expression cloning of new receptors used by simian and human immunodeficiency viruses. *Nature.* Jul 17 1997;388(6639):296-300.

131. Connor RI, Sheridan KE, Ceradini D, Choe S, Landau NR. Change in coreceptor use correlates with disease progression in HIV-1--infected individuals. *J Exp Med.* Feb 17 1997;185(4):621-628.

132. Zhang YJ, Dragic T, Cao Y, et al. Use of coreceptors other than CCR5 by non-syncytium-inducing adult and pediatric isolates of human immunodeficiency virus type 1 is rare in vitro. *J Virol.* Nov 1998;72(11):9337-9344.
133. Zerhouni B, Nelson JA, Saha K. Isolation of CD4-independent primary human immunodeficiency virus type 1 isolates that are syncytium inducing and acutely cytopathic for CD8+ lymphocytes. *J Virol.* Feb 2004;78(3):1243-1255.
134. Samson M, Labbe O, Mollereau C, Vassart G, Parmentier M. Molecular cloning and functional expression of a new human CC-chemokine receptor gene. *Biochemistry.* Mar 19 1996;35(11):3362-3367.
135. Raport CJ, Gosling J, Schweickart VL, Gray PW, Charo IF. Molecular cloning and functional characterization of a novel human CC chemokine receptor (CCR5) for RANTES, MIP-1beta, and MIP-1alpha. *J Biol Chem.* Jul 19 1996;271(29):17161-17166.
136. Cook DN. The role of MIP-1 alpha in inflammation and hematopoiesis. *J Leukoc Biol.* Jan 1996;59(1):61-66.
137. Schall TJ, Bacon K, Camp RD, Kaspari JW, Goeddel DV. Human macrophage inflammatory protein alpha (MIP-1 alpha) and MIP-1 beta chemokines attract distinct populations of lymphocytes. *J Exp Med.* Jun 1 1993;177(6):1821-1826.
138. Cocchi F, DeVico AL, Garzino-Demo A, Arya SK, Gallo RC, Lusso P. Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells. *Science.* Dec 15 1995;270(5243):1811-1815.
139. Mackewicz CE, Barker E, Greco G, Reyes-Teran G, Levy JA. Do beta-chemokines have clinical relevance in HIV infection? *J Clin Invest.* Aug 15 1997;100(4):921-930.
140. van Rij RP, Blaak H, Visser JA, et al. Differential coreceptor expression allows for independent evolution of non-syncytium-inducing and syncytium-inducing HIV-1. *J Clin Invest.* Oct 2000;106(8):1039-1052.
141. Bleul CC, Wu L, Hoxie JA, Springer TA, Mackay CR. The HIV coreceptors CXCR4 and CCR5 are differentially expressed and regulated on human T lymphocytes. *Proc Natl Acad Sci U S A.* Mar 4 1997;94(5):1925-1930.
142. Rubbert A, Combadiere C, Ostrowski M, et al. Dendritic cells express multiple chemokine receptors used as coreceptors for HIV entry. *J Immunol.* Apr 15 1998;160(8):3933-3941.
143. Zaitseva M, Blauvelt A, Lee S, et al. Expression and function of CCR5 and CXCR4 on human Langerhans cells and macrophages: implications for HIV primary infection. *Nat Med.* Dec 1997;3(12):1369-1375.
144. He J, Chen Y, Farzan M, et al. CCR3 and CCR5 are co-receptors for HIV-1 infection of microglia. *Nature.* Feb 13 1997;385(6617):645-649.
145. Valentin A, Rosati M, Patenaude DJ, et al. Persistent HIV-1 infection of natural killer cells in patients receiving highly active antiretroviral therapy. *Proc Natl Acad Sci U S A.* May 14 2002;99(10):7015-7020.
146. Agace WW, Roberts AI, Wu L, Greineder C, Ebert EC, Parker CM. Human intestinal lamina propria and intraepithelial lymphocytes express receptors specific for chemokines induced by inflammation. *Eur J Immunol.* Mar 2000;30(3):819-826.
147. Anton PA, Elliott J, Poles MA, et al. Enhanced levels of functional HIV-1 co-receptors on human mucosal T cells demonstrated using intestinal biopsy tissue. *AIDS.* Aug 18 2000;14(12):1761-1765.
148. Oppermann M. Chemokine receptor CCR5: insights into structure, function, and regulation. *Cell Signal.* Nov 2004;16(11):1201-1210.

149. Farzan M, Mirzabekov T, Kolchinsky P, et al. Tyrosine sulfation of the amino terminus of CCR5 facilitates HIV-1 entry. *Cell.* Mar 5 1999;96(5):667-676.
150. Cormier EG, Dragic T. The crown and stem of the V3 loop play distinct roles in human immunodeficiency virus type 1 envelope glycoprotein interactions with the CCR5 coreceptor. *J Virol.* Sep 2002;76(17):8953-8957.
151. Scarlatti G, Tresoldi E, Bjorndal A, et al. In vivo evolution of HIV-1 co-receptor usage and sensitivity to chemokine-mediated suppression. *Nat Med.* Nov 1997;3(11):1259-1265.
152. Casper C, Naver L, Clevestig P, et al. Coreceptor change appears after immune deficiency is established in children infected with different HIV-1 subtypes. *AIDS Res Hum Retroviruses.* Mar 20 2002;18(5):343-352.
153. Swanstrom R, Coffin J. HIV-1 pathogenesis: the virus. *Cold Spring Harb Perspect Med.* Dec 2012;2(12):a007443.
154. Isaacman-Beck J, Hermann EA, Yi Y, et al. Heterosexual transmission of human immunodeficiency virus type 1 subtype C: Macrophage tropism, alternative coreceptor use, and the molecular anatomy of CCR5 utilization. *J Virol.* Aug 2009;83(16):8208-8220.
155. Salazar-Gonzalez JF, Salazar MG, Keele BF, et al. Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *J Exp Med.* Jun 8 2009;206(6):1273-1289.
156. Alexander M, Lynch R, Mulenga J, Allen S, Derdeyn CA, Hunter E. Donor and recipient envs from heterosexual human immunodeficiency virus subtype C transmission pairs require high receptor levels for entry. *J Virol.* Apr 2010;84(8):4100-4104.
157. Ping LH, Joseph SB, Anderson JA, et al. Comparison of Viral Env Proteins from Acute and Chronic Infections with Subtype C Human Immunodeficiency Virus Type 1 Identifies Differences in Glycosylation and CCR5 Utilization and Suggests a New Strategy for Immunogen Design. *J Virol.* Jul 2013;87(13):7218-7233.
158. Schnell G, Joseph S, Spudich S, Price RW, Swanstrom R. HIV-1 replication in the central nervous system occurs in two distinct cell types. *PLoS Pathog.* Oct 2011;7(10):e1002286.
159. Spudich S, Gonzalez-Scarano F. HIV-1-related central nervous system disease: current issues in pathogenesis, diagnosis, and treatment. *Cold Spring Harb Perspect Med.* Jun 2012;2(6):a007120.
160. Liu R, Paxton WA, Choe S, et al. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell.* Aug 9 1996;86(3):367-377.
161. Dean M, Carrington M, Winkler C, et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science.* Sep 27 1996;273(5283):1856-1862.
162. Samson M, Libert F, Doranz BJ, et al. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature.* Aug 22 1996;382(6593):722-725.
163. Huang Y, Paxton WA, Wolinsky SM, et al. The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. *Nat Med.* Nov 1996;2(11):1240-1243.
164. Balotta C, Bagnarelli P, Violin M, et al. Homozygous delta 32 deletion of the CCR-5 chemokine receptor gene in an HIV-1-infected patient. *AIDS.* Aug 1997;11(10):F67-71.

165. O'Brien TR, Winkler C, Dean M, et al. HIV-1 infection in a man homozygous for CCR5 delta 32. *Lancet.* Apr 26 1997;349(9060):1219.
166. Theodorou I, Meyer L, Magierowska M, Katlama C, Rouzioux C. HIV-1 infection in an individual homozygous for CCR5 delta 32. Seroco Study Group. *Lancet.* Apr 26 1997;349(9060):1219-1220.
167. Glass WG, McDermott DH, Lim JK, et al. CCR5 deficiency increases risk of symptomatic West Nile virus infection. *J Exp Med.* Jan 23 2006;203(1):35-40.
168. Bracci PM, Skibola CF, Conde L, et al. Chemokine polymorphisms and lymphoma: a pooled analysis. *Leuk Lymphoma.* Mar 2010;51(3):497-506.
169. Bleul CC, Farzan M, Choe H, et al. The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and blocks HIV-1 entry. *Nature.* Aug 29 1996;382(6594):829-833.
170. Oberlin E, Amara A, Bachelerie F, et al. The CXC chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1. *Nature.* Aug 29 1996;382(6594):833-835.
171. Nagasawa T, Kikutani H, Kishimoto T. Molecular cloning and structure of a pre-B-cell growth-stimulating factor. *Proc Natl Acad Sci U S A.* Mar 15 1994;91(6):2305-2309.
172. Bleul CC, Fuhlbrigge RC, Casasnovas JM, Aiuti A, Springer TA. A highly efficacious lymphocyte chemoattractant, stromal cell-derived factor 1 (SDF-1). *J Exp Med.* Sep 1 1996;184(3):1101-1109.
173. Ostrowski MA, Chun TW, Justement SJ, et al. Both memory and CD45RA+/CD62L+ naive CD4(+) T cells are infected in human immunodeficiency virus type 1-infected individuals. *J Virol.* Aug 1999;73(8):6430-6435.
174. Nicholson JK, Browning SW, Hengel RL, et al. CCR5 and CXCR4 expression on memory and naive T cells in HIV-1 infection and response to highly active antiretroviral therapy. *J Acquir Immune Defic Syndr.* Jun 1 2001;27(2):105-115.
175. Tchou I, Misery L, Sabido O, et al. Functional HIV CXCR4 coreceptor on human epithelial Langerhans cells and infection by HIV strain X4. *J Leukoc Biol.* Aug 2001;70(2):313-321.
176. Valentin A, Trivedi H, Lu W, Kostrikis LG, Pavlakis GN. CXCR4 mediates entry and productive infection of syncytia-inducing (X4) HIV-1 strains in primary macrophages. *Virology.* Apr 10 2000;269(2):294-304.
177. Di Marzio P, Tse J, Landau NR. Chemokine receptor regulation and HIV type 1 tropism in monocyte-macrophages. *AIDS Res Hum Retroviruses.* Jan 20 1998;14(2):129-138.
178. Verani A, Gras G, Pancino G. Macrophages and HIV-1: dangerous liaisons. *Mol Immunol.* Feb 2005;42(2):195-212.
179. Tan Q, Zhu Y, Li J, et al. Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. *Science.* Sep 20 2013;341(6152):1387-1390.
180. Wu B, Chien EY, Mol CD, et al. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science.* Nov 19 2010;330(6007):1066-1071.
181. Brelot A, Heveker N, Montes M, Alizon M. Identification of residues of CXCR4 critical for human immunodeficiency virus coreceptor and chemokine receptor activities. *J Biol Chem.* Aug 4 2000;275(31):23736-23744.
182. Chabot DJ, Chen H, Dimitrov DS, Broder CC. N-linked glycosylation of CXCR4 masks coreceptor function for CCR5-dependent human immunodeficiency virus type 1 isolates. *J Virol.* May 2000;74(9):4404-4413.
183. Schuitemaker H, van 't Wout AB, Lusso P. Clinical significance of HIV-1 coreceptor usage. *J Transl Med.* 2011;9 Suppl 1:S5.

184. Bunnik EM, Quakkelaar ED, van Nuenen AC, Boeser-Nunnink B, Schuitemaker H. Increased neutralization sensitivity of recently emerged CXCR4-using human immunodeficiency virus type 1 strains compared to coexisting CCR5-using variants from the same patient. *J Virol.* Jan 2007;81(2):525-531.

185. Bunnik EM, Swenson LC, Edo-Matas D, et al. Detection of inferred CCR5- and CXCR4-using HIV-1 variants and evolutionary intermediates using ultra-deep pyrosequencing. *PLoS Pathog.* Jun 2011;7(6):e1002106.

186. Huang W, Eshleman SH, Toma J, et al. Coreceptor tropism in human immunodeficiency virus type 1 subtype D: high prevalence of CXCR4 tropism and heterogeneous composition of viral populations. *J Virol.* Aug 2007;81(15):7885-7893.

187. Irlbeck DM, Amrine-Madsen H, Kitrinos KM, Labranche CC, Demarest JF. Chemokine (C-C motif) receptor 5-using envelopes predominate in dual/mixed-tropic HIV from the plasma of drug-naive individuals. *AIDS.* Jul 31 2008;22(12):1425-1431.

188. Rucker J, Edinger AL, Sharron M, et al. Utilization of chemokine receptors, orphan receptors, and herpesvirus-encoded receptors by diverse human and simian immunodeficiency viruses. *J Virol.* Dec 1997;71(12):8999-9007.

189. Hwang SS, Boyle TJ, Lyerly HK, Cullen BR. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science.* Jul 5 1991;253(5015):71-74.

190. De Jong JJ, De Ronde A, Keulen W, Tersmette M, Goudsmit J. Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution. *J Virol.* Nov 1992;66(11):6777-6780.

191. de Jong JJ, Goudsmit J, Keulen W, et al. Human immunodeficiency virus type 1 clones chimeric for the envelope V3 domain differ in syncytium formation and replication capacity. *J Virol.* Feb 1992;66(2):757-765.

192. Fouchier RA, Groenink M, Kootstra NA, et al. Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. *J Virol.* May 1992;66(5):3183-3187.

193. Shioda T, Levy JA, Cheng-Mayer C. Small amino acid changes in the V3 hypervariable region of gp120 can affect the T-cell-line and macrophage tropism of human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A.* Oct 15 1992;89(20):9434-9438.

194. Chesebro B, Wehrly K, Nishio J, Perryman S. Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T-cell-tropic isolates: definition of critical amino acids involved in cell tropism. *J Virol.* Nov 1992;66(11):6547-6554.

195. Catasti P, Fontenot JD, Bradbury EM, Gupta G. Local and global structural properties of the HIV-MN V3 loop. *J Biol Chem.* Feb 3 1995;270(5):2224-2232.

196. Rini JM, Stanfield RL, Stura EA, Salinas PA, Profy AT, Wilson IA. Crystal structure of a human immunodeficiency virus type 1 neutralizing antibody, 50.1, in complex with its V3 loop peptide antigen. *Proc Natl Acad Sci U S A.* Jul 1 1993;90(13):6325-6329.

197. Tugarinov V, Zvi A, Levy R, Hayek Y, Matsushita S, Anglister J. NMR structure of an anti-gp120 antibody complex with a V3 peptide reveals a surface important for co-receptor binding. *Structure.* Apr 15 2000;8(4):385-395.

198. Huang CC, Tang M, Zhang MY, et al. Structure of a V3-containing HIV-1 gp120 core. *Science.* Nov 11 2005;310(5750):1025-1028.
199. Labrosse B, Treboute C, Brelot A, Alizon M. Cooperation of the V1/V2 and V3 domains of human immunodeficiency virus type 1 gp120 for interaction with the CXCR4 receptor. *J Virol.* Jun 2001;75(12):5457-5464.
200. Rizzuto CD, Wyatt R, Hernandez-Ramos N, et al. A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science.* Jun 19 1998;280(5371):1949-1953.
201. Cardozo T, Kimura T, Philpott S, Weiser B, Burger H, Zolla-Pazner S. Structural basis for coreceptor selectivity by the HIV type 1 V3 loop. *AIDS Res Hum Retroviruses.* Mar 2007;23(3):415-426.
202. De Wolf F, Hogervorst E, Goudsmit J, et al. Syncytium-inducing and non-syncytium-inducing capacity of human immunodeficiency virus type 1 subtypes other than B: phenotypic and genotypic characteristics. WHO Network for HIV Isolation and Characterization. *AIDS Res Hum Retroviruses.* Nov 1994;10(11):1387-1400.
203. Verrier F, Borman AM, Brand D, Girard M. Role of the HIV type 1 glycoprotein 120 V3 loop in determining coreceptor usage. *AIDS Res Hum Retroviruses.* May 20 1999;15(8):731-743.
204. Hu QX, Barry AP, Wang ZX, Connolly SM, Peiper SC, Greenberg ML. Evolution of the human immunodeficiency virus type 1 envelope during infection reveals molecular corollaries of specificity for coreceptor utilization and AIDS pathogenesis. *J Virol.* Dec 2000;74(24):11858-11872.
205. Milich L, Margolin B, Swanstrom R. V3 loop of the human immunodeficiency virus type 1 Env protein: interpreting sequence variability. *J Virol.* Sep 1993;67(9):5623-5634.
206. Milich L, Margolin BH, Swanstrom R. Patterns of amino acid variability in NSI-like and SI-like V3 sequences and a linked change in the CD4-binding domain of the HIV-1 Env protein. *Virology.* Dec 8 1997;239(1):108-118.
207. Ping LH, Nelson JA, Hoffman IF, et al. Characterization of V3 sequence heterogeneity in subtype C human immunodeficiency virus type 1 isolates from Malawi: underrepresentation of X4 variants. *J Virol.* Aug 1999;73(8):6271-6281.
208. Pollakis G, Kang S, Kliphuis A, Chalaby MI, Goudsmit J, Paxton WA. N-linked glycosylation of the HIV type-1 gp120 envelope glycoprotein as a major determinant of CCR5 and CXCR4 coreceptor utilization. *J Biol Chem.* Apr 20 2001;276(16):13433-13441.
209. Polzer S, Dittmar MT, Schmitz H, et al. Loss of N-linked glycans in the V3-loop region of gp120 is correlated to an enhanced infectivity of HIV-1. *Glycobiology.* Jan 2001;11(1):11-19.
210. Polzer S, Dittmar MT, Schmitz H, Schreiber M. The N-linked Glycan g15 within the V3 Loop of the HIV-1 External Glycoprotein gp120 Affects Coreceptor Usage, Cellular Tropism, and Neutralization. *Virology.* 2002;304(1):70-80.
211. Koch M, Pancera M, Kwong PD, et al. Structure-based, targeted deglycosylation of HIV-1 gp120 and effects on neutralization sensitivity and antibody recognition. *Virology.* Sep 1 2003;313(2):387-400.
212. Ogert RA, Lee MK, Ross W, Buckler-White A, Martin MA, Cho MW. N-linked glycosylation sites adjacent to and within the V1/V2 and the V3 loops of dualtropic human immunodeficiency virus type 1 isolate DH12 gp120 affect coreceptor usage and cellular tropism. *J Virol.* Jul 2001;75(13):5998-6006.

213. Biscone MJ, Miamidian JL, Muchiri JM, et al. Functional impact of HIV coreceptor-binding site mutations. *Virology.* Jul 20 2006;351(1):226-236.
214. Jansson M, Backstrom E, Scarlatti G, et al. Length variation of glycoprotein 120 V2 region in relation to biological phenotypes and coreceptor usage of primary HIV type 1 isolates. *AIDS Res Hum Retroviruses.* Oct 10 2001;17(15):1405-1414.
215. Nabatov AA, Pollakis G, Linnemann T, Kliphius A, Chalaby MI, Paxton WA. Intrapatient alterations in the human immunodeficiency virus type 1 gp120 V1V2 and V3 regions differentially modulate coreceptor usage, virus inhibition by CC/CXC chemokines, soluble CD4, and the b12 and 2G12 monoclonal antibodies. *J Virol.* Jan 2004;78(1):524-530.
216. Pastore C, Nedellec R, Ramos A, Pontow S, Ratner L, Mosier DE. Human immunodeficiency virus type 1 coreceptor switching: V1/V2 gain-of-fitness mutations compensate for V3 loss-of-fitness mutations. *J Virol.* Jan 2006;80(2):750-758.
217. Huang W, Toma J, Fransen S, et al. Coreceptor tropism can be influenced by amino acid substitutions in the gp41 transmembrane subunit of human immunodeficiency virus type 1 envelope protein. *J Virol.* Jun 2008;82(11):5584-5593.
218. Cecilia D, KewalRamani VN, O'Leary J, et al. Neutralization profiles of primary human immunodeficiency virus type 1 isolates in the context of coreceptor usage. *J Virol.* Sep 1998;72(9):6988-6996.
219. Keele BF, Giorgi EE, Salazar-Gonzalez JF, et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci U S A.* May 27 2008;105(21):7552-7557.
220. Parrish NF, Gao F, Li H, et al. Phenotypic properties of transmitted founder HIV-1. *Proc Natl Acad Sci U S A.* Apr 23 2013;110(17):6626-6633.
221. Whitcomb JM, Huang W, Fransen S, et al. Development and characterization of a novel single-cycle recombinant-virus assay to determine human immunodeficiency virus type 1 coreceptor tropism. *Antimicrob Agents Chemother.* Feb 2007;51(2):566-575.
222. Su Z, Gulick RM, Krambrink A, et al. Response to vicriviroc in treatment-experienced subjects, as determined by an enhanced-sensitivity coreceptor tropism assay: reanalysis of AIDS clinical trials group A5211. *J Infect Dis.* Dec 1 2009;200(11):1724-1728.
223. Reeves JD, Coakley E, Petropoulos CJ, Whitcomb J. An enhanced sensitivity Trofile HIV coreceptor tropism assay for selecting patients for therapy with entry inhibitors targeting CCR5: a review of analytical and clinical studies. *J Viral Entry.* 2009;3:94-102.
224. Koot M, van Leeuwen R, de Goede RE, et al. Conversion rate towards a syncytium-inducing (SI) phenotype during different stages of human immunodeficiency virus type 1 infection and prognostic value of SI phenotype for survival after AIDS diagnosis. *J Infect Dis.* Jan 1999;179(1):254-258.
225. Shepherd JC, Jacobson LP, Qiao W, et al. Emergence and persistence of CXCR4-tropic HIV-1 in a population of men from the multicenter AIDS cohort study. *J Infect Dis.* Oct 15 2008;198(8):1104-1112.
226. Wambui V, Kiptoo M, Kinyua J, et al. Predicted HIV-1 coreceptor usage among Kenya patients shows a high tendency for subtype d to be cxcr4 tropic. *AIDS Res Ther.* 2012;9(1):22.
227. Yu XF, Wang Z, Beyrer C, et al. Phenotypic and genotypic characteristics of human immunodeficiency virus type 1 from patients with AIDS in northern Thailand. *J Virol.* Aug 1995;69(8):4649-4655.

228. Utaipat U, Duerr A, Rudolph DL, et al. Coreceptor utilization of HIV type 1 subtype E viral isolates from Thai men with HIV type 1-infected and uninfected wives. *AIDS Res Hum Retroviruses.* Jan 1 2002;18(1):1-11.
229. To SW, Chen JH, Wong KH, Chan KC, Chen Z, Yam WC. Determination of the High Prevalence of Dual/Mixed- or X4-Tropism Among HIV Type 1 CRF01_AE in Hong Kong by Genotyping and Phenotyping Methods. *AIDS Res Hum Retroviruses.* May 31 2013.
230. Tscherning C, Alaeus A, Fredriksson R, et al. Differences in chemokine coreceptor usage between genetic subtypes of HIV-1. *Virology.* Feb 15 1998;241(2):181-188.
231. Abebe A, Demissie D, Goudsmit J, et al. HIV-1 subtype C syncytium- and non-syncytium-inducing phenotypes and coreceptor usage among Ethiopian patients with AIDS. *AIDS.* Jul 30 1999;13(11):1305-1311.
232. Bjorndal A, Sonnerborg A, Tscherning C, Albert J, Fenyo EM. Phenotypic characteristics of human immunodeficiency virus type 1 subtype C isolates of Ethiopian AIDS patients. *AIDS Res Hum Retroviruses.* May 1 1999;15(7):647-653.
233. Coetzer M, Nedellec R, Cilliers T, Meyers T, Morris L, Mosier DE. Extreme genetic divergence is required for coreceptor switching in HIV-1 subtype C. *J Acquir Immune Defic Syndr.* Jan 1 2011;56(1):9-15.
234. Johnston ER, Zijenah LS, Mutetwa S, Kantor R, Kittinunvorakoon C, Katzenstein DA. High Frequency of Syncytium-Inducing and CXCR4-Tropic Viruses among Human Immunodeficiency Virus Type 1 Subtype C-Infected Patients Receiving Antiretroviral Treatment. *Journal of Virology.* 2003;77(13):7682-7688.
235. Connell BJ, Michler K, Capovilla A, Venter WD, Stevens WS, Papathanasopoulos MA. Emergence of X4 usage among HIV-1 subtype C: evidence for an evolving epidemic in South Africa. *AIDS.* Apr 23 2008;22(7):896-899.
236. Singh A, Sunpath H, Green TN, et al. Drug resistance and viral tropism in HIV-1 subtype C-infected patients in KwaZulu-Natal, South Africa: implications for future treatment options. *J Acquir Immune Defic Syndr.* Nov 1 2011;58(3):233-240.
237. Green TN, Archary M, Gordon ML, et al. Drug resistance and coreceptor usage in HIV type 1 subtype C-infected children initiating or failing highly active antiretroviral therapy in South Africa. *AIDS Res Hum Retroviruses.* Apr 2012;28(4):324-332.
238. Engelman A, Cherepanov P. The structural biology of HIV-1: mechanistic and therapeutic insights. *Nat Rev Microbiol.* Apr 2012;10(4):279-290.
239. Arts EJ, Hazuda DJ. HIV-1 antiretroviral drug therapy. *Cold Spring Harb Perspect Med.* Apr 2012;2(4):a007161.
240. Sarafianos SG, Das K, Clark AD, Jr., et al. Lamivudine (3TC) resistance in HIV-1 reverse transcriptase involves steric hindrance with beta-branched amino acids. *Proc Natl Acad Sci U S A.* Aug 31 1999;96(18):10027-10032.
241. Bacheler LT, Anton ED, Kudish P, et al. Human immunodeficiency virus type 1 mutations selected in patients failing efavirenz combination therapy. *Antimicrob Agents Chemother.* Sep 2000;44(9):2475-2484.
242. Molla A, Korneyeva M, Gao Q, et al. Ordered accumulation of mutations in HIV protease confers resistance to ritonavir. *Nat Med.* Jul 1996;2(7):760-766.
243. Maeda K, Das D, Nakata H, Mitsuya H. CCR5 inhibitors: emergence, success, and challenges. *Expert Opin Emerg Drugs.* Jun 2012;17(2):135-145.

244. Hendrix CW, Collier AC, Lederman MM, et al. Safety, pharmacokinetics, and antiviral activity of AMD3100, a selective CXCR4 receptor inhibitor, in HIV-1 infection. *J Acquir Immune Defic Syndr.* Oct 1 2004;37(2):1253-1262.

245. Fransen S, Bridger G, Whitcomb JM, et al. Suppression of dualtropic human immunodeficiency virus type 1 by the CXCR4 antagonist AMD3100 is associated with efficiency of CXCR4 use and baseline virus composition. *Antimicrob Agents Chemother.* Jul 2008;52(7):2608-2615.

246. Moyle G, DeJesus E, Boffito M, et al. Proof of activity with AMD11070, an orally bioavailable inhibitor of CXCR4-tropic HIV type 1. *Clin Infect Dis.* Mar 15 2009;48(6):798-805.

247. FDA approves maraviroc tablets. *AIDS Patient Care STDS.* Sep 2007;21(9):702.

248. Caseiro MM, Nelson M, Diaz RS, et al. Vicriviroc plus optimized background therapy for treatment-experienced subjects with CCR5 HIV-1 infection: final results of two randomized phase III trials. *J Infect.* Oct 2012;65(4):326-335.

249. Nichols WG, Steel HM, Bonny T, et al. Hepatotoxicity observed in clinical trials of aplaviroc (GW873140). *Antimicrob Agents Chemother.* Mar 2008;52(3):858-865.

250. Klibanov OM, Williams SH, Iler CA. Cenicriviroc, an orally active CCR5 antagonist for the potential treatment of HIV infection. *Curr Opin Investig Drugs.* Aug 2010;11(8):940-950.

251. Hardy WD, Gulick RM, Mayer H, et al. Two-year safety and virologic efficacy of maraviroc in treatment-experienced patients with CCR5-tropic HIV-1 infection: 96-week combined analysis of MOTIVATE 1 and 2. *J Acquir Immune Defic Syndr.* Dec 15 2010;55(5):558-564.

252. Cooper DA, Heera J, Goodrich J, et al. Maraviroc versus efavirenz, both in combination with zidovudine-lamivudine, for the treatment of antiretroviral-naive subjects with CCR5-tropic HIV-1 infection. *J Infect Dis.* Mar 15 2010;201(6):803-813.

253. Wilkin TJ, Goetz MB, Leduc R, et al. Reanalysis of coreceptor tropism in HIV-1-infected adults using a phenotypic assay with enhanced sensitivity. *Clin Infect Dis.* Apr 1 2011;52(7):925-928.

254. Vandekerckhove LPR, Wensing AMJ, Kaiser R, et al. European guidelines on the clinical management of HIV-1 tropism testing. *The Lancet Infectious Diseases.* 2011;11(5):394-407.

255. McGovern RA, Thielen A, Mo T, et al. Population-based V3 genotypic tropism assay: a retrospective analysis using screening samples from the A4001029 and MOTIVATE studies. *AIDS.* Oct 23 2010;24(16):2517-2525.

256. Westby M, Smith-Burchnell C, Mori J, et al. Reduced maximal inhibition in phenotypic susceptibility assays indicates that viral strains resistant to the CCR5 antagonist maraviroc utilize inhibitor-bound receptor for entry. *J Virol.* Mar 2007;81(5):2359-2371.

257. Bussmann H, Wester CW, Ndwapi N, et al. Five-year outcomes of initial patients treated in Botswana's National Antiretroviral Treatment Program. *AIDS.* Nov 12 2008;22(17):2303-2311.

258. Ministry of Health, The Republich of Botswana, MASA. 2012 Botswana National HIV & AIDS Treatment Guidelines. 2012. <http://www.aidstar-one.com/sites/default/files/Botswana_2012_tagged.pdf>

259. Pillai S, Good B, Richman D, Corbeil J. A new perspective on V3 phenotype prediction. *AIDS Res Hum Retroviruses.* Feb 2003;19(2):145-149.

260. Jensen MA, Li FS, van 't Wout AB, et al. Improved Coreceptor Usage Prediction and Genotypic Monitoring of R5-to-X4 Transition by Motif Analysis

of Human Immunodeficiency Virus Type 1 env V3 Loop Sequences. *Journal of Virology.* 2003;77(24):13376-13388.

261. Lengauer T, Sander O, Sierra S, Thielen A, Kaiser R. Bioinformatics prediction of HIV coreceptor usage. *Nat Biotechnol.* Dec 2007;25(12):1407-1410.

262. Cashin K, Gray LR, Jakobsen MR, Sterjovski J, Churchill MJ, Gorry PR. CoRSeqV3-C: a novel HIV-1 subtype C specific V3 sequence based coreceptor usage prediction algorithm. *Retrovirology.* 2013;10:24.

263. Raymond S, Delobel P, Chaix ML, et al. Genotypic prediction of HIV-1 subtype D tropism. *Retrovirology.* 2011;8:56.

264. Raymond S, Delobel P, Rogez S, et al. Genotypic prediction of HIV-1 CRF01-AE tropism. *J Clin Microbiol.* Feb 2013;51(2):564-570.

265. Ndung'u T, Sepako E, McLane MF, et al. HIV-1 subtype C in vitro growth and coreceptor utilization. *Virology.* Apr 10 2006;347(2):247-260.

266. Clevestig P, Maljkovic I, Casper C, et al. The X4 phenotype of HIV type 1 evolves from R5 in two children of mothers, carrying X4, and is not linked to transmission. *AIDS Res Hum Retroviruses.* May 2005;21(5):371-378.

267. Leitner T, Korovina G, Marquina S, Smolskaya T, Albert J. Molecular epidemiology and MT-2 cell tropism of Russian HIV type 1 variant. *AIDS Res Hum Retroviruses.* Nov 20 1996;12(17):1595-1603.

268. Brinchmann JE, Albert J, Vartdal F. Few infected CD4+ T cells but a high proportion of replication-competent provirus copies in asymptomatic human immunodeficiency virus type 1 infection. *J Virol.* Apr 1991;65(4):2019-2023.

269. Rodrigo AG, Goracke PC, Rowhanian K, Mullins JI. Quantitation of target molecules from polymerase chain reaction-based limiting dilution assays. *AIDS Res Hum Retroviruses.* Jun 10 1997;13(9):737-742.

270. Palmer S, Kearney M, Maldarelli F, et al. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *J Clin Microbiol.* Jan 2005;43(1):406-413.

271. Marshall RD. Glycoproteins. *Annu Rev Biochem.* 1972;41:673-702.

272. Gavel Y, von Heijne G. Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng.* Apr 1990;3(5):433-442.

273. Mellquist JL, Kasturi L, Spitalnik SL, Shakin-Eshleman SH. The amino acid following an asn-X-Ser/Thr sequon is an important determinant of N-linked core glycosylation efficiency. *Biochemistry.* May 12 1998;37(19):6833-6837.

274. Kasturi L, Chen H, Shakin-Eshleman SH. Regulation of N-linked core glycosylation: use of a site-directed mutagenesis approach to identify Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. *Biochem J.* Apr 15 1997;323 (Pt 2):415-419.

275. Losman B, Biller M, Olofsson S, et al. The N-linked glycan of the V3 region of HIV-1 gp120 and CXCR4-dependent multiplication of a human immunodeficiency virus type 1 lymphocyte-tropic variant. *FEBS Lett.* Jul 2 1999;454(1-2):47-52.

276. Parczewski M, Leszczyszyn-Pynka M, Urbanska A, Bander D, Boron-Kaczmarska A. Genotypic tropism of antiretroviral-treated patients with drug resistant HIV-1. *J Med Virol.* Nov 2011;83(11):1869-1875.

277. Lin NH, Smeaton LM, Giguel F, et al. Prevalence and clinical associations of CXCR4-using HIV-1 among treatment-naive subtype C-infected women in Botswana. *J Acquir Immune Defic Syndr.* May 1 2011;57(1):46-50.

278. Wagner TA, Frenkel LM. Potential limitation of CCR5 antagonists: drug resistance more often linked to CXCR4-utilizing than to CCR5-utilizing HIV-1. *AIDS.* Nov 12 2008;22(17):2393-2395.

279. O'Meara D, Wilbe K, Leitner T, Hejdeman B, Albert J, Lundeberg J. Monitoring resistance to human immunodeficiency virus type 1 protease inhibitors by pyrosequencing. *J Clin Microbiol.* Feb 2001;39(2):464-473.

280. Archer J, Braverman MS, Taillon BE, et al. Detection of low-frequency pretherapy chemokine (CXC motif) receptor 4 (CXCR4)-using HIV-1 with ultra-deep pyrosequencing. *AIDS.* Jun 19 2009;23(10):1209-1218.

281. Buonaguro L, Tagliamonte M, Tornesello ML, Buonaguro FM. Genetic and phylogenetic evolution of HIV-1 in a low subtype heterogeneity epidemic: the Italian example. *Retrovirology.* 2007;4:34.