Regulation of HIV-1 mRNA Processing by Cellular Splicing Factors

ANNA TRANELL
Dissertation presented at Uppsala University to be publicly examined in C10:301, BMC, Husargatan 3, Uppsala. Thursday, April 26, 2012 at 09:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in Swedish.

Abstract

According to UNAIDS there were 34 million people living with human immunodeficiency virus (HIV) infection at the end of 2010. HIV is the causative agent of acquired immunodeficiency syndrome (AIDS) and the number of people dying of AIDS-related causes at the end of 2010 was 1.8 million. Due to the high mutability of the virus, there is a constant need for new approaches to attack the virus.

Splicing of HIV-1 pre-mRNA is a highly regulated process. In order to produce all mRNAs needed to be infectious HIV-1 utilizes alternative splicing -- from one single transcript more than 35 differently spliced mRNAs can be produced. A new approach to fight HIV-1 could be to interfere with the essential splicing. In this thesis, I describe the regulation of HIV-1 pre-mRNA splicing.

SR proteins are involved in the regulation of splicing, both in an early and a late stage. We find that the intracellular concentration of SR proteins is of great importance for HIV-1 to be able to produce the correct amounts of mRNAs. Variations in concentrations of SR proteins lead to big changes in the HIV-1 pre-mRNA splicing pattern.

The functions of HIV-1 protein Vpr are diverse and it is essential in vivo. HIV-1 vpr mRNA 13a7 is partially spliced, containing an intron, and the regulation of it is not fully understood. We find that SRp55 and SRp75 induce the production of HIV-1 vpr mRNA 13a7 by inhibiting splice donor 3. We also conclude that this inhibition at least for SRp55 is due to an interaction with the viral RNA element GAR. In the presence of SRp55 we also see an increase in cytoplasmic amounts of intron containing vpr mRNA due to increased nuclear export. Our results show that SRp55 can have several functions in the regulation of HIV-1 splicing: by inhibiting splice donors and by facilitating the export of incompletely spliced mRNAs to the cytoplasm.

In conclusion, this thesis describes SRp55 as a regulator of HIV-1 vpr mRNA, both in splicing as well as in nuclear export. These discoveries provide an insight into the regulation of HIV-1 mRNA processing.

Keywords: HIV-1, SR protein, SRp55, SFRS6, splicing, indole derivative

Anna Tranell, Uppsala University. Department of Medical Biochemistry and Microbiology, Box 582, SE-751 23 Uppsala, Sweden.

© Anna Tranell 2012

ISSN 1651-6206
urn:nbn:se:uu:diva-169256 (http://urn.kb.se/resolve?urn=nbn:se:uu:diva-169256)
Till mig själv och min familj
List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


Reprints were made with permission from the respective publishers.
## Contents

- **Introduction** .................................................................................................................. 11
- **Viruses** ......................................................................................................................... 11
- **Retroviruses** .................................................................................................................. 12
- **Human immunodeficiency virus** ................................................................................... 12
  - **AIDS** .......................................................................................................................... 13
  - **Treatment of HIV-1 infection** .................................................................................... 14
  - **HIV-1 genomic organization** ..................................................................................... 15
  - **Virion structure and infectious cycle** ......................................................................... 18
- **Post-transcriptional processing** ..................................................................................... 21
  - **Pre-mRNA splicing** .................................................................................................... 22
  - **SR proteins** ................................................................................................................. 27
  - **HIV-1 pre-mRNA splicing** ........................................................................................ 30
  - **Antiviral therapy and HIV-1 splicing** ........................................................................ 34
- **The present investigation** ............................................................................................... 35
  - **Aims of the study** ....................................................................................................... 35
  - **Experimental system** .................................................................................................. 35
  - **Results** ....................................................................................................................... 36
    - **Paper I** ..................................................................................................................... 36
    - **Paper II** .................................................................................................................... 38
    - **Paper III** .................................................................................................................. 39
- **Concluding remarks and future prospects** ................................................................... 41
- **Acknowledgements** ....................................................................................................... 44
- **References** .................................................................................................................... 47
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIDS</td>
<td>acquired immunodeficiency syndrome</td>
</tr>
<tr>
<td>ALV</td>
<td>avian leucosis virus</td>
</tr>
<tr>
<td>AZT</td>
<td>azidothymidine</td>
</tr>
<tr>
<td>CA</td>
<td>capsid</td>
</tr>
<tr>
<td>CTD</td>
<td>carboxy-terminal domain</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>ELISA</td>
<td>enzyme-linked immunosorbent assay</td>
</tr>
<tr>
<td>ER</td>
<td>endoplasmic reticulum</td>
</tr>
<tr>
<td>ESE</td>
<td>exonic splicing enhancer</td>
</tr>
<tr>
<td>ESS</td>
<td>exonic splicing silencer</td>
</tr>
<tr>
<td>FI</td>
<td>fusion inhibitor</td>
</tr>
<tr>
<td>GAR</td>
<td>guanosine-adenosine-rich</td>
</tr>
<tr>
<td>HAART</td>
<td>highly active antiretroviral therapy</td>
</tr>
<tr>
<td>HIV</td>
<td>human immunodeficiency virus</td>
</tr>
<tr>
<td>HIV-1</td>
<td>human immunodeficiency virus type 1</td>
</tr>
<tr>
<td>HIV-2</td>
<td>human immunodeficiency virus type 2</td>
</tr>
<tr>
<td>hnRNP</td>
<td>heterogeneous nuclear ribonucleoprotein</td>
</tr>
<tr>
<td>IN</td>
<td>integrase</td>
</tr>
<tr>
<td>ISE</td>
<td>intronic splicing enhancer</td>
</tr>
<tr>
<td>ISS</td>
<td>intronic splicing silencer</td>
</tr>
<tr>
<td>LTR</td>
<td>long terminal repeat</td>
</tr>
<tr>
<td>MA</td>
<td>matrix</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>NAb</td>
<td>neutralizing antibody</td>
</tr>
<tr>
<td>NC</td>
<td>nucleocapsid</td>
</tr>
<tr>
<td>NNRTI</td>
<td>non-nucleoside reverse transcriptase inhibitor</td>
</tr>
<tr>
<td>NRTI</td>
<td>nucleoside reverse transcriptase inhibitor</td>
</tr>
<tr>
<td>PBMC</td>
<td>primary blood mononuclear cell</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PPT</td>
<td>polypyrimidine tract</td>
</tr>
<tr>
<td>PR</td>
<td>protease</td>
</tr>
<tr>
<td>Py</td>
<td>polypyrimidines</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RRM</td>
<td>RNA recognition motif</td>
</tr>
<tr>
<td>RS</td>
<td>arginine-serine-rich domain</td>
</tr>
<tr>
<td>RT</td>
<td>reverse transcriptase</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>SA</td>
<td>splice acceptor (3’ splice site)</td>
</tr>
<tr>
<td>SD</td>
<td>splice donor (5’ splice site)</td>
</tr>
<tr>
<td>SIV</td>
<td>simian immunodeficiency virus</td>
</tr>
<tr>
<td>snRNA</td>
<td>small nuclear ribonucleic acid</td>
</tr>
<tr>
<td>snRNP</td>
<td>small nuclear ribonucleoprotein</td>
</tr>
<tr>
<td>SR protein</td>
<td>serine-arginine rich protein</td>
</tr>
<tr>
<td>SRPK</td>
<td>SR-specific protein kinase</td>
</tr>
<tr>
<td>TAR</td>
<td>transactivation response region</td>
</tr>
<tr>
<td>U2AF</td>
<td>U2 snRNP auxiliary factor</td>
</tr>
<tr>
<td>UNAIDS</td>
<td>the Joint Nations Programme on HIV/AIDS</td>
</tr>
<tr>
<td>U snRNA</td>
<td>uridine-rich small nuclear RNA</td>
</tr>
<tr>
<td>U snRNP</td>
<td>uridine-rich small nuclear RNP</td>
</tr>
</tbody>
</table>
Introduction

According to the Joint Nations Programme on HIV/AIDS (UNAIDS) there were 34 million people living with human immunodeficiency virus (HIV) infection at the end of 2010. HIV is the causative agent of acquired immunodeficiency syndrome (AIDS) and the number of people in 2010 dying of AIDS-related causes was 1.8 million. Thanks to changes in behavior and antiretroviral therapy the epidemic is declining in most countries. Apart from faster and broader treatments together with the spread of information there is also a constant need for new approaches and antivirals to attack the virus. In order to develop new strategies, we need to know more about the virus itself, its life cycle and the way it behaves in our cells. We also need to have more knowledge about how the cells and our immune system react during an HIV infection. There are two types of HIV: HIV-1 and HIV-2. HIV-1 is the main cause of AIDS and therefore this thesis is focusing on HIV-1. The splicing process of HIV-1 pre-mRNA is highly regulated and a prerequisite for the virus to be infectious. In this thesis I have therefore studied the regulation of HIV-1 pre-mRNA splicing, with the hope that these findings should open up new possibilities in the fight against the virus.

Viruses

A virus is a contagious material that can only replicate inside a cell. It is debated whether viruses should be considered as living or non-living. The first virus to be discovered was the tobacco mosaic virus. It was found to be the causative agent of tobacco mosaic disease in 1892 by Dimitri Ivanov’sky and independently by Martinus Beijerinck in 1898. Bacteria had already been discovered by that time. And the tobacco mosaic virus was stated by Beijerinck to be an agent that passed filters which normally trapped bacteria and that the agent could multiply within living plants. At the same time foot- and-mouth disease was found to be caused by an agent that could not be retained in filters, an observation made by Friedrich Loeffler and Paul Frosch. Together with the observation that both the foot-and-mouth disease and the tobacco mosaic disease causing agents were unable to grow in solutions where known bacteria could multiply, this made the researchers come to the conclusion that the agents must be something other than bacteria. (Flint and American Society for Microbiology., 2009).
In the beginning researchers thought that this new agent was a liquid. But they were described as particles after d’Herelle’s plaque assay in 1917 and the first electron micrographs of tobacco mosaic virus in 1939. The first human virus to be discovered was the yellow fever virus, by Walter Reed in 1901 (Lustig and Levine, 1992).

Ever since these discoveries 40 000 viruses have been classified, with varieties in size, shape, genome, host cell and life cycle. In addition, studies of viruses have given us knowledge not only about the viruses themselves, but also about the cellular machinery and processes like pre-mRNA splicing and gene silencing.

Retroviruses

One class of viruses that can reverse transcribe their RNA genome into DNA is called retroviruses, and belongs to the family Retroviridae. A retrovirus has the enzyme reverse transcriptase incorporated into the virion. Another enzyme incorporated into the virion is integrase that mediates integration of viral DNA into the host genome (Flint and American Society for Microbiology., 2009). Retroviral genomes are usually 6–11 kb in size and vary in complexity. The simplest retroviruses only have genes encoding the virion proteins, i.e. gag, pol and env, while the more complex retroviruses additionally have genes encoding accessory proteins that affect viral gene expression, fine tune replication and make it possible for the virus to adapt to host. These proteins could also be required in vivo. All retroviral genomes have LTRs (long terminal repeats) in their 5´- and 3´-end, where the promoter and sequences for polyadenylation are positioned (Vogt, 1997b).

Retroviruses were discovered at the beginning of the last century. These retroviruses were associated with disease in chickens and were subsequently named avian leucosis virus (ALV) and Rous sarcoma virus. Ever since then more and more retroviruses have been discovered, some of which infect humans (Vogt, 1997a). One of these viruses, the human immunodeficiency virus (HIV), was discovered in the beginning of 1980s and was later found to cause AIDS.

Human immunodeficiency virus

HIV is an enveloped retrovirus belonging to the retroviral sub-family lentivirus. There are two types of HIV, type 1 and type 2. HIV-2 is rather rare and originates from simian immunodeficiency virus (SIV) that infects the sooty mangabey (Cercocebus atys). HIV-1 is more common than HIV-2 and originates from SIV that infects chimpanzees, called SIVcpz. There are three different sub-groups of HIV-1: M, N and O, which represent transmissions
of SIVcpz to humans on three separate occasions. Sub-group M is the most widespread and is the cause of more than 95% of all HIV-1 infections. This group is further divided into subtypes A-K. Phylogenetic analyses date the common ancestor of the subtypes in the group M between 1902 and 1921. (Gao et al., 1999).

HIV is spread via sexual contact, blood and from mother to child during birth or via breast milk. The first transmission of SIV to humans occurred, probably by exposure to chimpanzee blood, in central Africa. The SIV strains in infected chimpanzees are closest related to HIV-1 group M that has been found in southeastern Cameroon. One theory is that HIV-1 infection spread from there to the great city of Leopoldville (Kinshasa) in Belgian Congo (Democratic Republic of the Congo) where the pandemic could start (Worobey et al., 2008). The spread of the virus was initially rather slow. However during the 1950s and 1960s the epidemic increased. This was probably due to great changes in Africa, with growing cities and the introduction of vaccine programs (reuse of needles). There was also increased travel to and from Africa and a sexual revolution may also have contributed to the spread of HIV (Sharp and Hahn, 2008).

At the end of 2010 34 million people were living with an HIV infection, but the total number is falling compared to the peak in 1997. Of all people infected, 68% are living in Sub-Saharan Africa, a region with only 12% of the global population. The epidemic is most severe in southern Africa. Even though the total number of infected people is decreasing, there is an increase in some parts of the world. For instance in Eastern Europe and Central Asia, there was a 250% increase in the number of infected people from 2001 to 2010 (UNAIDS, 2011).

AIDS

HIV is the causative agent of acquired immunodeficiency syndrome (AIDS) - the state where the immune system is degraded and the host easily acquires opportunistic infections. The first case of AIDS was described in 1981 and about two years later the causative agent of AIDS was found to be HIV. The main discoveries took place in two laboratories: Gallo in Bethesda, Maryland USA and Montagnier at the Pasteur Institute, Paris France (Gallo and Montagnier, 2003, Wainberg and Jeang, 2008). AIDS, caused by HIV, kills more people than any other infectious disease (Flint and American Society for Microbiology., 2009).

In an untreated patient AIDS usually occurs within 8-10 years after the initial HIV infection. AIDS is the state where the numbers of CD4+ T cell counts are below 200 per ml blood (normal count is around 1000 per ml blood) and the numbers of CD8+ T cell counts also decreases. The decrease in immune cells leads to conditions such as cancer (e.g. Kaposi’s sarcoma and B cell lymphomas), oral lesions, basal cell carcinoma and bacterial, fun-
gal and viral infections as well as neurological disorders such as AIDS dementia and meningitis. AIDS-related deaths in 2010 were estimated to 1.8 million. Due to increased treatment and changed behavior this number has been decreasing over the last years throughout the world apart from Eastern Europe and Central Asia (UNAIDS, 2011).

Treatment of HIV-1 infection

Antiviral therapy
The first treatment for HIV infection was azidothymidine (AZT), approved for marketing in 1987 (Mitsuya et al., 1985). There have been several antiviral drugs tested on humans and today more than 20 drugs have been approved for HIV infection treatment. The main targets for HIV drugs are Reverse transcriptase (RT), Protease (PR), Integrase (IN) and virus-cell interaction (attachment, fusion and entry). The RT inhibitors are either nucleoside (NRTI) or non-nucleoside reverse transcriptase inhibitors (NNRTI). The PR inhibitors (PIs) mimic the peptide linkage in the polypeptide gag-pol. The drugs targeting IN inhibit the integration process of the viral genome into the host genome (De Clercq, 2007). Virus-cell interaction inhibitors like fusion inhibitors (FIs) prevent the virus from entering the cell (Pugach et al., 2008). One approach to treating HIV infections is the HAART (highly active antiretroviral therapy) where the patients are treated with a combination of at least three anti-HIV drugs, leading to a lower risk of drug resistance.

The HIV genome is highly mutable due to the lack of proofreading activity of the reverse transcriptase. Mutations may lead to drug resistant viruses and there is a constant need of new types of therapy. Another problem in clearance of the infection is the integration of viral DNA into the host genome and no treatment available today can clear an infection. A third problem is the asymptomatic phase (with viral replication) during the first period of infection. Taken together, all these factors create a constant need for new antivirals and new strategies to discover infections, which demands further investigation of the virus life cycle.

Vaccine
Ever since HIV was discovered researchers have been trying to develop a vaccine in order to stop the epidemic. The major issues in designing a vaccine against HIV-1 are the genetic heterogeneity of the virus (due to the high mutation rate, RNA recombination and immune selection), the integration of the viral DNA into the host genome, virus can spread through cell-cell interaction, virus infects areas of reduced immune surveillance, the infection compromises immune function. Another issue is the ability of the virus to evade the immune system by ‘resting’ in lymphocytes and macrophages. Another problem is that a vaccine would have to protect against millions of
variants of virus particles and so far there is no vaccine available that can protect against HIV infection (Douek et al., 2006). A number of clinical trials have been performed with different vaccine candidates. One study was performed where researchers evaluated priming injections of recombinant canarypox HIV-vector vaccine (ALVAC-HIV) and booster injections of a recombinant gp120 envelope protein (AIDSVAX B/E) (Rerks-Ngarm et al., 2009). The vaccine showed a 31.2% efficacy but it did not have any effect on early viral load or the CD4 T cell count in HIV-1 infected patients. However, the study may provide new insight into the development of vaccines (Nossal, 2011). Another field in vaccine research is the production of neutralizing antibodies (NAbs). These studies mainly focus on antibodies against the viral envelope protein gp120. One prototype of a NAb called VRC01 has been shown to neutralize 90% of highly diverse HIV isolates (Walker and Burton, 2010, Wu et al., 2010, Zhou et al., 2010). A development of an efficient vaccine will need more knowledge about the viral life cycle as well as the immune response during an HIV-1 infection.

**HIV-1 genomic organization**

The genome of HIV-1 is approximately 9 kilobases (kb) in size and contains the genes: gag-pol, tat, rev, nef, vpr, vpu, vif and env (Fig. 1). As will be described later, the HIV-1 genome can produce mRNAs that are differently spliced. These mRNAs can be classified into unspliced (9 kb), partially spliced (4 kb) and fully spliced (2 kb). In the 5´- and 3´-ends are the Long Terminal Repeats (LTRs) that contain sequences for transcription and polyadenylation.
Figure 1. Schematic representation of the HIV-1 genome. Boxes represent open reading frames, circles indicate splice donors and triangles indicate splice acceptors. Black bars represent the different exons (Schwartz et al., 1990a).

Gag-Pol

The gag-pol gene encodes two polyprotein precursors, Gag and Gag-Pol. The Gag polyprotein gives rise to the matrix (MA), capsid protein (CA), nucleocapsid protein (NC) and p6, and the Gag-Pol polyprotein gives rise to protease (PR), reverse transcriptase (RT) and integrase (IN). The gag-pol mRNA is the same molecule as the pre-mRNA and may also be incorporated into the virion and serve as genome (Sierra et al., 2005).

Tat

The tat gene encodes the Tat protein that is essential for viral replication. Transcription from the viral promoter LTR is rather inefficient but in the presence of Tat, transcription is enhanced about 100-fold. Tat binds to an RNA hairpin in the 5’-end of the transcript called the transactivation re-
sponse region (TAR). The increase of transcription is probably due to enhanced phosphorylation of RNA polymerase II C-terminal domain (CTD), which increases the elongation of RNA transcription (Frankel and Young, 1998). The increase in elongation is due to Tat association with the elongation factor P-TEFb, and recruitment of P-TEFb to the viral promoter increases the elongation (Zhu et al., 1997). Tat may also be involved in regulation of viral splicing, see the section HIV-1 pre-mRNA splicing.

Rev

The Rev protein is translated from a fully spliced (2 kb) mRNA. After translation, Rev enters the nucleus where it binds to the Rev Response Element (RRE) on the partially spliced (4 kb) and unspliced (9 kb) HIV-1 mRNAs and thereby facilitates their export to the cytoplasm by interacting with Crm1 and other cellular factors (Nekhai and Jeang, 2006, Pollard and Malim, 1998).

Nef

The name of the Nef protein comes from negative factor because it was originally thought of as negatively affecting viral transcription but Nef has been found to enhance the infectivity of HIV-1. The enhancement is caused by down regulating the number of CD4 receptors on the cell surface to facilitate the release of new virions. Nef can also down regulate Major Histocompatibility Complex type I (MHC I), to prevent detection by the cytotoxic T cells (Benson et al., 1993, Schwartz et al., 1996).

Env

The env gene encodes the proteins gp120 (SU) and gp41 (TM) that sits in the envelope membrane. They are produced through cleavage of the Env polyprotein gp160 by cellular proteases. SU and TM are involved in entry and fusion between viral and cellular membranes. gp120 interacts with the CD4 receptor on T cells and macrophages and thereafter undergoes a conformational change that facilitates the interaction between gp120 and its co-receptor (CXCR4 and CCR5). Thereafter, gp41 acts as a fusion protein, mediating viral entry (Freed, 2001).

Vpu

The integral membrane viral protein U (Vpu) induces degradation of the CD4 receptor in the endoplasmic reticulum (ER). Vpu can also facilitate virion release (Bolduan et al., 2011, Bour and Strebel, 2003, Margottin et al., 1998, Schubert et al., 1998, Willey et al., 1992).
Vif

The viral infectivity factor (Vif) is required for infectivity of HIV-1 by inducing degradation of the innate antiviral response APOBEC3G and inhibition of its virion incorporation. Without Vif, the APOBEC3 enzymes deaminate C to U residues in the HIV-1 minus strand DNA produced by RT, causing G to A mutations in the proviral DNA reducing the infectivity of the virus (Gabuzda et al., 1992, Lecossier et al., 2003, Marin et al., 2003, Sheehy et al., 2002). It has also been reported that Vif is present in the virion and that it facilitates the reverse transcription process (Cancio et al., 2004, Carr et al., 2008).

Vpr

The viral protein R (Vpr) is a 96 amino acid, 14 kDa protein. It was isolated in 1990 and is highly conserved (Cohen et al., 1990, Planelles et al., 1996, Yuan et al., 1990). Vpr facilitates infection by mediating the nuclear import of HIV-1 DNA in non-dividing cells. This is achieved by Vpr functioning as a component of the pre-integration complex that interacts with the nuclear pore complex (Le Rouzic et al., 2002, Nie et al., 1998). Vpr also induces cell cycle arrest in the G2 phase by promoting phosphorylation of mitosis checkpoint proteins. This may result in apoptosis of infected cells and transactivation of the viral promoter LTR (Dehart and Planelles, 2008, He et al., 1995, Roshal et al., 2001). Arrest in G2 could also be caused by Vpr-induced breaks in the nuclear lamina structure, causing the cell nuclei to rapture (de Noronha et al., 2001). Vpr can also function as a transactivator of LTR by interacting with transcription factor IIB (TFIIB) and by facilitating the activation of LTR by p300 (Agostini et al., 1996, Felzien et al., 1998). In addition, there are several reports that suggest that Vpr aids in evasion of the immune response. This appears to occur by induction of cell death in immune system cells (Kogan and Rappaport, 2011).

Virion structure and infectious cycle

HIV-1 is an enveloped virus that carries two copies of the genomic RNA (+) strand within its virion. The RNA molecules are coated with nucleocapsid protein (NC). The virion also contains 50 to 100 molecules of reverse transcriptase (RT), 50 to 100 copies of integrase (IN), protease (PR) as well as cellular RNAs, e.g. tRNA. Viral proteins Vif and Vpr are also present in the virion. The capsid protein (CA) together with the matrix protein (MA) constitutes the viral building blocks (Flint and American Society for Microbiology., 2009) (Fig. 2).
Attachment and entry
HIV-1 enters the cell after binding with the viral gp120 envelope protein to
the cellular CD4 receptor present on T cells, monocytes, macrophages and
dendritic cells (Fig. 3). The viral protein gp120 thereafter interacts with the
co-receptor CXCR4 or CCR5. This leads to fusion, mediated by gp41, be-
tween the viral envelope and the plasma membrane. The viral core, consist-
ing of viral RNA genome (coated with NC protein), integrase (IN), reverse
transcriptase (RT), matrix (MA) and Vpr is released into the cytoplasm of
the cell (Frankel and Young, 1998, Hernandez et al., 1996).

Reverse transcription and integration
In the pre-integration complex, RT reverse transcribes the viral RNA mol-
ecules. This process is initiated by a cellular tRNA\textsuperscript{Lys} primer and results in (−)
DNA in a 3’ to 5’ direction. Following transcription, the RNase H of RT
degrades the template RNA. The (+) DNA strand is transcribed using the
first DNA strand as a template and viral RNA as a primer. The DNA toget-
er with IN and other proteins are called the pre-integration complex. This
complex is localized to the nucleus by Vpr and imported via the nuclear pore
complex, a mechanism that is still rather unclear. In the nucleus, the viral
protein IN catalyzes the insertion of viral DNA into the host cell genome.
Nicking of the two viral DNA-ends starts the integration. The new 3’-ends
are then joined to the target DNA. Cellular enzymes repair the resulting gaps
within the DNA. When integrated into the host genome the viral DNA is referred to as the provirus and behaves as a cellular gene (Freed, 2001, Turner and Summers, 1999).

**Transcription, pre-mRNA processing and translation**

Transcription of the HIV-1 genome is carried out by RNA polymerase II and is enhanced by the viral protein Tat. The promoter is positioned in the LTR region that contains the three regions U3, R and U5. The promoter and enhancer sequences are located in U3 and transcription starts from the R region. The cellular machinery initially transcribes at a low level. Transcription of the HIV-1 genome is dependent on cellular factors, especially NF-κb, which in T cells is only active in activated cells. Transcription may be enhanced as much as 100-fold by the viral protein Tat that recognizes a secondary structure in the 5’-end of the RNA, called transactivation response (TAR). Tat interacts with cellular factors, which leads to phosphorylation of the C-terminal domain of RNA Polymerase II, which stimulates the transcription by enhancing the elongation (Frankel and Young, 1998, Zhu et al., 1997).

The primary, full-length transcript encodes Gag and Gag-Pol proteins and also serves as the viral genome in newly synthesized virions. The other HIV-1 proteins are produced from mRNAs that have been alternatively spliced from the full-length transcript. The only mRNAs that are immediately exported to the cytoplasm are the fully spliced 2 kb mRNAs. The Tat and the Nef proteins are products of the 2 kb mRNAs. Another protein produced from 2 kb mRNAs is Rev, which facilitates the export of the full-length transcript and the 4 kb mRNAs. Full-length viral mRNAs are translated into Gag and Gag-Pol polyproteins that localize to and assemble at the cell membrane together with the viral genomic RNA, Vpr, Vif and Nef. The 4 kb mRNAs encode for the Env polyprotein that is synthesized and modified in the endoplasmic reticulum (ER) and Golgi apparatus to produce gp120 and gp41 (Arya and Gallo, 1986, Feinberg et al., 1986, Nekhai and Jeang, 2006, Pollard and Malim, 1998, Schwartz et al., 1990a, Schwartz et al., 1990b).

**Assembly, budding and maturation**

The viral protein Vpu assists the transport of the envelope proteins to the cell membrane by promoting degradation of newly synthesized CD4 in the ER (Dube et al., 2010). The cell surface becomes coated with the envelope proteins, which are prevented from binding to cell surface CD4 receptors through degradation of CD4 mediated by the viral protein Nef (Frankel and Young, 1998). Gag molecules associate with the plasma membrane and the RNA genome and assembly initiates. The virus assembles and buds from the infected cell as a noninfectious, immature virus particle. After budding, the Gag-Pol polyprotein is self-cleaved into RT, IN and PR. The Gag polypro-
tein is cleaved by the protease into matrix, capsid, nucleoprotein and p6 and the virion becomes mature (Gottlinger, 2001).

Figure 3. Schematic representation of the HIV-1 infectious cycle. 1) Virus attachment to CD4 receptor, fusion is mediated by binding to co-receptor. 2) Virus entry and core release. 3) Following reverse transcription proviral DNA is imported to the nucleus and 4) integrated into the host genome. 5) Viral DNA is transcribed and spliced by cellular machinery. 6) Fully spliced mRNAs are transported to the cytoplasm and translated. 7) The viral protein Rev returns to the nucleus and 8) mediates the export of partially spliced and unspliced mRNAs, which can be translated in the cytoplasm. 9) Unspliced mRNA is transported to the cell membrane, where Env, Gag and Gag-Pol polyproteins have assembled, and incorporates into new virus particles as the viral genome. 10) The new virus particle is released and after proteolytic cleaving it becomes mature.

Post-transcriptional processing

DNA is transcribed into mRNA, which will later be translated into protein. Before translation can occur several different modifications of the pre-mRNA have to take place. These modifications occur co-transcriptionally and include i) the addition of a 7-methyl guanosine (m⁷G) cap at the 5’-end of the RNA ii) the production of a poly(A) tail by cleavage and polyadenylation of the 3’-end of the pre-mRNA iii) as well as excision of introns by the process called splicing (Beyer and Osheim, 1988, Maniatis and Reed, 2002, Tennyson et al., 1995). The 5’-cap and the 3’-poly(A) tail have diverse functions: They protect the mRNA from degradation and facilitate translation by easing the export of mRNA, and in the cytoplasm they promote translation.
(Ford et al., 1997, Furuichi et al., 1977, Wakiyama et al., 2000). This thesis focuses on HIV-1 pre-mRNA splicing. The following section therefore gives a detailed description of this process.

Pre-mRNA splicing

Splicing is the process where the exons in the pre-mRNA are joined together and the introns in between are excised. In 1993 Philip Sharp and Richard Roberts were awarded the Nobel Prize for their discovery of splicing.

The biochemical process of splicing can be divided into two transesterification steps (Fig. 4). The first step is when the phosphodiester bond between the first exon and the 5’-end of the downstream intron is broken resulting in a free exon and a lariat RNA intermediate, consisting of the intron and the second exon. In the second step the phosphodiester bond between the second exon and the 3’-end of the intron is broken, the intron is excised and the two exons are joined together. A prerequisite for the splicing reaction to occur is the assembly of the spliceosome, which is partly directed by the recognition of sequences in the 5’ splice site also called splice donor (5’-end of intron), 3’ splice site also called splice acceptor (3’-end of intron), a polypyrimidine tract and a branch point (Burge, 1999). The 5’ splice site consensus sequence is AG/GURAGU (intron part is underscored) and is a binding site for U1 snRNP. The 3’ splice site contains the conserved sequence YAG/N and a polypyrimidine tract (PPT) (Py) upstream. The 3’ splice site is the binding site for the cellular splicing factor U2AF, where the subunit U2AF65 binds to the PPT and the subunit U2AF35 binds to the AG sequence. Upstream of the PPT is the branch point, YNYURAC that is bound by the branch point binding protein SF1/mBBP (Graveley, 2000).
The initial step of splicing in mammalian cells is the exon definition, a process in which factors at the 5’ splice site and the 3’ splice site interact. This interaction causes the assembly of the spliceosome into the early splicing complex, called E complex (Hoffman and Grabowski, 1992, Robberson et al., 1990). Splicing is not only regulated by transacting factors, but also by cis-elements within the pre-mRNA. These elements are called Exonic splicing enhancers, ESEs, Exonic splicing silencers, ESSs, Intronic splicing enhancers, ISEs, and Intronic splicing silencers, ISSs, which all recruit activators or repressors of splicing (Maniatis, 1991). Generally, many enhancer sequences recruit members of the serine-arginine rich (SR) protein family, which will be described further on, and most silencer sequences recruit members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family (Dreyfuss et al., 1993, Graveley, 2000, Pozzoli and Sironi, 2005), but this is by no means universally true. Splicing can be either constitutive, i.e. exons are always spliced or excised in the final mRNA, or alternative, i.e. exons
are joined together in multiple ways. Because HIV-1 uses alternative splicing the main focus in this thesis will be on alternative splicing, which will be explained further on.

**The spliceosome**

The spliceosome (i.e. U2-dependent spliceosome) (Fig. 5) contains four small nuclear ribonucleoprotein particles (snRNPs U1, U2, U4/U6 and U5) and several non-snRNPs like serine-arginine rich (SR) proteins. The U1 snRNP binds to the 5′ splice site, thereafter U2 snRNP auxiliary factor (U2AF) binds to the PPT and to the 3′-terminal AG, and splicing factor 1 (SF1) binds to the branch point. This leads to the formation of the E complex. These interactions promote the association of U2 snRNP with the branch point to form the ATP-dependent A complex. After U2 snRNP binding, U4/U6 snRNP and U5 snRNP association form the B complex. Further progress in the splicing reaction requires destabilization of the U4 snRNA carried out by RNA helicases. This step will create the catalytically active spliceosome, the C complex, which catalyzes the two chemical steps of splicing (Burge, 1999, Chen and Manley, 2009). There is also a minor class of spliceosome called the U12-dependent spliceosome but the work in this thesis is based on the U2-dependent spliceosome and the U12 will therefore not be presented any further.
Figure 5. Schematic representation of the splicing process. Splicing regulatory sequences (5’ splice site, 3’ splice site, branch-point and polypyrimidine tract) within the pre-mRNA are indicated in the upper panel. Boxes represent exons and thin line represents intron. The spliceosome assembly is indicated stepwise in the lower panel, from the formation of E complex to the formation of the catalytic C complex. Factors involved in the assembly are indicated: U2AF and U snRNPs (U1, U2, U4, U5 and U6). The final products in the process are mRNA and intron lariat, shown in the bottom of the panel.
Alternative splicing and splice site recognition

Most pre-mRNAs in higher eukaryotes contain multiple introns making it possible to create different mature mRNAs from the same pre-mRNA by alternative splicing patterns. Alternative splicing was first described in 1980 and since then we have come to understand the diversity of the expression from eukaryotic genes (Alt et al., 1980, Early et al., 1980). Sequencing analysis indicates that 92-94% of all human genes undergo alternative splicing (Wang et al., 2008). Two to several thousands of mRNAs can be produced from one single transcript by alternative splicing (Navaratnam et al., 1997, Rosenblatt et al., 1997, Schmucker et al., 2000). The effects of alternative splicing are versatile and can for example determine sex, as can be seen in probably the most studied system of alternative splicing, the splicing of RNA encoding the *Drosophila* Sex lethal protein. The RNA binding protein Sex lethal (Sxl) is expressed specifically in female flies and represses splicing patterns that would lead to development of male flies. The repression is due to Sxl preventing splicing factors from binding the RNA in a male splicing pattern. The Sxl transcript itself is alternatively spliced in a male or female manner, due to autoregulation (Baker, 1989). Other effects of alternative splicing can be seen in tissue-specific splicing, alternative splicing due to cell type or alternative splicing induced by external stimulus.

The different mechanisms for alternative splicing can be categorized into four different groups, alternative 5’ splice site selection, alternative 3’ splice site selection, cassette-exon inclusion or skipping and finally intron retention (Fig. 6) (Nilsen and Graveley, 2010). Alternative splicing is regulated by splice site recognition that is controlled by the strength of the splice site, the concentration of splicing factors and/or by the presence of splicing regulatory sequences within the pre-mRNA (Maniatis, 1991). Some splicing factors, like SR proteins, can facilitate splice site recognition by binding to splicing enhancer sequences and further recruit the splicing complex (Bourgeois et al., 1999, Graveley et al., 2001). There are several different ways in which inhibition of splice site recognition can occur. Factors binding to splicing silencers can sterically hinder other factors from binding to splice site recognition sequences, or even enhancer sequences, preventing further recruitment of the splicing complex. One example is hnRNP A1 binding to an intron in HIV-1 pre-mRNA that prevents binding of U2 snRNP to a downstream 3’ splice site (Tange et al., 2001). It has also been shown that there is a coupling between transcription and splicing and that this coupling could regulate splicing (Cheng et al., 2007, Kornblihtt, 2006, Lenasi et al., 2011, Smith et al., 2004). The outcome of a splicing event is therefore affected by the relative ratio of activators and inhibitors, the transcription event and the speed of elongation. Another possible influence of splicing is the secondary structure of the pre-mRNA and more studies are needed to further examine the effect of the secondary structure (Buratti and Baralle, 2004).
One type of splicing factors that are involved in alternative splicing is the family of serine-arginine rich proteins (SR proteins) and another type is the family of heterogeneous nuclear ribonucleoproteins (hnRNPs).

**Figure 6.** The different categories of alternative splicing. Boxes represent exons, thick lines represent introns and thin lines represent alternative splicing patterns. Adapted/modified from Nielsen et al 2010. Left panel shows pre-mRNAs, right panel shows spliced mRNAs.

**SR proteins**

The family of human serine-arginine rich (SR) proteins is phylogenetically conserved and currently contains 12 known members: SRp20, SRp30a-c, SRp38, SRp40, SRp46, SRp54, SRp55, SRp75, 9G8 and SRSF12. These proteins contain one or two RNA recognition motifs (RRM) at their N-
termini and a serine-arginine (RS) rich domain at their C-termini. The RRM(s) interacts in a sequence specific manner with the pre-mRNA, although with relatively low specificity. The RS domain functions as a protein interaction domain, especially with other RS-domain-containing splicing factors. The nomenclature for SR proteins has recently been changed into SR splicing factor (SRSF) and the new names are listed in Table 1 (Manley and Krainer, 2010).

Table 1. SR protein/gene names

<table>
<thead>
<tr>
<th>New protein/gene name</th>
<th>Previous name</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRSF1</td>
<td>ASF/SF2</td>
</tr>
<tr>
<td>SRSF2</td>
<td>SC35</td>
</tr>
<tr>
<td>SRSF3</td>
<td>SRp20</td>
</tr>
<tr>
<td>SRSF4</td>
<td>SRp75</td>
</tr>
<tr>
<td>SRSF5</td>
<td>SRp40</td>
</tr>
<tr>
<td>SRSF6</td>
<td>SRp55</td>
</tr>
<tr>
<td>SRSF7</td>
<td>9G8</td>
</tr>
<tr>
<td>SRSF8</td>
<td>SRp46</td>
</tr>
<tr>
<td>SRSF9</td>
<td>SRp30c</td>
</tr>
<tr>
<td>SRSF10</td>
<td>SRp38</td>
</tr>
<tr>
<td>SRSF11</td>
<td>SRp54</td>
</tr>
<tr>
<td>SRSF12</td>
<td>SRp35</td>
</tr>
</tbody>
</table>

**SR proteins as regulators of splicing**

SR proteins are involved in the regulation of splicing and can function both early and late in the splicing process. In constitutive splicing it has been suggested that they participate in the exon-definition process and a number of SR protein binding sites have been identified in constitutive exons (Graveley, 2000). SR proteins also play a role in alternative splicing: By binding to exonic splicing enhancers they can stimulate the usage of a 3′ splice site either by recruiting members of the splicing complex or by counteracting complexes that are inhibiting splicing (Kan and Green, 1999, Zuo and Maniatis, 1996). SR proteins bound to upstream splicing enhancers can stimulate the usage of the downstream 5′ splice site potentially by stimulating the binding of U1 snRNP to the splice site. SR proteins can also bind splicing silencer elements and thereby inhibit splicing by preventing recruitment of the spliceosome. This has for instance been shown in splicing regulation of adenovirus pre-mRNA. In this case, SR protein ASF/SF2 binds to a splicing silencer and prevents U2AF binding to the 3′ splice site (Kanopka et al., 1996).

**Other functions of SR proteins**

SR proteins participate in other RNA processing events such as targeting mRNAs with premature stop codons to the nonsense-mediated decay path-
way. Some SR proteins shuttle between the nucleus and the cytoplasm, indicating a function also in mRNA export. In fact, SRp20, 9G8 and ASF/SF2 have been shown to be involved in nuclear export of mRNA by interacting with TAP. Another example is the nuclear export of intronless histone H2A mRNA that is promoted by SRp20 and 9G8 (Hargous et al., 2006, Huang et al., 2003, Huang and Steitz, 2001, Huang et al., 2004). SR proteins can in addition stimulate translation, e.g. IRES-mediated translation by SRp20 (Bedard et al., 2007).

Phosphorylation determines the function of SR proteins

The serines in the RS domain can be phosphorylated and the state of phosphorylation determines the function of the SR proteins (Graveley, 2000, Kanopka et al., 1998, Xiao and Manley, 1998). The RS domain is phosphorylated by e.g. Clk/Sty kinase,cdc34, SR protein kinase 1 (SRPK1) and topoisomerase (Colwill et al., 1996, Mattaj, 1994, Okamoto et al., 1998, Soret and Tazi, 2003). It has been shown that phosphorylation is needed for efficient splice-site recognition, and that dephosphorylation is needed for splicing catalysis (Mermoud et al., 1992, Mermoud et al., 1994). It also seems like the dephosphorylated state more efficiently interacts with TAP and allows the proteins to shuttle from the nucleus to the cytoplasm. In the cytoplasm SR proteins are detached from the mRNA complex, followed by phosphorylation of the RS domain which mediates transport back to the nucleus (Caceres et al., 1997, Kataoka et al., 1999, Lai et al., 2000, Lai and Tarn, 2004, Swartz et al., 2007).

SRp55

One of the SR proteins, SRp55, is of particular interest in this thesis, and will be described in some detail. SRp55, or SFRS6 (or SRp55/B52 in Drosophila) as it is also called, is a serine-arginine rich protein with two RRMs. As for many SR proteins, the function of SRp55 is still not fully understood. But the protein has been reported to be involved in splicing regulation of several different transcripts. For instance the proinflammatory protein TF where the unspliced form is membrane bound and the spliced (promoted by SRp55) is soluble (Tardos et al., 2008). Another splicing event regulated by SRp55 is the calcitonin/CGRP RNA, where SRp55 regulates the splicing in a tissue specific manner (Tran and Roesser, 2003). Splicing regulation mediated by SRp55 has also been shown to act in an antagonistic way, competing with other SR proteins (Chandradas et al., 2010). Other studies have reported that the Drosophila protein SRp55/B52 targets Topoisomerase I to sites of transcription (Juge et al., 2010). SRp55/B52 has also been shown to play a role in Drosophila cell cycle control (Rasheva et al., 2006). SRp55 (SFRS6) can shuttle between the nucleus and the cytoplasm (Bjork et al., 2009, Sapra et al., 2009). It has also recently been shown that SRp55 enhances translation of unspliced HIV-1 mRNA (Swanson et al., 2010).
HIV-1 pre-mRNA splicing

In order to produce all the mRNAs needed for HIV-1 to be infectious, the virus utilizes alternative splicing (Arya and Gallo, 1986, Feinberg et al., 1986, Schwartz et al., 1990a, Schwartz et al., 1990b). HIV-1 produces one single transcript (9 kb) that apart from functioning as genomic RNA also encodes for the Gag and Gag-Pol proteins. This unspliced transcript can also be spliced into mRNAs within the 4 kb class or the 2 kb class (Fig. 7). In the 9 kb transcript there are several splice sites present: four 5′ splice sites, here referred to as splice donors (SD1-4), and nine 3′ splice sites, here referred to as splice acceptors (SA1-4, 4a-c, 5 and 7). The splice sites can all be used in different combinations to create more than 35 different mRNAs. The efficiencies of the splice donors depend on their complementarities to U1 snRNA while the splice acceptors are more depending on cis-regulatory elements present in the exons or the introns (O'Reilly et al., 1995). In order to produce the correct amounts of all mRNAs, the splice acceptors in HIV-1 are sub-optimal because of short and interrupted polypyrimidine tracts, branch points that are not canonical and inhibitory sequences that up or down regulate the usage of splice sites (Bilodeau et al., 2001, Dyhr-Mikkelsen and Kjems, 1995, Jacquenet et al., 2005, Marchand et al., 2002, O'Reilly et al., 1995). This allows temporal regulation by enhancers and silencers.

HIV-1 mRNAs

The mRNAs produced from HIV-1 can be divided into three classes: 9 kb, 4 kb and 2 kb. The full-length unspliced 9 kb transcript can (i) encode for the Gag and Gag-Pol proteins (ii) be incorporated into the virion as a genome (iii) be spliced into 4 kb and 2 kb class of mRNAs. The 4 kb class (partially spliced) includes the mRNAs encoding for Env, Vif, Vpr and Vpu and the 2 kb class (fully spliced) includes the mRNAs encoding for Tat, Rev Vif, Vpr and Nef. mRNAs that are present in high abundance are env and nef, while mRNAs like tat, vif and vpr are present in relatively low amounts. These differences are most likely due to differences in the splicing efficiency of the respective splice site (Purcell and Martin, 1993).
Viral proteins affecting splicing

Early in the HIV-1 infection only the proteins Tat, Rev and Nef are produced because the mRNAs encoding these proteins are the only ones completely spliced and therefore the only ones exported to the cytoplasm. Rev is an 18 kDa protein with a nuclear localization signal. Rev goes to the nucleus and interacts with a structured region in the env gene, the Rev response element (RRE), present in the unspliced and partially spliced mRNAs. Rev also interacts with the cellular protein Crml and thereby facilitates the export of the 4 kb and the 9 kb mRNAs to the cytoplasm. mRNAs that would otherwise be retained in the nucleus due to the presence of unused splice sites and cis-acting RNA elements (Chang and Sharp, 1990, Dreyfuss et al., 2002, Rosen et al., 1988, Schwartz et al., 1992). The export of incompletely spliced mRNAs mediated by Rev can compete with the splicing process in a time-dependent manner (Nekhai and Jeang, 2006, Pollard and Malim, 1998).

Tat has been shown to interact with the cellular protein p32, an interaction that is increased when Tat is acetylated. Tat has additionally been shown to interact with CDK13, a cellular factor that phosphorylates ASF/SF2, in a complex with p32. Phosphorylation of ASF/SF2 leads to an increase in splicing of viral RNA at the expense of unspliced RNA. Acetylated Tat has been shown to accumulate as the infection proceeds, and as a consequence this increases the amount of unspliced HIV-1 RNA at the expense of spliced viral RNA (Berro et al., 2006, Berro et al., 2008). Another study indicates that Tat can regulate splicing by interacting with cellular factors in a complex that recruits SR proteins AFS/SF2 and SC35 to the sequence in exon 5 called guanosine-adenosine-rich (GAR). This increases the expression of env mRNA (Jablonski et al., 2010).
Silencers and enhancers on the HIV-1 pre-mRNA

Several elements have been found \textit{(in vitro and in vivo)} to function as enhancers or silencers for HIV-1 pre-mRNA splicing. Some of those will be described in the following sections (Tazi et al., 2010). The splice sites are indicated in Fig. 8.

\textbf{Figure 8.} The HIV-1 splice sites. Boxes represent open reading frames, circles indicate splice donors and triangles indicate splice acceptors. Numbers of splice donors (SD) and splice acceptors (SA) are indicated.

\textit{Splice acceptor 2 and splice donor 2}

Splice acceptor 2 is regulated by a splicing enhancer within exon 2 that facilitates inclusion of exon 2, producing vif mRNA. This enhancer binds to SRp75. A splicing silencer controlling splice donor 2 is positioned downstream of exon 2 and down regulates the expression of vif mRNA (Exline et al., 2008).

\textit{Splice acceptor 3 and splice donor 3}

Splicing into splice acceptor 3 can result in vpr mRNA. A silencer positioned within exon 3 is called ESSV and inhibits the vpr splice acceptor SA3. This is due to hnRNP A/B binding to the silencer and antagonizing binding of U2AF65 to the viral polypyrrimidine tract (Bilodeau et al., 2001, Domsic et al., 2003). It has also been reported that splicing into SA3 is activated by ASF/SF2 and that this is in competition with hnRNP A/B (Ropers et al., 2004, Zhu et al., 2001). In our paper from 2010 (Paper I) we show that SRp55 inhibits SD3 by interacting with an RNA element in exon 5 (Tranell et al., 2010).

\textit{Splice acceptor 4}

Splicing into SA4 will result in tat mRNA production. Within exon 4 there are two silencers, ESS2 and ESS2p, both inhibiting splicing into exon 4 and
thereby the production of tat mRNA – the ESS2 by binding hnRNP A1 and the ESSp2 by binding hnRNP H (Amendt et al., 1994, Jacquenet et al., 2001). An enhancer within exon 4 has been named exonic splicing enhancer 2 (ESE2) because it attenuates ESS2 and thereby enhances splicing into SA4. This enhancer interacts with the SR protein SC35 (Zahler et al., 2004).

**Splice acceptor 5 and splice donor 5**

Another enhancer, named GAR, is positioned within exon 5 and has a positive effect on the usage of SA5, the nef splice acceptor, and SD5, used to produce all 2 kb mRNAs. GAR interacts with SRp40 and ASF/SF2 (Caputi et al., 2004, Kammler et al., 2001). Our results (Paper I) show that SRp55 interacts with GAR and inhibits SD3 (Tranell et al., 2010).

**Splice acceptor 7**

SA7 is used to produce all mRNAs in the 2 kb class. An intronic silencer (ISS) found to inhibit SA7 is located within the intron upstream of SA7 (Tange et al., 2001). The exonic splicing silencer ESS3 suppresses SA7 by blocking U2 snRNP. Both ISS and ESS3 interact with hnRNP A1 (Amendt et al., 1995). Close to ESS3 within exon 7 there is an enhancer element called ESE3 that activates SA7 that can also function as a silencer when binding to ASF/SF2 (Staffa and Cochrane, 1995, Zhu et al., 2001).

**SR proteins and HIV-1 pre-mRNA splicing**

SR proteins have been found to affect HIV-1 pre-mRNA splicing: SC35, 9G8 and SRp40 increase the amounts of tat mRNA. SC35 and SRp40 by binding to sequences in exon 4, which probably counteract the splicing inhibitory hnRNP A1 binding to the pre-mRNA (Hallay et al., 2006, Ropers et al., 2004). siRNA knockdowns of SR proteins Srp40 and SC35 in fact decreased the level of tat mRNA (Jablonski and Caputi, 2009). ASF/SF2 has been found to up-regulate the HIV-1 vpr mRNA and to increase splicing at SA2 and SA3 (Jacquenet et al., 2005, Ropers et al., 2004). The levels of Nef protein have been shown to decrease upon over expression of SC35, SRp40 and ASF/SF2 (Ropers et al., 2004).

**SR proteins and HIV-1 mRNA biogenesis**

Apart from the involvement in splicing of HIV-1 pre-mRNA, SR proteins have been reported to have other functions in HIV-1 mRNA biogenesis. For instance in our paper from 2011 (Paper II) there are indications that SR proteins can stabilize and increase nuclear export of viral mRNAs (Tranell et al., 2011). Another possible function for SR proteins is to facilitate translation of the viral unspliced mRNA encoding Gag and Gag-Pol protein. Unspliced mRNAs are not associated with exon junction complexes that are needed for an mRNA to be translated. Translation of the unspliced gag
mRNA must therefore be regulated in a different manner. This regulation is still unknown but there are reports indicating this could take place with the help of SR proteins (Swanson et al., 2010).

**Antiviral therapy and HIV-1 splicing**

A new approach to fight HIV-1 infection could be to interfere with the essential splicing, either by targeting a splice site or by targeting splicing factors. Researchers have used antisense molecules against HIV-1 splicing elements. One example is the antisense sequences targeted against splice donor 5 and the GAR sequences, which have been shown to induce exon skipping and thereby inhibit tat and rev mRNA splicing (Asparuhova et al., 2007). Another way of interfering is to target splicing factors, indirectly or directly. An indirect way would be to target factors that regulate splicing proteins e.g. kinases that phosphorylates SR proteins or using modified U1 snRNA to up or down regulate the usage of specific HIV-1 splice sites. A more direct way has been studied using small molecules called indole derivatives that can target SR proteins (Bakkour et al., 2007, Mandal et al., 2010, Soret et al., 2005, Wong et al., 2011).

**Indole derivatives**

Indole derivatives are small molecules that can interact with SR proteins and alter or inhibit their function in pre-mRNA splicing regulation. The most plausible mechanism to explain how they interfere with the splicing regulation is by changing the phosphorylation status of the SR proteins. Indole derivatives are potent inhibitors or re-directors of HIV-1 RNA production in chronically infected cells where they interact with SR proteins. One specific indole derivative called IDC16 was reported to block HIV-1 viral production in infected PBMCs or macrophages by interacting with ASF/SF2 (Bakkour et al., 2007, Fukuhara et al., 2006, Soret et al., 2005).
The present investigation

Aims of the study

In order to produce all the mRNAs needed for HIV-1 to be infectious the virus uses alternative splicing. From one single transcript more than 35 differently spliced mRNAs can be produced. This alternative splicing is highly regulated in order to generate the correct amount of each mRNA. Therefore more and more research is being done within this field. Splicing factors or cis-sequences within the HIV-1 RNA affecting the splicing pattern in HIV-1 are of great interest when designing new potential antiviral agents. We have studied the splicing regulation in the hope that this knowledge could be used to design new antiviral treatments. The first step in gaining this understanding is to identify regulatory elements or factors that are important for HIV-1 pre-mRNA splicing. We have identified cis-elements in the HIV-1 genome that are required for HIV-1 splicing regulation. We have also identified cellular factors that regulate HIV-1 mRNA splicing and we have studied how these factors affect the HIV-1 splicing pattern.

Experimental system

To study regulation of HIV-1 pre-mRNA splicing we used the sub-genomic HIV-1 plasmid pDP (Fig. 9), constructed in our lab. This plasmid contains the entire HIV-1 genome with the exception of a 2 kb deletion within the pol gene, which renders the genome non-infectious. The natural HIV-1 LTR promoter drives the transcription from this plasmid. We also used sub-genomic plasmids and cDNAs with different deletions and mutations to find the regions of importance for HIV-1 pre-mRNA splicing regulation.

For studies in cells the plasmids were transfected into cervical cancer cells (HeLa) and the splicing pattern was analyzed by northern blot analysis or RT-PCR. To identify factors that are functionally involved in HIV-1 pre-mRNA splicing, plasmids expressing the factors of interest were over-expressed in co-transfection experiments with sub-genomic HIV-1 plasmids in HeLa cells. For studies in cell lines containing integrated HIV-1 genome (HLfB cells) SR protein expression plasmids were transfected into these cells. The amount of HIV-1 p24gag protein in cell culture media was determined by p24gag antigen capture ELISA.
Figure 9. The HIV-1 sub-genomic plasmid pDP, containing a 2 kb deletion in the pol gene, and the mRNA classes it produces. Boxes represent open reading frames, triangles represent splice acceptors and circles represent splice donors. Thick lines represent exons and thin lines represent spliced introns.

Results

Paper I

Sera- and Arginine-rich Proteins 55 and 75 (SRp55 and SRp75) Induce Production of HIV-1 vpr mRNA by Inhibiting the 5'-Splice site of Exon 3

HIV-1 uses alternative splicing in order to produce all the mRNAs that are needed for the virus to be infectious. From one single transcript more than 35 differently spliced mRNAs are produced. Alternative splicing is highly regulated and in this study we tested the effect of the splicing factors SR proteins on HIV-1 pre-mRNA splicing. To study this we used sub-genomic HIV-1 expression plasmids. The plasmids were transfected into HeLa cells in the absence or presence of SR protein expression plasmids. Cytoplasmic RNA was extracted 24 hours post-transfection and the splicing pattern of HIV-1 was analyzed with northern blotting and RT-PCR.

The findings in this paper showed that the intracellular concentration of SR proteins is of great importance for HIV-1 to be able to produce the correct amounts of mRNAs. Variations in concentrations of SR proteins lead to great changes in HIV-1 pre-mRNA splicing pattern.

HIV-1 vpr mRNA 13a7 is a partially spliced mRNA because it contains an intron between exon 3 and exon 4. Incompletely spliced mRNAs are normally trapped in the nucleus where they are degraded. This would happen to the gag-pol and env mRNAs unless Rev would bind to RRE and aid in the
export of these mRNAs. In contrast, Rev does not bind to the vpr mRNA 13a7 that also contains an intron and it is therefore unclear how these mRNAs are made and how they reach the cytoplasm (Fig. 10). We found that SRp55 and SRp75 induced the production of HIV-1 tat mRNAs 147 and 14E as well as vpr mRNA 13a7.

Figure 10. Indication of Rev response element (RRE). The HIV-1 genome at the top. Boxes represent open reading frames, circles indicate splice donors and triangles indicate splice acceptors. Thick lines represent exons present in mRNA and thin lines represent spliced introns. The RRE is indicated on the 9 kb and the 4 kb mRNA classes. The 2 kb class and the vpr 13a7 mRNA does not contain the RRE and is therefore not exported to the cytoplasm with the help of Rev.

We wanted to study further the SRp55 regulation of vpr mRNA. By introducing mutations in the HIV-1 genome, we found that induction of vpr mRNA due to over expression of SRp55 was dependent on a previously identified sequence in exon 5 named GAR. We could also show that SRp55 interacts directly with the GAR RNA. The interaction inhibited the function of GAR as a splicing enhancer. More specifically, SRp55 inhibited SD3 in exon 3 and thereby induced vpr mRNA 13a7. We also concluded that SRp55 interaction with GAR is in competition with SRp40 and that the production of vpr mRNA is determined by the relative concentrations of SRp40 versus SRp55 (or SRp75). This competition is more complex than simply competing for binding because inhibition of SD3 requires not only the RRM of SRp55 but also the RS domain. This indicates that inhibition requires SRp55 interaction with an additional, still unidentified, factor.
Inhibition of Splicing by Serine-arginine Rich Protein 55 (SRp55) Causes the Appearance of Partially Spliced HIV-1 mRNAs in the Cytoplasm

In paper I we showed that SRp55 inhibits SD3 and thereby induces the production of 13a7 vpr mRNA. HIV-1 Vpr protein is incorporated into the virion and into the viral pre-integration complex. The functions of Vpr are diverse and it has been shown to stimulate transcription from the HIV-1 LTR promoter, affect the reverse transcription process, cause G2 cell cycle arrest, induce apoptosis and suppress immune response. These mechanisms require production of vpr both early and late in the viral life cycle. We therefore wanted to further study the regulation of HIV-1 mRNA processing by SRp55.

According to our results the majority of mRNAs in the nucleus produced from 13a7 cDNA plasmid pNL13a7 were unspliced 13a7 mRNA and a minority was the fully spliced 1357 mRNA. This showed that unspliced 13a7 mRNA was trapped in the nucleus. We have previously shown that SRp55 increases the production of 13a7 vpr mRNA by inhibiting SD3. And in this paper we concluded that this inhibition caused an increase in the amount of unspliced 13a7 vpr mRNA in the cytoplasm, due to an increased nuclear export of this mRNA.

Further on, we wanted to see if SRp55 could also induce the production of partially spliced 4 kb mRNAs that are dependent on the Rev protein and produced late in the viral life cycle, and more specifically vpr mRNA 13aE. We showed that SRp55 inhibited splicing of viral mRNAs spliced at SD5 suggesting that SRp55 inhibits SD5. This inhibition resulted in the appearance of partially spliced mRNAs vpu, env and vpr (13aE) in the cytoplasm. The increase in cytoplasmic amounts of these mRNAs also indicated an increase in mRNA nuclear export in the presence of SRp55.

The viral protein Rev is produced from a fully spliced mRNA and the protein associates with partially spliced and unspliced viral mRNAs and facilitates their nuclear export. Because intron-containing mRNAs are considered immature, these mRNAs survive with Rev, where they would otherwise be trapped in the nucleus and degraded. We studied whether SRp55 and SRp75 could increase the amount of partially spliced mRNAs in the cytoplasm of a cell line containing integrated copies of a Rev mutant virus and therefore is unable to produce infectious virus particles. The results showed that transfection of SRp55 and SRp75 into these cells increased the amount of p24-Gag, a viral protein present in HIV-1 virions and in the cell culture media. This protein is produced from an unspliced mRNA. The increase of p24-Gag was nevertheless as great as in Rev mediated export of mRNA.
The results from the transfection experiment together with the Rev mutant virus experiments suggested that high levels of SRp55 can replace the function of Rev in an early stage of infection, before the production of Rev has reached higher levels. It has previously been reported that Vpr has an effect on transcription of HIV-1 genome and that Vpr can suppress the immune system during an HIV-1 infection. High levels of SRp55 during early infection will lead to early production of Vpr and more efficient transcription. Vpr could therefore have important functions initially during an HIV-1 infection.

The increased nuclear export of vpr mRNA as could be seen during over expression of SRp55 could be due to two things: on the one hand SRp55 could mediate transport by binding mRNA and nuclear pore complex, on the other hand SRp55 could interfere with the interaction between partially spliced mRNA and splicing factors and thereby release vpr mRNA for export. Some SR proteins have previously been shown to interact with nuclear export factor TAP/Nxf1. Some, e.g. SRp55, can shuttle between the nucleus and the cytoplasm.

Paper III

**Indole-related Compounds from the National Cancer Institute (USA) Repository do not Inhibit HIV-1 Replication in Primary Blood Mononuclear Cells**

Indole derived substances have previously been shown to inhibit HIV-1 replication. We tested the effect of indole-related substances on HIV-1 gene expression in transient transfections and in primary cell cultures infected with HIV-1. For transfection assays we used a sub-genomic HIV-1 expression plasmid called pDP. This plasmid was transfected into HeLa cells in the absence or presence of 10 µM of each compound. The levels of HIV-1 mRNAs in cytoplasmic RNA extracts 24 hours post-transfection were determined by northern blot. HIV-1 gag protein levels were determined by western blot with proteins extracted 24 hours post-transfections.

In our study some of these substances had an inhibitory effect on HIV-1 gene expression in cells transiently transfected with the subgenomic plasmid pDP, which could be seen on northern blotting and western immunoblotting. For studies on HIV-1 infected cells we used primary peripheral blood mononuclear cells (PBMCs). To measure virus particles, the levels of extra cellular HIV-1 p24 gag protein were used in ELISA. The results showed that some of the substances were toxic to the cells, and were therefore not tested further. None of the substances tested inhibited replication of HIV-1 in primary cells. Although the promising results obtained in transfection studies, inhibition of HIV-1 gene expression was not seen in infected cells. These
results demonstrated important differences between transfection assays performed in cell lines and infection assays in primary cells.
Concluding remarks and future prospects

At the end of 2010, 34 million people were living with HIV-1 infection. HIV-1 is the causative agent of AIDS, which causes millions of deaths every year. There are several antivirals against HIV but due to the virus high mutability, causing resistance, there is a constant need for new antivirals. In order to develop new strategies of treatment and prevention, we need to know more about the virus and the way it behaves in our cells.

The splicing process of HIV-1 pre-mRNAs is highly regulated and is essential for the virus to be infectious. This process is a potential target in development of new treatments. It is therefore of great interest to understand how the splicing process is regulated for HIV-1 pre-mRNAs. In this thesis I have discussed the regulation of HIV-1 pre-mRNA splicing. We identify regulatory elements associated with splicing within the viral genome as well as cellular splicing factors that affect the HIV-1 pre-mRNA splicing pattern.

More specifically, we shed some light on how the HIV-1 vpr mRNA is regulated. The cellular protein SRp55 and SRp75 regulate the splicing of HIV-1 pre-mRNA by inhibiting splice donor 3 (SD3), leading to induction of HIV-1 vpr mRNA. We also discover that the inhibition caused by SRp55 is due to its interaction with an RNA element called GAR, located in the viral exon 5. According to our results the regulation of vpr mRNA is dependent on the relative concentrations of different SR proteins. We also show specifically that there is competition between SRp40 and SRp55 in the regulation of tat versus vpr mRNA, due to competitive interaction with GAR between these two SR proteins. This competition is not just due to steric hindrance, but requires other factors that are still unknown. It would be interesting to further study the mechanism of SRp55 splicing inhibition of SD3. What other factors are required for this inhibition? This could for instance be tested in vitro with immunoprecipitation by mixing vpr mRNA with tagged SRp55 and adding nuclear extract and also by fishing up tagged SRp55 and analyzing binding partners with mass spectrometry.

We find that inhibition of SD3 by SRp55 not only leads to induction of vpr mRNA but also to an increased export of vpr mRNA to the cytoplasm. The vpr mRNA contains an intron and would therefore be trapped in the nucleus and degraded. Somehow the vpr mRNA can overcome this and be exported to the cytoplasm for translation. The viral protein Rev facilitates export of unspliced and partially spliced viral mRNAs in the 9 kb and 4 kb class by interacting with RRE, an RNA element present on these mRNAs.
Vpr mRNA does not contain the RRE and can therefore not be exported with the assistance of Rev. We therefore speculate that SRp55 not only inhibits the SD3 to increase the production of vpr mRNA, but also that it can aid in the nuclear export of this mRNA. Another interesting question is therefore how SRp55 can facilitate the export of vpr mRNA to the cytoplasm. What mechanism is required for this export and is SRp55 associated with vpr mRNA in the cytoplasm? Is SRp55 mediating export by binding mRNA and nuclear pore complex or some other export factor, or is it interfering with the interaction between partially spliced mRNA and splicing factors thereby releasing export inhibitory factors from the mRNA? Some SR proteins have been shown to enhance translation. It would be interesting to see if SRp55 could, not only induce production and export of vpr mRNA, but also enhance the translation of this mRNA.

Our results show that SRp55 can have several functions in the regulation of HIV-1 splicing: by inhibiting a splice donor, and by facilitating the export of an incompletely spliced mRNA to the cytoplasm. It would be interesting to study the levels of SRp55 in the cell during HIV-1 infection. One could speculate that the levels would change during infection. It would also be interesting to see how the HIV-1 splicing pattern changes due to changes in SRp55 levels in infected cells. What happens during infection if SRp55 is knocked out? And how is the cell affected by this knock out? If SRp55 becomes a potential drug target it should preferably not be essential for the cell. One could generate a knock out mouse to observe the effects of SRp55 knock out.

Vpr is essential during viral infection and has several functions. It facilitates infection by mediating the nuclear import of HIV-1 pre-integration complex in non-dividing cells, it induces cell cycle arrest in the G2 phase, it transactivates the viral promoter and it induces cell death of immune cells (Agostini et al., 1996, Felzien et al., 1998, He et al., 1995, Kogan and Rappaport, 2011, Le Rouzic et al., 2002, Nie et al., 1998, Roshal et al., 2001). One possible antiviral strategy would be to interfere with the interaction between SRp55 and GAR, which could prevent vpr mRNA production. By using antisense oligos towards GAR, the interaction between SRp55 and GAR may be blocked. These antisense oligos would have to be specific for GAR and should preferably not interact with genomic sequences. In fact, antisense strategies targeting exon 5 have been tested with exon skipping and decreased viral production as a result (Asparuhova et al., 2007). Another approach could be to use small molecules that interact with SRp55. Small molecules called indole derivatives can interact with SR proteins and alter or inhibit their function in pre-mRNA splicing regulation. These molecules are potent inhibitors or re-directors of HIV-1 RNA production (Bakkour et al., 2007, Fukuhara et al., 2006, Soret et al., 2005). We test the effect of indole-related substances on HIV-1 gene expression in transient transfections and in primary cell cultures infected with HIV-1. Unfortunately none of these sub-
stances have any effect on HIV-1 replication. One reason is that the substances are toxic to the cells. To study this further we would probably use a more efficient screening method initially, where more substances could be tested at the same time.

Because splicing is essential for HIV-1 to be infectious it is important to find cis-elements and transacting factors that are regulating this splicing process. Both in vitro and in vivo methods are used when searching for these. The in vitro methods are fast and make it possible to study a specific RNA element without disturbing sequences that surround it. However, the splicing process is very complex. Interaction of a protein upstream of the sequence of interest could affect the interaction to the actual sequence. The secondary structure of the RNA could also be important in splicing regulation (Buratti and Baralle, 2004). In vivo experiments with full-length pre-mRNAs should give a better picture of the splicing due to the secondary structure. Taken together both in vitro and in vivo (transfections as well as infections) studies are important in the search for splicing regulatory elements or factors in HIV-1 pre-mRNA splicing.
Vägen till denna avhandling har varit ”long and winding”. Men som Gunde Svan brukar säga ”Går man den väg alla andra går, kommer man inte längre än någon annan”. Jag skulle vilja tacka alla som hjälpt mig:

Min handledare Stefan Schwartz för att jag fick möjligheten att forska i ditt lab! Du har alltid varit tillgänglig och jag har vågat ställa de mest korkade frågorna utan att känna mig dum.

Min bihandledare Göran Akusjärvi för alla råd och tips under denna tid. Du ger virusforskningen på IMBIM en stabilitet som är oerhört värdefull.

Catharina Svensson, för att du delat med dig av all din kunskap. Du är en inspirationskälla.

Min examinator Göran Magnusson för alla intressanta anekdoter på fredagsmötena.

Energiska och emotionella Monika Somberg, för att citera dig själv: ”partner in crime”. Vi visste inte vad som väntade när vi ungefär samtidigt började denna resa. Det har varit både upp och ner och hit och dit. Tack för alla hjälp och för alla skratt (och tårar) som du gett mig! Man blir aldrig uttråkad eller förvånad när man hänger med dig! (Och tack för alla lösningar jag ”fått låna”.)


Barbro, Erika, Kerstin, Marianne, Olav och Rehné. Utan er skulle IMBIM inte vara någonting.

Eva Maria Fenyö och Elzbieta Vincic för all hjälp på P3-labbet i Lund.

DA-gruppen för roliga lunchdiskussioner. Det får alltid plats några till vid ett runt bord.

Sara Brunt och Christer Harbeck för att ni gett mig redskap för resten av livet.

Biologfamiljen: Hanna S för alla klantigheter och alla roliga studentupplevelser. ”Man kan alltid tvätta händerna i lite NaOH”. Elisabet för att du alltid är ärlig och rak. Karin för roliga (och jobbiga) månader i Indien. Emma för din humor som alltid lockar fram skratt. Hanna B för att du är envis och smart. Tillsammans med alla andra medlemmar i biologifamiljen har ni gjort tiden i Uppsala till en av de bästa!


Jonas familj för att ni tagit emot mig med öppna armar. Ni är så härliga allihop!

Mina kära föräldrar. Tack mamma och pappa för all kärlek och allt stöd ni alltid har gett mig. Tack för att ni fyllt min uppväxt med trygghet och böcker, ”En barndom utan böcker, det vore ingen barndom”.


Ett stort tack också till morfar och mormor, farfar (var du än är) och farmor, Adolf.


through its interaction with the splicing regulator p32. *J Virol*, 80, 3189-204.


EARLY, P., ROGERS, J., DAVIS, M., CALAME, K., BOND, M., WALL, R. & HOOD, L. 1980. Two mRNAs can be produced from a single


HIV-1 gene expression: potential novel targets for therapy. Retrovirology, 8, 47.


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 749

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine.