THE ROLE OF THE HIV-1 PROTEASE SUBSTRATE IN THERAPY RESISTANCE

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Konstantin Kletenkov

von Russische Föderation

Basel, 2015

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von
Prof. Dr. Markus Affolter, Prof. Dr. Thomas Klimkait und PD Dr. Rainer Gosert

Basel, 08.12.2015

Prof. Dr. Jörg Schibler

ABBREVIATIONS

AIDS	Acquired immune deficiency syndrome
ART	Antiretroviral therapy
ATV	Atazanavir
BS	Basel
CA	Capsid / p24
DRV	Darunavir
EFV	Efavirenz
FPV	Fosamprenavir
HAART	Highly active antiretroviral therapy
IDV	Indinavir
IN	HIV integrase / p31
LPV	Lopinavir
MA	Matrix / p17
NC	Nucleocapsid / p7
NFV	Nelfinavir
NNRTI	Non-nucleosidanalogs reverse-transcriptase inhibitor
NRTI	Nucleosidanalogs reverse-transcriptase inhibitor
PI(s)	Protease inhibitor(s)
PR	HIV protease
RT	HIV reverse transcriptase / p51
SP1	Spacer protein 1 / p2
SP2	Spacer protein 2 / p1
SQV	Saquinavir
SHCS	Swiss HIV Cohort Study
TPV	Tripanavir
ZH	Zurich

1 CONTENTS

2 ABSTRACT

In Switzerland and Germany up to a half of the first-line regimens include protease inhibitors (PIs) [1, 2]. Although in the Swiss HIV Cohort Study (SHCS) most patients under antiretroviral therapy (ART) have suppressed viral loads [3], every third patient is or has been affected by drug resistances [4] which are one of major causes for therapy failure.

HIV resistance against PIs is typically characterized by the accumulation of structural alterations in the viral protease (PR). However, a number of cases of clinical therapy failure under PI-containing regimes have been reported, where genotypic resistance testing did not reveal sufficient explanation from information on the PR and regimen compliance [5, 6]. And certain alterations in the natural substrate of the PR, Gag polyprotein, have been associated with the development of PI resistance [7-13]. Nevertheless, until today most algorithms evaluating PI resistances take solely the protease gene itself into account.

In the SHCS protease inhibitor use and successful treatment are monitored regularly for all patients and every newly enrolled patient receives a genotypic resistance test. We used in vivo cross-sectional sequence data from SHCS patients to scrutinize PI resistance mutational pathways across Gag and PR. Roles of certain mutations as well as of their interactions were investigated.

Here we demonstrate that roughly every fifth of the SHCS patients carries resistance mutations in Gag. And since Gag is not considered by the current genotyping systems the overall level of PI resistance for these patients is underestimated. We report novel Gag mutations of potential clinical relevance and provide additional details on known resistance mutational patterns. Additionally our data support a new potential role of p6 alterations in PI resistance mediated by its phosphorylation. Taken together, our results suggest the relevance of Gag sequence information for the routine genotyping of PI-treated patients of the SHCS.

3 INTRODUCTION

3.1 Human Immunodeficiency Virus

3.1.1 Structure

HIV virions have a spherical morphology typical for most retroviruses (Figure 1), and the particle diameter is around 130 nm [14]. The external proteins gp120 are non-covalently associated with the transmembrane proteins gp41. Both gp120 and gp41 carry polysaccharide modifications. They are functionally active as trimeric complexes and responsible for target cell attachment and fusion. Observed numbers of such complexes vary between 4 and 35 [15], but more recent publications report 10 complexes per particle [16]. They tend to cluster in the mature viral particles but seem to be located randomly in the immature virions [17]. Gp41 proteins penetrate the membranous viral envelope, which originates from the cytoplasmic membrane of the host cell, and inside the membrane reach for the matrix proteins that cover the conical capsid of the virus. The contact between the envelope and the matrix trimeric proteins is provided by amino-terminally attached myristic acid residues [18]. Details about the contact between envelope proteins and matrix proteins is still under discussion [19]. Mature matrix proteins then form a lattice-like layer, which, at budding, becomes responsible for the shape of the virion.

The central core of the structure is represented by the conical (the shape is characteristic of the genus *Lentivirus*) capsid. It is built of matured capsid protein oligomers and protects two single-stranded RNA molecules. The RNA strands are associated with nucleocapsid proteins but neither possess covalent link between them nor exhibit any base-pair contact. HIV virions also include: Additional cellular components such as cyclophilin A bound to the capsid, actin, APOBEC3G, tRNA [20] and 7SL RNA [21]; essential virus-encoded enzymes, in according with features of the lifecycle: reverse transcriptase, integrase and protease; accessory proteins and factors, i.e. Tat, Vif, Vpr, Nef.

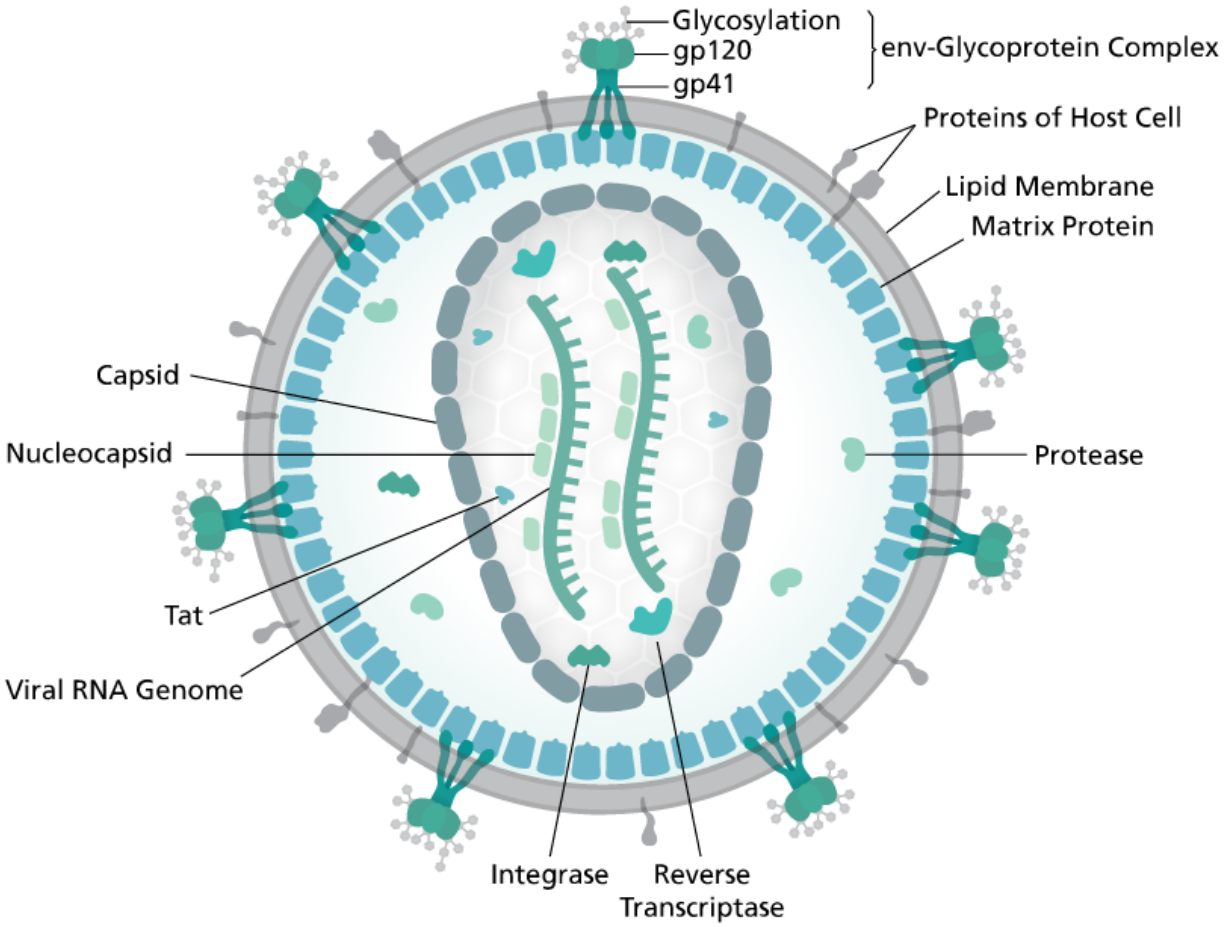


Figure 1. Structure of a mature HIV virion. Illustration by Th. Splettstoesser [22].

3.1.2 Genome organization

The HIV-1 genome is comprised of two linear plus-ssRNA molecules, both 5'-capped and 3'-polyadenylated, matching the key properties of eukaryotic mRNA. Each molecule is typically 9-10 kb in size and contains (Figure 2): Three genes encoding for viral structural proteins typical for all *Retroviridae*: gag, pol and env; two genes encoding for regulatory elements: tat and rev; four genes encoding for accessory regulatory proteins: vpr, vif, nef and vpu.

Encoded by the gag gene (corresponds to Gag protein, "group-specific antigen") are: matrix (MA / p17), capsid (CA / p24), spacer protein 1 (SP1 / p2), nucleocapsid (NC / p7), spacer protein 2 (SP2 / p1), and the p6 protein. Encoded by pol gene (corresponds to Pol protein, polymerase) are: protease (PR), reverse transcriptase (RT / p51), RNase H (p15), integrase (IN / p31) and transframe p6 protein. Envelope glycoproteins gp120 and gp41 are encoded by env gene and are synthesized as the protein precursor gp160.

Regulatory elements are responsible for transactivation while accessory proteins represent virulence factors [23]. Both 5' and 3' ends of the sequence harbor key elements necessary for reverse transcription and consequent integration of the viral DNA into the host chromosome: R ("redundant") is the fragment of identical sequence and orientation at the 3' and 5' termini; U5 ("unique") is located at the 5' terminus and is required for the correct integration process; PB is site responsible for the attachment of the 3' end of a Lys-tRNA molecule; Leader region with splice donor site; a polypurine tract is required for the initiation of the second strand DNA synthesis during reverse transcription; U3 region, which is a U5 analog but is positioned at the 3' terminus of the LTR and is followed by the R fragment. The complete Long terminal repeat (LTR) which regulates the gene expression is formed during reverse transcription of U3, R and U5 and represents the 3' and 5' ends only of the reversely transcribed genome.

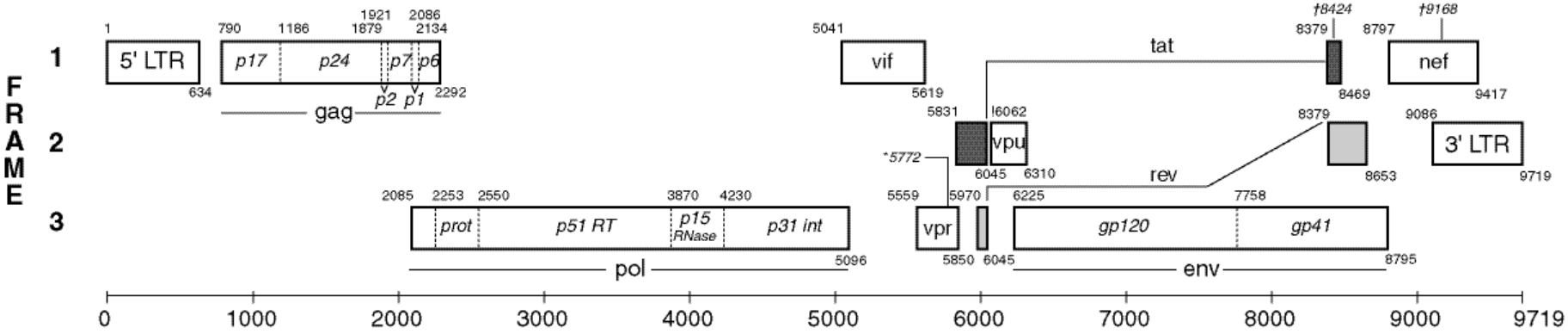


Figure 2. HIV-1 genome organization. Open reading frames are shown as rectangles. The gene start, indicated by the small number in the upper left corner of each rectangle records the position of the a in the ATG start codon for that gene, while the number in the lower right records the last position of the stop codon. For pol, the start is taken to be the first T in the sequence TTTTTTAG, which forms part of the stem loop that potentiates ribosomal slippage on the RNA and a resulting -1 frameshift and the translation of the Gag-Pol polyprotein. The tat and rev spliced exons are shown as shaded rectangles. Illustration from Los Alamos HIV Database [24].

3.1.3 Replication cycle

In order to initiate a new infection, the mature HIV viral particle has to attach to the target cell of T-cells or macrophages (Figure 3). This happens through the primary receptor CD4 as well as chemokine co-receptors CCR5 and CxCR4. Preferred co-receptor determines viral tropism with CxCR4 generally corresponding to T-cell-line tropic viruses (lymphotropic / X4 / "TCL"-tropic) and CCR5 – to the viruses replicating in macrophages (R5 / "M"-tropic). From the viral side, attachment is mediated by the envelope protein complexes. Their interaction with above-mentioned cellular receptors and conformational rearrangements allow entry of the virus capsid into the cell through the fusion between the viral envelope and cell membrane. This may reduce the effects of ART by allowing new infections to happen independently of production of infectious viral particles [25-28].

At this stage the viral capsid has to partially disassemble so reverse transcription process can be initiated in the cytoplasm. DNA is synthesized on the RNA matrix with the help of Lys-tRNA annealing to PB as a primer. Then this DNA binds to the U5 and R region of the RNA. RNase H removes U5 and R region of the RNA. Then the primer relocates to the 3' end of the viral genome which allows the extension of the first strand cDNA. After majority of viral RNA is degraded by RNase H the leftovers prime the synthesis of the second strand. The relocation happens when the two strands hybridize with their PB sequences which allows the extension for both of them.

Still bound to the viral components, dsDNA is then transported as a preintegration complex into the nucleus through the nuclear pores. Vpr, MA and cellular nuclear import factors are the key players at the stage. IN generates a 5' end overhang at both LTR by digesting a dinucleotide from the both 3' ends of the dsDNA. The enzyme also inserts a cut at a random site of the host DNA with overhangs at 5' termini. The 3' ends of the viral DNA genome then bind covalently to the 5' ends of the host DNA via phosphodiester bonds after which viral 5'

overhangs can be removed and single-stranded gaps are repaired by cellular systems. DNA ligases finish the process by adding missing covalent links.

After the first full-length mRNA is transcribed from the integrated provirus and spliced, Tat protein is translated and transcription rate of viral mRNAs is dramatically increased. This is explained by Tat being imported into the nucleus and binding to the TAR elements at the 5' termini of viral mRNAs. It has a stabilising effect and allows for efficient elongation. Variously spliced and unspliced mRNAs allow for production of (Figure 4): Tat, Rev and Nef; Vif, Vpr, Vpu and envelope proteins; structural and enzymatic viral components. Full-length unspliced mRNA are simultaneously used as viral genomes to be packaged into the particles.

Envelope proteins are initially translated as gp160 on the endoplasmatic reticulum. Then during the transport through the Golgi complex to the surface of infected cell gp160 is cleaved by cellular proteases into gp120 and gp41. Fusogenic activity of their complexes on the cell surface allows the virus to infect neighbouring cells in particle-independent manner.

Products of gag and pol genes are initially translated as Gag and Gag-Pol polyproteins on cytoplasmic ribosomes. Then the polyproteins are myristoylated at their p17 end and transported to the place of particle assembly, which is the cytoplasmic membrane in case of T-cells but can be intracellular membranes in the case of macrophages and monocytes. Gag and Gag-Pol polyproteins and their domains orchestrate the assembly and packaging of all components of the viral particle [29] which allows subsequent budding.

PR as a domain of Gag-Pol protein initiates autocatalytic process to excise itself. After that it produces cleavages at a number of fixed sites across the Gag and Gag-Pol polyproteins in order to transform precursor proteins into the proteins that build up the mature viral particle. This occurs when a particle is liberated from the host cell, and this step is necessary for the structural rearrangements that lead to the morphology of infectious virion.

The roles HIV proteins are summarized in the Table 1.

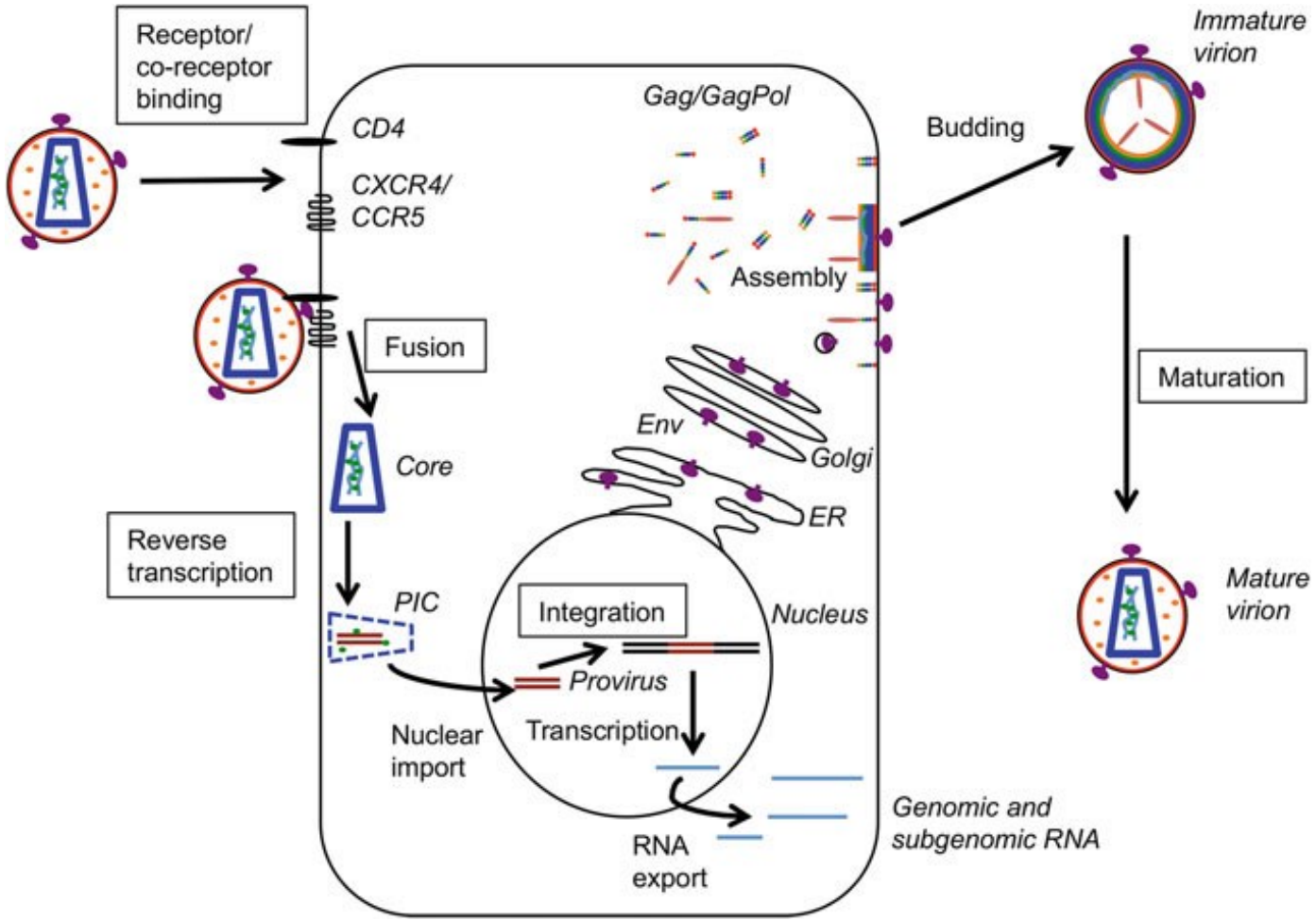


Figure 3. Scheme of the HIV replication cycle. Viral and cell components are labelled in *italics*, processes in plain text, and processes that can be inhibited by current antiretrovirals are boxed. MA, red; NC, green; p6, orange; Env, purple; viral RNA, cyan; viral cDNA, brown. Illustration by Tedbury and Freed [30].

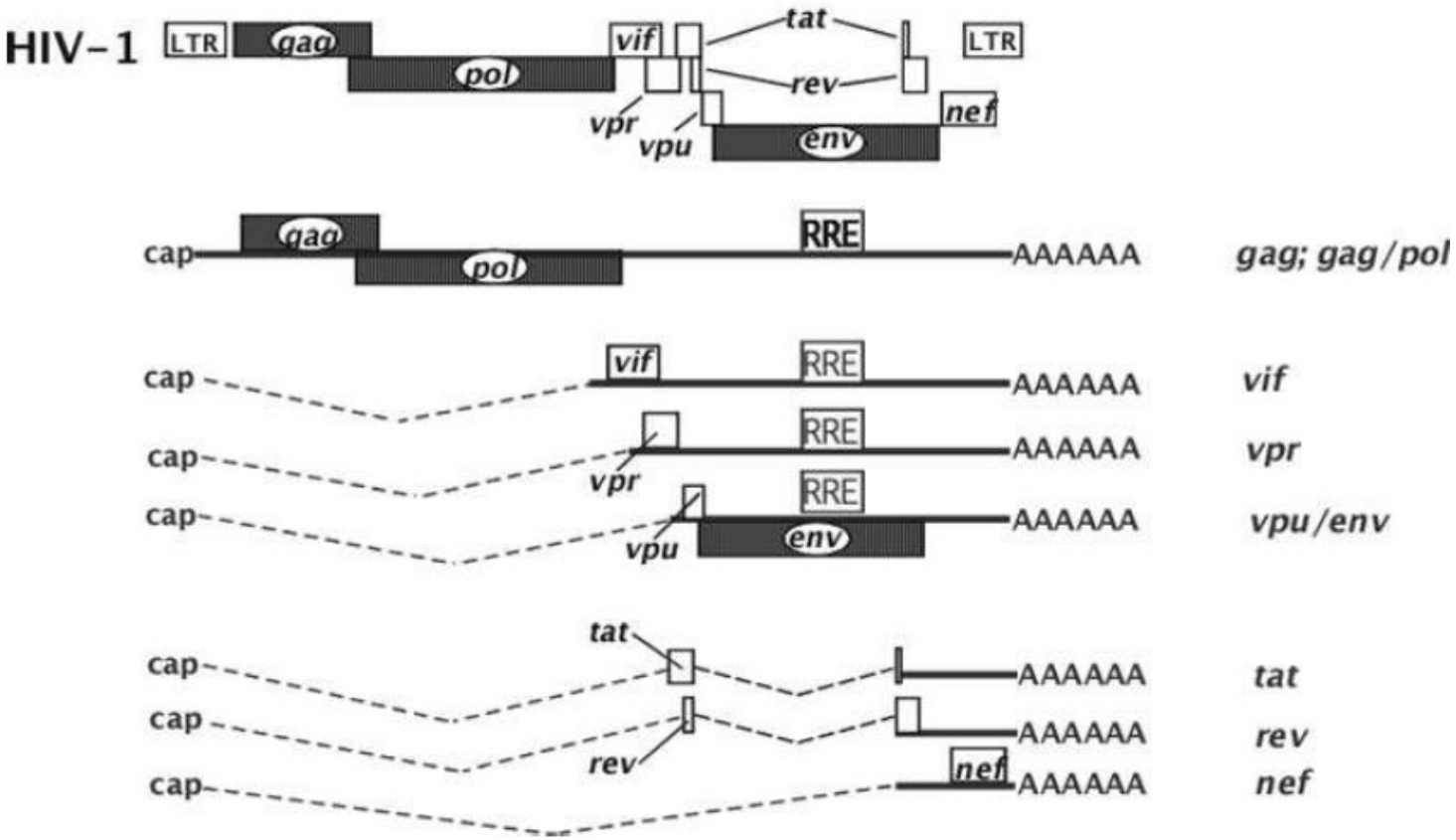


Figure 4. HIV-1 splicing patterns. The genomic organization of the proviral DNA and the location of protein coding sequences are indicated. The dashed lines connect the major splice donor to a downstream splice acceptor. Adapted from Fields, Knipe and Howley [31].

Table 1. HIV proteins and their respective roles. Adapted from Votteler and Schubert [32].

Protein class	Gene	Product	Processed product	Role
Structural	gag	Gag polyprotein	MA, CA, SP1, NC, SP2, P6	Virion assembly
	pol	Gag-Pol polyprotein	TF P6, PR, RT, RNase H, IN	Genome replication
	env	gp160	gp120, gp41	Attachment and entry
Regulatory	tat	Tat	Tat	Positive regulator of LTR transcription
	rev	Rev	Rev	Regulator of viral gene expression: splicing and transport
Accessory	nef	Nef	Nef	Downregulation of CD3, CD4 and MHC-1, signalling and T-cell activation, apoptosis, infectivity enhancement
	vpr	Vpr	Vpr	Virus associated membrane transduction, nuclear import of preintegration complex, differentiation, cell cycle arrest and apoptosis, regulation of glucocorticoid receptor
	vif	Vif	Vif	Infectivity factor: suppression of antiviral activity of APOBEC3G (cytitine deaminase)
	vpu	Vpu	Vpu	Augmentation of virus release, CD4 degradation, inhibition of NF-κB activation

3.1.4 History, diversity and classification

First clinical observations of the acquired immune deficiency syndrome (AIDS) were made in1981 in USA among an MSM cohort patients suffering from opportunistic infections [33]. And already in 1983 HIV was discovered as a causative infectious agent of AIDS by the two independent groups of Luc Montagnier and Robert Gallo [34, 35]. Several years passed before the term "HIV" was commonly accepted.

HIV belongs to the family *Retroviridae*, subfamily *Orthoretrovirinae*, genus *Lentivirus*. One of its characteristics is its high genetic variability that results from three major reasons: missing proofreading activity of the RT, that leads to high mutation rate ($3x10^{-5}$ per nucleotide base per cycle of replication); copy-choice recombination due to the ability of RT to switch between 2 RNA templates of a viral particle, which are not necessarily of identical sequence (2-20 events per genome per replication cycle); fast replication cycle (10^{10} virions a day) [36-38]. These factors along with the genome size of HIV allow multiple alterations at every nucleotide position every day. This is the basis of the observed vast viral diversity along with a rapid selection towards resistance under ART.

Two types of HIV have been described: more virulent and infective HIV-1 that causes most of the HIV infections and its less transmissive counterpart HIV-2 mostly observed in West Africa region [39, 40].

HIV-1 originated from Africa as a result of zoonotic transmissions of its phylogenetic "relative" SIV to humans. Natural SIV hosts, different simian species, generally do not develop disease upon infection. SIVcpz is considered to be the direct ancestor of HIV-1; this virus can cause AIDS-like symptoms in chimpanzees [41]. Sequence difference up to 30% allows classification of HIV-1 onto three major groups: M (major), O (outlier) and N (new / non-major) [42]. Group M isolates represent most of all cases of HIV infection. In this group several subtypes (or clades) have been identified: A, B, C, D, F, H, J and K [43]. They have different prevalence over geographic regions, and subtypes B and C are the

most frequent in the M group. Recombinant circulating forms (CRFs) have been also described and are the result of co-infection with the viral variants of different subtypes.

HIV-2 is phylogenetically distant from HIV-1: sequence difference can be as high as 40%. It is thought to have originated from SIVsmm that asymptomatically infects West-African sooty mangabey monkeys. This group can also be subdivided into subtypes: from A to H.

Since the discovery of the pathogen causing AIDS both basic and clinical research advanced extremely. Today one can effectively diagnose HIV infection and suppress viral replication using ART; HIV itself and the process of its pathogenesis are well understood [44, 45]. There has been also a major progress in decreasing HIV transmission, particularly mother-to child transmission [46].

There are three main avenues along which the healthcare community is now moving towards the goal of ending the HIV pandemic [47]: Complete and comprehensive global implementation of available treatment and prevention tools; research on elimination of the virus in patients or control of infection not dependent on lifelong ART; development of novel potent prevention tools that could complement and enhance the ones available currently.

Yet by the end of 2013 around 39 million people globally have died from HIV-related causes, and HIV infection continues to be one of the major unsolved global health problems as there is still no cure for the currently 35 million people living with it and for around 2 million newly infected (figure from 2013 [48]). In 20 years from 1990 to 2010 it went up in global ranks for causes of disability-adjusted life years from 33rd to 5th place [49, 50].

3.1.5 Infection, pathogenesis and clinical features

The clinical course of the disease is summarised in Figure 5 and generally divided into three phases: primary infection, latency (chronic infection) and AIDS.

Infection is carried out by mature HIV particles or virus-infected cells [25-27]. They get into the vaginal or intestinal mucosa and into the bloodstream through sexual contact or injuries. HIV initial targets are Langerhans cells of the skin, dendritic cells and macrophages. The latter ones allow the virus to establish reservoirs for long-term persistence. T lymphocytes are usually considered to be inoculated later. Infected cells travel through the lymphatic vessels and bloodstream. Lymph nodes with their dendritic cells represent a viral reservoir where monocytes, macrophages and primary T lymphocytes get infected. Macrophages also allow transportation of virus to the brain and other organs and infection of other cell types like astrocytes and endothelial cells.

Only one third of all HIV infections are described to manifest with the typical flu-like symptoms, unspecific rash and swollen lymph nodes in the first phase or "primary infection" several weeks after exposure; most cases are clinically unapparent. Up to 10^6-10^8 viral genome copies per millilitre of blood can be detected with quantitative PCR method at this stage. CD4 cell level drops below 500 cells per microliter of blood, and the CD4/CD8 ratio shifts below 0.5. Duration of several initial months is characteristic for this phase. Chronically unapparent infection or clinical latency may characterize a very prolonged (for up to more than two decades) phase often with no or mild observable symptoms: fever, weight loss, diarrhoea, fatigue and coughing may occur. One to three months post infection HIV-specific antibodies and T-lymphocytes can be detected. Viral load goes down often to only several thousand genome equivalents per millilitre of peripheral blood. Spleen, tonsils and Peyer patches are also the sites of viral replication during the stage. Viral proliferation occurs but still under control by the immune defence.

When CD4 cell level decline cannot be compensated by the differentiation of the bone marrow progenitor cells anymore, the control is lost. Damage to the CD4 cell population leads to the failure of associated immunological functions; CD8 lymphocytes are not indirectly activated anymore. HIV genetic variability also complicates immune recognition due to the alteration of epitopes. The immune system malfunction allows the development of opportunistic infections. The time when the CD4 cell count falls below 200 cells per μL of blood is the onset of clinical AIDS symptoms accompanied by fever, nocturnal sweating, swollen lymph nodes, weight loss and sometimes neurological problems.

There are several reasons for the loss of CD4 cells: Direct elimination by viral replication through necrotic pathway [51], mostly affects CD4 T lymphocytes; apoptotic processes induced by Tat expression, by cytokines and chemokines generated by infected macrophages and monocytes, and by gp120-antibody complexes bound to the uninfected cells; elimination by cytotoxic CD8 lymphocytes.

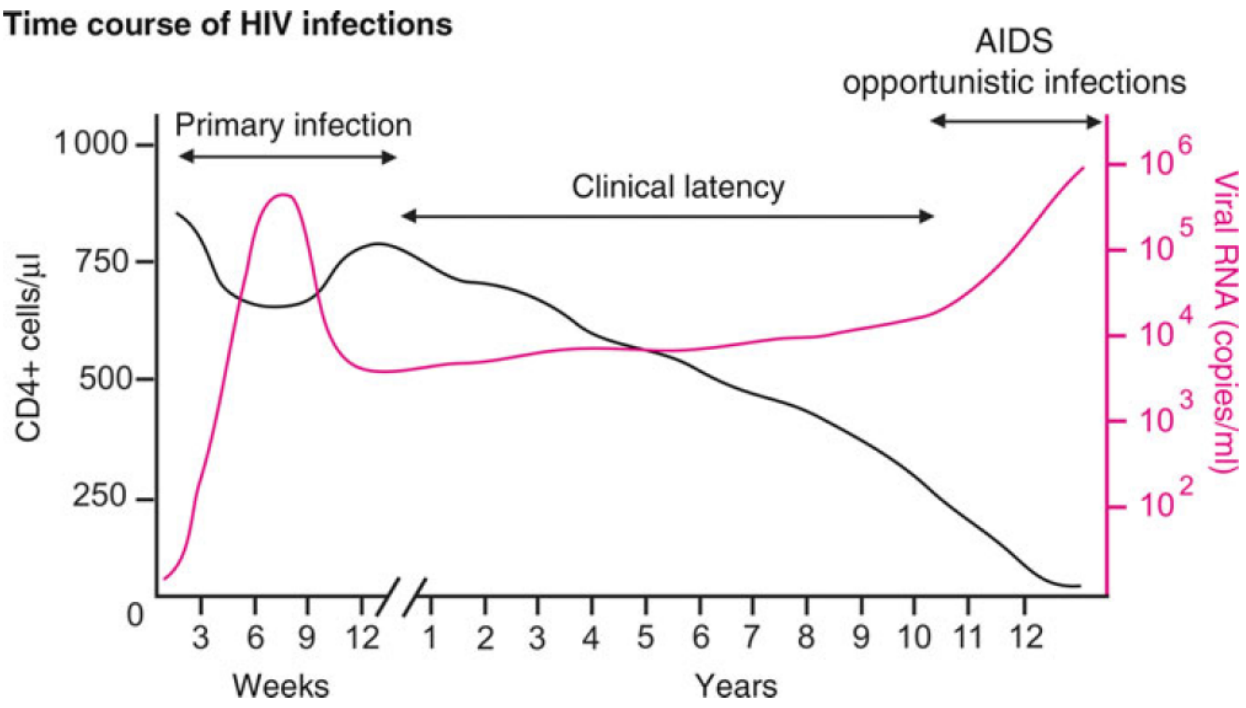


Figure 5. Clinical progression of HIV infection. The black curve represents the time course of the number of CD4 cells per microliter of blood; the red curve shows the number of viral genome copies per millilitre of blood. The time axis includes the first few weeks of infection up to a period of more than 10 years. Figure by Modrow *et al*. [52].

3.2 Gag, protease and protease inhibitor resistance

3.2.1 Gag structure and function

Initially recognized as a simple scaffold protein forming the viral particle, Gag has been shown to perform multiple functions in the HIV lifecycle and to be involved in multiple interactions with both cellular and viral components. Therefore today it is considered to be an emerging therapy target.

Gag or $Pr55^{Gag}$ (Figure 6) and Gag-Pol polyproteins are translated from the full-length RNA which serves as both the genome to be included into assembling viral particles [53]. Gag and Gag-Pol are produced at a rate of approximately 20:1. This is facilitated by the ribosomal slippery site in a uridine-rich region of the mRNA corresponding to the transframe p6 fragment [54]. In case of a frame-shift most of p6 is left out and PR, RT, RNase H and IN sequences are translated. Once Gag and Gag-Pol have been produced in the cytoplasm of a host cell they are guided by MA to the cholesterol-rich microdomains of plasma membrain [55, 56]. And such behaviour of MA is driven by its membrane-binding domain which includes an N-terminal covalently attached myristic acid and a basic region [18, 57-59].

Then binding of viral genomic RNA with NC domain of Gag renders Gag multimerization and assembly of the immature viral particle [60, 61]. The RNA association occurs via the overall positive charge of the NC Gag domain. And the specificity for the viral genomic RNA results from a direct interaction of the RNA packaging signal with two extremely conserved zinc finger motifs within the viral NC [61, 62].

CA as a part of Gag polyprotein is responsible for intermolecular interactions facilitating Gag multimerization and particle assembly [63]. The C-terminal domain (CTD) of CA containing a well-conserved major homology region is essential for this assembly process [64, 65]. The N-terminal domain (NTD) of CA carrying a proline-rich loop binds cyclophilins, in particular cyclophilin A [66]. It has been suggested that cyclophilin A binding to the capsid

core protects HIV-1 from being recognized by the cellular innate immune response [67].

Then glycosylated trimers of the Env glycoproteins (3 x gp120 + 3 x gp41) are incorporated into the immature Gag carcase through interaction of gp41 with MA [19, 68, 69]. Their origin is described in details elsewhere [19].

Further particle budding and release depends on the membrane scission which is also mediated by Gag. The p6 fraction of Gag contains two so-called late domains necessary for the recruitment of endosomal sorting complexes required for transport (ESCRTs: ESCRT-0, I, II, and III) to perform the scission. Normal functions of ESCRTs are discussed elsewhere [70, 71]. One of the late domains is Pro-Thr-Ala-Pro (PTAP) motif that binds directly to the ESCRT-I component Tsg101. Another late domain, Tyr-Pro-Xn-Leu sequence (YPXnL, where X may be any residue, and n = 1–4 amino acids) interacts with ESCRT-associated factor ALIX. The main late domain is PTAP motif [72-74], but YPXnL motif is indispensable for HIV-1 replication in some cell types [75].

Gag and Gag-Pol cleavage is performed by the PR resulting in the release of mature Gag and Gag-Pol derived proteins: MA, CA, SP1, NC, SP2, p6, PR, RT, RNase H, IN. This triggers viral particle maturation which happens during or shortly after the release of immature virion [63]. Maturation enhances the fusogenic potential of Env protein complexes [76, 77] and allows formation of MA lattice and mature CA conical core [16, 63]. MA forms hexamers of trimers so that the MA trimers orient themselves on top of the underlying hexameric lattice formed by CA [78]. Certain mutations in MA can completely block the incorporation of HIV-1 Env. Such block of Env incorporation can be rescued by truncations and alterations in the cytoplasmic tail of gp41 or by heterologous short-tailed Env glycoproteins. Efficiency of incorporation of the truncated HIV-1 Env depends on the cell type studied [79, 80]. Additional mutations at the MA trimer interface could compensate some of MA mutations that cause Env incorporation block. So MA trimer formation is suggested to play an important role in Env incorporation [81].

The structure of the hexagonal CA core is sealed with seven CA pentamers at the wide end and five - at the narrow end. Both CA hexamers and pentamers are formed based on NTD-NTD and intermolecular NTD-CTD interactions. At the same time CTD-CTD interactions allow the formation of the broad hexamer lattice [82]. Alterations of CA amino acid sequence have been shown to affect capsid core stability and influences virus infectivity [83]. Such rearrangements can be clearly seen on the pictures from electron microscopy of immature versus mature virions (Figure 7). Mature viral particles at this stage are finally ready to infect target cells and initiate another round of infection.

Gag derivatives additionally play an important role in post-entry events. So NC also functions as a nucleic acid chaperone which promotes reverse transcription and downstream stages of the viral lifecycle [60].

A further role of CA is its participation in processes of reverse transcription along with MA [84, 85]. CA interacts both with cellular transportins and nuclear pore components (karyopherin TNPO3, nuclear pore proteins Nup153 and Nup358) to control the nuclear import of pre-integration complex [86-88]. In line with this CA is considered to be a factor allowing lentiviral infection of non-dividing cells [89]. Cyclosporin A prevents the binding of cyclophilins to CA. This impairs HIV-1 replication [66, 90]. It has been suggested that by blocking the binding of cyclophilin A or Nup358 to CA, cyclosporin A can “unmask” the viral core, allowing it to be recognized by restriction factors [91] or other components of the host innate immune response [67].

A role of p6 as a Gag domain relevant for post-entry events is the recruitment of the HIV-1 accessory protein Vpr into a virion with the help of a specific binding sequence. Vpr impacts viral replication and pathogenesis. It participates in guiding of pre-integration complex to a nuclear pore and subsequent nuclear transport [92, 93].

As for the spacer peptides of Gag, they regulate kinetics of Gag processing. SP1 in addition forms part of the sequence following the C-terminus of CA which is crucial for Gag-Gag interactions at the particle assembly stage [94-96].

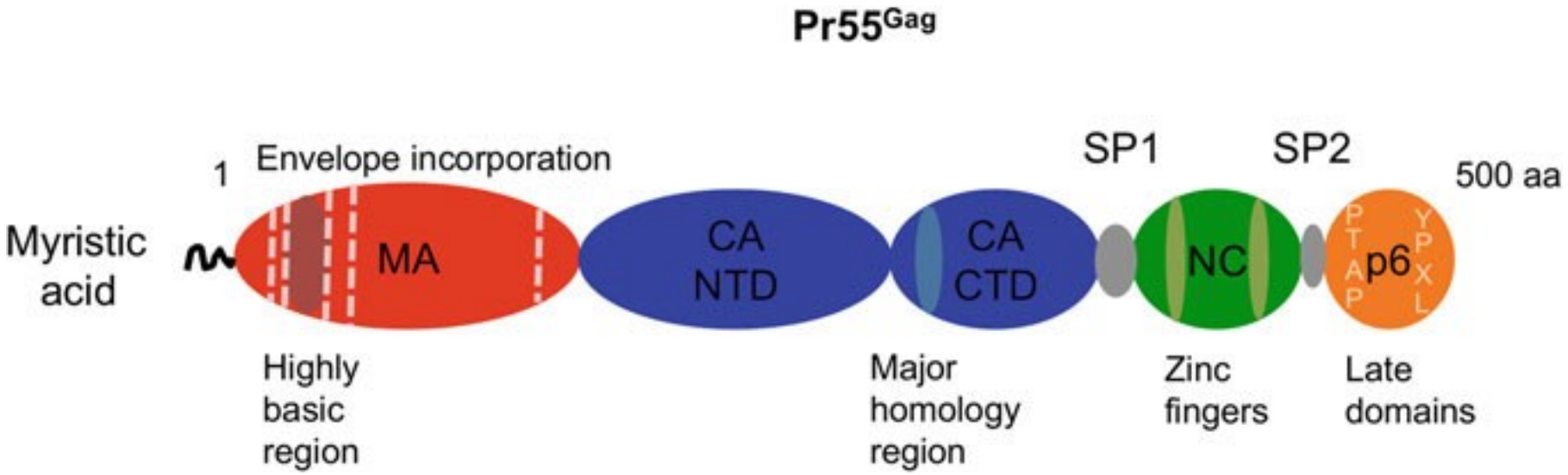


Figure 6. Schematic drawing of HIV-1 Gag indicating major functional motifs. The myristic acid and highly basic region of MA mediate membrane interactions of Gag. Residues in MA that have been shown to affect Env incorporation are indicated with dashed vertical lines. CA is divided into N-terminal and C-terminal domains, NTD and CTD, respectively. The NTD promotes pentamer formation, while the CTD, which also contains the major homology region, is required for CA dimerization and multimerization. NC contributes to Gag assembly by binding nucleic acid, typically the viral genome, via its zinc finger motifs, leading to long-range Gag multimerization. The p6 contains the late domains PTAP and YPXL, which bind TSG101 and ALIX, respectively, thereby recruiting the ESCRT machinery to facilitate virus budding from the cell membrane. MA, red; CA, blue; NC, green; p6, orange. Spacer peptides SP1 and SP2 are indicated, as is the approximate length of the Gag precursor (500 amino acids). By Tedbury and Freed [30].

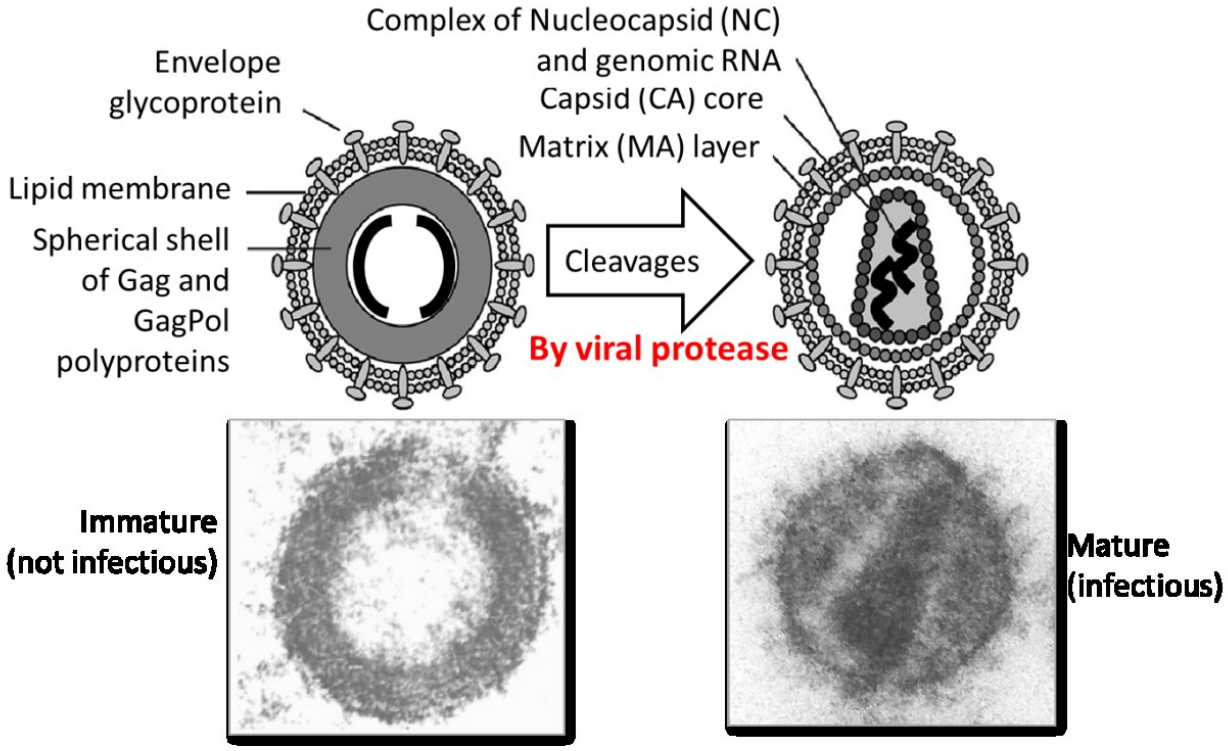


Figure 7. Effects of maturation. Transition between the immature, not infectious, and mature, infectious, state is initiated by the viral protease cleavages. Nothing new enters the viral particle at this stage, only rearrangements take place. Spherical shell of Gag and Gag-Pol polyproteins is converted into well-differentiated structures of infectious virion. Illustration was kindly provided by Th. Klimkait.

3.2.2 HIV-1 protease

The HIV-1 protease is critical for the viral life cycle. It initiates maturation of produced viral particles as described above. The natural substrates of PR are the Gag and Gag-Pol polyproteins that provide the virus with the key structural and enzymatic components. The HIV-1 protease functions as a homodimer. It is a retroviral aspartyl proteinase [97] with the active center including aspartic acid 25, threonine 26 and glycine 27 [98-102]. Though PR is a small protein and carries out critical function in viral maturation and infectivity, it possesses notable plasticity, and polymorphisms have been observed in one-third of its 99 amino acids [103, 104]. Three functional domains can be identified in the PR structure: active site cleft, two flaps above it and a dimerization interface [105].

HIV-1 PR possesses some activity already as a domain of Gag-Pol polyprotein. However such PR dimers are unstable and much less active than the released, excised form of protease. This is because embedded PR adopts the proper conformation only for short periods of time [106-108]. Initial intramolecular cleavage events are: SP1/NC then internal transframe protein cleavage site and transframe protein / PR cleavage site [108-111]. Now, liberated from one side, PR can gain proper conformation and therefore stability and catalytic activity[112]. Now cleaving becomes intermolecular and the PR monomers are completely liberated from the Gag-Pol precursor [113, 114].

When the N-terminus of the PR is bound it cannot cut intermolecularly [111], therefore Gag processing occurs subsequently to PR dimer maturation. The process of cleavage is highly specific and temporally and spatially regulated (Figure 8, Figure 9). However, PR needs to be somewhat promiscuous as it recognizes 12 Phe-Pro and Tyr-Pro containing cleavage sites (none of which is efficiently cleaved by mammalian proteases) with their individual sequences [54, 115-117] (Table 2). The order of cleavage is determined by the relative processing rates of individual cleavage sites. The “fastest” cleavage site SP1/NC is processed

400 more effectively than the “slowest” cleavage sites CA/SP1 and NC/SP2 [118-121]. And processing efficiency of each individual cleavage site is thought to depend in a complex way on its amino acid sequence, time of accessibility to the active center of the protease, formed shape, conformation of surrounding protein and contextual cues [96, 122, 123].

Table 2. HIV-1 M-group PR cleavage site decapeptides. TFP – transframe protein.

Site	P5	P4	P3	P2	P1	✂	P1'	P2'	P3'	P4'	P5'
MA/CA	V	S	Q	N	Y	/	P	I	V	Q	N
CA/p2	K	A	R	V	L	/	A	E	A	M	S
p2/NC	T	S	A	I	M	/	M	Q	R	G	N
NC/p1	E	R	Q	A	N	/	F	L	G	K	I
p1/p6^{gag}	R	P	G	N	F	/	L	Q	S	R	P
NC/TFP	E	R	Q	A	N	/	F	L	R	E	N
TFP/p6^{pol}	E	D	L	A	F	/	L	Q	G	K	A
p6^{pol}/PR	V	S	F	N	F	/	P	Q	V	T	C
PR/RTp51	C	T	L	N	F	/	P	I	S	P	I
RT/RTp66	G	A	E	T	F	/	Y	V	D	G	A
RTp66/INT	I	R	K	V	L	/	F	L	D	G	I
Nef	A	A	C	A	W	/	L	E	A	Q	E

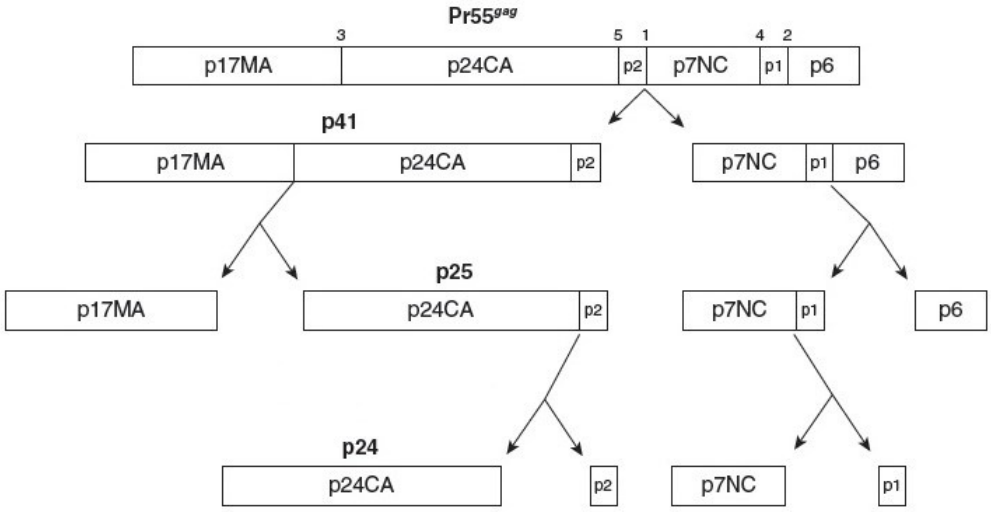


Figure 8. Gag polyprotein processing via ordered cleavages by PR. The order indicated at the top of the figure by the numbers over the cleavage sites is determined partly by the intrinsic processing rate of each cleavage recognition sequence and partly by preceding cleavages at neighboring sites. By Salzwedel, Martin and Sakalian [124].

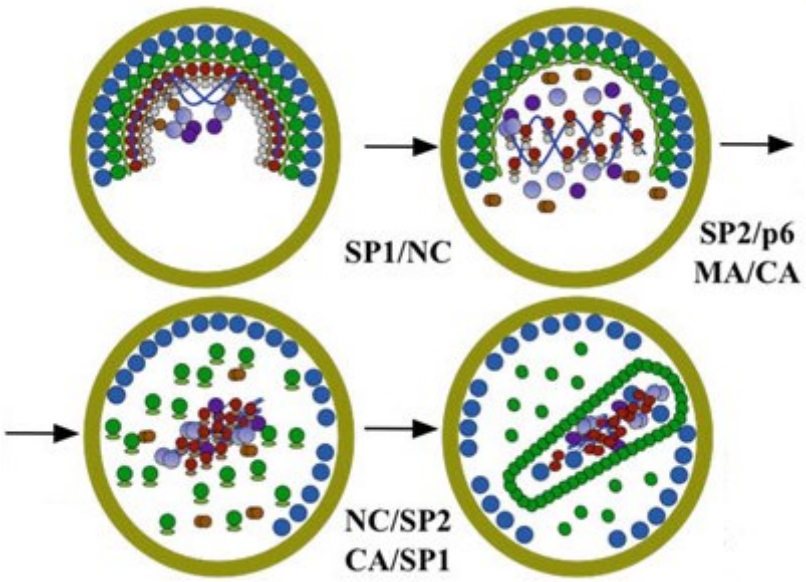


Figure 9. A model representation of the step-wise processing of HIV-1 Gag by the HIV-1 protease. Gag, comprising MA (blue), CA (green), SP1 (light green), NC (red), SP2 (tan), and p6 (gray), is extended in a radial orientation from the membrane (gold), as is Gag-Pro-Pol, which contains the viral enzymes PR (brown), RT (blue–gray), and IN (purple). By Potempa *et al*. [125].

3.2.3 Maturation inhibitors

There are two principle ways to block Gag and Gag-Pol cleavage by the PR and therefore prevent initiation of the maturation stage of viral lifecycle: to target the enzyme or the substrate. The former option is carried out by the protease inhibitors while the latter one – by maturation inhibitors.

Maturation inhibitors disrupt Gag cleavage. The currently only drug of the class, bevirimat, binds to and blocks the CA/SP1 cleavage site from being accessed by PR. Even if it occurs in an incomplete fashion, such a disruption causes the accumulation of a p25 precursor (see Figure 8) and loss of infectivity [126, 127]. Bevirimat, a derivative of betulinic acid (isolated from *Syzygium claviflorum*), was the first (and by now the only) compound of the class, which reached phase IIb clinical trials. It causes aberrant virion morphology and replication defects [128]. The compound proved to be safe with only mild and rare side effects such as headaches and throat discomfort [124, 129]. However, a large fraction (roughly half) of patients carried viruses with amino acid polymorphisms at the bevirimat binding site (SP1 residues 6–8 / Gag residues 369-371) that rendered the virus less susceptible to the drug [130, 131]. Some of the resistance mutations revealed subtype-specific consensus. Work on bevirimat as a potential therapeutic agent was discontinued due to the high prevalence of resistance-conferring polymorphisms [132-134]. Furthermore, it became obvious that only liquid formulation provided adequate drug levels, and such formulation is undesirable for commercial development.

Another chemically unrelated molecule, PF-46396, demonstrated similar anti-HIV effects mediated by a related mechanism applied to the same cleavage site. Its development encountered the same problem of resistance mutations clustering in the CA/SP1 junction region [135, 136].

3.2.4 Protease inhibitors

Since PR is indispensable for the HIV viability it represents extremely attractive and well-studied drug target. The substrate-mimicking compounds of the protease inhibitor class bind to the PR enzyme and, in contrast to maturation inhibitors, therefore block any stage of Gag and Gag-Pol processing. Nine PIs were approved for the treatment of HIV infection: saquinavir, ritonavir, indinavir, nelfinavir, amprenavir, lopinavir, atazanavir, tipranavir and darunavir (Figure 10). Low doses of ritonavir were used for boosting (to slow down metabolism of the drugs making up a regimen backbone).

The development of PI enabled the dual class triple combination therapy that became known as highly active antiretroviral therapy (HAART) [103, 137, 138]. All PIs except TPV are essentially analogues of the transition state of a natural PR substrate [139, 140]. They mimic a cleavage site recognized by the PR, but instead of natural and hydrolysable P1-P1' amide chemical bond they carry non-hydrolyzable transition state isosteres [141]. PI possesses special and distinct features compared to the compounds of other classes. One feature of PI is their cooperative inhibition of PR: Minor reductions in drug concentration or effectiveness cause nonlinear decreases in inhibition [125, 142-144]. This results in the steeper slopes of inhibition curves. A second feature is pleiotropic effect of the compound of the class (Figure 11): PIs are able to interfere with viral function at multiple stages of the viral lifecycle [145] (fusion [76, 77, 145, 146], reverse transcription [147, 148], nuclear import and integration [145]). The ability of PI to affect fusion might be mediated by HIV Env proteins. Since uncleaved Gag is stably linked to Env trimers [149] block of maturation prevents mobility and formation of a single cluster of Env molecules on the surface [17] which results in reduction of infectivity and fusogenic potential [76, 77, 146]. Several studies find the link between fusion and PR inhibition to be cell-type dependent [145] and co-receptor dependent[146]. PI is suggested to also affect the reverse transcription process in two ways: by decreasing RT activity and by interfering with the

assembly of the ribonucleoprotein (RNA and NC) and therefore a reverse-transcription complex. A simple way of PI control over initiating RT activity is by trapping RT in a precursor form where its activity has been estimated to be much lower [150, 151]. And the assembly of reverse transcription complex has been shown to depend heavily on the sequence of proteolytic processing of Gag and Gag-Pol polyproteins [119, 120, 152, 153]. In addition, normal assembly of a reverse transcription complex has been shown to be disrupted by the accumulation of intermediates of Gag processing [127, 154, 155].

As of nuclear import and integration, there are also several ways suggested that allow PI interference. The most obvious is trapping IN, CA [89, 156, 157], NC [158, 159] and MA as components of pre-integration complex within the precursors [85, 160-163]. However, in such a case the virus would not even make it through the reverse transcription. It is currently suggested that PI could affect CA assembly in such a way as to allow reverse transcription to occur, but then compromise its ability to facilitate nuclear import [125]. Another option is the interference with NC functions via its precursors. It has been reported that certain alterations of Gag C-terminal domain amino acid sequence block processing at the SP2/p6 site but still allow reverse transcription to occur [164]. Wrong timing of SP2/p6 cleavage could enable the CA cone to assemble before condensation of the core producing a reverse transcription-competent, but nuclear import-defective virus [125].

A third feature of PI that extends beyond their pleiotropic effects is the consequence of targeting the active site of the PR itself. Because PI mimic the transition state of the natural substrate of PR using their characteristic hydroxyl group the enzyme lowers the free energy of activation for the reaction [165]. While binding affinity of PR to its natural substrates are in the μM to mM range [166], PI bind the wild-type PR in the nM to pM range [167-171], so there are several orders of magnitude difference. Therefore PI are considered to be relatively potent drug class (Figure 12).

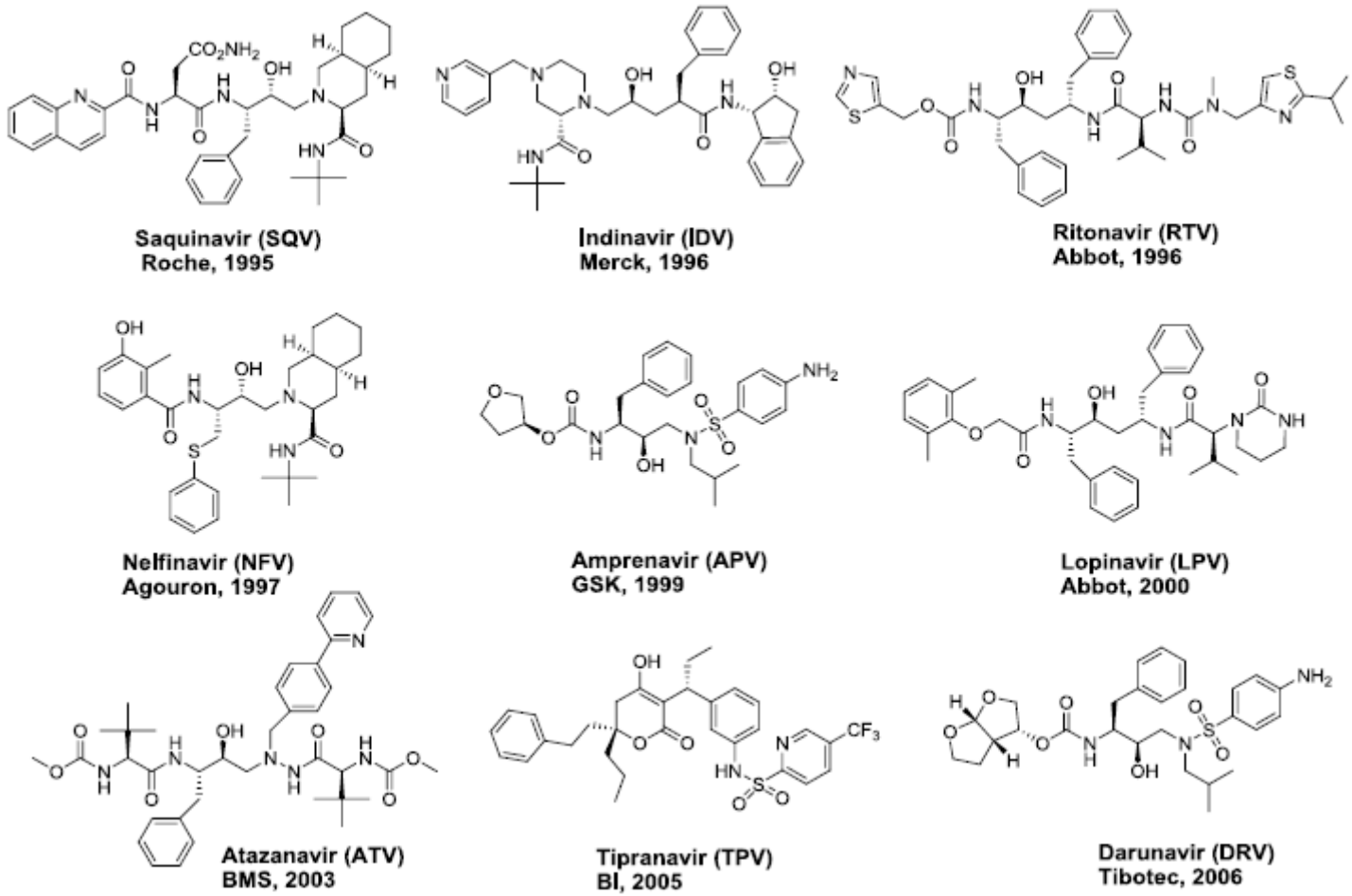


Figure 10. Chemical structures of the nine HIV-1 protease inhibitors approved for clinical use. Peptidomimetic protease inhibitors are characterized by a hydroxyethylene core. TPV, non-peptidomimetic protease inhibitor is characterized by a dihydropyrone ring. By Ali *et al*. [172].

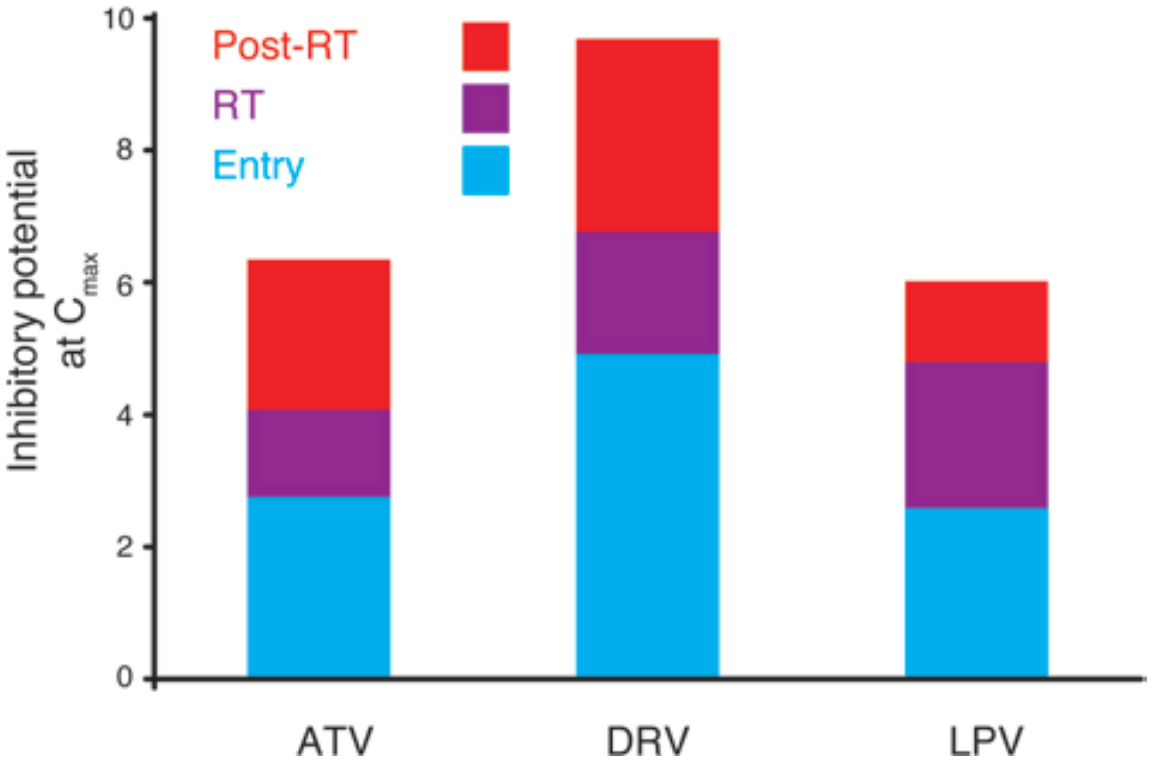


Figure 11. Contribution of the inhibitory effect of PIs on each step of viral life cycle to the overall inhibitory effect at *Cmax*. The linear dose-response curves of PIs at entry, reverse transcription, and post–reverse transcription steps were extrapolated to predict the inhibition of each step at *Cmax*. By Rabi *et al*. [145].

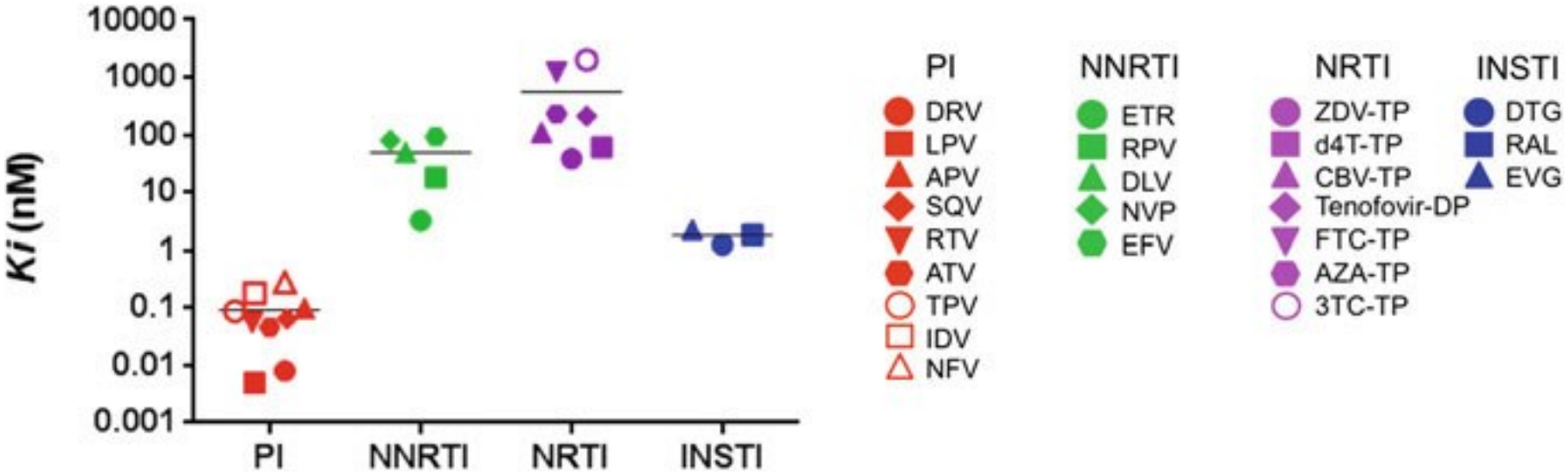


Figure 12. Comparison of the inhibitory constants for each inhibitor from four of the antiretroviral drug classes: protease inhibitors (PI), non-nucleoside reverse transcriptase inhibitors (NNRTI), nucleoside reverse transcriptase inhibitors (NRTI), and integrase strand transfer inhibitors (INSTI). By Potempa *et al*. [125].

3.2.5 Protease inhibitor resistance

High genetic variability of HIV allows it to develop resistances against antiretroviral therapies applied. In patients the virus develops a number of genetically distinct viral variants, referred to as a viral quasispecies [173]. The pool of quasispecies representing viral population in a patient is capable of evolving and responding to different selection pressures such as immune response and therapy applied.

The number of HIV variants that produces infectious progeny is relatively small [174-176]. Only 0.1%-1% of viral particles per generation is thought to be capable of carrying out new infection [177-182]. One of the reasons is the same mechanisms of genetic variability: make the virus to produce a high fraction of "dead" viruses carrying deleterious mutations. The number of resistance-associated mutations necessary to confer virological failure is defined as the barrier to resistance development [103]. But there are also other factors that have to be taken into account: baseline variability (groups and subtypes) and impact of the mutations on viral replication capacity.

Resistance against protease inhibitors has been observed and documented. This stepwise process often starts with substitutions that directly or indirectly alter the structure of the substrate-binding cleft of PR [183-185]. And the general tendency of the process is to widen the catalytic cleft for the enzyme. Due to such enlargement inhibitors lose affinity and drug susceptibility of the mutated virus is reduced. On the other hand and for the same reason the binding of the natural substrate may similarly be impaired which can lead to losses of viral replication capacity or fitness [8, 186-188]. Such a process is particularly apparent for the main ("primary") mutations. These are major or primary resistance mutations in protease. They tend to be selected first, located in the critical functional regions of PR, capable of reducing protease inhibitor susceptibility individually and extremely rare occurring in untreated isolates [104, 189, 190]. Other PR mutations are classified as secondary resistance mutations in protease, and in general, they

tend to emerge later and do not confer resistance effects by themselves in vitro. They are, however, capable of improving fitness of the viruses carrying primary resistance mutations and cooperatively enhance the degree of resistance. Secondary resistance mutations are usually located outside the critical functional regions of PR and can be observed in untreated isolates [186-188, 191].

There are mutations specific to certain drugs (D30N - NFV, I50L - ATV), however cross-resistances are very common (positions: 10, 46, 54, 82, 84 and 90) [104, 190]. The summary primary PI resistance mutations at 15 protease codons and secondary resistance mutations at 19 protease codons were summarized in Figure 13 [189].

It has been recently demonstrated that a second locus can be responsible for the resistance against protease inhibitors and for the compensation of resistance-associated fitness loss. Mutations in Gag located in or close to protease recognition sequences are more commonly found in treated viruses. They are thought to be an adaptation of the virus to the altered substrate-binding cleft of the mutant drug-resistant viral protease [7, 191, 192].

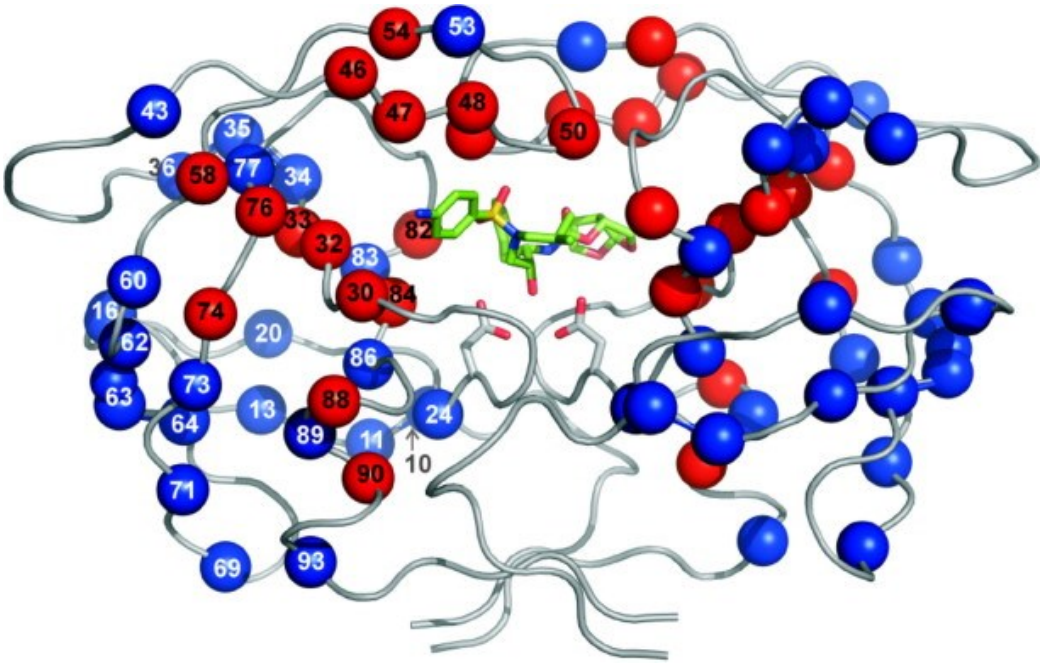


Figure 13. Three-dimensional structure of HIV PR dimer depicting the primary (major) and secondary (minor) mutations associated with resistance to protease inhibitors. Illustration by Johnson *et al*. [189]. Mutated residues are represented with their Cα atoms (spheres) and colored red and blue for major and minor mutations, respectively. Active site aspartates and DRV bound to the active site are represented in sticks. The figure was generated using the structure of highly mutated patient derived HIV PR [193] (PDB code 3GGU, doi:10.1128/JVI.00451-09) and program PyMol [194-196]. Depicted mutations do not occur all together in the same isolate, this is a synopsis.

3.2.6 Impact of Gag mutations on protease inhibitor resistance

In previous studies a number of both cleavage site and non-cleavage site mutations in Gag had been described to correlate with therapy failure [197-200]. These gag mutations have been described to associate with specific drug resistance profiles in the protease such as I437T/V with L76V [201], A431V with L24I-V82A-I54V, L449F-R452S-P453L with D30N-I84V, or P453L with I84V-L90M [202]. Moreover, Gag mutations can directly impact on PI susceptibility. This has been described for mutations immediately at cleavage sites but also at non-cleavage site positions; they have been reported to occur individually or in combination with further mutations; they appear in conjunction or complete absence of (enhancing) major resistance mutations in protease [10-13].

Mechanistically, alterations in Gag can restore the replication capacity of the affected virus, which could have been compromised by non-favorable protease mutations [192, 199, 203, 204]; certain Gag mutants may also retain such a replication advantages in the complete absence of protease inhibitor pressure [205]. On the molecular level Gag cleavage site mutations can exert their effect by affecting the processing rates of Gag cleavage sites [206], whereas non-cleavage site mutations could rather act indirectly through conformational changes of the polyprotein [13, 205]. Another likely mechanism affects functions of the mature cleaved Gag proteins [205].

4 RATIONALE AND AIMS OF THE STUDY

During the last two decades an extended arsenal of highly selective and potent antiretroviral drugs has become available. This turned HIV infection into a long-term manageable chronic condition. Moreover, the recent publication of the WHO goals 90-90-90 aims at controlling by the year 2020 HIV replication in 73% of all people infected by HIV on a global scale. Today on the Northern hemisphere diagnosis of HIV infection and drug availability with successful long-term suppression of viral replication are most common. Yet, there continues to be a major discrepancy for lower income regions, where older drugs with massive side effects are still in use, or where stock-outs contribute to unavoidable therapy interruption. Although HIV and the processes of its pathogenesis appear well understood [44, 45] mechanisms of viral escape seem to persist or even become more challenging among patients who are on therapy for very long [207-209].

Initially PIs were mainly used as part of second-line regimens. Today, however, up to 50% of the first-line regimens in Germany and Switzerland include PIs [1, 2], a number that further increases for second line regimens and beyond. And although in the SHCS most patients under ART have suppressed viral loads [3], every third patient is or has been affected by drug resistances [4]. Along with mal-compliance to a treatment regimen the development of viral drug resistances represents a key cause for therapy failure.

The detailed knowledge of HIV drug resistance mechanisms is indispensable for the development of robust and cost effective suppression strategies. Phenotypic drug resistance testing allows in vitro evaluation of possibly mutated virus population from a patient. Although it deals with a direct measurement of drug susceptibility and is particularly suitable to assess complex resistance patterns like coexisting quasispecies or the presence of minority variants today's standard of care is genotyping as a cheaper and more rapid approach. Genotyping utilizes previously obtained phenotypic information on numerous clinical samples and on engineered viruses paired with their underlying sequences.

Systems designed to interpret genotypic data and predict the therapy response take into account not only the presence of single resistance-associated mutations but also information on interactions between them [210, 211]. This is why it is crucial to document in details the role of as many therapy-associated mutations as possible and to investigate statistical, temporal and causal links between them adjusting for the role of natural HIV variation as well. Mechanistically, the resistance of HIV enzymes to certain substrate analogues is typically characterized by structural alterations in the viral target protein directly at the inhibitor binding site. But in addition to that, the viral protease offers an alternative route for the development of resistance – through the natural substrates of the enzyme. Earlier studies have shown that alterations near the protease recognition sites in Gag ("cleavage sites") can accompany or are responsible for viral drug resistance. Such an alternative escape route of HIV was demonstrated by the viral response to the maturation inhibitor bevirimat, where specific modifications in the p2 motif QVT of Gag [134] were responsible for inhibitor failure [10, 212].

A number of cases of clinical therapy failure under PI-containing regimes have been reported, where genotypic resistance testing did not reveal sufficient explanation from information on the protease gene [5, 6]. Nevertheless, until today most algorithms evaluating PI resistances take solely the PR itself into account. Meanwhile, also certain Gag mutations have been associated with the development of PI resistance, either by statistical analysis [11, 205, 213] or in studies analyzing patient-derived samples after PI exposure and failure and genotyping [7-13] or assessing viral replication [192, 199, 203, 204]. Further details on the role of Gag in protease inhibitor resistance and on the mutational patterns observed in Gag-PR might provide an additional argument to consider the inclusion of the gag gene for genotyping, particularly when complex PI resistance is suspected.

Therefore in this study we pursued the following **aims**:

1) **Assess the clinical and diagnostic importance of Gag mutations;**
2) **Describe Gag resistance mutations and their patterns in the HIV-1 isolates from patients in the SHCS;**
3) **Scrutinize the phenotypic impact of observed mutations.**

5 MATERIALS AND METHODS

5.1 Analytical part

5.1.1 Sequences and treatment information

Plasma samples were collected between 2004 and 2012 from patients in two centers of the Swiss HIV cohort study: Basel (n = 2022) and Zurich (n = 773).

The ViroSeq HIV-1 Genotyping System (Abbott Molecular, Illonois, USA) was used for Sanger sequencing of the HIV-1 pol region in the routine diagnostics setting of an accredited laboratory. Pol sequences were assembled and edited using the ViroSeq Genotyping software v2.5 (Abbott Molecular). For details on the sequencing procedure, see [214].

The F-primer, integral part of the system, produces a read that extends in reverse orientation from protease into the C-terminal gag region. Although not accessible with the standard ViroSeq software, this information was manually extracted from the raw sequencing data in the form of .ab1 files and analyzed to obtain Gag C-terminal sequences, which are disregarded in the standard setting. F-primer read chromatograms were processed with DNA Baser software (Heracle BioSoft SRL). The software performed base calling, base quality assessment, automatic ambiguity correction, homopolymer error correction and low quality end trimming. Default software settings for low quality reads were applied. Resulting nucleotide sequences were reverse-complemented, codon-aligned to subtype B consensus reference sequences [104, 190] and translated in the Gag and Pol reading frames using RegaDB Sequence analysis Tools [215] and Stanford HIVdb Program [104, 190], correspondingly. Amino acid substitutions were listed. Codons with more than 4 possible translations as well as preliminary stop codons were flagged and were excluded from statistical processing. Sequences with two or more adjacent flagged codons were trimmed to remove these and all the upstream codons. Different substitutions at the same single amino acid position were treated independently. We used the list and definitions of the Stanford HIV resistance

database for primary and secondary PI resistance mutations in the protease gene [104, 190].

Sequence information on protease and the C-terminal Gag region was paired with the corresponding patient treatment history. Treatment information came from the records on the order forms, on which the indication for resistance test is provided. Also the status of treatment history is categorized for each drug as "current" "previously" or "never". Every sample with a status "current" or "previously" for at least one protease inhibitor was considered protease inhibitor treatment experienced (further referred to as TE; n = 515). Indications of category "never" were rarely used. As a consequence, there was no explicit statement that a specific TE patient was never treated with any other protease inhibitor than those marked as "current" or "previously". As another consequence, there were complications with assigning samples to the protease inhibitor treatment naïve category, so we compared the blood collection date for every sample with the FDA approval dates of protease inhibitors. In case the former date was prior to the latter date for a given sample and protease inhibitor, treatment status was switched to "never". The group of protease inhibitor treatment naïve samples included those samples with the status "current" or "previously" for none of protease inhibitors (further referred to as TN, n = 825).

5.1.2 Statistical analysis

Data analysis was performed using R language [216]. Statistical associations were assessed using Fisher's exact test with a significance level of 0.95. Prevalence of Gag and protease mutations in the viral sequences of TE patients versus TN patients was assessed. We defined and analyzed these distinct types of mutations: Primary PI resistance mutations in protease; Secondary PI resistance mutations in protease; Other protease mutations; Treatment associated Gag mutations. Comparisons were performed for subtype B viruses (n = 890; of

those TE n = 369, TN n = 521). Processing and classification of our sequences are summarized in Figure 14.

5.1.3 Pairwise associations between mutations

Fisher's exact test was used to evaluate possible statistical associations between amino acid mutations. We selected significance level of 0.95 as critical cutoff. Obtained associations were visualized in heat-map fashioned correlation graph. Only mutations and mutational pairs occurring in more than 1% of total TE samples were included.

5.1.4 Construction of mutagenetic trees

Implementation of mtreemix software by Beerenwinkel *et al.* [217, 218] in R language was performed by Bogojeska *et al.* [219]. Resulting Rtreemix package allows modeling multiple paths of ordered accumulation of genetic changes from cross-sectional data. Assuming mutations occurred are permanent, it estimates local maximum likelihood mutagenetic tree using a combination of graph-theoretical method with an expectation-maximization approach. These models have been successfully used to scrutinize HIV resistance development characterized by ordered accumulation of resistance mutations in the viral genome under drug pressure [220].

To estimate stability of fit models we performed 1000 rounds of bootstrapping and selected tree branches by the number of bootstraps they were supported with.

5.1.5 Learning Bayesian network

We used Bayesian networks in order to model the role of mutations observed along with their interactions. R package pcalg [221, 222] implements this

probabilistic model describing statistical independencies between multiple variables [223]. We approached our dataset using two pairs of algorithms. FCI (Fast Causal Inference) algorithm [224, 225] paired to GBC (Generalized Backdoor Criterion) algorithm [226] were used to strictly approach our observational data with the assumption that it contains some hidden or selection variables. Alternative pair included PC (Peter-Clark) [225] and IDA (Intervention calculus when DAG is Absent) [227] algorithms with more relaxed assumption that our data contains no hidden and selection variable. Constructed models are visualized with directed acyclic graphs in which dependencies can be represented with edges. Applied algorithms extract the network capable of explaining a maximum of statistical correlations between the variables in the data using minimum edges. Binary representations of amino acid sequences labelled with a parameter for PI exposure were fed to the program.

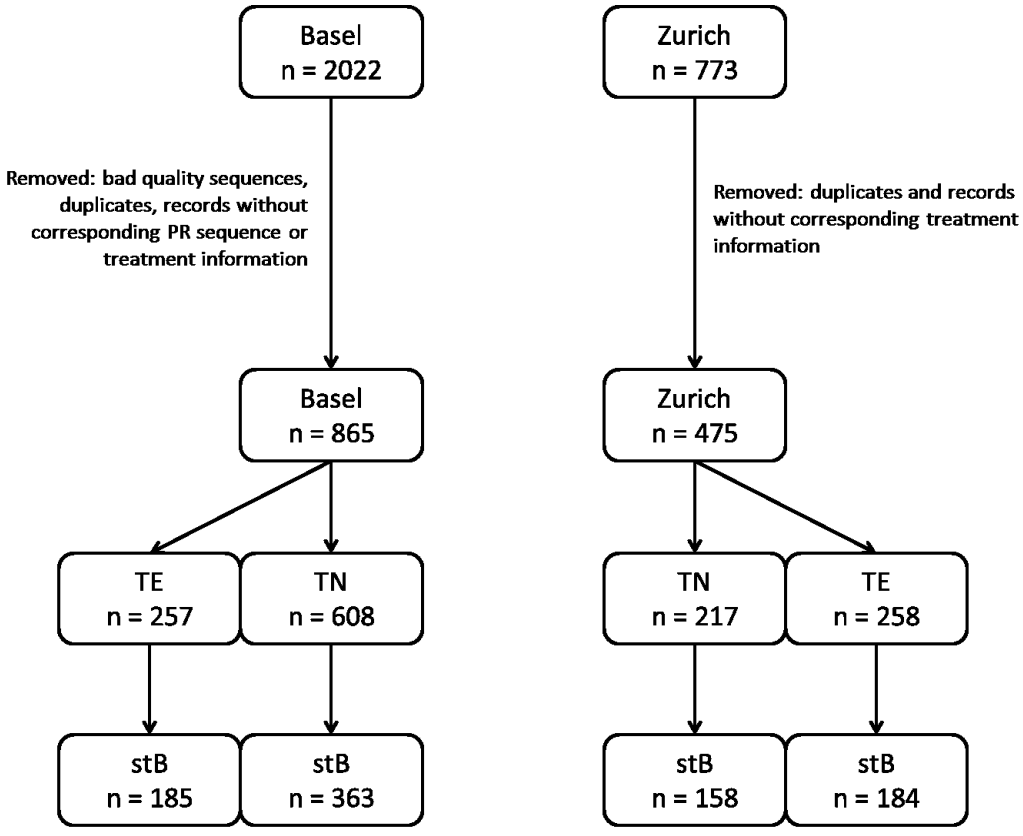


Figure 14. Processing and classification of the sequences obtained. TE – protease inhibitor treatment experienced, TN – protease inhibitor treatment naïve, stB – subtype B.

5.2 Experimental part - materials

5.2.1 Chemicals

CHEMICAL	SUPPLIER
PCR	
dNTPs (dATP, dCTP, dGTP, dTTP), 10mM	Sigma
PfuUltra II Fusion HS DNA Polymerase	Agilent
PfuUltra II Reaction Buffer, 10x	Agilent
Gel electrophoresis	
Agarose	Cambrex
TBE buffer, 10x	Amresco
Ethidium bromide solution (10mg/ml)	Sigma
1kb DNA ladder (1µg/µL)	Invitrogen
100bp DNA ladder (1µg/µL)	Invitrogen
DNA isolation and purification	
NucleoSpin® Plasmid / Plasmid (NoLid)	Macherey-Nagel
NucleoSpin® Gel and PCR Clean-up	Macherey-Nagel
Bacterial Culture, Competent Cells Preparation	
Bacto Agar (dehydrated)	Becton-Dickinson
Bacto Tryptone (dehydrated)	Becton-Dickinson
Bacto Yeast Extract (dehydrated)	Becton-Dickinson
NaCl	Fluka
Ampicillin (sodium salt)	Sigma
Glycerol (87%)	Fluka
$CaCl_2$ dihydrate	Fluka

One Shot® TOP10 cells	Invitrogen
HB101 cells	Promega

Cloning

BamHI	New England Biolabs
XmaI	New England Biolabs
Hind III	New England Biolabs
BssHII	New England Biolabs
Digestion buffers, 10x	New England Biolabs
Bovine Serum Albumin (BSA), 10x	New England Biolabs
Alkaline Phosphatase, Calf Intestinal (CIP)	New England Biolabs
Quick Ligation Kit	New England Biolabs

Sequencing

BigDye® Terminator v3.1 Cycle Sequencing	Applied Biosystems
BigDye® v1.1/3.1 Sequencing Buffer (5X)	Applied Biosystems

Cell culture

DMEM High Glucose (4.5g/L) with Stable Glutamine	BioConcept
RPMI 1640, with 25mM HEPES (w/o L-Glutamine)	BioConcept
L-Glutamine, 200mM (100x), liquid	Gibco
jetPRIME® transfection reagent	Polyplus Transfection
jetPRIME® buffer	Polyplus Transfection
Fetal Bovine Serum (Heat Inactivated)	Gibco
Trypsine/EDTA (w/o Ca2+/Mg2+)	Gibco
D-PBS (1x), liquid (w/o Ca2+/Mg2+)	Gibco
Trypan Blue Stain, 0.4%	Gibco

Pen/Strep	BioConcept

Virus Inactivation and Cells Fixation

Formaldehyde (36.5%)	Fluka
Glutaraldehyde (25%)	Fluka

ONPG assay

Buffer Z	NA
Buffer H	NA
o-Nitrophenyl-β-D-Galactopyranoside (ONPG)	Amresco
β-Mercaptoethanol (100%)	Fluka

Antiretrovirals

Atazanavir	Bristol-Myers Squibb
Darunavir	Tibotec
Indinavir	Abbott Laboratories
Lopinavir	Abbott Laboratories
Nelfinavir	Abbott Laboratories
Saquinavir	Abbott Laboratories
Amprenavir	Abbott Laboratories
Efavirenz	Abbott Laboratories
T-20	Abbott Laboratories
Dimethyl sulfoxide (DMSO)	Riedel-de Haën

General chemicals

Ethanol (100%)	Fluka
Sodium hypochlorite (10%)	Fluka
Dismozon	Bode Chemie

5.2.2 Cell lines

CELL LINE	DESCRIPTION	SOURCE
HeLa	Human adherent transfection cell line	Prof. Dr. Thomas Klimkait
293T	Human semi-adherent transfection cell line	
SxR5	HeLa – derived reporter cell line. Contains integrated plasmid with an HIV-1 LTR driving bacterial LacZ gene. Expresses CD4 receptor as well as CXCR4 and CCR5 co-receptors.	

5.2.3 Plasmids

PLASMID	DESCRIPTION	SOURCE
pNL 4-3	Provirus derived from NY5 (5') and LAV (3') HIV-1 isolates, cloned into pUC18 into the PvuII site (size: 14,877bp) [228]	Dr. Malcolm Martin (NIH)
pNL-NF	pNL 4-3 truncated in the flanking regions (cut with NaeI and FspI)	Prof. Dr. Thomas Klimkait
pNotI5-BX	Cloning cassette containing: BssHII, NotI/1141, BamHI/2424 and XmaI/2796	Prof. Dr. Thomas Klimkait

All mutants of interest were generated using the pNL-NF scaffold by overlap-PCR. Then generated and properly prepared Gag-PR inserts were ligated into the accordingly prepared cloning cassette pNotI5-BX.

The pNL-NF plasmid represents a shorter version of pNL4-3, the wild-type reference of an HIV-1 B subtype virus, in which the human flanking sequence had been reduced to a minimum. PNotI5-BX is a pNL-NF based cloning cassette for Gag and PR. The two restriction sites BssHII and XmaI of pNotI5-BX allowed the direct insertion of the mutated Gag-PR fragment. Since all plasmids contain as backbone the pUC18 plasmid, they can readily be amplified in *E. coli* and confer ampicillin resistance upon transformation.

5.2.4 Mutants

Mutations of interest were identified using analytical approach and introduced into standardized pNL-NF backbone using the overlap mutagenesis method. Amino acid substitutions were codon-optimized.

Protein	Mutation	Wild-type	Substitution	Nucleotide position	Nucleotide change	Number of nucleotide changes	Human codon usage
Gag	T427D	T (Threonine)	D (Aspartic Acid)	2068-2070	ACT->GAT	2	21.8
Gag	R429I	R (Arginine)	I (Isoleucine)	2074-2076	AGA->ATC	2	20.8
Gag	A431V	A (Alanine)	V (Valine)	2080-2082	GCT->GTG	2	28.1
Gag	I437V	I (Isoleucine)	V (Valine)	2098-2100	ATC->GTG	2	28.1
Gag	Y441Q	H (Histidine)	Q (Glutamine)	2110-2112	CAC->CAG	1	34.2
Gag	L449P	L (Leucine)	P (Proline)	2134-2136	CTT->CCT	1	17.5
Gag	L449V	L (Leucine)	V (Valine)	2134-2136	CTT->GTG	2	28.1
Gag	S451H	S (Serine)	H (Histidine)	2140-2142	AGC->CAC	2	15.1
Gag	S451N	S (Serine)	N (Asparagine)	2140-2142	AGC->AAC	2	19.1
Gag	R452S	R (Arginine)	S (Serine)	2143-2145	AGA->AGC	1	19.5
Gag	P453L	P (Proline)	L (Leucine)	2146-2148	CCA->CTG	2	39.6
PR	I47V	I (Isoleucine)	V (Valine)	2391-2393	ATA->GTG	2	28.1
PR	I54V	I (Isoleucine)	V (Valine)	2412-2414	ATC->GTG	2	28.1
PR	V82A	V (Valine)	A (Alanine)	2496-2498	GTC->GCC	1	27.7

5.2.5 Primers

All primers were named according to their 5' end position preceded by an "F" for forward or an "R" for reverse primers. This notation allows calculating directly the size of a fragment from the name of primers.

Name	Purpose	Sequence 5'-3'	Basic Tm, °C
"Border" primers for site-directed mutagenesis			
F-707L	BssHII RS	TGAAGCGCGCACRGCAAGAGGCGAG	64.2-65.9
R-2598	XmaI RS	CCATCCCGGGCTTTAATTTTACTGG	57.7
Mutagenesis primers			
F-2061-T427D	T427D	AGA TTG TGA TGA GAG ACA GGC TA	53.5
R-2083-T427D	T427D	TAG CCT GTC TCT CAT CAC AAT CT	53.5
F-2062-R429I	R429I	GAT TGT ACT GAG ATC CAG GCT AAT	54
R-2085-R429I	R429I	ATT AGC CTG GAT CTC AGT ACA ATC	54
F-2074	A431V	AGACAGGTGAATTTTTTAGGGAAGA	52.8
R-2094	A431V	CCCTAAAAAATTCACCTGTCTCTCAGTACAA	59.1
F-2084M	I437V	ATTTTTTAGGGAAGGTGTGGCCTTCC	58
R-2105	I437V	GGCCACACCTTCCCTAAAAAATTAGCCTGT	61.6
F-2099-Y441Q	Y441Q	TC TGG CCT TCC CAG AAG GGA A	56.3
R-2119-Y441Q	Y441Q	TTC CCT TCT GGG AAG GCC AGA	56.3
F-2128	L449P	AATTTTCCTCAGAGCAGACCAGAGC	57.7
R-2142	L449P	GCTCTGAGGAAAATTCCCTGGCCTT	59.3
F-2128-L449V	L449V	AATTTTGTGCAGAGCAGACCAGAGC	57.7

R-2142-L449V	L449V	GCTCTGCACAAAATTCCCTGGCCTT	59.3
F-2131-S451H	S451H	TTT CTT CAG CAC AGA CCA GAG	52.4
R-2151-S451H	S451H	CTC TGG TCT GTG CTG AAG AAA	52.4
F-2131-S451N	S451N	TTT CTT CAG AAC AGA CCA GAG CCA	55.7
R-2154-S451N	S451N	TGG CTC TGG TCT GTT CTG AAG AAA	55.7
F-2137-R452S	R452S	CAGAGCAGCCCAGAGCCAACAGCCC	65.9
R-2158-R452S	R452S	CTGTTGGCTCTGGGCTGCTCTGAAGAAAA	62.9
F-2140	P453L	AGC AGA CTG GAG CCA ACA GCC C	64.2 (60.4)
R-2158	P453L	C TGT TGG CTC CAG TCT GCT CTG AAG AAA A	57.7 (61.5)
F-2383-I47V	I47V	CAA AAA TGG TGG GGG GAA TTG GA	55.3
R-2405-I47V	I47V	TCC AAT TCC CCC CAC CAT TTT TG	55.3
F-2404	I54V	GAGGTTTTGTGAAAGTAAGACAGTATGATC	57.5
R-2424	I54V	GTCTTACTTTCACAAAACCTCCAATTCCCC	60.3
F-2483	V82A	AGGACCTACACCTGCCAACATAATTG	58
R-2504	V82A	TATGTTGGCAGGTGTAGGTCCTACTAATAC	60.3

Sequencing primers

F-620	BssHII region	GGA AAA TCT CTA GCA GTG GCG	54.4
F-1400	Gag region	CCA TCA ATG AGG AAG CTG CAG	54.4
F-1985	Gag-Pol region	TTA AGT GTT TCA AYT GTG GCA ARG AAG G	55.5-58.5
F-2084	XmaI region	ATT TTT TAG GGA AGA TCT GGC	48.5

5.3 Experimental part – methods

5.3.1 Overlap PCR

PCR-reactions were typically performed in a 50 µL volume. Primer stocks were stored at -20°C at a concentration of 100 µM. Working primer solutions were prepared at 10 µM. For each reaction 2µL of each primer are used, corresponding to 20pmol (final concentration of 400nM). The amount of DNA template ranged from 10 to 50 ng. Additionally 1.25 µL of 10mM dNTPs, 5 µL of 10x polymerase reaction buffer and 1 µL of polymerase were added. The volume then was filled up to 50 µL with autoclaved mQ water.

A standard amplification cycle is set up as it follows (in total 30 cycles):

Step	Time	Temperature, °C	
Denaturation	2’	98	
Denaturation	20’’	98	
Annealing	20’’	Primer Tm – 5°C	30 cycles
Elongation	30’’-60’’	72	
Elongation	3’’	72	
Pause	-	4	

The annealing temperature was selected according to the primer pair used. The extension time was adjusted according to the fragment amplified. Reactions were set up on ice to prevent unspecific primer annealing to the template.

The first step of overlap PCR included to parallel reactions on the same template in different tubes (Figure 15). One reaction used forward flanking and reverse mutagenic primer, another – forward mutagenic and reverse flanking primer. Mutagenic primers were designed so that they overlap on at least half of their annealing site length and introduce the same nucleotide changes from the selected template.

The second step included a single PCR reaction with purified products of the first step as a template and flanking primers (Figure 16).

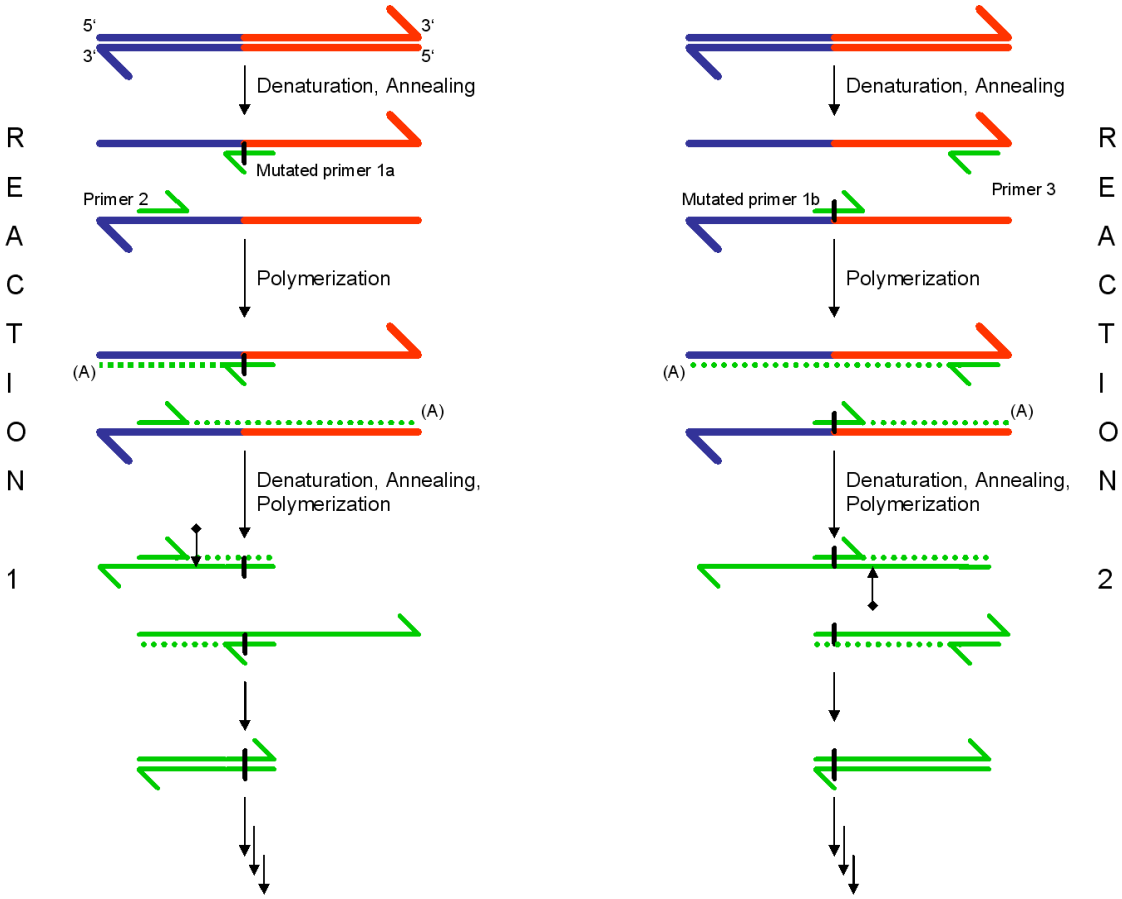


Figure 15. First step of the mutagenesis by overlap extension PCR. Primer 2 and Primer 3 – are flanking primers. Mutated primer 1a and mutated primer 1b – are the primers introducing desired mutations. Black bold mark indicated nucleotide difference introduced. Blue and red colors indicate template target regions for the two independent reactions at this stage. Green lines indicate products of the reaction while green dots visualize elongation of the primer on a given template. Illustration by Alessio Cremonesi [229].

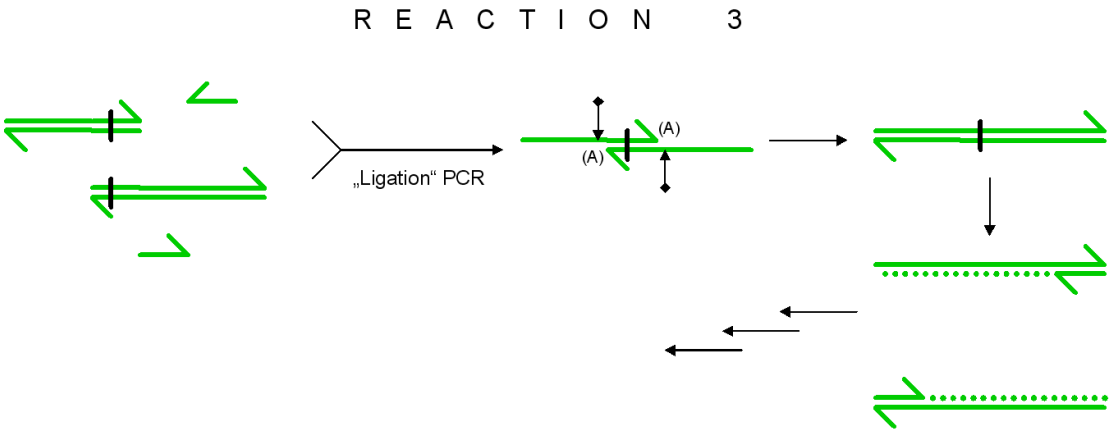


Figure 16. Second step of the mutagenesis by overlap extension PCR. Black bold marks indicate the nucleotide difference introduced. Green lines indicate products of the reaction while green dots visualize elongation of the primer on a given template. Illustration by Alessio Cremonesi [229].

5.3.2 Bacterial culture

LB medium with ampicillin: 5g NaCl, 10g Bacto Tryptone, and 7g Bacto Yeast Extract are dissolved in 1L milliQ H_2O and autoclaved. When the solution is at room temperature, 1mL ampicillin (200mg/mL) is added: final antibiotic concentration is 200µg/mL.

LB agar plates with ampicillin: 2.5g NaCl, 5g Bacto Tryptone, 3.5g Bacto Yeast Extract, and 6g Bacto Agar are dissolved in 0.5L milliQ H_2O and autoclaved. After cooling down the solution, 0.5mL ampicillin (200mg/mL) are added: final antibiotic concentration is 200µg/mL. Approximately 20mL medium are poured in each Petri dish.

5.3.3 Preparation of competent bacterial cells

Both HB101 and One Shot® TOP10 competent cells are AmpS and contain a recA mutation, which prevents undesirable recombination events. In addition, One Shot® TOP10 bacteria have an endA mutation that prevents carry-over of nucleases. Both cells are chemically competent.

Cells are plated on LB agar w/o ampicillin and incubated over night at 37°C. The next day one single colony is inoculated in 5mL LB medium and incubated overnight on a shaker. From the overnight culture, 5mL are transferred into 1L LB medium and let grown till OD260 of 0.2-0.5.Afterwards cells are transferred in 50mL tubes and put immediately on ice for 10'. Centrifugation is done at 4°C for 20' (2,500rpm). Supernatant is removed and cells are resuspended in 25mL ice cold 100mM $CaCl_2$ and centrifuged under the same conditions as before. After resuspending cells in 10mL ice cold 100mM $CaCl_2$, they are put on ice for 30' and then centrifuged. Following supernatant removal, cells are well mixed in 50mL ice cold 100mM $CaCl_2$ with 10% glycerol and aliquoted into PCR tubes, which are stored at -80°C. Bacteria are plated on LB agar with and w/o ampicillin to check for contaminations and concentration; moreover a transformation test with 10ng pNL-NF is done to check their transformation efficiency.

5.3.4 Plasmid DNA purification

DNA plasmid extraction is performed with Macherey-Nagel kit according to the enclosed protocol. For a miniprep, the starting amount of LB culture is 4 mL, for a midiprep 100mL, and for a maxiprep 250mL. The DNA content is quantified by UV spectrometry at 260nm using NanoDrop® ND-1000. Typically A260/A280 and A260/A230 ratios (for DNA 1.8 and 1.8-2.2, respectively) are monitored to estimate DNA purity.

5.3.5 Gel extraction

PCR products and prepared vector fragments (backbones and inserts) are extracted from agarose gel with Macherey-Nagel kit according to the enclosed protocol. The DNA concentration is normally not measured by UV spectrometry at 260nm, since the measurement is not very reproducible, due to low yield. At low

concentrations this measurement has a qualitative rather than quantitative character.

5.3.6 Vector preparation

Typically 2-4 µg of DNA were digested using 1-5 overdigestion as in 20 or 50 µl reaction as recommended by the enzyme manuals. High enzyme and glycerol (>5% v/v) concentrations can cause star activity and were therefore avoided. In case enzymes had different optimal temperatures they were added to the reaction mix and incubated sequentially for an hour each. Afterwards 1 µl of CIP phosphatase is added to remove 5´ phosphates from the cut plasmid, which prevents vector re-circularization. The reaction is incubated for 30' at 37°C. Then the sample was run in agarose gel and the band of expected size was excised and gel-purified.

5.3.7 Cloning and transformation

All inserts were digested for 1 hour with the appropriate combination of restriction enzymes at optimal conditions according to the enclosed manuals. The enzymes were heat-inactivated. The theoretical molar ratio between insert and vector should be approximately 3:1. Nevertheless, better results were obtained with higher ratios. In general 50-70ng of vector were used together with 2µL insert at 5-10ng/µL. An equivalent volume of 2x Quick ligation buffer was added and the reaction is incubated at 25°C for 20'. A negative control containing the vector alone is always performed to estimate the background of vector self-ligation.

Afterwards, 100µL of fresh thawed competent cells (either HB101 or One Shot® TOP10) were added to the ligation mix and the tubes are incubated on ice for 30' (bacteria were resuspended only few times to avoid mechanical lysis). Bacteria were heat-shocked at 42°C for 1'' and put back on ice for 10'. Depending on the aim of the experiment, bacteria are either put in liquid LB medium or plated

on LB agar. On one side, if a mixture of different inserts is cloned and has to be preserved, bacteria were incubated in 4mL liquid LB medium with ampicillin and incubated overnight at 37°C on a rocking platform. On the other side, if single clones had to be isolated after heat-shock, bacteria were plated on LB agar plates containing ampicillin and incubated overnight at 37°C. The next day colonies were picked up and grown in liquid culture.

5.3.8 Restriction digestion

Usually 1µL of plasmid DNA preparation (approximately 300ng/µL) was digested with 0.3U of each enzyme, 1µL of appropriate buffer 10x, 1µL of BSA 10x, and H2O up to total volume of 20µL. The reactions were incubated at appropriate temperatures for 1 hour. Subsequently samples were run on agarose gel and obtained restrictions patterns were compared to the ones expected.

5.3.9 DNA sequencing

Sequencing was performed in-house using Applied Biosystems 3130 Genetic Analyzer and the corresponding sequencing kit. Sequencing primers were purified through HPLC by the manufacturer. Sequencing reaction mix included:

5x Sequencing Buffer	3.0 µL
1mM primer	3.0 µL
5x Big Dye Terminator v3.1	1.5 µL
Sample	200-600 ng, typically 1.0 µL
mQ H_2O	Up to 20 µL

The PCR cycling is set up as follows:

Step	Time	Temperature, °C	
Denaturation	5’	95	
Denaturation	30’’	95	
Annealing	10’’	55	40 cycles
Elongation	4’	60	
Pause	-	4	

For subsequent purification 96 well filtration plates were used. They were filled with Sephadex and 300 µL mQ H_2O per well. Sephadex was let to swell for three hours at room temperature. Afterwards excess water was removed by centrifugation, and samples loaded on the filtration plate. During subsequent centrifugation samples were collected in 96 well sequencing plate and loaded to the sequencer.

5.3.10 Cell culture

Hela and SxR5 are adherent cells and 293T are semi-adherent cells. They were grown in DMEM High Glucose (4.5g/L). All cells were split three times a week. Before splitting, cells were examined by microscopy to check for confluence and possible contamination. For passaging medium was removed and cells were washed with PBS w/o Ca^{2+} and Mg^{2+}, and trypsinized with trypsin-EDTA. After incubation at 37°C for 5’ cells normally detached, were resuspended in DMEM and counted. Appropriate numbers of cells were resuspended in a flask with fresh medium. The cells were maintained until passage 20, before a new aliquot was thawed.

All three cell lines are stored in 1mL aliquots at -196°C (liquid nitrogen). They are thawed gently but at the same time quickly since cells are in a freezing solution containing 10% DMSO. Freshly thawed cells are added to culture medium, centrifuged, resuspended and finally incubated at 37°C and 7% CO_2 in fresh medium.

5.3.11 Preparation of antiretrovirals

Drug stocks were prepared from pills and then were diluted in two formats, either 6 or 10 concentrations depending on the experiment. Drugs were diluted according to their solubility in polar solvents. Moreover, each drug has its own range of concentrations depending on the IC50 value, in order to accurately extrapolate the inhibition curve.

5.3.12 deCIPhR

The HIV drug resistance phenotyping assay applied in the study was the deCIPhR system (dual enhancement of cell infection to phenotype resistance) [230, 231] as available through Th. Klimkait. It represents a proprietary assay of InPheno AG, permitting viral replication during assessment. Briefly (Figure 17), each proviral DNA is transfected into a human epitheloid cell line (HeLa or 293T) using a transfecting agent (jetPrime) leading to production of fully infectious HIV-1 particles. Cell-to-cell spread and replication of recombinant viruses is allowed for a period of four days in the absence or presence of specific drugs by co-culture with a reporter cell line (SxR5) expressing CD4 and both chemokine receptors CXCR4 and CCR5. In addition, the genome of this cell line contains the HIV-1 Long terminal repeat (LTR) fused upstream of the bacterial reporter gene LacZ coding for β-Galactosidase. Therefore in these cells, the activity of β -Galactosidase is proportional to the extent of viral replication. In the final step of the process, the cells are lysed to incubation with a chromogenic substrate for β -Galactosidase, ortho-nitrophenyl-galactopyranoside (ONPG).

In a sterile 2 mL tube, 1 μg of plasmid DNA (e.g., pNL4-3 or derivatives) are added to 100 μL of jetPRIME® buffer and mix by vortexing. The 2μL of jetPRIME® reagent are added on top and vortexing is performed. The solution is incubated at room temperature for 10’. A master mix is prepared for several

transfections. HeLa or 293T cells are harvested by trypsinization, counted, and an aliquot of 0.25 x106 cells in 500 µL of complete DMEM is set aside for each transfection. At the end of the incubation period, the cell aliquot is added to the transfection mix and the 2mL tube is returned to a 37°C cell incubator.

In a 96-well microtiter plate (96-w MTP) 10 µL of a 20x dilution of the test drug(s) (in cDMEM/DMSO) are added. In plate format positive and negative controls i.e., wells containing a reference inhibitor (e.g., 300 nM Efavirenz = 0% readout) or diluent (= 100% readout), are always included. SxR5 cells are harvested and counted (BSL2 cell culture lab). Then transfected cells are mixed with $1.1x10^6$ SxR5 cells in 42 mL of fresh medium. After that 190 µL of the mix are distributed to each well of two 96-w MTP with drug dilutions. Plates are returned into the incubator for 4 days (BSL3).

Then β-galactosidase development is performed: 96-well tissue culture plates are removed out of cell culture incubator and culture media is aspirated. Then 10 µL of Glo Lysis Buffer is added per well. Plates are cleaned with bleach from outside and on the inner surface of the lid and transferred to the main lab. Eighty µL of ONPG solution are dispensed into each well. The plates are read using reader set at 405nm wavelength. The optimal maximal absorbance is at least 0.4 to 0.8. The data obtained is analyzed using an Excel spreadsheet template with XLFit as add-in. In this template, raw data (absorbance at 405 nm) are converted into percent inhibition using the following formula: InhX = 100 – (((ReadoutX – Readout0%) / (Readout100% - Readout0%)) x 100), where: InhX – percent of inhibition; ReadoutX - absorbance of well containing substance ‘X’; Readout0% - absorbance average of 6 wells containing cells incubated in reference inhibitor (e.g., 300 nM Efavirenz); Readout100% - absorbance average of 6 wells containing cells incubated in diluent. Averages of all triplicates are then computed and XLFit determines EC50 and generates a graph plotting percent of inhibition as a function of concentration.

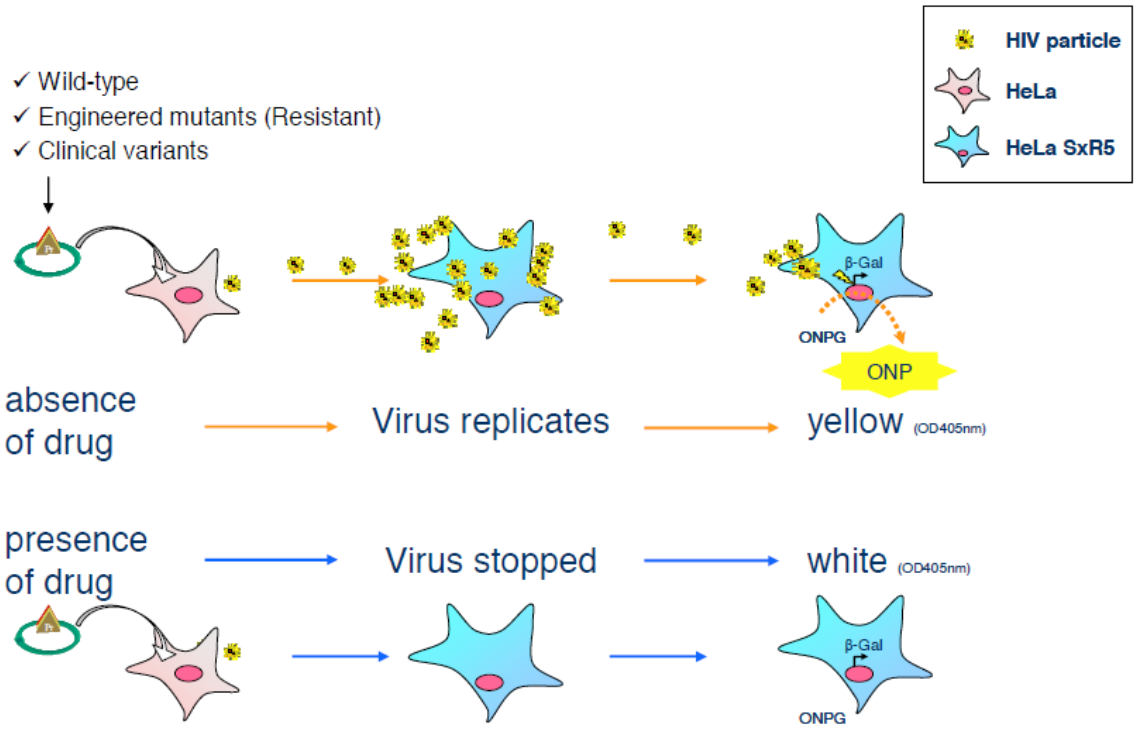


Figure 17. Principle of dual-enhancement of Cell-Infection for Phenotyping Resistance (deCIPhR). Illustration was kindly provided by Th. Klimkait [230, 231].

6 RESULTS

6.1 Prevalence of protease inhibitor resistance in protease

In order to validate our analytical approach we examined in a first step the prevalence of well-established PI resistance mutations for therapy-experienced (TE) and therapy-naïve (TN) samples. Among the collection of TE viruses 24.9% carried any kind of primary resistance mutations in the protease gene. This is in a good agreement with published data from Germany, reporting 30% (473 / 1586; $p>0.05$) [232] of resistance-associated changes. Eight primary resistance mutations were found to be significantly overrepresented in the TE over TN group: D30N, V32I, M46L/I, I54V, V82A, I84V, and L90M. In line with previous reports [233, 234], the frequency of none of these primary PI resistance mutations individually exceeded 4% in the TN group.

Also nine secondary PI resistance mutations in protease were found to significantly associate with the TE group of samples: L10F/I, L33F, Q58E, L63P, A71V/I, 73S, and N88D. For most of the secondary PI resistance mutations the frequency did not exceed 10% in TN group. Only the known polymorphism L63P occurred, as seen before, in 51.3% [104, 190, 235, 236], and the alterations A71V and L10I were identified in 12.9% and 10.7% of TN samples. Frequencies of these two mutations among TN patients of the SHCS have been reported to be 8.9% (108 / 1208; $p>0.05$) and 9.7% (118 / 1281; $p<0.05$) [236].

6.2 Correlation between PI exposure and Gag mutations

We assessed the prevalence of Gag mutations that have previously been reported to associate with PI-exposure or -resistance. Among the TE samples in our study 84.2% carried at least one of 48 known Gag mutations. Verheyen *et al.* reported the rate of known Gag mutations to be as high as 65.3% [202], but considered solely cleavage site mutations at p7/p1 and p1/p6.

Also other correlates of PI treatment with Gag changes were verified in the Swiss dataset: A431V (8.2%; OR=3.4), I437V (6.5%; OR=3.2), P453L (13.1%; OR=2.3), I479I (64.4%; OR=1.34). However, only for mutations A431V, I437V and P453L a phenotypic proof of PI resistance by in vitro mutagenesis has been published [10, 192, 237]. Among all analyzed TE sequences in our data set 17.9% carried at least one of the latter three mutations compared to the significantly lower number of 8.3% of the TN samples in the set. These figures are overall in agreement with the 19.1% (43 / 225; $p>0.05$) for TE and 12.1% for TN samples reported by Verheyen *et al*., [202] (33 / 275; $p>0.05$).

When assessing the association of those mutations with primary resistances in protease, only mutation A431V was found to correlate in the vast majority of cases (93.8%). Mutations P453L and I437V occurred along with primary PI resistances in 60.5% and 42.9%, respectively, and only 26.4% of those samples with I479I carried any known protease resistance mutation.

Like secondary PI resistance mutations in protease, several Gag alterations have been described in phenotypic in vitro experiments as being capable of compensating for fitness loss and of cooperatively decreasing PI susceptibility when they occurred in combination with primary PI resistance mutations [10-13]. We therefore assessed novel gag mutations emerging during PI exposure: Mutations T427D/N (10.4% vs 3.9% in TN; OR=2.9) and E467V/K (3.6% vs 0.8% in TN; OR=4.8) associated specifically with PI treatment. In addition, a link to the exposure to certain drugs could be demonstrated for these alterations: T427D/N to LPV (13.2%; OR=3.1), E467K to NFV (3.8%; OR=3.5), Q474H to DRV (5%; OR=3.5), and Y484P to DRV (5%; OR=5.3). With respect to combined Gag-mutations the occurrence of mutation Q474H correlated with the emergence of T427N (13.3%; OR=24.4) in our dataset.

6.3 Gag alterations in p2 and p7

For a subset of samples from Basel the analyses yielded longer gag sequences, which cover also the p24/p2 cleavage site. Among these TE samples two novel alterations were identified to be of potential relevance: A360S/P (7% = 6/86 vs 1.2% = 3/255; OR=6.3) and Q369L (8.1% = 7/86 vs 1.6% = 4/255; OR=5.5). Additionally, we found Q369L to be associated with other, established correlates of PI exposure: K418R (42.9%; OR=5.8), I437V (28.6%; OR=10.9) and P453T (42.9%: OR=15.0).

6.4 PI usage at the study centres

In the group of TE samples the respective frequencies of certain PIs were assessed. Three compounds were significantly more often applied in the Basel center (BS): NFV (60% vs 28.8% in ZH; OR=3.7), SQV (35.1% vs 12.5%; OR=3.8), IDV (54.1% vs 23.9%; OR=3.7). In contrast, Zurich (ZH) patients had been significantly more often exposed to ATV (36.4% vs 23.2% in BS; OR=0.5) and DRV (39.2% vs 8.3%; OR<0.1). This significant difference in the use of certain PI strongly hints differences in the choice of treatment regimens between the two centers, but it is likely that this also reflects differences in the average year of sampling since ATV and particularly DRV reflect more recent drugs than IDV or SQV.

6.5 Pairwise association of Gag and protease mutations in the group of TE sequences

In the TE sample group we observed links between 11 amino acid positions in Gag C-terminal region and 18 – in PR (Figure 18). In total we observed 188 associations between certain amino acid substitutions. And not surprisingly, the

most interconnected were residues in PR which resulted in the hottest region on the correlation heat map (Figure 19).

An absolute majority of protease resistances exhibited strong well-described positive interconnections with other alterations in protease. And even among other primary resistances L90M, M46I, V82A and I54V stood out with 24, 20, 19 and 18 observed significant links correspondingly.

The secondary PI resistance mutations in protease A71V and L33F were the most interconnected in their class with 19 significant associations for each of them. While most of their links were confounded with protease region, key Gag resistance mutations P453L and A431V were their only correlates in Gag.

Located at the p7/p1 and p1/p6 cleavage sites, these two mutations were the most interconnected of all Gag alterations with 18 significant correlations for each. Polymorphic L449P had 3 connections with mutations at the same p1/p6 cleavage site: S451G, P453T and P453L. I479I (previously reported to be PI-selected [203]) had as the only association the mutation of E468G.

Novel correlates of PI exposure observed were almost not interconnected. Their links were limited exclusively with other alterations in Gag. T427D/N was linked to E467V/K, Q474H and Y484P. For E467V/K we report association with T427D/N, S451G and P453L. Associated together, both Q474H and Y484P were linked to T427D/N and S451N.

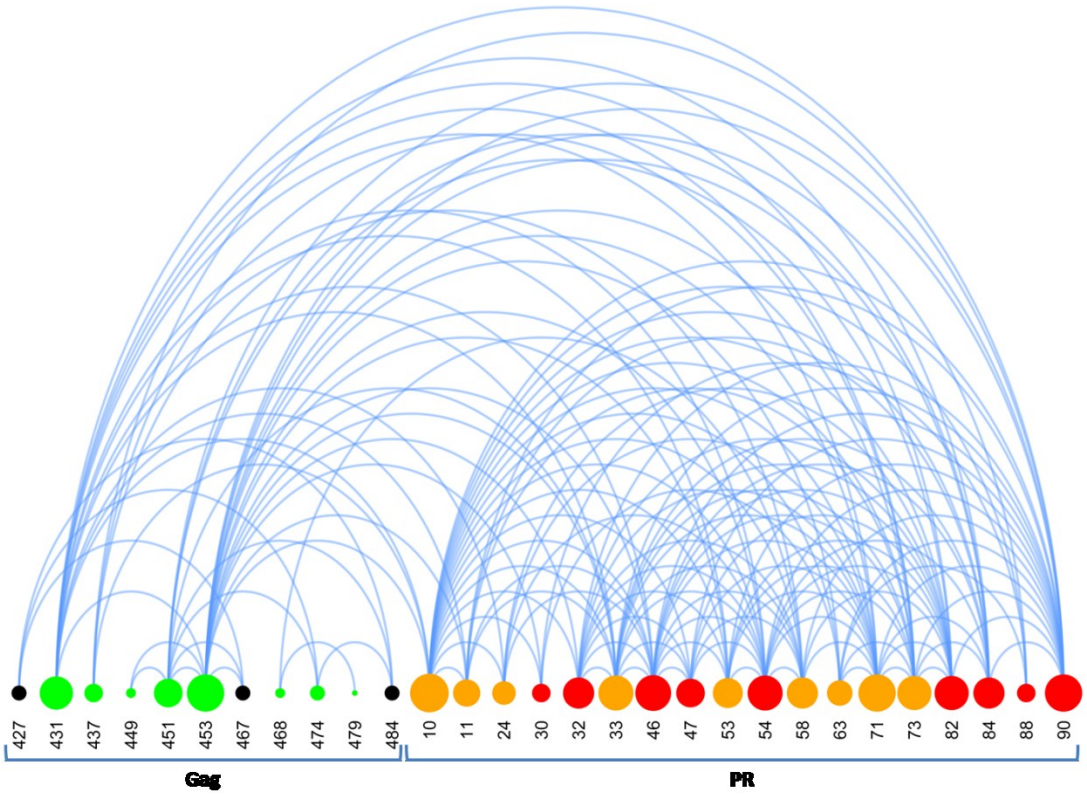


Figure 18. Arch diagram of covariation between Gag and protease. The line between two nodes indicates a covariation of corresponding residues. Node sizes are proportional to the degree of interconnection of a given residue. Red color indicates positions at which primary resistance mutations in protease have been described; yellow – secondary resistance mutations in protease; green – known Gag resistance- or treatment-associated mutations; black – newly identified treatment-associated mutations in Gag.

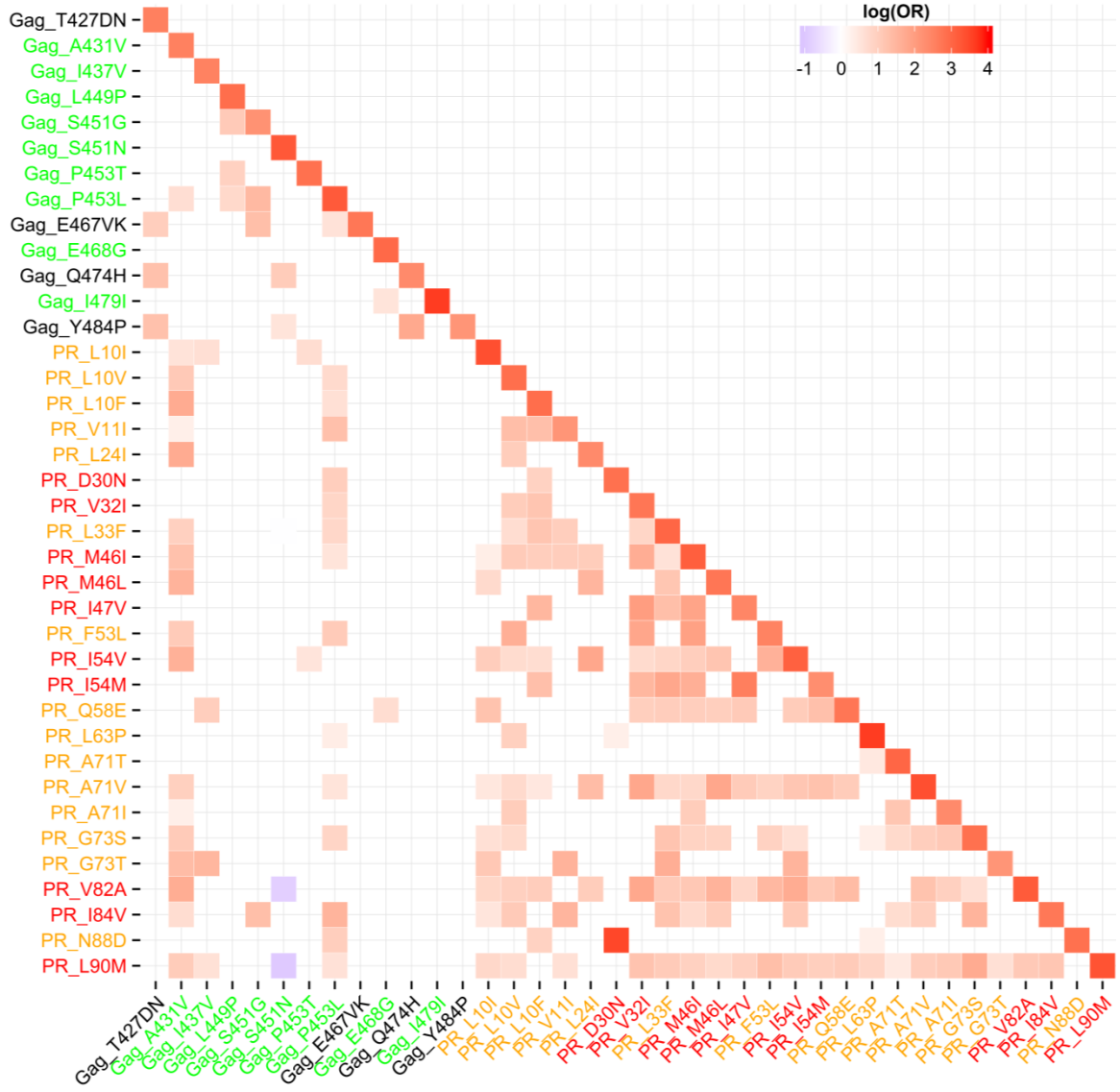


Figure 19. Heat map representing pairwise correlations in the Gag-PR region considered. Alterations were ordered according to their positions. Red labels indicate primary PI resistance mutations in protease; orange labels – secondary PI resistance mutations in protease; green labels – established correlates of PI exposure in Gag; black labels – novel correlates of PI exposure in Gag. Only significant associations were presented with colored squares. Alterations with no significant correlations were excluded. Odds ratios (OR) were plotted in logarithmic scale.

6.6 Patterns and predicted order of accumulation of mutations

We investigated clusters of mutations occurring among TE samples in a more detailed way. Multiple paths of ordered accumulation of genetic changes were modelled using random tree mixture approach. Those with the best bootstrap support and topology well corresponding to the observed pairwise associations were visualized as mutagenetic trees presented in Figure 20 and Figure 21.

The first model (Figure 20) explained up to 68% of observed sequence variants. Up to 38% of samples corresponded to the unordered model of accumulation of Gag and protease alterations or to the models that could not be robustly identified from our dataset. Second tree (Figure 21) was estimated to cover up to 56% of observed sequence variability leaving up to 50% for unordered or not observed mutational pathways.

Newly identified correlates of PI exposure in Gag were incorporated by the algorithm into the mutational pathways along with primary and secondary PI resistance mutations in protease and established treatment-associated alterations in Gag. Both trees obtained demonstrated the tendency of resistance-related mutations in protease to prime the occurrence of Gag alterations.

6.7 Bayesian networks analysis

We approached our data with two different pairs of algorithms to infer Bayesian networks. One, stricter, suggested our data contained hidden and selection variables while another, more relaxed, suggested the opposite. Nevertheless, an inferred background association skeleton was identical in both cases. Fourteen identified edges just supported observed pairwise associations among primary resistance mutations in protease (Figure 19) but did not provide any additional data on top of it.

We expected treatment exposure parameter to be directly linked to and identified as the cause for the presence of primary PI resistance mutations in

protease. Secondary PI resistance mutations in protease and established Gag resistances were expected to link to the PI exposure through the primary resistance mutations in protease. In contrast to our expectations, we could not observe direct causal links between the PI exposure and primary resistance mutations in protease.

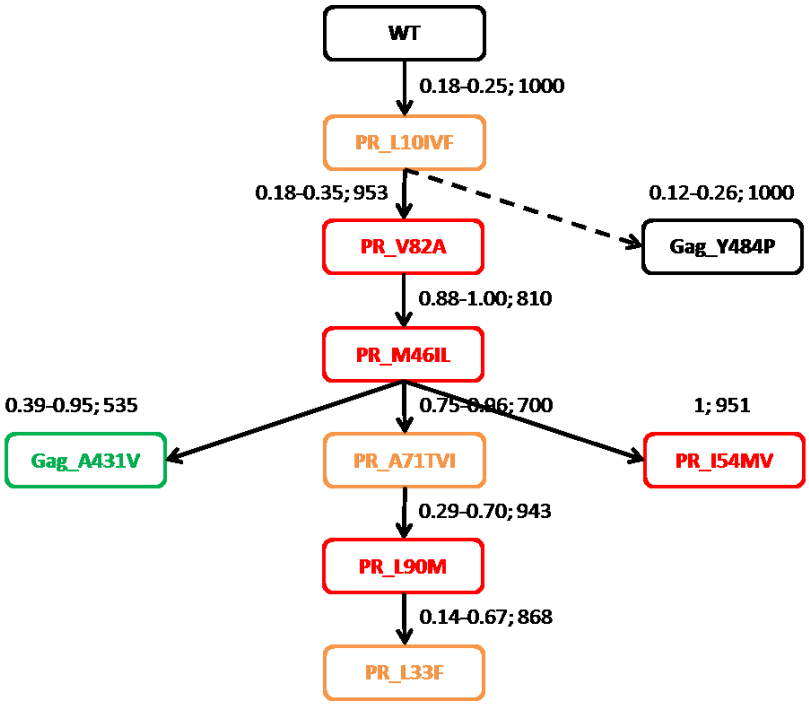


Figure 20. First mutagenetic tree illustrating mutational pathways for Gag and PR mutations observed in the dataset. Green boxes represent known treatment- and resistance-associated mutations in Gag, black – novel treatment-associated mutations in Gag, red and yellow – primary and secondary resistance mutations in protease correspondingly. Arrows indicate order of appearance. Simultaneous evolution along different pathways is possible, but a mutation can only occur in a sample, if all its predecessors (as seen from the root) were also present. First two numbers next to an arrow represent 95% confidence interval for the conditional probability of occurrence of the next respective genetic event. Third number indicates bootstrap support of a given element. Only tree structure supported by at least 500 out of 1000 bootstraps was taken into account. Dashed line indicates a branch with no support from the analysis of pairwise associations. From 62% to 68% of samples fit in the ordered accumulation model while the rest (32% - 38%) can be explained by an unordered appearance or other, not observed, pathways.

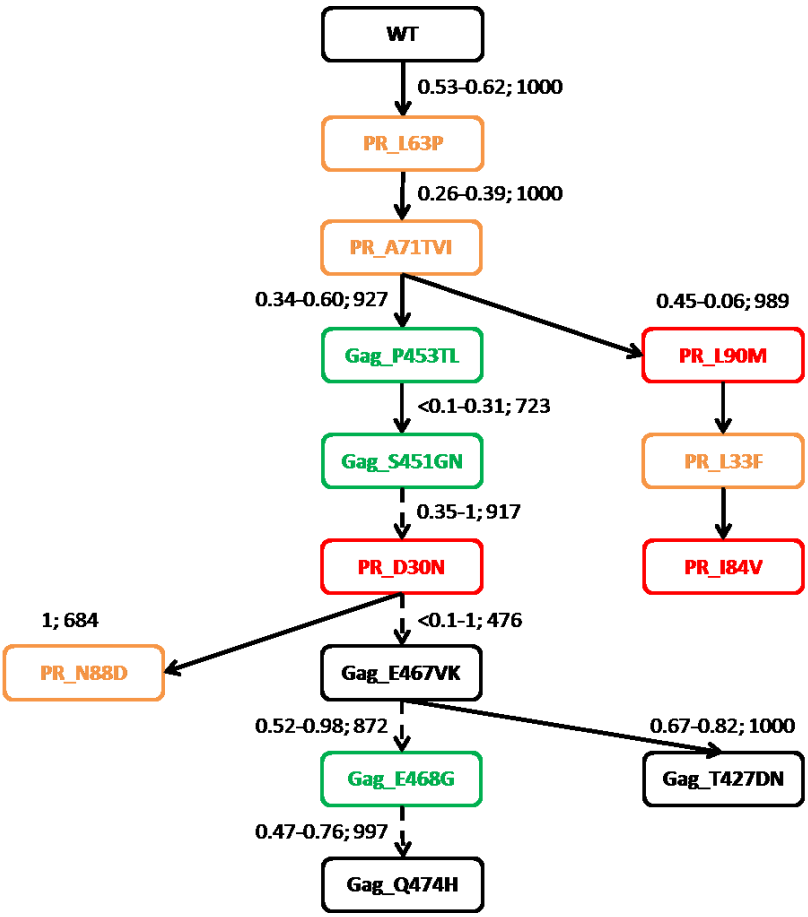


Figure 21. Second mutagenetic tree illustrating mutational pathways for Gag and PR mutations observed in the dataset. Green boxes represent known treatment- and resistance-associated mutations in Gag, black – novel treatment-associated mutations in Gag, red and yellow – primary and secondary resistance mutations in protease correspondingly. Arrows indicate order of appearance. Simultaneous evolution along different pathways is possible, but a mutation can only occur in a sample, if all its predecessors (as seen from the root) were also present. First two numbers next to an arrow represent 95% confidence interval for the conditional probability of occurrence of the next respective genetic event. Third number indicates bootstrap support of a given element. Only tree structure supported by at least 500 out of 1000 bootstraps was taken into account. Dashed line indicates a branch with the bootstrap support below selected cutoff with no support from the analysis of pairwise associations. From 50% to 56% of samples fit in the ordered accumulation model while the rest (44% - 50%) can be explained by an unordered appearance or other, not observed, pathways.

6.8 Selected phenotypes

We developed, verified by sequencing and phenotypically tested in duplicates several single-point mutants, double-mutants and broader combinations. Mutations in Gag A431V, I437V, L449V, L449P, S451N, S451H, R452S, P453L and mutations in PR I47V, I54V, V82A were selected as known resistance associated alterations. Mutation T427D was statistically identified to be associated to PI treatment. And alterations R429I and Y441Q showed tendency for overrepresentation in TE group.

In our experiments none of single mutants demonstrated any significant difference in PI susceptibility; however, we observed a decrease of viral fitness (Table 3). For none of emerging double mutants of interest an increase in drug resistance could be shown but rather a further fitness decrease: L449P-P453L – 81%, S451N-I47V – 59%, T427D-I54V – 37%. The first pair observed in our dataset had been discussed by Verheyn *et al*. [202]. The second and third pairs demonstrated association tendency in our dataset. Additionally, covariation of S451N and I47V was also supported by the structural proximity and presence of electrostatic contact between them (Figure 22).

Though tested mutations accumulated in PI treated sample group they did not confer any detectable resistance but rather decreased fitness. Our suggestion was that this might be due to the absence of necessary background alterations which might reveal the effects and functionality of selected mutations. Therefore we analyzed PR mutant L10I-M46I-I54V-A71V-V82A developed and characterized by Alessio Cremonesi as LPV (fold change IC50 2.4) and IDV (fold change 2.5) resistant and unfit [229] versus its Gag T427D, A431V, S451N and R452S derivatives. R452S has been described as potential Gag resistance and is particularly interesting due to its association with DRV and structural effects [238-240]. Every clone demonstrated expected LPV and IDV resistance (Figure 23), susceptibility of other PI tested were not affected. Mutant PR L10I-M46I-I54V-

A71V-V82A was 46% fit while addition of T427D, S451N or R452S increased the fitness up to 72-82%. In contrast, A431V decreased fitness down to 23%.

Table 3. Fitness summary on single mutants that showed no resistance in phenotyping tests.

Mutant	Fitness compared to pNL-NF, %	Mutant	Fitness compared to pNL-NF, %
Gag_T427D	60	Gag_S451N	86
Gag_R429I	47	Gag_S451H	77
Gag_A431V	52	Gag_R452S	100
Gag_I437V	57	Gag_P453L	75
Gag_Y441Q	91	PR_I47V	75
Gag_L449V	48	PR_I54V	90
Gag_L449P	60	PR_V82A	25

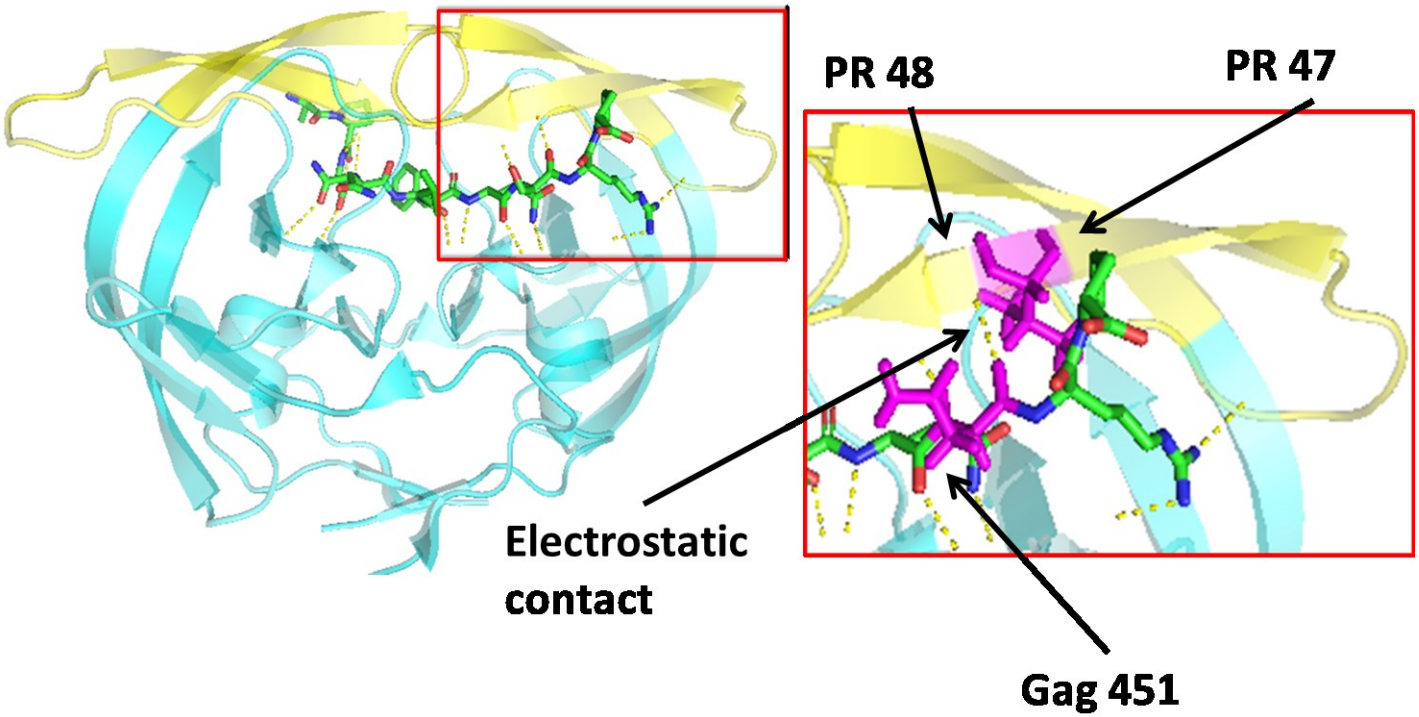


Figure 22. Structural aspect of the interaction between residues 47 in PR and 451 in Gag (in purple). HIV-1 protease dimer (cartoon representation) with a decapeptide of its natural substrate p1/p6 Gag (stick representation). PDB ID: 1MT9. Flexible PR flaps are in yellow. Yellow dotted lines represent electrostatic contacts. When residue PR 47 has been mutated to obtain resistance electrostatic contact with residue Gag 451 and therefore optimal substrate accommodation in the PR substrate binding cleft might be lost. But when residue Gag 451 mutates accordingly electrostatic contact is restored. Therefore processing efficiency of the cleavage site is restored as well.

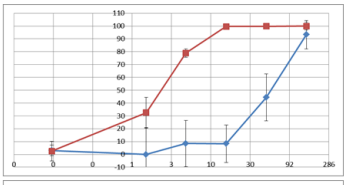
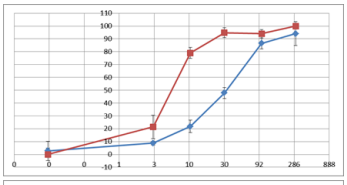
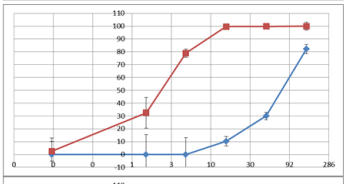
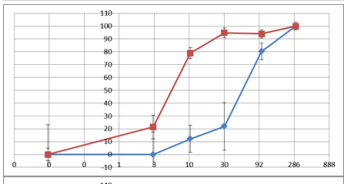
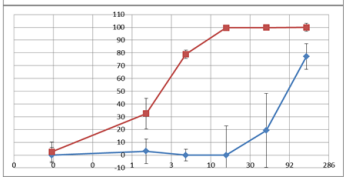
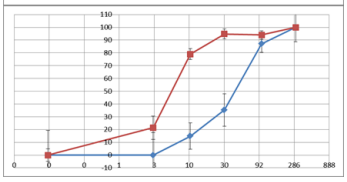
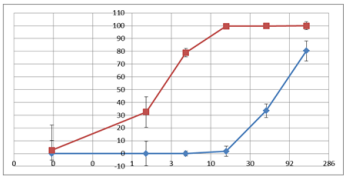
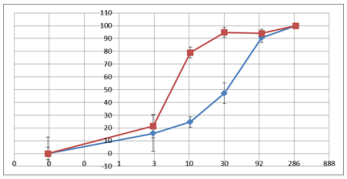
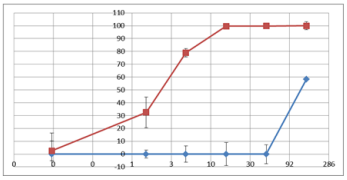
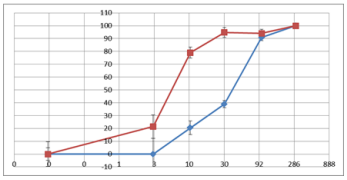
Mutant	Fitness, %	IDV	LPV
PR L10I-M46I-I54V-A71V-V82A	46		
PR L10I-M46I-I54V-A71V-V82A + T427D	82		
PR L10I-M46I-I54V-A71V-V82A + A431V	23		
PR L10I-M46I-I54V-A71V-V82A + S451N	78		
PR L10I-M46I-I54V-A71V-V82A + R452S	72		

Figure 23. Phenotypes of the mutants with resistant PR in combination with Gag alterations. On the graphs percent of inhibition is represented with vertical axis and drug concentration in nM – with horizontal axis. Blue line – mutant, red line – pNL-NF reference.

7 DISCUSSION

7.1 Validity of the chosen analytical approach

This study is the first in the SHCS to systematically investigate the possible role of HIV-1 Gag mutations during PI treatment and for the emergence of viral drug resistance. The pilot steps of our analysis validated the analytical approach selected as the observed rates of primary and secondary resistances in the viral protease matched with previously published data well. The only slight difference was observed for secondary resistances A71V and L10I, which, based on previous data, were more prevalent than expected in TN samples in the SHCS data set. These two natural polymorphisms have not been attributed to failures on certain protease inhibitors [236] but were rather associated with unspecific cooperative decrease of PI susceptibility in vitro [104, 190]. So the difference is rather explained by reported natural geographical variability in prevalence of mutations and by sampling than peculiarities of treatment regimens [104, 190, 241, 242]. These results confirm the validity of the statistical methods used in this study.

7.2 Frequency of established Gag resistance mutations in the SHCS

The frequencies of resistance-relevant known Gag mutations in the TE set of SHCS samples were in a good agreement with a report by Malet *et al*. [212], stating that at the variable positions at the p2/NC cleavage site alone 91% of PI-exposed subtype B viruses exhibited at least one mutation. That supports our observations being true tendencies rather than method of region-specific artifacts. The occurrence of mutations I437V, P453L and I479I among the TE samples in the absence of accompanying primary protease mutations could indicate several mechanistic roles: They could either have an independent role in resistance or they might prime as initial steps mutations in PR or they could cooperate with secondary protease mutations. Of interest, P453L has been described up to here to confer resistance only when primary protease mutations in protease had emerged

[192]. And for I479I no phenotypic proof in a standardized background is available. This leaves I437V to as the most likely candidate to contribute to the failure of PI-containing regimens in the absence of primary resistances. In a standardized in vitro background it decreased the susceptibilities of LPV, TPV, ATV, and APV [10]. In 17.9% of all TE samples the Gag region carried mutations capable of decreasing PI susceptibility alone or in combination with PR resistances. Yet, the Gag region is not taken into account by routine genotyping. Therefore the overall level of PI resistance might be underestimated.

7.3 Beyond established Gag resistances

Most of the currently known mutations in Gag that have been associated with PI resistance reside in or near the p7/p1 and p1/p6 cleavage sites. Our study set out to complement this information by including the entire C-terminal region of Gag available from our dataset. Several mutations were newly associated with PI exposure. They were mainly located at functionally critical points of the Gag polyprotein. Both positions A360 and Q369 are well conserved among M-group isolates of HIV-1. Alterations at position 360 affect p24/p2 cleavage and have been reported to influence virion assembly and release [243]. Residue 369 belongs to the bevirimat binding sequence and is required for proper virion formation and maturation in vitro, a step that is dependent on the conformation of p2 (SP1) [94]. The folding of this short spacer peptide is likely to be strongly affected by the observed substitution exchanging the hydrophilic glutamine with the rather hydrophobic leucine. Furthermore, the association of Q369L with other known resistance-related alterations in Gag such as K418R, I437V, and P453T, as observed here, underlines its potential critical role and lends support to the claim that Gag mutations may depend on the background sequence of the respective isolate.

The polar but uncharged T427 as the HIV-1 M-group consensus is neutral hydrophobic. Hence a change to asparagine or aspartic acid strongly increases the

hydrophilicity at this position, and such a drastic alteration in the immediate vicinity of a viral PR cleavage site is likely to affect processing efficiency. Additionally, alteration T427N has been previously recognized as a CTL escape mutation affecting viral replication capacity [244].

Of interest, p6 contains several predicted ERK-2 phosphorylation sites, i.e. at T456, S462, T469, T471, S473, Y484 and S499 [245], and the incorporation of ERK-2 into the HIV particle is thought to regulate the L-domain function of p6 [246]. One of possible ways to regulate mono-ubiquitination is via phosphorylation, and it has been suggested that cellular kinases regulate ubiquitination and thereby the structure of p6 [245]. Structure and conformation of p6 are critical for maintaining its hydrophobic interface. Most p6 functions are suggested to occur under hydrophobic conditions near the cytoplasmic membrane [247]. It is therefore interesting to note that the newly described mutations E467V/K, 474H and Y484P in this study were located in this crucial p6 protein, right in the center of the phosphorylation motifs and L-domains.

Gag amino acid 467 marks another potentially critical conserved position. It resides between a proline-rich region, responsible for interaction with TSG101, and an ALIX interaction motif. Additionally, the site is flanked by Vpr binding sequences. Glutamic acid at this position is the inter-subtype consensus. Therefore the observed change to valine would switch the character from hydrophilic to hydrophobic, while lysine would reverse the negative charge in the wild type to positive. It is currently not known what the precise contact points in the reaction partners are, but it is likely that either one of these drastic mutations would have implications for the protein-protein interaction.

Mutation Q474L has been described to occur during the acquisition of PI resistance. Alone, without any accompanying alterations, it was deleterious to viral replication. However, with the simultaneous introduction of corresponding Gag cleavage-site mutations the phenotype could be rescued [203]. In line with this Q474Q/L has also been linked in a subtype-dependent manner to the known PI resistance via P453L/T [248]. In addition, the P453L/T mutation has been

associated with a loss of an ERK-2 phosphorylation site at T471 [248]. In our SHCS dataset the mutation Q474H correlated with the T427N change in 13.3% of cases (OR=24.4), which supports a direct connection of the p6 phosphorylation motifs with the sequence surrounding the p7/p1 cleavage site. As histidine can serve as substrate for mammalian protein kinases the mutation Q474H might immediately be linked to a shift of the phosphorylation event at residue T471.

In the subtype B consensus amino acid Y484 belongs to the ALIX recognition motif, a site involved in the function of particle budding. Various alterations in the sequence of residues 482-484 have been described to affect ALIX binding as well as the accumulation of the Gag processing intermediates p41 and p25 [249, 250]. This observation highlights a connection between the p6 protein and the processing efficiency at the cleavage sites for p17/p24 and p24/p2. Additionally, residues 483 and 484 have been demonstrated to be subtype-dependently selected CTL epitopes [251]. Moreover, the mutation L483M has been reported to be selected in vitro in the presence of DRV [252].

In line with previous studies [203, 248, 253] our data support a new role of p6 alterations: its phosphorylation and folding are likely to affect the efficiency of Gag cleavage site processing and could hence directly associate with viral fitness and even PI susceptibility. P6 alterations have been linked to the mutations affecting p17/ p24, p24/p2 and p1/p6 [203, 248-250, 253] cleavage site processing efficiency and we report novel connection between p6 and p7/p1 cleavage site. Additionally we demonstrate a potential link of PI-associated p6 mutations to phosphorylation of the protein. Our data thereby indicate a key role of positions 360, 369, 427, 467, 474, and 484.

While mutations such as I437V have been reported to behave like primary protease mutations [10, 237], most other isolated Gag mutations do not confer resistance on their own but may rather act as amplifiers similarly to minor PI resistance mutations [254]. Therefore, in line with Doyon *et al*. [7], we speculate that some Gag mutations may perform a “fine tuning” function for HIV carrying protease mutations e.g. with a significant fitness cost. This is in agreement with

Zhang *et al*. [255] who found them to be common and important for viral adaptation pathways resembling secondary protease mutations [13]. For every clinically evolved virus the overall sequence of Gag and Gag-Pol is crucial for fully assessing the role of Gag mutations in a given virus isolate [13], where both assembly of mature virion proteins and polyprotein cleavage are coordinated by the balance of the processing rate between cleavage sites. Being relatively fragile, this balance limits variability of cleavage site sequences and therefore possibilities of compensation for resistance-associated fitness loss [119, 256]. And our results serve as an additional argument to support background-dependent role of Gag mutations.

7.4 Mutational patterns

Mutations in HIV-1 protease responsible for the development of resistance against protease inhibitors are known to form certain patterns. These patterns are dictated by epistatic interactions between them in terms of synergistic impact on drug susceptibility or compensation for resistance-associated fitness loss [104, 190]. However, Gag mutations have been identified as a second mechanism contributing to the decrease of PI susceptibility [197-201]. And their functionality crucially depends on the background sequence [13]. So this study is the first attempt to investigate Gag mutational clusters occurring in the isolates from SHCS patients.

Pairwise associations between the Gag and PR mutations observed in our dataset confirmed PR as the most interconnected region. Primary resistances L90M, M46I, V82A and I54V confirmed their major role in resistance and networking status [257-259] by showing the highest number of significant associations in PR. Secondary resistances in protease also demonstrated numerous significant associations. One of them, L33F, is a fitness-compensating alteration which decreases HIV susceptibility to several PI in the presence of various primary resistance mutations [104, 190, 260, 261]. L33F may be an important “crossroad”

for different resistance pathways. Such a role of L33F is supported by its direct contact with protease substrate / inhibitor binding residues, just like in case of L10I, I54LV and L90M [193]. Another secondary PI resistance mutation in protease A71V is located on the outer face of the protease dimer, but still was as interconnected as L33F and linked to the same established correlates of PI resistance in Gag A431V and P453L. This allows us to speculate that frequently observed secondary resistances represent one of the "bridges" that connect Gag and PR resistance patterns. We could not identify any significant link of novel PI-associated alterations in Gag to primary resistances in PR.

The next step was to analyze Gag-PR mutational landscape with a more sophisticated approach – random tree mixture models. High percentage of samples falling in either of proposed ordered models of accumulation of mutations with the bootstrap support numbers suggest that trees obtained describe interactions between the considered alterations within the selected patterns stably and comprehensively enough. Due to relatively low number of sequences included into the analysis we could not strictly identify the transition rates between the genetic events considered. But the first steps of the mutational pathways observed tended to represent a bottleneck, possibly related to fitness costs of further mutations.

Bayesian network inference has demonstrated its utility in the analysis of HIV resistance against PI [262, 263], but it could not provide any additional data for our analysis. It just confirmed some of identified pairwise associations between the primary resistance mutations in protease. The most likely reason is that our dataset was not large enough to provide the algorithm with a necessary signal to robustly dissect the system with potentially high number of confounding effects and complex interactions.

The backbone of the pattern involving Gag resistance mutation A431V and PR resistance mutations M46I/L, I54V and V82A observed in the current study has been reported by Verheyen *et al*. in the cohort of patients in Germany [202]. However, pattern reported here was broader and was primed by secondary resistance mutations at position 10 in PR. Additionally, A431V and I54V

represented rather final stages of the pattern development. By the PR mutations A71T/V/I the above-mentioned mutational cluster was linked to the pattern involving the p1/p6 cleavage site alterations P453TL and S451G/N, and which further involved D30N-N88D pair. In contrast to previous observations [202] L90M and I84V represented a branch alternative to P453T/L. For the third group of mutations we could not produce a reliable mutagenetic tree. This cluster seems to be rather isolated and includes Gag resistance I437V, major PI resistance L90M and three secondary mutations L10I, Q58E and G73T. In addition we show that resistance patterns are not limited with well-established and best-known primary resistances in PR and cleavage site alterations in Gag. Novel Gag mutations were included in the mutagenetic trees along with established Gag resistances that supports their potential importance in the context of PI resistance. Primary resistance mutations in protease seemed to initiate the development of resistance pattern. Then secondary resistances followed accompanied by the Gag alterations. L63P and L33F along with T427D/N, E467V/K, E468G, Q474H and Y484P were incorporated by the identified resistance patterns.

Suggested mutational patterns identified have already found a confirmation for their functionality. Kolli *et al*. analytically demonstrated their effects: Mutations at position 431 decreased IDV and LPV susceptibility in the presence of V82A. And in the presence of L90M, mutations at positions 431 and 453 decreased HIV susceptibility to every PI. However, in every case enhanced resistance came at a cost of reduced fitness. At the same time number of secondary resistances in protease positively correlated with the presence of Gag alterations [240]. A431V [264] and I437V [10] have been shown to confer PI resistance in the absence of primary resistance mutations, but we could not observe such an effect in our experiments. We could not detect previously reported resistance effects of single primary resistance mutations in protease I47V, I54V and V82A as well [104, 190]. This discrepancy to existing data might be explained by the difference between the phenotyping approaches. Nijhuis *et al.* [10] compared phenotypic drug susceptibility of Gag mutant K436E-I437V in multiple-cycle MTT assay and

single-cycle PhenoSense assay. So for every PI tested MTT assay gave IC50 fold change higher than that by PhenoSense (up to 2 times). In the current study and previous [229] deCIPhR measured the resistance effect of primary PR mutations I54V and L90M lower than PhenoSense (susceptible versus resistant) [104, 190]. Usage of heterologous envelope in an assay may increase infectivity and help the virus to pass by some of the lifecycle stages blocked by PI [145, 265-267]. Signal reading methods have their impact dependent on sensitivity and inherent preciseness. Cell lines utilized influence the development of HIV infection as well through the cellular factors involved in the interactions with the virus [79, 80].

In an attempt to make the effects of Gag alterations visible we continued with a more complex resistance in PR L10I-M46I-I54V-A71V-V82A which has demonstrated measurable resistance with our assay [229]. In the resistance background of PR L10I-M46I-I54V-A71V-V82A we could measure fitness-compensating effects of T427D, S451N and R452S. This is the first phenotypic characteristic produced for S451N and R452S previously recognized as resistance-associated. In addition to that we confirmed the importance of T427D alteration. Such results make sense from the clinical and molecular point of view. On the first place virus accumulates resistances in PR which allow it to survive the drug pressure but diminish the fitness. Then (and this is supported by the mutagenetic trees constructed: Figure 20, Figure 21) Gag mutations occur that change Gag accordingly to the structure of resistant PR in order to compensate for the fitness loss. However, such mutated Gag might not fit well with the wild-type PR that would result in low fitness which we observed in our experiments in case of isolated Gag mutations. The structural basis for this interplay between Gag and PR mutations is illustrated in Figure 22.

Initial steps to collect and structure comprehensive data on Gag region alterations have already been implemented [58, 59], which may help improvement of current genotypic algorithms. And this study provides additional details on the development and interplay of Gag and PR resistance patters.

7.5 Center dependence of the choice of PI-containing therapies

When analyzing the relative rates of use of certain PIs for the two Swiss centers in this study, we noted that samples from center BS were generally associated with a higher exposure to NFV, SQV, and IDV. On the other hand, for the ZH samples ATV and DRV were more often applied. It is conceivable that one center might have provided to the database a higher fraction of data for patients with successful viral suppression and with a lower need for a new therapy change, thereby remaining longer on a previous drug regimen. Alternatively, since both centers follow Swiss treatment guidelines with their timely updates, and as both centers have a regular exchange among treating physicians, it is very likely that these differences of more conservative data set originating from BS reflect a sampling effect, i.e. that a larger fraction of the analyzed sequence entities stems from an earlier period within the time of analysis. Indeed, the analysis date of ZH samples was on average significantly later than for BS samples, $p<0.05$. The fact that during the period from 2004 to 2012 in ZH center 1625 patients were recruited versus 527 in BS also supports this point of view [268].

8 CONCLUSIONS

This study confirms for the Swiss HIV Cohort previously published data from other European settings on the relevance of Gag mutations in the context of PI resistance. We demonstrate that 17.9% of SHCS patients carry resistance mutations in Gag. And since Gag is not considered by the current genotyping systems the overall level of PI resistance for these patients is underestimated.

We report novel Gag mutations which accumulate in PI-treated samples and reside in functionally important regions of Gag. They correlated with previously reported resistance patterns in Gag and PR. Some of them were capable of increasing viral fitness in the context of resistant PR.

The role of PI resistance mutations in Gag and PR critically depends on the background viral sequence. We show that secondary PI resistance mutations also contribute to the development of certain resistance patterns.

Additionally our data support a new role of p6 alterations: its phosphorylation and folding are likely to affect the efficiency of Gag cleavage site processing and could hence directly associate with viral fitness and even PI susceptibility.

Taken together, our data suggests the relevance of Gag sequence information for the routine genotyping of PI-treated patients of the SHCS.

9 OUTLOOK

We identified novel Gag mutations A360S/P, Q369L, T427D/N, E467V/K, Q474H and Y484P as potentially resistance-relevant. However, of those alterations, we managed to characterize phenotypically only T427D mutant alone and in combination with PR resistances. Phenotyping experiments on the rest of these alterations would help to reveal their true role. They should be tested alone and in combination with resistant PR. Alternative approach would be reversion of the mutations of interest in relevant patient viruses. In this case the impact of a given mutation can be observed in its naturally developed backbone. Recently Dr. Doris Chibo from HIV Characterisation Laboratory in Doherty Institute, Melbourne reported 6 cases when patients were failing PI-containing therapy without any sufficient resistances in PR and with controlled regimen compliance. She kindly agreed to share corresponding Gag-PR fragments. It would be interesting to see if these failures could be explained by Gag alterations and if we could find any of our novel Gag mutations involved here.

Different phenotyping protocols could be compared for the evaluation of the most frequently observed resistance patterns across Gag and PR. Nijhuis *et al.* [10] compared phenotypic drug susceptibility of Gag mutant K436E-I437V in multiple-cycle MTT assay and single-cycle PhenoSense assay. So for every PI tested MTT assay gave IC50 fold change higher than that by PhenoSense (up to 2 times). Additionally, in the current study and previous [229] deCIPhR measured the resistance effect of primary PR mutations I54V and L90M lower than PhenoSense (susceptible versus resistant) [104, 190]. Presence of heterologous envelope may increase infectivity and help the virus to pass by some of the lifecycle stages blocked by PI [145, 265-267]. Signal reading methods have their impact dependent on sensitivity and inherent preciseness. Cell lines utilized influence the development of HIV infection as well through the cellular factors involved in the interactions with the virus [79, 80].

In this study we concentrated our attention on subtype B isolates because the number of samples of other individual subtypes did not allow enough statistical power for the analysis even to confirm the status of primary PI resistance mutations in protease. In order to investigate subtype-specific aspects of Gag and PR resistance patterns one might try to build up a larger dataset of samples of a certain subtype by collecting them from different cohorts in comparable settings. An issue here is that different centers may use different genotyping systems that do not necessarily allow the extraction of Gag sequence information.

10 ACKNOWLEDGEMENTS

This PhD thesis was performed between October 2012 and December 2015 in the group of Molecular Virology at the DBM-Petersplatz site of the University of Basel.

I express my sincere gratitude to Prof. Thomas Klimkait for suggesting this exciting project and giving me the chance to work in the laboratories of the Institute of Medical Microbiology and InPheno. His kind guidance and patient tutoring allowed not only the project advancement but also my professional growth and development.

The completion of this thesis would be unthinkable without all-around support from: Joelle Bader, Sarah Wagner, Vincent Vidal, Severine Louvel, Isabell Seibert, Alexandra Haas, Adelaide Loureiro, Sabrina Steiner and every one of DBM-Petersplatz staff.

Prof. Daniel Hoffmann, his team and Dr. Jens Verheyen from the University of Duisburg-Essen generously shared their knowledge and experience with me for which I am very thankful.

Many thanks to Prof. Markus Affolter, who kindly assumed responsibility of my thesis as a Faculty Representative, PD Dr. Rainer Gosert, who was willing to support me by serving as an External Expert for the thesis, and Prof. Urs Jenal for accepting the work of chairing the doctoral examination.

11 REFERENCES

1. Zu Knyphausen F, Scheufele R, Kucherer C, *et al.* First line treatment response in patients with transmitted HIV drug resistance and well defined time point of HIV infection: updated results from the German HIV-1 seroconverter study. *PLoS One* 2014; **9:**e95956.
2. Elzi L, Erb S, Furrer H, *et al.* Choice of Initial Combination Antiretroviral Therapy in Individuals With HIV Infection: Determinants and Outcomes. *Arch Intern Med* 2012; **172:**1313-1321.
3. Kohler P, Schmidt AJ, Cavassini M, *et al.* The HIV care cascade in Switzerland: reaching the UNAIDS/WHO targets for patients diagnosed with HIV. *AIDS* 2015.
4. von Wyl V, Yerly S, Burgisser P, *et al.* Long-term trends of HIV type 1 drug resistance prevalence among antiretroviral treatment-experienced patients in Switzerland. *Clin Infect Dis* 2009; **48:**979-987.
5. Le T, Chiarella J, Simen BB, *et al.* Low-abundance HIV drug-resistant viral variants in treatment-experienced persons correlate with historical antiretroviral use. *PLoS One* 2009; **4:**e6079.
6. Simen BB, Simons JF, Hullsiek KH, *et al.* Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. *J Infect Dis* 2009; **199:**693-701.
7. Doyon L, Croteau G, Thibeault D, Poulin F, Pilote L, Lamarre D. Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *J Virol* 1996; **70:**3763-3769.
8. Croteau G, Doyon L, Thibeault D, McKercher G, Pilote L, Lamarre D. Impaired fitness of human immunodeficiency virus type 1 variants with high-level resistance to protease inhibitors. *J Virol* 1997; **71:**1089-1096.
9. Carrillo A, Stewart KD, Sham HL, *et al.* In vitro selection and characterization of human immunodeficiency virus type 1 variants with increased resistance to ABT-378, a novel protease inhibitor. *J Virol* 1998; **72:**7532-7541.
10. Nijhuis M, van Maarseveen NM, Lastere S, *et al.* A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism. *PLoS Med* 2007; **4:**e36.
11. Dam E, Quercia R, Glass B, *et al.* Gag mutations strongly contribute to HIV-1 resistance to protease inhibitors in highly drug-experienced patients besides compensating for fitness loss. *PLoS Pathog* 2009; **5:**e1000345.
12. Parry CM, Kohli A, Boinett CJ, Towers GJ, McCormick AL, Pillay D. Gag determinants of fitness and drug susceptibility in protease inhibitor-resistant human immunodeficiency virus type 1. *J Virol* 2009; **83:**9094-9101.
13. Parry CM, Kolli M, Myers RE, Cane PA, Schiffer C, Pillay D. Three residues in HIV-1 matrix contribute to protease inhibitor susceptibility and replication capacity. *Antimicrob Agents Chemother* 2011; **55:**1106-1113.
14. Briggs JA, Simon MN, Gross I, *et al.* The stoichiometry of Gag protein in HIV-1. *Nat Struct Mol Biol* 2004; **11:**672-675.
15. Zhu P, Liu J, Bess J, Jr., *et al.* Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* 2006; **441:**847-852.
16. Briggs JA, Krausslich HG. The molecular architecture of HIV. *J Mol Biol* 2011; **410:**491-500.
17. Chojnacki J, Staudt T, Glass B, *et al.* Maturation-dependent HIV-1 surface protein redistribution revealed by fluorescence nanoscopy. *Science* 2012; **338:**524-528.
18. Saad JS, Miller J, Tai J, Kim A, Ghanam RH, Summers MF. Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. *Proc Natl Acad Sci U S A* 2006; **103:**11364-11369.

19. Checkley MA, Luttge BG, Freed EO. HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. *J Mol Biol* 2011; **410:**582-608.
20. Ott DE. Cellular proteins detected in HIV-1. *Rev Med Virol* 2008; **18:**159-175.
21. Onafuwa-Nuga AA, Telesnitsky A, King SR. 7SL RNA, but not the 54-kd signal recognition particle protein, is an abundant component of both infectious HIV-1 and minimal virus-like particles. *RNA* 2006; **12:**542-546.
22. Splettstoesser T. SciStyle. Edited by Editor|. Year|; p.^pp. Pages|. City|: Publisher|.
23. Seelamgari A, Maddukuri A, Berro R*, et al.* Role of viral regulatory and accessory proteins in HIV-1 replication. *Front Biosci* 2004; **9:**2388-2413.
24. Los Alamos National Security L, for the U.S. Department of Energy's National Nuclear Security Administration. Los Alamos HIV Database. Edited by Editor|. Year|; p.^pp. Pages|. City|: Publisher|.
25. Zhang C, Zhou S, Groppelli E*, et al.* Hybrid spreading mechanisms and T cell activation shape the dynamics of HIV-1 infection. *PLoS Comput Biol* 2015; **11:**e1004179.
26. Gupta P, Balachandran R, Ho M, Enrico A, Rinaldo C. Cell-to-cell transmission of human immunodeficiency virus type 1 in the presence of azidothymidine and neutralizing antibody. *J Virol* 1989; **63:**2361-2365.
27. Mothes W, Sherer NM, Jin J, Zhong P. Virus cell-to-cell transmission. *J Virol* 2010; **84:**8360-8368.
28. Shimura K, Miyazato P, Oishi S, Fujii N, Matsuoka M. Impact of HIV-1 infection pathways on susceptibility to antiviral drugs and on virus spread. *Virology* 2015; **484:**364-376.
29. Ganser-Pornillos BK, Yeager M, Sundquist WI. The structural biology of HIV assembly. *Curr Opin Struct Biol* 2008; **18:**203-217.
30. Tedbury PR, Freed EO. HIV-1 gag: an emerging target for antiretroviral therapy. *Curr Top Microbiol Immunol* 2015; **389:**171-201.
31. Fields BN, Knipe DM, Howley P. *Fields virology*. 6 ed. Philadelphia: Wolters Kluwer Health

Lippincott Williams & Wilkins 2013.
32. Votteler J, Schubert U. Human Immunodeficiency Viruses: Molecular Biology. *Encyclopedia of Virology*. City 2008; p. 517-525.
33. In Mandell GL, Bennett JE, Dolin R (Mandell GL, Bennett JE, Dolin R(Mandell GL, Bennett JE, Dolin Rs). *Mandell, Douglas, and Bennett's principles and practice of infectious diseases*. City: Churchill Livingstone

Elsevier 2010.
34. Gallo RC, Sarin PS, Gelmann EP*, et al.* Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science* 1983; **220:**865-867.
35. 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* 1983; **220:**868-871.
36. Rambaut A, Posada D, Crandall KA, Holmes EC. The causes and consequences of HIV evolution. *Nat Rev Genet* 2004; **5:**52-61.
37. Perelson AS, Ribeiro RM. Estimating drug efficacy and viral dynamic parameters: HIV and HCV. *Stat Med* 2008; **27:**4647-4657.
38. Robertson DL, Hahn BH, Sharp PM. Recombination in AIDS viruses. *J Mol Evol* 1995; **40:**249-259.
39. Gilbert PB, McKeague IW, Eisen G*, et al.* Comparison of HIV-1 and HIV-2 infectivity from a prospective cohort study in Senegal. *Stat Med* 2003; **22:**573-593.
40. Reeves JD, Doms RW. Human immunodeficiency virus type 2. *J Gen Virol* 2002; **83:**1253-1265.

41. Keele BF, Jones JH, Terio KA*, et al.* Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. *Nature* 2009; **460:**515-519.
42. Thomson MM, Perez-Alvarez L, Najera R. Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. *Lancet Infect Dis* 2002; **2:**461-471.
43. Carr JK, Foley BT, Leitner T, Salminen M, Korber B, McCutchan F. *Reference sequences representing the principal genetic diversity of HIV-1 in the pandemic*. Los Alamos, New Mexico: Los Alamos National Laboratory 1998.
44. Gallo RC, Montagnier L. The discovery of HIV as the cause of AIDS. *N Engl J Med* 2003; **349:**2283-2285.
45. Organization WH. *Antiretroviral Therapy for HIV Infection in Adults and Adolescents: Recommendations for a Public Health Approach: 2010 Revision*. City 2010.
46. UNAIDS. Global Report: UNAIDS Report on the Global AIDS Epidemic 2010. Edited by Editor|. Year|; p.^pp. Pages|. City|: Publisher|.
47. Fauci AS, Folkers GK. Investing to meet the scientific challenges of HIV/AIDS. *Health Aff (Millwood)* 2009; **28:**1629-1641.
48. UNAIDS. Fact Sheet 2014. Edited by Editor|. Year|; p.^pp. Pages|. City|: Publisher|.
49. Ortblad KF, Lozano R, Murray CJ. The burden of HIV: insights from the Global Burden of Disease Study 2010. *AIDS* 2013; **27:**2003-2017.
50. Battegay M. Lecture "HIV - translational and clinical aspects". Edited by Editor|. Year|; p.^pp. Pages|. City|: Publisher|.
51. Lenardo MJ, Angleman SB, Bounkeua V*, et al.* Cytopathic killing of peripheral blood CD4(+) T lymphocytes by human immunodeficiency virus type 1 appears necrotic rather than apoptotic and does not require env. *J Virol* 2002; **76:**5082-5093.
52. Modrow S, Falke D, Truyen U, Schatzl H. *Molecular Virology*. Berlin

Heidelberg: Springer 2013.
53. Bell NM, Lever AM. HIV Gag polyprotein: processing and early viral particle assembly. *Trends Microbiol* 2013; **21:**136-144.
54. Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* 1988; **331:**280-283.
55. Ono A, Freed EO. Role of lipid rafts in virus replication. *Adv Virus Res* 2005; **64:**311-358.
56. Waheed AA, Freed EO. The Role of Lipids in Retrovirus Replication. *Viruses* 2010; **2:**1146-1180.
57. Chukkapalli V, Hogue IB, Boyko V, Hu WS, Ono A. Interaction between the human immunodeficiency virus type 1 Gag matrix domain and phosphatidylinositol-(4,5)-bisphosphate is essential for efficient gag membrane binding. *J Virol* 2008; **82:**2405-2417.
58. Ono A, Ablan SD, Lockett SJ, Nagashima K, Freed EO. Phosphatidylinositol (4,5) bisphosphate regulates HIV-1 Gag targeting to the plasma membrane. *Proc Natl Acad Sci U S A* 2004; **101:**14889-14894.
59. Chukkapalli V, Ono A. Molecular determinants that regulate plasma membrane association of HIV-1 Gag. *J Mol Biol* 2011; **410:**512-524.
60. Rein A. Nucleic acid chaperone activity of retroviral Gag proteins. *RNA Biol* 2010; **7:**700-705.
61. Muriaux D, Darlix JL. Properties and functions of the nucleocapsid protein in virus assembly. *RNA Biol* 2010; **7:**744-753.
62. Lu K, Heng X, Summers MF. Structural determinants and mechanism of HIV-1 genome packaging. *J Mol Biol* 2011; **410:**609-633.
63. Sundquist WI, Krausslich HG. HIV-1 assembly, budding, and maturation. *Cold Spring Harb Perspect Med* 2012; **2:**a006924.

64. Franke EK, Yuan HE, Luban J. Specific incorporation of cyclophilin A into HIV-1 virions. *Nature* 1994; **372:**359-362.
65. Gitti RK, Lee BM, Walker J, Summers MF, Yoo S, Sundquist WI. Structure of the amino-terminal core domain of the HIV-1 capsid protein. *Science* 1996; **273:**231-235.
66. Luban J, Bossolt KL, Franke EK, Kalpana GV, Goff SP. Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. *Cell* 1993; **73:**1067-1078.
67. Rasaiyaah J, Tan CP, Fletcher AJ*, et al.* HIV-1 evades innate immune recognition through specific cofactor recruitment. *Nature* 2013; **503:**402-405.
68. Johnson MC. Mechanisms for Env glycoprotein acquisition by retroviruses. *AIDS Res Hum Retroviruses* 2011; **27:**239-247.
69. Freed EO, Martin MA. Virion incorporation of envelope glycoproteins with long but not short cytoplasmic tails is blocked by specific, single amino acid substitutions in the human immunodeficiency virus type 1 matrix. *J Virol* 1995; **69:**1984-1989.
70. Votteler J, Sundquist WI. Virus budding and the ESCRT pathway. *Cell Host Microbe* 2013; **14:**232-241.
71. Weiss ER, Gottlinger H. The role of cellular factors in promoting HIV budding. *J Mol Biol* 2011; **410:**525-533.
72. Demirov DG, Orenstein JM, Freed EO. The late domain of human immunodeficiency virus type 1 p6 promotes virus release in a cell type-dependent manner. *J Virol* 2002; **76:**105-117.
73. Gottlinger HG, Dorfman T, Sodroski JG, Haseltine WA. Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proc Natl Acad Sci U S A* 1991; **88:**3195-3199.
74. Huang M, Orenstein JM, Martin MA, Freed EO. p6Gag is required for particle production from full-length human immunodeficiency virus type 1 molecular clones expressing protease. *J Virol* 1995; **69:**6810-6818.
75. Fujii K, Munshi UM, Ablan SD*, et al.* Functional role of Alix in HIV-1 replication. *Virology* 2009; **391:**284-292.
76. Murakami T, Ablan S, Freed EO, Tanaka Y. Regulation of human immunodeficiency virus type 1 Env-mediated membrane fusion by viral protease activity. *J Virol* 2004; **78:**1026-1031.
77. Wyma DJ, Jiang J, Shi J*, et al.* Coupling of human immunodeficiency virus type 1 fusion to virion maturation: a novel role of the gp41 cytoplasmic tail. *J Virol* 2004; **78:**3429-3435.
78. Alfadhli A, Barklis RL, Barklis E. HIV-1 matrix organizes as a hexamer of trimers on membranes containing phosphatidylinositol-(4,5)-bisphosphate. *Virology* 2009; **387:**466-472.
79. Akari H, Fukumori T, Adachi A. Cell-dependent requirement of human immunodeficiency virus type 1 gp41 cytoplasmic tail for Env incorporation into virions. *J Virol* 2000; **74:**4891-4893.
80. Murakami T, Freed EO. The long cytoplasmic tail of gp41 is required in a cell type-dependent manner for HIV-1 envelope glycoprotein incorporation into virions. *Proc Natl Acad Sci U S A* 2000; **97:**343-348.
81. Tedbury PR, Ablan SD, Freed EO. Global rescue of defects in HIV-1 envelope glycoprotein incorporation: implications for matrix structure. *PLoS Pathog* 2013; **9:**e1003739.
82. Ganser-Pornillos BK, Cheng A, Yeager M. Structure of full-length HIV-1 CA: a model for the mature capsid lattice. *Cell* 2007; **131:**70-79.
83. Forshey BM, von Schwedler U, Sundquist WI, Aiken C. Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *J Virol* 2002; **76:**5667-5677.
84. McDonald D, Vodicka MA, Lucero G*, et al.* Visualization of the intracellular behavior of HIV in living cells. *J Cell Biol* 2002; **159:**441-452.

85. Iordanskiy S, Berro R, Altieri M, Kashanchi F, Bukrinsky M. Intracytoplasmic maturation of the human immunodeficiency virus type 1 reverse transcription complexes determines their capacity to integrate into chromatin. *Retrovirology* 2006; **3:**4.
86. Matreyek KA, Engelman A. Viral and cellular requirements for the nuclear entry of retroviral preintegration nucleoprotein complexes. *Viruses* 2013; **5:**2483-2511.
87. Brass AL, Dykxhoorn DM, Benita Y*, et al.* Identification of host proteins required for HIV infection through a functional genomic screen. *Science* 2008; **319:**921-926.
88. Konig R, Zhou Y, Elleder D*, et al.* Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell* 2008; **135:**49-60.
89. Yamashita M, Perez O, Hope TJ, Emerman M. Evidence for direct involvement of the capsid protein in HIV infection of nondividing cells. *PLoS Pathog* 2007; **3:**1502-1510.
90. Thali M, Bukovsky A, Kondo E*, et al.* Functional association of cyclophilin A with HIV-1 virions. *Nature* 1994; **372:**363-365.
91. De Iaco A, Luban J. Cyclophilin A promotes HIV-1 reverse transcription but its effect on transduction correlates best with its effect on nuclear entry of viral cDNA. *Retrovirology* 2014; **11:**11.
92. Kogan M, Rappaport J. HIV-1 accessory protein Vpr: relevance in the pathogenesis of HIV and potential for therapeutic intervention. *Retrovirology* 2011; **8:**25.
93. Planelles V, Barker E. Roles of Vpr and Vpx in modulating the virus-host cell relationship. *Mol Aspects Med* 2010; **31:**398-406.
94. Datta SA, Temeselew LG, Crist RM*, et al.* On the role of the SP1 domain in HIV-1 particle assembly: a molecular switch? *J Virol* 2011; **85:**4111-4121.
95. de Marco A, Heuser AM, Glass B, Krausslich HG, Muller B, Briggs JA. Role of the SP2 domain and its proteolytic cleavage in HIV-1 structural maturation and infectivity. *J Virol* 2012; **86:**13708-13716.
96. Lee SK, Potempa M, Kolli M, Ozen A, Schiffer CA, Swanstrom R. Context surrounding processing sites is crucial in determining cleavage rate of a subset of processing sites in HIV-1 Gag and Gag-Pro-Pol polyprotein precursors by viral protease. *J Biol Chem* 2012; **287:**13279-13290.
97. Toh H, Ono M, Saigo K, Miyata T. Retroviral protease-like sequence in the yeast transposon Ty 1. *Nature* 1985; **315**.
98. Meek TD, Dayton BD, Metcalf BW*, et al.* Human immunodeficiency virus 1 protease expressed in Escherichia coli behaves as a dimeric aspartic protease. *Proc Natl Acad Sci U S A* 1989; **86:**1841-1845.
99. Wlodawer A, Miller M, Jaskolski M*, et al.* Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* 1989; **245:**616-621.
100. Lapatto R, Blundell T, Hemmings A*, et al.* X-ray analysis of HIV-1 proteinase at 2.7 A resolution confirms structural homology among retroviral enzymes. *Nature* 1989; **342:**299-302.
101. Navia MA, Fitzgerald PM, McKeever BM*, et al.* Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* 1989; **337:**615-620.
102. Miller M, Jaskolski M, Rao JK, Leis J, Wlodawer A. Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature* 1989; **337:**576-579.
103. Wensing AM, van Maarseveen NM, Nijhuis M. Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. *Antiviral Res* 2010; **85:**59-74.
104. Rhee SY, Gonzales MJ, Kantor R, Betts BJ, Ravela J, Shafer RW. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res* 2003; **31:**298-303.
105. Bhat TN, Baldwin ET, Liu B, Cheng YS, Erickson JW. Crystal structure of a tethered dimer of HIV-1 proteinase complexed with an inhibitor. *Nat Struct Biol* 1994; **1:**552-556.
106. Tang C, Louis JM, Aniana A, Suh JY, Clore GM. Visualizing transient events in amino-terminal autoprocessing of HIV-1 protease. *Nature* 2008; **455:**693-696.

107. Partin K, Zybarth G, Ehrlich L, DeCrombrugghe M, Wimmer E, Carter C. Deletion of sequences upstream of the proteinase improves the proteolytic processing of human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A* 1991; **88:**4776-4780.
108. Louis JM, Clore GM, Gronenborn AM. Autoprocessing of HIV-1 protease is tightly coupled to protein folding. *Nat Struct Biol* 1999; **6:**868-875.
109. Lindhofer H, von der Helm K, Nitschko H. In vivo processing of Pr160gag-pol from human immunodeficiency virus type 1 (HIV) in acutely infected, cultured human T-lymphocytes. *Virology* 1995; **214:**624-627.
110. Sluis-Cremer N, Arion D, Abram ME, Parniak MA. Proteolytic processing of an HIV-1 pol polyprotein precursor: insights into the mechanism of reverse transcriptase p66/p51 heterodimer formation. *Int J Biochem Cell Biol* 2004; **36:**1836-1847.
111. Pettit SC, Everitt LE, Choudhury S, Dunn BM, Kaplan AH. Initial cleavage of the human immunodeficiency virus type 1 GagPol precursor by its activated protease occurs by an intramolecular mechanism. *J Virol* 2004; **78:**8477-8485.
112. Louis JM, Wondrak EM, Kimmel AR, Wingfield PT, Nashed NT. Proteolytic processing of HIV-1 protease precursor, kinetics and mechanism. *J Biol Chem* 1999; **274:**23437-23442.
113. Wondrak EM, Nashed NT, Haber MT, Jerina DM, Louis JM. A transient precursor of the HIV-1 protease. Isolation, characterization, and kinetics of maturation. *J Biol Chem* 1996; **271:**4477-4481.
114. Pettit SC, Lindquist JN, Kaplan AH, Swanstrom R. Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates. *Retrovirology* 2005; **2:**66.
115. de Oliveira T, Engelbrecht S, Janse van Rensburg E*, et al.* Variability at human immunodeficiency virus type 1 subtype C protease cleavage sites: an indication of viral fitness? *J Virol* 2003; **77:**9422-9430.
116. Krausslich HG, Ingraham RH, Skoog MT, Wimmer E, Pallai PV, Carter CA. Activity of purified biosynthetic proteinase of human immunodeficiency virus on natural substrates and synthetic peptides. *Proc Natl Acad Sci U S A* 1989; **86:**807-811.
117. Swanstrom R, Wills JW. Retroviral gene expression. II. Synthesis, processing, and assembly of viral proteins. In Coffin JM, Hughes SH, Varmus HE (Coffin JM, Hughes SH, Varmus HE(Coffin JM, Hughes SH, Varmus HEs). *RetroViruses*. City: Cold Spring Harbor Laboratory Press 1997; p. 263-334.
118. Erickson-Viitanen S, Manfredi J, Viitanen P*, et al.* Cleavage of HIV-1 gag polyprotein synthesized in vitro: sequential cleavage by the viral protease. *AIDS Res Hum Retroviruses* 1989; **5:**577-591.
119. Pettit SC, Moody MD, Wehbie RS*, et al.* The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *J Virol* 1994; **68:**8017-8027.
120. Wiegers K, Rutter G, Kottler H, Tessmer U, Hohenberg H, Krausslich HG. Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites. *J Virol* 1998; **72:**2846-2854.
121. Pettit SC, Henderson GJ, Schiffer CA, Swanstrom R. Replacement of the P1 amino acid of human immunodeficiency virus type 1 Gag processing sites can inhibit or enhance the rate of cleavage by the viral protease. *J Virol* 2002; **76:**10226-10233.
122. Prabu-Jeyabalan M, Nalivaika E, Schiffer CA. Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. *Structure* 2002; **10:**369-381.
123. Perez MA, Fernandes PA, Ramos MJ. Substrate recognition in HIV-1 protease: a computational study. *J Phys Chem B* 2010; **114:**2525-2532.
124. Salzwedel K, Martin DE, Sakalian M. Maturation inhibitors: a new therapeutic class targets the virus structure. *AIDS Rev* 2007; **9:**162-172.

125. Potempa M, Lee SK, Wolfenden R, Swanstrom R. The triple threat of HIV-1 protease inhibitors. *Curr Top Microbiol Immunol* 2015; **389:**203-241.
126. Checkley MA, Luttge BG, Soheilian F, Nagashima K, Freed EO. The capsid-spacer peptide 1 Gag processing intermediate is a dominant-negative inhibitor of HIV-1 maturation. *Virology* 2010; **400:**137-144.
127. Muller B, Anders M, Akiyama H*, et al.* HIV-1 Gag processing intermediates trans-dominantly interfere with HIV-1 infectivity. *J Biol Chem* 2009; **284:**29692-29703.
128. Li F, Goila-Gaur R, Salzwedel K*, et al.* PA-457: a potent HIV inhibitor that disrupts core condensation by targeting a late step in Gag processing. *Proc Natl Acad Sci U S A* 2003; **100:**13555-13560.
129. Martin DE, Blum R, Wilton J*, et al.* Safety and pharmacokinetics of Bevirimat (PA-457), a novel inhibitor of human immunodeficiency virus maturation, in healthy volunteers. *Antimicrob Agents Chemother* 2007; **51:**3063-3066.
130. Adamson CS, Ablan SD, Boeras I*, et al.* In vitro resistance to the human immunodeficiency virus type 1 maturation inhibitor PA-457 (Bevirimat). *J Virol* 2006; **80:**10957-10971.
131. Van Baelen K, Salzwedel K, Rondelez E*, et al.* Susceptibility of human immunodeficiency virus type 1 to the maturation inhibitor bevirimat is modulated by baseline polymorphisms in Gag spacer peptide 1. *Antimicrob Agents Chemother* 2009; **53:**2185-2188.
132. Margot NA, Gibbs CS, Miller MD. Phenotypic susceptibility to bevirimat in isolates from HIV-1-infected patients without prior exposure to bevirimat. *Antimicrob Agents Chemother* 2010; **54:**2345-2353.
133. Seclen E, Gonzalez Mdel M, Corral A, de Mendoza C, Soriano V, Poveda E. High prevalence of natural polymorphisms in Gag (CA-SP1) associated with reduced response to Bevirimat, an HIV-1 maturation inhibitor. *AIDS* 2010; **24:**467-469.
134. Verheyen J, Verhofstede C, Knops E*, et al.* High prevalence of bevirimat resistance mutations in protease inhibitor-resistant HIV isolates. *AIDS* 2010; **24:**669-673.
135. Blair WS, Cao J, Fok-Seang J*, et al.* New small-molecule inhibitor class targeting human immunodeficiency virus type 1 virion maturation. *Antimicrob Agents Chemother* 2009; **53:**5080-5087.
136. Waki K, Durell SR, Soheilian F, Nagashima K, Butler SL, Freed EO. Structural and functional insights into the HIV-1 maturation inhibitor binding pocket. *PLoS Pathog* 2012; **8:**e1002997.
137. Palella FJ, Jr., Delaney KM, Moorman AC*, et al.* Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. *N Engl J Med* 1998; **338:**853-860.
138. Gulick RM, Mellors JW, Havlir D*, et al.* Treatment with indinavir, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy. *N Engl J Med* 1997; **337:**734-739.
139. Turner SR, Strohbach JW, Tommasi RA*, et al.* Tipranavir (PNU-140690): a potent, orally bioavailable nonpeptidic HIV protease inhibitor of the 5,6-dihydro-4-hydroxy-2-pyrone sulfonamide class. *J Med Chem* 1998; **41:**3467-3476.
140. Mehandru S, Markowitz M. Tipranavir: a novel non-peptidic protease inhibitor for the treatment of HIV infection. *Expert Opin Investig Drugs* 2003; **12:**1821-1828.
141. Roberts NA, Martin JA, Kinchington D*, et al.* Rational design of peptide-based HIV proteinase inhibitors. *Science* 1990; **248:**358-361.
142. Henderson GJ, Lee SK, Irlbeck DM*, et al.* Interplay between single resistance-associated mutations in the HIV-1 protease and viral infectivity, protease activity, and inhibitor sensitivity. *Antimicrob Agents Chemother* 2012; **56:**623-633.
143. de la Carriere LC, Paulous S, Clavel F, Mammano F. Effects of human immunodeficiency virus type 1 resistance to protease inhibitors on reverse transcriptase processing, activity, and drug sensitivity. *J Virol* 1999; **73:**3455-3459.

144. Shen L, Peterson S, Sedaghat AR, *et al.* Dose-response curve slope sets class-specific limits on inhibitory potential of anti-HIV drugs. *Nat Med* 2008; **14:**762-766.
145. Rabi SA, Laird GM, Durand CM, *et al.* Multi-step inhibition explains HIV-1 protease inhibitor pharmacodynamics and resistance. *J Clin Invest* 2013; **123:**3848-3860.
146. Jiang J, Aiken C. Maturation of the viral core enhances the fusion of HIV-1 particles with primary human T cells and monocyte-derived macrophages. *Virology* 2006; **346:**460-468.
147. Kaplan AH, Manchester M, Swanstrom R. The activity of the protease of human immunodeficiency virus type 1 is initiated at the membrane of infected cells before the release of viral proteins and is required for release to occur with maximum efficiency. *J Virol* 1994; **68:**6782-6786.
148. Lee SK, Harris J, Swanstrom R. A strongly transdominant mutation in the human immunodeficiency virus type 1 gag gene defines an Achilles heel in the virus life cycle. *J Virol* 2009; **83:**8536-8543.
149. Wyma DJ, Kotov A, Aiken C. Evidence for a stable interaction of gp41 with Pr55(Gag) in immature human immunodeficiency virus type 1 particles. *J Virol* 2000; **74:**9381-9387.
150. Lori F, Scovassi AI, Zella D, *et al.* Enzymatically active forms of reverse transcriptase of the human immunodeficiency virus. *AIDS Res Hum Retroviruses* 1988; **4:**393-398.
151. Kawamura M, Shimano R, Inubushi R, *et al.* Cleavage of Gag precursor is required for early replication phase of HIV-1. *FEBS Lett* 1997; **415:**227-230.
152. Mirambeau G, Lyonnais S, Coulaud D, *et al.* Transmission electron microscopy reveals an optimal HIV-1 nucleocapsid aggregation with single-stranded nucleic acids and the mature HIV-1 nucleocapsid protein. *J Mol Biol* 2006; **364:**496-511.
153. Mirambeau G, Lyonnais S, Coulaud D, *et al.* HIV-1 protease and reverse transcriptase control the architecture of their nucleocapsid partner. *PLoS One* 2007; **2:**e669.
154. Rulli SJ, Jr., Muriaux D, Nagashima K, *et al.* Mutant murine leukemia virus Gag proteins lacking proline at the N-terminus of the capsid domain block infectivity in virions containing wild-type Gag. *Virology* 2006; **347:**364-371.
155. Moore MD, Fu W, Soheilian F, *et al.* Suboptimal inhibition of protease activity in human immunodeficiency virus type 1: effects on virion morphogenesis and RNA maturation. *Virology* 2008; **379:**152-160.
156. Yamashita M, Emerman M. Cellular restriction targeting viral capsids perturbs human immunodeficiency virus type 1 infection of nondividing cells. *J Virol* 2009; **83:**9835-9843.
157. Qi M, Yang R, Aiken C. Cyclophilin A-dependent restriction of human immunodeficiency virus type 1 capsid mutants for infection of nondividing cells. *J Virol* 2008; **82:**12001-12008.
158. Thomas JA, Gagliardi TD, Alvord WG, Lubomirski M, Bosche WJ, Gorelick RJ. Human immunodeficiency virus type 1 nucleocapsid zinc-finger mutations cause defects in reverse transcription and integration. *Virology* 2006; **353:**41-51.
159. Buckman JS, Bosche WJ, Gorelick RJ. Human immunodeficiency virus type 1 nucleocapsid zn(2+) fingers are required for efficient reverse transcription, initial integration processes, and protection of newly synthesized viral DNA. *J Virol* 2003; **77:**1469-1480.
160. Bukrinsky MI, Sharova N, McDonald TL, Pushkarskaya T, Tarpley WG, Stevenson M. Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. *Proc Natl Acad Sci U S A* 1993; **90:**6125-6129.
161. Karageorgos L, Li P, Burrell C. Characterization of HIV replication complexes early after cell-to-cell infection. *AIDS Res Hum Retroviruses* 1993; **9:**817-823.
162. Heinzinger NK, Bukrinsky MI, Haggerty SA, *et al.* The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. *Proc Natl Acad Sci U S A* 1994; **91:**7311-7315.
163. Miller MD, Farnet CM, Bushman FD. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J Virol* 1997; **71:**5382-5390.

164. Coren LV, Thomas JA, Chertova E*, et al.* Mutational analysis of the C-terminal gag cleavage sites in human immunodeficiency virus type 1. *J Virol* 2007; **81:**10047-10054.
165. Wolfenden R. Analog approaches to the structure of the transition state in enzyme reactions. *Acc Chem Res* 1972; **5:**10–18.
166. Dreyer GB, Metcalf BW, Tomaszek TA, Jr.*, et al.* Inhibition of human immunodeficiency virus 1 protease in vitro: rational design of substrate analogue inhibitors. *Proc Natl Acad Sci U S A* 1989; **86:**9752-9756.
167. Krohn A, Redshaw S, Ritchie JC, Graves BJ, Hatada MH. Novel binding mode of highly potent HIV-proteinase inhibitors incorporating the (R)-hydroxyethylamine isostere. *J Med Chem* 1991; **34:**3340-3342.
168. Kempf DJ, Marsh KC, Denissen JF*, et al.* ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans. *Proc Natl Acad Sci U S A* 1995; **92:**2484-2488.
169. Kaldor SW, Kalish VJ, Davies JF, 2nd*, et al.* Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *J Med Chem* 1997; **40:**3979-3985.
170. Thaisrivongs S, Strohbach JW. Structure-based discovery of Tipranavir disodium (PNU-140690E): a potent, orally bioavailable, nonpeptidic HIV protease inhibitor. *Biopolymers* 1999; **51:**51-58.
171. Stoll V, Qin W, Stewart KD*, et al.* X-ray crystallographic structure of ABT-378 (lopinavir) bound to HIV-1 protease. *Bioorg Med Chem* 2002; **10:**2803-2806.
172. Ali A, Bandaranayake RM, Cai Y*, et al.* Molecular Basis for Drug Resistance in HIV-1 Protease. *Viruses* 2010; **2:**2509-2535.
173. Domingo E, Escarmis C, Sevilla N*, et al.* Basic concepts in RNA virus evolution. *FASEB J* 1996; **10:**859-864.
174. Brown AJ. Analysis of HIV-1 env gene sequences reveals evidence for a low effective number in the viral population. *Proc Natl Acad Sci U S A* 1997; **94:**1862-1865.
175. Brown AJ, Richman DD. HIV-1: gambling on the evolution of drug resistance? *Nat Med* 1997; **3:**268-271.
176. Nijhuis M, Boucher CA, Schipper P, Leitner T, Schuurman R, Albert J. Stochastic processes strongly influence HIV-1 evolution during suboptimal protease-inhibitor therapy. *Proc Natl Acad Sci U S A* 1998; **95:**14441-14446.
177. Shankarappa R, Margolick JB, Gange SJ*, et al.* Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol* 1999; **73:**10489-10502.
178. Rodrigo AG, Hanley EW, Goracke PC, Learn GH. Sampling and processing HIV molecular sequences: A computational evolutionary biologist's perspective. In Rodrigo AG, Learn GH (Rodrigo AG, Learn GH(Rodrigo AG, Learn GHs). *Computational and Evolutionary Analysis of HIV Molecular Sequences*. City: Kluwer Academic Publishers 2001.
179. Bourinbaiar AS. The ratio of defective HIV-1 particles to replication-competent infectious virions. *Acta Virol* 1994; **38:**59-61.
180. Dimitrov DS, Willey RL, Sato H, Chang LJ, Blumenthal R, Martin MA. Quantitation of human immunodeficiency virus type 1 infection kinetics. *J Virol* 1993; **67:**2182-2190.
181. Layne SP, Merges MJ, Dembo M*, et al.* Factors underlying spontaneous inactivation and susceptibility to neutralization of human immunodeficiency virus. *Virology* 1992; **189:**695-714.
182. Marozsan AJ, Fraundorf E, Abraha A*, et al.* Relationships between infectious titer, capsid protein levels, and reverse transcriptase activities of diverse human immunodeficiency virus type 1 isolates. *J Virol* 2004; **78:**11130-11141.
183. Chen Z, Li Y, Schock HB, Hall D, Chen E, Kuo LC. Three-dimensional structure of a mutant HIV-1 protease displaying cross-resistance to all protease inhibitors in clinical trials. *J Biol Chem* 1995; **270:**21433-21436.

184. Gulnik SV, Suvorov LI, Liu B*, et al.* Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure. *Biochemistry* 1995; **34:**9282-9287.
185. Mahalingam B, Louis JM, Reed CC*, et al.* Structural and kinetic analysis of drug resistant mutants of HIV-1 protease. *Eur J Biochem* 1999; **263:**238-245.
186. Mammano F, Trouplin V, Zennou V, Clavel F. Retracing the evolutionary pathways of human immunodeficiency virus type 1 resistance to protease inhibitors: virus fitness in the absence and in the presence of drug. *J Virol* 2000; **74:**8524-8531.
187. Nijhuis M, Schuurman R, de Jong D*, et al.* Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *AIDS* 1999; **13:**2349-2359.
188. Quinones Mateu ME, Arts EJ. HIV-1 fitness: implications for drug resistance, disease progression, and global epidemic evolution. In Kuiken C, Foley B, Hahn B, Marx PA, McCutchan F, Mellor J (Kuiken C, Foley B, Hahn B, Marx PA, McCutchan F, Mellor J(Kuiken C, Foley B, Hahn B, Marx PA, McCutchan F, Mellor Js). *HIV Sequence Compendium 2001*. City: Theoretical Biology and Biophysics Group 2001.
189. Johnson VA, Brun-Vezinet F, Clotet B*, et al.* Update of the drug resistance mutations in HIV-1: Spring 2008. *Top HIV Med* 2008; **16:**62-68.
190. Shafer RW. Rationale and uses of a public HIV drug-resistance database. *J Infect Dis* 2006; **194 Suppl 1:**S51-58.
191. Mammano F, Petit C, Clavel F. Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients. *J Virol* 1998; **72:**7632-7637.
192. Maguire MF, Guinea R, Griffin P*, et al.* Changes in human immunodeficiency virus type 1 Gag at positions L449 and P453 are linked to I50V protease mutants in vivo and cause reduction of sensitivity to amprenavir and improved viral fitness in vitro. *J Virol* 2002; **76:**7398-7406.
193. Saskova KG, Kozisek M, Rezacova P*, et al.* Molecular characterization of clinical isolates of human immunodeficiency virus resistant to the protease inhibitor darunavir. *J Virol* 2009; **83:**8810-8818.
194. Schrodinger, LLC. The AxPyMOL Molecular Graphics Plugin for Microsoft PowerPoint, Version 1.0. Edited by Editor|. Year|; p.^pp. Pages|. City|: Publisher|.
195. Schrodinger, LLC. The JyMOL Molecular Graphics Development Component, Version 1.0. Edited by Editor|. Year|; p.^pp. Pages|. City|: Publisher|.
196. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.3r1. Edited by Editor|. Year|; p.^pp. Pages|. City|: Publisher|.
197. Bally F, Martinez R, Peters S, Sudre P, Telenti A. Polymorphism of HIV type 1 gag p7/p1 and p1/p6 cleavage sites: clinical significance and implications for resistance to protease inhibitors. *AIDS Res Hum Retroviruses* 2000; **16:**1209-1213.
198. Kaufmann D, Munoz M, Bleiber G*, et al.* Virological and immunological characteristics of HIV treatment failure. *AIDS* 2000; **14:**1767-1774.
199. Bleiber G, Munoz M, Ciuffi A, Meylan P, Telenti A. Individual contributions of mutant protease and reverse transcriptase to viral infectivity, replication, and protein maturation of antiretroviral drug-resistant human immunodeficiency virus type 1. *J Virol* 2001; **75:**3291-3300.
200. Telenti A, Martinez R, Munoz M*, et al.* Analysis of natural variants of the human immunodeficiency virus type 1 gag-pol frameshift stem-loop structure. *J Virol* 2002; **76:**7868-7873.
201. Lambert-Niclot S, Flandre P, Malet I*, et al.* Impact of gag mutations on selection of darunavir resistance mutations in HIV-1 protease. *J Antimicrob Chemother* 2008; **62:**905-908.
202. Verheyen J, Litau E, Sing T*, et al.* Compensatory mutations at the HIV cleavage sites p7/p1 and p1/p6-gag in therapy-naive and therapy-experienced patients. *Antivir Ther* 2006; **11:**879-887.

203. Myint L, Matsuda M, Matsuda Z*, et al.* Gag non-cleavage site mutations contribute to full recovery of viral fitness in protease inhibitor-resistant human immunodeficiency virus type 1. *Antimicrob Agents Chemother* 2004; **48:**444-452.
204. Robinson LH, Myers RE, Snowden BW, Tisdale M, Blair ED. HIV type 1 protease cleavage site mutations and viral fitness: implications for drug susceptibility phenotyping assays. *AIDS Res Hum Retroviruses* 2000; **16:**1149-1156.
205. Gatanaga H, Suzuki Y, Tsang H*, et al.* Amino acid substitutions in Gag protein at non-cleavage sites are indispensable for the development of a high multitude of HIV-1 resistance against protease inhibitors. *J Biol Chem* 2002; **277:**5952-5961.
206. Prabu-Jeyabalan M, Nalivaika EA, King NM, Schiffer CA. Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. *J Virol* 2004; **78:**12446-12454.
207. Wu X, Li Y, Peng K, Zhou H. HIV protease inhibitors in gut barrier dysfunction and liver injury. *Curr Opin Pharmacol* 2014; **19:**61-66.
208. Caron-Debarle M, Lagathu C, Boccara F, Vigouroux C, Capeau J. HIV-associated lipodystrophy: from fat injury to premature aging. *Trends Mol Med* 2010; **16:**218-229.
209. Kim RJ, Wilson CG, Wabitsch M, Lazar MA, Steppan CM. HIV protease inhibitor-specific alterations in human adipocyte differentiation and metabolism. *Obesity (Silver Spring)* 2006; **14:**994-1002.
210. Shafer RW. Genotypic testing for human immunodeficiency virus type 1 drug resistance. *Clin Microbiol Rev* 2002; **15:**247-277.
211. Van Laethem K, De Luca A, Antinori A, Cingolani A, Perna CF, Vandamme AM. A genotypic drug resistance interpretation algorithm that significantly predicts therapy response in HIV-1-infected patients. *Antivir Ther* 2002; **7:**123-129.
212. Malet I, Roquebert B, Dalban C*, et al.* Association of Gag cleavage sites to protease mutations and to virological response in HIV-1 treated patients. *J Infect* 2007; **54:**367-374.
213. Ghosn J, Delaugerre C, Flandre P*, et al.* Polymorphism in Gag gene cleavage sites of HIV-1 non-B subtype and virological outcome of a first-line lopinavir/ritonavir single drug regimen. *PLoS One* 2011; **6:**e24798.
214. von Wyl V, Yerly S, Boni J*, et al.* Emergence of HIV-1 drug resistance in previously untreated patients initiating combination antiretroviral treatment: a comparison of different regimen types. *Arch Intern Med* 2007; **167:**1782-1790.
215. Institute R. RegaDB Sequence Analysis Tools. Edited by Editor|. Year|; p.^pp. Pages|. City|: Publisher|.
216. Team RDC. R: A language and environment for statistical computing. Edited by Editor|. Year|; p.^pp. Pages|. City|: Publisher|.
217. Beerenwinkel N, Daumer M, Sing T*, et al.* Estimating HIV evolutionary pathways and the genetic barrier to drug resistance. *J Infect Dis* 2005; **191:**1953-1960.
218. Beerenwinkel N, Rahnenfuhrer J, Kaiser R, Hoffmann D, Selbig J, Lengauer T. Mtreemix: a software package for learning and using mixture models of mutagenetic trees. *Bioinformatics* 2005; **21:**2106-2107.
219. Bogojeska J, Alexa A, Altmann A, Lengauer T, Rahnenfuhrer J. Rtreemix: an R package for estimating evolutionary pathways and genetic progression scores. *Bioinformatics* 2008; **24:**2391-2392.
220. Beerenwinkel N, Rahnenfuhrer J, Daumer M*, et al.* Learning multiple evolutionary pathways from cross-sectional data. *J Comput Biol* 2005; **12:**584-598.
221. Kalisch M, Maechler M, Colombo D, Maathuis MH, Buehlmann P. Causal Inference Using Graphical Models with the R Package pcalg. *Journal of Statistical Software* 2012; **47:**1-26.
222. Hauser A, Buehlmann P. Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs. *Journal of Machine Learning Research* 2012; **13:**2409-2464.

223. Pearl J. Graphical models for probabilistic and causal reasoning. In Gabbay DM, Smets P (Gabbay DM, Smets P(Gabbay DM, Smets Ps). *Handbook of Defeasible Reasoning and Uncertainty Management Systems*. Vol. 1. City: Kluwer Academic Publishers 1998; p. 367–389.
224. Spirtes P, Meek C, Richardson T. An Algorithm for Causal Inference in the Presence of Latent Variables and Selection Bias. *Computation, Causation and Discovery*. City: AAAI Press 1999; p. 211-252.
225. Spirtes P, Glymour C, Scheines R. *Causation, Prediction, and Search, Second Edition*. 2 ed. Cambridge: MIT Press 2000.
226. Maathuis MH, Colombo D. A generalized back-door criterion. *Annals of Statistics* 2015; **43:**1060-1088.
227. Maathuis MH, Colombo D, Kalisch M, Bühlmann P. Predicting causal effects in large-scale systems from observational data. *Nature Methods* 2010; **7:**247-248.
228. Adachi A, Gendelman HE, Koenig S*, et al.* Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. *J Virol* 1986; **59:**284-291.
229. Cremonesi A. A Replicative Phenotyping System Dissects Complex Resistances to HIV-1 Inhibitors. *In Department Biomedizin*. Edited by Editor|. Year|; p.^pp. Pages|. City|: Publisher|.
230. Klimkait T. A sensitive replicative system to assess HIV-1 drug resistance. *Am Clin Lab* 2002; **21:**20-24.
231. Klimkait T, Stauffer F, Lupo E, Sonderegger-Rubli C. Dissecting the mode of action of various HIV-inhibitor classes in a stable cellular system. *Arch Virol* 1998; **143:**2109-2131.
232. Schmidt D, Kollan C, Fatkenheuer G*, et al.* Estimating trends in the proportion of transmitted and acquired HIV drug resistance in a long term observational cohort in Germany. *PLoS One* 2014; **9:**e104474.
233. Shet A, Berry L, Mohri H*, et al.* Tracking the prevalence of transmitted antiretroviral drug-resistant HIV-1: a decade of experience. *J Acquir Immune Defic Syndr* 2006; **41:**439-446.
234. Buchacz K, Young B, Palella FJ, Jr.*, et al.* Trends in use of genotypic resistance testing and frequency of major drug resistance among antiretroviral-naive persons in the HIV Outpatient Study, 1999-2011. *J Antimicrob Chemother* 2015; **70:**2337-2346.
235. Kozal MJ, Shah N, Shen N*, et al.* Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. *Nat Med* 1996; **2:**753-759.
236. Scherrer AU, Ledergerber B, von Wyl V*, et al.* Minor Protease Inhibitor Mutations at Baseline Do Not Increase the Risk for a Virological Failure in HIV-1 Subtype B Infected Patients. *PLoS One* 2012; **7**.
237. Nijhuis M, Wensing AMJ, Bierman W*, et al.* A novel pathway involving Gag CS A431V followed by protease M46I & L76V leading to lopinavir resistance. *Antivir. Ther.* 2007; **12:**S146.
238. Verheyen J, Altmann A, Knops E*, et al.* Relevance of HIV gag cleavage site mutations in failures of protease-inhibitor therapies. *In XVII International HIV drug resistance workshop*. Edited by Editor|. Year|; p.^pp. Pages|. City|: Publisher|.
239. Ozen A, Lin KH, Kurt Yilmaz N, Schiffer CA. Structural basis and distal effects of Gag substrate coevolution in drug resistance to HIV-1 protease. *Proc Natl Acad Sci U S A* 2014; **111:**15993-15998.
240. Kolli M, Stawiski E, Chappey C, Schiffer CA. Human immunodeficiency virus type 1 protease-correlated cleavage site mutations enhance inhibitor resistance. *J Virol* 2009; **83:**11027-11042.
241. Rhee SY, Blanco JL, Jordan MR*, et al.* Geographic and temporal trends in the molecular epidemiology and genetic mechanisms of transmitted HIV-1 drug resistance: an individual-patient- and sequence-level meta-analysis. *PLoS Med* 2015; **12:**e1001810.
242. Dieudonne M, Garzaro D, Torres J*, et al.* High prevalence of secondary resistance mutations in Venezuelan HIV-1 isolates. *Invest Clin* 2006; **47:**27-34.

243. Melamed D, Mark-Danieli M, Kenan-Eichler M, *et al.* The conserved carboxy terminus of the capsid domain of human immunodeficiency virus type 1 gag protein is important for virion assembly and release. *J Virol* 2004; **78:**9675-9688.
244. Boutwell CL, Carlson JM, Lin TH, *et al.* Frequent and variable cytotoxic-T-lymphocyte escape-associated fitness costs in the human immunodeficiency virus type 1 subtype B Gag proteins. *J Virol* 2013; **87:**3952-3965.
245. Fossen T, Wray V, Bruns K, *et al.* Solution structure of the human immunodeficiency virus type 1 p6 protein. *J Biol Chem* 2005; **280:**42515-42527.
246. Hemonnot B, Cartier C, Gay B, *et al.* The host cell MAP kinase ERK-2 regulates viral assembly and release by phosphorylating the p6gag protein of HIV-1. *J Biol Chem* 2004; **279:**32426-32434.
247. Solbak SM, Reksten TR, Hahn F, *et al.* HIV-1 p6 - a structured to flexible multifunctional membrane-interacting protein. *Biochim Biophys Acta* 2013; **1828:**816-823.
248. Ojesina AI, Chaplin B, Sankale JL, *et al.* Interplay of reverse transcriptase inhibitor therapy and gag p6 diversity in HIV type 1 subtype G and CRF02_AG. *AIDS Res Hum Retroviruses* 2008; **24:**1167-1174.
249. Munshi UM, Kim J, Nagashima K, Hurley JH, Freed EO. An Alix fragment potently inhibits HIV-1 budding: characterization of binding to retroviral YPXL late domains. *J Biol Chem* 2007; **282:**3847-3855.
250. Lazert C, Chazal N, Briant L, Gerlier D, Cortay JC. Refined study of the interaction between HIV-1 p6 late domain and ALIX. *Retrovirology* 2008; **5:**39.
251. Peters HO, Mendoza MG, Capina RE, *et al.* An integrative bioinformatic approach for studying escape mutations in human immunodeficiency virus type 1 gag in the Pumwani Sex Worker Cohort. *J Virol* 2008; **82:**1980-1992.
252. Koh Y, Amano M, Towata T, *et al.* In vitro selection of highly darunavir-resistant and replication-competent HIV-1 variants by using a mixture of clinical HIV-1 isolates resistant to multiple conventional protease inhibitors. *J Virol* 2010; **84:**11961-11969.
253. Votteler J, Neumann L, Hahn S, *et al.* Highly conserved serine residue 40 in HIV-1 p6 regulates capsid processing and virus core assembly. *Retrovirology* 2011; **8:**11.
254. Banke S, Lillemark MR, Gerstoft J, Obel N, Jorgensen LB. Positive selection pressure introduces secondary mutations at Gag cleavage sites in human immunodeficiency virus type 1 harboring major protease resistance mutations. *J Virol* 2009; **83:**8916-8924.
255. Zhang YM, Imamichi H, Imamichi T, *et al.* Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites. *J Virol* 1997; **71:**6662-6670.
256. Krausslich HG, Facke M, Heuser AM, Konvalinka J, Zentgraf H. The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity. *J Virol* 1995; **69:**3407-3419.
257. Ozahata MC, Sabino EC, Diaz RS, Cesar-Jr RM, Ferreira JE. Data-intensive analysis of HIV mutations. *BMC Bioinformatics* 2015; **16:**35.
258. Rhee SY, Fessel WJ, Zolopa AR, *et al.* HIV-1 Protease and reverse-transcriptase mutations: correlations with antiretroviral therapy in subtype B isolates and implications for drug-resistance surveillance. *J Infect Dis* 2005; **192:**456-465.
259. Dykes C, Demeter LM. Clinical significance of human immunodeficiency virus type 1 replication fitness. *Clin Microbiol Rev* 2007; **20:**550-578.
260. Monno L, Scudeller L, Brindicci G, *et al.* Genotypic analysis of the protease and reverse transcriptase of non-B HIV type 1 clinical isolates from naive and treated subjects. *Antiviral Res* 2009; **83:**118-126.
261. Van Marck H, Dierynck I, Kraus G, *et al.* The impact of individual human immunodeficiency virus type 1 protease mutations on drug susceptibility is highly influenced by complex interactions with the background protease sequence. *J Virol* 2009; **83:**9512-9520.

262. Deforche K, Silander T, Camacho R*, et al.* Analysis of HIV-1 pol sequences using Bayesian Networks: implications for drug resistance. *Bioinformatics* 2006; **22:**2975-2979.
263. Deforche K, Camacho R, Grossman Z*, et al.* Bayesian network analysis of resistance pathways against HIV-1 protease inhibitors. *Infect Genet Evol* 2007; **7:**382-390.
264. Nijhuis M, Wensing AM, Bierman WF*, et al.* Failure of treatment with first-line lopinavir boosted with ritonavir can be explained by novel resistance pathways with protease mutation 76V. *J Infect Dis* 2009; **200:**698-709.
265. Aiken C. Pseudotyping human immunodeficiency virus type 1 (HIV-1) by the glycoprotein of vesicular stomatitis virus targets HIV-1 entry to an endocytic pathway and suppresses both the requirement for Nef and the sensitivity to cyclosporin A. *J Virol* 1997; **71:**5871-5877.
266. Luo T, Douglas JL, Livingston RL, Garcia JV. Infectivity enhancement by HIV-1 Nef is dependent on the pathway of virus entry: implications for HIV-based gene transfer systems. *Virology* 1998; **241:**224-233.
267. Yu D, Wang W, Yoder A, Spear M, Wu Y. The HIV envelope but not VSV glycoprotein is capable of mediating HIV latent infection of resting CD4 T cells. *PLoS Pathog* 2009; **5:**e1000633.
268. Swiss HIVCS. SHCS key data (tables): Demographical characteristics. Edited by Editor|. Year|; p.^pp. Pages|. City|: Publisher|.