SCUOLA NORMALE SUPERIORE – PISA

Single-cell and real-time analysis of transcription rates from integrated HIV-1 provirus

Thesis submitted for the degree of Doctor Philosophiae
(Perfezionamento in Genetica Molecolare e Biotecnologie)

Candidate:
Paolo Maiuri

Supervisor:
Dr. Alessandro Marcello

Academic Year 2008/2009

I. INDEX

II. INDEX OF FIGURES

III. ABSTRACT

Viral RNA biogenesis is a crucial step in the replication of retroviruses that require both the production of a genomic RNA as well as of translation templates. Cellular and viral factors concur in the biogenesis of RNA at the specific sub-nuclear chromatin site where the reaction takes place. The possibility of tracking viral RNA in living cells gives the unique possibility of measuring the kinetic parameters of RNA biogenesis as well as defining the dynamic recruitment of host and viral factors to the site of replication.

In order to study the activation of HIV-1 gene expression from the integrated viral promoter we exploited a method that allows the visualization of newly transcribed RNA in living cells through the specific recognition of an RNA consensus sequence for the bacteriophage MS2 coat protein tagged with an autofluorescent protein. We observed that transcription of HIV-1 occurred in discrete foci within the nucleus of cells carrying several tandem arrays of the HIV-1 construct. These foci, representing newly transcribed RNA, co-localized with the viral Tat transactivator as well as members of the positive transcription elongation factor (P-TEFb) and RNA polymerase II (RNAPII). This experimental setting was used to measure the dynamic of

HIV-1 RNA transcription in living cells. By fluorescence recovery after photobleaching (FRAP) we were able to demonstrate that following photobleaching the process reaches a steady state with a negligible immobile fraction allowing precise kinetic measurements of RNA polymerase elongation rates. We found that elongation proceeded at approximately 2 kb/min, and that 3'-end formation and release took another minute to complete. In addition we also analyzed the dynamic of RNAPII and the TAR:Tat:pTEFb complex at the site of HIV-1 transcription in living cells. Our data suggest that, while the residence time of RNAPII exceeds the time required for elongation through the viral template, the complex dissociates from the polymerase following transcription initiation, and may undergo subsequent cycles of association/dissociation.

This approach was extended to the analysis of single integrated HIV-1 transcription units by transduction of a HIV-based lentiviral vector into a human cell line and subsequent selection for Tat-induction from a latent state. Nascent RNAs from single integrated transcription units were detectable in living cells by MS2 RNA-tagging. At steady state a constant number of RNAs was measured at the transcription site corresponding to a minimal density of polymerases with negligible fluctuations over time both in space and intensity of the signal. Recovery of fluorescence after photobleaching of the transcription site was complete within seconds, much faster than what was observed

previously. However, the necessity of taking into account also the diffusion of the tagged MS2 protein required the development of novel analytical tools. To this end we developed a model that describes each polymerase sliding along the DNA like the peak of a positive progressive traveling wave (TranWave) to predict the number of MS2 RNA repeats at the transcription site in function of time. The outcome of this approach and its following improvements are being discussed.

This work provides for the first time a kinetic framework to analyze HIV-1 RNA biogenesis and RNA/protein dynamics in living cells.

List of candidate's publications during the Ph.D. training:

a) Strictly related to this thisis

S. Boireau, **P. Maiuri***, E. Basyuk, M. de la Mata, A. Knezevich, B. Pradet-Balade, V. Bäcker, A. Kornblihtt, A. Marcello, E. Bertrand. * first author, equal contribution
The transcriptional cycle of HIV-1 in real-time and live cells.
Journal of Cell Biology, 179 (2007): 291.

D. Molle, **P. Maiuri**, S. Boireau, E. Bertrand, A. Knezevich, A. Marcello, E. Basyuk.
A real-time view of the TAR:Tat:P-TEFb complex at HIV-1 transcription sites.
Retrovirology, 4 (2007): 36.

A. De Marco, C. Biancotto, A. Knezevich, **P. Maiuri**, C. Vardabasso, A. Marcello.
Intragenic transcriptional cis-activation of the human immunodeficiency virus 1 does not result in allele-specific inhibition of the endogenous gene.
Retrovirology, 5 (2008): 98.

b) Exploiting the MS2 tagging method

L. Miorin, **P. Maiuri**, V.M. Hoenninger, C.W. Mandl, A. Marcello.
Spatial and temporal organization of tick-borne encephalitis flavivirus replicated RNA in living cell.
Virology, 379 (2008): 64.

M. Dieudonné, **P. Maiuri**, C. Biancotto, A. Knezevich, A. Kula, M. Lusic and A. Marcello.
Transcriptional competence of the integrated HIV-1 provirus at the nuclear periphery.
EMBO J, 28 (2009): 2231.

IV. INTRODUCTION

IV.1 HIV

The acquired immunodeficiency syndrome (AIDS) is a life threatening disease caused by the human immunodeficiency virus type 1 (HIV-1). Following its discovery the virus has killed more than 25 million people world-wide and it remains a major threat to public health and a challenge for drug development [Ho and Bieniasz, 2008]. HIV-1 became the most studied virus in history and many years of research only partially unraveled the complexity of its life cycle. Studies on HIV led to critical discoveries in antiviral drug development but also gave rise to new concepts in viral and cellular biology. However, the plasticity of

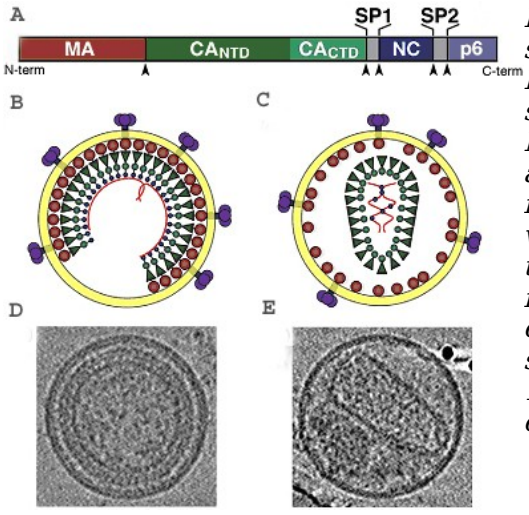

Figure IV.1: HIV-1 virion and structure of Gag polyprotein. HIV-1 Gag polyprotein (A) domain structure, showing the locations of MA, CANTD, CACTD, SP1, NC, SP2, and p6. Schematic models of the immature (B) and mature (C) HIV-1 virions. Central slices through tomograms of immature (D) and mature (E) HIV-1 particles derived by electroncryotomography. The spherical virions are approximately 130 nm in diameter [Ganser-Pornillos et al., 2008].

viral sequences, the establishment of a latent state, and the uncovered function of some auxiliary genes still remains a challenge in finding a cure for HIV infection [Richman et al., 2009].

IV.1.1 Virion structure

HIV-1 is composed of two copies of a single-stranded RNA enclosed by a proteic core (capsid) which is surrounded by the plasma membrane (envelope) of host-cell origin in which the viral glycoproteins are inserted. All structural proteins, which are components of the HIV-1 virion, are derived from the Gag polyprotein. Traditionally two forms of HIV particles, the immature form and the mature form, have been observed. The immature virion is a roughly spherical shell of radically extended uncleaved and multimerized Gag molecules (Fig. IV.1 B and D). During particle maturation, viral protease (PR) is activated and cleaves Gag generating a set of new proteins and spacer peptides (SP), termed MA, CA, SP1, NC, SP2 and p6 (Fig. IV.1 A). These newly processed proteins then reassemble to form the distinct layers of the mature virion: MA remains associated with the inner viral membrane (the ‘matrix’ layer), NC coats the viral RNA genome (the ‘nucleocapsid’ layer), and CA assembles into the conical capsid that surrounds the nucleocapsid and its associated enzymes: reverse transcriptase (RT) and integrase (IN) (Fig. IV.1 C and E).

IV.1.2 Genome organization

HIV-1 is a retrovirus, possessing an RNA genome that is replicated *via* a DNA intermediate. The full HIV genome is encoded on one long strand of RNA and contains approximately nine thousand nucleotides. HIV has several major genes encoding for structural proteins that are found in all retroviruses and several nonstructural (accessory) genes that are unique for HIV. The *gag* gene provides structural elements of the virus and the *pol* provides the replication enzymes. The *Env* gene codes for glycoproteins which are exposed on the surface of the viral envelope. Accessory proteins (Tat, Rev, Vif, Vpr and Vpu) help to enhance virus infectivity.

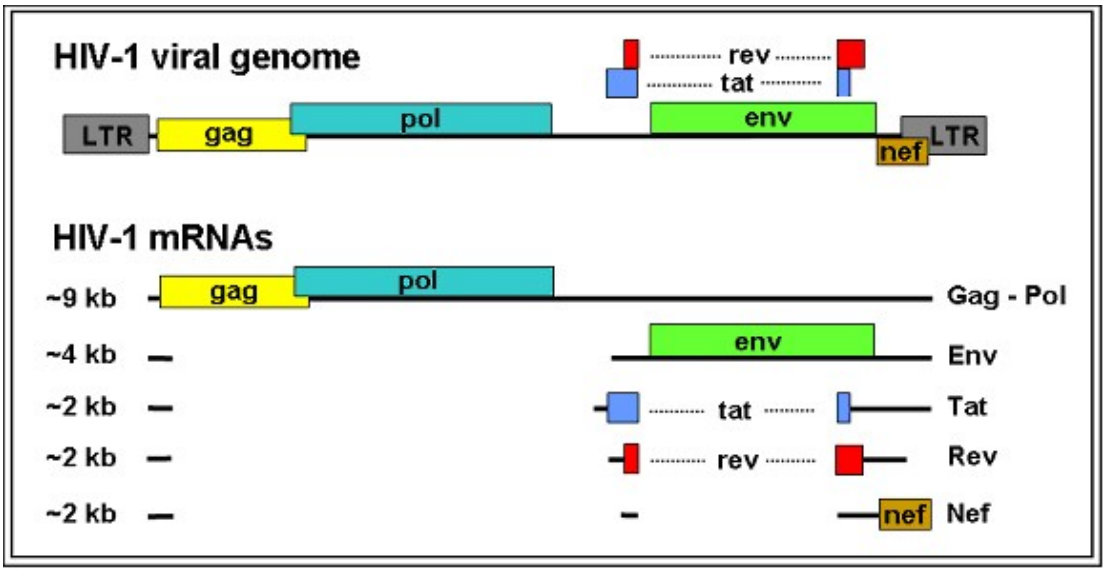

Figure IV.2: Structure of the HIV-1 genome and its mRNA products.
The open reading frames for various polypeptides are shown as rectangles. The HIV-1 viral transcript is spliced in different ways yielding unspliced (9 kb), singly spliced (4 kb) and fully spliced (2 kb) mRNAs. In order to simplify the figure only the most important viral proteins are shown [Rahbek, 2008].

IV.1.3 HIV-1 life cycle

After entry into a cell the single stranded RNA genome of the virus is

reverse transcribed into a double stranded DNA that is transported to the nucleus and integrated into the DNA of the host cell as a provirus [Greene and Peterlin, 2002]. Integrated provirus is transcribed by the host – cell RNA polymerase II (RNAPII) producing a copy of the viral RNA genome that can be used to infect other cells (Fig. IV.3) [Haseltine, 1991]. HIV infects cells of the immune system such as CD4+ T cells and cells of the monocyte – macrophage lineage. Productive infection starts when HIV-1 fuses its envelope with the cell plasma membrane. Entry is mediated by binding of the viral envelope glycoproteins gp120/gp41 to the CD4 receptor and to a co-receptor (the chemokine receptors CCR5 or CXCR4) [Berger et al., 1999]. These co-receptors provide a critical function for virus entry. CCR5 binds macrophage-tropic viruses (R5 viruses) and CXCR4 – T cell tropic viruses (X4 viruses). Following fusion, the viral core is released into the cell cytoplasm. Immediately after release into the cytoplasm the viral core undergoes a progressive disassembly, known as uncoating and the viral RNA genome is retrotranscribed into double stranded DNA by the viral reverse transcriptase. Finishing of reverse transcription originates the HIV preintegration complex (PIC). The viral proteins present in this complex are integrase (IN), MA, reverse transcriptase (RT), NC, and Vpr [Farnet and Haseltine, 1991]. In contrast to most retroviruses, HIV is able to infect nondividing cells such as differentiated macrophages, thus PIC needs to be translocated into the nucleus.

To reach the nuclear membrane, the particles must travel through the cytoplasm. HIV transport has been shown to exploit the cellular cytoskeleton. In particular, initial movements at the cell periphery occur in association with actin [Bukrinskaya et al., 1998], while subsequent translocation to the nucleus takes place along the microtubule network, likely by interaction of PIC with the dynein-dependent motor complex [McDonald et al., 2002]. After reaching the nuclear envelope PIC is translocated through the nuclear pore, most likely by relying on the cellular nuclear import machinery.

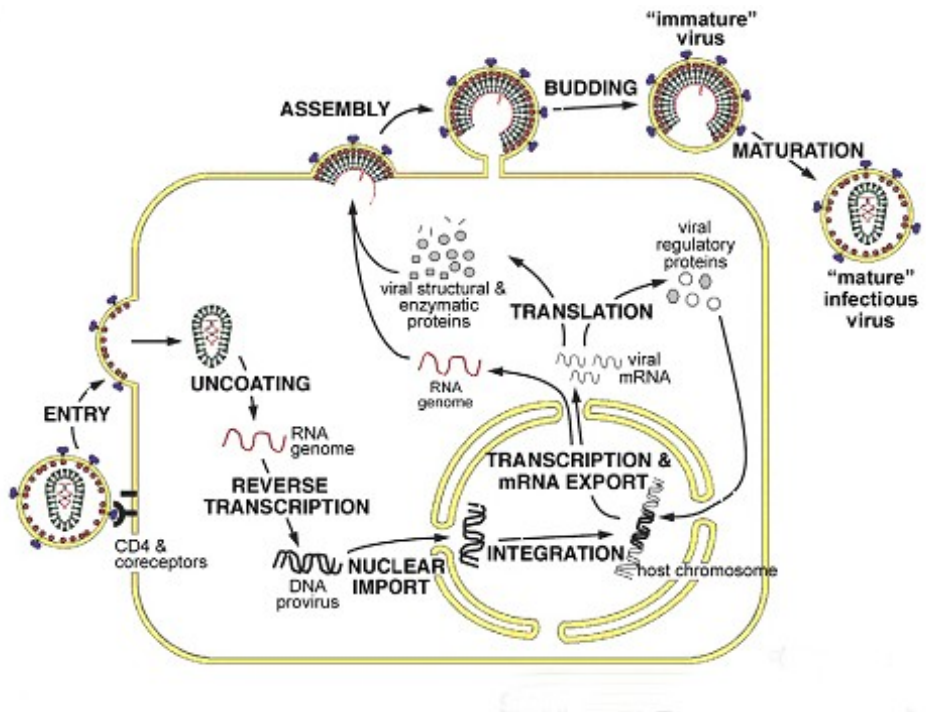

Figure IV.3: Summary of the HIV-1 replication cycle.
After binding and fusing to the surface of the host cell (entry), the HIV-1 core enters the host cell (uncoating). Reverse transcriptase synthesizes proviral double-stranded DNA, which is transported to the nucleus. Integration occurs and the ensuing provirus expresses viral RNA that will synthesize viral proteins. New virions are created by assembly and budding through the infected cell membrane and subsequent maturation due to the actions of protease [Ganser-Pornillos et al., 2008].

Like all retroviruses, HIV integrates into the host chromatin. The process of integration of the linear viral DNA is carried out by the viral integrase protein (IN) in coordination with several host cell factor.

For a long time integration of HIV and more in general of all retroviruses was believed to occur randomly into the host chromatin. Nevertheless, recent reports have challenged this notion showing a bias for integration into transcriptionally active genes [Schroeder et al., 2002]. Moreover, the chromatin status at the site of integration determines whether the provirus is transcriptionally active, poised for activation or inactive. The provirus once integrated into host genome can remain transcriptionally inactive, in a latent state or can undergo active transcription allowing continuous rounds of infection. After two months, in patients on HAART (antiretroviral therapy) the plasma levels of genomic RNA falls below the limit of detection. Therefore, it was initially assumed that prolonged treatment might lead to eradication of the virus in these patients [Marcello, 2006]. Unfortunately, it is now clear that long-live reservoirs of HIV-1 can persist for years in the presence of HAART. This reservoir is thought to consist mainly of latently infected resting memory CD4+ T cells [Chun and Fauci, 1999] [Pierson et al., 2000]. Both host transcription factors and the viral Tat *trans*-activator have been proposed as limiting factors for transcriptional reactivation of latent HIV-1. However, since HIV-1 is found integrated into the genome of resting memory T cells, it has been proposed that the

chromatin environment at the viral integration site may play a role in the transcriptional silencing of the HIV-1 genome [Jordan et al., 2003] [Jordan et al., 2001]. Indeed, integrated HIV-1 has nucleosomes positioned in its 5’ LTR that are remodeled by deacetylase inhibitors, cytokines and Tat [Lint et al., 1996][Lusic et al., 2003][Marcello et al., 2004]. Histones play an important role in regulating HIV-1 transcription since they integrate signals for repression, like the heterochromatin marker H3K9 trimethylation, Suv39H1 and HP1, and reactivation, like histone acetylation [Lusic et al., 2003][Chene et al., 2007]. It is now clear that also the three-dimensional (3D) nuclear architecture of the nucleus could be implicated in HIV-1 provirus regulation, as it has been proposed for cellular genes [Misteli, 2007][Dieudonne et al., 2009].

In an infectious state, the integrated provirus is actively transcribed from its promoter located in the U3 region of the HIV-1 LTR. Once the HIV-1 pre-mRNA transcript is produced, it can be spliced in alternative ways to yield three classes of mRNA in the nucleus: unspliced mRNA (9 kb), singly spliced mRNA (4 kb) and fully spliced mRNA (2 kb). The three classes of HIV-1 mRNA and the viral proteins they encode are shown in Figure IV.2. All three classes of mRNAs must be exported to the cytoplasm and translated into viral proteins for the viral life cycle to proceed. The fully spliced mRNAs, which are the first viral transcripts that appear after infection, are exported to the cytoplasm. They follow the same pathway of cellular RNA [Cullen, 1998] leading to expression

of the regulatory proteins Nef, Tat and Rev. The negative factor (Nef) is a modulator of cellular signaling pathways that optimizes the cellular environment for virus replication. The transcriptional activator (Tat) up-regulates the synthesis of viral mRNA. The regulator of viral gene expression (Rev) acts on a post-transcriptional level and plays a main role in Rev-dependent export of unspliced and incompletely spliced HIV transcripts from the nucleus to the cytoplasm.

The unspliced HIV transcript is the full-length viral RNA which acts both as a messenger and genomic RNA. Two copies of genomic RNA are encapsidated, usually linked at their 5’ end through the dimer linkage site [Darlix et al., 1990]. Dimerization is associated with encapsidation [Greatorex and Lever, 1998], but dimer stabilization occurs post-capture by the Gag protein involving the NC. Viral RNAs to be specifically packaged are identified by the presence of an RNA sequence named the packaging signal (ψ) placed within the 5’leader [McBride and Panganiban, 1996][Berkowitz et al., 1995].

To a first approximation, all the information necessary for retroviral particle assembly resides in the Gag polypeptide. For example, Gag alone can form extracellular virus-like particles in the absence of other viral proteins [Gheysen et al., 1989], and Gag molecules can spontaneously assemble into spherical, immature virus-like particles *in vitro* [Campbell and Rein, 1999]. All the viral proteins necessary for virion assembly and RNA genomes are transported to the plasma

membranes close to lipid raft membrane domains, where the building of new virions begins. The Gag-Pol precursor binds to plasma membrane through the myristol group of the MA domain. The resulting virions bud from the plasma membrane but they are still incomplete. Their maturation is finished by viral protease (PR) that cleaves Gag-Pol. Further, Gag and Pol precursors are processed to originate the single core proteins, marix and the viral enzymes. The proteolitic activity ends when the virion is already detached from plasma membrane and results in the formation of mature infectious viruses.

IV.1.3.1 Accessories genes

HIV-1 encodes at least nine genes which must be expressed during viral life cycle in the correct temporal order. Early in the viral infection, small, multiply spliced transcripts (2 kb) encoding Tat, Rev and Nef predominate in the cytoplasm. During the late phase of HIV infection, genomic (unspliced, 9kb) RNAs and singly spliced RNAs (4 kb) become leading species in the infected cell.

Nef

During the replication cycle, other auxiliary proteins of HIV-1 play a fundamental role in regulating the different steps of the intrecellular viral pathway. Completely spliced transcripts encode for Nef that is targeted to the plasma membrane by myristylation of its N-terminus. The first studied activity of Nef is the down-regulation of the CD4 receptor, a mechanism which is supposed to protect infected cell from

the immune system and to facilitate the release of HIV-1 virions. In addition to CD4, Nef is able to promote the down-regulation of cell surface MHC I (major histocompatibility complex 1) that presents viral epitopes to cytotoxic lymphocytes (CTL). As consequence, Nef impairs the CTL-mediated immune response that leads to the lysis of HIV-1 infected cell [Collins et al., 1998].

Vif

The viral infectivity factor Vif affects the infectivity but not the production of viral particles. The protein is encoded from singly spliced viral transcripts and its expression is therefore Rev-dependent. Once expressed by infected cell, the protein accumulates in the cytoplasm and is found in association with the plasma membrane. Mutant viruses lacking Vif are less infectious than the wild type, although cell-to-cell transmission is only slightly lower than normal. It has been demonstrated that Vif promotes the degradation of APOBEC3G [Sheehy et al., 2003]. APOBEC3G induces hypermutation in the retroviral plus strand DNA, in which up to 25% of all deoxyguanosine (G) sequenced are mutated to deoxyadenosine (A). This hypermutation then induces the degradation of the viral genome or, in the case of successful reverse transcription and insertion into the cell chromatin, the production of non-functional viral proteins. APOBEC3G restriction is overcome by the HIV-1 Vif protein which targets APOBEC3G to ubiquitination and thus proteasomal degradation [Sheehy et al., 2003].

Vpr

Vpr is expressed at later stages of HIV-1 replication and is packaged in all HIV-1 virions by the p6 core protein. After viral entry, Vpr is incorporated inside the PIC and enhances HIV-1 replication in nondividing cells by mediating its nuclear import. In addition, Vpr induces cell cycle arrest in the G2 phase through the inhibition of p34cdc2-cyclin B kinase activity increasing virion production [Re et al., 1995].

Vpu

Vpu is a transmembrane protein unique to HIV-1. Vpu binds to the cytoplasmic tail of CD4 and recruits the β-TrCP/Skp1p complex at the level of the endoplasmatic reticulum. This event induces the ubiquitination and the subsequent preteosomal degradation of CD4 while Vpu is recycled. In addition, Vpu enhances release by inducing the budding of virions from the plasma membrane and promoting the detachment of budding virions [Margottin et al., 1998].

Rev

Rev specifically recognizes an RNA element located within the coding sequence for Env. This sequence, called the RRE (Rev-responsive element), is present within the *env*-coding region and is about 200 nucleotides in size and forms secondary RNA structure. Rev binds to this region through amino acids in the basic arginine-rich domain [Zapp and Green, 1989].

Rev permits export of unspliced/partially spliced transcripts from the nucleus to the cytoplasm where they serve as a template for translation of the *gag-pol* open reading frame. On the other hand the unspliced transcript is the full-length genomic RNA which thanks to Rev can be translocated to the cytoplasm for encapsidation.

IV.1.4 Tat: structure and function

The *trans*-activator protein (Tat) is an unusual transcription factor because it interacts with a *cis*-acting RNA enhancer element, TAR, present at the 5’ end of all viral transcripts. The function of Tat has been described for the first time by Sodroski *et al* who noted that synthesis of reporter genes placed under the control of the viral long terminal repeat (LTR) increased 200- to 300-fold in the cells which have been previously infected by HIV [Sodroski et al., 1985]. They reasoned that the induction, or “transactivation” of transcription was due to the presence of a novel *trans* – activating factor, which they named Tat.

IV.1.4.1 The structure of Tat protein

Tat is a small nuclear protein of 86 to 101 amino acids (depending on the viral strain) which is encoded from two separate exons. Tat is translated from a group of multiply spliced monocistronic transcripts. On the basis of amino acids distribution in the sequence and their conservation with homologous proteins from other leniviruses, it is possible to identify five domains. In the first exon (aa 1–72), it can be

distinguished an N – terminal acidic domain (aa 1 – 21); a domain containing 7 cysteins (aa 22 – 37); a core region (aa 38 – 48); and a basic domain (aa 49 – 57) enriched in arginine and lysine amino acids, highly conserved among different strains. In the reporter plasmid co-transfection assays, the first 72 amino acids of Tat fully *trans*-activate transcription from the LTR [Garcia et al., 1988]. A basic domain RKKRRQRRR confers TAR RNA-binding properties of Tat [Dingwall et al., 1989][Chang and Jeang, 1992] and is important for nuclear localization of the protein [Ruben et al., 1989]. However, studies suggest that this short basic stretch is insufficient in determining the entire specificity of Tat-TAR binding since amino acids outside of the basic domain also contribute to this interaction [Churcher et al., 1993]. The second coding exon of Tat has been less studied. In the routine transfection of reporter plasmids, absence of the second exon does not alter greatly measurements of Tat activity in this type of assay. However, findings from Tat of HIV-2 or SIV are quite clear in demonstrating that this exon contributes towards optimal *trans*-activation [Tong-Starksen et al., 1993].

No crystal structure of Tat has been obtained to date; structural prediction and data from NMR spectrometry [Bayer et al., 1995] indicate that the protein has a highly flexible structure and does not exhibit obvious secondary structure elements.

IV.1.4.2 Role of Tat in HIV-1 transcription

Transcription of the HIV-1 provirus is characterized by an early, Tat-independent and a late, Tat-dependent phase. Transcription from the HIV-1 LTR is increased several hundred-fold in the presence of Tat and the ability of Tat to activate transcription is essential for virus replication.

IV.1.4.2.1 HIV-1 promoter

The U3 region of the HIV-1 LTR functions as the viral promoter. The viral LTR promoter has a structure typical of promoters activated by cellular RNA polymerase II. It contains several binding sites for general transcription factors as shown in Figure IV.4. The HIV-1 promoter contains the core promoter sequence (TATA box) and enhancer elements placed upstream and downstream of TATA box. Immediately upstream of the TATA box are two tandem NF-κB binding sites and three tandem SP–1 binding sites. Immediately downstream of the start of transcription is the transactivation response element (TAR). Three tandem binding sites for the constitutively expressed Sp1 transcription factor are necessary for basal levels of LTR-directed RNA synthesis. Mutation of individual or pairs of Sp1 sites has little, if any, effect on the basal or Tat-transactivated levels of expression [Harrich et al., 1990]. However, the mutation of all three Sp1 sites markedly reduces the response to Tat [Berkhout and Jeang, 1992]. Two tandemly arranged binding sites (κB sites) are recognized by the dimeric transcription factors composed of

several combinations of members of the Rel/NF-κB family of polypeptides [Baeuerle and Baltimore, 1996]. The predominant complex that binds to the LTR κB sites in activated cells is NF-κB (p50/p65 heterodimer). In unstimulated cells, the p65 subunit of NF-κB is retained

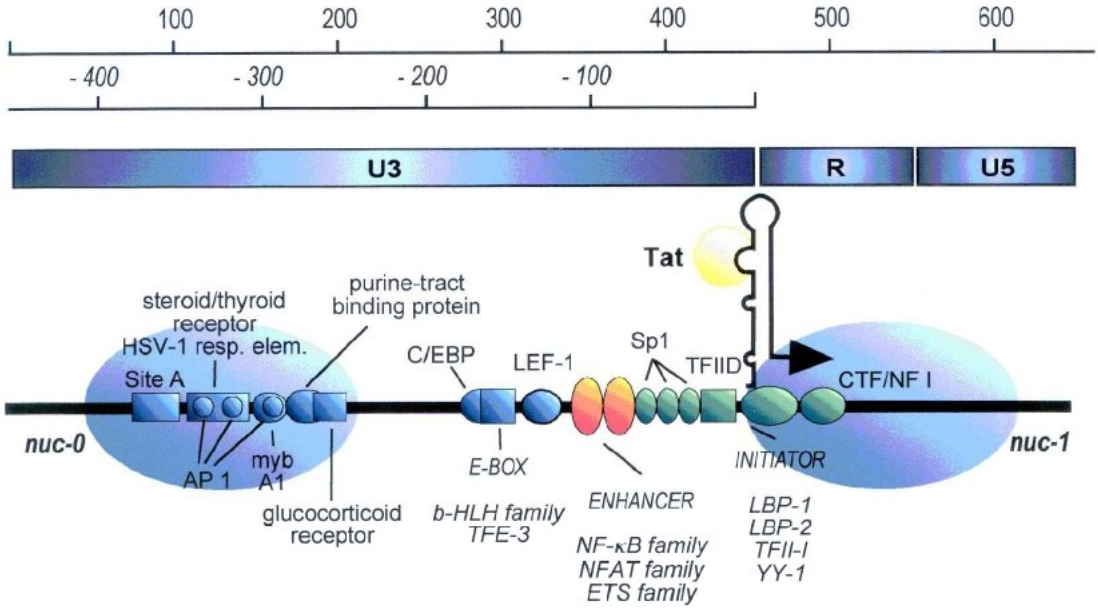

Figure IV.4: . Transcriptional control elements of the HIV-1 5' LTR.
Schematic representation of HIV-1 LTR promoter with the position of the identified binding sites for cellular transcription factors. Factors binding to the basal promoter region are in green, the enhancer is in orange, and upstream sites are in blue. The positions of nuc-0 and nuc-1 nucleosomes are indicated. Tat binds to a bulge region of a stem-loop structure formed by nascent RNA. The U3, R and U5 structures of the HIV 5' LTR are shown [Marcello et al., 2001].

in the cytoplasm through its interaction with inhibitor proteins belonging to the IκB family. The activation of NF-κB occurs through the phosphorylation and proteolysis of the IκB inhibitor, and the subsequent translocation of p65 to join p50 in the nucleus where the factor can bind to its cognate binding sites [Henkel et al., 1993]. Moreover, the arrangement of the transcription factor binding sites in the LTR may vary in different HIV-1 subtypes but the specific contribution of these

changes has not been yet thoroughly investigated.

IV.1.4.2.2 The Tat – TAR RNA interaction

The transactivation - responsive region (TAR) is located downstream of the initiation site for transcription (nucleotides +1 and +59). TAR RNA sequence forms a highly stable, nuclease – resistant, stem – loop structure. The stem – loop structure is located between +19 and +43 nucleotides which is the minimal sequence for transactivation. Because of its location in the R domain of the LTR, TAR is present in both viral RNA and DNA.

IV.1.4.2.3 Tat-mediated transcriptional activation

The interaction of Tat with TAR permits activation of HIV-1 transcription by promoting the assembly of transcriptionally active complexes at the LTR by multiple protein–protein interactions. Over the years, a number of cellular proteins have been reported to interact with Tat and to mediate or modulate its activity. These include general transcription factors, among which are TBP, TAFII250, TFIIB, TFIIH, hNAP-1 [Kashanchi et al., 1994][Parada and Roeder, 1996] [Vardabasso et al., 2008], RNA polymerase II [Wu-Baer et al., 1995]; transcription factor Sp1; the cyclin subunit of the positive transcription elongation factor complex (P-TEFb), cyclin T1 [Wei et al., 1998], and different transcriptional co-activators that possess histone acetyltransferase activity. These interactions with P-TEFb and HATs appear to be of particular importance in the transcriptional activation of the viral promoter.

- P-TEFb

Cyclin T1 is one of the cyclin subunits of P-TEFb (positive transcription elongation factor b) a general elongation cofactor originally identified in *D. melanogaster* for RNA polymerase II (RNAPII)-directed transcription [Price, 2000]. P-TEFb is known to be present in the early elongation complexes and is required for elongation after promoter clearance. Although the HIV-1 LTR contains DNA binding sites for several transcription factors, in the absence of Tat, there is no expression of viral genes. However, there is production of short, non-polyadenylated RNAs that include TAR RNA. The instability of synthesizing full-length viral transcripts is caused by the low processivity of RNAPII, which is overcome by P-TEFb. The P-TEFb complex is composed of one component of the cyclin T family and of the cyclin-dependent kinase Cdk9. Cyclin T1 participates in TAR RNA recognition. It has been demonstrated that sequences in the apical loop of the TAR RNA are required for cyclin T1 binding [Wei et al., 1998]. More recently it has been discovered that P-TEFb exists in two complexes: large and small. P-TEFb is inactive in the large complex which additionally contains HEXIM1, hnRNPs and the 7SK small nuclear RNA (snRNA). In response to stimuli such as stress or UV light 7SK RNA, as well as HEXIM proteins and hnRNPs are released by an unknown mechanism and P-TEFb is activated forming the small complex [Barrandon et al., 2007] [Nguyen et al., 2001].

During the pre-initiation of HIV-1 transcription the carboxy-terminal

domain of RNA polymerase II (RNAPII) is phosphorylated by CDK-7, a component of TFIIH initiation factor (Fig. IV.5 a). CTD phosphorylation is an early step associated with the clearance of the promoter (Fig. IV.5 b). Following clearance of the promoter, the phosphorylated RNAPII is able to transcribe through the TAR region. When the TAR RNA stem-loop structure is synthesized, the Tat protein binds to it and by interacting with cyclin T1, activates P-TEFb. The CDK9 kinase present in P-TEFb becomes constitutively activated. The actvated P-TEFb hyperphosphorylates the CTD domain of RNAPII (Fig. IV.5 c). This molecular event is associated with increased transcriptional processivity (Fig. IV.5 d). The formation of the P-TEFb-TAR-Tat tripartite complex is an essential step towards the assembly of the processive RNAPII machinery at the LTR promoter [Bieniasz et al., 1998] [Fujinaga et al., 1998].

- HATs

The LTR acts as a very strong promoter when analyzed as naked DNA *in vitro*, but when it is integrated into the cellular genome is almost silent. Therefore, chromatin conformation imposes inhibition onto the integrated promoter [Marzio et al., 1998]. Experiments performed both *in vivo* [Verdin et al., 1993][Kharroubi et al., 1998] and *in vitro* using the HIV promoter reconstituted into chromatin [Sheridan et al., 1997] have shown that, independent from the integration site, nucleosomes at the 5’ LTR are precisely positioned with respect to the *cis*-acting regulatory elements as shown in Figure IV.4. In the transcriptionally silent

provirus, these nucleosomes define two large nucleosome-free areas, one associated with the promoter/enhancer in the U3 region and the other spanning the primer-binding site immediately downstream of the 5’ LTR. These two open regions are separated by a single nucleosome called nuc-1 that is specifically and rapidly destabilized during transcriptional activation. Complexes containing histone acetyltransferase (HAT) activity facilitates transcriptional activation by modulating nucleosomal repression of specific promoters through acetylation of the N-terminal tails of histones. This event destabilizes the histone-DNA interaction. The HAT proteins responsible for the TAR-dependent Tat transactivation include the transcriptional co-activators p300, the highly homologue cAMP-response element binding protein (CREB) binding protein (CBP), the p300/CBP-associated factor (P/CAF), the general control non-derepressible-5 (GCN5) factor, the TIP60 protein, and the general transcription factor TAFII250. Lusic *et al*. has observed acetylation of histones H3 and H4 at distinct nucleosomal regions and the recruitment of mentioned HATs to the viral LTR promoter upon HIV activation either with Tat or phorbol esters [Lusic et al., 2003].

The Tat-TAR-P-TEFb interaction which increased the processivity of RNA Pol II confirms that Tat plays a role in transcription elongation. The initiation model quickly lost support when Kao *et al*. reported that in the absence of Tat the majority of RNA polymerases initiating transcription

stall near the promoter [Kao et al., 1987]. On the other hand it is also clear that optimal Tat transactivation of HIV-1 gene expression requires upstream transcription co-factors. Along these lines it has been reported that Tat physically interacts with the pre-initiation complex (PIC) including transcription factors such as Sp1 [Jeang et al., 1993], TATA

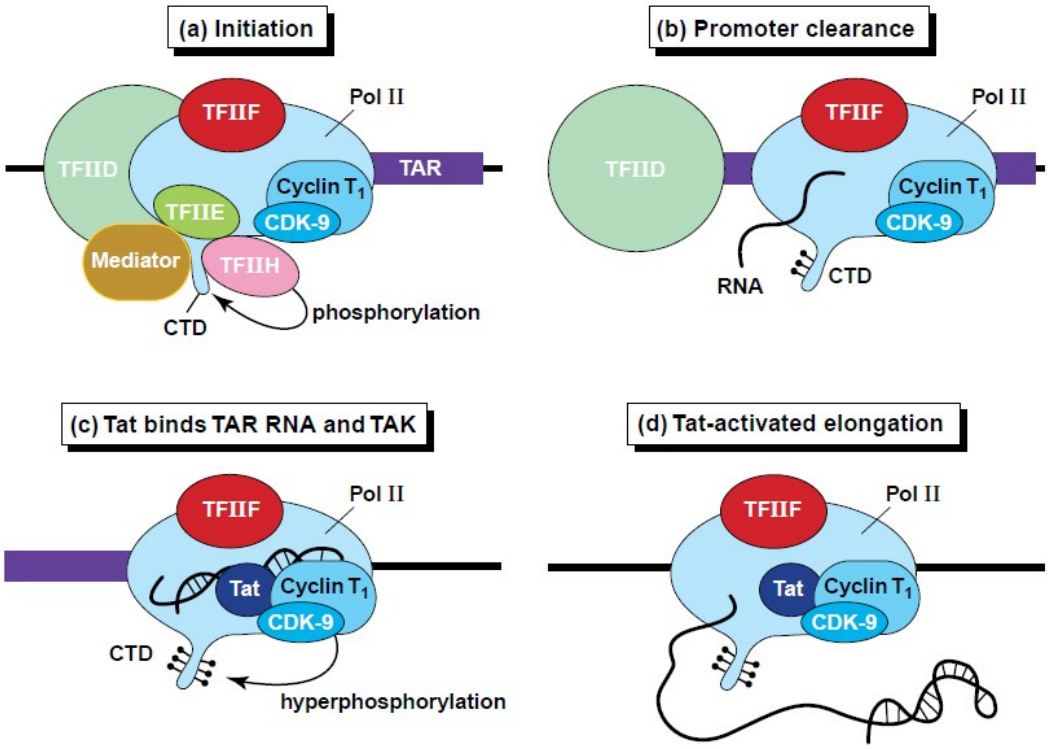

Figure IV.5: Tat activates HIV-1 transcription.
(a) The RNA polymerase II (Pol II) holoenzyme is recruited to the HIV LTR through its interactions with TFIID and other components of the basal transcription apparatus. The CTD domain of the RNA polymerase is phopshorylated by the CDK-7 kinase found in TFIIH and the modified polymerase clears the promoter and begins transcription of TAR. (b) The nascent RNA chain corresponding to the TAR RNA transcript folds into its characteristic stem-loop structure and binds the RNA polymerase. (c) Tat is recruited to the transcription complex through binding to TAR RNA structure and forms a ternary complex with TAK (contains CDK-9 kinase and Cyclin T1). The activated CDK-9 kinase catalyses phosphorylation of the CTD and TAR is displaced from the polymerase. (d) The activated transcription complex is able to transcribe the remainder of the HIV genome [Karn, 1999].

binding protein (TBP) [Majello et al., 1998], cylinE/cdk2 [Deng et al., 2002], TFIIH, Tip60 [Kamine et al., 1996] and RNAPII [Cujec et al.,

1997]. The evidence of Tat interaction with histone acetyltransferases support the role of Tat in chromatin remodeling before the onset of viral transcription. It is possible therefore that Tat facilitates chromatin remodeling, pre-initiation complex assembly and transcription elongation in the sequential manner that leads to high levels of HIV transcription. All of these activities of Tat are probably exerted by the transient formation of large subnuclear complexes at the site of HIV transcription and must therefore be spatially and temporally regulated. It has been demonstrated [Marcello et al., 2003] that cyclinT1 and p300 interact with the promyleocytic leukemia (PML) protein within specific subnuclear compartments that are coincident with PML nuclear bodies. This observation suggests that PML bodies could regulate Tat activity by modulating the availability of essential factors to the transcriptional machinery.

IV.2 Nuclear architecture

The nucleus is a complex organelle with an internal structure and component organization that is not fully identified. Understanding the molecular details of the organization of the nucleus including: how the chromosomes are arranged, how the nuclear compartments are formed and maintained, how the synthesis, processing, assembly and transport of molecules are coordinated and regulated are one of the main objectives of cell biology. It is well accepted that organization of

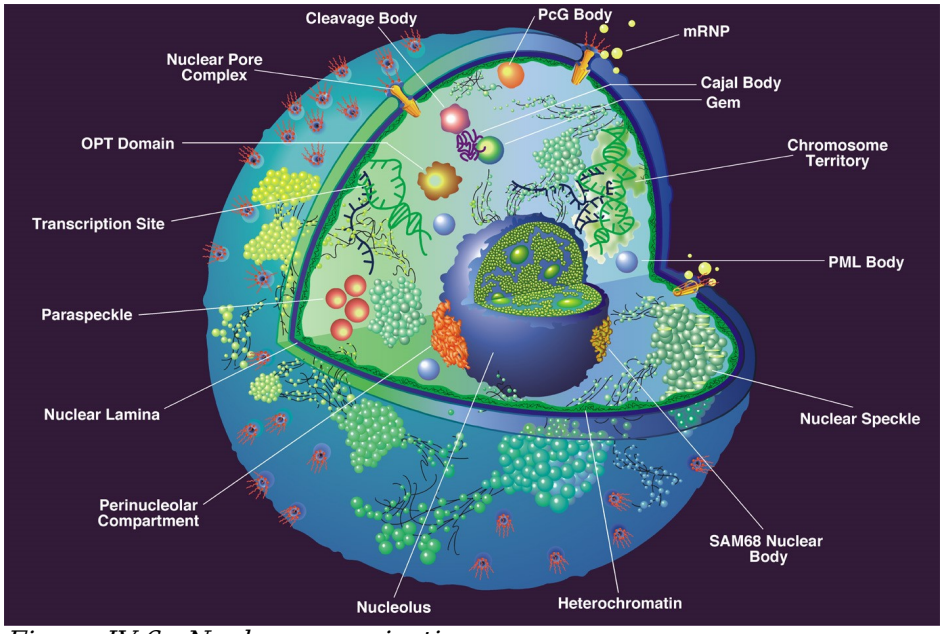

Figure IV.6: Nuclear organization.
The nucleus is a complex organelle which contains many specialized domains or nuclear sub-structures. Each nuclear domain is a highly dynamic structure which is involved in a certain nuclear process [Spector, 2001].

chromatin is not random and whole chromosomes reside in specific locations within the nucleus called chromosome territories [Cremer and Cremer, 2006][Cremer and Cremer, 2006a]. Within the nucleus, protein and RNA-protein complexes can move to and from chromatin through the interchromatic nucleoplasmic space by a diffusion mechanism until they encounter an appropriate binding site. In addition, many nuclear factors are able to form or interact with different types of subnuclear domains, which are either attached to chromatin or located in the interchromatic space. The nuclear bodies (also called nuclear speckles or domains) include nucleoli, splicing speckles, Cajal bodies, gems, PML bodies and most recently identified paraspeckles (Fig. IV.6). They are

highly dynamic nuclear structures not enclosed by membranes and their appearance is probably the result of a tight association of proteins and RNA and perhaps DNA components involved in the synthesis, assembly, and storage of macromolecules playing a role in gene expression. Moreover, the exact maintenance of subnuclear organization has important consequences for cellular function and perturbation of various nuclear domains may be associated with disease phenotype.

IV.2.1 Chromosome territories

The largest unit of organization of the eukaryotic genome is the chromosome. Chromosome painting approaches revealed that chromosomes have preferred positions with respect to the centre or the periphery of the nucleus and with respect to each other [Cremer and Cremer, 2006]. Thus, the spatial positioning of a chromosome territory (CT) in the nucleus is far from random. A correlation has been observed between CT location and human chromosome size, in which smaller chromosomes are generally situated towards the interior and larger chromosomes towards the periphery of the nucleus [Sun et al., 2000]. However, there is strong evidence that CTs with similar DNA content, but with very different gene densities, occupy distinct nuclear positions. This observation indicates that gene content is a key determinant of CT positioning. A striking example comes from the distribution of human chromosome 18 and 19. Although both chromosomes have a similar DNA content (85 and 67 Mb, respectively), the gene-poor chromosome

18 territories were typically found at the nuclear periphery, whereas the gene-rich chromosome 19 territories were located in the nuclear interior [Croft et al., 1999]. Genes sometimes move outside of their chromosome territory. This relocation occurs either in domains of constitutively high gene expression [Mahy et al., 2002] or, in some situation, when gene expression is induced [Volpi et al., 2000]. For example, the *Hoxb* gene cluster is activated at the same time as it moves away from the chromosome territory [Chambeyron and Bickmore, 2004]. A consequence of this relocation is that the 'looped out' *Hoxb* locus can now interact more with other chromosomes. This finding, together with the observation of extensive intermingling of DNA from different chromosomes at the boundary of, or just outside of, chromosome territories suggests that chromosome territories might be more dynamic then previously thought [Pombo and Branco, 2007].

IV.2.2 Transcription factories

Many nuclear processes are spatially compartmentalized. Transcription seems to be organized in subnuclear transcription centers, also called transcription factories. Several thousand of distinct transcription sites appear to be randomly dispersed through the nucleoplasm [Wansink et al., 1993]. These transcription sites represent hot spots of transcription that contain complex sets of transcription factors and polymerases to serve multiple genes [Cook, 1999]. It has been estimated that in the single nucleus there are 65˙000 active

RNAPII molecules but less that 10˙000 trascription sites in HeLa cells. Therefore transcription factories would contain 6 -8 actively elongating RNAPII that would probably transcribe multiple genes at the same time. Analogously, ribosomal RNA genes which are transcribed by RNA polymerase I within the nucleolus represent large specialized transcription centers. The compartmentalization of transcription has the obvious advantage of concentrating the needed factors to ensure efficient interactions among the components of the transcription machinery. Transcriptionally active genes have been proposed to loop out from their chromosome territory and associate with transcription factories. Work of Osborne et al. showed that widely separated genes are recruited to shared RNAP II sites if transcriptionally active or reposition away when inactive [Osborne et al., 2004]. They have postulated that genes with transcriptional potential often move out to existing transcription factories when activated rather than assemble their own transcription sites *de novo*. Transcription units might be associated through promoters and/or enhancers, with loops forming a ‘cloud’ around the factory. Transcription factories are highly dynamic structures and involved transient interaction of factors with chromatin [Misteli and Spector, 1998]. Many transcription factors undergo rapid exchange between chromatin and nucleoplasm and for instance the glucocorticoid- and estrogen-receptor transcriptional co-activators bind to their specific response elements in the promoter with residence times

of only a few seconds [McNally et al., 2000].

IV.3 MS2-based tagging of RNA

Labeling either by fusion with a fluorescent protein or with a fluorescent-dye moiety allows the dynamic study of a protein of interest inside the cell *in vivo*. In the case of nucleic acids, intra-molecular tagging approaches are required because the target DNA or RNA are typically not accessible for tagging *in vivo*. Labeling of chromosomes *in vivo* can be achieved by fluorescent tagging of chromosome binding factors. Usage of the lac operator (lacO) DNA sequence which is

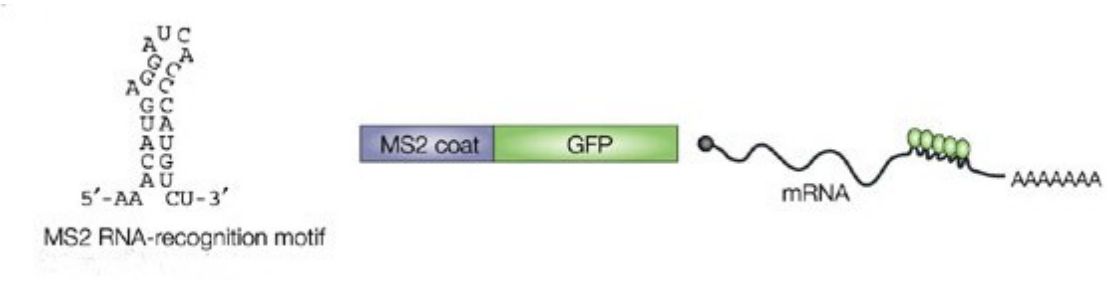

Figure IV.7: RNA detection MS2-based system.
RNA molecules are detected using a fusion protein that comprises a fluorescent tag, such as green fluorescent protein (GFP) and the MS2 bacteriophage coat protein. MS2 has an extremely high affinity for a 19 nt RNA-recognition motif that is derived from the phage genome [Shav-Tal et al., 2004a].

inserted into the DNA of interest provides an elegant way to label specific chromatic regions [Robinett et al., 1996]. The specific part of chromatin can be visualized by expression of a fluorescently tagged lac repressor (lacI) fusion protein which specifically binds to lacO sequence. This approach has been used to understand the chromosome positioning

and movements in living cells [Janicki et al., 2004]. The mRNA can be conjugated to fluorescent dyes and directly microinjected to the cells, but this is a rather invasive procedure [Ainger et al., 1993]. Several other techniques have been developed which include the use of fluorescently labeled probes for *in vivo* hybridization, the use of molecular “beacons” which are active only when bound to target RNA or the use of fluorescently tagged RNA-binding proteins [Politz et al., 1998]. However, none of those approaches provide information about the behavior of a single RNA molecule in real-time in living cells. An innovative approach was developed inserting a series of RNA aptamers–stem-loops into the transcript of interest. This RNA is then detected by high affinity specific interaction between RNA stem-loops and fluorescently labeled bacteriophage coat protein MS2. Therefore, the RNA detection system is comprised of two elements, a tag protein/peptide fused to the MS2 and a reporter RNA containing MS2-binding sites repeated in tandem. MS2 binds specifically to an RNA stem–loop structure consisting of 19 nucleotides (Fig. IV.7). This method allows to detect specific RNA in order to study RNA dynamics (i.e. biogenesis, processing, export) *in vivo*. This technique can be used in broad range of applications such as visualization and kinetics of RNA in single cells by tagging MS2 with a fluorescent mark; or affinity purification and further proteomic studies of endogenously assembled RNA-protein complexes in order to identify new factors involved in RNA

regulation.

The MS2-based system has been effectively used to tracking specific cytoplasmic and nuclear mRNA in yeast, *Dictyostelium*, plants, flies and mammalian cells. For instance MS2-GFP has been used to demonstrate that *ASH1* mRNA is localized to the bud tip in *S. cerevisiae* in a zip-code-dependent manner [Bertrand et al., 1998]. Zip-codes are *cis*-acting elements often placed in 3'UTRs of mRNAs that specify the spatial information for mRNA targeting to a certain subcellular location. Proper *ASH1* mRNA targeting required the SHE protein complex which is homologous to mammalian myosin motor proteins. Time-lapse imaging revealed that labeled mRNAs move to the bud tip at a rate of 0.2-0.44 µm/s in a directed fashion [Bertrand et al., 1998]. Using the MS2-system in living mammalian cells, it was possible to demonstrate that cytoplasmic mRNAs exhibit directed, corralled, diffusive and static movements [Fusco et al., 2003]. The *β-actin* zip-code increased the number and persistence of these directed movements relative to a non-zip-code containing control reporter. Single particle tracking and time-lapse imaging revealed directed mRNA movements along microtubules at an average rate of 1–1.5 µm/s. The MS2-system has also been used for the detection and kinetic studies of nuclear events such as transcription sites [Janicki et al., 2004]. Visualization and trafficking of single mRNP in the nucleus revealed that diffusion is the primary mechanism by which these molecules translocate from transcription site to nuclear

periphery [Shav-Tal et al., 2004]. The mRNP particles moved an overage of 5µm at velocities ranging from 0.03 to 0.08 µm /s. These movements were not affected by metabolic inhibitors of active transport and are consistent with diffusional rates. MS2-based approaches allowed the study of the dynamics of RNAPII in vivo [Darzacq et al., 2007]. Photobleaching and photoactivation of fluorescent MS2 proteins used to label nascent messenger RNAs provided sensitive elongation measurements. It was shown that RNAPII elongated at 4.3 kb/min on a synthetic gene. These data were consistent with elongation rates measured on integrated HIV-1 genes [Boireau et al., 2007]. Photobleaching of nascent HIV RNAs labeled with MS2-GFP revealed that the elongation rate of RNAPII was approximately 1.9 kb/min. Alternatively, the MS2 system can be converted from visualization tool to affinity purification coupled with proteomic studies. MS2 fused to the maltose binding protein (MBP) has been used to purify the spliceosome by affinity chromatography of cellular extracts [Zhou et al., 2002]. Thanks to this approach 145 distinct spliceosomal proteins were identified, making the spliceosome the most complex cellular machine so far characterized.

In conclusion, RNA detection by the MS2 system coupled with live cell imaging technologies is a powerful approach which allows the analysis of mRNA biogenesis in single cell in real time. Alternatively, MS2 system coupled with high affinity chromatography and mass

spectrometry-based techniques permits analysis of proteome with the identification of new components of RNA-protein complexes *in vivo*.

IV.4 The Green Fluorescent Protein

Green fluorescent protein was discovered by Shimomura et al. in 1962 in *Aequorea* jellyfish as a companion to the chemiluminescent protein aequorin [Shimomura et al., 1962]. Shimomura et al. also accurately described the appearance of GFP solutions: “a protein giving solution that look slightly greenish in sunlight through only yellowish under tungsten lights, and exhibiting a very bright greenish fluorescence in the ultraviolet of a Mineralite”. In the same year the same group showed that the emission spectrum of GFP has a maximum at 508 nm [Johnson et al., 1962]. They noted that the green bioluminiscence of living *Aequorea* tissue also peaked near this wavelength, whereas the chemiluminescence of pure aequorin was blue and peaked near 470 nm, which was close to one of the excitation peaks of GFP. Therefore the GFP converted the blue emission of aequorin to the green glow of the intact cells and animals. Morise et al. in 1974 purified and crystallized GFP, measured its absorbance

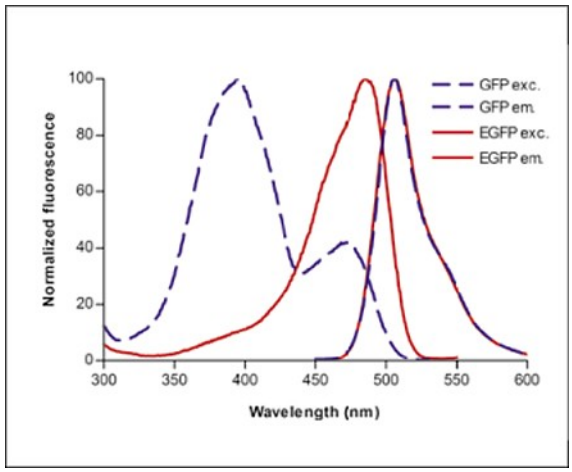

Figure IV.8: Two Aequorea victoria GFP variants were examined for their different excitation maxima

spectrum and fluorescence quantum yield, and showed that aequorin could efficiently transfer its luminescence energy to GFP when the two were coabsorbed onto cationic support [Morise et al., 1974]. The first estimate for the monomer molecular weight, 27 kD, was obtained by Prendergast and Mann in 1978 [Prendergast and Mann, 1978]. Fundamental historical landmarks came more than 30 years after the discovery of GFP, when in 1992 Prasher et al. cloned the 966 nucleotides long *gfp* gene [Prasher et al., 1992] and in 1994 Chalfie et al. [Chalfie et al., 1994] and Inouye and Tsuji [Inouye and Tsuji, 1994a] demonstrated that the protein expressed from this gene is fluorescent also in other organisms. Therefore the gene contains all the information necessary for the posttranslational synthesis of the chromophore, and no jellyfish-specific enzymes are needed [Tsien, 1998a].

IV.4.1 GFP structure and folding

GFP has a unique 11 β-sheet barrel-like structure with a diameter of about 24 Å and a height of 42 Å (Fig. IV.9). The β-sheets form the walls of a can, and α-helix runs diagonally through the can. The chromophore is in the center of the 11 β-sheet and is linked by the α-helical stretch that runs through the center of the barrel. The chromophore, a p-hydroxybenzylidebeimidazolinone, formed by an intramolecular cyclization of 65Ser-Tyr-Gly67, is well protected in the center of the barrel [Prasher et al., 1992][Cody et al., 1993]. By enclosing the chromophore in the can, it may be protected from quenching by oxygen

and attack by hydronium ions [Rao et al., 1980]. The GFP folding into the 11-strand β-barrel is probably the most crucial event for the formation of the chromophore and its bioluminescence. The formation is autocatalytic, and the only external requirement is the presence of oxygen [Chalfie et al., 1994]. The β-can structure protects the chromophore and is presumably responsible for GFP stability [Rao et al., 1980]. GFP can be reversibly denatured. Fluorescence is completely lost in the denatured GFP but is regained when the β-can structure is reformed [Ward and Bokman, 1982]. The onset of fluorescence can therefore be used as an indicator that the 11-strand β-barrel has been formed. GFP fluorescence is not observed until 90 min to 4 h after protein synthesis [Heim et al., 1994]. The protein folds quickly, but the subsequent fluorophore formation and oxidation is slow [Reid and Flynn, 1997].

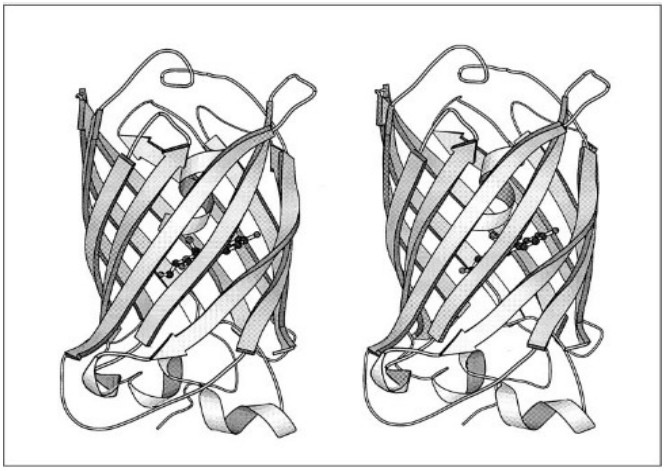

Figure IV.9: Stereoview of three-dimensional structure of GFP [Ormoe et al., 1996].

IV.4.2 GFP optimizations and spectral variants

Aequorea is found in the cold Pacific Northwest, and mature fluorescent GFP is most efficiently formed at temperatures well below 37 °C [Zimmer, 2002]. This has limited the uses of GFP and has led to the search for mutants that mature more efficiently at higher

temperature. Mutation of Phe64 to Leu results in improved maturation of fluorescence at 37 °C that is almost equivalent to maturation at 28 °C [Cormack et al., 1996].

Several modification were made to the original GFP to make it suitable for mammalian system,. The optimization of codon usage for mammalian expression has also improved overall brightness of GFP expressed in mammalian cells [Yang et al., 1996]. Over 190 silent mutations were introduced into the coding sequence to optimize expression in human tissues. A Kozak translation initiation was also introduced by insertion of a Val as the second amino acid [Kozak, 1989]. These improvements have resulted in a very useful probe for live cell imaging of mammalian cells and are common to all of the currently used FPs derived from *Aequorea* GFP [Goldman and Spector, 2005].

Although spectral variants with blue, cyan and yellowish-green emissions have been successfully generated from the *Aequorea* GFP, none exhibit emission maxima longer than 529 nm. The currently known GFP variants may be divided into seven classes based on the distinctive components of their chromophores [Tsien, 1998][Zimmer, 2002]:

1. *Wild-type GFP.* The chromophore is in an equilibrium between the phenol and phenolate form. It has a major excitation peaks at 395 nm and a minor peak at 475 nm.
2. *Phenolate anion (EGFP).* Ser65 has been substituted with Thr, Ala, or Gly. Does not have the 395 nm excitation peak.

3. *Neutral phenol (Sapphire).* Mutation of Thr203 to Ile almost suppresses the 475 nm excitation peak, leaving only the shorter wavelength peak at 399 nm. The emission peak is at 511 nm.
4. *Phenolate anion with stacked π-electron system (yellow fluorescent proteins).* Mutation of Thr203 to His, Trp, Phe or Tyr results in yellow fluorescent proteins. The YFPs have the most red-shifted absorption of all GFP mutant. YFPs have lower fluorescent quantum yield than GFPs, especially when excited at 475 nm.

   An enhanced YFP mutant, the E^2GFP, has been obtained that might be used as a single-biomolecule optical switch. As in most GFP molecules, prolonged or intense excitation, for E^2GFP at 476nm, results in photobleaching. However, in E^2GFP irradiation of the dark photobleached state at 350nm forms an excited state that photoconverts to a form which is fluorescent [Cinelli et al., 2001].
5. *Indole in chomophore derived from Y66W (cyan fluorescent proteins).* Substitution of Trp for Tyr66 produces a new chromophore with an indole instead of a phenol or phenolate. Excitation and emission wavelength are 436 and 476 nm, intermediate between neutral phenol and anionic phenolate chromophore. These protein are called cyan fluorescent proteins (CFPs) because their blue-green or cyan emission.
6. *Imidazole in chromophore derived from Y66H (blue fluorescent*

protein). Substitution of His for Tyr 66 shifts the wavelength yet shorter than Trp66. The excitation and emission peaks are around 383 and 447 nm, then the emission is blue. BFP suffers from a relatively low fluorescence quantum yield and is relatively easily bleached.

7. *Phenyl in chromophore derived from Y66F.* The shortest wavelength are obtained with Phe at 66.

The absorbance and fluorescent properties of the different classes are summarized in Table 1.

Mutation	Common name	λ_{exc} [nm]	λ_{em} [nm]
Class 1, wild-ype			
None or Q80R	Wild Type	395-397	504
		470-475	
F99S, M153T, V163A	Cycle 3	397	506
		475	
Class 2, phenolate anion			
S65T		489	509-511
F64L, S65T	EGFP	488	507-509
F64L, S65T, V163A		488	511
S65T, S72A, N149K, M153, I167T	Emerald	487	509
Class 3, neutral phenol			
S20F, T203I	H9	399	511
T203I, S72A, Y145F	H9-40	399	511
Class 4, phenolate anion with stacked π-electron system (YFP)			
S65G, S72A, T203F		512	522
S65G, S72A, T203H		508	518
S65G, V68L, Q69K, S72A, T203Y	10C Q69K	516	529
S65G, V68L, S72A, T203Y	10C	514	527
S65G, S72A, K79R, T203Y	Topaz	514	527
Class 5, indole in chromophore (cyan fluorescent proteins)			
Y66W		436	
Y66W, N146I, M153T, V163A	W7	434	476
		452	505
F64L, S65T, Y66W, N146I, M153T, V163A	W1B or ECFP	434	476
		452	505
S65A, Y66W, S72A, N146I, M153T, V163A	W1C	435	495
Class 6, imidazole in chromophore (blue fluorescent proteins)			
Y66H	BFP	384	448
Y66H,Y145F	P4-3	382	446
F64L, Y66H, Y145F	EBFP	380-383	440-447
Class 7, phenyl in chromophore			
Y66F		360	442

Table 1: Sprectral characterisitcs of the major classes of green fluorescent proteins [Tsien, 1998]. λexc and λem are the peak of the excitation and emission spectrum, respectively, in units of nanometers

IV.4.2.1 Fluorescent proteins not from Aequorea

The discovery of novel GFP-like proteins from *Anthozoa* have significantly expanded the range of colors available for cell biology applications. Despite the modest degree of sequence similarity, these GFP-like proteins probably share a β-can fold structure that is fundamental for the fluorescence of GFP. Proteins that fluoresce at red

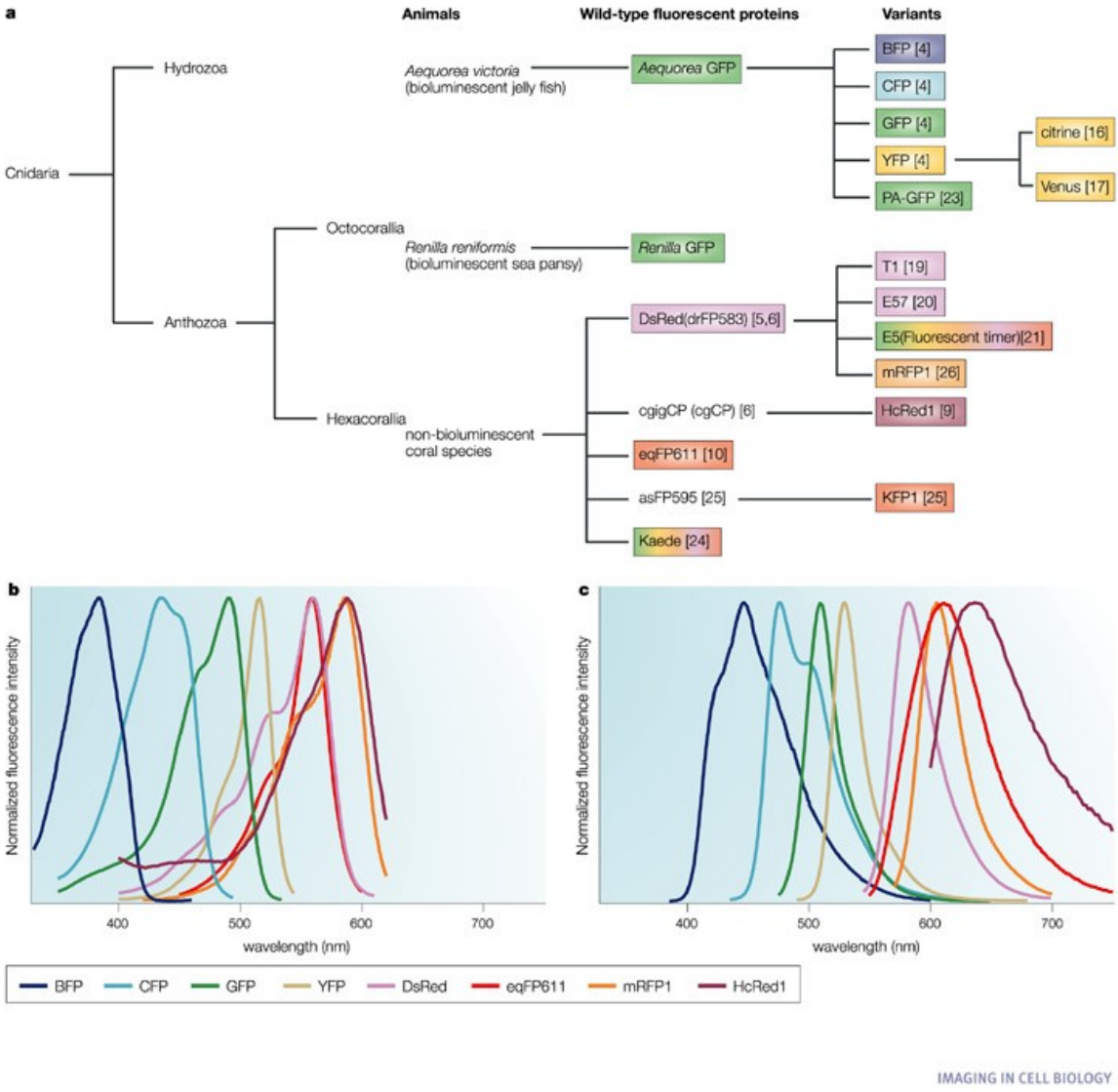

Figure IV.10: Wild-type fluorescent proteins and their spectral variants.
Phylogeny (a) and excitation (b) emission (c) specta of fluorescent proteins. Colors indicate emission maxima [Miyawaki et al., 2003].

or far-red wavelengths (red fluorescent proteins or RFPs) are of specific interest because they can be used in combination with other fluorescent protein that fluoresce at shorter wavelength for both multicolor labeling and fluorescence resonance energy transfer (FRET) experiment [Miyawaki et al., 2003].

IV.4.3 Applications

Fluorescent protein are quite versatile and have been successfully employed in almost every biological discipline to investigate a remarkable array of properties and behaviors.

IV.4.3.1 Reporter gene

The first applications of GFP were as a reporter gene [Chalfie et al., 1994]. Gene expression can be monitored by using a *gfp* gene that is under control of a promoter of interest and measuring during time the GFP fluorescence which directly indicates the gene expression level in living cells or tissue [Zimmer, 2002].

IV.4.3.2 Fusion Tags

The most successful and numerous applications of the GFPs have been as genetic fusion partner to host protein to monitor their localization and fate. The gene encoding the GFP is fused in frame with the gene encoding the protein of interest and the resulting chimera are expressed in the cell or organism. Indeed GFP fusion fusion tags can be used to visualize dynamic cellular events and to monitor protein

localization. GFP has been targeted successfully to almost every major organelle of the cell, including plasma membrane, nucleus, endoplasmic reticulum, Golgi apparatus, secretory vesicles, mitochondria, peroxisomes, vacuoles and phagosomes (reviewed in [Tsien, 1998]).

IV.4.3.3 FRET

Fluorescence resonance energy transfer (FRET) is a nonradiative exchange of energy from an excited donor fluorophore to an acceptor fluorophore that is within 10-100 Å from the donor. In order for FRET to occur, the emission spectrum of the donor has to overlap with the excitation spectrum of the acceptor. The FRET efficiency rapidly decreases with increasing distance between the two fluorophore, the efficiency is inversely proportional nearly to the sixth power of the distance. FRET is a powerful technique for studying protein-protein interactions in vitro and in vivo [Marcello et al., 2001][Marcello et al., 2003][Pantano et al., 2006].

IV.4.3.4 pH sensor

Wild-type GFP and many of its mutant display pH-dependent fluorescent behavior and have been used to monitor pH in *vivo* [Wachter et al., 1997]. A series of pH-sensitive GFP mutant have been developed by structure-direct combinatorial mutagenesis; they are often know as pHluorins [Miesenboeck et al., 1998].

IV.4.3.5 Photobleaching and photoactivation

Photobleaching and photoactivation when combined with time-lapse

imaging can be used to investigate protein dynamics in living cells by making its movement observable. Photobleaching is the photo-induced alteration of a fluorophore that extinguishes its fluorescence. Photoactivation works by converting molecules to a fluorescent state by using a brief pulse of high-intensity irradiation. After fluorescently highlighting specific populations of molecules can be followed as they re-equilibrate in the cell. The extent and rate at which this occurs can be quantified and used with computer-modeling approaches to describe the kinetic parameters of a protein [Lippincott-Schwartz et al., 2003].

IV.4.3.5.1 Photoconversion

An important phenomenon, relevant either in phobleaching and in FRET measurements, is the photoconversion of YFP in CFP-like-species after photobleaching [Valentin et al., 2005]. This process initially controversial [Thaler et al., 2006] has been later confirmed [Kirber et al., 2007].

IV.5 Fluorescence Recovery After Photobleaching

Fluorescence recovery after photobleaching (FRAP) is a method for measuring the bi-dimensional or three-dimensional mobility of fluorescent particles. A defined portion of the system containing mobile fluorescent molecules is exposed to a brief and intense focused laser beam, thereby causing irreversible, or almost irreversible, photochemical bleaching of the fluorophore in that region. The

subsequent kinetics of fluorescence recovery in the bleached region, which results from transport of fluorophore into bleached area from unirradiated parts of the system, provides a quantitative measure of fluorophore mobility [Axelrod et al., 1976].

The use of GFP, and principally of its engineered mutants with improved brightness, photostability and expression properties, as minimally invasive tool, allows the study of protein dynamics and function in vivo. Cells that express protein tagged with these GFPs can be imaged with low light intensity over many hours and can provide useful information about changes in protein localization and steady-state level over time. Time-lapse imaging alone, however, cannot reveal some of the arguably most interesting and informative features of proteins, their kinetic properties. FRAP is able to investigate the protein kinetic that underlie protein function within cells.

Before the bleach the concentration of fluorescent protein should be in equilibrium, at the bleach there isn't any change in physical or chemical properties of the protein tagged with the fluorophore, but the tagged protein amount is divided in two families, one set is dark and the other is still fluorescent. This artificial gradient of fluorescent protein concentration, that is not a true protein of interest concentration gradient, drives the rescue of the equilibrium by diffusion. Analysis of fluorescence recovery can be used to determine the kinetic parameters of a protein, including its diffusion constant, mobile fraction, transport

rate or binding/dissociation rate from other proteins [Lippincott-Schwartz et al., 2003]. When binding interactions are present, they retard a FRAP recovery in relation to what would be observed if only diffusion occurred. Indeed a bleached protein have to unbound its binding site before to be free to leave the bleached area. Usually GFP is considerate a good standard marker of mobility in absence of binding. GFP diffusion constant were estimated to be about 20 $\mu m^2/s$ in cytoplasm or nucleus [Sprague and McNally, 2005].

IV.5.1 FRAP model

In order to obtain quantitative informations from FRAP experiments, it is fundamental to develop a robust analysis of the physical phenomena involved. Diffusion and reaction, coupled or not, are the principal process of the fluorescence recovery after photobleaching.

IV.5.1.1 Diffusion

Diffusion is the net transport of molecules from a region of higher concentration to one of lower concentration, as a result of their kinetic energy of random motion. There will be a net flow of molecules due to the molecular motion only when there is a nonuniform distribution of the molecules, otherwise all the molecular motions would average to give no net flow. In a FRAP experiment the bleach produces a nonuniform distribution of the fluorescent proteins that riequilibrate by diffusion.

IV.5.1.1.1 Fick's laws

The first Fick's law (Equation 1) relates the diffusive flux of

molecules/proteins, J, to the concentration gradient, ∇f. It defines the diffusion coefficient, D, as the proportionality constant between flux and concentration gradient.

The minus sign in the formula reflects that the diffusive flux goes from higher concentration to lower concentration regions. In ideal conditions the Stoke-Einstein equation, $D=k_BT/(6\pi\eta R)$, where k_B is the Boltzmann constant, T the absolute temperature, η the medium viscosity and R the radius of the molecules, predicts the diffusion coefficient for

$$\vec{J}=-D\cdot\vec{\nabla}f$$

Equation 1: First Fuck's law

spherical or pseudo-spherical molecules. This relationship along with the rough approximation that the molecule radius is proportional to the cubic root of the molecule mass, possibly estimate the ratio of the diffusion coefficients of two specie, a and b, only in function of their mass, $D_a/D_b=(M_b/M_a)^{1/3}$.

The second Fick's law (Equation 2), or diffusion equation, that can be derived from the first law for mass conservation, describes how diffusion affects concentration in time. The diffusion equations have been solved

$$\frac{\partial f}{\partial t}=D\cdot\nabla^2 f$$

Equation 2: Second Fick's law

for FRAP with a uniform rectangular or circular bleach profile and also with a Gaussian-shaped bleach profile, including one, two and three-dimensional geometries [Axelrod et al., 1976][Soumpasis, 1983]

[Braeckmans et al., 2003]. Three-dimensional solutions of the diffusion equation has been proposed for a Gaussian-shaped bleach profile or taking into account diffusion during bleach [Braga et al., 2004] or based on multiphoton excitation [Mazza et al., 2008].

Equation 3 is a closed form diffusion equation solution in two dimension and for a circular uniformly bleached spot in an infinite domain, involving modified Bessel functions [Sprague et al., 2004].

$$frap(t) = e^{-\frac{\tau_D}{2t}}\left[I_0\left(\frac{\tau_D}{2t}\right) + I_1\left(\frac{\tau_D}{2t}\right)\right]$$

Equation 3: Diffusion equation solution

This equation describe how the fluorescence in the circular bleached area, frap(t), evolves after the bleach. In the formula $\tau_D = w^2/D$, where w is the bleached area radius and D is the diffusion coefficient of the fluorescent specie.

IV.5.1.2 Reaction-Diffusion

A general binding reaction can be represented as equation 4, where F is the free protein, S the vacant binding site, C the bound complex and

$$F + S \underset{K_{off}}{\overset{K_{on}}{\rightleftarrows}} C$$

Equation 4: Reaction equation

K_{on} and K_{off} are the on and off binding rates. In order to describe the reaction-diffusion behavior is necessary to link the reaction equation

(Equation 4) with the diffusion equation (Equation 2), as in equations system 5 [Sprague et al., 2004].

$$\frac{\partial f}{\partial t} = D_f \nabla^2 f - K_{on} f s + K_{off} c$$
$$\frac{\partial s}{\partial t} = D_s \nabla^2 s - K_{on} f s + K_{off} c$$
$$\frac{\partial c}{\partial t} = D_c \nabla^2 c + K_{on} f s - K_{off} c$$

Equation 5: Reaction-Diffusion equations system

In the reaction-diffusion equations system (Equation 5) f, s and c represent the concentration of free protein, of vacant binding sites and of bound complexes respectively; D_f, D_s and D_c are the diffusion coefficient of the free protein, of the binding sites and of the complexes respectively.

Two assumptions, applicable in many biological situations, can considerably simplify this equations system[Sprague et al., 2004].

The first simplifying assumption is that the biological system has reached equilibrium before photobleaching. This means that the total amount of both fluorescent fusion protein and its binding sites remains constant over the time course of the fluorescence recovery. Therefore it is possible to assume that the system is at the equilibrium and denote the corresponding equilibrium concentrations of F, S, and C by F_{eq}, S_{eq}, and C_{eq}. Although the act of bleaching changes the number of visible free and complexed molecules (F or C), it does not change the number of free binding sites. Therefore $s=S_{eq}$ is a constant throughout the

photobleaching recovery. This eliminates the second equation of the equations system 5 and also makes possible to replace the variable s in the remaining two equation with a constant S_{eq}. As a result, it is possible define a pseudo-first-order rate constant given by $K_{on}S_{eq}=K^{*}_{on}$.

The second simplifying assumption is that the binding sites are part of a large, relatively immobile complex, at least on the time- and length-scale of the FRAP measurement. Ignoring diffusion of the bound complex results in $D_c=0$ in the equations system 5.

With these two assumptions, equations system 5 reduce to equations system 6 [Sprague et al., 2004].

$$\frac{\partial f}{\partial t}=D_f \nabla^2 f - K^{*}_{on} s + K_{off} c$$
$$\frac{\partial c}{\partial t}=K^{*}_{on} f - K_{off} c$$

Equation 6: Simplified Reaction-Diffusion equations system

The measured FRAP recovery data are the sum of free and bound fluorescence, averaged over the bleach spot: frap(t)=avg[f(t)]+avg[c(t)]. Normalize the FRAP recovery as in equation 7 to the initial equilibrium condition is equivalent to fix that $F_{eq}+C_{eq}=1$.

$$frap(t)=\frac{I(t)\cdot T(t_0)}{I(t_0)\cdot T(t)}$$

Equation 7: Experimental FRAP curve normalization

In equation 7 I(t) is the average intensity of the region of interest at time point t, $I(t_0)$ is the average pre-bleach intensity of the region of

interest and $T(t_0)$ and $T(t)$ are the average total area fluorescence intensities at pre-bleach or at time point t, respectively [Phair and Misteli, 2000]. This expression accounts for the loss in total intensity caused by the scanning and the bleach itself. The consequence of either this normalization and the equilibrium condition in equations system 7, $\partial f/\partial t=\partial c/\partial t=0$, allow to obtain the value of $F_{eq}$and C_{eq} (Equation 8).

$$F_{eq}=\frac{K_{off}}{K^{*}_{on}+K_{off}}$$
$$C_{eq}=\frac{K^{*}_{on}}{K^{*}_{on}+K_{off}}$$

Equation 8: Equilibrium concentrations

A solution for the reaction-diffusion equations system (Equation 6) in 2 dimension and for an uniform circular bleach spot was proposed by Sprague et al. in 2004. Equation 9 is the average of the Laplace transform of the fluorescent intensity within the bleach spot [Sprague et al., 2004]; I_1 and K_1 are modified Bessel functions of the first ans second kind. They assume that the binding sites are homogeneously distributed and that the bleached spot is small relative to the size of fluorescent compartment. The 2D simplification is appropriate when the bleaching area forms a near-cylindric shape through the cell, as occurs for a circular bleach spot of reasonable diameter with a relatively low numerical aperture objective [Braeckmans et al., 2003]. In this case, axial terms disappear from Laplacian in equation 6 and only the radial

component remains [Sprague et al., 2004].

In the same conditions and with the same assumptions Sprague et al. propose also a reaction-diffusion model with two independent binding interactions [Sprague et al., 2004].

$$\overline{frap}(p)=\frac{1}{p}-\frac{F_{eq}}{p}(1-2K_1(qw)I_1(qw))\times \times(1+\frac{K^*_{on}}{p+K_{off}})-\frac{C_{eq}}{p+K_{off}};$$

$$q=\sqrt{(\frac{p}{D_f})(1+\frac{K^*_{on}}{p+K_{off}})}$$

Equation 9: Solutions of the reaction-diffusion equations system.Solution proposed by Sprague [Sprague et al., 2004].

IV.5.1.2.1 Effective diffusion

A simplified case for equations system 6 arise when the reaction process is much faster than diffusion. This implies that at any location within the bleach spot, the binding reaction rapidly achieves a local equilibrium. Under this condition the reaction-diffusion equations system 6 reduce to a simple diffusion equation (Equation 2) but recover at a slower rate determined by the strength of binding. In this case it is possible to fit the experimental data with equation 3, but the diffusion constant measured is an effective diffusion constant, D_{eff} (Equation 10), lower than the diffusion constant of the protein in absence of binding, D [Sprague et al., 2004].

$$D_{eff} = \frac{D}{1 + K_{on}^{*} / K_{off}}$$

Equation 10: Effective diffusion coefficient

IV.5.1.2.2 Reaction dominant

Another simplified scenario arises when diffusion is very fast compared both to binding and to timescale of the FRAP experiment. Free proteins immediately equilibrate after the bleach, so that diffusion is not detect in the FRAP recovery. Under this condition the solution of the reaction-diffusion equations system 6 reduces to equation 11 [Sprague et al., 2004].

$$frap(t) = 1 - C_{eq} e^{-K_{off} t}$$

Equation 11: Reaction dominant solution.

IV.5.1.3 FRAP model optimization

It is possible to solve the reaction-diffusion system also with less assumptions.

Beaudouin et al. propose a FRAP three-dimensional model taking into account an experimental nonuniform initial distribution of binding sites. They solve it numerically using a finite difference approach in a finite volume of the dimension of a typical cell nucleus, independently of the bleaching geometry [Beaudouin et al., 2006].

Braga et al. and Kang and Kenworthy solved the reaction-diffusion system taking into account that also the complexes can diffuse with a specific diffusion coefficient and solve the equations for a Gaussian-shaped bleach profile in a finite circular domain or in an infinity domain

respectively [Braga et al., 2007][Kang and Kenworthy, 2008].

Mueller et al. revisited the model proposed by Sprague et al. [Sprague et al., 2004] and derived a new FRAP model, solving the reaction-diffusion equations system for a Gaussian-shaped bleach profile in a finite domain [Mueller et al., 2008].

IV.5.2 FRAP variants

Alternative FRAP methods were developed in order to investigate the kinetic properties of proteins of interest in the most advantageous and informative conditions [Lippincott-Schwartz et al., 2003].

IV.5.2.1.1 iFRAP

Inverse FRAP (iFRAP) is performed as a normal FRAP experiment but in this case the molecules outside the region of interest are photobleached and the loss of fluorescence from the non-photobleached region is monitored over time.

IV.5.2.1.2 Photoactivation

A method complementary to iFRAP is the utilization of photoactivatable fluorescent proteins (PA-GFP), that allow to swich-on the fluorescence in a specific region and after follow its reallocation in the cell.

IV.5.2.1.3 FLIP

In fluorescence loss in photobleaching (FLIP) technique a fluorescent cell is repeatedly photobleached within a small region while the whole cell is continuously imaged. Any compartment of the cell that is

connected to the area being bleached will gradually lose fluorescence due to movement of mobile bleached proteins from the bleached area to the other compartments. By contrast, the fluorescence in unconnected regions will not be affected.

V. RESULTS

To analyze the biogenesis of HIV-1 mRNA, we developed two different HIV vectors carrying 24 MS2-binding sites: pExo-MS2x24 and pIntro-MS2x24 (Fig. V.1). In the pExo-MS2x24 vector the 24 MS2 repeats are in the 3' untranslated region; in the pIntro-MS2x24 vector they are within the intron. Each vector has the elements required for RNA production: the 5' LTR, the major splice donor (SD1), the packaging signal Ψ, the Rev-responsive element (RRE), the splice acceptor A7 flanked by its regulatory sequences (exonic splicing enhancers and silencers), and the 3' LTR that drives 3'-end formation. Moreover the pIntro-MS2x24 vector carries a reporter of gene expression (ECFP) fused to the peroxisome localization signal Ser-Lys-Leu (skl) and the selectable marker thymidine kinase (TK) of herpes simplex type 1.

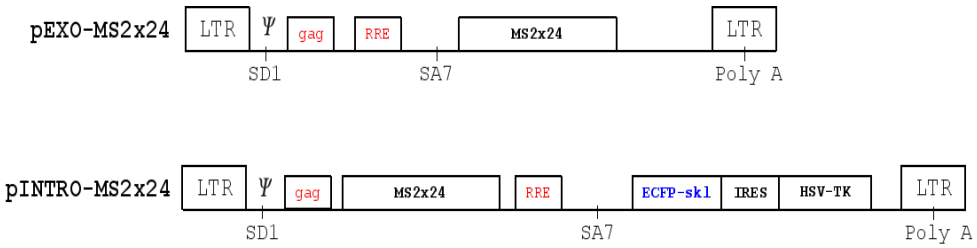

Figure V.1: p-Exo and P-Into HIV-1 reporters

V.1 pExo-MS2x24 array

Stable arrays of the pExo-MS2x24 construct were integrated in U2OS cells [Molle et al., 2007] [Boireau et al., 2007]. Two clones (Exo1 and Exo2) showed robust trans-activation by Tat and other stimuli (PMA/ionomycin) known to induce transcription of integrated HIV-1 promoters. When expressed, the RNAs were distributed homogeneously in the cytoplasm and concentrated in a bright spot in the nucleoplasm. This spot corresponds to the transcription site as it colocalized with RNA-Poll-II/P-TEFP and was labeled with probes directed against the non-transcribed strand of the vector (Fig. V.2). To visualize nascent HIV-1 RNAs, a nuclear MS2-GFP-nls fusion protein was transiently expressed in U2OS-Exo1 and U2OS-Exo2 cells. In live cells, MS2-GFP-nls was diffused in the nucleoplasm and concentrated in a spot at the transcription site (Fig. V.3).

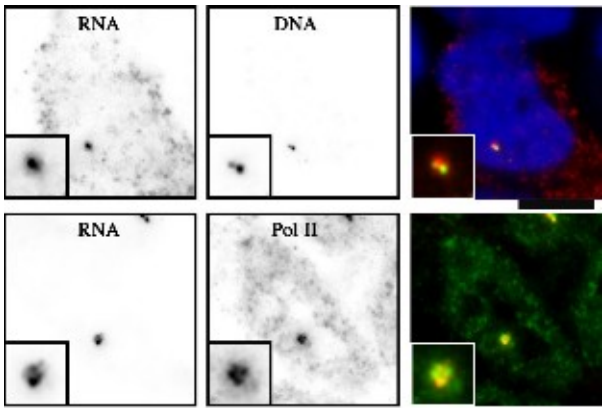

Figure V.2: RNA-DNA and RNA-PolII colocalization at the transcription spot

Figure V.3: Exo1 transfected with MS2-GFP-nls

V.1.1 Kinetic analysis of the transcription site

Fluorescence recovery after photobleaching (FRAP) [Axelrod et al., 1976] is a powerful technique to study the dynamic properties of a fluorescent molecule, and we used it to study mRNA synthesis by photobleaching the nascent RNAs labeled with MS2-GFP-nls. When a transcription site is bleached, incoming polymerases will synthesize new MS2 sites, and this will result in recovery of the fluorescence signal.

V.1.1.1 Transcription Spot

Diffusion of MS2-GFP-nls is rapid (15 μm²/s) and much faster than recovery at the transcription sites (Fig. V.4). For this reason the MS2-GFP-nls diffusion was neglected in the analysis of FRAP experiments. Recovery of Exo1 transcription sites (Fig. V.4) showed that synthesis of new RNAs occurred in < 3 min and that virtually no RNAs were retained at the transcription site for a long time (immobile fraction of 3%). A second clone expressing the same MS2-tagged RNA, Exo2, yielded similar kinetics (Fig. V.5), indicating that this was principally a property of the reporter. RNA biogenesis occurs in a series of steps: transcription, 3'-end formation (cleavage and

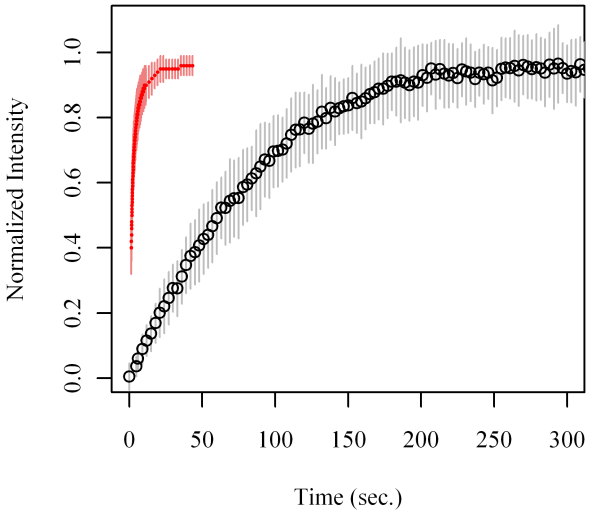

Figure V.4: Exo1 (Black) and MS2-GFP-nls (Red) recovery curve.

polyadenylation), and release from the transcription site (Fig. V.6). To test whether elongation was rate limiting, we used a slow version of RNA-Pol-II, the hC4 mutant [Mata et al., 2003]. This α-amanitin resistant mutant was transfected in U2OS-Exo1 cells, the endogenous enzymes were inactivated with α-amanitin, and the remaining transcription sites were analyzed by FRAP. As a control, we used another α-amanitin resistant version of RNA-Pol-II that was not defective (wild type [WT^{res}]; [Mata et al., 2003]). FRAP curves obtained with WT^{res} or the endogenous polymerase were nearly identical (Fig. V.7). In contrast, recoveries were markedly slower with the hC4 mutant: 129 s for half-recovery

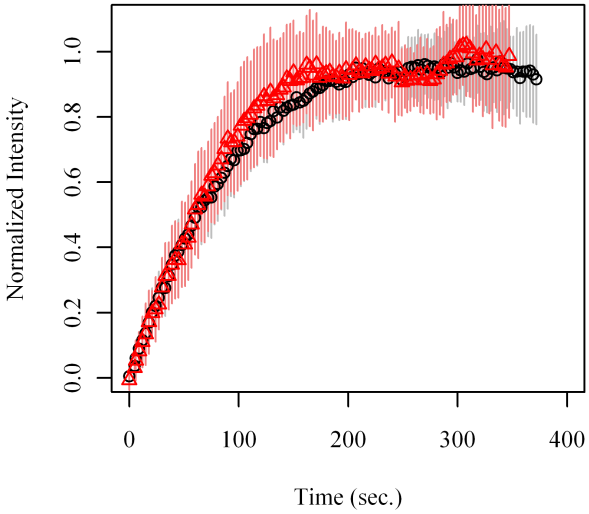

Figure V.5: Exo1 (Black) and Exo2 (Red) recovery curve.

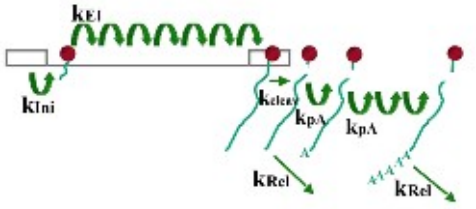

Figure V.6: Transcription and RNA processing

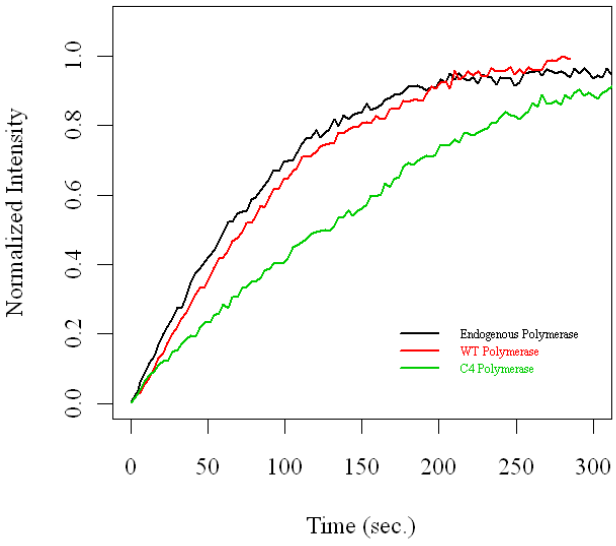

Figure V.7: Exo1 recovery curve with different RNAPII.

versus 75 s for the WT^{res} enzyme. To confirm that transcription elongation took a substantial part of the time that nascent RNAs spent at their transcription sites, we used camptothecin. This drug targets topoisomerase-I, and it transiently cross-links it to DNA. A previous study has shown that this creates steric road blocks for polymerases, which result in increased pausing and a net slowdown of transcription elongation [Capranico et al., 2007]. Indeed, when cells were treated with this drug, the recovery of nascent RNA was slower: 170 s for half-recovery in Exo1 cells compared with 65 s in the case of untreated cells (Fig. V.8). To further establish that elongation was rate limiting, we developed a cell line, Exo-long, wich differed by 2.4 Kb in the length of the sequence separating the MS2 repeats and the end of the gene. If

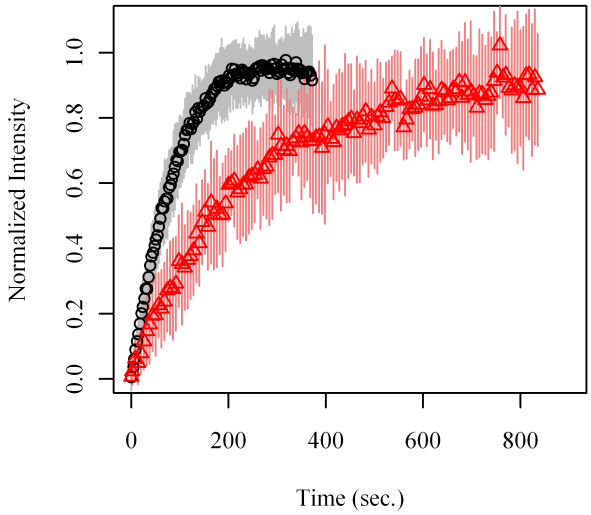

Figure V.8: Exo1 (Black) and Exo1 with Camptothecin (Red) recovery curve.

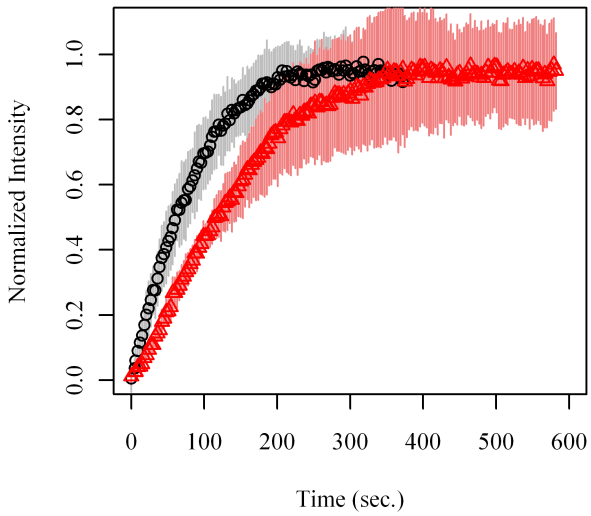

Figure V.9: Exo1 and Exo-Long recovery cureves.

elongation is rate limiting, the recovery should be slower for the long reporter, and the difference should represent the time taken to transcribe the additional sequences. As expected, transcription sites of Exo-long recovered more slowly: 126 s for half-recovery versus 65 s for the short reporter (Fig. V.9). Altogether, these observations demonstrated that elongation took a substantial fraction of the time dedicated to mRNA production.

V.1.1.2 A two step model for elongation and 3'-end processing

Transcription elongation is the repetition of an elementary step: the transcription of one nucleotide. The stochastic basis of the process, that is a typical Poisson process with a constant rate k_{el}, implies that the time taken to add a single nucleotide, or the time between two independent events, is variable and follows an exponential distribution. This distribution is described by the rate constant k_{el} and has an expected value of $1/k_{el}$. When polymerases transcribe a large number of nucleotides, the repetition of the elementary step creates a statistical average such that the time taken to synthesize n nucleotides will be virtually constant for all polymerases and equal to n/k_{el}. This consideration on the nature of the transcription process allows us to predict a linear increase in signal during FRAP recovery. Elongation is the first step of mRNA biogenesis that can be observed in our FRAP experiment. Indeed, once the polymerase reaches the PolyA site, the pre-mRNA is cleaved, polyadenylated and released. These precesses

occur as a series of single steps and, thus, may be modeled as exponentials. We attempted to fit the FRAP curve with two components: a straight line for elongation followed by a single exponential for 3'-end formation and release, equation 12, where t_{elong} is the time point at which the line converts to exponential. Since in equation 12 the numbers of constants were overestimate we reformulated equation 12 in equation 13, assuming the continuity of both the function and its first derivative. In equation 13 A is the function value at t=0, M the mobile fraction and α is the slope of the initial linear component.

$$f(t)=\begin{cases} A+\alpha\cdot t, & t\le t_{elong} \\ B+C[1-e^{-\beta}(t-t_{elong})] & t>t_{elong} \end{cases}$$

Equation 12: Basic linear plus exponential model

$$f(t)=\begin{cases} A+\alpha\cdot t, & t\le t_{elong} \\ A+\alpha\cdot t_{elong}+(M-A-\alpha\cdot t_{elong})[1-e^{-\alpha\frac{(t-t_{elong})}{M-A-\alpha\cdot t_{elong}}}], & t>t_{elong} \end{cases}$$

Equation 13: Linear plus exponential model

This two-step model (Equation 13) fitted well the experimental data (Figs. V.10 and V.12). The short reporter gave 65 to 73 s for elongation and 44 to 31 s for the half-time of the exponential in Exo1 and Exo2 cells, respectively

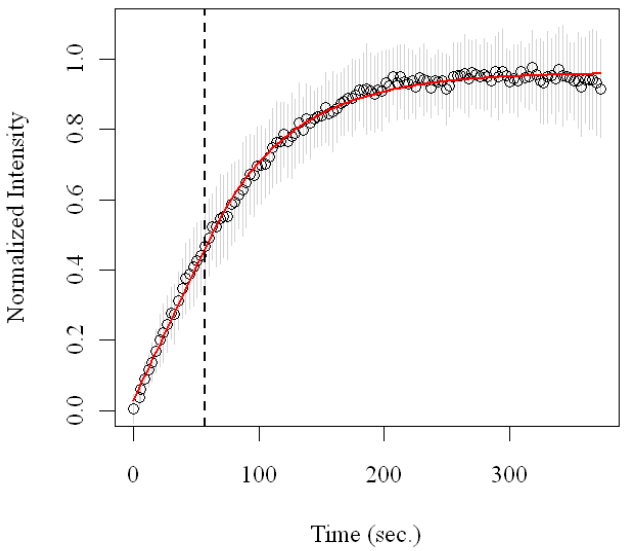

Figure V.10: Exo1 two step model fit.

(Table 2). The long reporter yielded 148 s for elongation and 65 s for the exponential (Fig. V.11). This translated into similar elongation rates of 1.79–2.03 kb/min. Cells treated with camptothecin or transfected with the slow mutant of RNAPII could also be fitted to this two-step model (Figs. V.13 and V.14). For the hC4 mutant, it yielded an elongation rate more than twofold slower than with WT RNAPII (0.8 kb/min), which is in agreement with in vitro data [Coulter and Greenleaf, 1985].

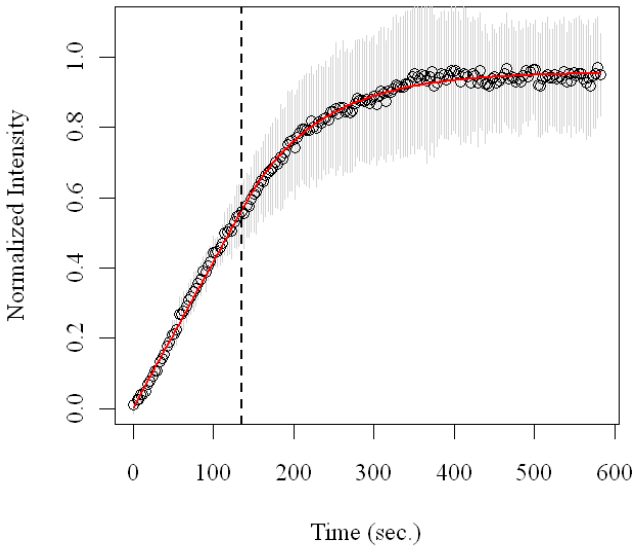

Figure V.11: Exo-Long two step model fit.

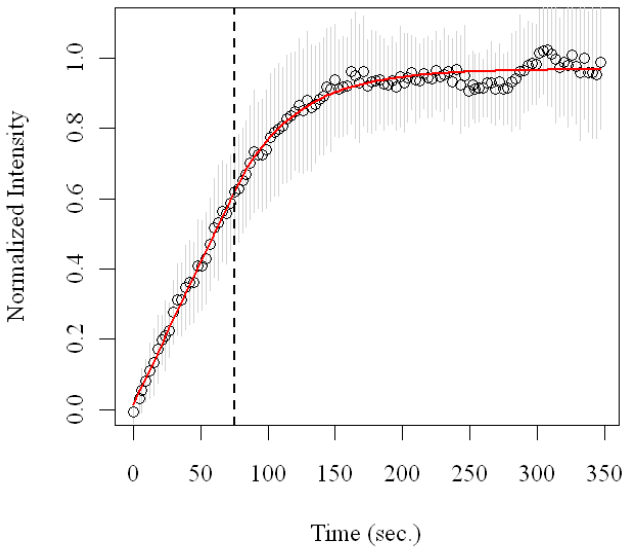

Figure V.12: Exo2 two step model fit

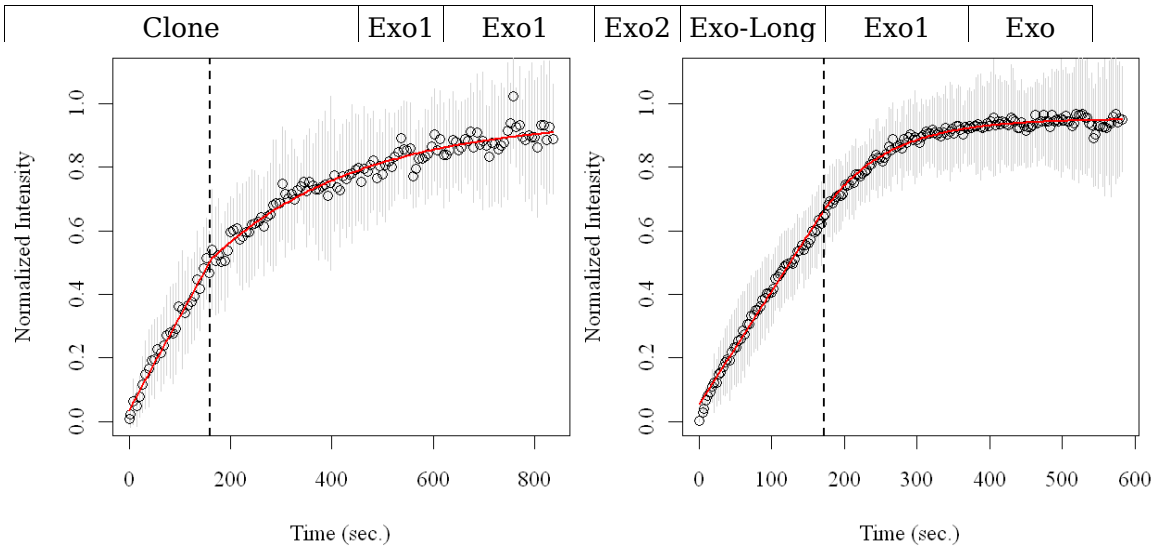

Figure V.13: Exo1 Camptothecin.

Figure V.14: Exo1 Pol-hC4

Table 2: Two step model fits results

V.1.1.3 Pol-II

Although the MS2-GFP FRAP assay provides a direct measurement of RNAPII activity, it provides little information on the events that occur either before the polymerase reaches the MS2 repeat or after it releases its mRNA. To gain a complete view of the transcription cycle, we repeated the FRAP assay with fluorescent subunits of RNAPII. HIV-1 transcription sites were identified with a red variant of MS2 (MS2-mCherry), and GFP-tagged subunit C of RNAPII was bleached in the nucleoplasm or at HIV-1 transcription sites. Recoveries at the HIV-1 transcription site were much slower than in the nucleoplasm, suggesting that most polymerases present at the HIV-1 gene array were engaged in transcription (Fig. V.15). Interestingly, recovery of the polymerase was substantially slower than recovery of nascent RNAs: in Exo1 cells, half-recovery took 200 s for the polymerase but only 66 s for nascent RNAs.

To extract more information from the recovery curves, they were fitted with the diffusion/reaction model developed by Sprague [Sprague et al., 2004]. This model assumes that the bleached spot contains identical binding sites that are equally distributed in space, also outside the area of bleach, and it allows to derive diffusion coefficients, binding time (τb), and the delay between two binding events (τd). Although probably oversimplifying we applied this model to describe our system. Indeed at the equilibrium/steady-state we have a cluster of immobile binding sites in an homogeneous compartment, the nucleoplasm. The recovery curves obtained with the subunit C of RNAPII were fitted with a reaction dominant version of this model (Equation 11). In Exo1 cells, we found that the polymerase resided for 333 s at the HIV-1 transcription site and diffused for 10 min before engaging a second transcription cycle. When the residency time of the polymerase was compared with that of nascent RNAs, it was obvious that the polymerase remained on the gene longer than expected. From the fit of Exo1 recovery curve with the two-step model, we know that

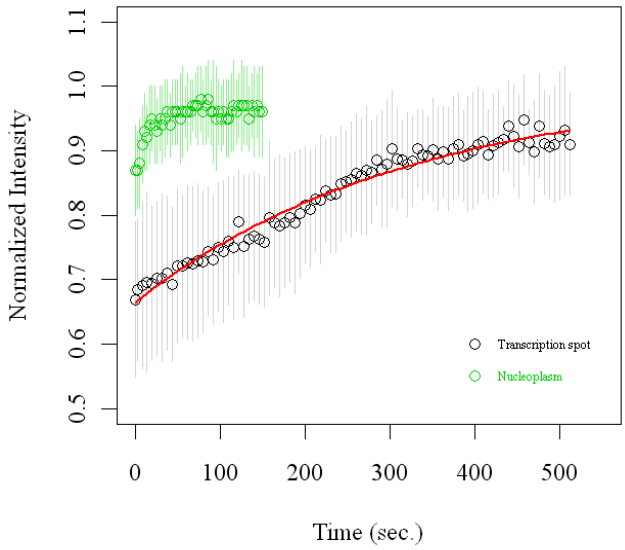

Figure V.15: RNAPII recovey curve at the transcription spot (Black) and in the nucleoplasm (Green).
The red line is the reaction dominant fit.

elongating through the reporter should take 114 s, and 3' end formation should take 63 s. Indeed, if the length of the full reporter is 3.852 Kb, it will take 114 s to transcribe it at a rate of 2 Kb/min. Thus, the difference between the total residence time of the polymerase at the transcription site, 333 s, and the time needed in transcription, 114 s, plus that for the 3' end formation, 63 s, is 156 s. Because chromatin immunoprecipitation experiments demonstrate that RNA-PolII stops rapidly after the end of the gene and is released without preceding much through neighboring sequences [Boireau et al., 2007], it is reasonable to suppose that this 156 s are employed at the transcription initiation step.

V.1.2 TAR:Tat:P-TEFb complex at HIV transcription sites

To evaluate the dynamic properties of the TAR:Tat:P-TEFb complex we identified HIV-1 transcription sites with a yellow or red fluorescent variant of MS2, and performed photobleaching experiments on CFP and GFP-tagged versions of Tat and Cdk9 (Figs. V.16 and V.17). When Tat or Cdk9 were bleached in the nucleoplasm of U2OS cells, the fluorescence recovered quickly, indicating that these proteins diffused rapidly through the nucleoplasm. We then bleach Tat and Cdk9 at the HIV-1

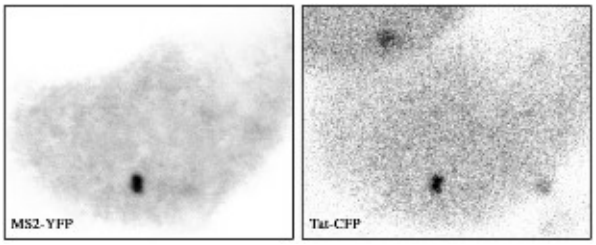

Figure V.16: MS2 and Tat at the transcription spot

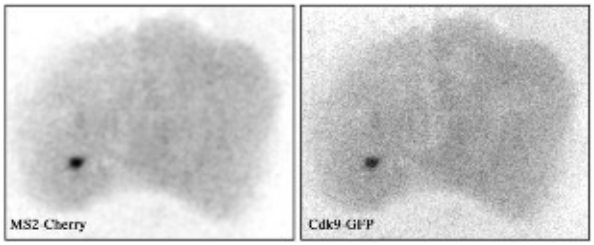

Figure V.17: MS2 and CDK9 at the transcription spot

transcription site, to analyze the dynamic properties of the complex formed on nascent RNAs.

V.1.2.1 Tat and Cdk9

The turn-over of Tat was slow, as complete fluorescence recovery took nearly three minutes. We observed two contrasting situations in the recovery curve of Cdk9 depending on the mode of activation of the HIV-1 promoter. In the presence of Tat, Cdk9 recovery was also slow and took several minutes to go to completion. In contrast, when HIV-1 transcription was activated by PMA/ionomycin, Cdk9 was highly dynamic, and recovery was within seconds. Tat and Cdk9 recovery curves were analyzed with the diffusion/binding model proposed by Sprague and MacNally (Equation 9) [Sprague et al., 2004], assuming that our conditions could be approximated to those of the model, and results are summarized in table 3.

	D [µm²/s]	$\boldsymbol{\tau_b}$ [s]	$\boldsymbol{\tau_d}$ [s]
Tat	8	55	60
Cdk9/Tat	7	71	142
Cdk9/PMA-ionomycin	9	11	15

Table 3: Kinetic parameters of the fitted Tat/Cdk9 FRAP curves.
D is the diffusion coefficient, τ_b the binding time and τ_d the time between two binding events.

The diffusion coefficients calculated for Tat and Cdk9 were similar and small, although they differed substantially in mass, pointing to a complex containing both protein. We found that Tat remained bound to nascent HIV-1 RNAs for 55 seconds, τ_b , and diffused for 60 seconds between two binding events, τ_d. Similarly, in the presence of Tat, Cdk9

resided for 71 seconds at HIV-1 transcription sites, but diffused for 142 seconds between two binding events. This situation changed dramatically when HIV-1 was activated by PMA/ionomycin, in the absence of Tat. In this case, PTEFb turn-over was rapid, as it remained bound to the HIV-1 transcription site for only 11s. Since PMA/ionomycin promotes transcription by activating NF-kB, which directly recruits P-TEFb, a major conclusion of this work is that the dynamic properties of P-TEFb depend on its mode of recruitment to the HIV-1 promoter. It is interesting to speculate that the higher stability of the Tat:TAR:P-TEFb complex may account for the stronger activation obtained with Tat. A similar case has been observed for the transcription factor HSF (heat shock factor) in Drosophila [Yao et al., 2006]. Indeed, while HSF bound its target gene whether it is activated or not, the dynamic properties of this interaction varied dramatically upon transcriptional activation. This raises the possibility that it is not only the binding of transcription factors to their target genes regulates transcription, but also the dynamic properties of these events.

V.2 pIntro-MS2x24 single integration

In order to study HIV-1 transcription in a system more similar to the natural viral infection condition, we developed a cell line carrying not an array but a single integration of the pIntro-MS2x24 construct (Fig. V.1). The HOS_A4 and HOS_B3 clones showed a robust and homogeneous

induction of the ECF-skl reporter in the cytoplasm upon treatment with GST-Tat [De Marco et al., 2008][Dieudonne et al., 2009]. After activation, the RNAs were concentrated in one (A4) or two (B3) bright spots in the nucleoplasm (Figs V.18 and V.19). When HOS_A4 or HOS_B3 were transfected with EYFP-MS2-nls and the virus transactivator Tat tagged with mDsRed many ECFP peroxisomes appeared in the cytoplasm and a bright yellow spot was clearly visible in the nucleus. This spot is compatible with a single site of pIntro-MS2x24 transcription that co-localized with Tat-mDsRed (Fig. V.20). Immunofluorescence with antisera against CyclinT1, the P-TEFb component recruited directly by Tat on the viral RNA, or RNAPII, demonstrated enrichment of such factors at these sites (Fig. V.21).

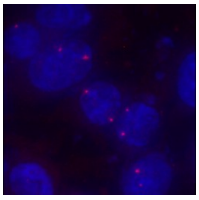

Figure V.18: HOS_B3 RNA FISH

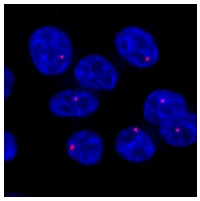

Figure V.19: HOS_A4 RNA FISH

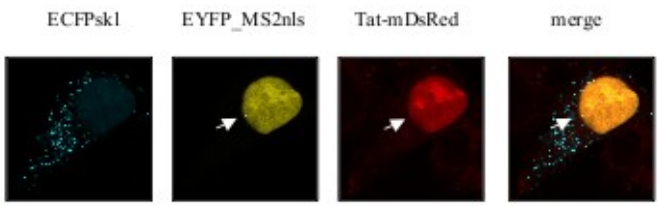

Figure V.20: HOS_A4

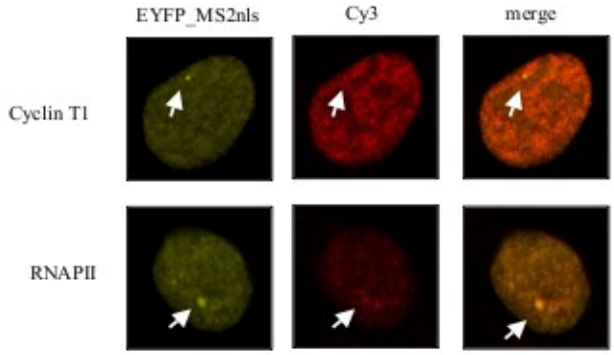

Figure V.21: HOS_A4 CyclinT1and RNA PolII

V.2.1 Kinetic analysis of the transcription site

We investigated the kinetic property of the pIntro-MS2-24x single integration transcription site by FRAP, as we did for the pExo-MS2-24x

array.

V.2.1.1 Transcription activation

When the pIntro-MS2-24x cells were activated by transient transfection of the viral transactivator Tat, the transcription of the pro-virus was constant for more than 10 hours (Figs V.22 and V.23) maintaining peripheral localization [Dieudonne et al., 2009]. The observation that the transcription spot intensity was constant over time suggests that also the RNA amount at that place was constant in time. Indeed, if the RNA amount at the transcription site is constant, when the system is perturbed by FRAP, any observed variation on the transcription spot fluorescence intensity can not be intrinsic but must be

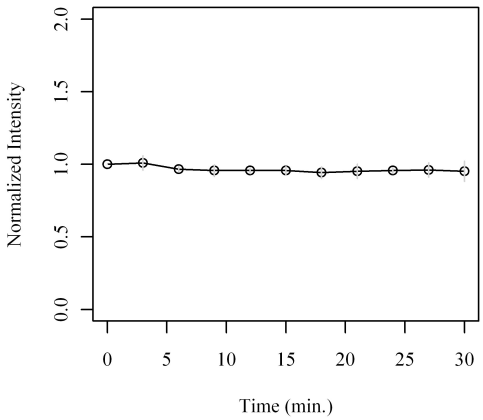

Figure V.22: Spot at 30min

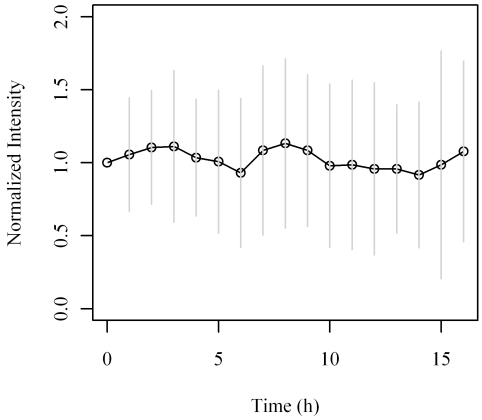

Figure V.23: Spot at 16h

generated by the bleach input.

V.2.1.2 Transcription spot

Surprising the recovery of fluorescence at the transcription spot for the single pIntro integration was very different as compared with that observed for the pExo array. The recovery of the single pIntro has an

half recovery time of almost 1 s compared to the 65 s needed for the pExo array (Fig V.4). Clearly in this case it is impossible to neglect the effect of diffusion, because the two processes have the same time scale recovery (Fig V.24). In order to estimate the transcription rate also for the pIntro single integration, it is therefore necessary to develop a model able to take into account both transcription and MS2-EYFP-nls diffusion.

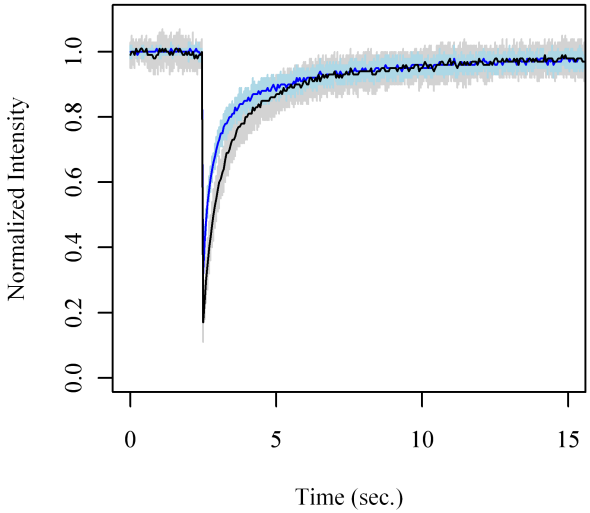

Figure V.24: B3 Clone, Ms2-YFP-nls at the transcription spot (black) and in the nucleoplasm (blue).

We treated pIntro-MS2-24x cells with several transcription inhibitors, with different inhibition mechanism, as DRB, actinomycin-D, α-amanitin, camptothecin and flavopiridol, and we always observed the fast disappearance of the spot. Although, if it was possible in the pExo array to slow down the fluorescence recovery at the transcription spot by inhibiting transcription with a drug (Fig. V.8), in the pIntro single integration cells we could not measure any slowing down in recovery rate. Until the spot was detectable, the rate of fluorescence recovery was unaffected and the recovery curve was indistinguishable from the control.

V.3 Transcription waves model

We developed a model to predict the number of MS2 RNA repeats at the transcription site in function of time.

We built our model on two assumptions:

1. all polymerases have the same transcription rate;
2. polymerases are equidistant.

We describe each polymerase sliding along the DNA like the peak of a positive progressive traveling wave (Equation 14). The transcription rate is the propagation speed (ω/k in equation 14), the maxima velocity. The amplitude of this wave is the number of MS2 RNA repeats transcribed by the single polymerase. This amplitude (Equation 15) is modulated following the position of the polymerase on the DNA: increases linearly during transcription of the 24 MS2 RNA repeats (l_1 in equation 15), is constant until the enzyme reaches the splicing acceptor site or the polyadenilation site (l_2 in equation 15), when the RNA processing begins and then exponentially decays (Fig. V.25). The model provides many important parameters of the system: the transcription rate, the polymerases density (π/k, equation 14) and the RNA processing mean life (k/(ωα), equations 14 and 15). The total number of MS2 RNA repeats at the transcription spot is the sum of the peak values (Equation 16).

$$f(x,t)=A(x)\cdot\left|\cos(k\cdot x-\omega\cdot t)\right|$$

Equation 14: Progressive wave.

$$A(x)=\begin{cases}\frac{24}{l_1}\cdot x & \forall\, x\le l_1\\ 24 & \forall\, x\le l_2\\ 24\cdot e^{-\alpha(x-l_2)} & \forall\, x>l_2\end{cases}$$

Equation 15: Transcription wave amplitude.

$$\mathrm{TranWave}(t)=\sum_0^{\frac{\omega}{\pi}t} A\left(\frac{\omega}{k}t-n\frac{\pi}{k}\right)$$

Equation 16: Transcription Wave model, total RNA.

At the equilibrium the total number of RNAs at the transcription site is constant, with polymerases sliding along the DNA and the transcription of new RNAs is compensated by the processing of that already transcribed. When we perturb the system by bleaching the transcription spot, the MS2-EYFP protein bound to the RNA repeats loses its fluorescence but newly synthesized repeats are freely accessible to fluorescent MS2-EYFP. The speed of fluorescence recovery at the transcription spot is related both to the rate of transcription and to that of RNA processing. Indeed, in order to recover the fluorescence signal at the transcription spot, new repeats have to be produced because the mean life of the binding of MS2 protein to the MS2 RNA

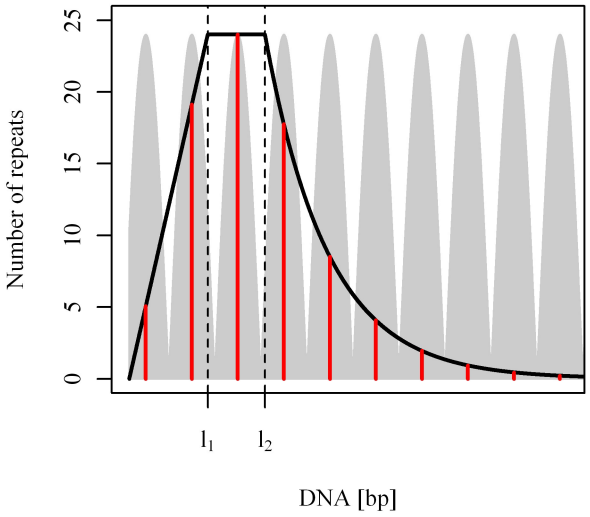

Figure V.25: Transcription Wave model.

repeats is many time higher than the recovery of the spot. It is also clear that, in order to rescue the initial fluorescence value, all the bleached MS2-EYFP proteins, bound to the MS2 RNA repeats, have to leave the transcription site. We initially tested the TranWave model by fitting the recovery curve data of U2OS-Exo1 carrying an array of integration sites, with equation 16 (Fig. V.26). The transcription rate obtained (2.4 Kb/min) was very similar to that obtained with the linear plus exponential model discussed previously (2.3 Kb/min). Also the time needed for the RNA processing obtained from the two models are compatible, 65 and 73 s, respectively. The transcription wave model allows also to estimate the PolymeraseII density at the transcription site, almost one each 100 bp.

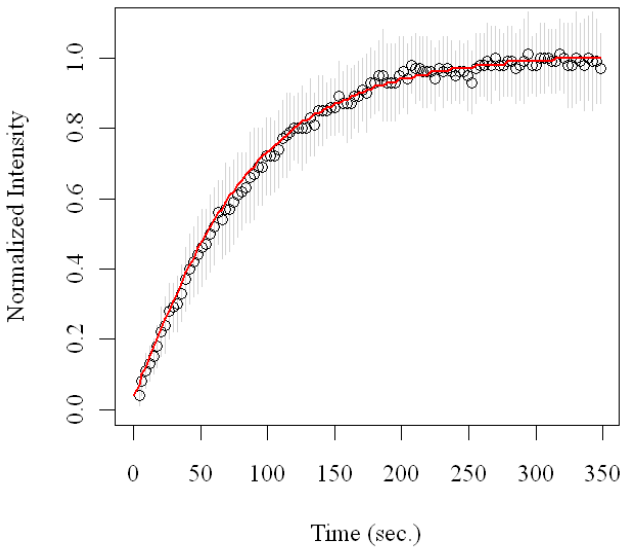

Figure V.26: Exo1 Transcription Wave fit.

V.3.1 Transcription Wave and Diffusion

In order to measure the recovery of the signal at the transcription site, after the bleach, the fluorescent MS2-EYFP proteins have to diffuse inside the area of bleach. The transcription spot can recover its intensity only if new fluorescent proteins are free to bind it.

Three different scenarios are possible: diffusion is faster than transcription, transcription is faster than diffusion or transcription and diffusion have the same time scale.

If diffusion is faster than transcription (ie. pExo-MS2x24 and pIntro-MS2x24 array), after the bleach free fluorescent MS2-YFP protein immediately diffuse at the transcription site before new RNA repeats are produced, and the recovery of the intensity of the transcription spot follows exactly the RNA production.

If transcription is faster than diffusion, after the bleach new RNA repeats are synthesized at the transcription site before fluorescent protein can bind to it, therefore only after MS2-EYFP diffusion it becomes possible to measure recovery at the spot. In this case the recovery of fluorescence at the transcription site would rise like that of MS2-EYFP diffusion. However, it is difficult to imagine that transcription can be faster than diffusion.

If transcription and diffusion have the same time scale both the processes participate to the recovery of the spot intensity.

When diffusion is faster than transcription, it is possible to take into account the effect of diffusion on the recovery of the transcription spot with a new assumption: the diffusion of MS2-EYFP protein is not influenced by the presence of the MS2 RNA repeats. This assumption is reasonable when the MS2 molecules free to bind are many time more abundant that their specific RNA binding sites. In this conditions we can

predict the concentration of free MS2-EYFP at the transcription site assuming that the protein diffuse within this region like in the nucleoplasm. We modified the equation 9 proposed by Sprague et al. [Sprague et al., 2004] integrating the solution of the reaction diffusion system for the free subset, not in a circle with the same radius of the bleached spot, w, but in a circle with a smaller radius, w_0. We used this equation (Equation. 17) to predict the concentration of MS2-EYFP-nls free to bind at the transcription spot. The solution for the total fluorescence in a smaller radius was used as a background of our measurements (Equation. 18).

$$\bar{f}(p)=\frac{2F_{eq}}{p}\frac{w}{w_0}K_1(qw)I_1(qw_0)$$

Equation 17: Modified solutions of the reaction-diffusion equations system for the free subset of fluorescent protein in a circle of radius smaller than that of bleach.

$$\overline{frap}(p)=\frac{1}{p}-[\frac{F_{eq}}{p}-\frac{2F_{eq}}{p}\frac{w}{w_0}K_1(qw)I_1(qw_0)]*$$
$$*(1+\frac{K_{on}}{p+K_{off}})-\frac{C_{eq}}{p+K_{off}}$$

Equation 18: Modified solution of the reaction-diffusion equations system for the total fluorescence in a circle of radius smaller than that of bleach.

Introducing this last assumption we are able to distinguish the two components of the fluorescence recovery at the transcription spot (Equation 19), diffusion of the MS2 protein and transcription. The parameters necessary to predict the diffusion of MS2, equations 17 and 18, were independently obtained by FRAP experiments in the

nucleoplasm outside the transcription spot. In equation 19: MS2(t) is, as in equation 18, the background of fluorescence, $MS2_{free}(t)$ is, from equation 17, the portion of fluorescent MS2 free to bind the RNA repeats at the transcription spot, and finally TranWave(t), as equation

$$TS(t)=(1-\alpha)\cdot\frac{MS2(t)}{MS2}+ \\ +\alpha\cdot\frac{MS2_{free}(t)\cdot TranWave(t)}{MS2_{free}\cdot TranWave}$$

Equation 19: Transcription spot fluorescence recovery.

16, is the number in time of MS2 RNA repeats; MS2, $MS2_{free}$ and TranWave are the equilibrium values. The parameters obtained from the experimental data fitted by equation 19 are all those involved in the transcription wave model, as described in paragraph V.3, plus α, that is a weight for the fluorescence background at the transcription spot.

V.3.2 pExo/pIntro-MS2x24 array

Since the single integration cell line carry the pIntro construct, we decided to develop a cell line with an array of pIntro similarly to what was done for pExo for direct comparison of the two models (arrays *vs* single integration). The fluorescence recovery at the transcription site of the pIntro-MS2x24 array, U2OS-pIntro, is similar to that of U2OS-pExo (Figs. V.27 and V.28). We applied equation 20 to fit FRAP data from the U2OS-pExo and U2OS-pIntro transcription spots (Figs. V.27 and V.28); this equation fits well the experimental data and results are summarized

in table 4.

TS arrays:	**ω** [1/s]	**k** [1/bp]	**ρ** [Kb/min]	**α** [1/bp]	**β** [1/bp]	**τ** [s]	**PolII** [n/bp]	**n MS2**
pExo	$6.2 \cdot 10^{-1}$	$1.6 \cdot 10^{-2}$	2.3	$1.5 \cdot 10^{-4}$	/	173	191	1000
pIntro	$6.4 \cdot 10^{-1}$	$4.5 \cdot 10^{-2}$	0.9	$5 \cdot 10^{-4}$	$2 \cdot 10^{-3}$	176	70	1300

Table 4: pExo/pIntro array transcription wave model results.

The transcription rate value obtained for U2OS-pExo, ρ, is in agreement with that obtained with the previous model, linear plus exponential (table 2).

In order to fit data from the pIntro construct we introduced an additional variable to the transcription wave model, β, to take into account that in this case RNA degradation can start just after the splicing acceptor site and not only at the polyadenilation site. In this

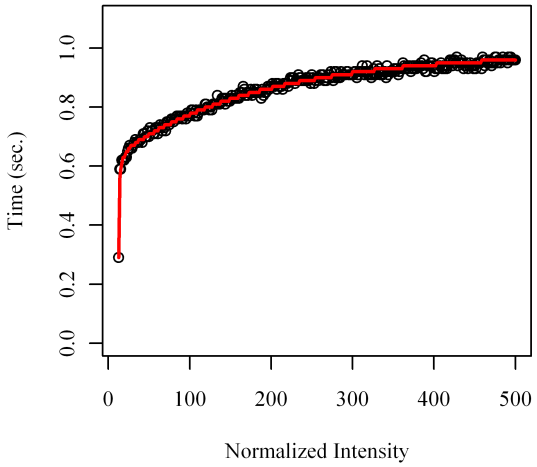

Figure V.27: pEXO

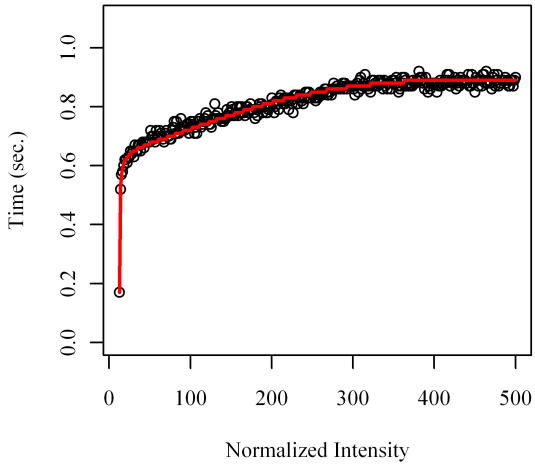

Figure V.28: pIntro

case the transcription rate was lower than what measured for U2OS-pExo, but still in the same order of magnitude.

V.3.3 pIntro-MS2x24 single integration

When we applied the transcription wave model as formulated in

equation 19 to fit the recovery curve of the Intro-MS2x24 single integration, HOS-Intro, we were unable to obtain a reasonable value for the variables. Indeed we obtained a transcription rate of 25 kb/min and 15 RNAPII at the transcription spot (Fig. V.29). We supposed that the model failed because the assumption on the MS2-EYFP diffusion was not respected being influenced by the presence of RNA specific binding sites at the transcription spot. Transcription spot fluorescence recovery in HOS_A4 or HOS_B3 is very fast, and has the same time scale of diffusion. It is possible that just after the bleach the number of available RNA repeats is major than that of MS2-EYFP free to bind it. Perhaps it is also important to considerate that the local concentration of MS2-EYFP bleached at the first time points after the bleach is higher than that of still fluorescent protein. Indeed at the beginning of fluorescence recovery some MS2-EYFP in the dark stage can bind the repeats introducing a variation from the model.

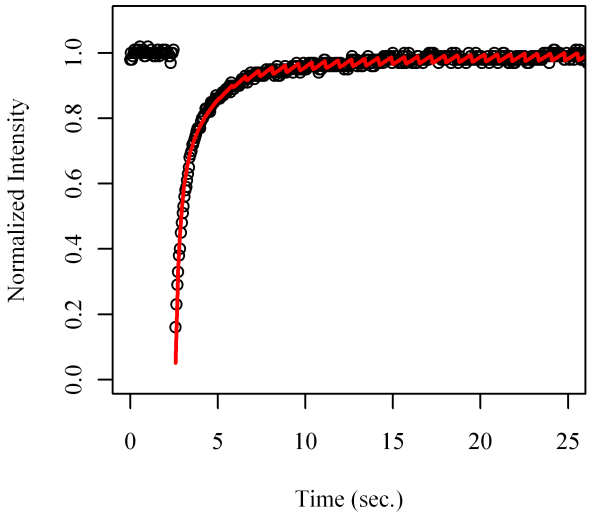

Figure V.29 B3 Transcription Wave Model

V.4 Monte Carlo simulation

Since the reaction-diffusion equations system (Equation 6) is not

analytically resolved for both clusterized and unstable/growing in time binding sites, we decided to reformulate our transcription wave model, equation 19, by replacing the diffusion component, obtained modifying the Sprague solution (Equations 17 and 18), with a Monte Carlo simulation. This approach allows us to reduce the model assumptions and to take into account effects previously neglected. With a Monte Carlo simulation it is possible to calculate the deviation on the MS2-EYFP-nls diffusion made by the RNA repeats at the transcription spot. This method also allows to take into account the probability that bleached MS2-EYFPs can bind the new synthesized RNA repeats at the first time steps after the bleach, when their local concentration is higher respect the still fluorescent protein subset.

We developed a two dimensions Monte Carlo system in a matrix and we fixed the cell dimensions as that of our FRAP experiment pictures pixel. The dimension of the matrix was chosen to have the same area of a typical cell nucleus. Fixing the cell length, the diffusion coefficient is uniquely determinate from the time step, as in equation 20.

$$D=\frac{\Delta x^2}{4\Delta t}$$

Equation 20: Diffusion coefficient in a 2D random walk.

We validated our Monte Carlo system simulating the MS2-YFP-nls diffusion in the nucleoplasm with the parameters obtained fitting the

experimental recovery curve with the Sprague model, equation 10. Indeed we fitted the simulated curve with the same equation and we obtained parameters pretty similar to that used to run the Monte Carlo simulation (Fig. V.30).

The next implementation to definitely tackle the transcription rate measurement in the single integration pIntro cell line will be linking the Monte Carlo simulation with the Transcription Wave model.

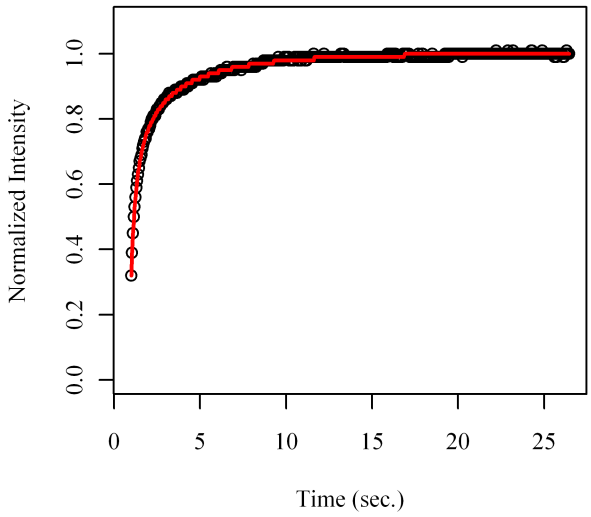

Figure V.30: Monte Carlo simulation fitted with the Sprague model.

VI. DISCUSSION

Transcription by RNA polymerase II (RNAPII) is the process that translates DNA into RNA leading to the expression of a specific gene in mammalian cells. The core of the process is the polymerizing activity of RNAPII that chemically catalyzes the addition of nucleotides to the RNA chain. Since transcription occurs in the environment of the living cells it became to be important to measure the transcription rates in such a context. However, very few studies were able to measure directly RNAPII activity in *vivo*. In *vitro* kinetic analysis of human RNAPII have yielded elongation rates between 0.9 and 1.8 kb/min and have suggested that the chemical step is rate limiting [Burton et al., 2005]. The giant gene of human dystrophin and the long (8 kb) yeast nonessential gene *YLR454*, have also been used to evaluate elongation rates [Tennyson et al., 1995][Mason and Struhl, 2005]. These studies yielded values of 2.4 and 2 kb/min, respectively. The range of observed elongation rates values in different biological systems went from 1.5 kb/min for living bacteria to 5.7 kb/min for eukaryotic ribosomal RNA genes [Dundr et al., 2002][Golding et al., 2005].

Several groups used the RNA reporter system that employs the coat protein of phage MS2 fused to the green fluorescent protein (GFP)

[Bertrand et al., 1998]. It was possible to monitor dynamics of mRNAs in a living cell by inserting of several MS2 binding sites in an inducible reporter gene [Shav-Tal et al., 2004a][Golding and Cox, 2004][Le et al., 2005][Forrest et al., 2003].

To analyze the biogenesis of mRNA in live cells and real time, we engineered cell lines that contained many copies of the reporter, a construct based on the human immunedeficiency virus 1 (HIV-1) (Figure V.1), integrated in a single place within the chromatin (75 copies for Exo1 cells). This amplifies the signal such that many events become visible by microscopy analysis [Janicki et al., 2004].

In our system we measured elongation rates of RNAPII, by bleaching MS2-GFP-nls bound to nascent RNAs, obtaining an elongation rate of approximately 2 kb min^{-1} and no evidence of exceeding pausing of the polymerase [Boireau et al., 2007](Fig. V.11). Three important controls support our result (Table 2). First, increasing the length of the mRNA increased the total synthesis time but yielded a similar elongation rate (Fig. V.14). Second, a slow mutant of RNAPII induced a reduced transcription rate (Fig. V.10). Third, treating cells with camptothecin, an inhibitor of elongation, also reduced the rate of RNA synthesis (Fig. V.12). In our analysis we neglected the possibility of a rapid dissociation of the RNA-bound MS2GFP-nls that may lead to rapid recovery rates unrelated to the synthesis of new RNAs. This issue was resolved by measuring the recovery of fluorescence on a relatively immobile

nucleolar RNA (U3) carrying an MS2 binding site (Fig. VII.1).It was shown that the recovery of bound MS2 on U3 exceeded ten minutes indicating a high affinity of the protein for the substrate in agreement with *in vitro* studies on the RNA binding properties of the MS2 protein [Johansson et al., 1998].

Interestingly in a similar system, carrying 100 tandem integrations of reporter, RNAPII was estimated to elongate at 4.3 kb min^{-1} with long paused intervals [Darzacq et al., 2007]. In vitro single-molecule analyses of bacterial RNA polymerases have shown that the time taken to transcribe a DNA segment can vary substantially between individual molecules, and at least part of this variability can be attributed to polymerase pausing [Herbert et al., 2006]. In higher eukaryotes, pausing is also a well-known phenomenon, and there are some evidences that the RNAPII elongation rate affects gene expression [Batsche et al., 2006]. However, as showed in paragraph V.1.1.2, where we presented our two step model, we described elongation as the repetition of an elementary step and represented it as a straight line in the recovery curves (Fig. V.11). This does not take into account the polymerase heterogeneity that can be caused by pausing that may directly contribute to the exponential component of the curve. Although in many cases determining polymerase pausing time could not be easy, it seems that pausing occurs when cells were treated with camptothecin. Indeed, in this case, the MS2-GFP-nls FRAP assay indicated a large

increase in the exponential component of the recovery curve (Fig. V.12). Because camptothecin is believed to physically arrest the polymerase, it is tempting to speculate that this increase represents pausing.

We studied also RNAPII kinetic at the transcription site and the polymerase recovery curves were fitted with a diffusion-binding model (Fig. V.15). The results were compared with that obtained for nascent RNAs bleaching. The data indicated that during the 333s that the polymerase resided at the HIV-1 transcription site, 114s could be attributed to elongation, and 65s could be attributed to 3'-end processing and transcription release. The remaining 154s could be the result of initiation and termination. This long residence time of the polymerase can be explained by several alternative hypothesis. It is possible that the polymerase would reengage transcription on the same gene by a looping mechanism, as described for yeast [Ansari and Hampsey, 2005][O'Sullivan et al., 2004] and also for HIV [Perkins et al., 2008]. Looping brings physically together the 5' and 3' ends of a gene coupling initiation with termination. Alternatively, as a result of the tandem array of integrations, RNAPII is not released from the DNA after termination sliding on the following template in the array. However, this hypothesis is unlikely given the observation by CHIP assays that the density of RNAPII on the HIV gene decrease after termination [Boireau et al., 2007]. One could imagine, however, that RNAPII, after termination and release, remains in the area able to reinitiate the

following gene. Finally we can not exclude that some polymerases present at the HIV-1 transcription sites are stalled or are involved in as yet uncharacterized but slow processes. As described in paragraph V.1.1.3, we found that although the residency time of the polymerase at the HIV-1 gene was 333 s, the diffusion time between two binding events was 10 min. Previous FRAP experiments of the large subunit of RNAPII at random nucleoplasmic sites have indicated that for cellular genes, polymerases should spend only a third of their time engaged in transcription [Kimura et al., 2002]. Because the half-life of transcription was estimated as 20 min, it was deduced that polymerases spend as much as 90 min diffusing between two transcriptional cycles [Kimura et al., 2002]. Thus, the initiation at the HIV-1 promoter is nearly 10 times more frequent than at cellular genes. This could be caused by a high efficiency of transcription initiation at the HIV-1 site or, alternatively, by a much higher density of active genes at the HIV-1 transcription site, although also in this case it is not possible to exclude a tandem integration effect. Interestingly, we did not detect a rapid component in the recovery curves of RNAPII, which corresponds to polymerases undergoing rapid binding and dissociation from the promoter [Darzacq et al., 2007]. In contrast, the vast majority of polymerases appeared transiently immobilized on the HIV-1 gene and engaged in productive transcription. This suggested that the initiation of HIV-1 transcription was indeed very efficient. A high efficiency of initiation would be

consistent with the biochemical studies that have shown that the HIV-1 promoter is constantly occupied by the polymerase [Jeang et al., 1999] [Marcello et al., 2001].

The integration of several independent transcription units increases the sensibility of our system and gives the possibility to visualize and study the recruitment of factors such as the Tat/P-TEFb complex at the HIV-1 transcription site (Figs. V.16 and V.17). Our data showed that Tat and P-TEFb remained bound to nascent RNAs for about a minute (Figs. V.26 andV.25, Table 3). If HIV-1 transcription proceeds with the previously described rate of 2 kb/min, then elongation through our reporter RNA would last more than 2 minutes. This raises the possibility that the TAR:Tat:P-TEFb complex is dissociated from the polymerase before the gene is completely transcribed. Following dissociation, the fate of Tat and Cdk9 is likely to differ, as shown by their significant difference in the time between two binding events (τ_d in Table 3). It is remarkable that Tat and Cdk9 have very similar dynamics. This supports the idea that they remain together in the elongating complex, rather then Tat being transferred to the polymerase while P-TEFb dissociating from it. Since chromatin immunoprecipitation data have shown that Tat and P-TEFb are present with elongating polymerases all along the gene [Bres et al., 2005], we suggest that Tat and P-TEFb could undergo constant association and dissociation cycles with TAR and the elongating polymerase. Altogether, our data showed that the TAR:Tat:P-TEFb

complex is remodeled during HIV-1 transcription.

Results discussed so far relied on the insertion of tandem arrays of the constructs, approximately 75 copies [Boireau et al., 2007], and almost 100 copies in a similar work [Darzacq et al., 2007], that is an artificial situation that has also been shown to induce heterochromatin at the repeats site [Wang et al., 2006]. Therefore, such pioneering studies could only open the way to further analysis to be carried out on single genes. Among retroviruses, the human immunedeficiency virus type 1 (HIV-1) maintains a robust transcription in the presence of the viral Tat transactivator possibly allowing the production of enough RNA to be visible by MS2-tagging. Hence, we decided to exploit HIV-1 derived vectors carrying all *cis*- information for packaging, integration and Tat-inducible transcription plus the MS2 binding sites inserted in the intron downstream of the major splice donor (Figure V.1). Transduced HOS cells were selected to be inducible by the viral transactivator and to carry single integrations of the vector. The HOS-B3 and HOS-A4 cell clones were chosen for subsequent work (Figures V.20 and V.21).

First we analyzed the intensity of the EYFP-MS2nls signal continuously in HOS_B3 live cells after transfection of Tat and EYFP-MS2-nls. As shown in figure V.22, little variation was observed in the intensity associated with the transcription sites for 30’. The bursts or pulses observed in other systems [Chubb et al., 2006][Golding et al.,

2005] were either not detected or extremely small, below 1% of the signal. Further analysis up to 17 hours did not show major fluctuations either (Figure V.23). Analysis was initiated fifteen hours after transfection indicating that, when the viral transactivator is not limiting, HIV transcription is continuously active. This situation mimics productive infection where the duration of the HIV gene expression pulse, determined by the positive feedback loop of Tat, exceeds the lifespan of the infected cell [Weinberger et al., 2005].

Next, we wished to explore the kinetic behavior of HIV RNA biogenesis from a single transcription unit. Therefore, we measured the recovery of fluorescent MS2 fusion protein bound to the MS2 stem-loops in the HIV RNA after photobleaching the transcription site. As shown in figure V.24, recovery of the signal in HOS_B3 occurred within seconds regaining pre-bleach equilibrium value after almost 10s. This finding was surprising and totally unexpected after previous observations made on tandem arrays that showed recovery at plateau after more than 200s [Boireau et al., 2007][Darzacq et al., 2007]. Indeed, when we compared directly the same HIV-Intro-MS2×24 construct either as single integrated provirus or in tandem arrays (Figs. V.24 and V.28), the previous recovered ten times quicker than the latter. In both cases recovery reached equilibrium with a negligible immobile fraction, indicating that no RNA was retained at the transcription site for longer times. Furthermore, this was not an intrinsic property of the HOS_B3

clone, since other clones derived with the same protocol, like the HOS_A4 clone, showed similar recovery curves.

After the bleaching of the transcription sites, new MS2 binding sites are synthesized resulting in the recovery of the fluorescent signal. However, RNA is visualized indirectly through MS2-EYFP-nls introducing a series of possible uncertainties in the interpretation of the FRAP data. Indeed a slow diffusion rate of MS2-EYFP-nls may mask the neo-synthesis of nascent RNAs. Diffusion of MS2-EYFP-nls was rapid and much faster than the recovery at the transcription sites and therefore neglected when measuring tandem arrays [Boireau et al., 2007][Darzacq et al., 2007]. Also in our case, diffusion of MS2-EYFP-nls didn't impact significantly the recovery of fluorescence in tandem arrays of HIV-Exo-MS2×24 (Fig. V.4). However, when comparing the recovery curve of free MS2-EYFP-nls in the nucleoplasm to the recovery of fluorescence at the transcription site in HOS_B3 cells the two curves were very close, although distinguishable (Fig. V.24). Hence, diffusion of MS2-EYFP-nls in the area of bleach cannot be neglected in the calculation of the transcription rate in single HIV transcription units.

We propose a model (TranWave) that describes the kinetic behavior of MS2-EYFP-nls binding to newly synthesized RNAs in order to measure RNAPII transcription rates. As a first step, in order to predict the kinetic of MS2-EYFP-nls at the transcription spot that is available for binding to the repeats, we took advantage of the well-described diffusion-binding

model proposed by Jim McNally and collaborators [Sprague et al., 2004]. With this information we hypothesized that MS2 binding sites were produced at a certain rate in the transcription spot, dependent on RNAPII activity. The intensity of the fluorescent signal generated by the MS2-EYFP-nls effectively bound to the RNA at the transcription site would be proportional both to the amount of free protein and to the available binding sites. At steady state there is equilibrium between nascent RNAs that is produced by RNAPII and its release from the transcription site. To predict this behavior, assuming that the polymerases are equidistant and that the transcription rate of RNAPII is constant, we imagined each polymerase trailed along the gene as the peak of a harmonic traveling wave. The amplitude of the wave would be linearly increasing during the synthesis of the repeats, then constant until the polymerase reaches signals for splicing or release from the transcription site and then exponentially decreasing (Fig. V.25 and Equation 15). Hence we can measure the transcription rate of RNAPII as the propagation speed of the wave, which is the speed of the sliding of the wave peak. In a FRAP experiment the recovery curve includes a component that is intrinsic of the diffusive behavior of MS2-EYFP-nls and a component that depends on the transcription rate of RNAPII, both of which are predicted by TranWave. We initially applied TranWave to cells carrying tandem arrays of HIV-Intro-MS2×24 and HIV-Exo-MS2x24 (Fig. V.1). We obtained transcription rates of approximately 2 and 1

kb/min respectively (Table 4), that were comparable to what we obtained with the two step model or what others group reported [Darzacq et al., 2007]. However, when we measured the transcription rate of the singularly integrated provirus we obtained the value of 25 kb/min that exceeds five to ten fold the values measured values so far. Interestingly, quantitative FISH showed that the number of RNA MS2 repeats presents at the transcription site in the single integration is almost one half of that in the array. This results is surprising if we consider that in the array there are 49 more copies of the same reporter than in the single integration. We can conclude that the number of RNAPII involved in transcription at the single integration site is higher than that of the tandem array. This higher concentration of RNAPII can also partially explain the higher transcription efficiency that we measured. Indeed it has been reported in bacteria that most transcriptional delays disappear if more than one RNA polymerase molecule initiates from the same promoter [Epshtein and Nudler, 2003] [Epshtein et al., 2003].

However we are still improving our model in order to decrease the number of assumptions and simplifications and to obtain stable and reliable fits of our experimental data. To this end we are developing a Monte Carlo simulation of the diffusion component of recovery curve.

VII. MATERIAL AND METHODS

The experimental part of this work dealing with plasmid and cell manipulation, was conducted with the precious help of Anna Knezevich and Alex De Marco.

VII.1 Cells and plasmids

VII.1.1 Plasmids

Plasmids expressing MS2-GFP, Tat-eGFP, CDK9-GFP, GFP-CycT1 and the hC4 and WTRES mutant of RNAPII have been described previously [Mata et al., 2003][Fusco et al., 2003][Marcello et al., 2003]. MS2-Cherry and GFP-PolII-C were created with the Gateway system (Invitrogen). U3-MS2 was created by inserting a single MS2 site in the apical loop of the rat U3B.7 gene. pExo-MS2×24, pIntro-MS2x24 and pExo-MS2×24-Long plasmids were derived from the plasmid pEV731 [Jordan et al., 2001][Marcello and Giaretta, 1998].

VII.1.2 pExo and pIntro HIV-1 based cells

U2OS and HOS cells were cultivated at 37°C in DMEM containing 10% FCS and antibiotics.

VII.1.2.1 pExo and pIntro-MS2x24 array

Stable U2OS transformants were obtained with the calcium-phosphate procedure by cotransfecting a 20-fold excess of the vector of interest, pExo, pExo-Long or pIntro (Fig. V.1), with Ptk-Hygro and selecting cells with 132 µg/ml hygromycin. Individual clones were expanded, and their gene copy number was measured by quantitative PCR using DNA from U1 cells as reference (two copies per genome). U2OS-Exo1, U2OS-ExoLong and U2OS-Intro contained approximately 75, 70 and 50 copies, respectively [Boireau et al., 2007].

VII.1.2.2 pIntro-MS2x24 single integration

Osteosarcoma HOS_143b cells, that are negative for tymidine kinase activity (TK-), were transduced with pIntro-MS2x24 vector (Fig. V.1) pseudotyped with the VSV-G envelope. To select for clones that carry an inducible integrated provirus, cells that constitutively expressed high level of HSV-TK were selected by treatment with 50 µg/ml ganciclovir. Surving cells, that were either non-transduced, or transduced but with a low level of TK expression, were treated with GST-Tat and briefly selected for inducible HSV-TK expression in hypoxanthine, aminopterin and thymidine (HAT) medium. Clonal populations were obtained by limiting dilutions and colonies were visually scored for low basal level of ECFP expression in the cytoplasm and to be higly inducible by GST-Tat by fluorescence microscopy [De Marco et al., 2008]. The HOS_A4 and HOS_B3 clones showed a robust and homogeneous induction of ECF-skl

in the cytoplasm upon treatment with GST-Tat. Southern blotting analysis and cloning of the integration sites by inverse PCR revealed that in HOS_A4 clone a single copy of the provirus integrated in the HMBOX1 gene of chromosome 8. HOS_B3 instead carry two single copy of the provirus integrated in the SIPA1L2 gene (chromosome 1) and in the DOC1 gene (chromosome 3) [De Marco et al., 2008][Dieudonne et al., 2009][Biancotto, 2006].

VII.2 Imaging and FRAP procedures

VII.2.1 Microscopes

The DMRA (Leica) wild-filed microscope was equipped with a 100x NA 1.4 objective and a CoolSNAP HQ (Roper Scientific) camera and was controlled by MetaMorph software (Universal Imaging Corp.).

The DMIRB (Leica) wild-filed microscope was equipped with a 63x NA 1.4 objective and a CoolSNAP K4 (Roper Scientific) camera and was controlled by MetaMorph software (Universal Imaging Corp.).

The LSM510 META (Zeiss) confocal microscope was equipped with a 63x NA 1.4 objective.

The Nikon TE200 microscope was equipped for both confocal and wide-field imaging. A 100x NA 1.45 objective was used. The wide-field port was equipped with an EM-CCD (Cascade 512K; Roper Scientific) camera controlled by MetaMorph (Universal Imaging Corp.).

VII.2.2 Immunofluorescence and in situ hybridization

Immunofluorescence and in situ hybridization were performed as previously described [Marcello et al., 2003][Fusco et al., 2003]. Fluorescent images of fixed cells were captured with different microscopes, either the DMRA and DMIRB (Leica) wide-field and the LSM510 META (Zeiss) confocal microscope.

Stacks of wide-field images were deconvolved with Huygens (Bitplane AG) or with the ImageJ (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://rsb.info.nih.gov/ij/, 1997-2009) plug-in "*Iterative Deconvolve 3D*".

VII.2.3 Live cell imaging and FRAP experiments

For live cell imaging, cells were maintained at 37°C in appropriate medium and humidity and CO2 air concentration were under control [Fusco et al., 2003].

VII.2.3.1 Wide-field FRAP

Recovery curve showed in figures V.4, V.5, V.7, V.8, V.9, V.10, V.11, V.12, V.13, V.14, V.15 and V.26 were obtained with the Nikon TE200 microscope. A circle of 2.5µm diameter was bleached in almost 1s using the confocal port and recovery was measured in the wide-field. Stacks of nine images 0.5µm apart were collected every 3s.

VII.2.3.1.1 Image processing

The fluorescence intensity recovery was automatically tracked in 3D.

A small parallelepiped of dimension 1x1x1.5 µm placed at the most intense pixel of the region of bleach was measured over time with a macro written for ImageJ (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://rsb.info.nih.gov/ij/, 1997-2009). In the transcription spot FRAP experiments the post-bleach value was taken at the 5s time point, when the diffusion component of the recovery is expected to be already at equilibrium, and was set to zero.

VII.2.3.2 Confocal FRAP

Recovery curve shoved in figures V.24, V.28, V.27, V.29 and VII.2 were obtained with the confocal LSM510 Meta (Zeiss) microscope. Images of 128x128 pixel (24.4x24.4 µm) and optical thickness of 1.5µm were acquired each 50ms (V.24, V.29 and VII.2) or 1.38s (V.28 and V.27). A circle of 2.28µm diameter was bleached in less than 100ms.

VII.2.3.2.1 Image processing

Images were analyzed with ImageJ (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://rsb.info.nih.gov/ij/, 1997-2009). When a transcription spot was bleached the *"SpotTracker"* plug-in [Sage et al., 2005] was used to track it in 2D and a macro was created to measure the fluorescence intensity in a circle of 1.14µm diameter at the center of the spot.

VII.2.3.3 Time lapse

Data in figures V.23 and V.22 were obtained from pictures acquired

with the DMIRB (Leica) microscope. Stacks of 21 images 0.5µm apart were collected every 3min in figure V.23 and every 1h in figure V.22.

VII.2.3.3.1 Image processing

The stacks were deconvolved with the ImageJ (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://rsb.info.nih.gov/ij/, 1997-2009) plug-in "*Iterative Deconvolve 3D*" using an experimental PSF and successively maximally projected. The transcription spot was tracked with the *"SpotTracker"* plug-in [Sage et al., 2005] and a macro was used to measure the fluorescence intensity in a circle of 1.14µm diameter at the center of the spot.

VII.3 Curve Fitting

For the mathematical analysis of the data experimental normalized recovery curves from 10 to 20 cells were averaged.

To fit the Linear plus Exponential model, equation 13, a C++ code was written and the Chi-square was minimized with the multidimensional minimization routine of the Gnu Scientific Library [GSL].

To fit Sprague equations, equations 4, 10, 18, 19 and the two independent binding site solution, a C++ code was written and the Chi-square was minimized with the multidimensional minimization routine of the Gnu Scientific Library [GSL]. To solve the equations the numerical inverse Laplace transform was computed by Euler summation as

proposed by Yu Wei, that translated the algorithm in C++ from the original in Abate et al. [Abate, 1996], and the complex Bessel function library was obtained from Jean-Pierre Moreau (http://pagesperso-orange.fr/jean-pierre.moreau/c_bessel.html).

To fit the the simplified Sprague reaction-dominant solution, equation 14, the R Nonlinear Least Squares (nls) function was used [R-project].

To develop Transcription Wave model, equation 17 and 20, and the Monte Carlo simulation, C++ codes were written based on GSL library [GSL]and when necessary previously described codes were used.

VII.4 MS2-G/YFP-nls

VII.4.1 Binding efficiency and U3

The MS2 binding efficiency to its specific 19 nucleotides RNA sequence is very high. In vitro the mean life of this binding is approximately 6h [Johansson et al., 1998]. To obtain an estimate for the rate of exchange of the MS2-G/YFP-nls protein on its target site in our system, we used an abundant non-coding RNA, U3. A single MS2 site was inserted in the apical loop of the rat U3B.7 gene. U3 is synthesized in the nucleoplasm and accumulates in nucleoli, where it plays essential functions during ribosomal RNA biogenesis [Kass et al., 1990]. It is also know that a fraction of the protein associated with U3 exchange slowly between nucleoli and nucleoplasm [Phair and Misteli, 2000]. When

nucleoli of cells expressing only MS2-G/YFP-nls were bleached, fluorescence recovery was nearly complete within seconds. In striking contrast, when nucleoli of cells expressing both U3-MS2 and MS2-G/YFP-nls (Fig. VII.1) were bleached, only a fraction of the fluorescence was recovered, even after 10 min [Boireau et al., 2007]. This indicated that a part of U3-MS2 was immobile and that bleached MS2-GFP-nls molecules stayed stably bound to the RNA.

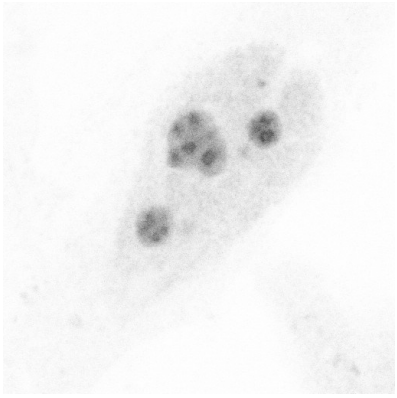

Figure VII.1: Exo1 transfected with U3

VII.4.2 Nuclear MS2-G/YFP-nls diffusion

FRAP experiments on nuclear MS2-EYFP-nls didn't show any substantial difference on kinetic properties of the protein in the two cell line HOS or U2OS.

HOS_B3 cells were tansfected with MS2-EYFP-nls but not with the viral transactivator Tat, and a circle of 1.1µm was bleached in the nucleus. The recovery curve obtained, normalized as in equation 8, was initially fitted with the Sprague model for single binding site (Equation

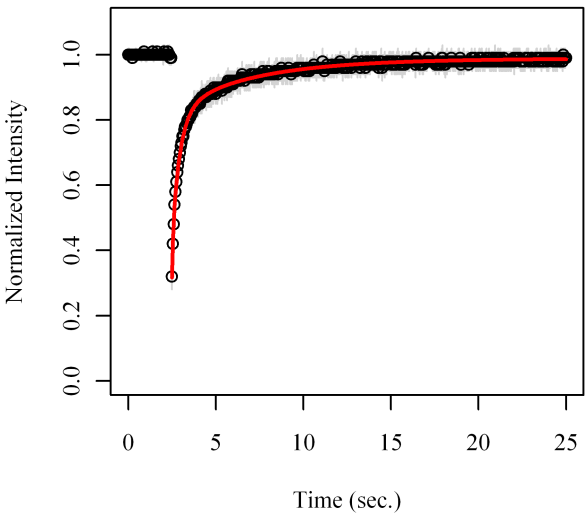

Figure VII.2: MS2-YFP-nls recovery curve in the nucleoplasm.

10,Fig. VII.2). Since the results suggest that the low diffusion coefficient value can reflects an effective diffusion condition, we fitted the curve also with the two independent binding site version of this model [Sprague et al., 2004], and we obtained a more reasonable value for the diffusion coefficient, 13 µm²/s. The results of the two model are summarized in table 5. It is possible to speculate that one binding site is the aspecifing binding of the MS2 protein with RNAs in the nucleus and the second is specific for nls, the nuclear localization signal.

***n* binding**	**D** [µm²/s]	τ_{d1} [s]	τ_{b1} [s]	τ_{d2} [s]	τ_{b2} [s]
I	2.4	333	77	/	/
II	13	20	4.7	0.6	0.4

Table 5: MS2-EYFP-nls FRAP in the nucleoplasm

VIII. BIBLIOGRAFY

Abate, 1996. Joseph Abate: On the Laguerrer Method for Numerically Inverting Laplace Transforms INFORMS Journal of Computing , 1996.

Ainger et al., 1993. K. Ainger and D. Avossa and F. Morgan and S. J. Hill and C. Barry and E. Barbarese and J. H. Carson: Transport and localization of exogenous myelin basic protein mRNA microinjected into oligodendrocytes. J Cell Biol ,1993.

Ansari and Hampsey, 2005. Athar Ansari and Michael Hampsey: A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping and Genes Dev ,2005.

Axelrod et al., 1976. D. Axelrod and D. E. Koppel and J. Schlessinger and E. Elson and W. W. Webb: Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. Biophys J ,1976.

Baeuerle and Baltimore, 1996. P. A. Baeuerle and D. Baltimore: NF-kappa B: ten years after. Cell ,1996.

Barrandon et al., 2007. Charlotte Barrandon and François Bonnet and Van Trung Nguyen and Valérie Labas and Olivier Bensaude: The transcription-dependent dissociation of P-TEFb-HEXIM1-7SK RNA relies upon formation of hnRNP-7SK RNA complexes. Mol Cell Biol ,2007.

Batsche et al., 2006. Eric Batsché and Moshe Yaniv and Christian Muchardt: The human SWI/SNF subunit Brm is a regulator of alternative splicing and Nat Struct Mol Biol ,2006.

Bayer et al., 1995. P. Bayer and M. Kraft and A. Ejchart and M. Westendorp and R. Frank and P. Rösch: Structural studies of HIV-1 Tat protein. J Mol Biol ,1995.

Beaudouin et al., 2006. Joël Beaudouin, Felipe Mora-Bermúdez and Thorsten Klee and Nathalie Daigle and Jan Ellenberg: Dissecting the contribution of diffusion and interactions to the mobility of nuclear proteins. Biophys J ,2006.

Berger et al., 1999. E. A. Berger and P. M. Murphy and J. M. Farber: Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. Annu Rev Immunol ,1999.

Berkhout and Jeang, 1992. B. Berkhout and K. T. Jeang: Functional roles for the TATA promoter and enhancers in basal and Tat-induced expression of the human immunodeficiency virus type 1 long terminal repeat. J Virol ,1992.

Berkowitz et al., 1995. R. D. Berkowitz and M. L. Hammarskjöld and C. Helga-Maria and D. Rekosh and S. P. Goff: 5' regions of HIV-1 RNAs are not sufficient for encapsidation: implications for the HIV-1 packaging signal. Virology ,1995.

Bertrand et al., 1998. E and Bertrand and P and Chartrand and M and Schaefer and S and M and Shenoy and R and H and Singer and R and M and Long: Localization of ASH1 mRNA particles in living yeast and Mol Cell ,1998.

Biancotto, 2006. Biancotto Chiara: Master thesis: “A method to monitor HIV-1 RNA in living cell”. Trieste University, 2006.

Bieniasz et al., 1998. P. D. Bieniasz and T. A. Grdina and H. P. Bogerd and B. R. Cullen: Recruitment of a protein complex containing Tat and cyclin T1 to TAR governs the species specificity of HIV-1 Tat. EMBO J ,1998.

Boireau et al., 2007. Stéphanie Boireau, Paolo Maiuri, Eugenia Basyuk, Manuel de la Mata, Anna Knezevich, Bérangère Pradet-Balade, Volker Bäcker, Alberto Kornblihtt, Alessandro Marcello, Edouard Bertrand: The transcriptional cycle of HIV-1 in real-time and live cells. J Cell Biol ,2007.

Braeckmans et al., 2003. Kevin Braeckmans and Liesbeth Peeters and Niek N Sanders and Stefaan C De Smedt and Joseph Demeester: Three-dimensional fluorescence recovery after photobleaching with the confocal scanning laser microscope and Biophys J ,2003.

Braga et al., 2004. José Braga and Joana M P Desterro and Maria Carmo-Fonseca: Intracellular macromolecular mobility measured by fluorescence recovery after photobleaching with confocal laser scanning microscopes and Mol Biol Cell ,2004.

Braga et al., 2007. José Braga and James G McNally and Maria Carmo-Fonseca: A reaction-diffusion model to study RNA motion by quantitative fluorescence recovery after photobleaching and Biophys J ,2007.

Bres et al., 2005. Vanessa Brès and Nathan Gomes and Loni Pickle and Katherine A Jones: A human splicing factor, SKIP, associates with

P-TEFb and enhances transcription elongation by HIV-1 Tat and Genes Dev ,2005.

Bukrinskaya et al., 1998. A. Bukrinskaya and B. Brichacek and A. Mann and M. Stevenson: Establishment of a functional human immunodeficiency virus type 1 (HIV-1) reverse transcription complex involves the cytoskeleton. J Exp Med ,1998.

Burton et al., 2005. Zachary F Burton and Michael Feig and Xue Q Gong and Chunfen Zhang and Yuri A Nedialkov and Yalin Xiong: NTP-driven translocation and regulation of downstream template opening by multi-subunit RNA polymerases and Biochem Cell Biol ,2005.

Campbell and Rein, 1999. S. Campbell and A. Rein: In vitro assembly properties of human immunodeficiency virus type 1 Gag protein lacking the p6 domain. J Virol ,1999.

Capranico et al., 2007. Giovanni Capranico, Francesca Ferri, Maria Vittoria Fogli, Alessandra Russo, Luca Lotito, Laura Baranello: The effects of camptothecin on RNA polymerase II transcription: roles of DNA topoisomerase I. Biochimie ,2007.

Chalfie et al., 1994. M and Chalfie and Y and Tu and G and Euskirchen and W and W and Ward and D and C and Prasher: Green fluorescent protein as a marker for gene expression and Science , 1994.

Chambeyron and Bickmore, 2004. Séverine Chambeyron and Wendy A Bickmore: Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. Genes Dev , 2004.

Chang and Jeang, 1992. Y. N. Chang and K. T. Jeang: The basic RNA-binding domain of HIV-2 Tat contributes to preferential trans-activation of a TAR2-containing LTR. Nucleic Acids Res ,1992.

Chene et al., 2007. Isaure du Chéné, Euguenia Basyuk, Yea-Lih Lin, Robinson Triboulet, Anna Knezevich, Christine Chable-Bessia, Clement Mettling, Vincent Baillat, Jacques Reynes, Pierre Corbeau, Edouard Bertrand, Alessandro Marcello, Stephane Emiliani, Rosemary Kiernan, Monsef Benkirane: Suv39H1 and HP1gamma are responsible for chromatin-mediated HIV-1 transcriptional silencing and post-integration latency. EMBO J , 2007.

Chubb et al., 2006. Jonathan R Chubb, Tatjana Trcek, Shailesh M Shenoy, Robert H Singer: Transcriptional pulsing of a developmental gene. Curr Biol ,2006.

Chun and Fauci, 1999. T. W. Chun and A. S. Fauci: Latent reservoirs of HIV: obstacles to the eradication of virus. Proc Natl Acad Sci U S A ,1999.

Churcher et al., 1993. M. J. Churcher and C. Lamont and F. Hamy and C. Dingwall and S. M. Green and A. D. Lowe and J. G. Butler and M. J. Gait and J. Karn: High affinity binding of TAR RNA by the human immunodeficiency virus type-1 tat protein requires base-pairs in the RNA stem and amino acid residues flanking the basic region. J Mol Biol ,1993.

Cinelli et al., 2001. R. A. Cinelli, V. Tozzini, V. Pellegrini, F. Beltram, G. Cerullo, M. Zavelani-Rossi, S. De Silvestri, M. Tyagi, M. Giacca: Coherent dynamics of photoexcited green fluorescent proteins. Phys Rev Lett ,2001.

Cody et al., 1993. C. W. Cody, D. C. Prasher, W. M. Westler, F. G. Prendergast, W. W. Ward: Chemical structure of the hexapeptide chromophore of the Aequorea green-fluorescent protein. Biochemistry ,1993.

Collins et al., 1998. K. L. Collins and B. K. Chen and S. A. Kalams and B. D. Walker and D. Baltimore: HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. Nature , 1998.

Cook, 1999. P. R. Cook: The organization of replication and transcription. Science ,1999.

Cormack et al., 1996. B. P. Cormack, R. H. Valdivia, S. Falkow: FACS-optimized mutants of the green fluorescent protein (GFP). Gene , 1996.

Coulter and Greenleaf, 1985. D. E. Coulter, A. L. Greenleaf: A mutation in the largest subunit of RNA polymerase II alters RNA chain elongation in vitro. J Biol Chem ,1985.

Cremer and Cremer, 2006. T and Cremer and C and Cremer: Rise, fall and resurrection of chromosome territories: a historical perspective and Part II and Fall and resurrection of chromosome territories during the 1950s to 1980s and Part III and Chromosome territories and the functional nuclear architecture: experiments and models from the 1990s to the present and Eur J Histochem ,2006.

Cremer and Cremer, 2006a. Thomas Cremer and C. Cremer: Rise, fall and resurrection of chromosome territories: a historical perspective. Part I. The rise of chromosome territories. Eur J Histochem ,2006.

Croft et al., 1999. J. A. Croft and J. M. Bridger and S. Boyle and P. Perry and P. Teague and W. A. Bickmore: Differences in the localization and morphology of chromosomes in the human nucleus. J Cell Biol ,1999.

Cujec et al., 1997. T. P. Cujec and H. Cho and E. Maldonado and J. Meyer and D. Reinberg and B. M. Peterlin: The human immunodeficiency virus transactivator Tat interacts with the RNA polymerase II holoenzyme. Mol Cell Biol ,1997.

Cullen, 1998. B. R. Cullen: Retroviruses as model systems for the study of nuclear RNA export pathways. Virology ,1998.

Darlix et al., 1990. J. L. Darlix and C. Gabus and M. T. Nugeyre and F. Clavel and F. Barré-Sinoussi: Cis elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. J Mol Biol ,1990.

Darzacq et al., 2007. Xavier Darzacq and Yaron Shav-Tal and Valeria de Turris and Yehuda Brody and Shailesh M Shenoy and Robert D Phair and Robert H Singer: In vivo dynamics of RNA polymerase II transcription and Nat Struct Mol Biol ,2007.

De Marco et al., 2008. Alex De Marco, Chiara Biancotto, Anna Knezevich, Paolo Maiuri, Chiara Vardabasso, Alessandro Marcello: Intragenic transcriptional cis-activation of the human immunodeficiency virus 1 does not result in allele-specific inhibition of the endogenous gene. Retrovirology ,2008.

Deng et al., 2002. Longwen Deng and Tatyana Ammosova and Anne Pumfery and Fatah Kashanchi and Sergei Nekhai: HIV-1 Tat interaction with RNA polymerase II C-terminal domain (CTD) and a dynamic association with CDK2 induce CTD phosphorylation and transcription from HIV-1 promoter. J Biol Chem ,2002.

Dieudonne et al., 2009. Mariacarolina Dieudonné, Paolo Maiuri, Chiara Biancotto, Anna Knezevich, Anna Kula, Marina Lusic, Alessandro Marcello: Transcriptional competence of the integrated HIV-1 provirus at the nuclear periphery. EMBO J ,2009.

Dingwall et al., 1989. C. Dingwall and I. Ernberg and M. J. Gait and S. M. Green and S. Heaphy and J. Karn and A. D. Lowe and M. Singh and M. A. Skinner and R. Valerio: Human immunodeficiency virus 1 tat protein binds trans-activation-responsive region (TAR) RNA in vitro. Proc Natl Acad Sci U S A ,1989.

Dundr et al., 2002. Miroslav Dundr and Urs Hoffmann-Rohrer and Qiyue Hu and Ingrid Grummt and Lawrence I Rothblum and Robert D Phair and Tom Misteli: A kinetic framework for a mammalian RNA polymerase in vivo and Science ,2002.

Epshtein and Nudler, 2003. Vitaly Epshtein and Evgeny Nudler: Cooperation between RNA polymerase molecules in transcription elongation and Science ,2003.

Epshtein et al., 2003. Vitaly Epshtein and Francine Toulmé and A and Rachid Rahmouni and Sergei Borukhov and Evgeny Nudler: Transcription through the roadblocks: the role of RNA polymerase cooperation and EMBO J ,2003.

Farnet and Haseltine, 1991. C. M. Farnet and W. A. Haseltine: Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. J Virol , 1991.

Forrest et al., 2003. Kevin M Forrest, Elizabeth R Gavis: Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in Drosophila. Curr Biol ,2003.

Fujinaga et al., 1998. K. Fujinaga and T. P. Cujec and J. Peng and J. Garriga and D. H. Price and X. Graña and B. M. Peterlin: The ability of positive transcription elongation factor B to transactivate human immunodeficiency virus transcription depends on a functional kinase domain, cyclin T1, and Tat. J Virol ,1998.

Fusco et al., 2003. Dahlene Fusco and Nathalie Accornero and Brigitte Lavoie and Shailesh M Shenoy and Jean Marie Blanchard and Robert H Singer and Edouard Bertrand: Single mRNA molecules demonstrate probabilistic movement in living mammalian cells and Curr Biol ,2003.

Ganser-Pornillos et al., 2008. Barbie K Ganser-Pornillos and Mark Yeager and Wesley I Sundquist: The structural biology of HIV assembly. Curr Opin Struct Biol ,2008.

Garcia et al., 1988. J. A. Garcia and D. Harrich and L. Pearson and R. Mitsuyasu and R. B. Gaynor: Functional domains required for tat-induced transcriptional activation of the HIV-1 long terminal repeat. EMBO J ,1988.

Gheysen et al., 1989. D. Gheysen and E. Jacobs and F. de Foresta and C. Thiriart and M. Francotte and D. Thines and M. De Wilde: Assembly and release of HIV-1 precursor Pr55gag virus-like particles from recombinant baculovirus-infected insect cells. Cell , 1989.

Golding and Cox, 2004. Ido Golding and Edward C Cox: RNA dynamics in live Escherichia coli cells and Proc Natl Acad Sci U S A ,2004.

Golding et al., 2005. Ido Golding and Johan Paulsson and Scott M Zawilski and Edward C Cox: Real-time kinetics of gene activity in

individual bacteria and Cell ,2005.

Goldman and Spector, 2005. R. D. Goldman, D. L. Spector: Live Cell Imaging. A laboratory Manual. ,2005.

Greatorex and Lever, 1998. J. Greatorex and A. Lever: Retroviral RNA dimer linkage. J Gen Virol ,1998.

Greene and Peterlin, 2002. Warner C Greene, B. Matija Peterlin: Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy. Nat Med ,2002.

GSL. M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, P. Alken, M. Booth, F. Rossi: GNU Scientific Library Reference Manual ,2009.

Harrich et al., 1990. D. Harrich and J. Garcia and R. Mitsuyasu and R. Gaynor: TAR independent activation of the human immunodeficiency virus in phorbol ester stimulated T lymphocytes. EMBO J ,1990.

Haseltine, 1991. W. A. Haseltine: Molecular biology of the human immunodeficiency virus type 1. FASEB J ,1991.

Heim et al., 1994. R. Heim, D. C. Prasher, R. Y. Tsien: Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proc Natl Acad Sci U S A ,1994.

Henkel et al., 1993. T. Henkel and T. Machleidt and I. Alkalay and M. Krönke and Y. Ben-Neriah and P. A. Baeuerle: Rapid proteolysis of I kappa B-alpha is necessary for activation of transcription factor NF-kappa B. Nature ,1993.

Herbert et al., 2006. Kristina M Herbert and Arthur La Porta and Becky J Wong and Rachel A Mooney and Keir C Neuman and Robert Landick and Steven M Block: Sequence-resolved detection of pausing by single RNA polymerase molecules and Cell ,2006.

Ho and Bieniasz, 2008. David D Ho and Paul D Bieniasz: HIV-1 at 25. Cell ,2008.

Inouye and Tsuji, 1994a. S and Inouye and F and I and Tsuji: Aequorea green fluorescent protein and Expression of the gene and fluorescence characteristics of the recombinant protein and FEBS Lett ,1994.

Janicki et al., 2004. Susan M Janicki and Toshiro Tsukamoto and Simone E Salghetti and William P Tansey and Ravi Sachidanandam and Kannanganattu V Prasanth and Thomas Ried and Yaron Shav-Tal and Edouard Bertrand and Robert H Singer and David L Spector: From silencing to gene expression: real-time analysis in single cells and Cell ,2004.

Jeang et al., 1993. K. T. Jeang and R. Chun and N. H. Lin and A. Gatignol and C. G. Glabe and H. Fan: In vitro and in vivo binding of human immunodeficiency virus type 1 Tat protein and Sp1 transcription factor. J Virol ,1993.

Jeang et al., 1999. K and T and Jeang and H and Xiao and E and A and Rich: Multifaceted activities of the HIV-1 transactivator of transcription, Tat and J Biol Chem ,1999.

Johansson et al., 1998. H. E. Johansson, D. Dertinger, K. A. LeCuyer, L. S. Behlen, C. H. Greef, O. C. Uhlenbeck: A thermodynamic analysis of the sequence-specific binding of RNA by bacteriophage MS2 coat protein. Proc Natl Acad Sci U S A ,1998.

Johnson et al., 1962. F. H. Johnson, O. Shimomura, Y. Saiga: Action of cyanide on Cypridina luciferin. J Cell Comp Physiol ,1962.

Jordan et al., 2001. A and Jordan and P and Defechereux and E and Verdin: The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation and EMBO J ,2001.

Jordan et al., 2003. Albert Jordan, Dwayne Bisgrove, Eric Verdin: HIV reproducibly establishes a latent infection after acute infection of T cells in vitro. EMBO J ,2003.

Kamine et al., 1996. J. Kamine and B. Elangovan and T. Subramanian and D. Coleman and G. Chinnadurai: Identification of a cellular protein that specifically interacts with the essential cysteine region of the HIV-1 Tat transactivator. Virology ,1996.

Kang and Kenworthy, 2008. Minchul Kang, Anne K Kenworthy: A closed-form analytic expression for FRAP formula for the binding diffusion model. Biophys J ,2008.

Kao et al., 1987. S. Y. Kao and A. F. Calman and P. A. Luciw and B. M. Peterlin: Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. Nature ,1987.

Karn, 1999. J. Karn: Tackling Tat. J Mol Biol ,1999.

Kashanchi et al., 1994. F. Kashanchi and G. Piras and M. F. Radonovich and J. F. Duvall and A. Fattaey and C. M. Chiang and R. G. Roeder and J. N. Brady: Direct interaction of human TFIID with the HIV-1 transactivator tat. Nature ,1994.

Kass et al., 1990. S. Kass, K. Tyc, J. A. Steitz, B. Sollner-Webb: The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. Cell ,1990.

Kharroubi et al., 1998. A. El Kharroubi and G. Piras and R. Zensen and

M. A. Martin: Transcriptional activation of the integrated chromatin-associated human immunodeficiency virus type 1 promoter. Mol Cell Biol ,1998.

Kimura et al., 2002. Hiroshi Kimura and Kimihiko Sugaya and Peter R Cook: The transcription cycle of RNA polymerase II in living cells and J Cell Biol ,2002.

Kirber et al., 2007. Michael T Kirber, Kai Chen, John F Keaney: YFP photoconversion revisited: confirmation of the CFP-like species. Nat Methods ,2007.

Kozak, 1989. M. Kozak: The scanning model for translation: an update. J Cell Biol ,1989.

Le et al., 2005. Thuc T Le, Sébastien Harlepp, Calin C Guet, Kimberly Dittmar, Thierry Emonet, Tao Pan, Philippe Cluzel: Real-time RNA profiling within a single bacterium. Proc Natl Acad Sci U S A , 2005.

Lint et al., 1996. C. Van Lint and S. Emiliani and M. Ott and E. Verdin: Transcriptional activation and chromatin remodeling of the HIV-1 promoter in response to histone acetylation. EMBO J ,1996.

Lippincott-Schwartz et al., 2003. Jennifer Lippincott-Schwartz and Nihal Altan-Bonnet and George H Patterson: Photobleaching and photoactivation: following protein dynamics in living cells and Nat Cell Biol ,2003.

Lusic et al., 2003. Marina Lusic, Alessandro Marcello, Anna Cereseto, Mauro Giacca: Regulation of HIV-1 gene expression by histone acetylation and factor recruitment at the LTR promoter. EMBO J , 2003.

Mahy et al., 2002. Nicola L Mahy and Paul E Perry and Wendy A Bickmore: Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. J Cell Biol ,2002.

Majello et al., 1998. B. Majello and G. Napolitano and L. Lania: Recruitment of the TATA-binding protein to the HIV-1 promoter is a limiting step for Tat transactivation. AIDS ,1998.

Marcello and Giaretta, 1998. A and Marcello and I and Giaretta: Inducible expression of herpes simplex virus thymidine kinase from a bicistronic HIV1 vector and Res Virol ,1998.

Marcello et al., 2001. A and Marcello and R and A and Cinelli and A and Ferrari and A and Signorelli and M and Tyagi and V and Pellegrini and F and Beltram and M and Giacca: Visualization of in vivo direct interaction between HIV-1 TAT and human cyclin T1

in specific subcellular compartments by fluorescence resonance energy transfer and J Biol Chem ,2001.

Marcello et al., 2003. Alessandro Marcello and Aldo Ferrari and Vittorio Pellegrini and Gianluca Pegoraro and Marina Lusic and Fabio Beltram and Mauro Giacca: Recruitment of human cyclin T1 to nuclear bodies through direct interaction with the PML protein and EMBO J ,2003.

Marcello et al., 2004. Alessandro Marcello, Marina Lusic, Gianluca Pegoraro, Vittorio Pellegrini, Fabio Beltram, Mauro Giacca: Nuclear organization and the control of HIV-1 transcription. Gene ,2004.

Marcello, 2006. Alessandro Marcello: Latency: the hidden HIV-1 challenge. Retrovirology ,2006.

Margottin et al., 1998. F. Margottin, S. P. Bour, H. Durand, L. Selig, S. Benichou, V. Richard, D. Thomas, K. Strebel, R. Benarous: A novel human WD protein, h-beta TrCp, that interacts with HIV-1 Vpu connects CD4 to the ER degradation pathway through an F-box motif. Mol Cell ,1998.

Marzio et al., 1998. G. Marzio, M. Tyagi, M. I. Gutierrez, M. Giacca: HIV-1 tat transactivator recruits p300 and CREB-binding protein histone acetyltransferases to the viral promoter. Proc Natl Acad Sci U S A ,1998.

Mason and Struhl, 2005. Paul B Mason and Kevin Struhl: Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo and Mol Cell ,2005.

Mata et al., 2003. Manuel de la Mata and Claudio R Alonso and Sebastián Kadener and Juan P Fededa and Matías Blaustein and Federico Pelisch and Paula Cramer and David Bentley and Alberto R Kornblihtt: A slow RNA polymerase II affects alternative splicing in vivo and Mol Cell ,2003.

Mazza et al., 2008. Davide Mazza, Kevin Braeckmans and Francesca Cella and Ilaria Testa and Dries Vercauteren and Jo Demeester and Stefaan S De Smedt and Alberto Diaspro: A new FRAP/FRAPa method for three-dimensional diffusion measurements based on multiphoton excitation microscopy. Biophys J ,2008.

McBride and Panganiban, 1996. M. S. McBride and A. T. Panganiban: The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures. J Virol ,1996.

McDonald et al., 2002. David McDonald and Marie A Vodicka and

Ginger Lucero and Tatyana M Svitkina and Gary G Borisy and Michael Emerman and Thomas J Hope: Visualization of the intracellular behavior of HIV in living cells. J Cell Biol ,2002.

McNally et al., 2000. J. G. McNally and W. G. Müller and D. Walker and R. Wolford and G. L. Hager: The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. Science ,2000.

Miesenboeck et al., 1998. G. Miesenböck, D. A. De Angelis, J. E. Rothman: Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. Nature ,1998.

Misteli and Spector, 1998. T. Misteli, D. L. Spector: The cellular organization of gene expression. Curr Opin Cell Biol ,1998.

Misteli, 2007. Tom Misteli: Beyond the sequence: cellular organization of genome function. Cell ,2007.

Miyawaki et al., 2003. Atsushi Miyawaki, Asako Sawano, Takako Kogure: Lighting up cells: labelling proteins with fluorophores. Nat Cell Biol ,2003.

Molle et al., 2007. Dorothée Molle, Paolo Maiuri, Stéphanie Boireau, Edouard Bertrand, Anna Knezevich, Alessandro Marcello, Eugenia Basyuk: A real-time view of the TAR:Tat:P-TEFb complex at HIV-1 transcription sites. Retrovirology ,2007.

Morise et al., 1974. H and Morise and O and Shimomura and F and H and Johnson and J and Winant: Intermolecular energy transfer in the bioluminescent system of Aequorea and Biochemistry ,1974.

Mueller et al., 2008. Florian Mueller, Paul Wach, James G McNally: Evidence for a common mode of transcription factor interaction with chromatin as revealed by improved quantitative fluorescence recovery after photobleaching. Biophys J ,2008.

Nguyen et al., 2001. V. T. Nguyen and T. Kiss and A. A. Michels and O. Bensaude: 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature ,2001.

O'Sullivan et al., 2004. Justin M O'Sullivan and Sue Mei Tan-Wong and Antonin Morillon and Barbara Lee and Joel Coles and Jane Mellor and Nick J Proudfoot: Gene loops juxtapose promoters and terminators in yeast and Nat Genet ,2004.

Ormoe et al., 1996. M. Ormö, A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien, S. J. Remington: Crystal structure of the Aequorea victoria green fluorescent protein. Science ,1996.

Osborne et al., 2004. Cameron S Osborne and Lyubomira Chakalova and Karen E Brown and David Carter and Alice Horton and Emmanuel

Debrand and Beatriz Goyenechea and Jennifer A Mitchell and Susana Lopes and Wolf Reik and Peter Fraser: Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet ,2004.

Pantano et al., 2006. Sergio Pantano and Alessandro Marcello and Aldo Ferrari and Daniele Gaudiosi and Arianna Sabò and Vittorio Pellegrini and Fabio Beltram and Mauro Giacca and Paolo Carloni: Insights on HIV-1 Tat:P/CAF bromodomain molecular recognition from in vivo experiments and molecular dynamics simulations and Proteins ,2006.

Parada and Roeder, 1996. C. A. Parada and R. G. Roeder: Enhanced processivity of RNA polymerase II triggered by Tat-induced phosphorylation of its carboxy-terminal domain. Nature ,1996.

Perkins et al., 2008. Kelly J Perkins, Marina Lusic, Ivonne Mitar, Mauro Giacca, Nick J Proudfoot: Transcription-dependent gene looping of the HIV-1 provirus is dictated by recognition of pre-mRNA processing signals. Mol Cell ,2008.

Phair and Misteli, 2000. R and D and Phair and T and Misteli: High mobility of proteins in the mammalian cell nucleus and Nature , 2000.

Pierson et al., 2000. T. Pierson and J. McArthur and R. F. Siliciano: Reservoirs for HIV-1: mechanisms for viral persistence in the presence of antiviral immune responses and antiretroviral therapy. Annu Rev Immunol ,2000.

Politz et al., 1998. J. C. Politz and E. S. Browne and D. E. Wolf and T. Pederson: Intranuclear diffusion and hybridization state of oligonucleotides measured by fluorescence correlation spectroscopy in living cells. Proc Natl Acad Sci U S A ,1998.

Pombo and Branco, 2007. Ana Pombo, Miguel R Branco: Functional organisation of the genome during interphase. Curr Opin Genet Dev ,2007.

Prasher et al., 1992. D and C and Prasher and V and K and Eckenrode and W and W and Ward and F and G and Prendergast and M and J and Cormier: Primary structure of the Aequorea victoria green-fluorescent protein and Gene ,1992.

Prendergast and Mann, 1978. F and G and Prendergast and K and G and Mann: Chemical and physical properties of aequorin and the green fluorescent protein isolated from Aequorea forskålea and Biochemistry ,1978.

Price, 2000. D. H. Price: P-TEFb, a cyclin-dependent kinase controlling

elongation by RNA polymerase II. Mol Cell Biol ,2000.

R-project. {R Development Core Team}: R: A Language and Environment for Statistical Computing ,2009.

Rahbek, 2008. Rahbek, Ulrik Lytt: Small interfering rna delivery, trafficking and gene silencing using polymeric nanoparticles. , 2008.

Rao et al., 1980. B. D Nageswara Rao, M. D. Kemple, F. G. Prendergast: PROTON NUCLEAR MAGNETIC RESONANCE AND FLUORESCENCE SPECTROSCOPIC STUDIES OF SEGMENTAL MOBILITY IN AEQUORIN AND A GREEN FLUORESCENT PROTEIN FROM AEQUOREA FORSKALEA. Biophys J ,1980.

Re et al., 1995. F. Re and D. Braaten and E. K. Franke and J. Luban: Human immunodeficiency virus type 1 Vpr arrests the cell cycle in G2 by inhibiting the activation of p34cdc2-cyclin B. J Virol ,1995.

Reid and Flynn, 1997. B. G. Reid, G. C. Flynn: Chromophore formation in green fluorescent protein. Biochemistry ,1997.

Richman et al., 2009. Douglas D Richman and David M Margolis and Martin Delaney and Warner C Greene and Daria Hazuda and Roger J Pomerantz: The challenge of finding a cure for HIV infection. Science ,2009.

Robinett et al., 1996. C. C. Robinett, A. Straight, G. Li, C. Willhelm, G. Sudlow, A. Murray, A. S. Belmont: In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. J Cell Biol ,1996.

Ruben et al., 1989. S. Ruben and A. Perkins and R. Purcell and K. Joung and R. Sia and R. Burghoff and W. A. Haseltine and C. A. Rosen: Structural and functional characterization of human immunodeficiency virus tat protein. J Virol ,1989.

Sage et al., 2005. Daniel Sage, Franck R Neumann, Florence Hediger, Susan M Gasser, Michael Unser: Automatic tracking of individual fluorescence particles: application to the study of chromosome dynamics. IEEE Trans Image Process ,2005.

Schroeder et al., 2002. Astrid R W Schröder, Paul Shinn, Huaming Chen, Charles Berry, Joseph R Ecker, Frederic Bushman: HIV-1 integration in the human genome favors active genes and local hotspots. Cell ,2002.

Shav-Tal et al., 2004. Yaron Shav-Tal and Xavier Darzacq and Shailesh M Shenoy and Dahlene Fusco and Susan M Janicki and David L Spector and Robert H Singer: Dynamics of single mRNPs in nuclei of living cells. Science ,2004.

Shav-Tal et al., 2004a. Yaron Shav-Tal and Robert H Singer and Xavier Darzacq: Imaging gene expression in single living cells and Nat Rev Mol Cell Biol ,2004.

Sheehy et al., 2003. Ann M Sheehy and Nathan C Gaddis and Michael H Malim: The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. Nat Med ,2003.

Sheridan et al., 1997. P. L. Sheridan and T. P. Mayall and E. Verdin and K. A. Jones: Histone acetyltransferases regulate HIV-1 enhancer activity in vitro. Genes Dev ,1997.

Shimomura et al., 1962. O. Shimomura, F. H. Johnson, Y. Saiga: Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. J Cell Comp Physiol ,1962.

Sodroski et al., 1985. J. Sodroski and C. Rosen and F. Wong-Staal and S. Z. Salahuddin and M. Popovic and S. Arya and R. C. Gallo and W. A. Haseltine: Trans-acting transcriptional regulation of human T-cell leukemia virus type III long terminal repeat. Science ,1985.

Soumpasis, 1983. D. M. Soumpasis: Theoretical analysis of fluorescence photobleaching recovery experiments. Biophys J ,1983.

Spector, 2001. D. L. Spector: Nuclear domains. J Cell Sci ,2001.

Sprague and McNally, 2005. Brian L Sprague and James G McNally: FRAP analysis of binding: proper and fitting. Trends Cell Biol , 2005.

Sprague et al., 2004. Brian L Sprague and Robert L Pego and Diana A Stavreva and James G McNally: Analysis of binding reactions by fluorescence recovery after photobleaching and Biophys J ,2004.

Sun et al., 2000. H. B. Sun and J. Shen and H. Yokota: Size-dependent positioning of human chromosomes in interphase nuclei. Biophys J ,2000.

Tennyson et al., 1995. C and N and Tennyson and H and J and Klamut and R and G and Worton: The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced and Nat Genet ,1995.

Thaler et al., 2006. Christopher Thaler, Steven S Vogel, Stephen R Ikeda, Huanmian Chen: Photobleaching of YFP does not produce a CFP-like species that affects FRET measurements. Nat Methods , 2006.

Tong-Starksen et al., 1993. S. E. Tong-Starksen and A. Baur and X. B. Lu and E. Peck and B. M. Peterlin: Second exon of Tat of HIV-2 is

required for optimal trans-activation of HIV-1 and HIV-2 LTRs. Virology ,1993.

Tsien, 1998. R. Y. Tsien: The green fluorescent protein. Annu Rev Biochem ,1998.

Tsien, 1998a. R and Y and Tsien: The green fluorescent protein and Annu Rev Biochem ,1998.

Valentin et al., 2005. Guillaume Valentin, Céline Verheggen, Tristan Piolot, Henry Neel, Maïté Coppey-Moisan, Edouard Bertrand: Photoconversion of YFP into a CFP-like species during acceptor photobleaching FRET experiments. Nat Methods ,2005.

Vardabasso et al., 2008. Chiara Vardabasso and Lara Manganaro and Marina Lusic and Alessandro Marcello and Mauro Giacca: The histone chaperone protein Nucleosome Assembly Protein-1 (hNAP-1) binds HIV-1 Tat and promotes viral transcription and Retrovirology ,2008.

Verdin et al., 1993. E. Verdin and P. Paras and C. Van Lint: Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation. EMBO J ,1993.

Volpi et al., 2000. E. V. Volpi and E. Chevret and T. Jones and R. Vatcheva and J. Williamson and S. Beck and R. D. Campbell and M. Goldsworthy and S. H. Powis and J. Ragoussis and J. Trowsdale and D. Sheer: Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. J Cell Sci ,2000.

Wachter et al., 1997. R. M. Wachter, B. A. King, R. Heim, K. Kallio, R. Y. Tsien, S. G. Boxer, S. J. Remington: Crystal structure and photodynamic behavior of the blue emission variant Y66H/Y145F of green fluorescent protein. Biochemistry ,1997.

Wang et al., 2006. Dong Wang and David A Bushnell and Kenneth D Westover and Craig D Kaplan and Roger D Kornberg: Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis and Cell ,2006.

Wansink et al., 1993. D. G. Wansink and W. Schul and I. van der Kraan and B. van Steensel and R. van Driel and L. de Jong: Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. J Cell Biol ,1993.

Ward and Bokman, 1982. W. W. Ward, S. H. Bokman: Reversible denaturation of Aequorea green-fluorescent protein: physical separation and characterization of the renatured protein.

Biochemistry ,1982.

Wei et al., 1998. P and Wei and M and E and Garber and S and M and Fang and W and H and Fischer and K and A and Jones: A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA and Cell ,1998.

Weinberger et al., 2005. Leor S Weinberger, John C Burnett, Jared E Toettcher, Adam P Arkin, David V Schaffer: Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. Cell ,2005.

Wu-Baer et al., 1995. F. Wu-Baer and D. Sigman and R. B. Gaynor: Specific binding of RNA polymerase II to the human immunodeficiency virus trans-activating region RNA is regulated by cellular cofactors and Tat. Proc Natl Acad Sci U S A ,1995.

Yang et al., 1996. T. T. Yang, L. Cheng, S. R. Kain: Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. Nucleic Acids Res ,1996.

Yao et al., 2006. Jie Yao, Katherine M Munson, Watt W Webb, John T Lis: Dynamics of heat shock factor association with native gene loci in living cells. Nature ,2006.

Zapp and Green, 1989. M. L. Zapp and M. R. Green: Sequence-specific RNA binding by the HIV-1 Rev protein. Nature ,1989.

Zhou et al., 2002. Zhaolan Zhou and Lawrence J Licklider and Steven P Gygi and Robin Reed: Comprehensive proteomic analysis of the human spliceosome. Nature ,2002.

Zimmer, 2002. Marc Zimmer: Green fluorescent protein (GFP): applications, structure, and related photophysical behavior. Chem Rev ,2002.