Structure and function of the repeat region in the HIV-1 RNA genome

Vrolijk, M.M.

Citation for published version (APA):
Vrolijk, M. M. (2010). Structure and function of the repeat region in the HIV-1 RNA genome
Structure and function of the repeat region in the HIV-1 RNA genome

Martine Vrolijk

Uitnodiging
Voor het bijwonen van de openbare verdediging van het proefschrift van
Martine Vrolijk

Structure and function of the repeat region in the HIV-1 RNA genome
donderdag 28 januari 2010
14.00 uur
Agnietenkapel
Universiteit van Amsterdam
Oudezijds Voorburgwal 231
Amsterdam

Aansluitend bent u van harte welkom op de receptie in de Agnietenkapel

Info/Contact:
Martine Vrolijk
martinevrolijk@gmail.com
06-43986767

Paranimfen:
Marie-José Claessen
mjclaessen@gmail.com
06-22827741

Ying Poi Liu
y_pliu@hotmail.com
06-50607167

Bregje Vrolijk
bregjevrolijk@hotmail.com
06-44169180
Structure and function of the repeat region in the HIV-1 RNA genome

Martine Vrolijk
The research described in this thesis was performed at the Laboratory of Experimental Virology, Department of Medical Microbiology, Center of Infection and Immunity Amsterdam (CINIMA), Academic Medical Center, University of Amsterdam, The Netherlands. The research is financially supported by The Netherlands Organization for Scientific Research (NWO).

The printing of this thesis was financially supported by:
Boehringer Ingelheim B.V.
J.E. Jurriaanse Stichting
Tibotec, een divisie van Janssen-Cilag B.V.
University of Amsterdam

ISBN: 978-90-9025015-1
Cover design: Danielle de Graaff
Printed by: Gildeprint

Copyright © 2009 by Martine Vrolijk. All rights reserved. No part of this publication may be reproduced, stored or transmitted in any way without prior permission from the author.
Structure and function of the repeat region in the HIV-1 RNA genome

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit van Amsterdam op gezag van de Rector Magnificus prof. Dr. D.C. van den Boom ten overstaan van een door het college voor promoties ingestelde commissie, in het openbaar te verdedingen in de Agnietenkapel op donderdag 28 januari 2010, te 14:00 uur

door

Martine Mireille Vrolijk

geboren te Arnhem
PROMOTIECOMMISSIE

Promotor Prof. dr. B. Berkhout
Co-promotor Dr. A.T. Das
Overige leden Prof. dr. F. Baas
Prof. dr. J.M.A. Lange
Dr. R.C. Olsthoorn
Prof. Dr. P. Rottier
Prof. Dr. H. Schuitemaker

Faculteit der Geneeskunde
## CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHAPTER 1</td>
<td>General introduction / Scope of the thesis</td>
<td>7</td>
</tr>
<tr>
<td>CHAPTER 2</td>
<td>The TAR hairpin of human immunodeficiency virus type-1 can be deleted when not required for Tat-mediated activation of transcription</td>
<td>29</td>
</tr>
<tr>
<td>CHAPTER 3</td>
<td>Destabilization of the TAR hairpin affects the structure and function of the HIV-1 leader RNA</td>
<td>43</td>
</tr>
<tr>
<td>CHAPTER 4</td>
<td>Destabilization of the TAR hairpin of the HIV-1 genome causes aberrant RNA dimerization and packaging</td>
<td>65</td>
</tr>
<tr>
<td>CHAPTER 5</td>
<td>Destabilization of the TAR hairpin leads to extension of the polyA hairpin and inhibition of HIV-1 polyadenylation</td>
<td>79</td>
</tr>
<tr>
<td>CHAPTER 6</td>
<td>Summary and Discussion</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Samenvatting</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>Dankwoord</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>Curriculum Vitae</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>Publications</td>
<td>111</td>
</tr>
</tbody>
</table>
CHAPTER ONE

General Introduction
Scope of the Thesis
Chapter 1

General Introduction

The retrovirus HIV-1

Human immunodeficiency virus type I (HIV-1) was identified as the causative agent of acquired immunodeficiency syndrome (AIDS) in 1983 (1, 2). A related virus, which is now known as HIV-2, was isolated from AIDS patients in West-Africa in 1986 (3). HIV-1 and HIV-2 belong to the lentivirus genus, a subfamily of the Retroviridae. Retroviruses carry a diploid positive stranded RNA genome, which is copied into double-stranded DNA by the viral reverse transcriptase (RT) enzyme. Upon infection of the host cell, the proviral DNA is integrated into one of the host chromosomes. The provirus may remain dormant for some time or induce the expression of new virus particles.

Organization of the HIV-1 genome and structure of the virus particle

The proviral DNA genome is approximately 9.5 kbp in length and encodes nine genes of which three (gag, pol and env) are present in all retroviruses. The gag gene encodes the Gag polyprotein that is proteolytically processed during virion assembly in the structural proteins matrix (MA), capsid (CA) and nucleocapsid (NC) (Fig. 1A). The pol gene encodes the viral protease (PR), RT and integrase (IN) enzymes. These proteins are produced from a Gag-Pol polyprotein precursor that is processed by the protease during particle maturation, ensuring incorporation of all three enzymes into the virus particle. The env gene encodes the envelope (Env) glycoprotein that is present on the surface of the viral particle and that mediates attachment to specific cell surface receptors and fusion of the viral membrane with the cell membrane. The Env protein is expressed as a precursor protein (gp160) that is cleaved into the gp120 surface protein and the gp41 trans-membrane domain (4). The HIV-1 genome also contains open reading frames for the two essential regulatory proteins Tat and Rev. The viral transactivator protein Tat is required for the activation of HIV-1 transcription (5, 6). Rev is required for nuclear export of unspliced and singly spliced HIV-1 transcripts to the cytoplasm (7-11). In addition, the HIV-1 genome encodes the accessory viral proteins Vif, Vpr, Vpu and Nef. These proteins are not strictly required for viral replication in some cell culture systems, but are important for efficient virus spread and pathogenesis in vivo (12-14).

Besides the open reading frames, the HIV-1 genome contains important non-coding domains. The long terminal repeat (LTR), present at the 5’ and 3’ end of the proviral DNA genome, contains the viral promoter and encodes the untranslated leader and trailer sequences of all HIV-1 mRNAs. The LTR is subdivided in the U3 (unique 3’ terminus), R (repeat) and U5 (unique 5’ terminus) domains (Fig. 1A). The first residue of the R region in the 5’ LTR marks the transcriptional start site and thus the 5’ end of all HIV-1 mRNAs. The RNA transcript is polyadenylated at the last nucleotide of the R region encoded by the 3’ LTR.
Figure 1. The HIV-1 genome and viral particle. (A) The HIV-1 DNA genome showing the open reading frames for the viral proteins Gag, Pol, Vif, Vpr, Vpu, Tat, Rev, Env and Nef. The LTRs encode regulatory DNA and RNA elements that are required for viral replication. The 5' LTR acts as a promoter for the synthesis of viral RNA. The 3' LTR encodes an active polyadenylation signal. The 5' end of the genomic RNA consists of R and U5 sequences, while the 3' end is formed by U3 and R. (B) Schematic of an HIV-1 particle, see the text for further details.

HIV-1 is an enveloped RNA virus that contains two copies of the positive stranded viral RNA genome (Fig. 1B). In the mature virus particle, the RNA genomes are closely associated with the NC protein to form an RNA-protein complex. This complex and the viral enzymes PR, RT and IN are packaged in the cone-shaped virion core, which is mainly formed by the CA protein. Outside the core, the MA protein forms a shell underneath the lipid membrane bilayer that is derived from the host cell. The Env gp41 subunit is anchored in the virion membrane and the gp120 subunit is non-covalently linked to gp41. This gp41-gp120 complex is expressed as an Env trimer on the surface of the viral particles.
The retroviral replication cycle
A schematic of the HIV-1 replication cycle is shown in figure 2. HIV-1 replicates in CD4-positive cells, including T lymphocytes, monocytes and macrophages. The HIV-1 virion attaches to the host cell via its gp120 protein that binds to the CD4 receptor. This interaction triggers a conformational change in gp120 that facilitates binding to the CCR5 or CXCR4 co-receptor. Subsequently, the gp41 subunit undergoes a conformational change that mediates fusion of the viral membrane with that of the host cell (15, 16). Within the virion core, the viral RNA genome is copied into DNA by the viral RT enzyme (17). The virion core is targeted to the cell nucleus where the full-length DNA is inserted into the host genome by the viral IN enzyme (18). The integrated proviral DNA serves as a template for the production of different viral mRNA species. Initially, spliced mRNAs that encode the regulatory Tat and Rev proteins are produced. Accumulation of the Rev protein triggers a shift towards functional expression of unspliced and singly spliced mRNAs that encode the Gag, Pol and Env proteins. These proteins, together with two copies of unspliced viral RNA,
assemble at the cell membrane to form the progeny virions, which are released from the cell by budding. The progeny virions mature by processing of the Gag and Gag-Pol precursor proteins to generate infectious particles that can infect a new host cell. A detailed description of the HIV-1 replication cycle can be found in (19).

**The HIV-1 leader RNA**

This thesis addresses the structure and function of motifs that are present at the extreme 5' and 3' end of the HIV-1 RNA genome. All HIV-1 mRNAs have an untranslated region at both the 5’ and 3’ end, to which we refer as leader and trailer region, respectively. The sequences and structures within the HIV-1 leader RNA are highly conserved and contain several key signals that are necessary in the regulation of various HIV-1 replication steps, such as transcription, translation, polyadenylation, RNA dimerization and packaging (Fig 3). Two of these structured RNA signals are present in the 97-nt R region and thus repeated in the 3' trailer region of all HIV-1 mRNAs. In the next paragraphs of this introduction I will give an overview of the regulatory functions that are encoded by the leader and trailer RNA.

**Transcription**

The HIV-1 DNA genome is transcribed by RNA polymerase II from the 5' LTR promoter that spans the U3 and R regions. The regulation of HIV-1 transcription has been studied extensively and requires remodeling of the chromatin and binding of cellular transcription factors (reviewed in (6, 20-23)). In addition, the viral Tat protein strongly enhances transcription. This Tat-dependent trans-activation requires an RNA hairpin that is located at the 5’ end of the untranslated leader RNA, the Trans Activation Response (TAR) element (Fig. 3).

Although early studies on HIV-1 gene expression identified the TAR sequence as an important regulatory element for viral gene expression (24). It was initially unclear whether TAR acted as a regulatory DNA or RNA element. Subsequent experiments demonstrated that proper folding of the RNA hairpin as part of the nascent transcript is required for Tat-dependent transcription, thus providing the first evidence that TAR acts as an RNA enhancer (25). The 57-nt TAR hairpin (Fig. 3) contains a highly conserved 3-nt pyrimidine bulge that directly interacts with the basic domain of the Tat protein and an apical 6-nt loop that binds cyclin T1 in a Tat-dependent manner. Cyclin T1 is a component of the positive Transcription Elongation Factor complex (pTEFb), which includes CDK-9 kinase that phosphorylates the C-terminal domain of RNA polymerase II (26-29). This phosphorylation increases the processivity of the elongating polymerase (30-32). Initiation of transcription is also stimulated by Tat (33, 34) and it was recently demonstrated that pTEFb directs the recruitment of TATA-box-binding protein (TBP) to the LTR promoter to stimulate the assembly of new transcription complexes (35).
Splicing
The Gag and Pol proteins are translated from the unspliced HIV-1 transcript and the Env protein is made from a singly spliced mRNA. The other viral proteins (Tat, Rev, Nef, Vpu, Vif and Vpr) are expressed from multiply spliced mRNAs. The HIV-1 RNA genome contains a large number of competing splice donor and acceptor sites that give rise to over 40 differentially spliced HIV-1 mRNAs (36). There is recent evidence that splicing is regulated by RNA structure (37). Early in the infected cells, the HIV-1 RNA is multiply spliced to produce more than 20 mRNAs that are exported to the cytoplasm and translated to produce the Rev, Tat and Nef proteins (Fig. 4). In the late phase, unspliced and singly spliced mRNAs for the Gag, Pol and Env proteins are generated (reviewed in (38)). These late mRNAs all contain the Rev Responsive element (RRE), a structured RNA element within the Env open reading frame to which the Rev protein binds. The Rev-RRE complex subsequently interacts with cellular factors, such as CRM1 (39, 40), Ran-GTP (41) and possibly eIF2A (42) to facilitate nuclear export of unspliced and singly spliced transcripts to the cytoplasm (10, 43, 44). The unspliced transcript is of particular interest since it serves both as the mRNA for the Gag and Gag-Pol proteins and as the viral genome that is packaged in virions.

The large number of differentially spliced HIV-1 mRNAs share a common untranslated leader sequence of 290 nucleotides, which is determined by the use of the major splice donor site (SD) in all spliced products (36) (Fig. 4). The unspliced HIV-1 transcript contains the uninterrupted untranslated leader sequence of 335 nucleotides. At the 3' end, all HIV-1 mRNAs share an untranslated trailer sequence of 97 nucleotides.

Capping and polyadenylation
Eukaryotic RNA polymerase II transcripts are covalently modified at the 5' triphosphate end by enzymatic addition of a methylated guanosine monophosphate, which is called the cap. Capping occurs rapidly after transcription initiation and is required for splicing, export of the RNA from the nucleus, protection from exonucleolytic decay and initiation of translation (45-50).

The 3' end of most eukaryotic mRNAs is processed by means of polyadenylation, which provides a tail of adenosine residues. This post-transcriptional
Figure 4. Differently spliced forms of HIV-1 RNA. Transcription of the proviral DNA genome produces spliced and unspliced viral RNAs. In the early phase of infection, multiply spliced mRNAs for the Tat, Rev and Nef proteins are produced (as an example the Tat mRNA is shown). Late mRNAs are singly spliced or unspliced and contain the Rev Responsive element (RRE) that mediates nuclear export of the transcript upon binding of Rev.

Modification requires cleavage of the primary transcript and addition of the adenosine tail by polyA polymerase. The AAUAAA hexameric sequence located upstream of the cleavage site serves as a near-universal polyadenylation signal in combination with a downstream positioned, more variable GU- or U-rich sequence (51, 52). The AAUAAA hexamer binds the cleavage and polyadenylation specificity factor (CPSF) and the downstream element interacts with the cleavage stimulation factor (CstF) (53-57).

HIV-1 and HIV-2, as well as several other retroviruses, encode the AAUAAA hexamer sequence in the R region that is present at both 5' and 3' ends of the viral transcript (Fig. 1A). This requires a regulatory mechanism to prevent recognition of the 5' polyadenylation signal and to ensure efficient polyadenylation at the 3' end. Several factors have been implicated in the regulation of the 5' and 3' polyadenylation site usage. Initial studies demonstrated that the 3' polyadenylation signal in HIV-1 is activated by U-rich sequences that are uniquely present in the U3 region upstream of 3' R (58, 59). In addition, downregulation of the 5' polyA signal has been associated with its proximity to the LTR promoter, where the assembly of RNA polymerase II initiation complexes may interfere with polyadenylation (60-62). Other studies demonstrated that suppression of the 5' polyadenylation site is strongly dependent on the proximity of the major splice donor site (SD) (63-65). In HIV-1, the major SD is located approximately 200 nucleotides downstream of the AAUAAA hexamer and serves as a binding site for U1 small nuclear ribonucleoprotein (snRNP).
Binding of U1 snRNP to the HIV-1 splice donor, but not active splicing, was shown to result in with the occlusion of the upstream polyadenylation site (65).

In addition, there is ample evidence that the structure of the HIV-1 leader RNA is involved in regulating the polyadenylation site activity (66-68). A hairpin that contains part of the AAUAAA hexamer in the basepaired stem region can fold in the R region of HIV-1 and is essential for virus replication (69) (Fig. 3). The structure of this polyA hairpin is highly conserved among different HIV-1 strains and actively represses premature polyadenylation at the 5’ end through occlusion of the polyadenylation signal AAUAAA (70). This ensures that premature polyadenylation at the 5’ end is suppressed through rapid folding of the polyA hairpin in the nascent RNA transcript. Studies with purified polyadenylation factors and cellular extracts suggest that additional suppression is obtained by sequences that are exclusively present in the 5’ leader RNA, possibly through an RNA structural effect (71). This rather stable hairpin is formed exclusively in 5’ R and not in 3’ R, thus allowing polyadenylation at the 3’ end.

Translation
Protein synthesis in eukaryotic cells is mainly regulated at the initiation level. The majority of mRNA translation events is initiated through recognition of the 5’ cap structure by translation initiation factors (72, 73). These factors attract the 40S ribosomal subunit and migrate or scan along the RNA until an AUG start codon in a favorable sequence context is encountered (74). Alternatively, translation can be initiated independent of cap recognition by direct binding of the 40S subunit to structured RNA signals termed Internal Ribosome Entry Sites (IRES) (75, 76). Many viruses, including a large number of retroviruses, initiate translation by an IRES-dependent mechanism (77-82).

The HIV-1 untranslated leader RNA usually does not contain any AUG start codons preceding the Gag open reading frame (83), which is consistent with the scanning mode for ribosomes. On the other hand, stable secondary structures in the leader RNA could potentially interfere with ribosome scanning. Several studies have demonstrated a repressive effect of the HIV-1 TAR hairpin on translation that can be alleviated by cellular proteins and the viral Tat protein (84, 85). Indeed, a translational component of the TAR-Tat mediated activation of transcription has been suggested (86-89). TAR has been shown to interact directly with the initiation factor eIF2 and this interaction stimulates protein synthesis from TAR-containing RNAs (90). Interactions of both Tat and TAR with the interferon induced double-stranded RNA-dependent protein kinase (PKR), which inactivates eIF2, have also been linked to the regulation of protein synthesis. TAR RNA induces PKR activity and the interaction of Tat with PKR counters this effect (91-97), but another study demonstrated that TAR RNA blocks PKR dimerization when present in large quantities (98). It is currently not known how these opposing effects on protein synthesis are integrated to regulate HIV-1 gene expression.
There have been indications that translation of the HIV-1 Gag protein involves IRES activity. One study reported IRES activity of the HIV-1 leader RNA in the G2/M phase of the cell cycle (99). The presence of an IRES within the Gag open reading frame has also been suggested (100). Studies with spliced HIV-1 transcripts generally favor a cap-dependent scanning mechanism for the initiation of translation (101, 102), but the translational efficiency differs markedly with the identity of the untranslated leader RNA in differentially spliced RNAs (36). For example, it has been suggested that the translation of the Tat mRNA is dependent on IRES activity and stimulated by the Tat protein (103). In addition, leaky scanning, translational interference and re-initiation have been proposed to affect the translation of downstream open reading frames in multicistronic spliced mRNAs (101, 102, 104). It thus appears that initiation of translation may occur by different mechanisms on spliced and unspliced HIV-1 transcripts. Translation of the full length HIV-1 RNA results in the production of both Gag and Pol. The Pol protein is produced as the Gag-Pol polyprotein. This Gag-Pol precursor is generated by a ribosomal frameshifting event, which is triggered by specific cis-acting RNA motifs, including a slippery sequence and a short stem-loop structure at the 3’ end of the Gag open reading frame (105-107). During Gag translation, 5% of the ribosomes shift their reading frame to produce the Gag-Pol protein without interrupting translation.

RNA genome dimerization

The genomic RNA is present in retroviral particles as an RNA dimer. This dimeric status is thought to help circumvent the effect of physical damage to the RNA genome as the RT enzyme can switch to the other RNA copy when a nick in the RNA is encountered during reverse transcription. Such template switching can also contribute to the genetic diversity of retroviruses (108). The dimer is maintained by RNA-RNA interactions since proteins can be removed without affecting the integrity of the dimer. Furthermore, the dimeric RNA can be dissociated by heat treatment, indicating that the dimer linkage is mediated by non-covalent interactions. Electron microscopy studies with partially denatured RNA demonstrated that the dimer linkage site maps to the 5’ end of the retroviral RNA and that the strands are arranged in parallel fashion (109, 110). It has been observed that the RNA dimer matures during particle assembly, which involves a change in thermostability of the dimer from a loose to a tight interaction (110). Several studies indicate that RNA dimerization may be linked to packaging of the viral genome (111-113).

In vitro synthesized RNA corresponding to retroviral leaders dimerize readily in vitro in response to cations, heat-treatment or the presence of the viral NC protein (114-117). This observation has caused intensive efforts to identify the cis-acting RNA elements that trigger RNA dimerization in vitro. Initial studies suggested the involvement of G-rich sequences, which are reminiscent of G-tetrad structures that have been implicated in maintaining the integrity of chromosome telomeres (118). In subsequent studies, the involvement of G-rich elements has been dismissed (119) in favor of a dimerization mechanism mediated by a stem-loop structure with an
autocomplementary (palindromic), sequence that is exposed in the hairpin loop (120-122). This hairpin has been termed the dimer initiation site hairpin (DIS) and is located upstream of the major splice donor site in the HIV-1 leader RNA (Fig. 3). Direct base pairing between the loop-exposed palindromes of two DIS hairpins yields an instable RNA dimer termed the kissing loop complex (Fig. 3). In the presence of the viral NC protein or upon heat treatment, the kissing loop complex rearranges such that the hairpin stems are melted, allowing the formation of an extended duplex dimer (120, 123-127).

DIS-like stem-loop structures are present in the untranslated leader of many retroviral genomes, suggesting a common mechanism of RNA dimerization (128-133). However, it has proven difficult to correlate the results of in vitro dimerization studies with virus replication experiments. Mutations in the DIS palindrome have surprisingly small effects on RNA dimers in virus particles, but do affect viral replication due to defects in RNA packaging and reverse transcription (113, 134-136). Several studies demonstrated that regions outside the DIS motif also contribute substantially to RNA dimerization in vivo (111, 112, 137), although this thesis will argue that this is likely due to indirect RNA structure effects.

RNA packaging
The retroviral RNA genome is selectively packaged into progeny virions by a highly specific process that involves interactions between the viral Gag protein and RNA elements in the full length viral transcript. Important packaging signals have been mapped to the 5′ untranslated leader RNA for a number of retroviruses (138). The core packaging signal is located downstream of the major splice donor, which ensures that splicing removes a critical part of the packaging signal (4). Thus, the location of the packaging signal provides the means by which unspliced RNA is discriminated from spliced viral RNA.

Early studies confirmed that RNA elements important for packaging of the HIV-1 genome are located downstream of the major SD (139-142). Several conserved stem-loop structures were identified within this domain, including a hairpin with a GGAG tetraloop (the Ψ hairpin, Fig. 3) that interacts strongly with the Gag-derived NC protein and that has been regarded as the core HIV-1 packaging signal (128, 143-146). However, other studies demonstrated that the HIV-1 packaging signal is in fact multipartite and includes several domains upstream of the major SD (111, 147-149). These data may suggest that the overall structure of the untranslated leader is recognized for selective packaging of the genome. But as will be argued in this thesis, part of these data may also be explained by indirect effects at the level of the leader RNA structure.

HIV-1 is not unique in having such an extensive packaging requirement. Analogous observations have been made for several other retroviruses, including HIV-2 (150-156). It is currently not known how genomic RNA is selectively recognized in this context, since packaging signals will be present on both the spliced and unspliced transcripts. Indeed, it has been reported that spliced HIV-1 transcripts
are efficiently packaged when the Ψ domain is mutated (112, 157-159), confirming that the spliced RNA does contain at least part of the packaging determinant. It was shown that reduced packaging of full length HIV-1 RNA coincides with increased packaging of spliced molecules, indicating that these RNAs compete for the same trans-acting factor for packaging (160, 161).

An alternative packaging mechanism was described for the genomic RNA of Moloney murine leukemia virus genomic RNAs (162). It was shown that RNA dimerization results in the exposure of UCUG elements, which are specifically recognized by NC protein. Although several studies indicate that active translation is not required for packaging (147, 163-165), translated transcripts appear to have a packaging advantage (166, 167). Possibly, the unspliced transcript, which also serves as the mRNA for the Gag and Gag-Pol proteins, associates with the Gag protein cotranslationally. However, for HIV-1 and HIV-2 there appear to be no different pools of translated versus packaged RNA (165).

**Reverse transcription**

The dimeric RNA genome present in virus particles serves as the template for the viral RT enzyme, which generates the double-stranded DNA that is subsequently integrated into a host chromosome. The highly complicated process of reverse transcription has been reviewed in detail by several authors (17, 19, 168-171). Reverse transcription is initiated near the 5' end of the RNA genome and elongation proceeds in the 3' to 5' direction until the extreme 5' end of the RNA is reached. The RT enzyme initiates reverse transcription using a cellular tRNA as a primer. The tRNA primer anneals to the RNA at the primer binding site (PBS), this PBS is located directly downstream of the U5 and is complementary to the 5' terminal nucleotides of the tRNA. HIV-1 and HIV-2 both use tRNA\(^{\text{lys3}}\) for priming of reverse transcription, and both viruses have a PBS with an 18 nucleotide complementarity to the tRNA. Selection of the tRNA\(^{\text{lys3}}\) from the pool of cellular tRNAs is not mediated by the PBS alone, as the RT enzyme specifically recognizes its cognate primer (172, 173). In addition, there is evidence that the corresponding tRNA synthetase is specifically packaged into virions (174-176).

During DNA synthesis the RT enzyme removes the RNA template by means of its RNaseH activity, which degrades the RNA in RNA-DNA heteroduplexes. The initial single stranded cDNA product is called strong-stop cDNA. The 5' and 3' terminal ends of the RNA genome contain the R-sequence, which allows the strong stop cDNA to base pair with its complementary part at the 3' end of the genome, where elongation is resumed. This process of translocating the cDNA from the 5' end to the 3' end is termed strand transfer. Subsequent elongation and a second transfer event results in generation of a double stranded DNA provirus flanked by two complete LTRs.

Several additional interactions between the tRNA primer and the viral genome have been proposed that contribute to the specificity of primer usage. On the basis of RNA structure probing it has been suggested that an A-rich sequence
upstream of the HIV-1 PBS interacts with the anticodon loop of the tRNA\textsubscript{lys3} primer (177-181). Indirect virological evidence has been presented in support of this interaction (182-184). However, the role of this interaction in reverse transcription was not confirmed by studies from another laboratory (185). Furthermore, the ability to form base pairs between the tRNA anticodon loop and the viral genome is poorly conserved among other retroviruses, including the closely related SIV isolates. It has been argued that base pairing between the A-rich sequence and the tRNA\textsubscript{lys3} anticodon loop introduces highly unusual structural consequences such as severe bending, extensive knotting and unwinding of helices (186). It seems more likely that the conservation of the A-rich sequence motif in HIV-1 RNA is due to the role of the corresponding DNA sequences in integration of the provirus DNA (187-192).

An alternative base pairing interaction between sequences in the TΨC arm of the tRNA\textsubscript{lys3} and an eight nucleotide sequence element in the HIV-1 U5 region has been proposed (193, 194). Mutations in this sequence severely decrease the initiation of reverse transcription and it has hence been termed the Primer Activation Signal (PAS). Base pairing of the tRNA primer TΨC arm and the viral genome is highly conserved among retroviruses (195) and has been shown to regulate reverse transcription of the avian Rous Sarcoma Virus (RSV) (196) and yeast TY1 and 3 retrotransposons (197). In HIV-1, PAS-mediated initiation of reverse transcription appears to be regulated by RNA structure since the PAS element is occluded by base pairing within the leader RNA before binding of the tRNA primer (194).

In addition to these sequence and structural requirements for the initiation of reverse transcription, elongation by the RT enzyme imposes further constraints on structured elements within the RNA. Extremely stable helices pose an obstacle to the elongating RT enzyme and result in the accumulation of premature stop products (198, 199). On the other hand, there is also evidence that the secondary structure within the 5’ R region actively contributes to the strand transfer mechanism, in addition to the required sequence complementarity between the 5’ R region and the 3’ R-cDNA (200). Furthermore, the integrity of the RNA dimer in virus particles seems required to perform the first strand transfer \textit{in vivo} (201).

**The HIV-1 leader RNA structure**

The HIV-1 leader RNA is involved in several essential steps of the HIV-1 replication cycle. Several groups made efforts to elucidate the structure of this RNA domain. Two studies proposed two rather different RNA structure models. Baudin proposed an elongated fold with many internal loops based on structure probing assays (202). Berkhout identified a number of small stem-loop structures that are conserved among HIV and SIV isolates by comparative sequence analysis (128). Both models are different except for the presence of the TAR hairpin at the 5’ end of the leader RNA. A third study demonstrated that the results of the structure probing study are in fact compatible with most of the proposed stem-loop structures (145). Although \textit{in silico} analysis of the structures displayed low probability of folding (203), the TAR hairpin again emerged as the only consistently predicted RNA structure. Several of
the stem-loop structures proposed by Berkhout have since been confirmed in biological experiments (143, 204-210). A recent study using SHAPE technology largely confirmed the proposed structure motifs in the HIV-1 leader RNA (211, 212).

Our laboratory observed a strange behavior for HIV-1 transcripts corresponding to the complete leader RNA in non-denaturing gel electrophoresis. More specifically, larger transcripts migrated faster through the gel than shorter transcripts (213). It was initially thought that a compact tertiary structure caused this behavior, but it was later shown that the HIV-1 leader RNA structure can adopt two alternatively folded RNA structures. The originally proposed branched multiple hairpin (BMH, Fig 3) conformation can switch into an energetically more stable conformation in which the polyA and DIS hairpins are opened and their sequences interact (214-216). This long-distance interaction (LDI, Fig. 3) prevents exposure of the DIS element and the formation of RNA dimers in vitro (214, 217, 218). The BMH and LDI conformers may thus provide the virus with a riboswitch that controls leader sRNA functions like dimerization and packaging. In the BMH conformation, the AUG start codon can interact with sequences downstream of the polyA hairpin (216, 219, 220), which opens up the possibility that translation is also regulated by the same riboswitch. However, a more recent study showed that the HIV-1 leader RNA conformational switch does regulate in vitro RNA dimerization, but not mRNA translation (217). A similar mechanism was proposed for HIV-2 since this leader can also adopt an alternative conformation that regulates RNA dimerization in vitro (130, 221).

The pleiotropic TAR hairpin
The TAR hairpin that is present at the 5’ end of all viral transcripts plays an essential role in transcription (Fig. 3) and several studies have shown that the structure of the stem and the sequence of the bulge and loop are critical for the Tat-mediated trans-activation of transcription (222). In addition to its role in transcription, the 5' TAR hairpin has been suggested to be important for dimerization of the viral RNA genome (223) and packaging of the viral RNA into virions (112, 148, 149, 215, 224). The TAR hairpin at the 3’ end of viral transcripts has been proposed to have a spacer function in the process of polyadenylation by juxtaposing the AAUAAA polyadenylation signal located in the polyA hairpin and an upstream sequence element (USE) that enhances CPSF binding (71, 225). Both the 5' and 3' TAR hairpin have been suggested to be important for the strand transfer step of reverse transcription (200), and as a possible HIV-1 derived miRNA with a role in latency (226, 227). However, many of these additional TAR functions remain to be validated in the physiological setting of virus replication.
SCOPE OF THE THESIS

Given the multitude of structured RNA signals in the untranslated leader of the HIV-1 RNA genome, it may be clear that one should be extremely careful when inactivating a single signal by mutation, as this may have an unwanted effect on other signals by a change of the overall RNA structure. Such a danger of unwanted RNA structure effects is even more prominent as this part of the viral genome acts as a riboswitch that is very sensitive to mutations. We will argue in this thesis that several indirect leader defects are induced through RNA structure changes when one mutates the 5' TAR hairpin. This result may be surprising as the TAR hairpin is present as a separate hairpin structure in both the LDI and BMH conformations of the riboswitch. Moreover, we will demonstrate that such TAR modifications do not only affect the structure of the 5' untranslated leader, but also the 3' R trailer region. In this thesis, we focus on a set of TAR mutants and determine the effects of these mutations on different processes involved in HIV-1 replication using a combination of in vitro and in vivo approaches.

We created a set of HIV-1 variants with deletions in TAR. These variants were based on a genetically engineered HIV-1 strain that does not need TAR for transcription. This special viral context allowed us to address the other putative functions of the TAR hairpin (228, 229). Deletions were introduced in the left or right side of the TAR hairpin, which resulted in the opening of the RNA structure. In addition, these deletions were combined to truncate the TAR hairpin or TAR was completely deleted (Chapter 2). We observed that deletions in one side of the stem-loop structure abolished virus replication, whereas truncation or complete removal of the hairpin did not seriously affect replication. This important result set the stage for follow-up studies to find out what goes wrong when the TAR hairpin is opened, either at the 5' or 3' end of the HIV-1 RNA genome.

We used the same set of mutants to synthesize transcripts of the untranslated leader RNA in vitro and subjected them to denaturing and non-denaturing gel electrophoresis and thermal melting assays (Chapter 3). The biochemical profiles demonstrated a switch from the LDI to the BMH structure of the HIV-1 leader RNA when TAR is opened, with a concomitant deregulation of in vitro RNA dimerization. Several mutants were subjected to RNA structure probing analysis and this confirmed the switch to the BMH structure. Prolonged culturing of the mutant viruses resulted in the selection of revertant viruses that restore RNA folding and the wild type RNA dimerization phenotype.

Next, we analyzed the RNA genome in virions and observed that destabilization of the 5' TAR hairpin also resulted in a packaging defect (Chapter 4). This defect in packaging of full-length HIV-1 RNA seems to trigger the packaging of spliced HIV-1 transcripts. We show that the TAR hairpin is not essential for RNA packaging, but that destabilization of TAR does indirectly affect this process, likely due to the structural rearrangements within the leader RNA.
We next focused on TAR in the 3’ trailer of HIV-1 transcripts. We analyzed HIV-1 RNA from infected cells on a denaturing Northern blot and observed transcripts of increased size for the TAR-destabilized mutants, which is indicative of a 3’ polyadenylation defect (Chapter 5). We demonstrate that opening of the TAR hairpin frees nucleotides that subsequently trigger an extension of the adjacent polyA hairpin. This provides a molecular explanation for the polyadenylation defect as the stability of the polyA hairpin modulates the recognition of the AAUAAA signal, which is partially occluded by base pairing.
REFERENCES


The TAR hairpin of human immunodeficiency virus type-1 can be deleted when not required for Tat-mediated activation of transcription

Atze T. Das, Alex Harwig, Martine M. Vrolijk and Ben Berkhout

JOURNAL OF VIROLOGY (2007)


**ABSTRACT**

The human immunodeficiency virus type-1 (HIV-1) RNA genome contains a terminal repeat (R) region that encodes the trans-acting responsive (TAR) hairpin, which is essential for Tat-mediated activation of gene expression. TAR has also been implicated in several other processes during viral replication, including translation, dimerization, packaging and reverse transcription. However, most studies in which replication of TAR-mutated viruses was analyzed, were complicated by the dominant negative effect of the mutations on transcription. We therefore used an HIV-1 variant that does not require TAR for transcription to reinvestigate the role of TAR in HIV-1 replication. We demonstrate that this virus can replicate efficiently upon complete deletion of TAR. Furthermore, evolution of a TAR-deleted variant in long-term cultures indicates that HIV-1 requires a stable stem-loop structure at the start of the viral transcripts, in which the 5’-terminal nucleotides are base-paired. This prerequisite for efficient replication can be fulfilled by the TAR hairpin, but also by unrelated stem-loop structures. We therefore conclude that TAR has no essential function in HIV-1 replication other than to accommodate Tat-mediated activation of transcription.

**INTRODUCTION**

Retroviral RNA genomes contain a sequence repeat (R) that forms the extreme 5’ and 3’ ends of the viral transcripts. The trans-acting responsive (TAR) region of the 97-nt R region of the human immunodeficiency virus type 1 (HIV-1) RNA can fold a stable hairpin structure (1)(Fig. 1), which has been suggested to play multiple important roles in viral replication. The best-studied function of this TAR hairpin is its essential role in the activation of transcription from the promoter in the 5’ long terminal repeat (LTR) of the proviral genome (reviewed in (2, 3)). Important features in TAR are the highly conserved 3-nucleotide pyrimidine bulge in the stem, which binds the viral Tat transactivator protein (4), and the apical 6-nucleotide loop to which the cyclin T1 subunit of the positive transcriptional elongation factor (pTEFb) binds in a Tat-dependent manner (5, 6). Upon binding, the kinase component of P-TEFb, CDK9, can phosphorylate the C-terminal domain of RNA polymerase II, which enhances the processivity of the elongating polymerase (7, 8). Furthermore, it was recently demonstrated that pTEFb also directs the recruitment of TATA-box-binding protein (TBP) to the LTR promoter and thus stimulates the assembly of new transcription complexes (9).

There have been numerous reports in which additional functions of TAR have been proposed in translation, dimerization, packaging and reverse transcription of the viral transcripts (reviewed in (3, 10-12)). More recently, it has been suggested...
that TAR may also affect the cellular RNA interference process (13, 14). Thus, a pleiotropy of functions has been attributed to the TAR motif in a variety of experimental systems. The biologically most relevant assay system is that of the replicating virus, and the importance of TAR is underlined by the observation that mutations within TAR cause severe replication defects. However, these studies are complicated by the fact that nearly all mutations in TAR affect the transcription process. This dominant effect on transcription makes it difficult or impossible to distinguish the effect of such mutations on other processes during virus replication.
We and others previously reported the construction of an HIV-1 variant that does not depend on the Tat-TAR interaction for activation of transcription (15-17). In our HIV-rtTA variant, both the Tat protein and its TAR binding site were inactivated by mutation (Fig. 1A) and functionally replaced by the components of the Tet-On gene regulation system (18). The gene encoding the rtTA transcriptional activator protein was inserted in place of the 3’-terminal nef gene, and tet operator (tetO) binding sites were introduced in the LTR promoter. Administration of the effector doxycycline (dox) induces a conformational switch in the rtTA protein, which subsequently can bind to the tetO-LTR promoter region and activate transcription of the proviral genome. Thus, transcription and replication of HIV-rtTA are critically dependent on dox. Since this virus does not require TAR for the activation of transcription, it is the ideal tool to study non-transcriptional functions of TAR in virus replication. We here present the development of an efficiently replicating HIV-rtTA variant that completely lacks the TAR-hairpin. Our results demonstrate that TAR is not essential for processes other than transcription during HIV-1 replication.

**MATERIALS AND METHODS**

**Construction of HIV-rtTA variants**

Construction of the HIV-rtTA molecular clone was described previously (15, 17). The HIV-rtTA variant used in this study (HIV-rtTA\textsubscript{F86Y A209T}\textsubscript{2ΔtetO}) contains the 2ΔtetO promoter configuration in both the 5’ and 3’ LTR (19) and the optimized rtTA\textsubscript{F86Y A209T} gene (20). Deletions in TAR were introduced in both the 5’ and 3’ LTR of the HIV-rtTA plasmid in three steps as described below. Briefly, we first introduced the deletions by PCR mutagenesis in the 3’ LTR of a shuttle vector encompassing the 3’ half of the HIV-rtTA genome, subsequently introduced the mutations in the 5’ LTR of HIV-rtTA, and finally combined the 5’ and 3’ LTR mutated fragments. For the construction of the A variant, the 3’ LTR sequence was amplified with primers tTA-tetO-1 (ctccccgggtaactaagtaaggat; sense primer annealing at 3’ end of rtTA gene) and TAR-A (cagagagctc-Δ-atgctccagagacccagtacaggc; SacI site underlined; Δ indicating position of deletion), and with plasmid pBlue3’LTRext-ΔU3-rtTA\textsubscript{F86Y A209T-2ΔtetO}, which includes Env, rtTA and 3’ LTR sequences of the HIV-rtTA genome (20), as template. The PCR product was digested with BspE1 and SacI, and ligated into the corresponding sites of the shuttle vector pBlue3’LTRext-ΔU3-rtTA\textsubscript{F86Y A209T-2ΔtetO-mPL}, which is a derivative of pBlue3’LTRext-ΔU3-rtTA\textsubscript{F86Y A209T-2ΔtetO} in which the SacI site has been removed from the vector sequence by digestion with BssHII and BamHI, blunting of the sticky ends with Klenow polymerase and dNTPs, and subsequent religation. The C variant was constructed in the same way, but PCR was performed with primers tTA-tetO-1 and TAR-C (cagagagctccaatgctcctttctgg-Δ-cccagtacaggcaaaaagcag). Similarly, for the construction of the B variant, PCR was
performed with primers TAR-B (attggagctc-∆-tagggaacccactgcttaagcc) and pLAI-3’seq (tgctctagaggggtagtacata; antisense primer annealing to vector sequences downstream of 3’ LTR), the PCR product was digested with SacI and AatII, and subsequently ligated into the corresponding sites of pBlue3’LTRext-ΔU3-rtTA\textsubscript{F86Y A209T} 2ΔtetO-mPL. The D and F variants were constructed in the same way, but with primers TAR-D (attggagctctctgg-∆-cccactgcttaagcctcaata) and TAR-F (attggagctc-∆-cccactgcttaagcctcaata), respectively. For the construction of the E variant, we used primers tTA-tetO-1 and TAR-E (agggcaagcttattgggtaagcctgag-∆-cccagctacgg caaaaagca; HindIII site underlined), digested the PCR product with BspEI and HindIII, and subsequently ligated this fragment into the corresponding sites of pBlue3’LTRext-ΔU3-rtTA\textsubscript{F86Y A209T} 2ΔtetO-mPL. For the construction of the double mutant AB, we combined the A and B deletions. The A variant of pBlue3’LTRext-ΔU3-rtTA\textsubscript{F86Y A209T} 2ΔtetO-mPL was digested with BamHI and SacI, and the 1463-bp env-rtTA-LTR fragment was used to replace the corresponding sequences in the B variant of this shuttle vector. Similarly, to combine the C and D deletions in the CD double mutant, the BamHI-SacI env-rtTA-LTR fragment of the C variant was used to replace the corresponding sequences in the D variant. For the construction of the double mutant AB, we combined the A and B deletions. The A variant of pBlue3’LTRext-ΔU3-rtTA\textsubscript{F86Y A209T} 2ΔtetO-mPL was digested with BamHI and SacI, and the 1463-bp env-rtTA-LTR fragment was used to replace the corresponding sequences in the B variant of this shuttle vector. Similarly, to combine the C and D deletions in the CD double mutant, the BamHI-SacI env-rtTA-LTR fragment of the C variant was used to replace the corresponding sequences in the D variant. For the introduction of the TAR-mutations into the 5’ LTR of HIV-rtTA, we PCR-amplified the LTR region from the pBlue3’LTRext-ΔU3-rtTA\textsubscript{F86Y A209T} 2ΔtetO-mPL variants A-F with primers U3-Xba-Not (acgtctagagccccagcactggaagggctattcactc, positions -331 to -313) and U5-Nar (ttcgggcgccactgctagagattttgccacatg, positions +192 to +160), digested the PCR product with NotI and NarI, and used this fragment to replace the corresponding 5’ LTR sequences in HIV-rtTA. To construct 5’ plus 3’ TAR-mutated HIV-rtTA variants, the BamHI-BglI fragments of the pBlue3’LTRext-ΔU3-rtTA\textsubscript{F86Y A209T} 2ΔtetO-mPL variants were used to replace the corresponding Env-rtTA-3’ LTR sequences in the 5’ LTR-mutated HIV-rtTA variants. All mutations were verified by sequence analysis.

**Proviral DNA analysis and cloning of evolved sequences**

Virus infected cells were pelleted by centrifugation at 4,000 rpm for 4 min and washed with phosphate-buffered saline. DNA was solubilized by resuspending the cells in 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 0.5% Tween 20, followed by incubation with 200 µg of proteinase K per ml at 56°C for 30 min and at 95°C for 10 min. Proviral DNA sequences were PCR amplified from total cellular DNA with primers U3-Xba-Not (acgtctagagccccagcactggaagggctattcactc, positions -331 to -313) and U5-Nar (ttcgggcgccactgctagagattttgccacatg, positions +192 to +160), digested the PCR product with NotI and NarI, and used this fragment to replace the corresponding 5’ LTR sequences in HIV-rtTA. To construct 5’ plus 3’ TAR-mutated HIV-rtTA variants, the BamHI-BglI fragments of the pBlue3’LTRext-ΔU3-rtTA\textsubscript{F86Y A209T} 2ΔtetO-mPL variants were used to replace the corresponding Env-rtTA-3’ LTR sequences in the 5’ LTR-mutated HIV-rtTA variants. All mutations were verified by sequence analysis.
pBlue3’LTRext-ΔU3-rtTA<sub>F86Y A209T</sub>-2ΔtetO-mPL. The BamHI-BglI fragment of these 3’ LTR-mutated plasmids were used to replace the corresponding sequences in the 5’ LTR-mutated HIV-rtTA variants, which resulted in HIV-rtTA variants with ER1-3 sequences in both the 5’ and 3’ LTR.

**rtTA activity assay**

In the plasmid pLTR-2ΔtetO-X/H-luc<sup>ff</sup> the expression of firefly luciferase is under the control of the LTR-2ΔtetO-promoter of HIV-rtTA. For the construction of this plasmid, we isolated the XbaI-HindIII fragment encompassing the LTR-2ΔtetO promoter sequences (-333 to +82) of HIV-rtTA<sub>F86Y A209T</sub>-2ΔtetO, and ligated this fragment into the compatible NheI and HindIII sites of the luciferase-reporter construct pGL3-Basic (Promega). For cloning purposes, the SalI site in pGL3-basic had been removed by digestion with SalI, blunting of the sticky ends with Klenow polymerase and dNTPs, and subsequent religation.

For the introduction of the TAR-mutations into pLTR-2ΔtetO-X/H-luc<sup>ff</sup>, we PCR amplified the LTR region from the pBlue3’LTRext-ΔU3-rtTA<sub>F86Y A209T</sub>-2ΔtetO-mPL variants A-F with primers U3-Xba-Not and U5-Nar, digested the PCR product with SalI and HindIII, and used this fragment to replace the corresponding sequences in pLTR-2ΔtetO-X/H-luc<sup>ff</sup>. Plasmid pBlue3’LTR-luc<sup>ff</sup> contains the complete U3 region and R sequences up to position +82 of the wild-type HIV-1 LAI proviral DNA coupled to the firefly luciferase reporter gene (21). The plasmid pRL-CMV (Promega), in which the expression of renilla luciferase is controlled by a CMV promoter, is co-transfected into the C33A cells to allow correction for differences in transfection efficiency.

C33A cells were cultured in 2-cm<sup>2</sup> wells and transfected with 20 ng pLTR-2ΔtetO-luc<sup>ff</sup> construct (TAR<sup>wt</sup> and mutants A-F) or 20 ng pBlue3’LTR-luc<sup>ff</sup> (TAR<sup>mut</sup>), 0.4 ng rtTA-expression plasmid pCMV-rtTA<sub>F86Y A209T</sub> (20) and 0.5 ng pRL-CMV. pBluescript was added to the transfection mix to have a total of 1 µg of DNA. The cells were cultured after transfection for 48 hours with 0-1000 ng/ml dox (Sigma D-9891). Cells were lysed in Passive Lysis Buffer and firefly and renilla luciferase activities were determined with the Dual-Luciferase assay (Promega). The expression of firefly and renilla luciferase was within the linear range and no squelching effects were observed. The promoter activity was calculated as the ratio between the firefly and renilla luciferase activities, and corrected for between session variation (22).

**Cells and viruses**

SupT1 T cells were cultured and transfected by electroporation, as previously described (20). To assay virus replication, 5 x 10<sup>6</sup> SupT1 cells were transfected with 1 µg of the proviral constructs and cultured in 5 ml medium with 1 µg/ml dox. For the selection of viruses with improved replication capacity, virus-cell cultures were continued for up to 168 days. When virus-induced cytopathic effects were observed, high level virus replication was maintained by passage of the cell-free culture supernatant onto uninfected SupT1 cells. Cell and supernatant samples were stored at -80°C for subsequent analysis. C33A cervix carcinoma cells (ATCC HTB31) were
cultured in 2-cm² wells and transfected with 1 µg HIV-rtTA construct by calcium phosphate precipitation, as previously described (20). Virus production was measured by CA-p24 ELISA on culture medium samples (23).

RNA analysis
For the isolation of viral transcripts, C33A cells were cultured in 10-cm² wells and transfected with 5 µg HIV-rtTA construct. After 48 hours, cells were washed with PBS, shortly incubated with 0.5 ml 0.05% trypsin-EDTA (Invitrogen) till cells detached, resuspended in 1 ml 10% FBS-containing medium to inactivate trypsin, and subsequently centrifuged at 2750 x g for 5 min. Cells were washed in 1 ml PBS, centrifuged at 2750 x g for 5 min, lysed in 0.6 ml RLT buffer (Qiagen), and homogenized with a QIAshredder column (Qiagen). Total cellular RNA was isolated with the RNAeasy kit (Qiagen) and contaminating DNA was removed with RNase-free DNase during isolation (Qiagen). When indicated, RNA was de-capped with tobacco acid pyrophosphatase (Epicentre Biotechnologies, BIOzymTC, Landgraaf, The Netherlands).

We used the 5’ RACE (rapid amplification of cDNA ends) system Version 2.0 (Invitrogen) to analyze the 5’ end of the RNA transcripts. Briefly, the primer AD-GAG was annealed to the viral RNA at 85°C for 2 min and 70°C for 10 min. The RNA was reverse transcribed with SuperScript II reverse transcriptase (RT; Invitrogen) at 50°C for 50 min. After inactivation of RT at 70°C for 15 min, RNA was degraded with RNase H and RNase T1. The cDNA product was purified with the QIAquick PCR purification kit (Qiagen) and dA-tailed with terminal deoxynucleotidyl transferase (TdT) and dATP. After inactivation of TdT at 65°C for 10 min, the dA-tailed cDNA was amplified by PCR with primers AD-SD (catgatccagtcgcctcccctcgcctc, positions +290 to +270) and 3’RACE Abridged Primer (ggccacgcgtcgactagtac[t]17), and ligated into the pCR2.1 TA-cloning vector. Cloned cDNA fragments were sequenced with the primer BB3.

RESULTS
Deletions in TAR do not affect gene expression
HIV-rtTA carries multiple nucleotide substitutions in TAR (TARm in Fig. 1B) that completely abolish trans-activation of the viral promoter by Tat. This virus no longer requires the Tat-TAR axis for the activation of gene expression, but the TAR stem-loop structure may still have other roles in viral replication. Previous attempts to identify such secondary functions of TAR were complicated by the indispensability of the wild-type TAR structure for the activation of transcription. We therefore set out to determine if the TAR structure in the HIV-rtTA variant is required for viral replication, and introduced more rigorous mutations in the hairpin motif. Partial deletion of the left-hand side (mutants A and C) or the right-hand side (mutants B, D and F) of the
TAR stem severely reduces the stability of this hairpin (Fig. 1B). Combination of these deletions in the double mutants AB and CD results in truncated stem-loop structures. In the most rigorous mutant E, we deleted nearly the complete TAR structure except for the bottom two base pairs, which were left in place to preserve the G residues at the transcription initiation site.

To determine the effect of these TAR deletions on gene expression, we made plasmid reporter constructs in which the LTR-tetO-promoter of HIV-rtTA, carrying the original (TAR\textsuperscript{m}) or mutated TAR sequence (A-F), was coupled to the luciferase reporter gene. We assayed dox-responsiveness of these constructs upon co-transfection with an rtTA-expression plasmid into C33A cervix carcinoma cells (Fig. 2A). The TAR\textsuperscript{m} construct demonstrates low basal luciferase expression in the absence of dox, and this activity gradually increases with increased dox concentration. Similar dox-dependent expression levels are observed for the mutants A-F, demonstrating that the TAR deletions do not affect gene expression in this promoter context. A control construct with the wild-type HIV-1 LTR promoter (TAR\textsuperscript{wt}) obviously does not respond to dox. We also assayed Tat-responsiveness of these promoter-luciferase constructs upon co-transfection with a Tat-expression plasmid. As anticipated, TAR\textsuperscript{m} and all TAR-deletion constructs (A-F) are not activated by Tat, and only the control TAR\textsuperscript{wt} promoter construct is greatly responsive to Tat (data not shown).

**Replication of TAR-deleted variants**

To reveal additional roles of TAR in viral replication, we introduced the TAR deletions A-F into the HIV-rtTA genome. The TAR sequence is part of the 97-nt repeat (R) region that is present at both ends of the viral RNA genome and plays an important role in the first strand-transfer during reverse transcription (24). Since sequence differences between the 5’ and 3’ R regions may hamper the reverse transcription process, we introduced the TAR mutations at both ends of the proviral genome. We transfected these HIV-rtTA plasmids into C33A cells, which support HIV-1 gene expression and virion production, but not HIV-1 infection due to a lack of CD4 receptor. Cells were cultured for 48 hours with dox, and we subsequently measured CA-p24 production in the culture supernatant. The TAR\textsuperscript{m} and TAR-deletion variants all produced a high CA-p24 level (data not shown), which confirms that viral gene expression is not significantly affected by partial or complete deletion of TAR.

We transfected the HIV-rtTA plasmids into the HIV-1 susceptible SupT1 T-cell line to determine the effect of the TAR deletions on viral replication. These cells support efficient replication of the original HIV-rtTA in the presence of dox (Fig. 2B, TAR\textsuperscript{m}). While replication of the variants with a deletion in either the left or right hand side of the TAR stem (A, B, C, D and F) was below the detection level, the double mutants AB and CD replicated as efficiently as the original HIV-rtTA, and the TAR-deleted E mutant showed delayed replication kinetics. These results demonstrate that whereas destabilization of TAR abolishes viral replication, truncation or complete removal of TAR does not, or only moderately, affect viral replication.
Chapter 2

Figure 2. Effect of TAR mutations on gene expression and virus replication. (A) The LTR-tetO promoter region of HIV-rtTA with the original (TARm) or mutant TAR sequence (A-F) was placed upstream of the firefly luciferase gene. To determine dox-responsiveness, C33A cells transfected with these plasmid reporter constructs and an rtTA-expression plasmid were cultured at different dox levels (0-1000 ng/ml). A plasmid constitutively expressing renilla luciferase was co-transfected to correct for differences in transfection efficiency, and the ratio of the firefly and renilla luciferase activities measured two days after transfection reflects the promoter activity. A construct with the wild-type HIV-1 LTR promoter (TARwt) was included as a control. Average values of three transfections are shown with the error bars indicating the standard deviation. (B) SupT1 T cells were transfected with the original HIV-rtTA (TARm) and TAR-mutated variants (A-F), and cultured with 1 µg/ml dox for several weeks. CA-p24 in the culture supernatant was measured by ELISA. No virus replication was observed in the absence of dox (not shown). This experiment has been repeated three times with similar results.

Improved replication of TAR-deleted HIV-rtTA upon viral evolution

Although the complete removal of the TAR structure does not abolish replication, the E mutant replicates much slower than the original HIV-rtTA (TARm), and we anticipated that this mutant could evolve to a better replicating variant when cultured for a prolonged period. We therefore started three long-term cultures of the E mutant and passaged the virus onto fresh cells at the peak of infection when massive syncytia were observed. We noticed that the time interval between infection and the appearance of syncytia became shorter upon prolonged culturing, suggesting that the viral replication capacity was improved. Sequence analysis of the proviral genome present in these long-term cultures revealed that the E mutant had acquired multiple nucleotide substitutions, deletions and insertions at the U3-R boundary in cultures I and II, while part of the R sequence had been duplicated in culture III (Fig. 3A).

To demonstrate that these mutations improve viral replication, the sequences that were most abundant in culture I at day 104 and 168 were re-cloned into the HIV-rtTA proviral genome (variants ER1 and ER2, respectively). Similarly, we re-cloned the duplicated R sequence observed in culture III at day 97 (variant ER3). Upon transfection of these plasmids into C33A cells, a similarly high CA-p24 production was observed with the original construct (TARm), the TAR-deleted E mutant and the evolved ER1, ER2 and ER3 variants (data not shown). Replication of these HIV-rtTA variants was assayed in the SupT1 T-cell line. The ER1, ER2 and ER3 variants replicated much more efficiently than the E mutant, and almost as efficient as the original HIV-rtTA (TARm)(Fig. 3B). These results demonstrate that the mutations selected during evolution at the U3-R boundary of the E mutant significantly improve replication of this virus.
Figure 3. Evolutionary repair of a hairpin structure at the 5’ end of the viral RNA. (A) The TAR-deleted E virus was cultured for up to 168 (cultures I and II) or 97 days (culture III). Cellular proviral DNA was isolated at 104 and 168 days (culture I), at 99 and 168 days (culture II) or at 97 days (culture III), and the LTR region was subsequently PCR amplified and cloned into the TA-cloning vector. The nucleotide sequence of the TAR region was determined for 3-7 clones for each sample. The -10 to +78 region (with +1 indicating the transcription initiation site) is shown for the original HIV-rtTA (TAR\textsuperscript{m}), the E mutant and the evolved viruses (with the frequency at which each sequence is observed indicated on the left; #). Nucleotide substitutions, insertions and deletions (Δ) are boxed in grey. Arrows indicate the duplicated R sequence observed in culture III. At the right, ER1, ER2 and ER3 indicate the evolved sequences re-cloned into the HIV-rtTA virus. (B) To assay replication of the original HIV-rtTA (TAR\textsuperscript{m}), the E mutant and the evolved variants (ER1, ER2 and ER3), SupT1 T cells were transfected with the proviral constructs and cultured with 1 µg/ml dox. This experiment has been re-
Evolutionary repair of a hairpin structure at the 5’ end of the viral RNA

Transcription of the proviral genome starts at the U3-R boundary in the 5’ LTR promoter. Both in wild-type HIV-1 and in HIV-rtTA, the 5’ end of the RNA transcript will fold the TAR and polyA hairpin structures (25), with the 5’ terminal nucleotides being included in the base-paired TAR stem (Fig. 3C). It has previously been shown that the sequence GG\(^+1\)GTCT (with +1 indicating the major transcription start site) is an important element for initiation of HIV-1 transcription (26). Deletion of TAR created the GGGCCC sequence at this position in the E variant, which may have affected the transcription process. The multiple mutations selected near the U3-R border do indeed suggest an evolutionary adaptation of the transcription start site. We therefore set out to identify the transcription initiation site for the E mutant and the evolved ER1, ER2 and ER3 variants.

We transfected the proviral clones into C33A cells and analyzed the 5’ terminal sequence of the RNA transcripts by 5’ RACE (rapid amplification of cDNA ends). As a control, we identified the transcription initiation site for the original HIV-rtTA clone (TAR\(^m\)), which corresponds with the wild-type G\(^+1\) start (results not shown). Transcripts of the E mutant were also found to initiate at this G\(^+1\) residue or the adjacent G\(^+2\) residue (Fig. 3D). The 5’ end of the E transcript is predicted to fold the polyA hairpin structure, but the 5’ terminal nucleotides (GGCC or GCC) will remain single stranded (Fig. 3C). Transcription initiation of the ER1 variant was more diffuse. This variant used the original start site (ER1a) and the position three nucleotides further downstream (ER1b), corresponding to nucleotide +59 in the original HIV-rtTA, with a similar efficiency (Fig. 3D). The ER2 variant predominantly initiated transcription at this A\(^+59\) position. Whereas the ER1a transcript is predicted to fold the polyA hairpin with two single stranded nucleotides at the 5’ end, the ER1b/ER2 transcript can fold the nearly complete polyA hairpin but effectively removed the 5’ dangling end (Fig. 3C).

The ER3 variant did not change the transcription initiation site (Fig. 3D). Nevertheless, this variant was able to remove the 5’ dangling end by alternative means, as the sequence insert triggers the formation of a novel mini-hairpin immediately upstream of the polyA hairpin (Fig. 3C). These results demonstrate that the nucleotide substitutions, deletions and insertions at the U3-R boundary observed during evolution of the E mutant either cause a reallocation of the transcription initiation site (culture I, ER1 and ER2 variants) or introduce a new hairpin structure.

(C) Secondary structure of the 5’ end of the viral RNA transcripts. (D) Determination of the transcription initiation site. C33A cells were transfected with HIV-rtTA proviral clones carrying the mutant E or the evolved ER1, ER2 or ER3 sequence. After 2 days of culturing with 1 µg/ml dox, intracellular RNA was isolated and de-capped, and the 5’-terminal sequence of the viral RNA transcripts was analyzed by 5’ RACE. The cDNA fragments were cloned in the TA-cloning vector and 5-11 clones were sequenced for each sample. The transcription start site (>) observed for each clone is shown on the corresponding proviral U3-R sequence (*: Transcription of the E mutant may have started one nucleotide downstream of the indicated position because the corresponding RNA sample had not been de-capped prior to reverse transcription, and RT may have copied the cap-G nucleotide into cDNA (36)).
CHAPTER 2

at the 5’ end of the RNA transcripts (culture III, ER3 variant). Both evolutionary routes effectively result in inclusion of the 5’ terminal nucleotides of the transcripts in a base-paired stem structure.

These results demonstrate that the nucleotide substitutions, deletions and insertions at the U3-R boundary observed during evolution of the E mutant either cause a reallocation of the transcription initiation site (culture I, ER1 and ER2 variants) or introduce a new hairpin structure at the 5’ end of the RNA transcripts (culture III, ER3 variant). Both evolutionary routes effectively result in inclusion of the 5’ terminal nucleotides of the transcripts in a base-paired stem structure.

DISCUSSION

The TAR hairpin, which is present at the 5’ and 3’ ends of the HIV-1 genomic and messenger RNAs, has been extensively studied in the past decades. The best-studied function of TAR is its essential role in the activation of transcription from the 5’ LTR promoter. Multiple studies in which replication of TAR-mutated HIV-1 variants in T-cell lines was analyzed, suggested several additional functions of TAR in other processes during viral replication, including translation, dimerization, packaging and reverse transcription. However, most of these studies are hampered by the fact that mutation of TAR significantly reduced Tat-mediated activation of transcription, which made it difficult to distinguish effects on other replication processes. In this study, we demonstrate that complete deletion of TAR does not abolish replication of an HIV-1 variant that does not require TAR for the activation of transcription. This result demonstrates that TAR has no essential function in HIV-1 replication other than to accommodate Tat-mediated activation of transcription. However, our studies focused on replication in T-cells and we cannot exclude that TAR may have an accessory function under specific conditions or in specific cell types in vivo.

Our results suggest that efficient HIV-1 replication requires a stable stem-loop structure at the start of the viral transcripts, in which the 5’ terminal nucleotides are base-paired. This structure can either be the wild-type or truncated TAR hairpin (AB and CD mutant), the polyA hairpin (ER1 and ER2 variant) or a new stem-loop structure (ER3 variant), indicating that the nucleotide sequence of this 5’ hairpin is not important. These results are in agreement with a previous forced evolution study in which HIV-1 replication was significantly reduced by opening of the lower TAR stem, and a strong evolutionary pressure restored the base-pairing of this TAR region (27). Furthermore, structure analysis of multiple HIV-1, HIV-2 and SIV isolates revealed that the 5’ end of the RNA is always base-paired, despite variation in overall leader and TAR structures (28).

Whereas truncation of TAR in the double-mutated variants (AB, CD) had no effect on virus replication, destabilization of this stem-loop structure in the single-mutated variants (A, B, C, D, F) blocked replication. The untranslated RNA leader region can fold either an extended duplex through long-distance base pairing (long distance
interaction; LDI) or a branched conformation in which the RNA locally folds into hairpin structures (branched multiple hairpin; BMH) (29). Although both conformations have the TAR hairpin, the unpaired nucleotides in the destabilized structure may interact with other regions of the leader and alter the LDI-BMH equilibrium, and thus indirectly affect viral replication (29, 30). Similarly, the unpaired nucleotides present at the 5’ end of the transcripts of the E mutant may affect the leader structure and this may explain the reduced replication capacity of this variant. Alternatively, the presence of a stable stem-loop structure at the 5’ end of the transcripts may be important for RNA longevity, as previously described for bacterial and organelle mRNAs (31-34).

The TAR DNA sequence in the proviral 5’ LTR promoter region has been shown to bind various cellular transcription factors (reviewed in (35)). We observed efficient replication of the evolved E variants in which TAR was replaced by non-related sequences or completely removed. Thus, although the binding of transcription factors to TAR DNA may be important for Tat-controlled transcription, these interactions are apparently not essential for virus replication in T cells when transcription is controlled by rtTA.

During reverse transcription the sequence complementarity between the 5’ and 3’ R regions facilitates the first strand transfer in which a cDNA copy of the 5’ R-U5 region (strong-stop minus-strand DNA) is translocated to the 3’ end of the viral RNA genome. Our results demonstrate that for this function the R region can be significantly shorter than the 97-nt wild type element. The shortest R region, only 39 nt, is observed in the efficiently replicating ER2 variant. These results are in agreement with a previous study that demonstrated that strand transfer can be efficient with a minimal R-overlap region of approximately 30 nt (24).

The dox-dependent HIV-rtTA variant was proposed as a novel strategy toward a safe live attenuated HIV vaccine. In this study, we used HIV-rtTA to clarify the role of TAR in virus replication, which demonstrates that this variant can also be a powerful tool to study HIV-1 biology.

**ACKNOWLEDGMENTS**

We thank Stephan Heynen for CA-p24 ELISA. This research was sponsored by the Dutch AIDS Foundation (Aids Fonds grant 2005022), Technology Foundation STW (applied science division of NWO and the technology program of the Ministry of Economic Affairs, Utrecht, The Netherlands), ZON-Medical Sciences (MW; VICI grant) and NWO-Chemical Sciences (CW; TOP grant).
REFERENCES

CHAPTER THREE

Destabilization of the TAR hairpin affects the structure and function of the HIV-1 leader RNA

Martine M. Vrolijk, Marcel Ooms, Alex Harwig, Atze T. Das and Ben Berkhout

NUCLEIC ACIDS RESEARCH (2008)
**ABSTRACT**

The TAR hairpin of the HIV-1 RNA genome is essential for virus replication. TAR forms the binding site for the transcriptional trans-activator protein Tat and multiple additional TAR functions have been proposed. We previously constructed an HIV-1 variant in which the TAR-Tat transcription control mechanism is replaced by the components of the Tet-ON regulatory system. In this context, the surprising finding was that TAR can be truncated or even deleted, but partial TAR deletions that destabilize the stem structure cause a severe replication defect. In this study, we demonstrate that the HIV-1 RNA genome requires a stable hairpin at its 5’ end because unpaired TAR sequences affect the proper folding of the untranslated leader RNA. Consequently, multiple leader-encoded functions are affected by partial TAR-deletions. Upon evolution of such mutant viruses, the replication capacity was repaired through the acquisition of additional TAR mutations that restore the local RNA folding, thus preventing the detrimental effect on the leader conformation.

**INTRODUCTION**

Human immunodeficiency virus type 1 (HIV-1) is a retrovirus with an RNA genome of approximately 9 kb that contains 9 open reading frames and untranslated regions at the 5’ and 3’ end (Fig. 1A). The untranslated leader region at the 5’ end encodes several regulatory RNA motifs involved in both early and late replication steps (reverse transcription, transcription, splicing, polyadenylation, translation, RNA dimerization and packaging) (1). This leader RNA can adopt two mutually exclusive structures *in vitro* (Fig. 1B). The energetically favored structure is formed by a long-distance base pairing interaction (LDI) between the polyA and dimerization initiation site (DIS) (2). Both the polyA and DIS motifs fold a hairpin in the alternative structure that is termed the branched multiple hairpin (BMH) conformation. Exposure of the DIS hairpin is important for the formation of RNA dimers by kissing-loop base pairing (Fig. 1B) (3, 4). The kissing-loop complex can rearrange to form an extended dimer that is more stable (5, 6).

In both the LDI and BMH conformation the TAR hairpin is located at the extreme 5’ end of the HIV-1 transcripts. This hairpin has an essential role in the activation of transcription from the promoter in the 5’ long terminal repeat (LTR) of the proviral DNA genome (7, 8). The TAR hairpin has a highly conserved 3-nucleotide pyrimidine bulge that binds the viral Tat transactivator protein (9), and an apical 6-nucleotide loop that binds the cellular Cyclin T1 protein in a Tat-dependent manner (10, 11). The Tat-TAR-Cyclin T1 axis results in enhanced transcription through phosphorylation of the RNA polymerase. In addition to its role in transcription, the TAR hairpin has also been suggested to be involved in dimerization of the viral genome (12), packaging of the viral RNA into virions (13-16) and in the strand transfer step of reverse transcription (17). Mutation of TAR causes severe replication
Figure 1. The HIV-1 TAR hairpin as part of the untranslated leader RNA. (A) The HIV-rtTA provirus genome is shown, with the Tat-TAR axis of transcription regulation inactivated by mutation of both Tat and TAR (tat\textsuperscript{m} and TAR\textsuperscript{m}; crossed boxes). The LTR region is subdivided into the U3, R and U5 region. Transcription and replication of the virus was made dox-dependent by the introduction of tetO elements in the U3 promoter region, and replacing the Nef gene by the rtTA gene. The domain structure of the untranslated leader RNA is shown. The untranslated leader RNA of HIV-1 (+1/+368) consists of several regulatory domains (TAR; polyA: polyadenylation hairpin; PAS: primer activation signal; PBS: primer-binding site; DIS: dimerization initiation site; SD: splice donor; \(\Psi\): RNA packaging signal; AUG: translation start codon of gag). (B) The leader RNA can adopt two conformations. The polyA and DIS sequences are base paired in the LDI structure. The same sequences form the polyA and DIS hairpin in the BMH structure. The DIS hairpin allows the formation of RNA dimers by kissing-loop base pairing. The riboswitch model for regulated dimerization argues that the ground-state LDI structure must first rearrange into the BMH conformation to expose the DIS hairpin, which mediates subsequent RNA dimerization.

defects due to a lack of LTR transcription (18-20). This dominant effect on transcription precluded the careful analysis of the putative non-transcriptional functions of TAR in HIV-1 replication.

We previously constructed an HIV-1 variant that does not depend on the Tat-TAR interaction for activation of transcription (21-23). In this HIV-rtTA variant, both the Tat protein and its TAR binding site were inactivated by mutation (Fig. 1A) and functionally replaced by the components of the Tet-ON gene regulation system (24).
The gene encoding the rtTA transcriptional activator protein was inserted in place of the 3'-terminal nef gene, and tet operator (tetO) binding sites were introduced in the LTR promoter. Administration of the effector doxycycline (dox) induces a conformational switch in the rtTA protein, which subsequently can bind to the tetO-LTR promoter and activate transcription of the proviral genome. Thus, transcription and replication of HIV-rtTA are critically dependent on the presence of dox.

HIV-rtTA has a mutated TAR hairpin (TAR\textsuperscript{m}) with five nucleotide substitutions in the bulge and loop domains to prevent binding of Tat and Cyclin T1 (Fig. 2). This virus does not require TAR for activation of gene expression and thus allows us to test for additional TAR functions in viral replication by deleting parts of TAR. We introduced deletions on either the left or the right side of the TAR\textsuperscript{m} hairpin (Fig. 2: mutants A and C, and mutants B, D and F, respectively). We also made the double mutants (AB and CD) and a nearly complete TAR deletion (mutant E) (Fig. 2). As previously shown, the mutations did not affect doxycycline-dependent transcription. However, whereas the double mutants did not affect viral replication, the single mutants showed a severe replication defect (25).

It was previously observed that mutations in the leader RNA can affect the equilibrium between the dimerization-incompetent LDI structure and the dimerization-prone BMH structure (2, 15, 26, 27). Although the TAR hairpin itself is not directly involved in the switch from LDI to BMH, as it is present in both structures
(Fig. 1B), destabilization of TAR may affect the LDI-BMH equilibrium and thus indirectly influence the function of other RNA motifs. To understand why the single TAR deletions abolish viral replication, we analyzed the effect of these mutations on the leader RNA conformation.

**MATERIAL AND METHODS**

**Mutant and evolved HIV-1 sequences**

All TAR mutations were introduced in the infectious HIV-rtTA (25) molecular clone, which is based on the HIV-1 LAI plasmid (28). To start virus cultures, the SupT1 cell line was transfected by electroporation as described previously (29). In brief, 5 x 10^6 SupT1 cells were transfected with 10 µg of the proviral constructs and cultured in 5 ml medium with 1 µg/ml doxycycline (Sigma D-9891). For the selection of virus variants with improved replication capacity, virus-cell cultures were continued for up to 175 days (25). Cell samples were stored at -80°C for PCR-amplification of HIV-1 genome segments and subsequent sequence analysis. Proviral DNA sequences were PCR amplified from total cellular DNA with primers U3-Xba-Not (5’-ACG TCT AGA GCG GCC GCA CTG GAA GGG CTA ATT CAC TC-3’, position -331 to -313) and AD-GAG (5’-ATG GAT CCG TTC TAG CTC CCT GCT TGC CC-3’, position +463 to +442), ligated into pCR2.1-TOPO TA-cloning vector (Invitrogen), and sequenced with primer BB3 (5’-GAG TCC TGC GTC GAG AGA GCT CCT CTG GTT-3’, position +245 to +216). We used the 5’ RACE (rapid amplification of cDNA ends) system Version 2.0 (Invitrogen) to analyze the 5’ end of the RNA transcripts, as described previously (25).

**In vitro transcription of leader RNA**

RNAs were in vitro transcribed from DNA templates in which a T7 promoter was positioned directly upstream of the transcription start site as previously described (27). For the production of these DNA templates, we PCR-amplified the HIV-1 LAI (TARwt) and HIV-rtTA plasmids (TARM, mutant and evolved variants) with a forward primer encoding the T7 promoter sequence and the 5’ end of the transcript, and a reverse primer R368 (5’-TCC CCC GCT TAA TAC TGA CGC T-3’) that is complementary to nucleotides +347 to +368 in the wild-type leader sequence. We used forward primer T7-2 (5’-CTA ATA CTA CGA CTC AGT ATA GGG TCT CTC TGG AGC ATT GGA-3’) for the production of TARwt, TARM, B, D and F transcripts, T7-A (5’-CTA ATA CGA CTC AGT ATA GGG TCT CTC TGG AGC ATT GGA-3’) for transcripts A, AB and AR2, T7-C (5’-CTA ATA CGA CTC AGT ATA GGG CCA GAA AGG AGC ATT GGA-3’) for the transcripts C and CD, T7-E (5’-CTA ATA CGA CTC AGT ATA GGG CCC ACT GCT TAA GCC TCA-3’) for transcript E, T7-AR1 (5’-CTA ATA CGA CTC AGT ATA GGG TCT CTC TAG AGC ATT GGA-3’) for transcript AR1, T7-FG3 (5’-CTA ATA CGA CTC AGT ATA GGG GCT CTC TGA CTA GAC CAG-3’) for transcript FG3, T7-FG4 (5’-CTA ATA
CGA CTC AGT ATA GGG GGC TCT CTG ACT AGA CCA-3’) for transcript FG4, T7-FG5 (5’-CTA ATA CGA CTC AGT ATA GGG GGC TCT CTC TCT GAC TAG ACC A-3’) for transcript FG5, T7-FG6 (5’-CTA ATA CGA CTC AGT ATA GGG GGC TCT CTC TGA CTA GAC CA-3’) for transcript FG6, T7-ER1 (5’-cta ata cga ctc ctc gta agc ctc aat-3’) for transcript ER1, T7-ER2 (5’-cta ata cga ctc ctc gta agc ctc aat aaa-3’) for transcript ER2 and T7-ER3 (5’-cta ata cga ctc ctc gta ggg cca ctg ctt aag cct cca ggc-3’) for transcript ER3. In vitro transcription was performed with the Megashortscript T7 transcription kit (Ambion). For the production of 32P-labeled transcripts, 1 µl [α32P] UTP (0.33 Mbq/µl, Amersham) was added to the reaction. Transcripts were excised from a 6% denaturing polyacrylamide gel (visualized by UV-shadowing) and eluted from the gel fragment by overnight incubation in 1x TBE-buffer at room temperature. The RNA was ethanol precipitated and redissolved in water. Quantification of the RNA was done by UV-absorbance measurements or scintillation counting in case of 32P-labeled transcripts. The size and integrity of the transcripts was analyzed on a denaturing sequence gel.

**In vitro RNA dimerization and conformation analysis**

100 ng 32P-labeled RNA was incubated with or without oligonucleotide CN1 at 85°C for 2 min in 10 µl MO-buffer (125 mM KAc, 2.5 mM Mg Acetate, 25 mM HEPES pH 7.0) and slowly cooled from 65°C to room temperature for renaturation and dimerization (30). CN1 is complementary to nucleotides 123-151 of the HIV-1 leader RNA and CN1 annealing effectively prevents folding of the LDI conformation. After incubation, the samples were chilled on ice and 4 µl non-denaturing loading buffer (MO-buffer with 30% glycerol and BFB dye) was added to the samples. Samples were analyzed on a 4% polyacrylamide gel in 0.25x TBE (22.5 mM Tris, 22.5 mM boric acid, 0.625 mM EDTA) for LDI/BMH conformation analysis and 0.25x TBM (22.5 mM Tris, 22.5 mM boric acid, 0.1 mM MgCl2) for dimerization analysis (27, 31). Gels were run at 150V at room temperature, dried and subjected to autoradiography. Quantification of the percentage RNA dimerization was performed on a Phosphor Imager (Molecular Dynamics). The dimerization yield was determined by dividing the amount of dimer by the total amount of RNA (dimer and monomer).

**RNA secondary structure prediction**

Computer-assisted RNA secondary structure prediction was performed using the Mfold v3.2 algorithm (32, 33). Settings were the same for all folding jobs, using standard conditions (37°C and 1.0 M NaCl), a 5% suboptimal range and a maximum distance between paired bases of 200 nucleotides. To analyze the BMH structure of these transcripts, we forced the folding of the DIS hairpin by using the constraint option in the Mfold program. Folding was performed with wild-type, mutant and evolved sequences that started at position +1 and ended at the nucleotide corresponding to position +368 in wild-type HIV-1.
RNA absorbance measurements

UV melting of the T7 transcripts was measured by monitoring the absorbance of UV light at 260 nm on a Varian Cary 300 Bio UV-visible spectrophotometer in a quartz cuvet with a standard 1 cm path length. Prior to the measurement, the 40 ng/µl RNA in 50 mM Na-cacodylate buffer (pH 7.2) was incubated at 85°C for 5 min and then slowly cooled to room temperature for renaturation.

RNA structure probing

The T7 transcribed leader RNAs (nt 1-368) corresponding to the TARm, A, B and AB variants were used for RNA structure probing as described (34). These RNAs (50 pmol) were diluted in 37.5 µl water and mixed with 12.5 µl 4x MO-buffer. The samples were incubated at 85°C for 3 min and slowly cooled to room temperature. The transcripts were incubated with 10 mM lead (Pb2+) nitrate at room temperature. Samples (10 µl) were taken at 0 and 10 min after the addition of lead (Pb2+) nitrate, and the reaction was stopped by adding 2 µl 0.5 M EDTA. RNA products were purified over a NucAway™ spin column (Ambion), ethanol precipitated and dissolved in 10 µl H2O. Oligonucleotides (30 pmol) CN1, R368 and lys21 (5’-CAA GTC CCT GTT CGG GCG CCA -3’; complementary to nucleotides +182 to +202 in the wild-type leader sequence) were 5’ end labeled with the kinaseMax kit (Ambion) in the presence of 3 µl of [γ-32P] ATP (0.37 MBq/µl, Amersham). Two picomoles of 32P-labeled oligonucleotide was annealed to 1 pmol of Pb2+-treated RNA by incubation at 85°C for 1 min, followed by slow cooling. These primers were extended at 60°C for 2h using the Thermoscript reverse transcriptase (Invitrogen). After adding 20 µl gel-loading buffer II (Ambion), the samples were heated to 95°C and 3 µl was analyzed on a denaturing 6% polyacrylamide gel. A sequence ladder was produced with the thermo sequenase cycle sequencing kit (USB) and 32P-labeled oligonucleotide primer and the DNA fragment corresponding to the TARm leader region as template (T7-2/R386 PCR product).

RESULTS

Single-side TAR deletions abolish HIV-rtTA replication

We previously presented variants of the doxycycline-dependent HIV-rtTA with deletions in the TAR hairpin (Fig. 2). Whereas these mutations did not significantly affect doxycycline-dependent transcription, the single mutants (A, B, C, D, F) abolished viral replication (25). In contrast, the double mutants (AB, CD) replicated efficiently and the truncated mutant E showed delayed replication (summarized in Fig. 2). The single deletions are possibly toxic because unpaired TAR sequences are generated, which may interfere with the proper folding of viral RNA. In particular, these mutations may alter the structure of the leader RNA that encodes many
important replication signals. To test this hypothesis, we made a set of leader transcripts and performed several analyses to determine the impact of the TAR mutations on the structure and function of the leader RNA.

Replication defect of TAR mutants corresponds with an altered LDI-BMH equilibrium

We probed the conformational state of the wild-type and mutant leader RNAs by inspection of the migration on a native gel. The TAR\textsuperscript{wt} transcript (corresponding to nucleotides +1 to +368) exhibits the characteristic fast electrophoretic mobility of the LDI conformation, which is converted to a slower migrating BMH structure after annealing of the antisense oligonucleotide CN1 (Fig. 3). This antisense DNA oligonucleotide binds the U5 region of the leader RNA and effectively prevents the LDI interaction between the polyA and DIS motifs, thus forcing the RNA into a BMH-like structure that exposes the DIS hairpin (30). The TAR\textsuperscript{m} transcript, which contains the TAR bulge and loop modifications as present in HIV-rtTA, displays the same gel migration pattern as TAR\textsuperscript{wt} and thus also adopts the LDI conformation. The analysis of the other transcripts is obviously complicated by the fact that the transcripts differ in size due to the TAR deletions. We therefore focus on the migration differences of the transcripts with and without CN1. Some transcripts display a partial shift from LDI to BMH (mutants A and B) or a complete shift (mutants C and E). We marked the transcripts that adopt the slow-migrating BMH conformation with an asterisk (*) in Figure 3. These BMH-structured RNA molecules migrate somewhat faster than the CN1-induced BMH-folded RNAs because they lack the oligonucleotide. While transcripts A and B show a diffuse pattern, which indicates a partial shift from LDI to BMH, only the LDI is formed by the double mutant AB. Likewise, the C mutant adopts the BMH structure, while the double mutant CD folds the LDI conformation. No BMH structure is observed for the mutants D and F. When most of the TAR hairpin is deleted as in mutant E, the BMH conformation is formed. Thus, whereas the majority of the single mutants display the BMH structure, the double mutants fold only the LDI conformation. Apparently, a deletion in one side of the TAR stem can shift the LDI-BMH equilibrium from LDI to BMH.
We next compared the conformation of the TAR\textsuperscript{m}, A, B and AB transcripts by measuring its UV absorbance at increasing temperature. In this assay, an increase in absorbance corresponds to the melting of RNA structure, and we previously reported that this method can accurately distinguish the LDI from the BMH conformation (2, 35). The melting curves in the 30-80°C trajectory are shown as -\(\Delta A/\Delta T\) plot, which allows the determination of the melting transition and the \(T_m\). The TAR\textsuperscript{m} leader transcript (+1/+368) shows two discrete transitions in the melting profile at \(T_m = 52.7°C \) and \(T_m = 71.1°C\) (Fig. 4). Previous studies showed that the transition at \(T_m = 71.1°C\) is due to melting of the TAR hairpin (2). The transition at \(T_m = 52.7°C\) corresponds with denaturation of the extended long distance interaction of the LDI structure. These two transitions were also observed for the TAR\textsuperscript{wt} transcripts (results not shown, see also ref. (2, 30, 31)). Mutant A and in particular mutant B exhibited a grossly different UV-melting pattern (Fig. 4). First, both transcripts lack the prominent TAR signal at \(T_m = 71.1°C\), which is consistent with the partial TAR deletion. The remnant of TAR may still fold a TAR-like structure that explains minor transitions at \(T_m = 72.7°C\) and \(T_m = 66.9°C\) for mutant A and B, respectively. Second, the typical resonance of the LDI structure at \(T_m = 52.7°C\) is partially lost in mutant A and completely lost in mutant B. Other resonances at lower temperature (\(T_m = 39.0°C\) and \(T_m = 37.3°C\) for mutants A and B, respectively) and higher temperature (\(T_m = 60.0°C\) and \(T_m = 58.4°C\), respectively) indicate the BMH folding pattern (2). Strikingly, the UV-melting pattern reverses to a characteristic LDI signature with two peaks for the AB double mutant. The transition at \(T_m = 78.6°C\) is likely to represent the TAR hairpin of this mutant. Compared to the TAR\textsuperscript{m} control, the resonance is much reduced because the hairpin is severely truncated. Interestingly, the \(T_m\) of this truncated TAR hairpin of mutant AB is greater than that of TAR\textsuperscript{m} or TAR\textsuperscript{wt}, which is likely due to removal of destabilizing elements such as the 3-nt and 1-nt bulges.

To better understand the effects of the TAR mutations on leader folding, the structure of the transcripts (+1/+368) was also studied with the Mfold algorithm (32, 33). This RNA folding program predicts alternative structures and their thermodynamic stability (\(\Delta G\) in kilocalories per mole). These values were used to calculate the \(\Delta\Delta G\) (\(\Delta G_{\text{LDI}} - \Delta G_{\text{BMH}}\)), which describes the status of the LDI/BMH equilibrium (Table 1). Both conformations are predicted to be present in equimolar ratio when the \(\Delta\Delta G\) is zero. The \(\Delta\Delta G\) value for TAR\textsuperscript{wt} and TAR\textsuperscript{m} is -3.2 kcal/mole (Table 1), indicating that these transcripts adopt mainly the LDI conformation. The single mutants A, B, C, D and F show a significant shift to \(\Delta\Delta G = -0.1\) kcal/mole, a state in which both LDI and BMH structures are present in approximately equimolar amounts. Most importantly, Mfold predicted a complete reversal to the TAR\textsuperscript{wt}/TAR\textsuperscript{m} state for the double mutant AB and CD with a \(\Delta\Delta G\) of -3.2 kcal/mole, which indicates that these transcripts fold predominantly the LDI conformation. These results are consistent with the gel migration and UV-melting experiments. There is one notable exception, Mfold predicts the LDI conformation for the E mutant, but the \textit{in vitro} experiments indicated a BMH-like conformation. Taken together, the RNA structure and replication results demonstrate that folding of the LDI conformation correlates...
with efficient virus replication, whereas an altered LDI-BMH equilibrium correlates with a replication defect.

**Distorted leader RNA conformation affects dimerization**

The single-side mutations in TAR affect the LDI-BMH equilibrium. Since the DIS motif, which is required for RNA dimerization, is exclusively exposed in the BMH conformation, one may expect an impact of these mutations on the dimerization capacity. We therefore performed RNA dimerization assays with $^{32}$P-labeled transcripts. These transcripts showed the expected differences in size on a denaturing polyacrylamide gel (Fig. 5A). To determine the RNA dimerization
efficiency, the transcripts were incubated in dimerization buffer and analyzed on a native polyacrylamide gel in the presence of Mg$^{2+}$ (Fig. 5B). The monomer and dimer bands were quantified to determine the dimerization efficiency (Fig. 5C). The TAR$^{wt}$ and the TAR$^{m}$ transcripts are predominantly present as monomer under these conditions, which is consistent with LDI folding of these transcripts. The single mutants A and B show increased RNA dimerization, but the AB double mutant returned to the wild-type level. Similarly, the single mutants C and D show increased dimerization efficiency, which is restored to the wild-type level for the CD double mutant. The extended single mutant F also displays increased dimerization. Thus, HIV-1 leader transcripts dimerize more efficiently when only one side of the TAR hairpin is partially deleted. Also when almost the complete hairpin is removed as in mutant E, dimerization is increased significantly. These results are in agreement with the RNA folding experiments, as mutants that adopt the BMH conformation are dimerization-prone (3, 4, 30, 36).

Evolution of TAR-deleted HIV-rtTA variants restores the LDI/BMH equilibrium

Our results thus far demonstrate that a modified leader configuration correlates with reduced replication capacity. To obtain further evidence for this correlation, we tried to select fast-replicating variants for the replication-impaired HIV-rtTA mutants carrying the single TAR deletions. For each variant, we started four long-term cultures by transfecting the SupT1 T cell line with 10 µg of the molecular clones. Only in one of the A cultures and all four F cultures we observed virus replication after
prolonged culturing, which was apparent from an increased CA-p24 level in the culture supernatant and formation of virus induced multi-nuclear syncytia. The virus was passaged on fresh cells when massive syncytia were observed. Cell samples were stored at several points as a source for integrated HIV-rtTA DNA, of which the U3-R-U5 region was PCR-amplified and sequenced (Fig. 6A).

Mutant A acquired several point mutations in the R region (Fig. 6A). At day 52, the viral quasispecies consisted of a mixture of two sequences that we named AR1 and AR2, each with two mutations. At later times, only the AR2 sequence was observed. We performed 5’ RACE analysis to verify that the wild-type transcription start site (+1 in Fig. 6) is used by the A mutant (data not shown). The Mfold program was used to predict the structure and stability of the mutant and evolved TAR structures (Fig. 6B). AR1 acquired 2 mutations that convert weak G-U into more stable A-U base pairs, concomitant with a change in ∆G of the TAR hairpin from -13.0 to -14.2 kcal/mole. AR2 shares one of these mutations, but acquired a unique mutation that changes a G-U into a G-C base pair. The latter mutation reorganizes the top of the stem to form a more stable hairpin of -16.3 kcal/mole. This AR2 variant seems superior over AR1, as it dominates the viral quasispecies at later culture times (days 87 and 169).

The F mutant acquired multiple nucleotide substitutions and insertions of G residues near the transcription start site in all four independent cultures (numbered I to IV in Fig. 6A). Culture II acquired four substitutions and a G-insertion over time, which resulted in a gradual increase in the number of G nucleotides at the transcription start site. Because of these changes around the U3/R border, we used 5’ RACE analysis to determine the actual transcription start site of the virus in this culture at day 169. Transcription of the original F mutant, which was included as a control, started at the same G nucleotide as observed for the TARm and A mutant. The evolved F variant started transcription predominantly at the original +1 nucleotide and the immediate upstream G nucleotide, which resulted in the presence of 5 or 6 G nucleotides at the 5’ end of the transcript (FG5 and FG6 in Fig. 6B, respectively). The structure and stability of the mutant and evolved TAR hairpins was analyzed with the Mfold program. Since we observed an increase in the number of G nucleotides at the transcription start site in this culture, we also analyzed the folding of transcripts with 3 and 4 G nucleotides at the 5’ end as putative evolution intermediates (FG3 and FG4, respectively). The TAR sequence of mutant F is predicted to fold two mini-hairpins (Fig. 6B). The three point mutations in variant FG3 destabilize the upstream mini-hairpin and reduce the overall stability from -14.9 to -10.9 kcal/mole (Fig. 6B). The additional G residues resulting from base-substitution or G-insertion cause a structure switch to an alternative single hairpin of which the stability is progressively increased from -14.4 (FG4) to -16.7 (FG5) and -20.8 (FG6) kcal/mole.

The stability of complete leader transcripts of the evolved variants was also analyzed with Mfold and the ∆∆G of the LDI-BMH equilibrium was calculated (Table
1). Whereas the $\Delta\Delta G$ of mutants A and F is -0.1 kcal/mole, the AR1 and AR2 variants show a $\Delta\Delta G$ of -3.2 kcal/mole, which is in fact an identical value as determined for TAR$^{wt}$, TAR$^{m}$, AB and CD. These $\Delta\Delta G$ calculations thus indicate that LDI folding is restored for the AR1 and AR2 transcripts. However, the evolved F variants show a $\Delta\Delta G$ of -0.1 kcal/mole, which does not indicate such a shift in the LDI-BMH equilibrium.

To test whether the additional mutations observed in the evolved A and F variants restore leader RNA folding, we determined the dimerization capacity of leader transcripts. On a denaturing gel the transcripts show migration differences according to their size (Fig. 6C, top panel). Upon incubation in dimerization buffer, the transcripts were analyzed on a non-denaturing gel (Fig. 6C, lower panel) and RNA dimer and monomer bands were quantitated (Fig. 6D). While mutant A has an increased dimerization capacity due to the LDI-to-BMH shift, the evolved AR1 and AR2 variants return to the low level of dimerization as seen for TAR$^{m}$ and TAR$^{wt}$. The F mutant also showed increased RNA dimerization. Whereas RNA dimerization efficiency was further increased for the putative FG3 and FG4 intermediates, the wild-type level was restored for the FG5 and FG6 variants. Notably, the 5’ RACE analysis revealed that only the FG5 and FG6 transcripts were produced in the virus-infected cells. In fact, the increased dimerization efficiency of the FG3 and FG4 transcripts may explain why these variants were not observed as intermediates during virus evolution. Thus, both the evolved A and F variants show a restored dimerization capacity. These results indicate that evolution of both the A and the F


**CHAPTER 3**

A

<table>
<thead>
<tr>
<th>TAR</th>
<th>A</th>
<th>+1</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>52</td>
<td>(2/5) gcctagtggctctc7g</td>
<td>GACATGGAGCTCTGCTGCTGCACTAGGAAACACTGCTTT</td>
<td>GACATGGAGCTCTGCTGCTGCACTAGGAAACACTGCTTT</td>
<td>AR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>(4/4) gcctagtggctctc7g</td>
<td>GACATGGAGCTCTGCTGCTGCACTAGGAAACACTGCTTT</td>
<td>GACATGGAGCTCTGCTGCTGCACTAGGAAACACTGCTTT</td>
<td>AR2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>169</td>
<td>(4/4) gcctagtggctctc7g</td>
<td>GACATGGAGCTCTGCTGCTGCACTAGGAAACACTGCTTT</td>
<td>GACATGGAGCTCTGCTGCTGCACTAGGAAACACTGCTTT</td>
<td>AR3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B

C

\[ \Delta G = -29.6 \text{ (kcal/mole)} \]

\[ +\text{Mg}^{2+} \]

D

<table>
<thead>
<tr>
<th>m</th>
<th>0</th>
<th>20</th>
<th>40</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAR</td>
<td>A</td>
<td>AR1</td>
<td>AR2</td>
<td>F</td>
</tr>
<tr>
<td>FG3</td>
<td>FG4</td>
<td>FG5</td>
<td>FG6</td>
<td></td>
</tr>
</tbody>
</table>
mutant resulted in a stabilization of the TAR-like hairpin, which restored the leader RNA folding and thereby dimerization control.

We also tested the dimerization capacity of leader transcripts corresponding to evolved variants of the TAR-deleted mutant E that were described recently (25). We observed distinct evolution paths in two independent cultures. In one culture, the virus acquired additional mutations at the transcription initiation site, which resulted in the modified ER1 and ER2 transcripts (Fig. 7A). At an early stage in evolution, both the ER1 and ER2 transcripts were produced, whereas the ER2 variant dominated at a later stage. The mutant E transcript has four unpaired nucleotides upstream of the polyA hairpin and the ER1 variant has only two unpaired nucleotides at this position. In the ER2 transcript these latter nucleotides are deleted as well, and as a result the transcript starts precisely at the first nucleotide of the polyA hairpin. In the other culture a sequence insertion at the transcription start site resulted in the presence of a new hairpin upstream of the polyA hairpin (ER3 in Fig. 7A), which also effectively removes the unpaired nucleotides at the 5’ end of the transcript. We already mentioned that mutant E has a non-wild-type phenotype with a high level of spontaneous RNA dimerization, which is indicative for a BMH-like leader RNA structure. We observed a strong downmodulation of the RNA dimerization capacity of the three evolved variants (Fig. 7B and 7C), which suggests that a wild-type-like LDI structure was restored.

<<<< Figure 6. Evolutionary repair of the hairpin structure at the 5’ end of the viral RNA of mutants A and F restores wild-type in vitro dimerization of leader RNAs. (A) Upon long-term culturing of the A and F mutant viruses, the sequence of the TAR region was analyzed. The viruses in culture A and F II were cultured for up to 169 days. The cultures F I, III and IV were maintained for up to 175, 97 and 91 days, respectively. The LTR region was PCR amplified and cloned into the TA-cloning vector. The nucleotide sequence of the TAR region was determined for 2-5 clones in each culture. The frequency at which each sequence is observed is indicated (#). The -10 to +66 region is shown for the original TARm, the A and F mutants, and the evolved variants of A and F (with +1 indicating the transcription start site). Nucleotide substitutions and insertions are in bold and underlined. (B) Secondary structure of the TAR sequence in the A and F mutant and evolved viruses as predicted by the Mfold program. (C, upper panel) In vitro transcribed leader RNA molecules with either the original TARm, the A and F mutants, and the evolved variants of A and F (with +1 indicating the transcription start site). Nucleotide substitutions and insertions are in bold and underlined. (B) Secondary structure of the TAR sequence in the A and F mutant and evolved viruses as predicted by the Mfold program. (C, upper panel) In vitro transcribed leader RNA molecules with either the original TARm, mutated (A, F) or evolved TAR sequences (AR1-2, FG3-6) were analyzed on a denaturing polyacrylamide gel. (C, lower panel) The transcripts were allowed to dimerize in the presence of Mg2+ and monomeric (m) and dimeric (d) RNAs were resolved on a native polyacrylamide gel containing Mg2+. (D) The monomeric and dimeric bands were quantified and the dimerization efficiency was determined by calculating the fraction RNA present as dimer. The average of 3-5 experiments is shown with error bars indicating the standard deviation.
Figure 7. Re-establishment of a stable hairpin structure at the 5’ end of mutant E restores wild-type in vitro dimerization level of leader RNAs. (A) Secondary structure of the remaining TAR sequences and the polyA hairpin of mutant E and evolved variants. (B, upper panel) In vitro transcribed RNA molecules corresponding to the leader region of HIV-1 rtTA (TARm), the E mutant and the evolved variants (ER1, ER2, ER3) were analyzed on a denaturing polyacrylamide gel. (B, lower panel) The transcripts were allowed to dimerize and analyzed as described in (Fig. 6C). (C) The dimerization efficiency was determined as described in (Fig. 6D). The average of 2-4 experiments is shown with error bars indicating the standard deviation.

Extension of the polyA hairpin triggers the BMH leader conformation.
Although the TAR hairpin is not directly involved in the LDI-BMH conformational switch, as it is present in both conformations, our results demonstrate that the single side TAR deletions result in increased BMH folding and dimerization of the leader RNA. To understand how the TAR deletions affected the LDI-BMH equilibrium, we performed RNA structure probing of the TARm, A, B and AB leader transcripts.
The RNAs were incubated with lead (Pb$^{2+}$) nitrate that mainly cleaves at single-stranded nucleotides. The RNA fragments were subsequently analyzed by primer extension analysis. Probing of the polyA region in the TAR$^m$ transcript (Fig. 8A) revealed high Pb$^{2+}$-induced cleavage at positions G$^{102}$ to U$^{105}$, indicating that these nucleotides are predominantly single-stranded, which is characteristic for the LDI conformation (Fig. 8B). In contrast, these nucleotides are protected in the A and B mutants, which is in agreement with formation of the polyA hairpin and consequently the BMH conformation. Combined with the results of Mfold analyses, an extended version of the polyA hairpin is proposed for the A and B mutants (Fig. 8B). In fact, 3′-terminal TAR nucleotides, in particular the G triplet (positions +51 to +53 of the wild type sequence), interacts with the C nucleotides that are present 3′ of the polyA hairpin (at positions +109 to +111). The Pb$^{2+}$-reactivity of these nucleotides was restored in the AB double mutant, demonstrating that the LDI conformation was re-established. Nucleotide U$^{75}$ in the polyA signal was highly reactive in all transcripts, which makes sense as this nucleotide is single stranded in both conformations. With the A and B mutants, a strong stop product is observed both with and without Pb$^{2+}$ treatment, which probably results from pausing of reverse transcription at C$^{111}$ at the base of the extended polyA hairpin in the BMH conformation (Fig. 8B). This pause product is not observed with the TAR$^m$ and AB transcripts, which is in agreement with the absence of this hairpin in the LDI conformation. Analysis of the upstream TAR region was more complex because of the introduced deletions. The TAR$^m$ and AB transcripts fold a stable hairpin with a highly reactive loop sequence, whereas the A and B mutants fold a destabilized structure in which many more nucleotides are reactive (results not shown). Probing of the DIS region, revealed reduced reactivity of U$^{244}$ and increased reactivity of A$^{255}$ in the A and B mutants compared to the TAR$^m$ and AB transcripts (Fig. 8C). This reactivity pattern is in agreement with the formation of the DIS hairpin and the BMH conformation by the A and B transcripts, and the absence of this hairpin in the LDI conformation of the TAR$^m$ and AB transcripts (Fig. 8D).

**DISCUSSION**

The untranslated leader region, of which TAR is a component, is an evolutionary conserved region of the HIV-1 RNA genome. In previous studies, most of the mutations in the TAR hairpin resulted in a replication incompetent virus (20, 37, 38). Since the TAR hairpin plays a crucial role in Tat-mediated activation of transcription, it was not possible to study other functions of TAR in HIV-1 replication by mutation or (partial) deletion of the TAR sequence. By replacing the TAR-Tat transcription control mechanism with the rtTA-tetO (Tet-ON) system it became possible to study non-transcriptional functions of TAR (23). Our recent study indicates that TAR can be deleted completely when not required for Tat-mediated activation of transcription (25). However, partial TAR deletions resulted in a non-replicating virus, suggesting
Figure 8. RNA structure probing of TAR mutated HIV-1 leader RNA. Transcripts corresponding to the TAR\textsuperscript{m}, A, B and AB variants were incubated with Pb\textsuperscript{2+} (+) or mock treated (-). The RNA fragments were analyzed by primer extension with oligonucleotide lys21 to analyze the polyA region (A) and with R368 to analyze the DIS region (C). Reactive (black arrows) and protected nucleotides (open arrows) that are discussed in the text are indicated in the LDI and BMH conformation of the leader RNA (Fig 8. B and D). These structures were described previously, with minor adaptations (39). The polyA hairpin (nt 58-103 in wild-type HIV) is extended in the A and B mutants, resulting in a strong reverse transcription stop at nucleotide C\textsuperscript{111} (grey arrow in Fig. 8B).
that left-over TAR sequences may interfere with another viral mechanism. We here demonstrate that the HIV-1 RNA genome requires a stable stem-loop structure at its 5’ end because unpaired nucleotides interact with downstream leader sequences. This interaction results in an unnatural folding of the leader RNA and causes a severe replication defect.

We performed \textit{in vitro} assays with leader transcripts encoding the full-length or truncated TAR elements to screen for an effect on the LDI/BMH equilibrium. In these assays the wild-type leader adopts predominantly the LDI conformation. Strikingly, transcripts with a deletion in one or the other side of the TAR hairpin demonstrate a shift towards the alternative BMH conformation, in which the DIS hairpin motif is exposed. As a consequence we measured enhanced RNA dimerization for these mutants. In the double mutants, the wild-type LDI/BMH equilibrium is restored and these mutants show a low level of dimerization, similar to that of the wild-type leader. RNA structure probing revealed that, upon single-side deletion of TAR sequences, some TAR nucleotides base pair with nucleotides directly downstream of the polyA hairpin, resulting in a stabilization of this structure. As a consequence, the leader RNA shifts towards the BMH structure in which the DIS hairpin is exposed, which explains the increased dimerization of the A and B mutants. In the double mutant AB, the TAR nucleotides fold a truncated, but relatively stable TAR-like hairpin and do not disturb the typical LDI folding. Mfold analysis indicate that a similar interaction between TAR sequences and downstream leader sequences can explain the shift toward the BMH conformation for the C, D, E and F mutants. This scenario is strengthened by the evolution of mutant A, E and F viruses. Using alternative evolutionary strategies, the remnant TAR sequences in these mutants are either removed or modified by mutations such that a stable hairpin structure can be formed at the 5’ end of the viral genome, which prevents an interaction with downstream nucleotides. These changes restore LDI folding and restrict RNA dimerization, as in the wild-type RNA.

For the large set of HIV-1 RNA mutants and evolved variants, we observed a strong correlation between RNA properties (LDI/BMH folding, RNA dimerization) and the Mfold analysis. The notable exception is the deletion mutant E, which is predicted to adopt the LDI conformation in Mfold, yet clearly displays the BMH conformation and increased RNA dimerization in the experiments. The unpaired GGCCC nucleotides at the 5’ end of the E transcript may also interact with downstream sequences and stabilize the polyA hairpin. Alternatively, this dangling 5’ end may induce a tertiary base pairing interaction that stabilizes the BMH conformation, but such interactions are ignored by Mfold. Mutant E displayed delayed replication and the evolved variants remove this “toxic” GGCCC sequence by different means (Fig. 7A). These evolved variants restore the wild-type pattern of restricted RNA dimerization. There are also some minor discrepancies between the MFold analysis, in particular the $\Delta\Delta G$ calculation, and the observed dimerization efficiency of these evolved variants, although one has to realize that quite different structural solutions were selected in cultures ER1/2 and ER3.
Also the phenotype of the FG5 and FG6 evolution products of the F mutant is not supported by the MFold prediction. Whereas MFold analysis predicts an equilibrium for LDI/BMH formation and increased dimerization for the evolved variants similar to that of the F mutant compared to TAR\(^m\), the dimerization assay demonstrates that they dimerize less efficiently than the F mutant, indicating that LDI folding is restored. MFold analysis of the TAR region (Fig. 6B), however, demonstrates that the additional G nucleotides in FG5 and FG6 do restore the stability of this hairpin. This TAR stabilization may prevent the interaction between TAR nucleotides and downstream sequences that probably caused the increased BMH formation of the F mutant. As a result, the evolved leader transcripts can fold the LDI conformation more efficiently and show a wild-type like dimerization efficiency.

We previously discussed that premature RNA dimerization may not be beneficial for the virus (30). If an increased dimerization level in our *in vitro* assays correlates with increased dimerization *in vivo*, the unnatural dimerization capacity of the single-side mutants may explain their replication defect. However, it cannot be excluded that an aberrant folding of the viral leader RNA in these mutants does not only affect dimerization but also other processes like RNA stability, transport, packaging, processing and translation. Notably, extension of the polyA hairpin results in partial disruption of the U5-AUG duplex, the interaction of nucleotides +105/+115 (downstream of the polyA hairpin) with nucleotides +334/+344, including the gag-AUG start codon (27, 39). The U5-AUG duplex is formed exclusively in the genomic HIV-1 RNA that contains the gag sequences and not in the subgenomic spliced transcripts (39). As said, these adverse effects are neutralized in the TAR double mutants and the evolved variants of mutants A, E and F. In these variants the remnant TAR sequences can either fold a TAR-like structure that is stable enough to avoid an interaction with downstream sequences or the remnant TAR sequences are completely removed.

In conclusion, we demonstrate that extreme caution is warranted when mutating the highly structured HIV-1 leader RNA. When structured elements are modified, single-stranded regions may be created that cause unwanted side effects by affecting the proper folding of the other leader RNA replication signals. An effect of leader mutations on the BMH/LDI equilibrium of the HIV-1 leader RNA was previously suggested in a theoretical study that used experimental data obtained for 38 HIV-1 leader mutants (15). We here demonstrate that partial deletions in TAR do indeed affect overall leader structure and thus have detrimental effects on viral replication.

**ACKNOWLEDGEMENTS**

We thank Wim van Est for art work. This research was sponsored by NWO-CW (Top grant) and the Dutch AIDS Foundation (Aids Fonds grant 2005022).
REFERENCES

CHAPTER FOUR

Destabilization of the TAR hairpin of the HIV-1 genome causes aberrant RNA dimerization and packaging

Martine M. Vrolijk, Alex Harwig, Ben Berkhout and Atze T. Das
CHAPTER 4

ABSTRACT

The TAR hairpin is present at both the 5’ and 3’ end of the HIV-1 RNA genome. The 5’ element binds the viral Tat protein and is essential for Tat-mediated activation of transcription. We recently observed that complete TAR deletion is allowed in the context of an HIV-1 variant that does not depend on this Tat-TAR axis for transcription. Destabilization of the 5’ hairpin did however affect the leader RNA conformation and resulted in a severe replication defect. In this study, we demonstrate that 5’ TAR destabilization results in reduced packaging of the full-length RNA genome into virions and increased encapsidation of spliced viral transcripts. Furthermore, dimerization of the packaged full-length RNAs was affected. In contrast, truncation of the TAR hairpin did not affect RNA packaging and dimer formation. These results demonstrate that, although the TAR hairpin is not essential for RNA packaging and dimerization, mutations in TAR can significantly affect these processes through misfolding of the relevant RNA signals.

INTRODUCTION

Human immunodeficiency virus type-1 (HIV-1) is a retrovirus with an RNA genome of approximately 9 kb that contains nine open reading frames and untranslated regions at the 5’ and 3’ end. The highly conserved leader RNA at the 5’ end contains several important regulatory RNA motifs that are involved in both early and late replication steps (1). The first 97 nt of this leader RNA consists of a repeat region (R) that is also present at the 3’ end of viral transcripts (Fig.1A). This repeat allows the first strand transfer step during reverse transcription and can fold two stem-loop structures: the trans-acting responsive (TAR) element and the polyA hairpin. The 5’ TAR hairpin has an important role in transcription activation by binding the viral Tat protein and the cyclin T1 subunit of the positive transcriptional elongation factor (pTEFb) (2, 3). The polyA hairpin masks the polyadenylation signal AAUAAA and its stability is delicately balanced to prevent premature polyadenylation at the 5’ end, yet allow efficient polyadenylation at the 3’ end (4, 5). Two important RNA elements involved in the initiation of reverse transcription, the primer binding site (PBS) and the primer activation signal (PAS), are positioned downstream of the 5’ R region in the untranslated leader (Fig. 1A) (6, 7). Additional RNA signals include the dimerization initiation signal (DIS), the major splice donor site (SD) and the packaging signal Ψ. The DIS hairpin presents a palindromic loop sequence for kissing-loop base pairing and RNA dimerization. The SD site is used for the production of all spliced transcripts and the stability of this hairpin modulates the splicing efficiency (8). The Ψ signal is exclusively present in unspliced transcripts and is important for packaging of the full-length genome into virions. This packaging signal is still partly defined, and other cis-acting sequences in the HIV-1 leader RNA
Figure 1. The HIV-rtTA genome and mutations in the TAR hairpin. (A) The HIV-rtTA proviral DNA genome and the organization and secondary structure of the leader RNA are shown. In HIV-rtTA the Tat-TAR axis of transcription regulation was inactivated by mutation of both Tat and TAR (tat<sup>m</sup> and TAR<sup>m</sup>; crossed boxes). Transcription and replication of the virus was made dox-dependent by the introduction of tetO elements in the U3 promoter region and replacing the Nef gene by the rtTA gene. The untranslated leader RNA of HIV-1 (+1/+368) consists of several regulatory domains (TAR; polyA: polyadenylation signal; PAS: primer activation signal; PBS: primer binding site; SD: splice donor; Ψ: RNA packaging signal; AUG: translation start codon of gag) and can fold the branched multiple hairpin (BMH) and long distance interaction (LDI) conformation (see text for details). (B) The wild-type TAR hairpin (TAR<sup>wt</sup>) and the TAR<sup>m</sup> version with bulge and loop mutations as present in HIV-rtTA are shown. The TAR<sup>m</sup> sequence is partially deleted in the mutants A, B and the double mutant AB. The deleted nucleotides are indicated by a grey box.
have been suggested to contribute to the packaging efficiency, including the upstream TAR and polyA hairpins (9-14).

In vitro studies demonstrated that the HIV-1 leader RNA can not only fold this branched multiple hairpin (BMH) conformation but also an alternative structure in which the polyA and DIS hairpins have to be opened in order to interact (15-17) (Fig. 1A). This long-distance interaction (LDI) prevents exposure of the DIS element and the premature formation of RNA dimers (15, 18, 19). The BMH and LDI conformers may provide the virus with a riboswitch that controls leader RNA functions like dimerization and packaging. Although the TAR hairpin is present in both leader RNA conformations, it has previously been suggested that TAR changes can affect the LDI-BMH riboswitch and consequently several leader RNA functions (16). Indeed, TAR has been shown to influence dimerization of the viral RNA genome (20, 21), packaging of the genomic RNA into virions (16, 18-20, 22, 23) and the strand transfer step of reverse transcription (24-27). Previous attempts to dissect the functions of TAR in vivo, i.e. in the context of the replicating virus, were hampered because mutations in this element cause a severe transcription problem and replication defect (9, 22, 28). We created an HIV-1 variant with an alternative transcription axis and demonstrated that the TAR hairpin can be truncated and even completely deleted when not required for Tat-mediated activation of transcription (23). Surprisingly, virus variants in which the TAR hairpin was partially deleted were found to have a severe replication defect. We showed that destabilization of the 3’ TAR hairpin severely affects polyadenylation of the viral transcripts (29). Furthermore, we demonstrated that opening of the 5’ TAR element caused aberrant folding of the leader RNA in vitro, resulting in uncontrolled dimerization of leader RNA transcripts due to induced exposure of the DIS hairpin (30). In this study, we used the same set of TAR mutants to demonstrate that the processes of RNA dimerization and packaging in virus particles are seriously affected by misfolding of the leader RNA.

**Material and Methods**

**Cells and viruses**

C33A cervix carcinoma cells (ATCC HTB31) (31) were grown as a monolayer in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 10% FCS and minimal essential medium nonessential amino acids, penicillin (100 U/ml) and streptomycin (100 µg/ml) at 37°C and 5% CO₂. Construction of the infectious HIV-rtTA molecular clone and variants with a deletion in the 5’, the 3’ or both TAR elements was described previously (23, 32). In all constructs used in this study an SV40 polyadenylation site is positioned downstream of the viral genome as described previously (29).
**RNA isolation**

C33A cells were cultured to 60% confluency in 10-cm² wells and transfected with 5 µg HIV-rtTA plasmid by calcium phosphate precipitation as previously described (32). Cells were cultured in the presence of 1 µg/ml doxycycline (dox) (Sigma D-9891) and the culture medium was changed after 16 h. The virus level in the culture medium was quantitated at 2 days after transfection by CA-p24 enzyme-linked immunosorbent assay (ELISA) (33). RNA was isolated from the culture supernatant by the method of Boom et al. (34). The cells were washed with phosphate buffered saline (PBS), briefly incubated with 0.5 ml 0.05% trypsin-EDTA (Invitrogen) till they detached and resuspended in 1 ml 10% fetal calf serum-containing medium to inactivate trypsin. Cells were subsequently centrifuged at 2,750 x g for 5 min, washed in 1 ml PBS, centrifuged at 2,750 x g for 5 min, resuspended in 0.6 ml RLT buffer (QIAGEN) and homogenized with a QIAshredder column (QIAGEN). Total RNA was isolated with an RNeasy kit (QIAGEN) and contaminating DNA was removed with RNase-free DNase (QIAGEN) during isolation.

For the isolation of RNA from purified virions, C33A cells were cultured in 75-cm² plates to 60% confluency, transfected with 40 µg HIV-rtTA and cultured with 1 µg/ml dox. The culture supernatant was harvested 2 days after transfection and cells were removed by low-speed centrifugation (10 min at 1500 rpm). The supernatant was filtered through a 0.45-µm filter and virion particles were pelleted by centrifugation at 32,000 rpm (175,000 g) for 90 min at 4°C in a Beckman SW32 Ti rotor. The virions were resuspended in 0.6 ml RLT buffer. The RNA was isolated as described above and resuspended in 50 µl water.

**Northern blot analysis**

For denaturing Northern blot analysis, 5 µg cell-derived RNA or 10 µl virion-derived RNA was mixed with 10 µl denaturing sample buffer, heated at 65°C for 10 min and subsequently electrophoresed on a 1% agarose gel in MOPS buffer (40 mM MOPS, 10 mM sodium acetate, pH 7.0) with 7% of formaldehyde at 100 V for 4 h. For non-denaturing Northern blot analysis, 10 µl virion-derived RNA was mixed with 5 µl sample buffer (30% glycerol and 0.25% bromophenol blue), incubated at 37°C or 55°C for 10 min and electrophoresed on a 0.9% agarose gel in 1 x Tris-borate-EDTA buffer at 100 V for 4 h. The RNA was subsequently denatured by soaking the gel in 3 volumes of 10% formaldehyde at 65°C for 30 min.

The RNA was transferred onto a positively charged nylon membrane (Boehringer Mannheim) with 20 x SSC by means of capillary force for 16 h. The RNA was linked to the membrane using a UV crosslinker (Stratagene). A ³²P-labeled probe corresponding to the nearly complete HIV-rtTA genome was generated by random-primed labeling (High Prime DNA Labeling kit; Roche Diagnostics) of the 3.9 and 5.5 kb SalI fragments derived from the HIV-rtTA plasmid. Prehybridization and hybridization was done in ULTRAhyb buffer (Ambion) at 55°C for 1 and 16 hours, respectively. The membrane was then washed two times at room temperature for 5
min with low-stringency buffer (2×SSC, 0.2% SDS) and two times for 10 min at 50°C in high stringency buffer (0.1×SSC, 0.2% SDS). Images were obtained using the PhosphorImager (Amersham Biosciences) and data analysis was performed with the ImageQuant software package.

**RT/PCR analysis of viral RNA**

RNA was reverse transcribed with ThermoScript reverse transcriptase at 50°C (Invitrogen) using the supplied oligo(dT)25 and random hexamers primers. The cDNA product was used as template in a PCR assay with primers 1 (GAG ACC ATC AAT GAG GAA GCT GCA GAA TGG GAT; position +942/+974 with +1 as the transcription start site) and 2 (GCC CGC CCC TTG TAG GCC GGC CAG ATC TTC CC; +1663/+1638) to detect unspliced RNA, with primers 3 and 5 (CTC CGC AGA TCG TCC CAG AT; +8102/+8083) to detect the double spliced RNA. The cDNA was denatured at 94°C for 5 min and PCR-amplified in 30 cycles of 1 min 95°C, 1 min 55°C, 2 min 72°C and a final extension time of 7 min at 72°C. The PCR products were visualized on a 1% agarose gel stained with ethidium bromide. In the co-transfection experiment, the full length transcript was detected with primers 1T/1B (CTA ATA CGA CTC AGT ATA GGG TCT CTC TGG Δ11/24-AGC ATT GGA; original HIV-rtTA positions +1/+33) and 2T/2B (CAT CGA TCT AAT TCT CCC CTT AAT ACT GAC GC; +382/+348), and the splice products were detected with primers 1 and 4 (single spliced) or 5 (double spliced). The PCR products were digested with NarI before gel analysis.

**RESULTS**

We previously designed an HIV-1 variant in which the Tat-TAR transcription mechanism was inactivated through mutation and functionally replaced by the doxycycline (dox)-inducible Tet-On gene regulation system (Fig. 1A). This HIV-rtTA variant replicates exclusively in the presence of dox and does not require TAR for the activation of transcription. To study additional functions of the TAR hairpin in HIV replication, we generated HIV-rtTA variants in which the right or the left side of the TAR stem was partially deleted (A and B mutants in Fig. 1B, respectively). These deletions open the TAR stem and thus destabilize this RNA element. The A and B deletions were combined in the AB double mutant, which resulted in truncation of the TAR hairpin. Whereas the A and B virus mutants showed a severe replication defect, the AB mutant replicated as efficiently as the original HIV-rtTA virus (23). Further analysis revealed that opening of the 3′ TAR hairpin (A3 and B3 variants) caused reduced polyadenylation of the viral transcripts, whereas truncation of this element (AB3) did not (29). Moreover, using in vitro dimerization assays, we observed that opening of 5′ TAR (A5 and B5) caused increased dimerization of leader RNA transcripts, whereas the AB5 transcripts showed the regulated wild-type dimerization
pattern (30). We here set out to analyze which step of the HIV-1 replication cycle is affected by 5’ TAR-destabilization.

We first analyzed the effect of the TAR mutations on the production of HIV-1 RNA and its packaging into virions. C33A cervix carcinoma cells were transfected with the HIV-rtTA molecular clones that contain either the wild-type or modified TAR sequence in the 5’ LTR, the 3’ LTR or both the 5’ and 3’ LTR, and cultured with dox for two days. Because the A3 and B3 mutations reduce 3’ polyadenylation of the viral transcripts, all molecular clones contained an SV40-derived polyadenylation site...
downstream of the viral genome. As previously demonstrated, transcripts that are not polyadenylated at the 3’ LTR site will be polyadenylated at this SV40 site, resulting in a 276-nt extension (29).

To quantify the viral RNA present in cells and virions, RNA was isolated from the cell and culture medium, respectively, and analyzed in a reverse transcription/polymerase chain reaction (RT/PCR) assay. The isolated RNA was used for the synthesis of cDNA, which was PCR amplified with primer combinations that detect unspliced or double spliced viral RNAs (primers indicated in Fig. 2A). The intracellular level of unspliced and double spliced viral RNAs was similar for all HIV-rtTA variants (Fig. 2B), indicating that the TAR deletions do not significantly affect the production and splicing of the viral transcripts. However, marked differences were apparent in the viral particle compartment (Fig. 2C). A reduced level of unspliced RNA was observed for HIV-rtTA variants with the A or B deletion in 5’ TAR (A⁵, B⁵, A⁵⁺³ and B⁵⁺³). Intriguingly, the level of spliced HIV-1 transcripts was concomitantly increased in these virions. No such effects were observed upon mutation of 3’ TAR. Moreover, the AB mutated viruses, which have a truncated TAR motif instead of a destabilized hairpin, also did not reveal this pattern. These results indicate that destabilization of the 5’ TAR hairpin results in reduced packaging of the full-length RNA and a concomitant increase in packaging of spliced viral transcripts.

We next transfected C33A cells with the original and 5’-mutated HIV-rtTA variants and purified the virions from the culture supernatant by ultracentrifugation.
We isolated RNA from both the cells and virions to confirm the effect of 5’ TAR destabilization on RNA packaging by means of Northern blotting. The intracellular level of unspliced, single spliced and double spliced viral RNAs was similar for the wild-type and 5’ mutated constructs (Fig. 3A, left panel). We again observed that the level of 9 kb genomic RNA in virions was reduced for the A5 and B5 mutants, but not for the AB5 variant (Fig. 3A, right panel). Efficient packaging of the genomic RNA was also observed for variants in which TAR had been completely deleted or replaced by an unrelated hairpin (ER2 and ER3 variants described in ref. (23); results not shown). Whereas the level of the single spliced and double spliced RNA in the wild-type and AB5 virions was too low to detect by Northern blot analysis, the 2 kb double spliced RNAs were apparent for the A5 and B5 mutants (indicated with an arrow in Fig. 3A). RT/PCR analysis of these RNA samples with primers that specifically detect the single and double spliced transcripts demonstrated that not only the level of the double spliced RNAs but also the level of the single spliced RNAs was increased by TAR destabilization (Fig. 3B, lower and middle panels). Furthermore, the concomitant decrease in the level of virion genomic RNA was confirmed (Fig. 3B, upper panels).

It has previously been demonstrated that spliced HIV-1 transcripts can be packaged into virion particles, although full length transcripts are packaged much more efficiently (35-38). Both transcripts are encapsidated through the same mechanism and compete for the same trans-acting packaging factors (37). Accordingly, the increased packaging of spliced transcripts upon 5’ TAR destabilization may be a direct consequence of the reduced packaging of the full length transcript. Alternatively, TAR destabilization may specifically improve packaging of the spliced transcripts. To discriminate between these two possibilities, we co-transfected cells with mutants A and AB and analyzed intracellular and virion RNA by RT/PCR. We used primers that anneal upstream and downstream of the TAR deletion, which resulted in differently sized products for the A and AB variants (AB products being 10 nucleotides smaller than the A products). Analysis of the intracellular RNA revealed a similar production of full length and spliced transcripts for the A and AB variants (Fig. 4, left panel). Analysis of the full length and spliced RNAs present in the virus particles resulting from this co-transfection showed that these virions contain predominantly AB transcripts (Fig. 4, right panel). These results demonstrate that destabilization of the 5’ TAR element reduces packaging of both the full length and the spliced viral RNAs. Since TAR destabilization does not improve packaging of spliced transcripts, it seems likely that the increased level of spliced RNAs in the A and B mutated virions is a consequence of the reduced packaging of the full length transcripts.

We previously observed that the A and B mutations affect the in vitro dimerization of HIV-1 leader transcripts (30). We therefore examined the dimeric state of the virion RNA by non-denaturing Northern-blot analysis (Figure 5, left panel). Whereas the genomic RNA molecules of the HIV-rtTA virions were predominantly present as dimer, the A5 and B5 mutated virions contained no discrete dimeric molecules. The latter virions produce a smear in the non-denaturing RNA gel,
including seemingly large complexes that run high in the gel. The formation of discrete RNA dimers was restored for the AB5 double-mutated virions, indicating that the aberrant dimerization is caused by opening of the TAR structure. Upon denaturation of the RNA at 55°C prior to gel analysis, a strongly reduced level of 9 kb RNA genome was measured for the A5 and B5 mutants (Fig. 5, right panel), which confirms the observed packaging defect. These results suggest that destabilization of the 5' TAR element does not only reduce packaging of the genomic RNA but also affects the process of RNA dimerization.

**DISCUSSION**

This study is based on the surprising finding that the HIV-rtTA variant, which does not need the TAR RNA element for the activation of transcription, is severely replication impaired by deletions in TAR that open the hairpin structure, but not by larger deletions that truncate this stem-loop element. We demonstrate that muta-
tions that destabilize the 5’ TAR structure do not influence the production and splicing of viral RNA. However, we measured reduced packaging of the full length RNA into virions. This reduced level of full length RNA in virions coincided with an increased level of the spliced RNAs. In addition, no discrete RNA dimers were formed upon 5’ TAR destabilization. In contrast, the viruses with a truncated TAR hairpin exhibited normal RNA packaging and dimerization. These RNA packaging and dimerization results are thus fully consistent with the virus replication phenotype.

The full length viral RNA is used as mRNA for the synthesis of the Gag and Pol proteins and as genomic RNA that is encapsidated in the viral particle. During this packaging process, the genomic RNA must be selected from a multitude of cellular and spliced viral RNAs. This selection involves several cis-acting RNA elements and the trans-acting nucleocapsid domain of the Gag protein (39). The major packaging signal ψ is positioned downstream of the splice donor site and is thus exclusively present in the unspliced transcripts (40). Other leader RNA motifs upstream and downstream of the SD site have also been implicated in the process of RNA packaging (9, 12, 13, 41). The upstream elements (including the TAR, polyA, PBS and DIS elements) are present in both the full length and spliced RNAs, suggesting that these elements cannot contribute to the packaging specificity. However, although the spliced HIV RNAs are packaged much less efficiently than the unspliced RNA, they are packaged more efficiently than cellular RNAs, which argues for a contribution of the upstream region in packaging. Opening of the 5’ TAR structure caused reduced packaging of the full length RNA and increased packaging of spliced RNAs.
RNAs. These results suggest that the spliced and unspliced viral RNAs compete for packaging, which is in agreement with previous observations (35-37). We demonstrated that TAR-destabilization did not selectively improve packaging of the spliced transcripts, but rather reduced the packaging efficiency of both the unspliced and spliced RNAs. Apparently, packaging of the full length transcripts is affected more seriously by TAR opening, as the relative amount of spliced transcripts increases. Possibly, the 5’ packaging signals upstream of the SD are differently exposed in the unspliced and spliced RNAs and – as a consequence – they may be differently affected by the TAR mutations.

Previous in vitro RNA structure probing experiments showed that opening of the TAR structure resulted in an interaction between unpaired TAR nucleotides and downstream sequences, which caused extension of the adjacent polyA hairpin (30). Stabilization of this hairpin results in occlusion of the polyadenylation signal and reduced polyadenylation at the 3’ end of the viral transcripts (4, 5, 29, 42). An altered RNA conformation may also explain the deleterious effect of TAR opening in the 5’ leader RNA. Whereas the wild-type leader transcript adopts the LDI conformation, the A and B mutated RNAs shift toward the alternative BMH conformation in which the DIS sequence is exposed (30). This LDI to BMH switch is triggered by the same polyA hairpin extension. Ooms et al. (16) previously showed that a change in the LDI/BMH equilibrium correlates with reduced packaging of the genomic RNA into virus particles. In agreement with this, we observed that the A and B mutations do indeed reduce packaging of the full length RNA. Moreover, we found that the packaged A^5 and B^5 genomic RNA molecules do not form discrete RNA dimers but instead form large complexes of variable size. We previously measured aberrant dimerization of TAR mutated leader RNA molecules in in vitro assays. The A and B RNAs were found to dimerize more efficiently than the wild-type and AB molecules, which may be due to the increased exposure of the DIS signal (30). The absence of discrete RNA dimers in virions may be due in part to interactions between the 9 kb genomic RNA and co-packaged 2 and 4 kb spliced HIV RNAs. Moreover, the TAR-destabilizing mutations may have triggered additional interactions between the packaged viral RNAs. For example, opening of the TAR structure may have induced an additional intermolecular RNA interaction between the TAR domains that was previously observed in vitro by Andersen et al. (20). This interaction involves the sequence from the upper part of the TAR structure that is still largely present in the A and B mutated viral RNAs and requires the presence of the viral nucleocapsid protein. The observed reduction of both RNA packaging and RNA dimer formation is in agreement with the correlation between these parameters as described by Ooms et al. (16). Other studies have suggested a coupling of the processes of RNA dimerization and RNA packaging (43-45).

We previously showed that HIV-rtTA and a similarly designed dox-controlled SIV-rtTA variant replicate efficiently upon complete deletion of the TAR hairpin (23, 46). These results indicate that TAR has no additional essential function in the HIV and SIV life cycle other than its role in Tat-mediated activation of transcription.
However, opening of the 5’ TAR hairpin perturbs the leader RNA structure and affects both packaging and dimerization of the viral RNAs (this study). Furthermore, similar destabilization of the 3’ TAR structure results in masking of the polyadenylation signal in the adjacent polyA hairpin and inhibition of polyadenylation of the viral transcripts (29). Our combined studies thus demonstrate that although the TAR hairpin is not required for packaging, dimerization and polyadenylation, mutations in TAR can affect these processes in an indirect way by disturbing the structure of the 5’ and 3’ ends of the viral RNA. Apparently, the wild-type TAR hairpin is sufficiently stable to prevent detrimental interactions with other RNA domains. This insight explains the crucial role of TAR as structured RNA motif and the strong evolutionary pressure to close disrupted TAR elements (23, 30, 47, 48). These studies illustrate how mutations that are designed to explore the function of a specific RNA element can affect viral replication in an indirect way through unforeseen conformational perturbation of the viral RNA genome.

REFERENCES

Destabilization of the TAR hairpin leads to extension of the polyA hairpin and inhibition of HIV-1 polyadenylation

Martine M. Vrolijk, Alex Harwig, Atze T. Das and Ben Berkhout
ABSTRACT

Background
Two hairpin structures that are present at both the 5’ and 3’ end of the HIV-1 RNA genome have important functions in the viral life cycle. The TAR hairpin binds the viral Tat protein and is essential for Tat-mediated activation of transcription. The adjacent polyA hairpin encompasses the polyadenylation signal AAUAAA and is important for the regulation of polyadenylation. Specifically, this RNA structure represses polyadenylation at the 5’ side, and enhancer elements on the 3’ side overcome this suppression. We recently described that the replication of an HIV-1 variant that does not need TAR for transcription was severely impaired by destabilization of the TAR hairpin, even though a complete TAR deletion was acceptable.

Results
In this study, we show that the TAR-destabilizing mutations result in reduced 3’ polyadenylation of the viral transcripts due to an extension of the adjacent polyA hairpin. Thus, although the TAR hairpin is not directly involved in polyadenylation, mutations in TAR can affect this process.

Conclusion
The stability of the HIV-1 TAR hairpin structure is important for the proper folding of the viral RNA transcripts. This study illustrates how mutations that are designed to study the function of a specific RNA structure can change the structural presentation of other RNA domains and thus affect viral replication in an indirect way.

BACKGROUND
All retroviral RNA genomes contain a repeat (R) region at the extreme 5’ and 3’ end. This sequence repeat allows the first strand transfer step of the reverse transcription process, which results in the formation of long terminal repeat (LTR) regions in the proviral DNA. The 97-nt R region in HIV-1 RNA can fold two stem-loop structures, the TAR and polyA hairpins (Fig. 1A). Both motifs have important functions in the biosynthesis of viral transcripts. The TAR hairpin contains a highly conserved 3-nucleotide pyrimidine bulge that binds the viral Tat transactivator protein (1) and an apical 6-nucleotide loop that binds the cyclin T1 subunit of the cellular transcriptional elongation factor (pTEFb) in a Tat-dependent manner (2-4). The TAR bound CDK9 kinase component of pTEFb phosphorylates the C-terminal domain of RNA polymerase II, which enhances the processivity of the elongating polymerase (5, 6). Furthermore, it was demonstrated that pTEFb directs the recruitment of TATA-box-
binding protein (TBP) to the LTR promoter to stimulate the assembly of new transcription complexes (7, 8). In addition to its role in transcription, the TAR hairpin has been suggested to be important for dimerization of the viral RNA genome (9), packaging of the viral RNA into virions (10-14), the strand transfer step of reverse transcription (15), and as a possible HIV-1 derived miRNA with a role in latency (16, 17).

The polyA hairpin encompasses the AAUAAA polyadenylation signal that is recognized by the cleavage polyadenylation specificity factor (CPSF), resulting in polyadenylation of the viral transcripts. Whereas TAR should exert its function in the 5’ LTR promoter context, the polyadenylation signal should be recognized exclusively in the 3’ LTR context. Previous studies indicated that usage of the 3’ polyadenylation site is promoted by an upstream sequence element (USE) in the U3 region that is uniquely present at the 3’ end of viral transcripts (18-22). This element enhances binding of CPSF to the AAUAAA motif (23). The 5’ polyadenylation site may also be inactive because it is positioned close to the transcription initiation site, such that polyadenylation factors have not yet gained access to the nascent transcript through the RNA polymerase II complex (24-26). Moreover, binding of U1 snRNP to the major splice donor site that is uniquely present downstream of 5’ R represses polyadenylation at the 5’ polyadenylation signal (27, 28). We previously demonstrated that the polyA hairpin masks the AAUAAA signal from recognition by CPSF and that the stability of the polyA hairpin is delicately balanced to allow 5’ repression and 3’ activation of polyadenylation (29-31).

We recently used the designed HIV-rtTA variant that does not need TAR for activation of transcription (Fig.1A) (32, 33) to study additional functions of TAR in viral replication by deleting parts of this motif (34). We observed that virus mutants with a deletion on either the left or right side of the TAR stem (mutants A and B in Fig. 1B, respectively) are replication deficient, whereas the double mutant with a truncated TAR stem (AB) and variants with a complete TAR deletion replicate efficiently. This latter result indicates that TAR has no essential function in the viral life cycle other than to accommodate Tat-mediated activation of transcription. To understand why the single-side deletions abolished replication, we previously analyzed the effect of these mutations on the HIV-1 RNA structure (35). These assays with in vitro produced transcripts revealed that the 5’ TAR-destabilizing mutations affect the proposed riboswitch of the leader RNA, the so called LDI-BMH equilibrium (36-38). Whereas the wild-type transcript adopts predominantly the LDI conformation, the A and B mutants demonstrate a shift toward the alternative BMH conformation. As a result, the DIS hairpin that mediates RNA dimerization is more exposed, which affects dimerization (35) and packaging of the transcripts into virion particles (unpublished results). We now demonstrate an effect of 3’ TAR destabilization on 3’ polyadenylation of the viral transcripts in vivo. We propose that unpaired TAR nucleotides extend the polyA hairpin, thus restricting the availability of the AAUAAA signal for CPSF binding and polyadenylation.
**Figure 1. The HIV-rtTA genome and mutations in the TAR hairpin.** (A) The HIV-rtTA proviral DNA genome and the viral RNA transcript are shown. In this virus the Tat-TAR axis of transcription regulation was inactivated by mutation of both Tat and TAR (tat<sup>m</sup> and TAR<sup>m</sup>; crossed boxes) and functionally replaced by the doxycycline(dox)-inducible Tet-ON gene regulation system (32, 33). The tetO elements were introduced in the U3 promoter region and the Nef gene was replaced by the rtTA gene. The R region that is present at both the 5' and 3' end of the viral transcript folds the TAR and polyA hairpin elements. The latter structure is truncated upon polyadenylation at the 3' R. (B) The wild-type TAR hairpin (TAR<sup>wt</sup>) and the TAR<sup>m</sup> version with bulge and loop mutations as present in the HIV-rtTA virus are shown. The TAR<sup>m</sup> sequence is partially deleted in the mutants A, B and AB. The deleted nucleotides are indicated by a grey box. The transcription and replication properties of these mutant viruses are indicated as previously presented (34, 35).
CHAPTER 5

METHODS

Construction of HIV-rtTA variants

Construction of the infectious HIV-rtTA molecular clone and variants with a deletion in the 5’ TAR or both the 5’ and 3’ TAR elements were described previously (32, 34). For the construction of the 3’ TAR-mutated variants, the BamHI-BglI fragment of the constructs with both 5’ and 3’ deletions, which encodes the 3’ half of the viral genome, was used to replace the corresponding fragment in HIV-rtTA. The SV40 polyadenylation signal was inserted downstream of the 3’ LTR in these constructs as follows. We first made a pBlue-NANB cloning vector with NotI, AatII, NcoI and BamHI sites in the multiple cloning site. For this, the primers MCS-NANB- Fw (ggc cgc gac gtc cat ggt cta gat ctg gat cca cgt) and MCS-NANB-Rev (gga tcc aga tct aga cca tgg acg tcg c) were annealed and ligated into the NotI-AatII digested pBluescript fragment of pBue3’LTRext-ΔU3-rtTAF86Y A209T-2ΔtetO (32). The NcoI-BamHI fragment of pGL3-basic (Promega) that encodes the firefly luciferase gene and SV40 polyadenylation site was ligated into the NcoI and BamHI sites of pBlue-NANB, resulting in the pBlue-MCS-luc-SV40pA plasmid. The luciferase gene was removed by digestion with NcoI and XbaI, blunting of the sticky ends with Klenow polymerase and self-ligation of the vector to produce pBlue-MCS-SV40pA. The AatII-BglI fragment of this plasmid, which contains the SV40 polyadenylation signal, was inserted into the AatII and BglII sites downstream of the 3’ LTR in the different HIV-rtTA clones.

Cell culturing

HeLa X1/6 is a HeLa-derived cervix carcinoma cell line containing chromosomally integrated copies of the CMV-7tetO luciferase reporter construct pUHC13-3 (39). HeLa X1/6 and C33A (ATCC HTB31) (40) were grown as a monolayer in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 10% FCS and minimal essential medium nonessential amino acids, penicillin (100 U/ml) and streptomycin (100 µg/ml). All cell cultures were kept at 37°C and 5% CO2.

Transient transfection and RNA isolation

C33A cells were cultured in 10-cm2 wells, grown to 60% confluency and transfected with 5 µg HIV-rtTA construct by calcium phosphate precipitation as previously described (32). Cells were cultured in the presence of 1 µg/ml doxycycline (dox) (Sigma D-9891) and the culture medium was changed after 16 h. The virus level in the culture medium was quantitated by CA-p24 enzyme-linked immunosorbent assay (ELISA) after two days (41). The cells were subsequently washed with phosphate buffered saline (PBS), briefly incubated with 0.5 ml 0.05% trypsin-EDTA (Invitrogen) till cells detached from the plate and resuspended in 1 ml 10% fetal bovine serum-containing medium to inactivate trypsin. Cells were pelleted at 2,750 x g for 5 min,
washed in 1 ml PBS, centrifuged at 2,750 x g for 5 min, resuspended in 0.6 ml RLT buffer (QIAGEN) and homogenized with a QIAshredder column (QIAGEN). Total RNA was isolated with the RNeasy kit (QIAGEN) procedure, and contaminating DNA was removed with RNase-free DNase (QIAGEN) that was added during the isolation procedure as described in the RNeasy protocol.

**Northern blot analysis**

Gel electrophoresis of RNA was performed on a 1% agarose gel in MOPS buffer (40 mM MOPS, 10 mM sodium acetate, pH 7.0) with 7% formaldehyde at 100 Volt. The RNA was transferred overnight onto a positively charged nylon membrane (Boehringer Mannheim) by means of capillary force. RNA was attached to the membrane with a UV crosslinker (Stratagene). The 373-bp EcoRV-HinDIII fragment of HIV-rtTA encoding the U3/R region was 32P-labeled with the High Prime DNA Labeling kit (Roche Diagnostics) and used as HIV-rtTA probe. To generate the SV40 probe, the HIV-rtTA-SV40 TARm molecular clone was digested with AatII and BamHI and the 276-bp fragment was isolated and labeled as described above. Prehybridization and hybridization was done in ULTRAhyb buffer (Ambion) at 55°C for 1 and 16 h, respectively. The membrane was then washed two times at room temperature for 5 min in low-stringency buffer (2×SSC, 0.2% SDS) and two times for 10 min at 50°C in high stringency buffer (0.1×SSC, 0.2% SDS). Images were obtained using the PhosphorImager (Amersham Biosciences) and data analysis was performed with the ImageQuant software package. The Northern blot was stripped by boiling the membrane at 70°C in 0.1% SDS for 3 times 1h. The stripping efficiency was controlled by scintillation counting and the blot was hybridized with the SV40 probe after prehybridization.

**Splicing and polyadenylation assay**

To analyze splicing and polyadenylation of the viral transcripts, the isolated RNA was reverse transcribed with ThermoScript reverse transcriptase at 50°C (Invitrogen), using the oligo(dT)25 and random hexamers primers. The cDNA product was used as template in a polymerase chain reaction (PCR) with primers 1 (GAG ACC ATC AAT GAG GAA GCT GCA GAA TGG GA) and 2 (GGC CGG CCC TTG TAG GCC GGC CAG ATC TTC CC) to detect the unspliced RNAs, with primers 3 (TCA ATA AAG CTT GCC TTG AGT GC) and 4 (CTC CGC AGA TCG TCC CAG AT) to detect the single spliced RNAs, and with primers 3 and 5 (CTA TGA TTA CTA TGG ACC ACA CA) to detect the double spliced RNAs. The polyadenylated RNAs were detected with primer 6 (CTG TGT CAG AGT AT) that anneals to the polyA tail. The cDNA was denatured at 94°C for 5 min and amplified in 35 cycles of 1 min 95°C, 1 min 55°C, 2 min 72°C and a final extension time of 7 min at 72°C. The PCR products were visualized on a 1% agarose gel stained with ethidium bromide.
rtTA assay
HeLa X1/6 cells were cultured in 2-cm2 wells to 60% confluency and transfected with 1 µg of the HIV-rtTA constructs and 0.5 ng pRL-CMV (Promega), in which the expression of Renilla luciferase is controlled by the CMV immediate-early enhancer promoter, to allow correction for differences in transfection efficiency. The cells were cultured with dox (1 µg/ml) and the medium was refreshed after 16 h. The culture medium was collected after 48 hours for CA-p24 measurement. The cells were washed with 1 ml PBS and subsequently lysed with passive lysis buffer (Promega). Firefly and Renilla luciferase activities were determined with the dual-luciferase assay (Promega). The rtTA level was calculated as the ratio between the firefly and Renilla luciferase activities and corrected for between session variation (42).

RESULTS

HIV-rtTA expression is reduced by destabilization but not by truncation of the TAR hairpin
We previously demonstrated that the TAR-destabilizing A and B mutations induce an alternative folding at the 5’ end of the HIV-rtTA transcripts. Since TAR is part of the R region that is present at both ends of the viral RNA, the TAR deletions may also affect 3’ RNA functions such as polyadenylation. We therefore analyzed the effect of the TAR deletions on viral gene expression, RNA production and processing. C33A cervix carcinoma cells were transfected with the HIV-rtTA molecular clones that contain either the original TARm hairpin or modified TAR sequences at both the 5’ and 3’ LTR. After culturing the cells with dox for two days, we quantified virus production by measuring the CA-p24 level in the culture medium (Fig. 2A). CA-p24 production was reduced for the A and B mutants, and it was restored for the AB variant. In addition, we transfected the HIV-rtTA variants into HeLa X1/6 cells, which contain an integrated rtTA/dox-responsive luciferase reporter construct. The luciferase level measured after two days of culturing with dox reflects the production of the virus-encoded rtTA protein. This analysis revealed that rtTA production was also reduced for the A and B mutants and restored to the wild-type level for the AB mutant (Fig. 2B).

Northern blot analysis of RNA isolated from transfected C33A cells revealed that the reduced viral protein production of the A and B mutants correlated with a reduced level of unspliced (9 kb), single spliced (4 kb) and double spliced (2 kb) HIV-rtTA transcripts, whereas normal amounts were observed for the AB mutant (Fig. 2C). Several novel viral transcripts were detected for the A and B mutants, which were not observed with the TARm and AB viruses (open triangles in Fig. 2C). In addition, RNA molecules with an unexpected size were detected for all virus constructs (grey triangles). Because the viral transcripts were produced from transfected circular HIV-rtTA plasmids, these artificial RNAs may be the product of
improper initiation of transcription at the 3' LTR or incomplete 3' LTR polyadenylation of correctly initiated transcripts. Both events will result in the production of odd-size transcripts that comprise vector sequences, which complicates the analysis.

Destabilization of the 3' TAR element hinders polyadenylation of viral transcripts. To avoid inclusion of vector sequences in the viral RNAs, we made a novel set of HIV-rtTA constructs in which the SV40-derived polyadenylation signal is positioned downstream of the 3' LTR (Fig. 3A). Transcripts starting at the 3' LTR of these constructs will be polyadenylated at the SV40 polyadenylation site and such short RNAs will not be detected on the Northern blot. Transcripts starting at the natural 5' LTR promoter that are not polyadenylated at the 3' LTR will be polyadenylated at the SV40 site, which will result in a discrete 276-nt extension. To distinguish 5' LTR from 3' LTR effects, we made a complete set of 5', 3' and 5'+3' TAR mutants.

C33A cells were transfected with the new HIV-rtTA-SV40 constructs. After two days of culturing with dox, the original and all TAR-mutated HIV-rtTA-SV40 constructs showed no significant variation in the production of CA-p24 (Fig. 3B),
which contrasts with the reduced protein production of the A and B variants that lacked the SV40 element (Fig. 2A). Analysis of the intracellular RNA by Northern blotting did indeed produce a more standard RNA pattern with only the three major RNA classes (9, 4 and 2 kb) (Fig. 3C). Within the set of 5’+3’ TAR mutants an increase in the size of the A and B transcripts was apparent, whereas the size of the AB transcript was similar to that of the original (TARm) virus. This size increase corresponds with what one would expect for read-through transcription to the SV40 polyadenylation site, and was most prominent for the shorter multi-spliced transcripts. The same RNA shift was observed for 3’ mutants A and B, but again not for the AB double mutant. To confirm that the longer transcripts are the result of polyadenylation at the SV40 polyadenylation site, the Northern blot membrane was stripped and hybridized with a probe that specifically detects the SV40 sequences present downstream of the 3’ LTR. This analysis revealed that the extended transcripts do indeed contain this sequence (Fig. 3D).

To rule out aberrant splicing of the viral transcripts, we analyzed the splice pattern of the TAR-deleted HIV-rTAT-SV40 variants in more detail. The isolated cellular RNA was used for the synthesis of cDNA, which was PCR amplified with primer combinations that detect unspliced or spliced viral RNAs (Fig. 3E). This analysis did not reveal any difference between the original virus and modified variants indicating that the 5’ and 3’ mutations do not affect splicing. To confirm that the single-side TAR deletions do affect polyadenylation, the 3’ end of the viral RNA was further analyzed by 3’ RACE (rapid amplification of cDNA ends). The RNA was reverse transcribed using an oligo-dT primer that anneals to the polyA tail and the cDNA product was PCR amplified. For constructs with the original TARm sequence at the 3’ LTR, polyadenylation will result in a PCR product of 939 bp, whereas polyadenylation at the SV40 site will result in a product of 1215 bp (Fig. 3A). For constructs with the A, B and AB deletion in the 3’ TAR hairpin, these fragments will be 14, 10 and 24 bp shorter, respectively. A product corresponding to polyadenylation at the 3’ LTR was observed for all viruses with a TARm or AB sequence at the 3’ LTR (Fig. 3F, 939-bp product for TARm and the 5’ A, B and AB mutants; 915-bp product for the 3’ and 5’+3’ AB variants). This demonstrates that these viruses do efficiently polyadenylate at the 3’ LTR. In contrast, viruses with a single-side deletion in the 3’ TAR element polyadenylated predominantly at the SV40 site, which resulted in the longer PCR product (Fig. 3F, 1201-bp product for 3’ and 5’+3’ A mutants; 1205-bp product for 3’ and 5’+3’ B variants). These results demonstrate that the single-side deletions in the 3’ TAR element reduce the usage of the adjacent 3’ LTR polyadenylation site. Consequently, the downstream SV40 polyadenylation site is used more frequently, resulting in transcripts that are extended by 276 nt. Quantification of the “short” and “long” forms of the double-spliced transcripts on the Northern blot (Fig. 3C) revealed that 80-90% of the viral transcripts with a complete (TARm) or truncated (AB) 3’ TAR structure are processed at the natural 3’ polyadenylation site, whereas only 30-40% of the transcripts with a single-side deletion in 3’ TAR (A, B) are polyadenylated at this position (Fig. 3G).
Figure 3. Destabilization of the 3' TAR hairpin affects polyadenylation. (A) In the HIV-rtTASV40 constructs the SV40 polyadenylation site was placed downstream of the viral genome. The position of the oligonucleotides that were used as primer in the RNA analyses (panels E and F) are indicated. (B) C33A cells were transfected with 5', 3' and 5’+3’ mutated constructs and the CA-p24 level in the culture medium was measured after culturing with dox for 48 h. Average values obtained in three transfections are shown, with the error bars indicating the standard deviation. (C) Intracellular RNA was isolated and analyzed by Northern blotting with a probe against the U3/R region of HIV-rtTA. The position of the 18S and 28S RNA bands, and the unspliced (9 kb), single spliced (4 kb) and double spliced (2 kb) viral transcripts are indicated. (D) The Northern blot was stripped and rehybridized with a probe against the downstream SV40 sequences. Only the extended RNA transcripts observed for the variants with a 3’ A or 3’ B mutation hybridized with this probe. The residual staining of the normally sized transcripts is due to incomplete stripping of the blot. (E) The isolated RNA was used as template
for the production of viral cDNA. The cDNA products were amplified with indicated primers for the unspliced (1+2), single-spliced (3+4) and double-spliced transcripts (3+5). (F) Polyadenylation site usage was analyzed by PCR amplification of the cDNA with primers 6 and 7. Polyadenylation at the 3’ LTR results in a 939-bp product, whereas polyadenylation at the SV40 sequence results in a 1215-bp product. For constructs with the A, B and AB deletion in the 3’ TAR hairpin, these fragments will be 14, 10 and 24 bp shorter, respectively. The identity of these PCR products was confirmed by sequence analysis. (G) The polyadenylation efficiency at the 3’ LTR was calculated by quantification of the 2 kb RNA bands in Fig. 3C.


**DISCUSSION**

We demonstrate that destabilization of the TAR hairpin in HIV-1 RNA results in reduced polyadenylation at the 3’ end of the viral transcripts. Incomplete polyadenylation of the transcripts will result in read-through transcripts that contain vector or cellular genome sequences when proviral plasmids or integrated proviruses are transcribed, respectively. These improperly polyadenylated transcripts may be less stable and we did indeed observe reduced levels of viral RNA and protein (Fig. 2). The extended viral transcripts may face additional problems in the viral life cycle. For example, they may not be efficiently packaged into virions due to size restriction. Incomplete polyadenylation can thus, at least partially, explain the replication defect of the TAR-destabilized HIV-rtTA variants with a single-side deletion in the TAR stem (A and B mutants). In contrast, the double mutant with a truncated TAR hairpin (AB mutant) demonstrated efficient polyadenylation at the 3’ end and viruses with this mutation replicated efficiently (34). Thus, the TAR hairpin itself is not needed for proper 3’ LTR polyadenylation, but TAR destabilization does negatively influence this process.

We recently showed that destabilization of the 5’ TAR element does affect the LDI-BMH equilibrium of the leader RNA (35). Whereas in vitro produced wild-type and AB-mutated leader transcripts adopted predominantly the LDI conformation, the A and B mutants demonstrated a shift toward the alternative BMH conformation. Probing of the RNA structure showed that TAR destabilization liberates TAR-nucleotides that can pair with unpaired nucleotides downstream of the polyA hairpin to extend this structure (35). This extension increases the thermodynamic stability (ΔG) of the polyA hairpin from -17.5 kcal/mole to -24.7 kcal/mole, as predicted by the Mfold program (43). Stabilization of the 5’ polyA hairpin, which is part of the BMH structure, did indirectly affect the LDI/BMH equilibrium and leader-mediated RNA dimerization. Stabilization of the 3’ polyA hairpin explains the reduced polyadenylation of the 3’-TAR destabilized transcripts (A and B mutants), as we previously demonstrated that such an increase in the polyA hairpin stability hinders the binding of polyadenylation factors to the AAUAAA polyadenylation signal (44) and decreases the efficiency of polyadenylation (29-31).

It has previously been suggested that folding of the TAR hairpin is important to appropriately space the upstream sequence element (USE) and 3’ polyadenylation site in the RNA transcript (20). This would resemble the situation in the human T-cell leukemia virus type-I (HTLV-1), where folding of a 276-nt spacer functionally juxtaposes the AAUAAA sequence and the polyadenylation cleavage site (45). The AB-mutation will result in a truncated but stable stem-loop structure that will not affect the spacing between the USE and polyadenylation site. RNA structure probing studies (35) indicated that the A and B mutants fold an alternative, significantly less stable stem-loop structure that may effectively increase the spacer and thus contribute to the reduced polyadenylation efficiency.
Several studies suggested that the TAR hairpin has other functions in HIV-1 replication in addition to its role in transcription, such as in translation, dimerization, packaging and reverse transcription of the viral RNA (9-15, 46-48) (and references therein). Most of these studies were complicated by the fact that mutations in TAR have a dominant negative effect on viral transcription, which obscures other effects in the viral life cycle. Using the HIV-rtTA variant that does not need TAR for the activation of transcription, we recently demonstrated that complete deletion of TAR does not abolish in vitro replication, which indicates that TAR has no other essential function in HIV-1 replication (34). Moreover, our TAR deletion studies demonstrate that TAR destabilization is risky because it induces unwanted side effects. TAR opening triggers the formation of an extended and more stable polyA hairpin, which affects the structure and function of both the 5’ leader and the 3’ end of the viral RNA. These TAR mutations indirectly affect dimerization, packaging and polyadenylation of the viral transcripts, but TAR is not directly involved in these processes. Apparently, the wild-type TAR element is sufficiently stable to prevent the TAR-nucleotides from interacting with other RNA domains.

CONCLUSIONS

Although the TAR hairpin is not directly involved in polyadenylation of the HIV-1 RNA transcripts, destabilization of TAR does affect this process. This study demonstrates that the stability of TAR structure is important for proper folding of the adjacent polyA hairpin.

ACKNOWLEDGEMENTS

We thank Stephan Heynen for performing CA-p24 ELISA. This research was sponsored by NWO-CW (Top grant) and the Dutch AIDS Foundation (AIDS Fonds grant 2005022).
REFERENCES

CHAPTER SIX

Summary and Discussion
SUMMARY AND DISCUSSION

This thesis describes studies performed with a set of HIV-1 variants in which the TAR RNA element was mutated. Because the TAR element is essential for viral transcription through binding of the Tat protein, we purposely used a modified virus that does not need the Tat-TAR axis for transcription. This allowed us to dissect the significance of the many non-transcriptional functions that have been attributed to the TAR element. In fact, one major message from these studies is that there are no other TAR functions that are essential for virus replication in vitro because complete TAR deletion is allowed. Nevertheless, we document severe virus replication defects for several other TAR-mutated HIV-1 variants. To explain this seemingly paradox ("why do certain TAR mutants yield a dead virus as this motif can be deleted"), we have to realize that the TAR RNA hairpin is part of the highly structured untranslated leader region of the HIV-1 RNA genome that contains many important replication signals. The leader encodes for instance essential RNA sequences such as the primer-binding site (PBS) and important RNA structures such as the RNA dimerization initiation signal (DIS hairpin). As such, another major message from this thesis is that mutational analysis of TAR is dangerous because of indirect effects on the structure of the complete leader RNA and its encoded signals. The overall structure of the HIV-1 leader RNA, the different RNA signals and their function in virus replication are introduced in chapter 1.

Many previous studies in which TAR was mutated to study additional roles of this hairpin were complicated by the fact that mutation resulted in abolished transcription and consequently a severe virus replication defect due to disruption of the Tat-TAR axis (Fig. 1). We therefore studied the additional roles of TAR in the background of the HIV-rtTA virus, in which viral transcription is controlled by the introduced Tet-On system (chapter 2). In this context, we deleted large parts of the TAR hairpin on either the left side, the right side or on both sides of the base paired stem, resulting in destabilization or truncation of the hairpin. Surprisingly, we demonstrated that truncation and even nearly complete deletion of the TAR hairpin is compatible with virus replication, but destabilization of the hairpin creates a major replication problem. We cultured the mutated viruses for several months, this resulted in HIV-1 evolution and the selection of virus variants that regained replication competence by the acquisition of additional mutations (1-3). For instance, we described that prolonged culturing of the virus mutant with a nearly complete TAR deletion resulted in the deletion of the remaining TAR sequences. In another evolution experiment we observed the insertion of a new sequence segment that creates a novel stable stem-loop structure at the 5' end of the viral transcript. These TAR-deleted variants replicated efficiently, which demonstrates that, apart from its critical primary role in transactivation of Tat-mediated transcription, there is no essential second TAR function, at least in these in vitro virus replication experiments. Another lesson learned is that HIV-1 replication is severely inhibited by a destabilized
RNA structure at the 5’ end of the viral RNA. In addition, as TAR is repeated near the 3’ end of the viral RNA, additional problems may surface here. We set out to study these issues and to dissect the molecular mechanism(s) that underlie these 5’ and 3’ problems.

Several studies showed that the 5’ TAR hairpin is an autonomous hairpin structure in the untranslated leader RNA that does not interact with downstream sequences. The leader RNA has been proposed to adopt two alternative conformations. In the dimerization-competent branched multiple hairpin (BMH) conformation, the DIS and polyA domains fold as individual hairpins to execute their functions as RNA dimerization motif and suppressor of polyadenylation, respectively. In the alternative folding of the long distance interaction (LDI), these DIS and polyA sequences interact with each other, thus effectively preventing the formation of the individual hairpins. The TAR mutations described above were assessed in vitro for their impact on the LDI-BMH equilibrium of the HIV-1 leader RNA, which is believed to function as a riboswitch (chapter 3). The wild-type HIV-1 leader RNA adopts the LDI conformation and is therefore restricted in RNA dimerization. We demonstrated that TAR destabilized mutant transcripts display the alternative BMH conformation, concomitant with an increased ability to dimerize because the DIS hairpin is exposed. These results were confirmed by probing of the actual leader RNA structure. Very importantly, we documented extension and stabilization of the polyA hairpin by usage of TAR-derived nucleotides that were freed by destabilization of the TAR hairpin (Fig. 1, lower panel). Stabilization of the polyA hairpin explains the LDI-to-BMH switch because this hairpin forms an integral part of the latter conformation. We also studied the RNA properties of two evolved HIV-1 variants that overcome the TAR-destabilization by additional mutations. These variants stabilize what is left of the TAR hairpin and consequently do not affect the polyA hairpin. This also means that the LDI-BMH riboswitch returns to the wild-type position (LDI configuration) that is restricted in RNA dimerization because the DIS hairpin is not exposed.

Two copies of the full length HIV-1 RNA genome are packaged in virion particles as a non-covalently linked dimer. Some studies indicated that spliced HIV-1 transcripts are packaged when leader RNA mutations cause a packaging defect of the full-length genomic RNA. We observed decreased packaging of HIV-1 RNA genomes with a destabilized TAR hairpin and indeed witnessed increased packaging of such spliced HIV-1 transcripts. Consistent with the replication phenotype, this packaging problem was not apparent when TAR was truncated (chapter 4). To determine which of the two TAR elements caused the packaging problem, we repeated the analysis for virions that either contained the TAR mutation at the 5’ or 3’ end of the viral genome. The observed packaging defect was exclusively due to destabilization of the 5’ TAR hairpin. In addition, dimerization of the packaged genomic RNA was reduced by these 5’ TAR mutations. Thus, 5’ TAR destabilization causes unwanted side effects in the leader RNA that affect functions like packaging and dimerization of the viral RNA. Deregulation of these viral processes may be triggered by the changes in the LDI-BMH riboswitch of the HIV-1 leader RNA that were observed in vitro.
Figure 1. Processing of HIV-1 RNA and the impact of a destabilized TAR hairpin. See text for details.
Polyadenylation of the primary HIV-1 transcript is an important measure to prevent degradation of the viral RNA genome. As said, RNA structure probing indicated that the polyA hairpin is stabilized when the TAR hairpin is destabilized. This would create a problem at the 3’ end of the HIV-1 RNA genome because of inaccessibility of the polyadenylation signal, which is embedded in the polyA hairpin, for recognition by polyadenylation factors. We therefore set out to test whether 3’ TAR destabilization would impose such problems. The RNA of HIV-producing cells was analyzed on Northern blot, which indeed showed read-through transcripts of increased size upon 3’ TAR destabilization (Chapter 5). We confirmed by introduction of an exogenous polyadenylation signal that these extended transcripts are caused by read-through transcription due to a lack of polyadenylation.

In our studies we focused on the signals in the HIV-1 leader and trailer RNA that are affected by destabilization of the TAR hairpin on either the 5’, 3’ or on both sides of the HIV-1 RNA genome. Although we observed wild-type levels of transcription for all variants several other processes were severely affected by the deletions in TAR. Figure 1 summarizes the processes involved in the processing of viral RNA that are addressed in this thesis. The upper panel shows the situation for wild-type HIV-1, with transcriptional enhancement by the Tat protein through TAR, polyadenylation at the 3’ end and packaging of the viral RNA in newly assembled virus particles. In the lower panel we illustrate the effects of TAR destabilization in HIV-rtTA for the 5’ and 3’ ends, which affect different RNA processes. Destabilization of 5’ TAR results in stabilization of the 5’ polyA hairpin and a shift in the LDI-BMH equilibrium. As a result, we observed increased dimerization of leader transcripts in vitro. In the context of HIV-1 replication, destabilization of the 5’ TAR hairpin resulted in decreased packaging of the 9 kb viral genome, such that less infectious virus is made (Fig. 1, lower left panel). When deleting parts of the TAR hairpin at the 3’ end of HIV-rtTA, the 3’ polyA hairpin is stabilized and consequently less accessible for polyadenylation (lower right panel).

What does this novel insight, especially the multitude of HIV-1 RNA processes that go wrong when one mutates the TAR hairpin, mean for the interpretation of previous TAR mutagenesis studies? Several previous studies mutated or deleted parts of the TAR hairpin and a wide range of effects were observed (1, 4, 5). However, these studies were seriously hampered because the introduced mutations often affected the Tat-TAR axis, which resulted in abolished replication. The interpretation of these studies is even more complex because the TAR hairpin is also present at the 3’ end of the genome. In the system described in this thesis we were able to separate the TAR hairpin from its essential role in transcriptional activation and to study the other proposed functions of the TAR hairpin. We dissected the defects caused by the deletions in the TAR hairpin at three different levels. We studied the properties of the 5’ TAR-mutated leader RNA in vitro and examined the TAR functions as part of the complete viral RNA genome in cells and virions. Furthermore, we were able to separate the 5’and 3’ TAR effects.
It is generally known that deleting parts of RNA structures is a dangerous act that can have striking structural consequences. However, since the TAR hairpin forms an autonomous RNA structure within the leader and trailer RNA, we did initially not expect major structural problems. As said, we nevertheless witnessed gross problems upon mutation of the 5’ or 3’ TAR element. It may be of interest to broaden these findings to observations made with other viral systems. For example, two RNA deletion studies were performed with the positive stranded RNA genome of porcine reproductive and respiratory syndrome virus (6, 7). The first study demonstrated that some small deletions within ORF7 can be tolerated. However, other deletions resulted in abolished virus production. In the second study smaller deletions were introduced and RNA structure analysis was performed with the Mfold program. The results led to the identification of a secondary RNA structure in the open reading frame that interacted with sequences in the 3’ UTR, and this long-distance RNA interaction appears necessary for virus production.

In fact, it was difficult to find literature about misfolding of RNA due to introduced deletions. With the following search items RNA + virus + misfolding, one obtains 41 hits in PubMed, but most of these studies are in fact about protein misfolding. We suspect that many cases of unwanted RNA structural misfolding go unnoticed or are not published. The results presented in this thesis indicate that such unappreciated adverse RNA structure effects may have serious consequences, in particular because wrong conclusions will be drawn on the biological function of the RNA motif under study.

There are other complications to the study of RNA molecules that relate to the in vivo presence of cellular co-factors that interact with the RNA. We will present a few examples. RNA molecules fold into three-dimensional structures that are essential for processes including translation, mRNA processing and viral replication. In most cases, RNAs obtain their biologically active structure through interactions with protein cofactors or chaperones. It was thought that the functionally active secondary structure is also the thermodynamically most stable secondary structure. The study by Duncan et al. showed by SHAPE analysis of the yeast b13 group I intron holoenzyme that the secondary structure is rearranged extensively by protein cofactors (8). This rearrangement is necessary for formation of the catalytically active tertiary structure. This study demonstrated that the secondary RNA structure determined in vitro can be quite different from the functionally active structure that is acquired in vivo, putting another restriction on the study of RNA molecules. Another example of chaperone activity are DExD/H-box proteins that are involved in RNA-mediated processes and use ATP to accelerate conformational changes in the RNA. Some of the proteins within this family function as RNA chaperones. These RNA chaperones are also involved in refolding of the RNA from the misfolded to the native form. One example is the CYT-19 protein of Neurosporra crassa, which is required for the efficient splicing of several mitochondrial group I introns (9). In this study a ribozyme was used that is derived from a group I intron of Tetrahymena thermophila, this ribozyme is mainly in the misfolded conformation and then slowly refolds into the
native structure. CYT-19 accelerates this folding reaction by promoting partial unfolding of misfolded ribozyme.

Both monovalent and divalent metal ions also have a large effect on the folding of complex RNA molecules (10, 11), although they seem to play distinct roles. Magnesium is more efficient than sodium in inducing RNA tertiary folding and stabilizing RNA structures, but can cause nonspecific RNA collapse. It seems that monovalent cations prevent ribozyme misfolding by altering the initial folding states (12). It was demonstrated that monovalent cations are capable of preventing misfolding of a peripheral structure of the Candida ribozyme and overcome the misfolding caused by magnesium (13). These effects may be missed when performing in vitro studies.

The possibility of alternative RNA folding also greatly complicates the study of RNA structures and their biological function. Such alternative RNA foldings are essential for function in several regulatory mRNA elements or riboswitches. These elements are mainly present in bacteria, but may have escaped identification in other model systems. Riboswitches are usually encoded within the transcript that is regulated in cis to control the expression of the coding sequence(s) of that transcript. Riboswitches have been found to be responsive to a diversity of molecular triggers such as guanine, adenine, thiamine, lysine, glycine (14-17). Several thiamine pyrophosphate sensing riboswitches have been found in fungal and plant genes, where they modulate mRNA splicing and stability (18). We proposed the LDI-BMH riboswitch in the HIV-1 RNA genome is triggered by the viral nucleocapsid chaperone protein (19).

In relation to viral RNA folding another major concern is that important RNA structures can be missed when small or subgenomic RNA segments are studied. For instance, long distance base pairing interactions are common in positive stranded RNA viruses that infect plants (20-22). These interactions are now recognized for different classes of viruses and can span more than 10 kb. A role for long-range interactions in viral replication has also been demonstrated for different flaviviruses (23-25), as well as for the related hepatitis C virus (26-28). Also in minus-strand RNA viruses such RNA interactions may play an important role in the viral replication cycle (29-32). In HIV-1 three such interactions have been described, a long distance interaction between the 5' UTR and the AUG of Gag (33), between leader and Gag sequences (34) and an interaction between the extreme 5’ and 3’ ends, which leads to circularization of the HIV-1 RNA genome (35).

The examples presented above show that many intricate biological processes are involved in RNA structure folding. In our studies we showed that deletions in the HIV-1 TAR hairpin, which forms a separate hairpin structure, nevertheless influence many RNA signals in the surrounding area. As for studies that look at RNA secondary structures by deleting parts of a structure, one should realize that this might have serious unwanted side effects.
REFERENCES

SAMENVATTING

Het humane immunodeficiëntie virus type 1 (HIV-1) veroorzaakt AIDS ("acquired immune deficiency syndrome"). HIV-1 is een virus met een RNA-genoom van ongeveer 9200 nucleotiden. Het grootste deel van dit RNA-molecuul codeert voor de virale eiwitten. Aan beide uiteinden zijn zogenaamde niet-coderende regio’s aanwezig. Deze “leader” (5’ kant, voorkant) en “trailer” (3’ kant, achterkant) regio’s bevatten verschillende regulerende RNA-sequenties en structuren zoals haarspelden (hairpins) die essentieel zijn voor de replicatie van dit virus. In dit proefschrift worden de structuur en functie van deze niet-coderende regio’s bestudeerd.

**Hoofdstuk 1** geeft een overzicht van de HIV-1 replicatiecyclus en van de leader en trailer RNA-sequenties en structuren die later in dit proefschrift onderwerp van studie zijn. De leader-regio bevat verschillende haarspeldstructuren, waaronder TAR, polyA en DIS. Deze RNA-conformatie wordt daarom wel de “branched-multiple-hairpin” (BMH)-conformatie genoemd. De trailer-regio bevat een herhaling (repeat) van een gedeelte van de leader, waaronder de TAR en polyA-sequenties, die hier waarschijnlijk dezelfde haarspeld-structuren aannemen. De leader kan naast de BMH-conformatie ook een zogenaamde “long-distance interaction” (LDI)-conformatie aannemen, waarbij de TAR-structuur nog aanwezig is, maar de polyA en DIS-haarspelden “geopend” zijn zodat deze sequenties met elkaar een interactie kunnen aangaan via baseparing. De TAR-haarspeld in de leader is essentieel voor de productie van de nieuwe HIV-1 RNA-transcripten (transcriptie) omdat het virale Tat-eiwit, de activator van dit proces, aan deze TAR-structuur bindt. Zonder deze binding van Tat aan TAR vindt er normaal onvoldoende RNA-productie plaats en kan het virus zichzelf niet vermengvuldigen (repliceren). De polyA-haarspeld, die dus zowel in de 5’ leader als 3’ trailer kan worden gevormd, bevat een sequentiesignaal dat belangrijk is voor polyadenylering van de RNA-producten plaats en kan het virus zichzelf niet vermengvuldigen (repliceren). De HIV-1 transcripten behoren uiteraard alleen aan het 3’ uiteinde gepolyadenyleerd te worden en de stabiliteit van de polyA-haarspeld is belangrijk voor het voorkomen van vroegtijdige polyadenylering aan de 5’ kant. Een door mutaties te stabiel gemaakte 3’ polyA-haarspeld resulteert ook in remming van de polyadenylering aan de 3’ kant. In elk virusdeeltje zijn twee volledige HIV-1 RNA-moleculen als een dimer aanwezig en de DIS-structuur in de leader is noodzakelijk voor de RNA-RNA interactie, een proces dat RNA-dimerisatie wordt genoemd. Omdat de DIS-haarspeld alleen in de BMH-conformatie aanwezig is, kunnen alleen RNA-moleculen die zich in deze conformatie bevinden dimeriseren.

In **hoofdstuk 2** hebben we bepaald of de TAR-haarspeld essentiële functies vervult tijdens de HIV-1 replicatie, anders dan de bekende en belangrijke functie in de stimulatie van het transcriptie proces. Hierbij hebben we TAR gemuteerd en gekeken naar het effect op de virusreplicatie. Echter, omdat mutatie van TAR normaliter de transcriptie blokkeert, wat elk ander effect op replicatie zal maskeren, hebben we gebruik gemaakt van een speciale HIV-1 variant waarin wij een alternatief systeem hebben geïntroduceerd voor de regulatie van het transcriptie
SAMENVATTING

proces. In dit zogenaamde HIV-rtTA virus hebben wij een set TAR-mutanten gemaakt waarin de linkerzijde, de rechterzijde, of beide kanten van de haarspeld gedeeltelijk zijn weggehaald. De mate van virusreplicatie van deze mutanten is uitgebreid bestudeerd. Het verwijderen van het linker of rechter deel van de haarspeld bleek de replicatie stil te leggen. Het inkorten van de haarspeld in de dubbelmutanten had echter nauwelijks effect op de virusreplicatie. Zelfs na het bijna volledig verwijderen van de TAR-haarspeld replicaerde het virus nog vrij goed. Door middel van "geforceerde evolutie" hebben we geprobeerd het replicatievermogen van deze mutant virussen te verbeteren. Deze methode maakt gebruik van het feit dat het virale "reverse transcriptase" enzym, dat tijdens de replicatie het RNA-genoom omzet in een DNA-molecuul, fouten maakt waardoor er nieuwe virusvarianten ontstaan. In langdurende virus kweken worden de best replicerende virussen geselecteerd door positieve Darwiniaanse selectie volgens het "survival of the fittest"-principe. Op deze manier hebben wij goed replicerende virus varianten verkregen waarin het TAR-element volledig afwezig was. Deze bevinding leidde tot de conclusie dat het TAR-motief niet nodig is voor virusreplicatie in het HIV-rtTA systeem, m.a.w. dat TAR geen andere essentiële functie heeft tijdens HIV-1 replicatie dan de bekende rol in het transcriptieproces.


In hoofdstuk 4 is het effect van de veranderde leader RNA-structuur in de context van het replicerende virus, dus in vivo, nader geanalyseerd. Deze studie laat zien dat de enkelzijdige mutaties die de leader RNA-structuur en het LDI-BMH evenwicht bevloeiden, niet alleen de dimerisatie maar ook het inpakken ("packaging") van de virale RNA-genomen in virusdeeltjes beïnvloeden. De dubbelzijdige TAR-mutaties die de leader RNA-structuur niet aantasten hebben geen
effect op deze processen. Deze studie toont aan dat hoewel het TAR-element niet noodzakelijk is voor HIV-1 RNA-dimerisatie en packaging, mutaties in dit element deze processen wel op een indirecte wijze kunnen beïnvloeden door misvouwing van de leader-RNA.

De TAR en polyA-haarspelden zijn niet alleen in de leader-regio aanwezig maar ook in de trailer-regio. In hoofdstuk 5 laten we zien dat het openen van TAR ook problemen geeft aan deze 3’ kant. Een enkelzijdige TAR-mutatie die de naastgelegen polyA-haarspeld verlengt en stabiliseert resulteert in een verminderd gebruik van het polyadenyleringssignaal dat in deze haarspeld is gelegen. Een dubbelzijdige mutatie in 3’ TAR, die de structuur van de polyA-haarspeld niet beïnvloedt, heeft geen nadelig effect op het proces van polyadenylering. Deze studie maakt duidelijk dat hoewel TAR geen directe invloed heeft op het proces van polyadenylering, destabiliserende TAR-mutaties in de trailer wel indirect het proces van de polyadenylering kunnen beïnvloeden.

Dit proefschrift beschrijft hoe verschillende stappen van de HIV-1 replicatiecyclus beïnvloed worden door mutaties in de TAR haarspeld. Echter, wij laten vervolgens zien dat de TAR-haarspeld geen directe rol speelt in deze processen, maar dat het openen of destabiliseren van TAR een indirect effect sorteert op de structuur van de naastgelegen RNA-signalen, zowel in de leader als de trailer-regio van het HIV-1 RNA-genoom. Dit onderzoek toont hoe gevaarlijk het kan zijn om RNA-structuren te bestuderen met behulp van mutaties. Wanneer de mutaties niet voorzichtig worden gekozen, bestaat het reëlle gevaar dat er door RNA-misvouwing andere problemen ontstaan. Zulke indirecte effecten op de virusreplicatie kunnen gemakkelijk leiden tot misinterpretatie van de resultaten.
DANKWOORD

Dan is het nu toch zover gekomen. Het laatste en misschien wel het lastigste gedeelte van het proefschrift. Het is ook het stuk wat waarschijnlijk het meest gelezen wordt. Ik wil hier iedereen bedanken die van belang was voor het onderzoek of voor mij persoonlijk. En het is o zo belangrijk niemand te vergeten, wat bijna onmogelijk is. Mocht ik iemand vergeten zijn dan zeg ik hier alvast, bedankt!

Allereerst wil ik Raoul de Groot bedanken. Door zijn goede zorgen en zijn groot enthousiasme tijdens mijn eerste stage heb ik gekozen voor de virologie. Ik weet dat deze woorden niet geheel orgineel zijn, maar ik wilde ze wel even kwijt.

Prof. dr. B. Berkhout, beste Ben, ik vind het fijn dat ik onder jouw leiding mijn promotietraject heb doorlopen. Jij zat altijd vol ideeën en theorieën waarbij je altijd trouw zei: "..., maar ik zeg niet dat je dat moet doen". Ik wil je bedanken voor de prettige samenwerking.

Dr. A.T. Das, beste Atze, ik heb veel opgestoken van jouw kritische blik en je praktische tips. Ik wil je bedanken voor je energie en tijd en ook dat je deur altijd open stond voor vragen.

Marcel ik wil jou bedanken voor de introductie in de RNA wereld. Zonder jou zou dat zeker minder succesvol geweest zijn.

Alex, bedankt je hulp op het lab en voor alle gezelligheid op het lab en erbuiten. Soms is er toch wat tijd te doden en dan had je altijd wel wat grappigs.

Hierbij wil ik ook mijn kamergenoten van het laatste uur bedanken voor de gezelligheid en het vullen van de snoeppot. Karin, Martijn, Nick en Ying Poi bedankt! Ying Poi je bent een erg goede vriendin geworden de afgelopen 4 jaar, ik vind het fijn dat je mijn paranimf wilt zijn.

Daarnaast zijn er nog een groot aantal mensen op het lab geweest die ik wil bedanken voor hun gezelligheid. Mijn mede-AOIs Thijs, Walter, Ellen, Karolina, Edwin, Dirk, Mark, Julia, Ronald, Michel, Elly, Gisela en Renee. En de andere mensen op het lab Monique, Bep, Ilja, Wendy, Mireille, Rienk, Lia, Joost, Maarten, Jeroen, Stef, en alle anderen.

Prof. Dr. P. Rottier, beste Peter, bedankt dat ik de mogelijkheid heb gekregen mijn proefschrift af te maken tijdens mijn huidige werk.

Jildou, bedankt voor je steun, toewijding en gezelligheid bij het nieuwe onderzoek, waardoor ik tijd kon vrijmaken voor het aframen van mijn proefschrift.
DANKWOORD

Lieve Josse, Marie-José, Marko en Wendy, we zijn alweer meer dan 10 jaar geleden begonnen met dezelfde studie en sindsdien hebben we vele etentjes en spelletjes-avonden gehad. Ook de uitstapjes naar de uithoeken van Nederland en zelfs Barcelona waren erg gezellig en inspirerend. Het was erg prettig om het soms niet, maar soms ook juist wel over werk en wetenschap te hebben. Bedankt voor jullie vriendschap! Marie-José ik vind het fijn dat ik altijd zo goed met je heb kunnen kletsen en dat je altijd onbevooroordeeld je mening geeft. Ik vind het dan ook een eer dat je mijn paranimf wilt zijn.

Danielle, bedankt voor de hulp bij het maken van de omslag van dit boekje. Het was een erg gezellige middag en het ziet er geweldig uit.

Beste Joost, Bas en Tim, bedankt voor het organiseren van het feest in Brasserie Blazer.

Beste mede-SIGMA-leden, bedankt voor de gezelligheid en afleiding van mijn werk tijdens alle lessen en oefendagen.

Lieve papa en mama, ik wil jullie bedanken voor alle mogelijkheden die jullie mij gegeven hebben en voor de steun de afgelopen jaren. Lieve Xander, Bregje en Afra bedankt voor jullie gezelligheid tijdens verjaardagen en alle spontane gekke knutsel acties. En natuurlijk moet ik dan ook Jolanda, Niek, Tijn, Jeroen en Peter bedanken. Bregje, ik vind het heel fijn dat jij ook een van mijn paranimfen wilt zijn.

Huub, Bernadette, Tijs en Deborah, dank jullie voor alle borrels en lekkere etenjes de afgelopen jaren, dat heeft zeker geholpen te ontspannen tussen al het werk door.

Lieve Luc, ik wil je bedanken voor je enorme relatifervermogen en alle pep-talks. Ik heb je leren kennen toen ik al AIO was en je kent me dus eigenlijk niet zonder stress en rennen. Ik hoop dat ik nu langzaamaan wat meer rust in mijn donder krijg en dat we daar samen van kunnen genieten.

Proost!
Martine Mireille Vrolijk was born on 15th of July 1978 in Arnhem, The Netherlands. She graduated in Biomedical Science at the Utrecht University in 2004. During her first internship at the department of Virology at the faculty of Veterinary Science at the Utrecht University she was introduced in virology. Martine studied the role of the hemagglutinin esterase protein in mouse hepatitis virus. The second internship was at the department of Biomedical Genetics of the University Medical Center in Utrecht. She performed a genomewide screen in a four-generation Dutch family with celiac disease and determined a region involved in this illness in this family. In 2004 Martine started her PhD training at the laboratory of Experimental Virology of the Academic Medical Center in Amsterdam. The results of this research are described in this thesis. In January 2009, Martine joined the Virology Division of the faculty of Veterinary Science at the Utrecht University and started with the development of a new vaccine against porcine reproductive and respiratory syndrome virus.
PUBLICATIONS

Vrolijk MM, Harwig A, Berkhout B, Das AT.
Destabilization of the TAR hairpin leads to extension of the polyA hairpin and inhibition of HIV-1 polyadenylation.

Vrolijk MM, Ooms M, Harwig A, Das AT, Berkhout B.
Destabilization of the TAR hairpin affects the structure and function of the HIV-1 leader RNA.

Das, AT, Harwig A, Vrolijk MM, Berhout B.
The TAR hairpin of human immunodeficiency virus type 1 can be deleted when not required for Tat-mediated activation of transcription.

Luxury at a cost? Recombinant mouse hepatitis viruses expressing the accessory hemagglutinin esterase protein display reduced fitness in vitro.

A genomewide screen in a four-generation Dutch family with celiac disease: evidence for linkage to chromosomes 6 and 9.