Regulation of HIV-1 reverse transcription
Beerens, N.

Citation for published version (APA):
Beerens, N. (2002). Regulation of HIV-1 reverse transcription
Chapter 1
1. GENERAL INTRODUCTION

1.1 The human immunodeficiency virus type 1
The causative agent of acquired immunodeficiency syndrome (AIDS) was first identified in 1983 when a new retrovirus was isolated from patients (5). This retrovirus is now known as the human immunodeficiency virus type 1 (HIV-1), and in 1986 the related virus HIV-2 was isolated (20). The HIV-1 and HIV-2 viruses belong to the lentivirus genus, a subfamily of the Retroviridae. These viruses carry a diploid positive-strand RNA genome. The retroviral life cycle is characterized by reverse transcription of the viral RNA genome by the viral enzyme reverse transcriptase (RT) into double-stranded proviral DNA, which subsequently integrates into the host cell genome.

1.2 The HIV-1 genome and the structure of the virus particle
The HIV-1 proviral DNA genome is approximately 10 kb in size and encodes nine open reading frames (Fig 1A). The gag gene encodes the matrix (MA), capsid (CA) and nucleocapsid (NC) structural proteins that are translated as a Gag polyprotein. After assembly of the virus particle, proteolytic processing of this precursor protein by the viral protease enzyme yields the mature structural proteins (Fig 1B). The virus particle is enveloped by a lipid membrane that originates from the cell that produced the virus particle. In the mature particle, the matrix protein lines the inner surface of the virion membrane. The capsid protein forms the inner core structure that surrounds the RNA genome (23,60,85,92), and the nucleocapsid protein is associated with the viral RNA (Fig 1B). The NC protein has been suggested to be involved in RNA dimer formation, RNA packaging and reverse transcription (75,92). The pol gene encodes the viral enzymes reverse transcriptase (RT), integrase (IN) and protease (PR) (23,60,92). These proteins are produced as a Gag-Pol precursor protein that is cleaved into the functional enzymes by the viral protease during virus maturation (Fig 1B) (85). The reverse transcriptase copies the viral RNA genome into double stranded DNA (87), which subsequently

---

Fig 1. (A) The HIV-1 proviral DNA genome. The positions of the open reading frames in the double stranded DNA genome are indicated. HIV-1 encodes the basic set of retroviral proteins Gag, Pol and Env and the additional proteins Vif, Vpr, Vpu, Tat, Rev and Nef. The coding region is flanked by Long Terminal Repeats (LTRs) that encode regulatory signals essential for virus replication. The 5' LTR acts as a promoter for the synthesis of viral RNA, and the start site of transcription is indicated by an arrow. (B) Structure of the HIV-1 virus particle. The virus particle is enveloped by a lipid membrane that contains the Env glycoprotein subunits SU-gp120 and TM-p41. The gag-encoded matrix proteins (MA) line the inside of the envelope, and the capsid (CA) proteins form the cone-shaped virion capsid. The viral nucleocapsid protein (NC) is associated with the RNA genome (vRNA) that is present as a dimer in the capsid structure. The pol-encoded proteins integrase (IN), protease (PR) and reverse transcriptase (RT) are also incorporated into the virion capsid.
integrates into the host genome through the action of the integrase enzyme (15). The env gene encodes a glycoprotein that is expressed as gp160 precursor protein that is subsequently cleaved into a surface domain (SU-gp120) and a transmembrane domain (TM-gp41), which remain associated through non-covalent interactions (Fig 1B) (85). The surface protein mediates the attachment to the cellular CD4 receptor and a chemokine coreceptor, which results in fusion of the viral membrane with the cell membrane (34,60). In addition to this standard set of retroviral genes, the HIV-1 genome encodes six additional proteins (Fig 1A). The accessory proteins Vif, Vpr, Vpu and Nef are not absolutely required for virus replication, but stimulate virus replication in certain cell-types, including primary cells (60,88). The essential Tat and Rev proteins regulate HIV-1 gene expression (Fig 1A). The viral transactivator protein Tat is responsible for the activation of HIV-1 transcription by binding to the TAR element located in the untranslated leader region of the viral RNA (74). The Rev protein binds to viral mRNAs via the Rev responsive element (RRE) located within the env gene, and induces nuclear export of these HIV-1 transcripts. Accumulation of Rev results in a shift from the production of multiply spliced transcripts (encoding Tat, Rev and Nef) in the early stage of infection to singly spliced and unspliced transcripts (encoding Gag, Pol and Env) in the late stage of infection (74).

1.3 The HIV-1 replication cycle
HIV-1 replicates in CD4-positive cells, including T-lymphocytes, monocytes and macrophages. The HIV-1 replication cycle is schematically shown in figure 2. The viral envelope protein gp120 binds to the cellular CD4 receptor and a specific co-receptor (CXCR4 for T-cell tropic strains and CCR5 for macrophage-tropic cells, see (34)). This interaction mediates fusion of the viral and cellular membranes, and the viral core is released into the cytoplasm of the host cell. The viral RNA genome is converted into double stranded DNA by the viral RT enzyme, a process that is initiated from a cellular tRNA\textsuperscript{lys} molecule that is used as primer for reverse transcription. This process takes place within the viral core, and the core is transported into the nucleus. Integration of the proviral DNA into the host genome is mediated by the viral integrase protein that recognizes the specific attachment sites (att) at the ends of the viral DNA molecule (15). The integrated proviral DNA subsequently serves as a template for the production of RNA. Transcription is mediated by the cellular RNA polymerase II and is strongly enhanced by the Tat protein (85). In the early stage of infection, the regulatory proteins Tat and Rev and the accessory proteins are produced. Accumulation of Rev results in a shift towards singly and unspliced transcripts that encode the Gag, Pol and Env proteins (85). The structural viral proteins and two copies of the full-length viral RNA assemble at the cell membrane and new virus particles are released from the cell by budding. The virus particles mature through processing of the Gag and Gag-Pol precursor proteins by the virus-encoded protease enzyme to produce infectious particles that can infect new host cells. A detailed overview of the HIV-1 replication cycle is presented in ref. (60).

2. REVERSE TRANSCRIPTION OF THE HIV-1 GENOME

2.1 Mechanism of reverse transcription
The process of reverse transcription is a complex, multi-step reaction. The mechanism of reverse transcription has been reviewed in detail by several authors (50,60,86,87). I will present a general overview of this process and focus in more detail on the initiation phase of reverse transcription. Retroviral reverse transcription is initiated from a specific tRNA molecule that is annealed to a complementary sequence in the viral genome, which is called the primer binding site (PBS). Reverse transcription of the HIV-1 genome is primed by tRNA\textsuperscript{lys} (Fig 3) (50,90). The 3' terminal 18 nucleotides of the tRNA primer anneal to the PBS that is located in a highly structured region of the untranslated 5' leader of the viral genome (Fig 4). The presence of the PBS near the 5' terminus of viral RNA was unexpected since extension of the tRNA primer will result in a short cDNA product only. To synthesize a full-length minus-strand DNA, the reverse transcription complex must be repositioned to the 3' end of the viral RNA. The mechanism to accomplish this involves the repeat (R) regions at the 5' and 3' ends of the viral RNA genome, as is illustrated in figure 3. The initial cDNA product that is complementary to the 5' end of the viral genome contains the 5' R region and is called (-) strand strong-stop cDNA.
Fig 2. The HIV-1 viral life cycle. See the text for details.
Chapter 1

Ribonuclease H (RNase H) activity of the RT enzyme degrades the RNA template within the RNA-DNA duplex (19), and the cDNA strand is subsequently transferred to the 3’ R region (1st transfer). The minus strand strong-stop cDNA is elongated, and the RNA template is degraded by the RNase H activity. Plus-strand cDNA synthesis is initiated from two RNase H-resistant fragments of the viral RNA template, the polypurine tract in the U3 region (U3-PPT) and the central polypurine tract (cPPT). Plus-strand synthesis stops at the first modified base in tRNA^{3’}, that is after copying of the 3’-terminal 18 nucleotides. The tRNA primer is removed from the plus-strand strong-stop DNA by RNase H cleavage. This plus-strand strong-stop cDNA is subsequently transferred to the 5’ end of the viral genome (2nd transfer), and plus- and minus-strand DNA synthesis continues to complete the double stranded DNA genome. The final product is a linear double stranded proviral DNA flanked by two long terminal repeats (LTRs) that consist of the U3, R and U5 regions.

2.2 The reverse transcriptase enzyme

The HIV-1 RT enzyme is a heterodimer consisting of a 51 kDa subunit (p51) and a 66 kDa subunit (p66) (59,87). The p51 subunit of the p66/p51 heterodimer is generated by cleavage of p66 that is mediated by the viral protease enzyme. The RT enzyme possesses both RNA-dependent and DNA-dependent DNA polymerase activity that enables the enzyme to copy the viral RNA genome into a double stranded DNA copy. In addition, the p66 subunit contains an RNase H domain that degrades the RNA template during reverse transcription. Both the DNA polymerase and the RNase H activity reside within the large p66 subunit of the enzyme (32,48,49). The interaction between both subunits is necessary for optimal polymerase activity. Monomeric or dimeric enzymes formed with either p51 or p66 have reduced activity compared to the heterodimer p51/p66 (82). The X-ray structure of the HIV-1 RT enzyme resembles a right hand and the distinct subregions are termed fingers, palm, thumb and connection domain, with a separate RNase H domain (40,41,68,77,87).

The HIV-1 RT enzyme has a relatively low fidelity of nucleotide incorporation. Incorrect dNTPs that are incorporated during reverse transcription are not removed due to the absence of 3’ exonuclease activity (7). The absence of proofreading, combined with the high viral replication rate, is responsible for rapid genetic variation in the virus population. HIV-1 RT error frequencies reported in literature vary from 10^{-4} to 10^{-2} mutations per base per replication cycle. The consequence of this genetic diversity is that the virus population can adapt rapidly to changes in the environment such as the presence of antiviral drugs, leading to the appearance of drug-resistant variants.

Fig 3. Mechanism of reverse transcription. See the text for details.
2.3 The tRNA primer

The retrovirus family uses diverse tRNA molecules as primer for reverse transcription. Avian retrovirus use tRNA\(_{\text{lys}}\), most murine retroviruses and the human T-cell leukemia viruses HTLV-1 and HTLV-2 use tRNA\(_{\text{lys}}\), and HIV-1 and HIV-2 and the simian immunodeficiency viruses (SIV) use tRNA\(_{\text{lys}}\). Although the nucleotide sequence varies among different tRNA species, they all form the characteristic L-shaped three-dimensional tRNA structure. A particular feature of tRNAs is the high content of modified nucleotides. Some modified bases are thought to be important for retroviral reverse transcription, and reverse transcription is less efficiently initiated from \textit{in vitro} synthesized, unmodified tRNA primers (3,4,35,37,38,65,79,97). The 3’ end of all tRNAs consists of a CCA-OH triplet that is added posttranscriptionally. The 3’-terminal 18 nucleotides of the tRNA molecule hybridize to the complementary PBS in the retroviral genome, and the first dNTP is added to the extreme 3’ end of the tRNA to initiate reverse transcription.

The tRNA primer that is used for reverse transcription by a particular retrovirus determines the sequence of the PBS. Mutation of the PBS-lys3 motif in the HIV-1 genome to sequences complementary to other tRNAs resulted in viruses with reduced replication capacity that use the new tRNA primers inefficiently to initiate reverse transcription (27,53,71,93,94). In addition, the PBS-mutated viruses revert to the wild-type PBS-lys3 sequence after several rounds of virus replication. The mutant viruses appear to recruit tRNA\(_{\text{lys}}\) as a primer at a low frequency, despite several mismatches with the mutant PBS sequence. As the PBS sequence is copied from the tRNA primer during plus-strand strong-stop DNA synthesis, this results in reversion to the natural PBS-lys3 sequence. Similar reversion events have been described for mutant avian retroviruses (96). These results indicate that retroviruses have a strong preference for the self-tRNA primer and demonstrate that the PBS is not the only determinant of tRNA primer selection. The viral RT enzyme and additional interactions between the tRNA primer and the viral RNA genome have been implicated in tRNA primer selection. These multiple levels of specificity may restrict aberrant reverse transcription from non-self tRNA primers.

2.4 Selective tRNA-packaging and tRNA-usage

Retrovirus particles contain approximately 30 tRNA molecules per virion, which represent a non-random subset of the cellular tRNA pool (33,61,63). The tRNA that is used as primer is usually packaged with the highest efficiency and represents 20-30% of the tRNA population in the virion (52,61,63,72,95). Although selective tRNA packaging could theoretically occur through interaction with the PBS in the viral RNA, there is no evidence for this. Virions with a PBS-mutated genome or virions lacking an RNA genome have a wild-type tRNA content (52,61,73). There is accumulating evidence that selective tRNA packaging is determined by the RT protein (26,45,58,61,62,70,73,80). RT-deficient viruses were found to contain a random subset of cellular tRNAs, suggesting a critical role for the RT enzyme in selective packaging of the tRNA primer. Primer selection most likely occurs in the early stages of virion assembly, when RT is part of the Gag-Pol precursor protein. Consistent with this idea, protease-deficient HIV-1 virions have a wild-type tRNA content, demonstrating that mature RT is not required for tRNA packaging (58,61).

The RT enzyme is also involved in placement of the tRNA primer onto the PBS (17,18). The HIV-1 RT enzyme stimulates annealing of the tRNA primer on the viral RNA genome \textit{in vitro}, and RT-deficient HIV-1 virions contain an RNA genome without an associated tRNA primer (61,71). In addition, the HIV-1 RT enzyme is strongly committed to the self-tRNA\(_{\text{lys}}\) primer for initiation of reverse transcription. Reverse transcription is not efficiently initiated from non-self tRNA primers that are annealed to PBS-mutated templates (71). These combined results demonstrate that the viral RT enzyme is involved in tRNA primer selection.

2.5 Additional viral RNA-tRNA contacts involved in reverse transcription

Interactions between viral RNA sequences flanking the PBS and the tRNA primer have been suggested to be important for reverse transcription. The PBS is located in the highly structured untranslated leader region at the 5’ end of the HIV-1 RNA genome (Fig 4), and several RNA secondary structure models have been proposed for the PBS region (6,8,25,76). One model is based primarily on
Fig 4. The 5' terminus of the HIV-1 proviral DNA genome and the viral RNA transcript. The HIV-1 proviral DNA genome is flanked by two long terminal repeats (LTRs) that consist of the U3, R and U5 regions. The transcriptional start site at +1 marks the border between the upstream (untranscribed) U3 region and the repeat (R) region of the LTR. The R region (positions +1 to 97) is present at each end of the RNA genome. The U5 region (positions 98 to 181) is encoded by the LTR, but unique for the 5' end of HIV-1 transcripts. The leader RNA ends at the AUG start codon of the gag open reading frame at position +336. The PBS (positions +182 to 199) is marked by a shaded box. The leader RNA is highly structured with distinct hairpin motifs that are named after their (putative) function in HIV-1 replication. The trans-activating responsive (TAR) hairpin regulates viral transcription from the LTR promoter by binding of the Tat protein. The polyA hairpin contains the polyadenylation site that mediates polyadenylation of viral transcripts at the 3' end of the viral genome. The dimer initiation site (DIS) is important for dimerization of the viral genome, the SD encodes the major splice donor site, and the packaging signal (Ψ) mediates selective encapsidation of the genome into assembling virions.

The phylogenetic analysis of different virus isolates, and predicts folding of the U5-PBS hairpin with an A-rich loop sequence. This hairpin occludes part of the PBS sequence in base-pairing (Fig 5A) (8). Sequences in the upstream U5 and downstream leader region form the extended U5-leader stem. An alternative model is based on biochemical probing experiments and predicts the U5-top hairpin, in which the A-rich sequence is present as an internal loop, and a shorter U5-leader stem (Fig 5B) (6,25). RNA secondary structures similar to that in the HIV-1 PBS region have been predicted for other retroviruses, including HIV-2, the simian immunodeficiency virus (SIV), the avian Rous sarcoma virus (Fig 5C) and the murine leukemia virus (MLV) (1,9,10,66,76).

Evidence for additional viral RNA-tRNA interactions was initially described for the avian Rous sarcoma virus (RSV). In RSV, initiation of reverse transcription is stimulated by an interaction between a 7-nucleotide sequence motif in the U5 region of the viral RNA genome and the TryC arm of the trNAbp primer (Fig 5C). Furthermore, RNA secondary structure in the RSV genome was demonstrated to be important for initiation of reverse transcription (21,22,24,64). Mutations that disrupt the U5-IR hairpin or U5-leader stem structure impair reverse transcription, whereas mutations that alter the sequence but that retain the structure have no effect. In addition, the distance between these structural elements appears critical for function (2,21,69). Studies with MLV support a role for the RNA secondary structure elements in reverse transcription (67).

A profoundly different viral RNA-tRNA interaction model has been proposed for HIV-1, based mainly on biochemical probing studies (35-37,39,47). The A-rich loop of the U5-PBS hairpin (A168-A171) was suggested to interact with the U-rich anticodon loop of the trNA primer (U33-U36), and the 3' portion of the anticodon stem of the tRNA molecule may interact with viral RNA sequen-
Fig 5. Secondary RNA structure models for the PBS region of HIV-1 and RSV RNA genomes and interactions with the corresponding tRNA primers. (A) Model for the PBS region of HIV-1 that is based primarily on phylogeny and that predicts folding of the U5-PBS hairpin directly upstream of the PBS. Sequences in the upstream U5 and downstream leader region form the extended U5-leader stem. (B) Model for the HIV-1 PBS region that is based primarily on biochemical probing experiments and that predicts folding of the U5-top hairpin directly upstream of the PBS and a shorter U5-leader stem. (C) Model for the PBS region of RSV that predicts folding of the U5-IR hairpin and the U5-leader stem. In addition, the cloverleaf structures of the corresponding tRNA molecules (tRNA^lys for HIV-1 and tRNA^pol for RSV) are shown. Base modifications are indicated according to standard nomenclature (81) (AC, anticodon loop; D, D loop). The (putative) interactions between the vRNA and tRNA molecule are marked. The interaction between the PBS and the 3’ 18 nucleotides of the tRNA molecule is marked by a light gray box. The interaction between the A-rich loop in the HIV-1 genome and the anticodon loop in tRNA^lys is marked by a dark gray box and the interaction between positions C142-U145 in the HIV-1 genome and tRNA^pol is marked by an open box. The interaction between the PAS in the viral RNA and the antiPAS sequence in the TPC arm of the tRNA molecule is marked by a black box.

2.6 Co-factors involved in reverse transcription

Virus encoded proteins or cellular factors may influence the process of reverse transcription. In particular, the viral nucleocapsid protein has been
suggested to play a role in reverse transcription. Extensive RNA structure is present in both the HIV-1 viral RNA template and the tRNA$^{\text{lys3}}$ primer. Therefore, partial unfolding of both RNA molecules is required for formation of the viral RNA-TRNA initiation complex. Although the RT enzyme itself may be able to disrupt secondary structure of the viral RNA and the tRNA primer (70), the viral NC protein has been proposed to be specifically involved in this process (reviewed in ref. (75)). The NC protein is considered to be a nucleic acid chaperone that catalyses structural rearrangements in the RNA. NC binds preferentially to single-stranded nucleic acids and unwinds tRNA in vitro, thereby stimulating the annealing of the tRNA primer onto the template (43,44,84) and initiation of reverse transcription (54,78). Stable secondary structure in the template RNA interferes with the elongation of reverse transcription (31,46,83), and the NC protein stimulates the processivity of DNA synthesis on such templates (42,46). Furthermore, NC stimulates the two strand-transfer steps during reverse transcription (16,30,54).

3. SCOPE OF THIS THESIS

The 5'-untranslated leader region of the HIV-1 RNA genome encodes multiple sequence motifs that are important for virus replication (Fig 4) (8). These sequences do not encode proteins, but regulate different steps in the viral replication cycle. This region of the RNA genome is highly structured with distinct hairpin motifs that regulate e.g. dimerization (DIS) and selective packaging (Y) of the viral genome into assembling virions. Furthermore, processes such as mRNA splicing (SD), polyadenylation (PolyA) and transcription (TAR) are controlled by sequence elements in the 5'-untranslated leader. This region also contains the PBS, the site at which the cellular tRNA$^{\text{lys3}}$ molecule anneals and reverse transcription is initiated.

Relatively little is known about the role of the sequences that flank the PBS motif in virus replication. These sequences have been proposed to basepair and fold the U5-PBS hairpin and U5-leader stem (Fig 5A). However, an alternative RNA secondary structure model was suggested for this region in which the PBS is flanked by the more extended U5-top hairpin and a somewhat shorter U5-leader stem (Fig 5B) (6,8,25,76). The function of the leader sequences 3' of the PBS (positions 200 to 243) is unknown. In the proviral DNA, sequences in the U5 region 5' of the PBS (positions 112 to 182) were demonstrated to be important for integration of the proviral genome into the host cell DNA (12,29,91), and several transcription factor binding sites were also proposed to be located in this region (28,51,89). In the RNA genome, a GU-rich enhancer of polyadenylation has been reported to be present in the U5 region (13,14). Several RNA motifs near the PBS have been suggested to interact with the tRNA$^{\text{lys3}}$ primer to stimulate reverse transcription (37,47,55-57,98). The A-rich loop of the U5-PBS hairpin (A168-A171) was suggested to interact with the U-rich anticodon loop of the tRNA primer (U33-U36), and the 3' portion of the anticodon stem of the tRNA molecule may interact with a sequence motif in the U5 region of the viral RNA (C142-U145) (35-37,39,47) (Fig 5A/B). However, these interaction models are based primarily on biochemical probing studies and were not tested in detail by mutational analysis. The interpretation of virus replication studies with A-rich loop deletion mutants (56) is complicated by the presence of overlapping sequence motifs for integration of the viral DNA genome.

To obtain information on the RNA secondary structure of the PBS region and to study the role of this region in HIV-1 replication, we generated various mutants based on the RNA secondary structure model shown in Figure 5A. The mutants were tested for virus replication, and subsequent analyses were performed to identify the viral function that is affected by the mutations, e.g. packaging of the RNA genome, transcription or reverse transcription. We also performed in vitro structure probing experiments and reverse transcription assays with purified HIV-1 reagents (RNA template, tRNA primer, RT enzyme). In addition, we used the "forced evolution" approach that has proven valuable in the analysis of regulatory RNA motifs. This approach is based on the finding that prolonged culturing of replication-impaired virus mutants can result in the selection of phenotypic revertants with an increased replication capacity. The appearance of revertant viruses is driven by spontaneous mutation of the viral genome due to reverse transcription errors during virus replication. The beneficial changes are subsequently selected by positive Darwinian evolution. Analysis of these revertant
genres can help to identify important RNA sequences and/or structures.

In chapter 2 we study the role of the U5-PBS hairpin in virus replication by introducing mutations that affect the stability of the hairpin. Both stabilization and destabilization inhibit virus replication and several virus revertants were obtained that restore the stability of the U5-PBS hairpin to approximately wild-type level. Stabilization of the hairpin inhibits tRNA annealing, and destabilization affects the correct placement of the tRNA primer. In chapter 3 we performed in vitro reverse transcription assays to further analyze the effect of the mutations on tRNA annealing and reverse transcription. In addition, the effect of the viral co-factor NC on reverse transcription was studied.

In chapter 4, we accumulate the HIV-1 leader sequence data of several “forced evolution” experiments that were performed over the years in our laboratory. G-to-A mutations are most frequently observed in the viral RNA genome, which probably reflect G-dT mispairing during the initial phase of reverse transcription. We argue that the frequent G-to-A mutations, which were recently reported to result from a new RNA editing mechanism (11), are in fact introduced by error-prone reverse transcription.

In chapter 5 we stabilize the U5-leader stem of HIV-1 to study its role in virus replication and reverse transcription. Stabilization of the base-paired stem inhibits the initiation and elongation steps of reverse transcription. The replication capacity of the mutant viruses is reduced, and a virus revertant was obtained that restores virus replication and reverse transcription by opening of the stabilized U5-leader stem. In chapter 6 we study the role of the U5-leader stem in further detail by a mutational analysis. We performed in vitro reverse transcription assays and identified a novel sequence motif in the U5 region that is important for tRNA-primed reverse transcription. Mutation of this motif inhibits the initiation of reverse transcription, whereas mutation of the “opposing” leader sequence stimulates reverse transcription. This enhancer motif was termed tRNA primer activation signal (PAS). We propose that the PAS interacts with the antiPAS sequence in the TΨC arm of tRNA^m3 (Fig 5A/B), similar to the interaction proposed for RSV and its tRNA^mp primer (Fig 5C). RNA structure probing experiments of the wild-type and mutant templates are consistent with the partially modified RNA secondary structure model that juxtaposes the critical PBS and PAS motifs (6,25) (Fig 5B). In chapter 7 we demonstrate that the PAS is also important for initiation of reverse transcription in vivo. We selected revertant viruses that partially overcome the reverse transcription defect of a PAS-deletion mutant. All revertants acquired a single nucleotide substitution that does not restore the PAS sequence, but that stimulates elongation of reverse transcription. These results indicate that the additional PAS-antiPAS interaction is needed to assemble an initiation-competent and processive reverse transcription complex. In chapter 8 we provide further evidence for the PAS-antiPAS interaction. We show that the efficiency of initiation of reverse transcription can be modulated by PAS mutations that strengthen or weaken the interaction with antiPAS. Furthermore, we demonstrate that the identity of the priming tRNA species can be switched by simultaneous alteration of the PBS and PAS sequences, indicating that the PAS is also involved in tRNA selection.

The studies described in this thesis provide novel insight into the process of reverse transcription. The PAS motif is important for initiation of reverse transcription, but is occluded by base-pairing in the U5-leader stem and mutation of the “opposing” leader sequence stimulates reverse transcription. Thus, reverse transcription appears to be restricted by RNA secondary structure in the wild-type HIV-1 genome. This may provide a mechanism for temporal regulation of reverse transcription. For instance, premature reverse transcription may be precluded in the virus-producing cell, such that the viral RNA genome is only copied upon infection of a new host cell. The additional PAS-antiPAS interaction also increases the specificity of reverse transcription and may restrict aberrant reverse transcription from non-self primers. Most importantly, extensive phylogenetic analysis of the different retrovirus genera that use other tRNA primers to initiate reverse transcription suggests that the PAS-antiPAS interaction is conserved in evolution. These combined results suggest that the process of reverse transcription is regulated by a common mechanism in all retroviridae.
REFERENCES


