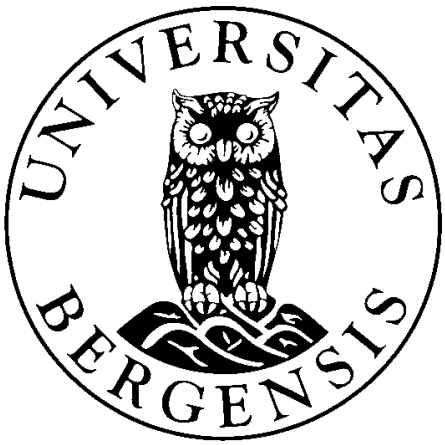
Functional studies of two wild type variants of HIV-1 Tat

A thesis in partial fulfilment of the requirements for the degree of ***Candidatus scientiarum.*** Submitted to the University of Bergen.

Håvard Nyhagen Henriksen



University of Bergen

Department of Molecular Biology

2003

Acknowledgements

I would like to thank my supervisor, professor Dag E. Helland, for giving me something to do, for supporting me in, and never abandoning the hope that I would eventually finish, doing it; also my assistant supervisor, professor Anne Marie Szilvay, for encouragement and for pushing me, especially the latter; I sincerely doubt you would be reading this for another century if she had not. I am also beholden to them for reading this manuscript. I am indebted to Bente Skjellstad, for much help, and for countless procedures, without which I would probably have known how to look up things for myself a great deal sooner; Vibeke Andresen, for vast amounts of help, and for her non-violent nature, which hindered her in kicking my ass with her mad kickboxing skills when I was being too much of a nuisance; and Vidar Staalesen, among other things, for mentally preparing me for being stuck in an elevator. Thanks go also to numerous others in the laboratory and the rest of MBI, not only for tolerating me when I pestered them, but also for actually being nice to me while doing that.

Extra special thanks go to Håvard Valvatne for having the same first name as me. This made it very convenient for me to steal the clogs he left behind.

Thanks go also to the Internet, for wasting my time; and also my parents, for letting me leech of them while I was conducting this work.

Bergen, July 31st, 2003

Håvard Henriksen

Abbreviations

AIDS	Acquired Immunodeficiency Syndrome
Amp	Ampicillin
APS	Ammonium Persulphate
A_x	Absorbance at λ=X
BFP	Blue Fluorescent Protein
BSA	Bovine serum albumine
CA	HIV Capsid Protein
CAT	Chloramphenicol Acetyl Transferase
CDK9	Cyclin Dependent Kinase 9
CFP	Cyan Fluorescent Protein
Chl	Chloramphenicol
CIAP	Calf Intestinal Alkaline Phosphatase
CTD	C-Terminal Domain of the major subunit of RNA Polymerase II
Da	Dalton
dH_2O	Sterilized Water
ECL	Enhanced Chemoluminiscense
EDTA	Ethylenediamminetetraacetate
EGFP	Enhanced Green Fluorescent Protein
ESS2	Exon splicing silencer 2
E. coli	Eschericia coli
FPLC	Fast Pressure Liquid Chromatography
GFP	Green Fluorescent Protein
gp	Glycoprotein
HAART	Highly Active Anti-Retroviral Therapy
HAT	Histone Acetyl Transferase
His-tag	Polyhistidine sequence
HIV	Human Immunodeficiency Virus
HIV-1	Human Immunodeficiency Virus type 1
HIV-2	Human Immunodeficiency Virus type 2
HSPG	Heparan Sulfate Proteoglycan
HXB2	A particular strain of HIV-1
HXB3	A particular strain of HIV-1

IMAC	Immobilized metal ion affinity chromatography
IN	HIV-1 Integrase
IPTG	Isopropyl β-D-thiogalactoside
kDa	Kilodalton
LAI	A particular strain of HIV-1
LB	Luria Bertani/Luria Broth
LTR	Long Terminal Repeat
MA	Matrix protein
mAb	Monoclonal Antibody
MCAC	Metal Chelate Affinity Chromatography
MCS	Multiple Cloning Site
MVP5180	A particular strain of HIV-1
NC	HIV Nucleocapsid Protein
NES	Nuclear Export Signal
NLS	Nuclear Localization Signal
NMR	Nuclear Magnetic Resonance
NOS	Nucleolar Localisation Signal
NTEF	Negative Transcription Elongation Factors
O/n	Over Night
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate-buffered saline
PCR	Polymerase Chain Reaction
PFA	Paraformaldehyde
PIC	HIV-1 Preintegration Complex
PKR	A dsDNA Dependent Protein Kinase
PMSF	Phenyl Methyl Sulfonyl Fluoride
POD	Horseradish Peroxidase
PR	HIV-1 Protease
PTEF	Positive Transcription Elongation Factor
RFP	Red Fluorescent Protein
RGD	The amino acid triplet arginine-glycine-aspartic acid sequence
RNAP II	RNA Polymerase II
RT	HIV-1 Reverse Transcriptase

SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SIV	Simian Immunodeficiency Virus
SU	Surface Protein
TAR	*Trans*-Activating Response Element
Tat	*Trans*-cellular Transactivator Protein
Tat72	One-exon encoded Tat protein
Tat86	Two-exon encoded Tat protein comprising 86 amino acids
Tat101	Two-exon encoded Tat protein comprising 101 amino acids
Tat115	Two-exon encoded Tat protein comprising 115 amino acids
TEMED	Tetramethylendiamid
T_m	Melting temperature of dsDNA
TM	Transmembrane Protein
u	Units of enzyme activity
YFP	Yellow Fluorescent Protein

Contents

Aims and summary of the present study

The HIV-1 encoded Tat protein is essential for viral propagation. It transactivates the viral LTR promoter, and it also has transcellular activities as it can be transferred from one cell to another. Most studies of the Tat protein have been performed using Tat from a few laboratory-adapted strains of HIV-1. These strains encode an 86 amino acid variant of the protein. However, most strains encode a 101 amino acid long variant. It has been speculated that the 86 amino acid version is a truncated form of the protein, and that it does not have all the properties of the full-length variant.

The main aims of this study were (I) to construct two full-length wild type variants of the *tat* gene. This would allow the study of the properties of these full-length versions of Tat. (II) To perform a computational analysis of the proteins, to predict their properties. (III) To insert the genes into bacterial expression vectors, which would allow the purification of the Tat protein, and (IV) the Tat proteins would then tested for transcellular activation. (V) The genes were also to be inserted into vectors allowing production of fusion proteins with the fluorescing protein EGFP. (VI) The intracellular distribution of these proteins was then to be studied in transfected cells by fluorescence microscopy, and (VII) the transactivating properties of the EGFP-Tat fusion proteins was then to be verified using an LTR-reporter gene construct.

(I) The HXB3 clone encodes an 86 amino acid Tat protein and was available in our laboratory. It was decided to construct a 101 amino acid Tat protein based on the sequence of this clone. Then the abilities of the protein would be comparable to those of the Tat86 protein available for study. A point mutation was made to alter the stop codon at codon 87. This gene was then named *tat101*. A second *tat* gene was constructed in order to study differences among Tat proteins from different strains. This second *tat* gene was made based on the sequence of the MVP5180 strain of HIV, and was called *tat115*.

(II) The computer analyses showed a possible novel domain, named the EK domain, as well as a potential new nuclear localization signal in the Tat101 protein.

(III) The *tat* genes were successfully inserted into the pRSET-C vector. The Tat101 protein was purified by the aid of an N-terminal 6xhis tag. There were problems purifying the Tat115 protein, so this was not done.

(IV) The Tat101 protein was added to the media of cultured COS-7 cells transfected with and LTR-CAT plasmid. And it was shown that this induced the expression of the reporter protein.

(V) The *tat* genes were successfully inserted into the pEGFP-C1 vector.

(VI) By fluorescence microscopy the Tat101 and Tat115 proteins fused to EGFP were shown to be located in the nucleus and the nucleolus when expressed in COS-7 cells.

(VII) It was shown that the EGFP-Tat115 and EGFP-Tat101 fusion proteins induced the expression of the CAT reporter protein when the LTR-CAT plasmid was co-transfected with either of the EGFP-Tat plasmids.

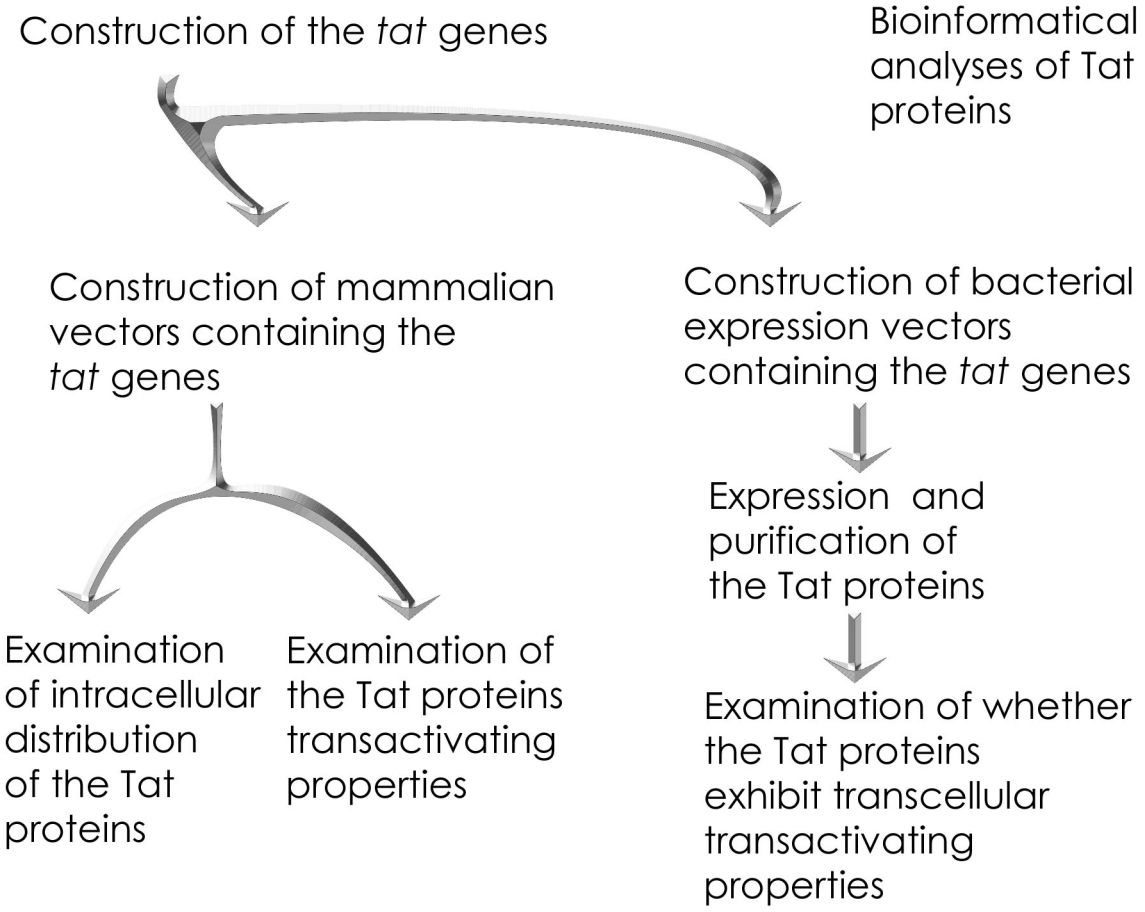


Figure 0.1: An overview of the aims of this study.

1 Introduction

1.1 HIV and AIDS

The Human Immunodeficiency Virus (HIV) is a retrovirus of the lentivirus family. Lenti means slow, and the name reflects the long incubation period, the time it takes from infection until the associated disease starts developing. The virus is mainly transmittable through bodily fluids, primarily blood and semen. A person infected with the virus has a gradual inactivation of the immune system. Characteristically, the CD4+ T cells are depleted. This causes the disease Acquired Immunodeficiency Syndrome (AIDS). With their immune system in disarray the bearer of the virus is more susceptible to opportunistic infections and several, otherwise rare, types of cancer (Blattner, 1999; Tirelli et al., 2002; http://www.who.int).

AIDS is a global epidemic. At the end of 2002 the number of people with HIV or AIDS in Norway was estimated to 2555. Globally it was estimated to 42 million. Of these 29.4 million were located in the southern parts of Africa where the disease is most widespread. More than 25 million people have died AIDS-related deaths since the discovery of this disease in 1981 (http://www.unaids.org; www.fhi.no). The infrastructure of those countries where the disease is most prevalent is threatened by a future collapse (reviewed by Piot *et al.*, 2001).

Several steps are being taken to combat this disease. One important step is health education campaigns that inform the public of how AIDS is transmitted, thereby trying to limit the number of people being exposed to the virus. Another step is the development of drugs to treat the disease. Research that gives us knowledge of the molecular mechanisms of how the virus interacts with its host will drive forward the development of novel therapies. No known therapy can completely rid the body of the virus. Several vaccines are currently in different stages of clinical trials in humans. These vaccines are, however, therapeutic vaccines, aimed at controlling the infection rather than preventing it (reviewed by Amara and Robinson, 2002). A combination of drugs called HAART (Highly Active Anti-Retroviral Therapy) has proved very efficient in stalling the disease, but there are serious problems with toxic side effects (Louie *et al.,* 2002; http://hivmedicine.com/index.htm).

A major problem in the development of treatments and therapies for HIV infection is the high frequency with which the virus mutates. Mutations also cause the changes that make the immune system unable to recognize the virus (Wei *et al*., 2003). The viral encoded reverse transcriptase protein (RT) transcribes genomic RNA to double stranded DNA, which is

integrated into the host genome. RT does not have a proof-reading activity, and is the main cause of mutations of the virus (Preston and Dougherty, 1996). This accounts for the many different strains and subtypes of HIV.

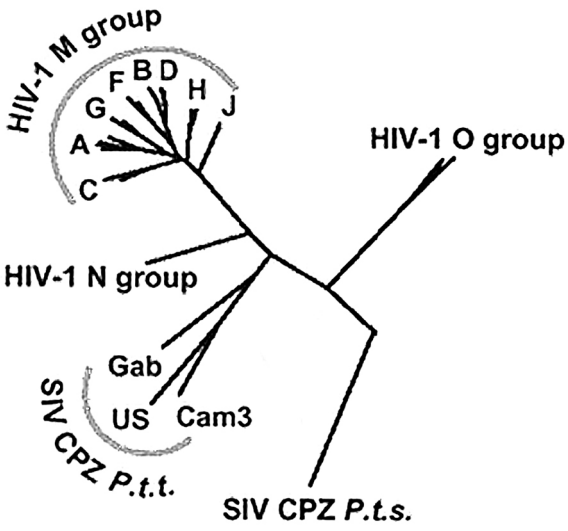


Figure 1.1: Phylogenetic (i.e. evolutionary) tree of different strains of HIV-1 and SIV CPZ, showing how they are related. The length from one indicated point to another shows relative genetic distance. HIV-1 group M is the main cause of the global pandemic and all the subtypes within are likely to have descended after a single zoonotic transfer (cross-species transfer from the natural host, a non-human, to a human) of a simian virus. HIV groups N and O are likely to have arisen after separate zoonotic transmissions. SIV CPZ P.t.t.: strains of SIV found in *Pan troglodytes troglodytes*, a subspecies of chimpanzee. SIV CPZ P.t.s.: a strain of SIV found in a chimpanzee of the subspecies *Pan troglodytes schweinfurthii*. (Modified from Reeves and Doms, 2002)

HIV is classified as two major types: HIV-1 and HIV-2. HIV-1 is the most common and infectious of these two, and can further be divided into 3 groups: M, N and O. HIV-1 Group M is by far the most important contributor to the AIDS pandemic and can further be divided into 10 subtypes, A-H and J-K (http://hiv-web.lanl.gov). When comparing the sequences of different strains of HIV with those of Simian Immunodeficiency Virus (SIV), the interspersion in the evolutionary tree (see figure 1.1) suggests shared viral lineages (Hahn *et al.*, 2000; Korber *et al.*, 2000), and that HIV has arisen when SIV has been transmitted from its natural simian host to humans. It also suggests that this sort of transmission has occurred several times. For instance, it is likely that HIV-1 groups M, N and O have arisen from separate zoonotic transmissions of SIV. HIV-1 and HIV-2 are related to quite different strains of SIV, SIV from chimpanzees, $SIV_{CPZ,}$ and sooty mangabey, SIV_{SM} respectively. Despite its name, most strains of SIV do not cause disease in their natural host.

A sample from 1959 shows that HIV-1 was in a human population already at that time (Zhu *et al.* 1998) and sequence analyses estimate that the last common ancestor of the HIV-1 M group existed in a human host some time between 1915 and 1941 (Korber *et al.* 2000).

What effect the divergence in sequence between the different strains has on virulence, transmission rates and general epidemiology has not been clarified, but the different HIV-1

types all have the same basic structure (shown in figure 1.2), and the same life cycle (figure 1.3).

1.2 The physical and genetic structure of HIV

A model of the virus particle is depicted in figure 1.2.A. It is spherical with a membrane that originates from the host cell plasma membrane. Located on the surface of the membrane is a protein called gp120 or the Surface protein (SU). It is anchored through gp41, the Transmembrane protein (TM). TM is embedded in the membrane. Underneath the plasma membrane is an icosahedron made by the Matrix protein (MA). Inside the icosahedron the Capsid protein (CA) forms a cone in the mature virus particle. This cone is called the core particle. Within this cone are the proteins Reverse Transcriptase (RT), Integrase (IN), Protease (PR), and Nucleocapsid (NC). The latter is associated with the two RNA copies of the HIV genome. This association serves to stabilize the genome. In addition the viral particle

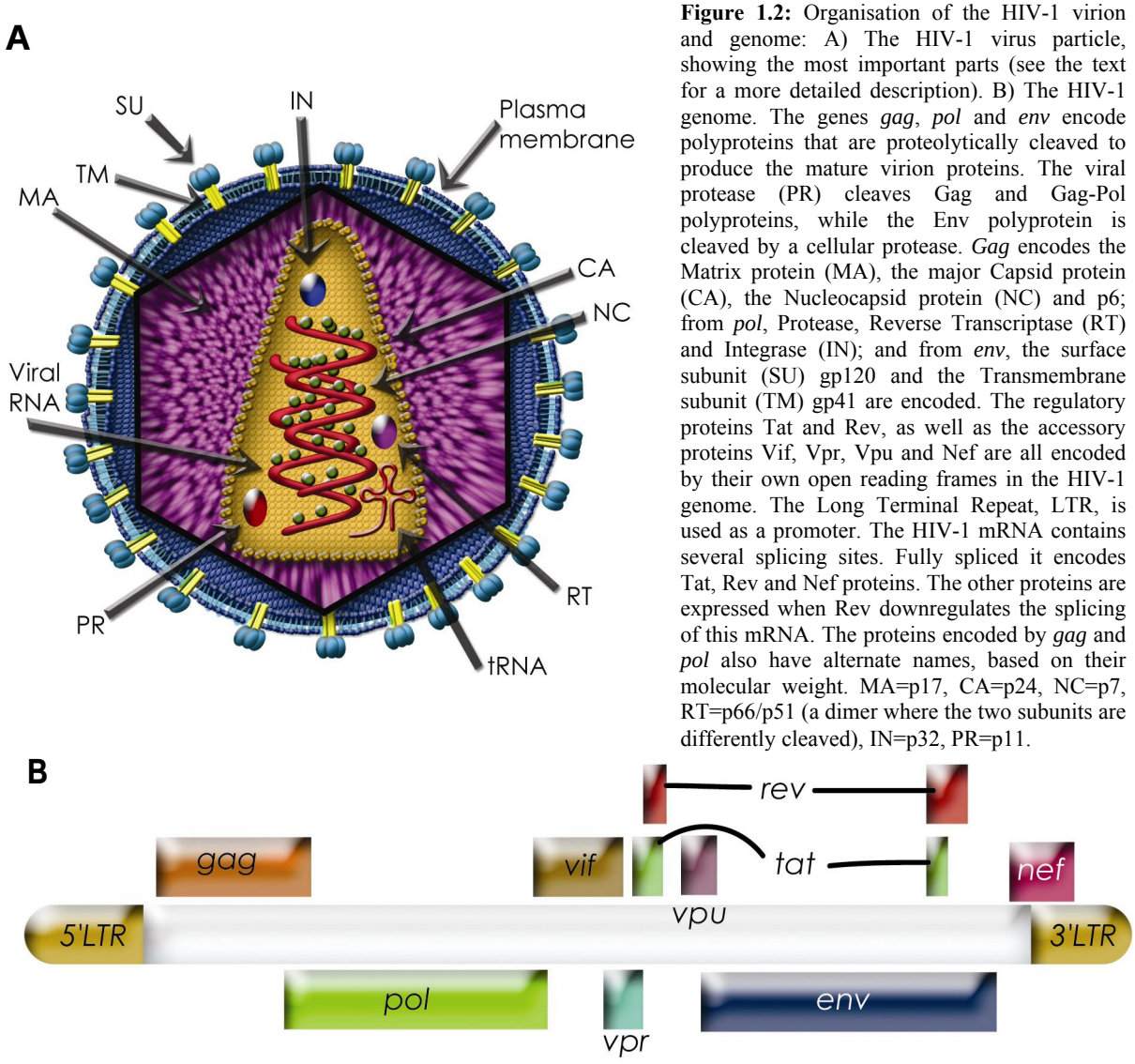


Figure 1.2: Organisation of the HIV-1 virion and genome: A) The HIV-1 virus particle, showing the most important parts (see the text for a more detailed description). B) The HIV-1 genome. The genes *gag*, *pol* and *env* encode polyproteins that are proteolytically cleaved to produce the mature virion proteins. The viral protease (PR) cleaves Gag and Gag-Pol polyproteins, while the Env polyprotein is cleaved by a cellular protease. *Gag* encodes the Matrix protein (MA), the major Capsid protein (CA), the Nucleocapsid protein (NC) and p6; from *pol*, Protease, Reverse Transcriptase (RT) and Integrase (IN); and from *env*, the surface subunit (SU) gp120 and the Transmembrane subunit (TM) gp41 are encoded. The regulatory proteins Tat and Rev, as well as the accessory proteins Vif, Vpr, Vpu and Nef are all encoded by their own open reading frames in the HIV-1 genome. The Long Terminal Repeat, LTR, is used as a promoter. The HIV-1 mRNA contains several splicing sites. Fully spliced it encodes Tat, Rev and Nef proteins. The other proteins are expressed when Rev downregulates the splicing of this mRNA. The proteins encoded by *gag* and *pol* also have alternate names, based on their molecular weight. MA=p17, CA=p24, NC=p7, RT=p66/p51 (a dimer where the two subunits are differently cleaved), IN=p32, PR=p11.

contains the viral proteins Nef, Vif, Vpr and p6, as well as several cellular factors (Frankel and Young 1998; Ott *et al.*, 2000). Among these is the tRNA ($tRNA^{Lys}$) that is used as primer for initiation of reverse transcription.

Depicted in figure 1.2.B is the organisation of the HIV-1 genome, an approximately 9200 bp long single stranded RNA. It contains the *gag*, *pol* and *env* genes. This is typical for all retroviruses. These genes encode polyproteins, which are cleaved by proteolysis into individual proteins. The Gag polyprotein is cleaved into the proteins MA, CA and NC that make up the core of the virion, as well as p6 found within the virion. *Env* encodes the membrane proteins SU and TM. *Pol* encodes the enzymes PR, RT and IN. The HIV genome further contains the genes for Tat and Rev, proteins that regulate HIV gene expression, as well as genes for four accessory proteins, Nef, Vif, Vpr and Vpu. As seen in the figure, these additional proteins are encoded by separate and overlapping ORFs (reviewed by Frankel and Young, 1998).

1.3 The life cycle of HIV

The life cycle of HIV is outlined in figure 1.3. HIV-1 infects a cell as the plasma membrane of the virus particle fuses with the cellular plasma membrane. This is a result of the interaction between SU in the viral membrane and the cell-surface receptor CD4. The fusion of these membranes also requires a coreceptor. The coreceptors commonly used by HIV-1 are CXCR4, which is located on the surface of CD4 T Helper cells; or CCR5, which is located on the surface of macrophages and a subset of the CD4 T Helper cells. Which coreceptor a virus can use determines its tropism. The fusion of viral and cellular membranes leads to the release of the viral core particle into the cytoplasm. The particle develops into a looser structure. Within this structure the RNA genome is reverse transcribed. This renders a linear double stranded DNA molecule. The transformed core particle, a DNA protein complex called the preintegration complex (PIC), moves from the cytoplasm into the nucleus. There IN integrates the viral DNA into a host chromosome, thus making the viral genome a stable genetic element of the infected cell, a provirus.

The first full-length viral mRNAs to be produced will for the most part be doubly spliced. These mRNAs encode the Tat, Rev and Nef proteins. Tat and Rev function in feedback loops and will travel to the nucleus. Tat will increase the production of functional HIV mRNAs through a mechanism discussed later. Rev transports unspliced and singly spliced viral

mRNAs to the cytoplasm where the mRNAs are translated. This leads to the expression of the other HIV proteins.

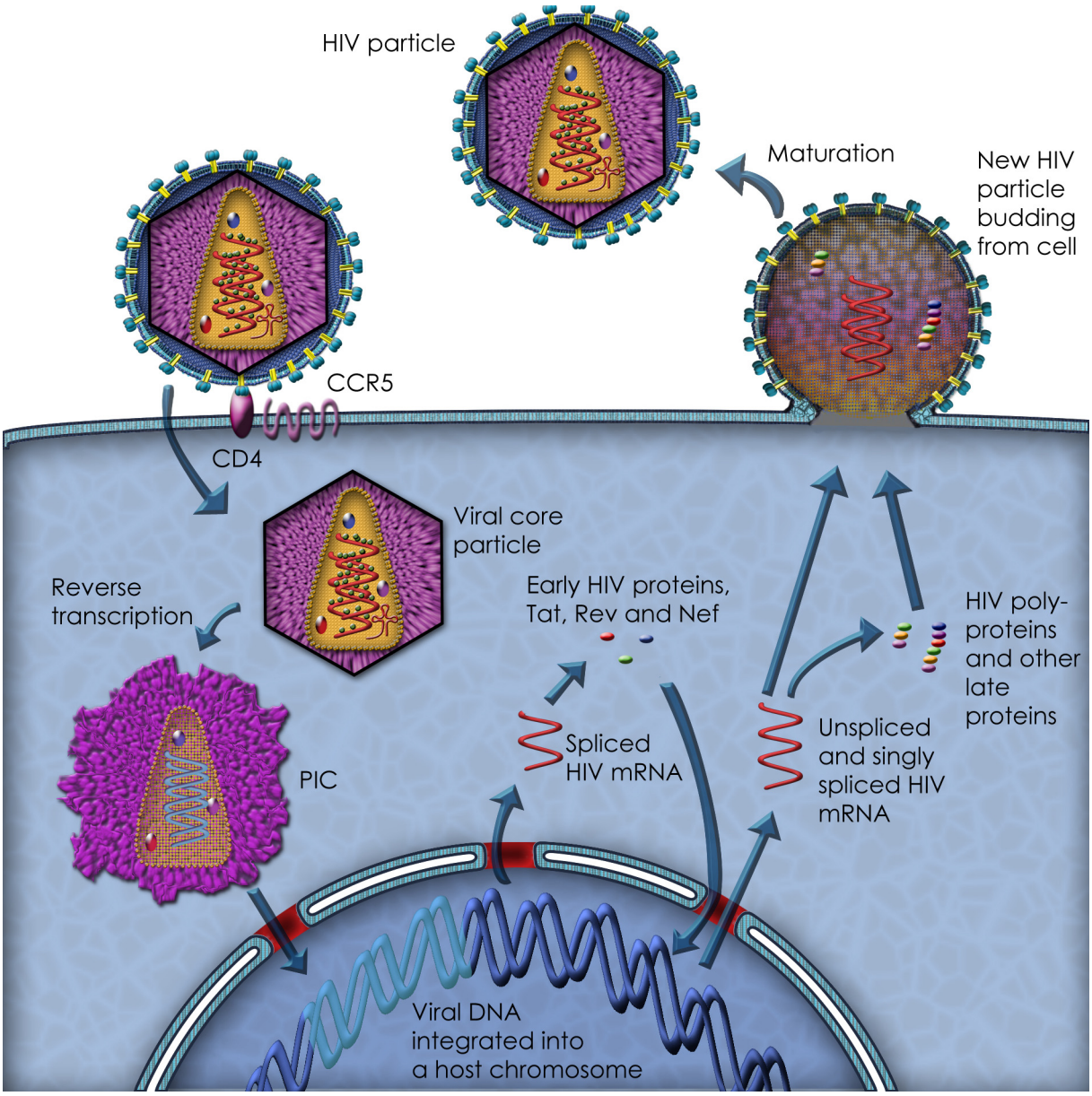


Fig 1.3: The HIV life cycle. The details are outlined in the text.

Full-length, unspliced mRNA transcribed from the provirus is packaged into the new virions and used as genomic RNA. It is also used as messenger RNA for synthesis of both Gag and Gag-Pol polyproteins. The latter is made when the ribosome shifts reading frames while translating this mRNA. This happens approximately every 20th round of translation. The Gag and Gag-Pol polyproteins will associate at the plasma membrane and initiate the assembly of the core particle. The plasma membrane where they assemble is enriched with the SU and TM proteins. These two proteins are formed as gp160 polyprotein and cleaved by a cellular protease in the Golgi. The core particle becomes encapsulated within plasma membrane as it

is budded off from the cell. Maturation takes place during or immediately after the budding of this particle. A functional core particle is formed in this process when the viral protease cleaves the Gag and Gag-Pol polyproteins (reviewed by Frankel and Young, 1998).

1.3.1 Regulation of transcriptional activity

The Tat protein has several functions. Most importantly, it regulates the extent to which HIV mRNA is elongated during transcription. As illustrated in figure 1.4, in the absence of Tat most transcripts are short RNAs less than 100 bases in length. These are degraded by cellular mechanisms. A few transcripts are full-length. This means that the production of HIV proteins is very low, and the production of full-length HIV mRNA is at a minimum. The first full-length transcripts that are produced become fully spliced and encode the HIV early proteins. Among these is the Tat protein. The presence of Tat protein leads to the production of more full-length transcripts; a 200-300-fold increase of these when Tat is present has been reported (Jeang *et al.*, 1999). This indicates that Tat functions in a positive feedback loop. It upregulates its own expression along with that of the other HIV proteins. Thus it accelerates the viral life cycle and the virus production.

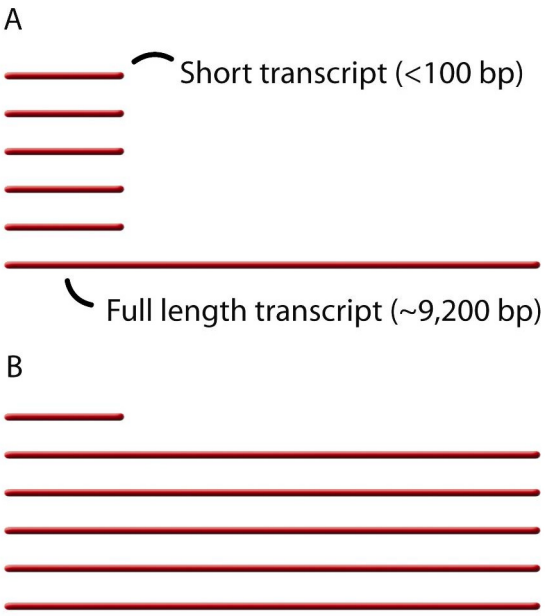


Figure 1.4: The net effect of Tat in the regulation of HIV mRNA. A: In the absence of Tat most RNA transcripts of the HIV genome are less than 100 basepairs, and will quickly be degraded. A few transcripts are full-length. B: When The Tat protein is present the situation changes. The amount of full-length transcripts can be increased some hundredfold.

Studies of the mechanism behind this regulation have revealed previously unknown mechanisms of regulation of gene expression.

Transcription as performed by RNA Polymerase II (RNAP II) can be divided into three parts: initiation of transcription, elongation of the RNA molecule, and termination. When the HIV genome is transcribed the RNA Polymerase II will usually stop synthesizing the mRNA after approximately the first 59 nucleotides. At this point the initiation complex will be transformed into an elongation complex, which includes different factors. Among the factors that associate with the elongation complex are Negative Transcription Elongation Factors (NTEFs). These are

thought to be responsible for the halt in transcription by dephosphorylating the C-Terminal Domain (CTD) of the major subunit of RNA Polymerase II. Positive Transcription Elongation Factors (PTEFs), are required to overcome this halt in transcription and Tat functions by recruiting one such PTEF to the elongation complex (Price, 2000).

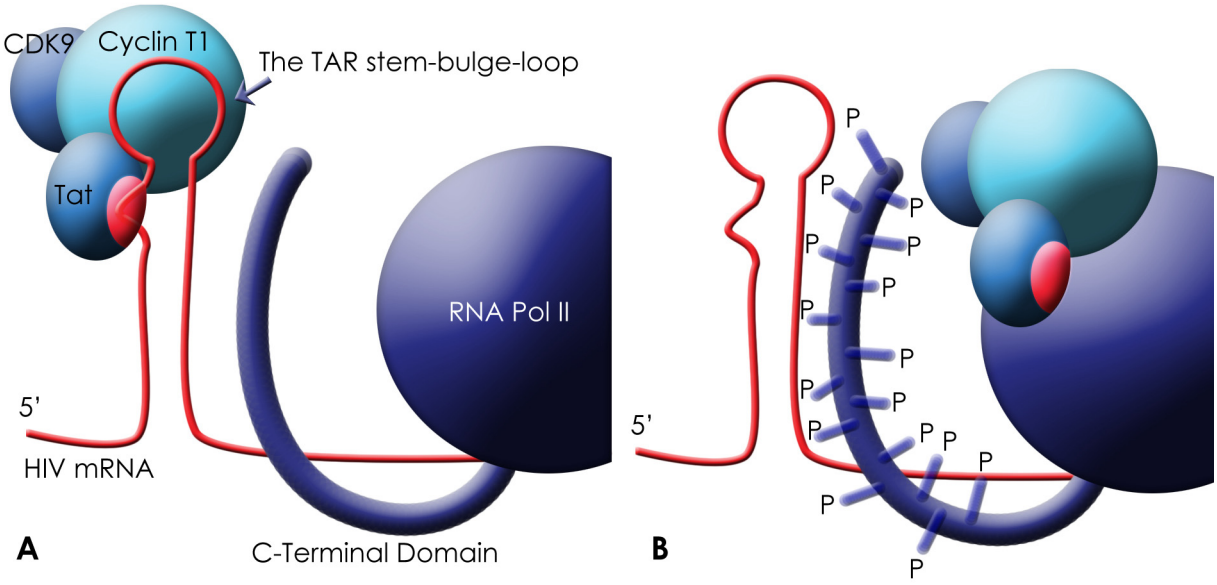


Figure 1.5: A representation of how Tat helps to overcome the block of the transcription of the HIV genome. A) The RNA polymerase II complex has transcribed the stem-bulge-loop TAR, which lies just downstream of the promoter. Tat has bound to TAR and to Cyclin T1, thus starting the formation of a complex with the PTEFb. PTEFb includes a kinase, cdk9 (which binds to Cyclin T1). B) When the Tat-PTEFb complex becomes part of the RNAP II elongating complex, TAR is released and cdk9 will phosphorylate the C-terminal domain of the major subunit of RNAPII. This makes the transcription complex able to continue transcribing the HIV-1 genome.

If the RNA that has been transcribed up to this point is at least 44 nucleotides long it will form the stem-loop structure known as TAR, the transactivation response element. When present, Tat will bind to a small bulge in the TAR structure (figure 1.5.A). The protein Cyclin T1 will bind to Tat and the loop of TAR. Cyclin T1 is a part of the Positive Transcription Elongation Factor b complex, PTEFb. Another component of PTEFb is Cyclin Dependent Kinase 9 (CDK9), which binds directly to Cyclin T1. The Tat-TAR-PTEFb complex will associate with the stalled RNAP II complex. Subsequently TAR will be released from the complex (Keen *et al.*, 1997). CDK9 will then phosphorylate the CTD (figure 1.5.B) as well as other factors. These other factors include at least one subunit of an NTEF (Ping and Rana, 2001; reviewed by Kobor and Greenblatt, 2002). This phosphorylation seems to be required to overcome the halt in transcription, and this allows the RNAP II complex to proceed to make a complete transcript.

In addition to this it has been demonstrated that Tat interacts with several other cellular factors involved in transcription (Kobor and Greenblatt, 2002). Tat has also been shown to associate with Histone Acetyl Transferases, HATs (Benkirane *et al.,* 1998; Kiernan *et al.,* 1999). When DNA is to be transcribed these participate in unravelling the nucleosome structures in which the DNA is packed.

1.3.2 The transcellular activity of Tat

The Tat protein accumulates in the nucleus and the nucleoli of the cell. It is a nucleocytoplasmic shuttle protein, moving back and forth between the nucleus and the cytoplasm (Stauber and Pavlakis, 1998). It is secreted from cells (Ensoli et al., 1993) and can be detected in sera from HIV positive patients (Westendorp *et al.*, 1995). It can also enter cells from the outside (Helland *et al.* 1991; Mann and Frankel 1991; Ensoli *et al.* 1993; Verhoef *et al.* 1996; Valvatne *et al.,* 1996; Staalesen, 1999). It is not known how Tat can be transported from the nucleus to the cytoplasm since a nuclear export signal (NES) has not been identified. Nor is it known by which mechanism Tat is secreted from cells. However, it is known that the basic domain of Tat mediates import into the cell as well as into the nucleus (Dang and Lee, 1989). This is the same domain as the one with which Tat binds to TAR. Tat enters cells by binding to Heparan Sulfate Proteoglycans (HSPG) on the cell surface (Tyagi *et al.*, 2001). These are part of the extracellular matrix and nearly all types of cells have them, hence Tat can enter almost any kind of cell.

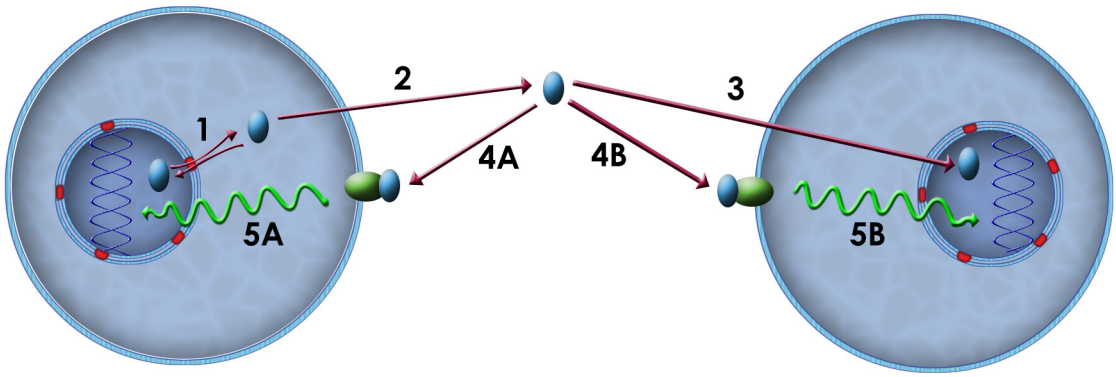


Figure 1.6: Trafficking of the Tat protein. 1) Tat is a shuttle protein moving back and forth between the nucleus and the cytoplasm. 2) Tat may be secreted from cells and 3) be taken up by another cell, or alternately 4) interact with cell-surface receptors of either A) the cell from which it originated in an autocrine loop or B) a nearby cell. 5) Interaction with cell-surface receptors may trigger signal transduction pathways that affect the gene expression of the cell.

Figure 1.6 shows an overview of the trafficking of the Tat protein. Tat's ability to traverse from one cell to another means that Tat may affect gene expression in neighbouring cells, cells that might not be infected with HIV. This can be a result of Tat interacting with

transcription factors inside this other cell, but it can also be the result of Tat's interaction with either of several receptors on the cell surface (Rusnati *et al.*, 2000). This would trigger signal chains that ultimately affect the gene expression of the cell. Through these Tat may be linked to several of the pathogenic (i.e. disease-causing) effects caused by HIV infection.

1.3.3 Implicating Tat in the pathogenic effects of HIV

Tat has been implicated in a multiplicity of functions related to the HIV life cycle and the diseases associated with HIV. The effects of Tat on the immune system can be divided into two parts: Tat as a transcriptional activator of the HIV genome has a direct effect on the speed with which the virus replicates, and thus also how fast the disease progresses; and secondly, Tat affects the expression of cellular genes, directly and indirectly.

Tat can affect cellular gene expression in two ways: intracellularly by direct contact with components of the transcriptional machinery, or extracellularly through receptors on the surface of the cell. The latter will trigger signalling pathways that will lead to a change in cell function/cellular gene expression. Extracellular Tat functions as a chemoattractant on several different cell types of the immune system (Benelli *et al.* 2000). By moving into the vicinity of a cell productively infected with HIV these attracted cells are more easily infected. Extracellular Tat also induces immune cells to produce several cytokines, which regulate the immune response (Nath *et al.* 1999). This may lead to a dysregulation of the immune response. Tat induces the expression of monocyte chemoattractant protein-1 in astrocytes. If there has been a productive infection of HIV in the brain the monocytes that are indirectly attracted by Tat will also be induced to secrete eicosanoids and inflammatory cytokines that are toxic to the cells of the central nervous system. This induces dementia (Conant *et al.*, 1998). Tat is also an angiogenic factor and induces the production of angiogenic factors in macrophages. This means that the presence of Tat leads to a production of blood vessels. This is important in the development of tumours. For instance, Tat increases the growth rate of cells derived from the characteristic skin cancer Kaposi's Sarcoma which affects an uncommonly high percentage of HIV patients, especially gay men (Blattner, 1999; Louie *et al.*, 2002).

The depletion of $CD4^+$ T cells is the most characteristic trait of HIV infection. When HIV infects these cells it does lead to their eventual death, but the massive loss of these cells cannot be explained by this alone. Depending on which cell surface coreceptor the SU protein can interact with, HIV can infect $CD4^+$ T cells, macrophages or both. Usually when a person

is infected with HIV it will be with a macrophage tropic virus. In less than half of the patients there will be a shift in tropism of the virus so that it can also infect the $CD4^+$ T cells (Weiss, 2000). The loss of these cells in the absence of their infection has been ascribed to apoptosis, programmed cell death (Gougeon and Montagnier, 1993). Extracellular Tat has been reported to induce several effects on uninfected $CD4^+$ T cells: activation, functional unresponsiveness and apoptosis (Westendorp *et al.,* 1995; Ott et al. 1997; Bartz and Emerman, 1999). This would make them more susceptible to infection by HIV, ineffective or dead, respectively.

The phenomenon illustrated in figure 1.4 is also dependent on the levels of Cyclin T1 of PTEFb. An important part of the immune response is that promonocytic cell lines differentiate to macrophages and that T Cells are activated. When this happens the levels of Cyclin T1 increase dramatically (Price, 2000). The expression of HIV in these cells is hampered until this has happened. But as the Cyclin T1 levels rise, the virus production of the cell will increase and normal functions of the cell will be downregulated so that the cell cannot take part in the intended immune response.

The virulence of HIV varies among strains, as does the sequence of the Tat protein. When HIV-1 Tat has been studied mostly variants derived from viruses of group M, subtype B have been used.

1.3.4 An overview of the Tat protein

In most wild type isolates of HIV-1, the Tat protein comprises 101 amino acid residues. The first 72 of these are encoded by *tat* exon-1, and residues 73-101 are encoded *tat* exon-2 (figure 1.2.B). In the first stages of HIV infection the cell will produce Tat protein encoded by both exons, but in the later stages as Rev transports unspliced and singly spliced mRNA to the cytoplasm, the cell will produce one exon Tat protein (Purcell and Martin, 1993). The loss of the second exon in late stages of HIV infection might be to slow down or shut off the attributes provided by the exon-2 encoded sequence.

In some isolates the Tat protein contains additional residues. In these the second exon is expanded. In a strain from group O, the MVP5180 isolate, comprises 115 amino acid residues. An 86 amino acid form exists for some laboratory-passaged strains, like LAI and HXB2, from HIV-1 group M. This variant has been used in many studies and a general map of the protein can be made based on this, where the attributes of the protein can be assigned to different parts of the protein (figure 1.6). However, since most wild-type strains have 101

amino acids, it has been suggested that the 86 amino acid variant represents a truncated non-natural form of the protein. Most likely it arose as the virus was passaged in tissue culture. This postulation is corroborated by the fact that a single nucleotide change in HXB2 and LAI at the 87th codon, which in these isolates is a stop codon, results in a 101 amino acid protein (Neuveut and Jeang, 1996; Jeang *et al*, 1999). However, other studies indicate that the 86 amino acid version does exist *in vivo* in viruses from subtypes D and H (Kleivbo, 2001). The preservation of residues 87-101 in wild-type isolates hints at some biological importance, even if they do not contribute significantly to the *ex vivo* propagation of the virus. Furthermore, it has been shown that exon 2 is important in several biological assays (Howcroft et al. 1993; Neuveut and Jeang, 1996; Ott et al. 1997; Verhoef et al. 1998; Jeang *et al*, 1999; Col *et al.*, 2002).

A multiplicity of functions has been ascribed to the Tat protein (Rubartelli *et al.*, 1998). The HIV genome is only 9 kbp. This restricts the number of genes and their size, and a possible consequence of this is that each HIV-1 protein might have evolved to have multiple functions. The different functions and tasks of the Tat protein can be assigned to different domains (figure 1.6). These include the proline rich domain, the cystein rich domain, the core domain, the basic domain (which is also known as the RNA binding domain) and the glutamine rich domain. Also marked are two lysines, 28 and 50. These have been shown to be acetylated by HATs, and this modulates Tat's transactivating properties (Benkirane *et. al.,* 1998). Amino acids 78-80 have also been marked. This is the RGD sequence, which is postulated to bind to integrins on the cell surface (Valvatne *et al.,* 1996).

Phosphorylation by PKR, a dsRNA dependent protein kinase, has been suggested to modulate Tat function. Differences in the extent to which Tat can be phosphorylated by this protein kinase might have an effect on the virulence of the HIV strain (Péloponèse et al., 1999). The amino acid residues that are phosphorylated by this kinase have been marked in figure 1.7.

The most studied region of Tat is the basic domain. It consists of amino acids 49-57, the sequence of which is RKKRRQRRR. This is the motif with which Tat binds to TAR (Karn, 1999). It also functions as a nuclear localization signal, NLS, as well as a nucleolar localization signal, NOS. In addition, it seems to be responsible for the uptake of Tat by cells (Jeang *et al.*, 1999). The residues flanking this basic domain have been shown to significantly influence the specificity of the interaction between Tat and TAR (Karn, 1999). This may be

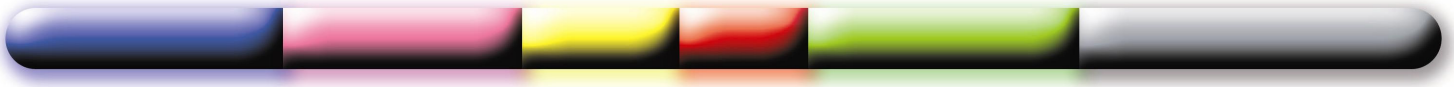


Figure 1.7: An overview of the Tat protein. P, prolines in domain 1, the proline rich domain; C, cysteines in domain 2, the cysteine rich domain; R/K, arginines and lysines of domain 4, the basic domain; Q, glutamines of domain 5, the glutamine rich region; RGD, a motif postulated to confer binding of Tat to the extracellular matrix; S/T, serines and threonines that can be phosphorylated by PKR, a protein kinase dependent on dsRNA for its activation; K, Lysines that are acetylated by Histone Acetyl Transferases. Domain 1, amino acids 1-20, is a proline rich region. Domain 2, amino acids 21-37, is a cysteine rich region; it confers binding to the PTEFb complex via zinc ions. Domain 3, amino acids 38-48, is very conserved in Tat from different strains of HIV-1, and even HIV-2 and SIV. Domain 4, amino acids 48-57, is the basic domain of Tat, also known as the transduction domain. It is this basic domain that binds to TAR; it is also responsible for the nuclear and nucleolar localization of the protein as well as its uptake by cells. Domain 5, amino acids 58-77, is the glutamine rich region. Its properties are not well defined, but domain 3 and 5 enhances the binding specificity between Tat and TAR. Domain 6, amino acids 78-101, is the remainder of the protein and consists of most of what is encoded by exon 2. The part of the protein encoded by exon 2, amino acids 73-101. Its function is not understood. All essential functions of the protein have been located to the part encoded by exon 1. The domains have been highlighted in different colours.

the case both for the glutamine rich region as well as the core domain. The core domain has along with the first two domains been shown to be necessary for transcriptional function of the Tat protein. The cysteine rich domain is thought to bind to the PTEFb complex through metal ion binding (Karn, 1999).

2 Materials

2.1 Chemicals

Chemical name:	Supplier:
2-Mercaptoethanol	Merck
Acetic acid	Sigma
Acrylamide	Merck
Bacto-Agar	Difco
Agarose NA	Pharmacia Biotech.
Ammonium persulphate	Biorad
Ampicillin	Sigma
Aprotinin	Sigma
Bis-acrylamide	Biorad
Bovine Serum Albumine (BSA)	Sigma
Bromphenol blue	IBI
Chloramphenicol	Sigma
Chloroform	Merck
Chrome sulphuric acid	Merck
Coomassie Brilliant Blue R	Sigma
D(+)-Glucose-Monohydrate	Merck
Diethyl ether	Merck
Dimethyl sulfoxide	Merck
Di-Sodium hydrogen phosphate dihydrate	Merck
Ethanol	Arcus Produkter
Ethylenediaminetetraacetic acid, (EDTA)	Sigma
Ethylphenyl polyethylene glycol, (Nonidet P-40)	United States Biochemical corp.
Ficoll 400	Sigma
Formic acid	Merck
Gelatine	Difco
Glycerol	Merck
Glycine	Merck
Hydrochloric acid	Merck
Isopropanol	Arcus produkter
Isopropyl-1-thio-β-D-galactoside (IPTG)	SERVA
Kanamycin	Sigma
Leupeptin	Sigma
Magnesium chloride hexahydrate	Merck
Magnesium sulphate	Merck
Methanol	Prolabo
N,N,N',N'-Tetramethylendiamin, (TEMED)	Merck
PEG 6000	KEBO Lab
Phenyl methyl sulfonyl fluoride (PMSF)	Sigma
Polyoxyethylene sorbitan monolaurate, (Tween-20)	Sigma
Polyvinylpyrrolidon	Merck
Sodium acetate	Merck
Sodium azide	Riedel-De Häen ag.
Sodium chloride	Merck
Sodium dihydrogen phosphate monohydrate	Merck
Sodium dodecyl sulphate, (SDS)	United States Biochemical corp.

(cont.)

Sodium hydroxide pellets	Merck
Trizma base	United States Biochemical corp.
Tryptone	Difco
Yeast extract	Difco

2.2 Media solutions and chemicals for cell culturing

Name:	Supplier:
Foetal Calf Serum (FCS)	Bio Whittaker, USA
Gentamicin Sulphate (50 mg/ml)	Bio Whittaker, USA
Iscove's Modified Dulbecco's Medium (IMDM)	Bio Whittaker, USA
Trypsin-Versene mixture	Bio Whittaker, USA

2.3 Cell lines and bacteria

COS-7 cells

This cell line stems from the kidney of an African green monkey (Cercopithecus aethiops). An origin defective mutant of SV40 has been used to transform these cells, leading them to produce the large T antigen. This enables the replication of plasmids containing the SV40 ori.

E.coli XL1-blue (*recA1, endA1, gyrA96, thi, hsdR17 (rk[+], mk[+]), supE44, relA1, [λ][-], lac, [F', proAB, lacIqZ[δ]M15, Tn10 (tet[R])])* (Stratagene)

E.coli DH5α (F^- Φ80d*lacZ*ΔM15, Δ(*lacZYA-argF*)U169, *deoR*, *recA*1, *endA*1, *hsdR*17(r_K^-, m_K^+), *phoA*, *supE*44, λ^-, *thi*-1, *gyrA*96, *relA*1) (Stratagene)

E. coli BL21 CodonPlus(DE3)-RIL (F^- *ompT hsdS*(r_B^- m_B^-) *dcm*$^+$ Tet^r *gal* λ(DE3) *endA* The [*argU ileY leuW* Cam^r]) (Stratagene)

These latter bacteria were heat shock competent when bought from the supplier, but were made electroporation competent for use in the experiments in this thesis.

2.4 Plasmids

Plasmid name:	Comments:	Provider:
pEGFP-C1	Cloning vector. Expresses the EGFP protein in mammalian cell lines.	Clontech
pEGFP-Tat72	Expresses a fusion protein in mammalian cell lines: The one-exon encoded Tat protein (from the HIV-1 strain HXB3) fused to EGFP.	Vibeke Andresen, MBI, UiB, Norway
pEGFP-Tat86	Expresses a fusion protein in mammalian cell lines: The two-exon encoded Tat protein (from the HIV-1 strain HXB3) fused to EGFP.	Vibeke Andresen, MBI, UiB, Norway

(Cont.)

pCR-Blunt II-TOPO	Cloning vector, used to clone blunt-ended PCR products.	Invitrogen.
pRSET-C	Bacterial expression vector. Expresses a fusion protein with an N-terminal his-tag.	Invitrogen
LTR-CAT	Reporter plasmid. Expresses the CAT protein when the LTR promoter is activated by Tat protein.	Vibeke Andresen, MBI, UiB, Norway

Plasmid maps and further information is shown in the appendix (section A.3)

2.5 Enzymes

Enzyme:	Supplier :
*Bam*HI	Promega
*Eco*RI	Promega/NEB
*Hin*dIII	Promega
*Mlu*I	Promega
*Xho*I	Promega
Calf Intestinal Alkaline Phosphatase (CIAP)	Promega
T4 DNA Ligase	Promega
RNase A	Boehringer Mannheim
T4 Polynucleotide Kinase	Promega
Pfu Turbo	Stratagene
Taq DNA polymerase	Promega

2.6 Antibodies used in immunofluorescence and Western blot

Name:	Specificity:	Host:	Dilution:	Supplier/reference:
1D9D5	Tat (aa 2-21)*	Mouse	1:10 (WB), undiluted (IF)	Valvatne *et al.*, 1996
2D9E7	Tat (aa 49-86)	Mouse	1:10	Valvatne *et al.*, 1996
4B4C4	Tat (aa 49-86)	Mouse	1:10	Valvatne *et al.*, 1996
1D2F11	Tat (aa 49-86)	Mouse	1:10	Valvatne *et al.*, 1996
5G7D8	Tat (aa 49-86)	Mouse	1:10	Valvatne *et al.*, 1996
Anti-polyhistidine	his-tag	Mouse	1:2000	Sigma
Anti-CAT	CAT	Rabbit	1:2000	Sigma

* The antibody 1D9D5 is reported to have a specificity of amino acids 1-20 of Tat (Valvatne *et al.*, 1996). However, the synthetic peptide Valvatne *et al.* showed that the antibody bound to comprises amino acids 2-21 of Tat.

2.7 Conjugates used in immunofluorescence and Western blot

Label:	Specificity:	Method:	Dilution:	Supplier:
Horseradish peroxidase (POD)	Mouse	WB	1:2000-2500	Amersham
POD	Rabbit	WB	1:2000	Amersham
$Alexa_{568}$	Mouse	IF	1:600	Molecular Probes

2.8 Oligonucleotides

Name:	Sequence(5'►3'):	Comments:
Tat115f	CGC *CTC GAG* CC **ATG** *GAT CC*A GTA GAT CCT	T_M=60° C *Xho*I site, used to amplify *tat115* from the pEGFP-Tat115 vector
Tat115r	CGC *GAATTC* **TTA** TTG TCC TGA GAT CCG TGT AC	T_M=60° C *Eco*RI site, used to amplify *tat115* from the pEGFP-Tat115 vector
Tat101f	CGC *CTC GAG* CC **ATG** GAG CCA GTA GAT C	T_M=60° C *Eco*RI site, used to amplify *tat101* from the pEGFP-Tat101 vector
Tat101r	CGC *GAA TTC* **TTA** ATC GAA TGG ATC TGT CTC TGT	T_M=60° C *Eco*RI site, used to amplify *tat101* from the pEGFP-Tat101 vector
Primer A	GAC GTC TGA TCG G *GAATTC* C **ATG** GAG CCA GTA GAT C	Used to construct the *tat101* gene.
Primer B	CTC CAC CTT CTT CTT GGA TTC CTT CGG GCC TGT CG	Used to construct the *tat101* gene.
Primer C	TCG AAG AAG AAG GTG GAG AGA GAG ACA GAG ACA GATC	Used to construct the *tat101* gene.
Primer D	GGG *GGATCC* ATC GAA TGG ATC TGT CTC TGT CTC TCT	Used to construct the *tat101* gene.
Tatmvp1	CCC *AAG CTT* CC **ATG** GAT CCA GTA GAT CCT GAG ATG CCC CCT TGG CAT CAC CCT GGG AGC AAG CCC CAA ACC CCT TGT	Used to construct the *tat115* gene
Tatmvp2	CCTT CTT TGT GAA ACA AAC ATA GCA ATG ATA GCA GCA TCT TTT GCA ATA GCA ATT ATT ACA AGG GGT TTG GGG CTT	Used to construct the *tat115* gene
Tatmvp3	ATGTT TGT TTC ACA AAG AAG GGT TTG GGA ATC TCC CAT GGC AGG AAG AAG CGA AGA AGA CCA GCA GCT GCT GCA AGC	Used to construct the *tat115* gene
Tatmvp4	TT CCT GCC GGT GTG GGA CAG GGA TTG CTC TGG TAC AGG ATC TTT ATT ATC TGG ATA GCT TGC AGC AGC TGC TG	Used to construct the *tat115* gene
Tatmvp5	TCC CAC ACC GGC AGG AAG CAG AAA CGC CAG GAA GAA CAG GAG AAG AAG GTG GAG AAG GAG ACA GGC CCA AGT GGA C	Used to construct the *tat115* gene

(Cont.)

Tatmvp6	CCC *GGA TCC* TTG TCC TGA GAT CCG TGT ACA ACT GTT GCA AGA ATC CTG GTG GCA AGG CTG TCC A**G*** T TGG GCC TGT C	Used to construct the *tat115* gene
Tatmvpa	CCC *AAG CTT* CC **ATG** GAT CCA GTA G	T_M=56°C, used to amplify the *tat115* gene
Tatmvpb	CCC *GGATCC* TTG TCC TGA GAT CC	T_M=56°C, used to amplify the *tat115* gene

Written in italics are restriction enzyme cutting sites. Written in bold are start and stop codons. Written in bold and marked with an asterisk is one nucleotide in the Tatmvp6 oligonucleotide. This one nucleotide erroneously replaces a C, and makes the sequence of the *tat115* gene differ from the *tat* gene from MVP5180.

2.9 Commercial reagents

Name:	Supplier:
Big Dye™ Terminator (Terminator Ready Reaction Mix)	Perkin Elmer Applied Biosystems, USA
6X Loading dye	Promega
dNTP	Promega

2.10 Molecular weight markers

Name:	Supplier:
Lambda/*Hin*dIII digest	Promega
pGEM Markers	Promega
Prestained Broad Range Standard	NEB
Kaleidoscope markers	Biorad

2.11 Instruments, computers and software

Amersham pharmacia biotech Hyperfilm Autoradiography film
ASSAB CO2 incubator
Bachofer Speed Vac Concentrator
Beckman CPR centrifuge
Beckman GS-15R Centrifuge
Bender & Hobein AG Vortexer Model K-550-GE
Biorad Mini-PROTEAN 3 Acrylamide gel equipment
Biorad Power supply Model 200/2.0
Biorad Mini Trans-blot Cell blotting equipment
Biorad Gene Pulser ® Cuvette Electroporation cuvettes
Biorad Gene Pulser Electroporator
Biorad Pulse Controller
Biorad Gene Pulser II
Biorad Gene Pulser II RF module
Centrifuge A 14 Eppendorf centrifuge Jouan
Clifton heater (Alfa-lab A/S)
Costar minicentrifuge

(Cont.)

EC575 and EC105 Voltagesupplier (E-C Apparatur corporation)
Fedegari autoclave (Max Planck Institut)
Fuji X-ray film developer RGII
GeneAmp PCR System 2400 thermocycler (Perkin Elmer)
Gilson Pipettes: 0.1-2 µl, 2-20 µl, 20-100 µl, 50-200 µl, 200-1000µl
Heigar weight (Precision standard)
Heraeus incubator (Max Planck Institut)
HorizonTM58 Gel electrophoresis system (Bethesda Research Laboratory)
HT Incubator with shaking Infors AG
Kontes Micro Ultrasonic Cell Disrupter
Leica DM Irbe fluorescence microscope
Mettler weight AJ100
Microwave oven
MiliQ-plus (Millipore)
Multiscan MS ELISA-plate reader Labsystems
Owl Separation Systems Inc. Model B1 Agarose gel equipment
Radiometer Copenhagen pH-system
Savant SpeedVac SVC100 Vacuum centrifuge
Shimadzu UV-1201 uv-vis spectrophotometer
Sigma Laboratory centrifuge GmbH model 112
Sorvall RC5B Refrigerated Superspeed Centrifuge
Sorvall RC-53 centrifuge (DuPont Instruments), rotors: SS-34, GSA, HB-4
Techne dri-block DB1 and DB2a (HOUM A/S)
Temaks incubator
Thermomixer compact JOUAN A14 (Eppendorf)
Whatman Chromatography paper 3MM Chp Filterpaper

FPLC System (Pharmacia):

Monitor UV-1
P-500 pumps
Controller
LCC-500
LCC-500 plus
Fraction collector Frac-100
Valve MV-7
Mixer
Peristaltic pump P-1
1 ml HiTrap Chelating column
IBM compatible PC, OS/2
FPLC Director Software

Computers:

Apple MacIntosh Computer
Leica DM IRBE linked to a MacIntosh Computer. The Openlab program was used for image capturing.
IBM Compatible PC, Windows OS

Software:

Microsoft Office
Image Editing programs:
Adobe Photoshop 7.0
Adobe Illustrator 9.0

2.12 Commercial kits

Name:	Supplier:
Qiagen Plasmid Mega kit	Qiagen, USA
Qiaex II Gel extraction Kit	Qiagen, USA
Slow Fade™ Antifade kit	Molecular Probes, USA/Holland
ECL™ Western blotting detection reagents, 1 and 2	Amersham Pharmacia
Concert Rapid Miniprep	Gibco BRL Life Technologies
Concert rapid gel extraction	Gibco BRL Life Technologies
Concert Rapid PCR purification system	Gibco BRL Life Technologies

2.13 Consumables

Name:	Size:	Supplier:
Nunc Easy Flask	25 cm^2 75 cm^2	NUNC Brand Products, Nunc A/S, Denmark
Nunclon™ Multidishes	6 wells 24 wells	NUNC Brand Products, Nunc A/S, Denmark
Transfer pipette	3.5 ml	Sarsted
High performance Chemoluminesence film Hyperfilm™ ECL™		Amersham Life Science, UK
Microscope slides	76 X 26 mm	Merck
Microscopical cover-glasses	12 mm	Assistent
Millipore Express sterile filter	0.22 µm	Millipore
FP30 sterile filter	0.45 µm	Schleicher & Schuell
Nitrocellulose transfer membrane	300 mm X 3 m, 0.45 µm	Schleicher & Schuell

2.14 Buffers and solutions:

Buffer/Solution:	Recipe:
30% Acrylamide/bis	291 g Acrylamide 9 g Bis-acrylamide 1 liter dH_2O
Agarose gel (1%)	1% (w/v) Agarose in TAE-buffer (1X) with 0.5 µg/ml EtBr
10% Ammonium Persulphate (APS)	0.1 g Ammonium Persulphate dH_2O was added to 1 ml
Blocking solution	1 g Ficoll 1 g Polyvinylpyrrolidone 10.3 g Boric acid 1.1 g NaOH 8.8 g NaCl 1 ml Nonidet P-40 0.5 g Gelatin 15 g BSA (Bovine serum albumine) dH_2O was added to 1 l.
10 X blotting buffer	144 g Glycine 30.3 g Trizma base dH_2O was added to 1 liter

(Cont.)

1 X blotting buffer	7 parts of dH_2O 2 parts of Methanol 1 part of 10X blotting buffer
BSA (0.5%) - blocking buffer:	0.5 g Bovine serum albumine 100 ml PBS
Buffer A (for FPLC)	100 mM NaCl 50 mM Na_2HPO_4 pH 8.0 (adjusted with NaOH and HCl) The buffer was sterilised by filtration.
Buffer B (for FPLC)	100 mM NaCl 50 mM NaH_2PO_4 pH 4.0 (adjusted with NaOH and HCl) The buffer was sterilised by filtration.
Coomassie blue staining solution	40% (v/v) EtOH, 10% (v/v) HAc and 0.1% (w/v) Coomassie Brilliant Blue R in dH_2O
De-staining solution:	10% Acetic acid 10% Methanol in dH_2O
Ethidium bromide solution (EtBr)	0.5 mg/ml Ethidium bromide in dH_2O
Frozen Storage Buffer (FSB)	Stock Solution
GTE-buffer	50 mM Glucose 25 mM Tris-Cl, pH=8.0 10 mM ETDA, pH=8.0 Sterilized by autoclaving
GYT buffer	10 ml 10% Glycerol 0.125 g Yeast extract 0.250 g Tryptone dH_2O was added to 100 ml Sterilized by autoclaving
2 x HBS	8.0 g NaCl 6.5 g HEPES (sodium salt) 10 ml 1.0 M Na_2HPO_4 (Merck) The pH was adjusted with NaOH and HCl to 7.0. The volume was adjusted with dH_2O to 500 ml, and the buffer was sterilised by filtration.
Hypotonic buffer (lysis buffer)	10 mM Tris-Hcl pH 7.4 10 mM NaCl 1.5 mM $MgCl_2$ 0.5% NP-40

(Cont.)

LB medium:	5 g NaCl 5 g Yeast extract 10 g Tryptone 1 ml 1 M NaOH dH_2O was added to 1 liter Sterilized by autoclaving Added if required: Antibiotics 15g Bacto agar (for generation of agar plates.)
1x PBS, per liter dH_2O:	8 g NaCl 0.2 g KCl 1.81 g Na_2HPO_4 x $2H_2O$ (or 1.44 g Na_2HPO_4) 0.24 g KH_2PO_4 pH was adjusted to 7.4 with 1 M HCl.
13% PEG 6000	13% PEG 6000 (w/v) in dH_2O Sterilized by autoclaving
4% PFA (Paraformaldehyde)-solution	1 g PFA in 25ml PBS
100 mM PMSF	17.4 mg/ml in Isopropanol
Potassium acetate (3M), pH~5.5:	3 M KAc 1.18 M Formic acid in dH_2O
4 x Sample buffer:	2 ml 0.5 M Tris-HCl pH 6.8 1.6 ml Glyserol 3.2 ml 10% SDS 0.8 ml Mercaptoethanol 0.4 ml 0.1% Bromphenolblue 0.4 ml 0.01% Pyronin B
SOB	10 g Tryptone 2.5 g Yeast extract 0.29 g NaCl (10 mM) 0.093 g KCl (2.5 mM) Sterilized by autoclaving
SOC	0.5 ml 1 M MgCl2 0.5 ml 1 M MgSO4 0.5 ml 2 M Glucose SOB medium till 50 ml
Sucrose buffer	27 mM Na_2HPO_4 150 mM Sucrose Sterilized by autoclaving
Separating gel (20% acrylamide):	6.7 ml 30% (w/v) Acrylamide/0.8% (w/v) Bis-acrylamide in dH_2O 2.5 ml 4xTris-Cl, pH 8.8 0.6 ml dH_2O 100 µl 10% SDS (Sodium Dodecyl Sulphate) Immediately before casting the following was added: 50 µl 10% (w/v) ammonium persulfate (fresh) 10 µl TEMED

(Cont.)

Stacking gel solution, 2 gels	3.05 ml dH_2O 1.25 ml 0.5 M Tris-HCl, pH 6.8 50 µl 10% SDS 0.65 ml 30% Acrylamide/bis Immediately before casting the following was added: 25 µl 10% APS 5 µl TEMED
10 x Running buffer (for SDS-PAGE)	18 g Trizma base 86.4 g Glycin 60 ml 10% SDS
3M Sodiumacetate, pH 4.8	40.81 g Sodium acetate 80 ml H_2O pH was adjusted to 4.8 with glacial acetic acid dH_2O was added to 100 ml
1X TAE:	40 mM Tris-acetate 1 mM EDTA
TB (Terrific Broth)	12 g Tryptone 24 g Yeast extract 4.0 ml Glycerol dH_2O was added to 900 ml Sterilized by autoclaving 100 ml of the following autoclaved buffer was added: 0.17 M KH_2PO_4 (2.38 g/100 ml) 0.72 M K_2HPO_4 (12.4 g/100 ml)
10 x TBS	200 ml 1 M Tris-HCl, pH 7.5 300 ml 5 M NaCl 500 ml dH_2O The medium was autoclaved.
TBS-Tween:	100 ml 10 x TBS 500 µl Tween-20 900 ml dH_2O
TE-buffer (1X):	10 mM Tris-Cl, pH=8.0 1 mM EDTA, pH=8.0
0.5 M Tris-HCl, pH 6.8	60.6 g Trizma base 1 l dH_2O pH was adjusted to 6.8 with 1 M HCl
1.5 M Tris-HCl, pH 8.8	181.7 g Trizma base 1 l dH_2O pH was adjusted to 8.8 with 1 M HCl

3 Methods

3.1 Gene cloning techniques

3.1.1 The PCR reaction

The Polymerase Chain Reaction is a method for amplification of DNA. It is based on cyclical temperature variation. Each cycle starts with the reaction mixture being heated so that the strands of the DNA to be copied, the template, break apart (denaturation). This is followed by a lower temperature where a pair of small oligonucleotides anneals to the template DNA. These oligonucleotides are designed such that one of them, the forward primer, is similar to the 5' end of the sequence, and thus binds to the 3' end of the reverse strand. The other oligonucleotide, the reverse primer, is similar to the reverse strand at the end of the sequence to be amplified. Thus it binds to the 3' end of the sequence on the first strand. The cycle continues with a temperature optimal for a DNA polymerase to elongate the oligonucleotides from the 5' end to the 3' end (elongation). The elongation of these oligonucleotides is then based on the sequence of the template DNA. The oligonucleotides of one type that are elongated in each cycle can be used as template for the other type of oligonucleotide in the subsequent cycles. Thus with each cycle the amount of product/template can theoretically be doubled. In the reaction one uses a thermostable DNA polymerase, able to withstand the temperature needed to denature the template DNA. Most commonly used is the *Taq* polymerase.

The temperatures used in the PCR cycle are rarely altered when it comes to the denaturation temperature (which is commonly 94° C), and the elongation temperature (which is commonly 72° C). However the annealing temperature is usually varied. This is done according to the calculated melting temperature, T_M, of the stretch where the oligonucleotides overlap with the template. A simple formula used to calculate the melting temperature is: $T_M=[4*(G+C)+2*(A+T)]° C$

A is the number of Adenines, T is the number of thymines, C is the number of cytosines and G is the number of guanines in the part of the oligonucleotide that overlaps with the template. The annealing temperature is then given as T_M-5° C.

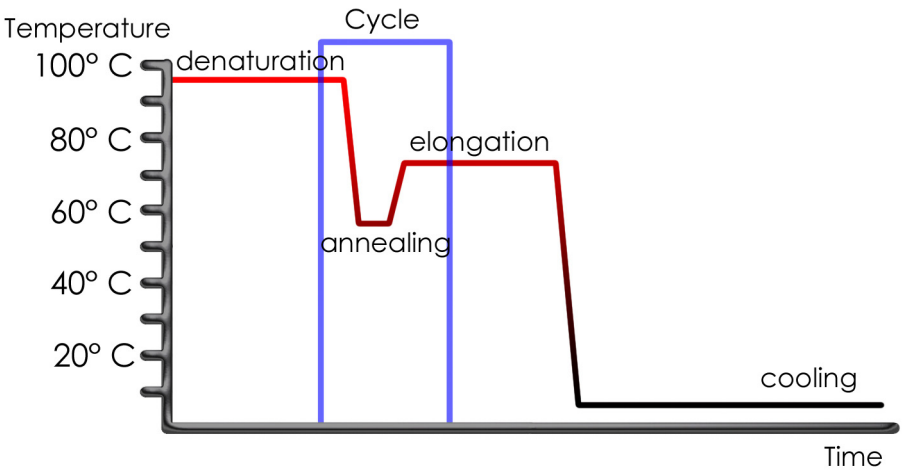


Figure 3.1: Schematic view of a PCR program showing how the temperature in the reaction mixture is varied with time. There are 4 main phases: An initial heating to denature the DNA, then the cycle, then a final elongation step, before cooling at 4° C pending further work. The cycle is usually repeated 25-35 times, and consists of a denaturation step, where the temperature is optimal for separating the DNA strands of the template, an annealing step, where the temperature is optimal for hybridisation between the oligonucleotides and the template DNA, and an elongation step, where the temperature is optimal for the thermostable DNA polymerase to elongate the oligonucleotides.

3.1.2 Amplification of genes by PCR

The *tat101* and *tat115* genes were amplified by PCR from the EGFP-vectors containing them.

The reaction mixture was as follows:

DNA template (50 ng/µl)	10 µl
10X PCR buffer	10 µl
10 mM dNTP	4 µl
Primers (1µg/µl)	2 µl of each
Taq DNA Polymerase (5 U/µl)	1 µl
dH_2O	78 µl
Total volume	100 µl

The DNA template used were the EGFP-Tat plasmids diluted in dH_2O. The EGFP-Tat101 plasmid and the Tat101F and Tat101R primers were used when the *tat101* gene was amplified. The EGFP-Tat115 plasmid and the Tat115F and Tat115R primers were used when the *tat115* gene was amplified. The reaction cycle was as follows:

Initial denaturation step	94° C	7 min
Cycle:		
Denaturation	94° C	30 sec
Annealing	59 (+0.2 per cycle) ° C	30 sec
Elongation	72° C	2 min 30 sec
Repeat cycle 35 times		
Final extension step	72° C	8 min
Cooling	4° C	∞

A similar reaction was also used to remove the 3' T added by the *Taq* polymerase from the amplified *tat115* gene. The polymerase used was then *Pfu Turbo*, the template was the amplified *tat115* gene, and the primers were Tat115f and Tat115r.

3.1.3 Addition of nucleotides to the 3' end of a gene by PCR

In one PCR reaction the original gene is amplified. The reverse primer used in this reaction is made such that a small part is added to the gene. This will ensure that this PCR product is overlapping with a different PCR product that is generated in another reaction. In this other PCR reaction the sequence to be added is made by two oligonucleotides that are overlapping. These two PCR products are then used as template in a third PCR reaction. Seeing how they partially overlap they will in the first cycles generate the new gene. In the later cycles the final product is amplified by the forward primer from the first PCR reaction and the reverse primer from the second PCR reaction. A schematic view of this is shown in figure 3.2. This procedure was used to create the *tat101* gene, by adding nucleotides encoding the last 15 amino acids of the Tat101 protein to the *tat86* gene.

For the two parallel PCR reactions the reaction mixture was as follows:

DNA template (50 ng/µl)	10 µl
10X PCR buffer	5 µl
10 mM dNTP	2 µl
Primers (0.2 µg/µl)	2.5 µl of each
Taq DNA Polymerase (2.5 U/µl)	0.5 µl
dH_2O	22.5 µl
Total volume	50 µl

In one of the reactions, the DNA template was the EGFP-Tat86 plasmid, and the primers used were Primer A and Primer B (section 2.8). In the other reaction the primers used were primer C and Primer D, and no DNA template was used. In this reaction 32.5 µl dH_2O was used. The reaction cycle was as follows:

Initial denaturation step	94° C	5 min
Cycle:		
Denaturation	94° C	90 sec
Annealing	55 ° C	2 min
Elongation	72° C	3 min
Repeat cycle 7 times		
Cooling	4° C	∞

The generation of the *tat101* gene was completed by using the products of these two PCR reactions as template in a new PCR reaction. In this the reaction mixture was as follows:

Template	1 µl of each of the two preceding reactions
10X PCR buffer	5 µl
10 mM dNTP	1 µl
Primers (0.2 µg/µl)	5 µl of each
Taq DNA Polymerase (2.5 U/µl)	0.5 µl
dH_2O	31.5 µl
Total volume	50 µl

The primers used were Primer A and Primer D. The reaction cycle was as follows:

Initial denaturation step	94° C	5 min
Cycle:		
Denaturation	94° C	90 sec
Annealing/ Elongation	72° C	5 min
Repeat cycle 25 times		
Final extension step	72° C	7 min
Cooling	4° C	∞

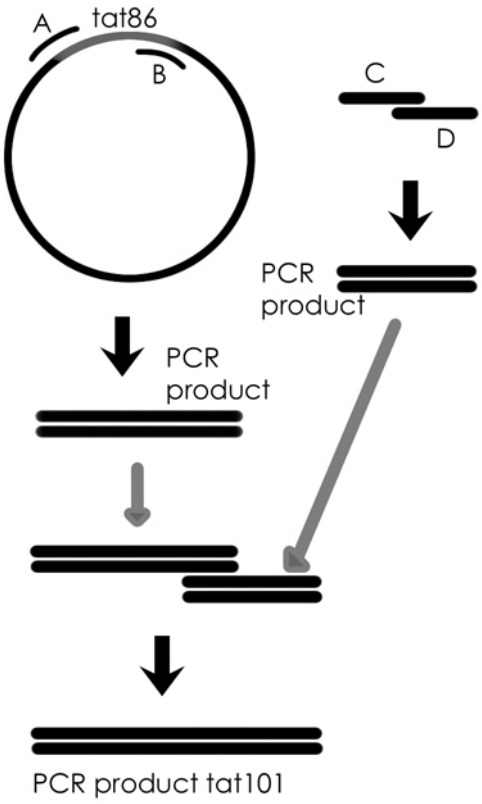


Figure 3.2: A schematic view of how the *tat101* gene was synthesized. A black arrow indicates the product of a PCR reaction, while a grey arrow indicates that this product was used as a template in a new PCR reaction.

3.1.4 Constructing a synthetic gene by PCR

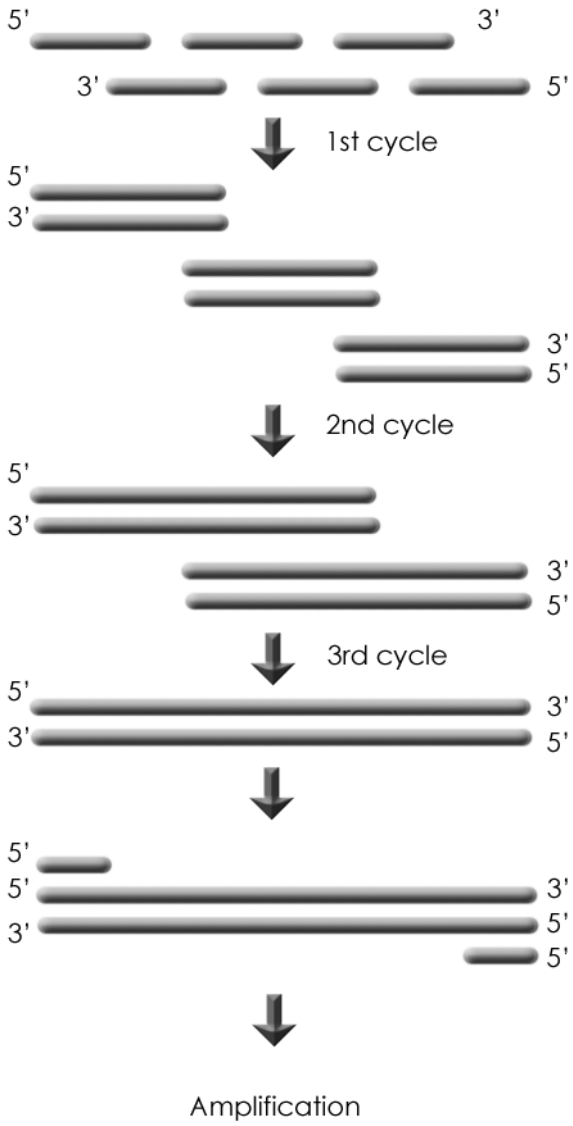


Fig 3.3: The principle of gene construction by PCR. This technique was used to construct the *tat115* gene. Separate large overlapping oligonucleotides are mixed, and this figure shows the intermediate products after the first, second and third cycle in a hypothetical experiment where 6 oligonucleotides together hold the entire genetic sequence. After the final product has been formed (after the third PCR cycle), this is amplified with a set of smaller outer primers.

If the gene to be studied is not available, a synthetic version of the gene can be constructed. This is also a practical option if the gene needs to be mutated so extensively that site-directed mutagenesis would be a time-consuming process. The gene can be made by oligonucleotides that are overlapping. The oligonucleotides represent the gene in its entirety, but only partially each strand. The double stranded gene is then synthesized in a PCR reaction (Dillon and Rosen 1990; Prodromou and Pearl 1992; Sandhu *et al.* 1992; Stemmer *et al.*, 1995). Overlapping nucleotides from each strand form successively larger products with each PCR cycle. In the end they will produce the entire gene, which will then be amplified by flanking primers. The principle of how the gene is gradually constructed with each cycle of the reaction is shown in figure 3.3.

The *tat115* gene was made with this technique. In a first PCR reaction the gene was generated using the primers called Tatmvp1-6. The reaction mixture was as follows:

10X PCR buffer	5 µl
10 mM dNTP	1 µl
Primers (0.2 µg/µl)	1 µl of each
Taq DNA Polymerase (2.5 U/µl)	0.5 µl
ddH_2O	37.5 µl
Total volume	50 µl

And this reaction mixture was used in the following reaction cycle:

Initial denaturation step	94° C	7 min
Cycle:		
Denaturation	94° C	90 sec
Annealing	55 ° C	2 min
Elongation	72° C	3 min
Repeat cycle 7 times		
Final extension step	72° C	3 min
Cooling	4° C	∞

The gene was then amplified in a new PCR reaction, where the product of the previous PCR reaction was used as a template, and the Tatmvpa and Tatmvpb primers were used. The reaction mix was as follows:

Template	1 µl
10X PCR buffer	10 µl
10 mM dNTP	4 µl
Primers (0.2 µg/µl)	5 µl of each
Taq DNA Polymerase (2.5 U/µl)	1 µl
dH_2O	64 µl
Total volume	100 µl

And the PCR program was as follows:

Initial denaturation step	94° C	5 min 30 sec
Cycle:		
Denaturation	94° C	1 min 30 sec
Annealing/Elongation	72° C	5 min
Repeat cycle 30 times		
Final extension step	72° C	2 min
Cooling	4° C	∞

3.1.5 The construction of the EGFP-Tat101 and EGFP-Tat115 plasmids

The flanking primers that were used to construct the *tat* genes (sections 3.1.3 and 3.1.4) had been designed so that the final products of these PCR reactions had restriction enzyme sites at the ends. These restriction enzyme sites had been chosen so that it would be easy to clone the genes into the MCS of the EGFP-C1 plasmid. Additional nucleotides had been added between the restriction enzyme sites and the *tat* genes so that the genes would be in frame with the gene encoding EGFP, and an EGFP fusion protein would be generated. The enzymes that were chosen were *Hin*dIII and *Bam*HI for *tat115* and *Eco*RI and *Bam*HI for *tat101*. The PCR products were treated with these restriction enzymes (section 3.1.9). After this they were purified (section 3.1.10), to remove anything in the reaction that could interfere with the ensuing enzymatic reactions. The plasmid pEGFP-C1, in which these restriction enzyme sites were unique, was also treated with a

phosphatase (section 3.1.12) in order to prevent it from re-ligating later, and purified by gel extraction (section 3.1.11). This would hinder any uncut plasmid from being present in the ligation mixture that would be used to transform bacteria. The genes were then ligated into the plasmid in separate reactions (section 3.1.13).

Bacteria were then transformed with a few µl of these ligation reactions (sections 3.1.17 and 3.1.18). The pEGFP-C1 plasmid and the expected derivatives of it would confer resistance to kanamycin, so this antibiotic was used in the process of selecting for these plasmids in these procedures. Bacteria that survived the selection process in the transformation procedure were then grown in 5 ml LB or TB medium, and analysed by miniprep (section 3.1.19). The purified plasmid from these was then checked to see if it was correct by restriction enzyme analysis. This analysis was performed by cutting the plasmid with the restriction enzymes that had been used when the insert was ligated into the vector. So the assumed EGFP-Tat115 plasmid was cut with *Bam*HI and *Hin*dIII, while the assumed EGFP-Tat101 plasmid was cut with *Eco*RI and *Hin*dIII. If a DNA piece of the expected size were excised when both enzymes were used and not when only one enzyme was used then that would indicate that the correct gene had been inserted into the vector. A few µl of these reactions were then run on agarose gels to visualize the results.

Purified plasmid in which the appropriate gene seemed to have been inserted was used in DNA sequencing to verify that the insert had the correct sequence (section 3.1.22). A culture from which a such plasmid had come was then used in the Qiagen Megaprep protocol to purify large amounts of this plasmid (section 3.1.20). The identities of the plasmids that were purified with this procedure were then established by restriction enzyme analysis.

3.1.6 The construction of the pRSET-Tat101 plasmid

In order to clone the *tat101* and *tat115* sequences into the bacterial expression vector pRSET-C, these genes were amplified by PCR as described in section 3.1.2. The constructed EGFP-Tat plasmids were used as template, and the primers used were Tat101f and Tat101r, for the amplification of *tat101*; and Tat115f and Tat115r for the amplification of *tat115* (section 2.8).

When it came to the construction of the pRSET-Tat101 plasmid, the procedure was very much like the one described in section 3.1.5. The PCR product containing the *tat101* gene was purified (section 3.1.10), treated with the appropriate restriction enzymes (section 3.1.9), purified again and ligated (section 3.1.13) into the pRSET-C plasmid, that had been digested

with the same restriction enzymes and treated with phosphatase (section 3.1.12). Then this ligation mixture was used to transform bacteria (sections 3.1.17 and 3.1.18), plasmid was purified from these bacteria (3.1.19), and the plasmid was subjected to restriction enzyme analysis. Plasmids in which the correct bands were excised were sequenced (section 3.1.22), and the plasmid containing the correct product was purified by megaprep (3.1.20). Afterwards the identity of this plasmid was verified by restriction enzyme analysis.

3.1.7 The construction of the pRSET-Tat115 plasmid

The PCR product containing the *tat115* gene was subcloned into a TOPO vector, pCR-BLUNT II-TOPO. This vector requires the insert to be blunt ended. The PCR product with the *tat115* gene was originally amplified with *Taq* DNA polymerase, which leaves a 3' T overhang. To get a blunt PCR product the original PCR product was amplified in a new PCR reaction using the *Pfu* Turbo DNA polymerase (section 3.1.2). The PCR product from this reaction was ligated into the TOPO vector (section 3.1.14), and the gene was cut out of the plasmid using the *Xho*I and *Eco*RI enzymes (section 3.1.9), and purified by gel extraction (section 3.1.11). However, the vector also contained these restriction enzyme sites, and not too far from the sites that were placed outside the gene by the primers (although on opposite sides). So when the gene was purified from the vector by gel extraction, some not completely digested fragments could contaminate the fragment that was to be ligated into the pRSET-C vector. This was checked for in the screening of colonies by looking for fragments that were cut out of the vector with only one of the restriction enzymes. A colony with the correct insert was then used to purify the plasmid with an alkaline lysis midiprep procedure (section 3.1.19). The gene was then cut out of this plasmid with *Eco*RI and *Xho*I enzymes, and purified by gel extraction. The purified gene was then ligated into the pRSET-C vector, in a similar manner as when this was done with the *tat101* gene (section 3.1.6). Figure 4.9 shows the restriction enzyme analysis of the plasmid preps generated in this procedure. When DNA sequencing of the minipreps affirmed the identity of the plasmid, an alkaline lysis midiprep procedure was used to purify the plasmid.

3.1.8 Agarose gel electrophoresis

The PCR products, as well as plasmids digested with restriction enzymes, were analysed by agarose gel electrophoresis. The gels contained 1-2% agarose, 1xTAE-buffer, as well as 0.5 µg/ml EtBr, for the later visualization of the DNA. The gels were run in 1x TAE-buffer. The gels were typically 10 cm long, and electrophoresis was performed for 30 minutes to 2 hours

at 100-200 V. The time varied with the resolution required. Visualization was done by UV-light.

3.1.9 Restriction enzyme digestion of DNA

Restriction enzymes recognize specific sequences of DNA and cleave the DNA at this vicinity. Enzymes and buffers provided by Promega and NEB were used. The volume used for a restriction enzyme digest was usually 20 µl. The DNA was diluted in dH_2O, the buffer that was supplied with the restriction enzyme was added, and typically 5-10 U restriction enzyme was added. The mix was incubated at 37° C for one hour, or over night. To avoid the restriction enzymes interfering in the reactions that followed, the DNA was purified. The purification of DNA was then done either by gel extraction or by the PCR purification kit.

An example of a restriction enzyme digest with two enzymes:

DNA (5µg/µl)	1 µl
Enzyme 1 (10 U/µl)	1 µl
Enzyme 2 (10 U/µl)	1 µl
10X multicore buffer	2 µl
dH_2O	15 µl
Total volume	20 µl

3.1.10 Purification of PCR products and removal of restriction enzymes

The Concert Rapid PCR Purification kit was used in accordance with the manufacturer's protocol to purify PCR products. This kit was also used to purify these PCR products after they had been digested with restriction enzymes.

3.1.11 Purification of DNA by extraction from gel

The Concert Rapid Gel Extraction kit was used to purify plasmid DNA. This was done when the DNA had been treated with restriction enzymes and a phosphatase. The manufacturer's protocol was followed.

3.1.12 Phosphatase treatment of digested DNA

A phosphatase is used to remove the 5' phosphate group of a DNA fragment. In this study this was routinely done with plasmids that had been digested with restriction enzymes to prevent re-ligation of the plasmid in subsequent ligation reactions. A typical reaction mixture was as follows:

Digested plasmid solution	20 µl
Phosphatase (CIAP 1U/µl)	1 µl
10X phosphatase buffer	3µl
dH_2O	6 µl
Total volume	30 µl

The mixture was incubated at 37° C for 30 minutes, then the plasmid was purified by gel extraction.

3.1.13 Ligation of DNA fragments

In a ligation reaction a DNA ligase chemically couples two DNA ends, the 5' and 3' ends. It can ligate so-called blunt ends or staggered ends. In this study this was used to ligate inserts into plasmids. The inserts were most often PCR products that had been digested with restriction enzymes. The plasmids had been treated with restriction enzymes that would render the same staggered ends as the ones used to digest the PCR products. The insert and plasmid DNA was added in amounts corresponding to a ratio of 3:1, 1:1, or 1:3. This was estimated by running both fragments on agarose gels for comparison of the relative molar amounts. Ligation buffer, ligase enzyme and dH_2O were added to a total volume of 10-30 µl, and the ligation mix was incubated over night at room temperature. An example of a ligation reaction mix, where the plasmid contains about 15 times as many base pairs as the insert:

Plasmid (50 ng/µl)	2 µl
Insert (2 ng/µl)	3.5 µl
10X Ligase buffer	1 µl
T4 DNA ligase (1 U/µl)	0.3 µl
dH_2O	3.2 µl
Total volume	10 µl

3.1.14 TOPO™ Cloning

The Zero Blunt™ TOPO™ PCR Cloning kit from Invitrogen is used to clone blunt-end PCR products. The pCR-Blunt II-TOPO vector is supplied linearized with vaccinia virus topoisomerase I covalently bound to the 3' end. This means that ligation with a blunt-end PCR product will occur spontaneously. A PCR product containing the *tat115* gene was cloned this way. The manufacturer's protocol was followed.

3.1.15 Electrocompetent cells

The bacteria used in this procedure were *E. coli* BL21 CodonPlus(DE3)-RIL from Stratagene. These bacteria were originally heat shock competent. The bacteria were streaked out on a 2x YT-G agar plate containing 25 µg/ml chloramphenicol (this antibiotic and this concentration of it is used in the media throughout the procedure), and were incubated at 37°C O/N. One colony was then picked to inoculate 5 ml 2xYT-G media, which was then incubated at 37°C O/N. 2.5 ml of this culture was then transferred to 1 litre of 2x YT-G media, and this was grown at 37°C until the OD_{600} reached ~0.6, at which point the culture was cooled on ice for 30 minutes. The bacteria were then harvested in 4 GSA tubes by centrifugation at 4000 g for 5

minutes at 4°C. From this point on the bacteria were kept on ice at all times. The supernatant was discarded, and each pellet was resuspended in 100 ml of ice-cold 10% glycerol. The resuspended bacteria were divided between two GSA tubes, and the centrifugation was repeated. Each of the two pellets was then resuspended in 100 ml ice-cold 10% glycerol, and the centrifugation was repeated. Each bacteria pellet was then resuspended in 2 ml GYT media, and 50 µl aliquots were put in eppendorf tubes. The aliquots were then flash-frozen on dry ice in ethanol, before they were stored at -80°C until they were to be used.

3.1.16 Heat shock competent cells

One colony of *E.coli* DH5α Bacteria was grown over night (o/n) at 37°C in SOB-medium with 20 mM $MgSO_4$. From this culture 2 ml was added to 100 ml of the same medium, and this was incubated at 37°C until the A_{600} was ~ 0.4 - 0.6. The bacteria were transferred to four pre-chilled 50 ml centrifuge-tubes and incubated on ice for 10 min. They were then harvested by centrifugation at 1500 g for 10 min at 4°C. The pellets were then resuspended in 15 ml ice-cold FSB-buffer and incubated on ice for 10 min before another centrifugation at 1500 g for 10 minutes. Each pellet was resuspended in 2 ml FSB and 70 µl DMSO was added. This mix was then incubated on ice for 15 min before 100 µl aliquots of it was added to eppendorf tubes. These tubes were then flash-frozen on dry ice in ethanol.

3.1.17 Transformation by electroporation

Plasmid DNA (from a few ng up to 1 µg) was mixed with 40 µl of electro-competent cells. This mixture was then transferred to a pre-chilled electroporation cuvette, which was then placed in an electroporator (Biorad Gene Pulsar). A pulse of 2.5 kV was applied, 1 ml of SOC-medium was added to the mixture which was then transferred to a sterile eppendorf-tube and incubated for approximately 45 minutes at 37°C with moderate shaking. Subsequently 50 µl of the culture was plated out on a LB-plate containing the appropriate antibiotics for selection of bacteria containing the plasmid. The rest of the bacteria were harvested by 1 minute of centrifugation at 13000 rpm in a Sigma laboratory centrifuge (model 112), the supernatant was poured of, and the pellet was re-dissolved in 100 µl SOC-medium and this was plated out on another LB-plate containing the same antibiotics. The plates were incubated at 37°C over night.

3.1.18 Transformation by heat shock

1-100 ng plasmid was added to and mixed with 50 µg heat-shock competent bacteria. This was then incubated on ice for 30 min. The mixture was then incubated at 42°C for 60-90

seconds and put directly on ice for 2 min. 1 ml SOC-medium was added and the bacterial suspension was incubated on a shaker at 37°C for approximately 45 min. The suspension was transferred to LB-plates containing the proper antibiotics, just as in the procedure for electroporation.

3.1.19 Small-scale preparations of plasmid DNA

The alkaline lysis/PEG method

This procedure was used for detection and characterization of plasmids. All centrifugations in this procedure, except for the vacuum centrifugation, were done at 13000 rpm in a table centrifuge. One single bacterial colony was scraped from an LB-plate and inoculated into 5-10 ml of TB-medium containing the appropriate antibiotic, and grown to saturation over night at 37°C with moderate shaking. 4.5 ml (1.5 ml of the culture three times) of the culture was spun down for 1 minute in a Sigma laboratory centrifuge and the supernatant removed. The pellet was resuspended in 200 µl GTE-solution by pipetting up and down. 300 µl of 0.2 M NaOH/1% SDS solution was added, the solution was mixed by inverting the tube a few times, and the tube was placed on ice for 5 minutes. Subsequently, 300 µl of KAc-solution was added. The mixture was vortexed and placed on ice for 5 minutes, before centrifugation for 10 minutes in a Sigma laboratory centrifuge to pellet cell debris and chromosomal DNA. The supernatant was transferred to a clean eppendorf tube, RNase A was added to a final concentration of 20 µg/ml, and this was incubated for 20 minutes at 37°C. The supernatant was extracted with chloroform, 400 µl chloroform was added to the tube and this was vortexed for 30 seconds before centrifugation for 1 minute, the top aqueous phase was then transferred to a clean tube. The chloroform extraction was repeated until there was no debris visible between the organic and aqueous phases (In other words this chloroform extraction was done at least twice, some times thrice). The DNA was then precipitated by the addition of an equal volume of 100% isopropanol, this was swiftly mixed before it was centrifuged for 10 minutes at room temperature. The supernatant was removed, the pellet was washed with 500 µl 70 % ethanol, the tube was centrifuged for two minutes, the ethanol was decanted, and subsequently the pellet was dried in a vacuum centrifuge for 10 minutes. The pellet was then dissolved in 32 µl of deionized H_2O. To precipitate the DNA, 8 µl of 4 M NaCl was added, followed by the addition of 40 µl of autoclaved 13% PEG 8000 to precipitate the DNA. This was mixed thoroughly before the sample was incubated on ice for 20 minutes. The DNA was pelleted by centrifugation for 15 minutes at 4°C. The supernatant was carefully decanted and the pellet was washed with 500 µl 70% ethanol (as above), before it was dried in a vacuum

centrifuge (as above) and resuspended in 20 µl TE-buffer. (http://www.biotech.iastate.edu/Facilities/DSSF/plasmid_prep.html). By scaling up the amounts of the various buffers and solutions, this proecudre was also used as a midiprep procedure.

Concert Rapid Miniprep

As an alternate way to prepare small amounts of plasmid DNA, the Concert Rapid Miniprep system was used. The manufacturers protocol was followed with one exception: the bacteria was occasionally grown in TB-medium in stead of in LB-medium in order to assure a sufficient yield of plasmid.

3.1.20 Large-scale preparations of plasmid DNA (Megaprep)

In order to attain a high yield of very pure plasmid (which would be needed in experiments involving cell cultures) the Qiagen Megaprep kit was used, and the manufacturers protocol was followed.

3.1.21 Determining DNA concentration

Determining DNA concentration was done by spectrophotometry. The A_{260} was used to measure this, as it is 1 when the DNA concentration is 50 µg/ml for dsDNA and 37 µg/ml for single stranded DNA. The purified DNA was diluted in dH_2O for measurements by specrophotometry, and dH_2O was used in the reference cell. The ratio between A_{260} and A_{280} was used to evaluate the purity of the DNA. Optimally this ration should be about 1.8 for pure dsDNA. Higher values indicate presence of RNA; lower values indicate the presence of proteins or phenol.

3.1.22 DNA sequencing reaction

The following reagents were mixed by short vortexing in PCR tubes and spun down in a table centrifuge, before thermocycling:

4 µl Big Dye
500-1000 ng purified DNA template
1.6 pmol primer
dH_2O to a total volume of 10 µl in the reaction.

The following PCR program was used:

Initial denaturation step	94° C	5 min
Cycle:		
Denaturation	96° C	10 sec
Annealing	50 ° C	5 sec
Elongation	72° C	4 min
Repeat cycle 25 times		
Cooling	4° C	∞

The samples were delivered directly to the sequencing facility at this point.

3.2 Computer analyses

The work with identifying the properties of proteins, their structure, and what molecules they can interact with; is facilitated by computer tools and databases. These contain vast amounts of sequences of proteins where the functions, the structures and domains with specific properties have been identified. These computer tools may elucidate properties of the protein you are working with by comparing these sequences to that of your protein. The results gained are only theoretical, but may lead towards the final characterization of the protein more rapidly.

Several programs were used to analyse the sequences of Tat101 and Tat115. The predict protein tool at the PHD server at Columbia (http://dodo.cpmc.columbia.edu); The PSIPRED protein structure prediction server (http://bioinf.cs.ucl.ac.uk/psiform.html), which predicts the secondary structure of a protein based on its primary structure; The ScanProsite program found at the ExPASy server (http://us.expasy.org/prosite); The ProtParam tool (http://us.expasy.org/tools/protparam.html), which calculates theoretical molecular weight, pI, and other features of a user provided sequence; PSORT II (http://psort.nibb.ac.jp/form2.html), a program that predicts the subcellular localization sites of proteins; Blast and "Blast 2 sequences" from NCBI (http://www.ncbi.nlm.nih.gov/), which is a sequence alignment tool.

3.3 Protein purification methods

3.3.1 SDS polyacrylamide gel electrophoresis

In order to analyse samples separate proteins were separated by size by SDS polyacrylamide gel electrophoresis (SDS-PAGE). The results were visualized by Coomassie blue staining or Western blot analysis. The Biorad Mini-PROTEAN 3 Acrylamide gel equipment was used to cast and run minigels (8.3 cm width, 7.3 cm height, 0.5 mm thickness). The gels were cast with a 4% stacking gel and a 20% separation gel, and 10 wells. The samples were boiled at

95° C for 5-15 minutes with after addition of sample buffer to a concentration of 1X. The gels were loaded with 10µl sample. The electrophoresis was always carried out at a constant current (25-45 mA/gel).

3.3.2 Coomassie blue staining of gels

The Coomassie blue colour was routinely used to visualize the proteins on an SDS-PA gel. The detection limit of this method is approximately 100 ng. After electrophoresis the gel sandwich was opened and the stacking gel cut off. The separation gel was covered with the coomassie staining solution and microwaved for 1 minute (until boiling). It was then incubated with shaking for 10 minutes, before the staining solution was removed. After staining the gel was covered in fix solution, and then microwaved for 1 minute. A piece of paper towel was added to the outskirts of the container (so that it did not come in contact with the gel) to absorb the dye. The gel was incubated until satisfactory staining was achieved (2-8 hours).

3.3.3 Expression of the Tat101 protein

In this study the Tat proteins were to be expressed in *E.coli.* The bacteria were transformed with a vector, pRSET-C, in which the *tat* genes were under control of the T7 late promoter. In the bacteria used the T7 polymerase gene is regulated by the *lac* promoter. In the presence of lactose or a lactose analogue, such as isopropylthiogalactoside, IPTG, large amounts of T7 RNA polymerase is produced. Then the polymerase will produce vast amounts of mRNA from the *tat* genes, which will then be used for translation. IPTG was the preferred inducer since it cannot be metabolised and the concentration will remain constant through the experiment.

The antibiotics referred to in this procedure are 34 µg/ml CA (chloramphenicol) and 200 µg/ml amp (ampicillin).

Electrocompetent *E. coli* BL21 CodonPlus (DE3)-RIL were transformed by electroporation with the plasmid pRSETC-Tat101 and were plated on LB-agar containing antibiotics and incubated over night at 37° C. A single colony were inoculated in 5-10 ml LB-medium with antibiotics and grown over night at 37° C. 2-3 ml of this culture was the added to 500 ml of LB-medium with antibiotics and this was grown at 37° C until the A_{600} reached 0.5-0.8, and at this point IPTG was added to a final concentration of 1 mM, and the culture was incubated over night at 30° C (This procedure is modified from Staalesen, 1999).

The bacteria were transferred to GSA tubes and harvested by centrifugation at 4000 g for 20 minutes at 4° C in a sorvall centrifuge, and either the pellet was frozen at -20° C for later use, or it was immediately used for purification of the Tat protein.

3.3.4 Purification of Tat101

Immobilized Metal ion Affinity Chromatography (IMAC), also known as Metal Chelate Affinity Chromatography (MCAC), can be used to quickly and specifically purify a protein or peptide. This is a type of adsorption chromatography. The stationary phase consists of a binding substance (ligand) immobilized on an insoluble support (matrix). In IMAC the ligand is metal ions, for instance Zn^{2+}, Cu^{2+}, Co^{2+}, or Ni^{2+}, which is chelated to the matrix. At a pH around neutral these will form complexes with certain amino acids, for example histidine and cysteine. It is especially histidine that is responsible for the binding to the chelated metal; hence this technique is very effective in the purification of recombinant proteins that have had a poly-histidine tag added. The protein can be eluted by reducing the pH, which will alter the ionic charge of histidine; by including EDTA in the buffer, which will scavenge the metal ions from the matrix; or by adding imidazol to the buffer, which will compete with the protein for binding to the metal ions. (Amersham Pharmacia Biotech Handbook).

Preparation of sample

The pellet of bacteria in which Tat protein had been expressed was resuspended in 10 ml cold Buffer A (section 2.14). To prevent degradation of the protein the following protease inhibitors were added: 2 μg/ml aprotinin, 2 μg/ml leupeptin, and 2 mM PMSF. The purification of protein was performed immediately after the bacteria had been resuspended. This was done to prevent degradation, and the suspension of bacteria was kept at ice during the further experiments. The cells were lysed by sonication 3 x 30 seconds with 30 seconds intervals while the suspension was kept on ice. The lysate was centrifuged 40 minutes at 23,000 g at 4° C, to remove cellular debris. The supernatant was then filtrated through a 0.45 μm sterile filter.

Affinity Chromatography (FPLC)

The Tat101 protein was purified in one step by affinity chromatography on an FPLC system (Pharmacia). A 1 ml HiTrap Chelating column loaded with Zn^{2+} ions were used. The flow rate was set to 1.0 ml/min and 1-5 ml sample was applied to the column. The column was then washed with 10-15 volumes of Buffer A (section 2.14). Ten volumes of Buffer B (section 2.14) were then used to elute the his-tagged Tat protein from the column. With this buffer

change, the collection of 1 ml fractions were started. 110 µl 1 M Tris-Cl pH 8.0 was added to each fraction, to raise the pH above 7.0.

3.3.5 Estimation of protein concentration

Bradford analysis gave a rough estimate of the amounts of Tat101. The following procedure was used: A standard curve was made based on samples of BSA in 0.15 M NaCl in the range of 25-100 µg/ml. 100 µl each of the Tat101 solution and the standards were added 1 ml Coomassie brilliant blue solution and vortexed. After exactly 10 minutes at room temperature A_{595} was measured using 100 µl 0.15 M NaCl plus 1 ml Coomassie brilliant blue as blank.

3.3.6 Storage of Tat protein

The Tat101 protein was frozen at –20°C or kept at 4°C until it was to be used.

3.3.7 Western blot analysis.

Two filter papers were put on each of the pads and were heavily moistened with 1X blotting buffer. The SDS-PA gel, on the glass plate, was moistened with blotting buffer and put onto one of the filterpapers. A blotting sandwich was then made as folllows:

From the bottom: The black plate of the blotting plates, a pad, 2 pieces of filter paper, the gel, the nitrocellulose membrane, two more pieces of filter paper, a pad and then the whole thing was enclosed by the white plate of the blotting plates. While putting together this sandwich, bubbles of air between the layers within were removed carefully. The sandwich was put into the blotting apparatus, which was then filled with 1X blotting buffer. The blotting apparatus was put on ice and the electrophoretic transfer was conducted for 30-45 minutes at 100V.

After the transfer of proteins, the nitrocellulose membrane was quickly rinsed twice with TBS. Then the membrane was incubated in Blocking Solution (section 2.14) for 1 hour at room temperature to block the rest of the protein binding sites on the membrane. The membrane was then quickly rinsed twice, then once for 15 minutes and twice for five minutes with TBS-Tween. The membrane was then incubated with primary antibody at room temperature for 1 hour, or over night at 4^{o}C. The primary antibodies were diluted 1:10 (the anti-Tat antibodies) or 1:2000 (the anti-polyhistidine antibody). The membrane was rinsed twice quickly with TBS-Tween, once for 15 minutes and twice for 5 minutes. Further the membrane was incubated for approximately an hour at room temperature with secondary antibody (conjugated to Horseradish peroxidase (POD)), which had been diluted 1:2000 in

Blocking Solution The membrane was then washed quickly thrice, and once for 15 minutes with TBS-Tween, then twice for 5 minutes with TBS.

The protein bands recognized by the antibodies were visualized with ECL™ Western blotting detection reagents from Amersham Life Science by following the producer's guidelines.

The membrane was incubated in equal volumes (0.5 ml) of ECL solutions 1 and 2 (in the kit from Amersham, section 2.12) (which were mixed by vortexing seconds before their application) for 1 to 3 minutes at room temperature. The fluid was allowed to drain off, and the membrane was wrapped in plastic. In the dark, light sensitive film was then put upon the membrane inside of an X-ray film cassette. Depending on the strength of the signal, the film was developed anywhere between 10 seconds and several hours later.

3.4 Cell culture techniques

3.4.1 COS-7 Cells

These are adherent monkey kidney cells. They were grown in IMDM medium with 5-10% FCS, and 50 µg/ml Gentamicin. The growth medium was changed 3 times a week. The cells were subcloned when 80% confluence was reached.

3.4.2 Subcloning and harvesting of COS-7 cells

After a short wash with sterile PBS, 1 ml Trypsin was added to loosen the cells. The trypsin was inactivated after 1-2 minutes by adding FCS containing medium. The cells were then loosened from the flask wall by aspiration with a sterile pipette. The cell-containing medium, or parts of it, was discarded or transferred to another cell flask for expansion. For harvesting the cells were collected by centrifugation at 190 g for 2 minutes and the medium was discarded.

3.4.3 Transfection by electroporation

Cells were trypsinated and harvested by centrifugation at 190 g for 2 minutes. They were then washed in 10 ml sucrose buffer. The cells were harvested again, and resuspended in 1-2 ml sucrose buffer. For each electroporation 2.5-5 µg plasmid was added to 150 µl of the cell suspension and transferred to an electroporation cuvette. The mix was put on ice for 5 minutes before electroporation. The settings on the electroporator (Bio-Rad Gene Pulser II with an RF module) are shown on the next page:

Voltage	200
% Mod	100
DC Ampl	100
RF (kHz)	40
Burst Dur (msec)	2
No burst	5
Burst interv.	1

3.4.4 Transfection by $CaSO_4$ precipitation

The cells to be transfected were grown in either a 6-well plate or 24-well plate, to a density of 50-75%. The plasmid with which the cells were to be transfected was diluted with dH_2O to a total volume of 439 µl. 61 µl of 2M $CaCl_2$ were added, and the solution was vortexed. 500 µl of 2X HBS (section 2.14) was added, the solution was vortexed, and incubated at room temperature for 10-30 minutes. The mixture was added drop by drop to the medium above the cells (250 µl per well in a 24 well plate and 1 ml per well in a 6 well plate). The cells were incubated with the mixture for 18-24 hours. The medium was removed, and the cells washed 1–2 times with PBS before fresh medium was added to the cells. The cells were examined the next day.

3.4.5 A biological assay for Tat function

The HIV-1 LTR promoter is heavily dependent on Tat to produce functional mRNAs. This makes it possible to study the transcriptional activity of Tat. An LTR-reporter gene construct is transfected transiently or permanently into a cell culture and the activity of Tat can be measured by measuring the increased amounts or increased activity of the reporter protein. The reporter gene may for instance be EGFP, β-Galactosidase or Chloramphenicol Acetyl Transferase (CAT). In this study the LTR-CAT plasmid was used. Tat can then be expressed from a plasmid that has been co-transfected into these cells. Alternately this assay can be used to measure the uptake of extracellular Tat. This can be done by adding purified protein into the media in which these cells grow, or by transfecting other cells with the plasmid containing the *tat* gene and co-culture these cells with the cells containing the reporter plasmid. In the latter case one can detect a transcellular activity of Tat.

3.4.6 Preparation of cell lysates for Western blot analysis

The cells were lysed after having been incubated at 37° C for 48 hours. The cells were washed twice with cold PBS on ice. To each well 150 µl hypotonic buffer was added. The cells were collected with a cell scraper and transferred to an eppendorf tube containing 50 µl 4X sample buffer. The samples were then run on SDS-PAGE, or stored at –20° C for later use (Szilvay *et al.*, 1995).

3.4.7 Detecting cellular localization by fluorescence microscopy

To determine the intracellular location of a protein, it can be labelled with a fluorescing agent and viewed by fluorescence microscopy. This can be done by immunofluorescence techniques, where antibodies specifically binding to an antigen have been conjugated with a fluorescing molecule; or by expressing the protein of interest from a vector, where the gene for the protein to be viewed is fused with the gene for a fluorescing protein, such that a fusion protein is generated. Green Fluorescent Protein (GFP), Red Fluorescent Protein (RFP), Blue Fluorescent Protein (BFP), Cyan Fluorescent Protein (CFP), and Yellow Fluorescent Protein (YFP) are all available fusion partners in commercial vectors. In this study the *tat* genes were fused to EGFP in the pEGFP-C1 vector.

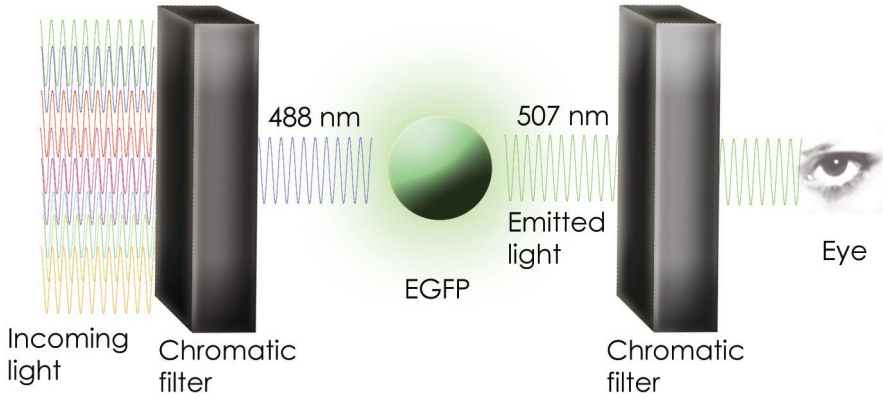


Fig 3.4. The principle of fluorescence microscopy. A chromatic filter makes sure that light of only one wavelength reaches the sample. If this wavelength corresponds to the excitation maximum of a chromophore (which in this figure is EGFP), then the chromophore will emit light of a different wavelength. The light will then pass through a second chromatic filter that makes sure only light emitted by the chromophore is visible.

3.4.8 Indirect immunofluorescence analysis

In order to visualize specific proteins in the cells, the coverslips on which the COS-7 cells were growing were washed twice with approximately 1 ml PBS in the wells, before they were incubated for 20 minutes at room temperature in 0.5 ml cold 4% paraformaldehyde (PFA) in PBS. The cells were subsequently washed once in PBS, and incubated for a minimum of 20 minutes at –20°C covered in pre-chilled methanol. The methanol was removed and the cells were washed three times with PBS.

Following this the coverslips were removed from their wells and placed, cells down, onto drops (30 µl) of a blocking buffer, 0.5% BSA in PBS, and were incubated for 15 minutes. The coverslips were then moved onto drops (25 µl) of primary antibody (the dilutions would vary, 1:10, 1:50, 1:400 depending on the antibody) and incubated for an hour at room temperature. The coverslips were washed thrice for 10 minutes on drops of PBS, and incubated 1 hour in the dark on 25µl drops of secondary antibody conjugated with $Alexa_{568}$ (diluted 1:100, 1:200,

or 1:600) at room temperature. When cells were to be examined only by the fluorescence of EGFP, then the steps in this paragraph were not performed.

The coverslips were washed thrice for 10 minutes on drops of PBS before washed briefly in dH_2O and mounted onto an object glass in 5 µl Slow-Fade mounting medium (Molecular Probes Inc.). The coverslips were then examined by fluorescence microscopy.

3.4.9 Microscopy and images

The immunostained and EGFP containing cells were visualised by using microscopes with filters for $Alexa_{568}$ and EGFP. The images were directly captured and stored digitally by using Open Lab imaging (Leica Microscope linked to a MacIntosh computer). The figures were created using Adobe Photoshop 7.0.

4 Results

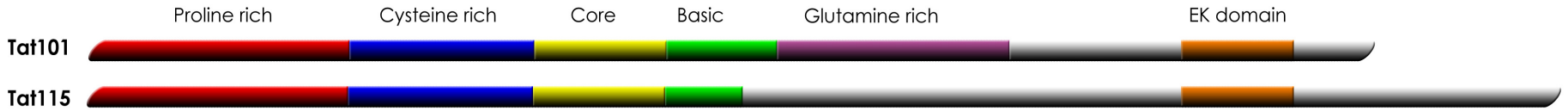
The Tat protein of HIV-1 was to be studied. Of particular interest was its second exon. A truncated 86 amino acid version of the Tat protein with the sequence from the HXB3 clone had already been extensively studied. It was decided to construct a full-length 101 amino acid long Tat variant based on the sequence of this clone. A point mutation at codon 87 would replace the premature stop codon, and this would render a gene encoding 101 amino acid residues. Any differences in the protein's abilities could then only be ascribed to whether the protein had a complete exon 2. This 101 amino acid variant was named Tat101. Another variant of the Tat protein was also to be studied in order to possibly gain new insights from the comparison of the two variants. It was decided to use the MVP5180 sequence as a basis for the other Tat protein to be studied. The gene to be constructed is called *tat115*, since it encodes a 115 amino acid protein. The sequence used in the following computer analyses is based on the *tat115* gene that was created, and not on the *tat* gene from MVP5180. One nucleotide differs between the *tat115* gene and the *tat* gene from MVP5180. This difference leads to an amino acid change, from serine at position 97 to threonine.

4.1 Computer analyses of the different Tat proteins

Different bioinformatical approaches were taken to analyze the amino acid sequences of Tat101 and Tat115. Results obtained from the Blast program, PSIPRED, PSORT II, and ScanProsite were assembled and compared (figure 4.1). The data that were used in the creation of this figure are included in the appendix (sections A.2.3 to A.2.8).

The Blast program aligned the two proteins so that maximum homology was obtained. This alignment made the comparison between the protein sequences easier. It resulted in a gap of 6 amino acids in the Tat115 sequence and a gap of 5 amino acids in the Tat101 sequence. This is shown in figure 4.1 A. In figure 4.1 B the regions that were not included by the Blast program have been manually added. The aligned sequence in this figure has been expanded with the appropriate amino acids. The homology investigation showed that the proteins have a high degree of similarity for the first 54 amino acids, 37 of these being identical, and 8 of the remaining 17 amino acids in this region have similar chemical properties. In the remainder of the proteins there is very little similarity, except for a stretch of ten amino acids, 86-95 of Tat101 and 85-94 of Tat115. Seven of these ten amino acids are identical, while 2 of the others have similar chemical properties. The apparent conservation indicates that this domain

```
a)
Tat101: MEPVDPRLEP WKHPGSQPKT ACTNCYCKKC CFHCQVCFIT KALGISYXXX XXXXXXXXXX XSQTHQVSLS KQPTSQSRGD PTGPK----- ESKKKVERET
        M+PVDP + P W HPGS+P+T  C NCYCK+C C+HC VCF   K LGIS+                       + K P  +      TG K      E +KKVE+ET
Tat115: MDPVDPEMPP WHHPGSKPQT PCNNCYCKRC CYHCYVCFTK KGLGISHGRK KRRRPAAAAS YPD------N KDPVPEQSLS HTGRKQKRQE EQEKKVEKET
b)
Tat101: MEPVDPRLEP WKHPGSQPKT ACTNCYCKKC CFHCQVCFIT KALGISYGRK KRRQRRRPPQ GSQTHQVSLS KQPTSQSRGD PTGPK----- ESKKKVERET ETDPFD
Align : M+PVDP + P W HPGS+P+T  C NCYCK+C C+HC VCF   K LGIS+GRK KRR+                 + K P  +      TG K      E +KKVE+ET
Tat115: MDPVDPEMPP WHHPGSKPQT PCNNCYCKRC CYHCYVCFTK KGLGISHGRK KRRRPAAAAS YPD------N KDPVPEQSLS HTGRKQKRQE EQEKKVEKET GPTGQPCHQD SCNSCTRISG Q
c)
```




```
d)
Tat101: CCCCCCCCCC CCCCCCCCCC CCCCCCCHHH HHHHHHHHHH CCCEECCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCC----- HHHHHHHHHH CCCCCC
Tat115: CCCCCCCCCC CCCCCCCCCC CCCCCCCHHH HHHHHHHHHH CCCEEECCCC CCCCCCCCCC CCC------C CCCCCCCCCC CCCCCCCCCC CCEEEEEEEC CCCCCCCCCC CCCCCCCCCC C
```

Figure 4.1: An assembly of the results of the computer analyses of Tat101 and Tat115.
a) Result of pairwise Blast
b) A figure based on the blast alignment of the two proteins and the predicted signal sequences from the Scan Prosite tool and the PSORT II tool. In b the amino acids that were not considered by the blast engine have been added, both in the Tat101 sequence as well as in the aligned sequence.
S/T Serines and threonines that might be phosphorylated by protein kinase C, PKC or Casein Kinase 2, CK2
X Amino acids that are part of a cystein rich domain
K/R Amino acids that are predicted to play a part in nuclear localization
c) Organization of the Tat protein. The domains have been marked with colour and are mostly defined based on studies of the Tat protein from group M. The glutamine rich domain is only found in Tat101; its properties are not well defined. Amino acids 86-95 of Tat101 and 85-94 of Tat115 have all been marked in orange. The Blast alignment shows a resemblance in the sequence of these proteins here, possibly indicating a domain with a preserved function that is yet unknown. This part of the proteins have therefore been called the EK domain.
d) PSIPRED Prediction of secondary structure (with sequences aligned as based on the blast result) (H=helix, E=β-strand, C=coil)

has a biological function. This novel domain was named the EK domain to reflect its high content of glutamate and lysine.

In figure 4.1 B the results from the ScanProsite tool as well as the PSORT II tool have been marked in colour. They show that both proteins have a cystein rich domain comprising amino acid residues 22-37, and a nuclear localization signal, NLS. In Tat101 this NLS was predicted to consist of amino acids 49-53, while in Tat115 it was predicted to be amino acids 49-54. The Tat101 protein has previously been reported to have an NLS consisting of amino acids 49-57 that also functions as a Nucleolar localization signal, NOS (Jeang *et al.*, 1999). The PSORT II program also indicated a novel second NLS, PKESKKK, in the Tat101 protein, at amino acids 84-90. This sequence is not found in Tat115, even though it is partially located in the EK domain, where there is homology between the two proteins. The PSORT II program also indicated that the NLS in Tat115 at residues 49-54 may be a part of a bipartite NLS that includes the lysines 40 and 41. Furthermore, this program predicted that some threonines and serines may be phosphorylated in these proteins. Phosphorylation of these could be a mechanism for modulating the functions of the Tat protein. However, the amino acids that were predicted to be phosphorylated are not located at the same positions in the two different proteins.

Figure 4.1.C gives an overview of the different domains of the proteins. The different domains have been marked in different colours. The naming and location of the domains are based on earlier studies of Tat protein from HIV-1 group M, like Tat101. There is a high degree of homology between the first 54 amino acids of the two Tat proteins. The proline rich, the cystein rich, the core and the RNA binding domain; have been marked similarly in the two proteins. The RNA binding domain of Tat101 is shown to stretch three residues further than that of Tat115 in order to include three arginines. Then a glutamine rich domain is shown in purple. No such glutamine rich domain was found in Tat115, and the remainder of that protein is shown in grey to indicate the lack of any known domain. The exception is the EK domain mentioned previously. This domain has been marked in orange.

PSIPRED was used to predict the secondary structure of the two proteins. The results are shown in figure 4.1.D, and they were aligned in accordance with the Blast result in 4.1.A. Large parts of these two proteins do not seem to have a secondary structure. Amino acids 28-40 were predicted to be helical in both proteins, and amino acids 44-46 of Tat115 and 44-45

of Tat101 were predicted to be β-strands. Furthermore, the program predicted that there is secondary structure in the EK domain: helical in Tat101 and β-strand structure in Tat115.

4.2 Construction of genes, cloning of genes into vectors

4.2.1 The construction of the *tat101* and *tat115* genes

The *tat101* gene was constructed by amplifying the *tat86* gene from the vector EGFP-Tat86 with Primers A and B, and by constructing the remainder of the *tat101* gene with Primers C and D (Materials, section 2.8). These two sequences were designed to be overlapping and were amplified with Primers A and D, as described in Methods, section 3.1.3. The sequence for the *tat86* gene, as found in the EGFP-Tat86 vector, comes from the HXB3 isolate and encodes the first 86 codons of the resulting *tat101* gene.

The *tat115* gene was created by using 6 large oligonucleotides called Tatmvp1-6, and was amplified by using two smaller ones, Tatmvpa and Tamvpb (Materials, section 2.8), as described in Methods, section 3.1.4.

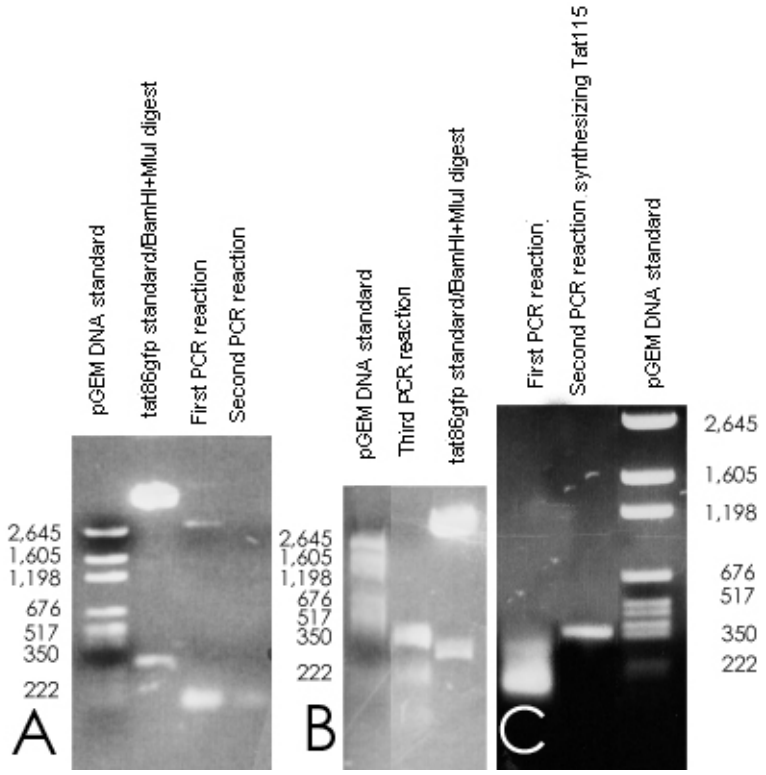


Figure 4.2: Agarose gel analysis of the PCR reactions resulting in the construction of *tat101* and *tat115*. The pGEM size markers were used, and the relative sizes are inidcated at the side of the gels. A) In the first reaction the *tat86* gene is amplified, but only the EGFP-Tat86 plasmid and the primers used can be seen on the gel. In the second PCR reaction the primers primer C and D have been run to create the end fragment of *tat101*. Only the primers can be seen on the gel. B) In the third PCR reaction, 1 µl of each of the two former PCR reactions have been used as template, and the agarose gel analysis shows that the product of this reaction is as expected a little over 300 bp. In C the PCR reactions resulting in the construction of *tat115*. The first reaction yields a smear on the gel. In the second reaction this smear is used as template, and the size of the PCR-result in this reaction was as expected approximately 370 bp.

Agarose gels (section 3.1.5) were run to confirm that the PCR products were of expected size, and were also used as a means to estimate the yield of the PCR reaction (figure 4.2). The PCR product containing the *tat101* gene was expected to be about 320 bp. In figure 4.2.A there is no visible middle product in the two first PCR reactions that were to generate *tat101*. But in

the third PCR reaction (figure 4.2.B) where the products of these reactions are used as template, the expected band of ~320 bp is visible. The PCR product containing the *tat115* gene was expected to be about 370 bp. In figure 4.2.C it can be seen that the first PCR reaction that was to generate *tat115* generated a smear, and that in the second PCR reaction, in which the product of the first was used as a template, the expected *tat115* band of ~370 bp appears.

4.2.2 The construction of the EGFP-Tat101 and EGFP-Tat115 plasmids

The EGFP-Tat101 and EGFP-Tat115 plasmids were constructed as outlined in section 3.1.5. The *tat* genes were ligated into the pEGFP-C1 plasmid such that the gene product would be fused with the EGFP protein. This allowed for the determination of the intracellular distribution of the Tat proteins.

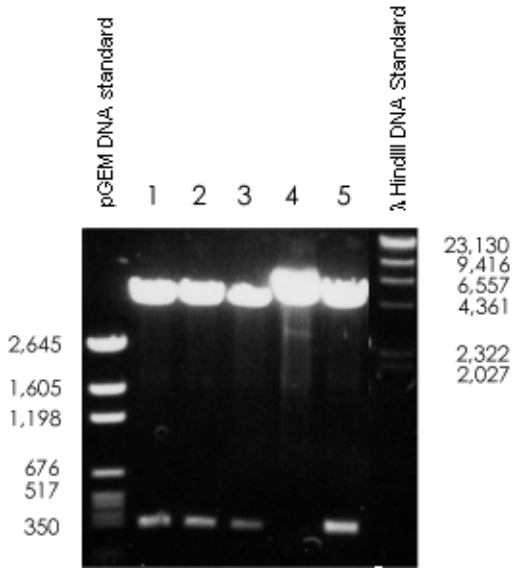


Figure 4.3: Restriction enzyme analysis of EGFP-Tat115 plasmid preps. 1-5 are different minipreps digested with *Bam*HI and *Hin*dIII. If they contained the *tat115* gene, a band of approximately 370 bp in addition to the plasmid band of approximately 4700 bp would be seen. As seen in the figure, 1-3 and 5 show this, while 4 does not, and that indicates that it is in all likelihood the religated EGFP-C1 plasmid. The pGEM Markers and λ/HindIII digest have been used as size markers. The relative sizes are indicated at each side of the gel.

The purified plasmid from these was then investigated by restriction enzyme analysis. The assumed EGFP-Tat115 plasmid was cut with *Bam*HI and *Hin*dIII, while the assumed EGFP-Tat101 plasmid was cut with *Eco*RI and *Hin*dIII. The restriction enzyme analysis of the possible EGFP-Tat115 plasmids is shown in figure 4.3. A band of the expected size, ~370 bp, excised from several of the plasmids was observed (figure 4.3, lanes 1-3 and 5). The corresponding data for EGFP-Tat101 is not shown.

The plasmids were then sequenced, and purified in larger amounts. The identities of these purified plasmids were established by restriction enzyme analysis (fig 4.4). In figure 4.4.A and Bit can be seen that a band of correct size, ~300 bp for *tat101* and ~370 bp for *tat115*, is excised from the purified EGFP-Tat plasmids when they are digested with the restriction enzymes that were used to ligate the *tat* genes into the pEGFP-C1 plasmid. An outline of the resultant plasmids is shown in figure 4.11. These constructed vectors were used in the experiments in sections 4.4.1 and 4.4.2.

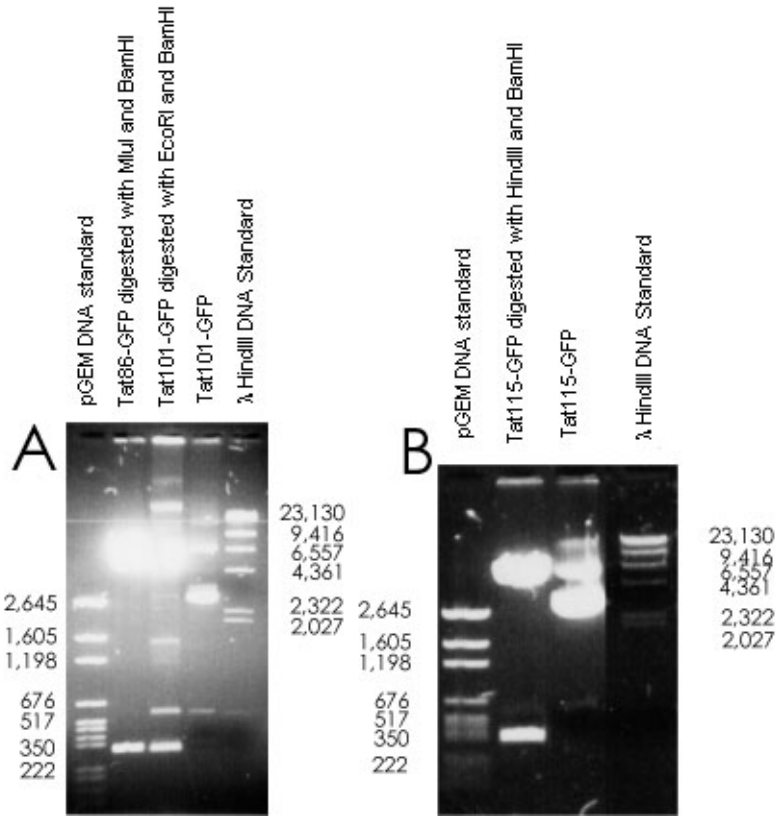


Figure 4.4: Restriction enzyme analyses of the EGFP-Tat101 and the EGFP-Tat115 plasmids. The pGEM Markers and λ/HindIII digest have been used as size markers. The relative sizes are indicated at each side of the gels. A) Analysis of the EGFP-Tat101 plasmid after purification by megaprep. When digested with *Bam*HI and *Eco*RI, the two restriction enzymes used prior to ligation, a fragment of ~300 bp, the expected size of *tat101*, was visible. This is also shown by comparison to the band cut out of the EGFP-Tat86plasmid with the enzymes *Mlu*I and *Bam*HI, which is supposed to yield a band of that size. B) Analysis of the EGFP-Tat115 plasmid after purification by megaprep. The expected size of *tat115* is ~370 bp, and when this plasmid is digested with *Bam*HI and *Hin*dIII, the two restriction enzymes used prior to ligation, a fragment of that size is visible.

4.2.3 The construction of the pRSET-Tat101 plasmid

The *tat* genes were inserted into the pRSET-C vector. When the pRSET-Tat101 vector was constructed, this was done as outlined in section 3.1.6.

The *tat101* and *tat115* were amplified by PCR. Figure 4.5 shows that the PCR products were of the expected size (~300 and ~370 bp). The PCR product containing the *tat101* gene was ligated into the pRSET-C plasmid. The purified plasmid from this reaction was subjected to restriction enzyme analysis (figure 4.6). This figure shows that a band of the correct size was excised from two of the three plasmids analysed on that gel, when these plasmids were digested with both of the two restriction enzymes used when the *tat101* gene was ligated into the plasmid, but not when only one of these was used. These plasmids were then sequenced to establish their identity, and one of them were purified in large amounts. As a

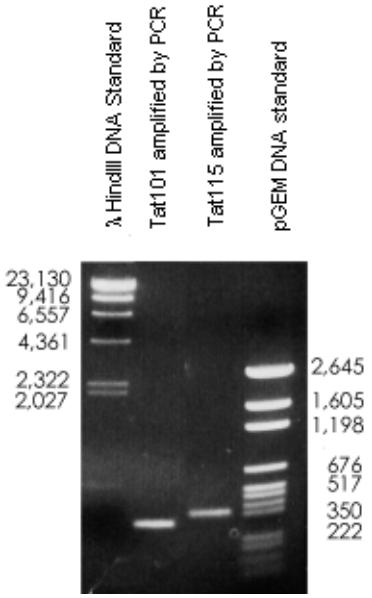


Figure 4.5: Agarose gel analysis of PCR products. The gel shows the PCR reactions where *tat101* and *tat115* have been amplified. The size of the PCR-results were as expected, approximately 300 and 370 bp, and no other products can be seen. The pGEM Markers and λ/HindIII digest have been used as size markers. The relative sizes are indicated at each side of the gel.

final insurance that it was the correct plasmid that had been obtained, the identity of this plasmid was afterwards verified by restriction enzyme analysis (figure 4.7), which showed the plasmid to have the correct pattern like two of the plasmids in figure 4.6. This plasmid was further used in the experiments in section 4.3.1.

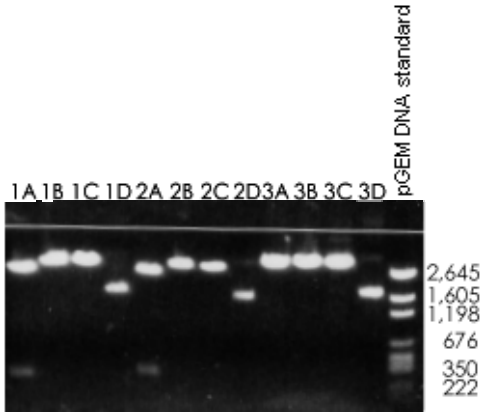


Figure 4.6: Restriction enzyme analysis of the pRSET-Tat101 plasmid. 1-3 are different minipreps of bacteria transformed with pRSET-C ligated with the *tat101* gene. In A they have been digested with *Xho*I and *Eco*RI, in B *Eco*RI, in C *Xho*I, and in D they have not been digested with any restriction enzymes. *Xho*I and *Eco*RI were the enzymes used prior to ligation and one would expect a band of approximately 300 bp to have been cut out of the pRSET plasmid with those two enzymes if the gene had been ligated into the vector. This can be seen for 1 and 2, but not 3. The pGEM Markers has been used as a size marker. The relative sizes are indicated to the right of the gel.

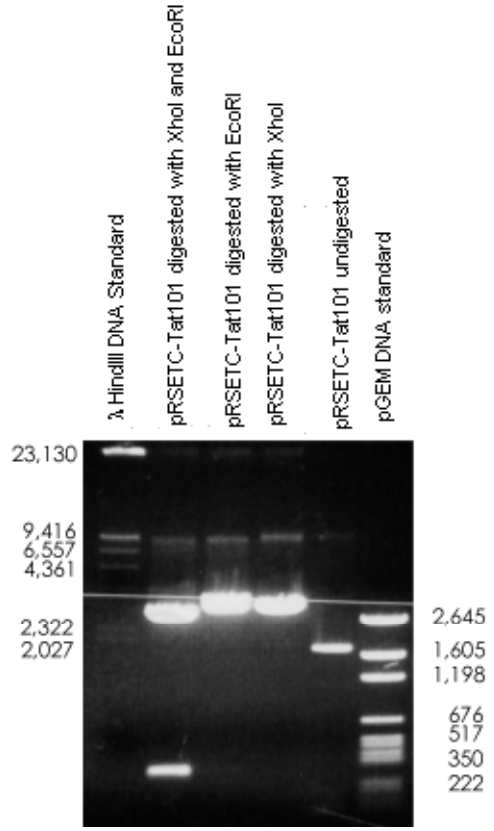


Figure 4.7: Restriction enzyme analysis of the pRSET-Tat101 plasmid after purification by megaprep. When digested with *EcoRI and XhoI*, the two restriction enzymes used prior to ligation, a fragment of ~300 bp, the same as *tat101*, was visible in addition to the main plasmid band of ca. 2700 bp. The pGEM Markers and λ/HindIII digest have been used as size markers. The relative sizes are indicated to each side of the gel.

4.2.4 The construction of the pRSET-Tat115 plasmid

The pRSET-Tat115 plasmid was constructed as outlined in section 3.1.7. The PCR product containing the *tat115* gene was subcloned into a TOPO vector, pCR-BLUNT II-TOPO. This was followed by restriction enzyme analysis. The results (figure 4.8) showed that only one clone contained the *tat115* gen. The *tat115* gene was then ligated into the pRSET-C vector, in a similar manner as done with the *tat101* gene (section 4.2.3). Figure 4.9 shows the restriction

enzyme analysis of the plasmid preps. This figure shows that the *tat115* gene had been cloned into two of three clones. Figure 4.10 shows an agarose gel analysis of the purified pRSET-Tat115 plasmid. DNA sequencing affirmed the identity of the plasmid. Figure 4.11 outlines the most important features of the different Tat plasmids created. These plasmids were used in the experiments described in sections 4.3.1, 4.4.1 and 4.4.2.

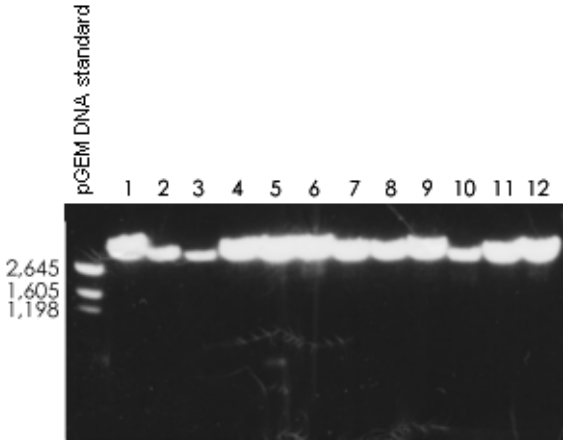


Figure 4.8: Restriction enzyme analysis of the pCR-BLUNT II-TOPO vector after the *tat115* gene had been subcloned into it. Lanes 1-12 contain different minipreps. All the minipreps have been digested with *Xho*I, which should lead to the release of a fragment of approximately 450 bp. This can be seen in lane 5. The pGEM Markers has been used as a size marker. The relative sizes are indicated at the right of the gel.

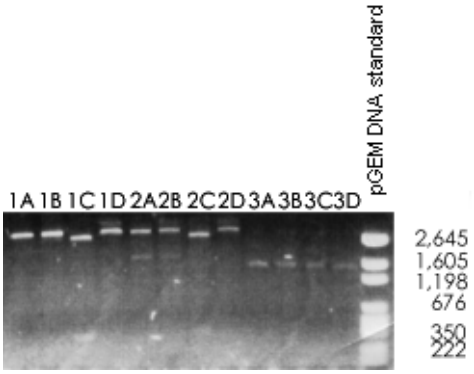


Figure 4.9: Restriction enzyme analysis of the pRSET-Tat115 plasmid. 1-3 are different minipreps. In A they have been digested with *Bam*HI (there is a cleavage site for this enzyme within the *tat115* gene, and if the plasmid is linearised by treatmeant with this enzyme, this would show that the correct gene has been inserted. In B *Eco*RI, in C *Eco*RI and *Xho*I, and in D *Xho*I. The restriction enzymes used prior to ligation were *Eco*RI and *Xho*I, and a fragment of approximately 350 bp when the plasmid had been digested with both these enzymes would show the correct insert. This can be seen for minipreps 1 and 2. 3 seems to be a failed miniprep. The pGEM Markers has been used as a size marker. The relative sizes are indicated at the right of the gel.

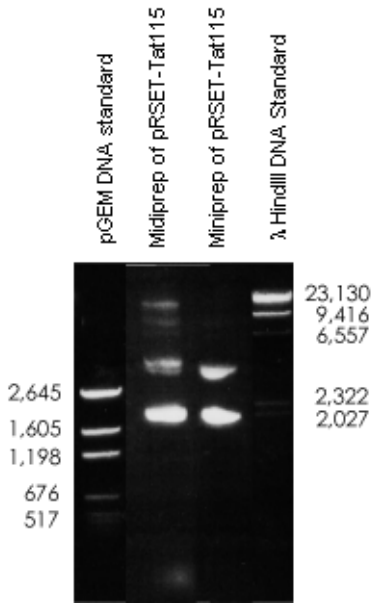


Figure 4.10: Agarose gel analysis of thepRSET-Tat115 plasmid. After the miniprep, pRSET-Tat115 was purified by midiprep. A plasmid of the same size as that in the miniprep was purified, and identity was established by DNA sequencing of the miniprep. The pGEM Markers and λ/HindIII digest have been used as size markers. The relative sizes are indicated at each side of the gel.

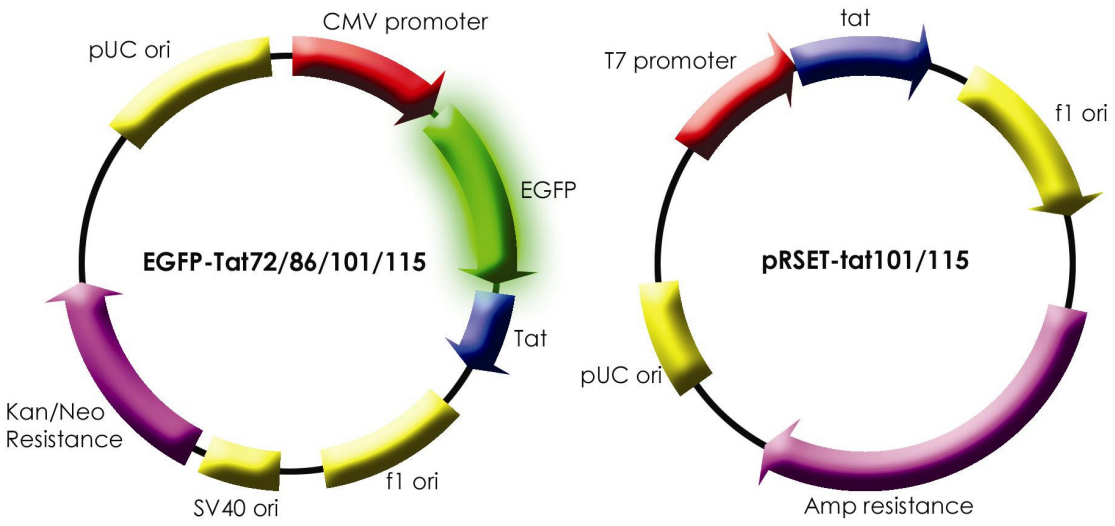


Figure 4.11: The resultant plasmids to be used in the further experiments (sections 4.3.1, 4.4.1 and 4.4.2). EGFP-Tat72 and EGFP-Tat86 were kindly provided by Vibeke Andresen and have the same basic elements as the EGFP-Tat115 and EGFP-Tat101 plasmids constructed in this thesis.

4.3 Purification of Tat protein

4.3.1 Expression and affinity chromatography of Tat101

The Tat101 protein was expressed recombinantly using the pRSET-Tat101 vector (3.3.3), and was successfully purified by affinity chromatography (3.3.4). The supernatant of bacterial lysate from the bacteria transformed with pRSET-Tat101 was run through the Hitrap chelating column on the FPLC system. The chromatogram showed a large peak after the sample has been applied to the column, which indicated that most of the proteins in the lysate were washed out with the flowthrough. After the pH was changed to 4.0, a small peak emerged, indicating that the Tat101 protein was indeed expressed, and was purified (figure 4.12). When bacteria transformed with the empty pRSET-C vector, there was no such peak released at pH 4.0 (data not shown). This indicates that it was indeed the 6Xhis-Tat101 protein that had been purified. The same procedure was attempted to purify the Tat115 protein. However, the amount of purified Tat115 protein was too small to be used in the transcellular activity assay. In figure 4.13 samples from the different stages of purification of the Tat101 protein have been run on SDS-PAGE (section 3.3.1). In A the proteins have been visualized by Coomassie blue colour (section 3.3.2), and in B the Tat protein is detected by Western blot (section 3.3.7).

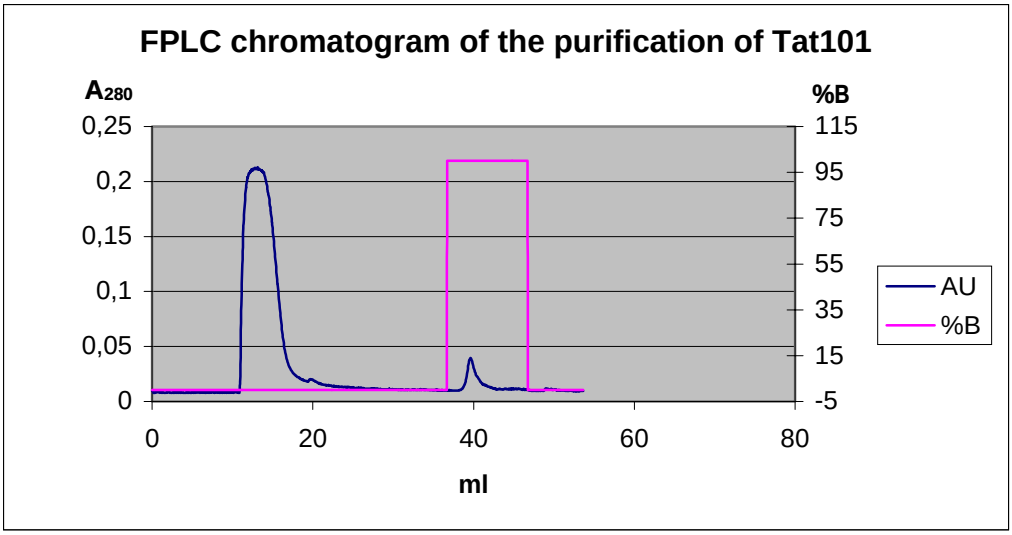


Figure 4.12: A chromatogram showing the purification of the Tat101 protein. The left axis shows the scale of the A_{280} values, which are shown drawn in blue. The right axis shows the percentage of Buffer B used, which is drawn in pink. At 0% B, the pH is 8.0, at 100% B, the pH is 4.0. Most of the filtrated bacterial lysate was flushed through with Buffer A (pH 8.0). The buffer was changed to Buffer B (pH 4.0). Anything that had bound to the Zn^{2+} on the column was then released. It can be seen that a small top does emerge in the A_{280} after addition of Buffer B.

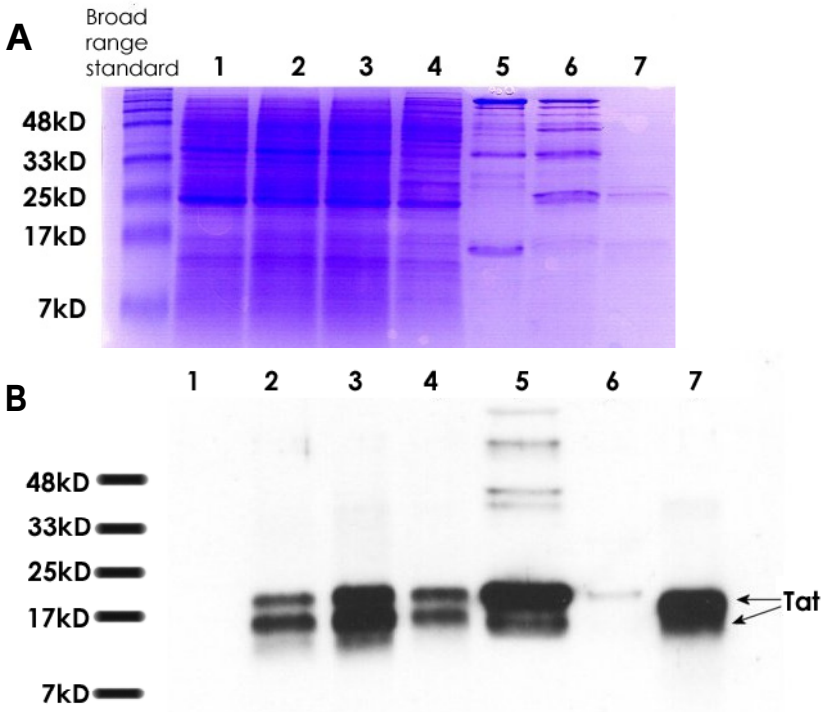


Figure 4.13: Two similar SDS-PAGE were run to document the purification of the Tat101 protein. In A the proteins on the gel have been visualised by Coomassie blue staining, while in B the contents of the gel have been subject to Western blot analysis using the mAb 1D9D5 as primary antibody, and anti-mouse conjugated to POD as secondary antibody. The Western blot was developed by ECL. In B lanes 1-4 had been diluted 1:450 when compared to the amounts applied to the gel shown in A. The NEB prestained broad range standard was used as a size marker, and the relative sizes are indicated to the left. What was applied in the different lanes: Lane 1: Total protein from *E.coli* transfected with pRSET-C, Lane 2: Total protein from *E.coli* transfected with the pRSET-Tat101 plasmid, Lane 3: Pellet after sonication of *E.coli* transfected with the pRSET-Tat101 plasmid, Lane 4: Supernatant after sonication of *E.coli* transfected with the pRSET-Tat101 plasmid, Lane 5: Fraction 7 eluted after the filtrated supernatant after sonication of *E.coli* transfected with the pRSET-Tat101 plasmid on a HiTrap chelating FPLC column, Lane 6: Fraction 8 eluted after the filtrated supernatant after sonication of *E.coli* transfected with the pRSET-Tat101 plasmid on a HiTrap chelating FPLC column, Lane 7: Previously purified 6Xhis Tat86 protein (Mendonca and Lavoura, 1998)

In A, none of the lanes 2, 3 and 4, where the samples come from bacteria transformed with the pRSET-Tat101 plasmid, differ significantly from lane 1, where the sample comes from bacteria transformed with the empty pRSET-C vector. But the fractions taken when the top emerged in figure 4.12 contained a few clear bands. One of these is slightly smaller than 17 kD, as seen in lanes 5 and 6, showing that something had indeed been purified. When the sizes of these bands were compared with those in lane 7, where previously purified Tat86 protein had been run, the bands were of expected and comparable size, indicating that it was Tat protein that had been purified. On the Western blot (figure 4.13.B) there are two bands in all the lanes where the samples originated from bacteria transformed with pRSET-Tat101, except from lane 6 where only one is visible and barely so, while no such bands are visible when the bacteria had been transformed with pRSET-C. This showed that the Tat101 protein had been expressed and purified, and also suggested that in addition to the full fusion protein, a smaller, possibly degraded version of the protein was purified along with it. If this was indeed a degraded version of the protein, then this was a good opportunity to test the epitope specificity of some monoclonal anti-Tat antibodies previously shown to bind to a peptide spanning amino acids 49-86 (Valvatne *et al*., 1996). This was done in section 4.3.2, and the purified protein was further used in section 4.4.1.

4.3.2 Attempted epitope mapping of anti-Tat antibodies

A Western blot was run to test the epitope specificity of mAbs 4B4C4, 1D2F11, 2D9E7, and 5G7D8 (section 3.3.7). Each lane contained extract from bacteria induced to express the Tat101 protein. The Western blot was cut into strips and each of the antibodies to be tested was incubated with one of the strips. An antibody shown to bind to a peptide spanning amino acids 1-20 of Tat, 1D9D5, was used as a control, as was an antibody binding to the his-tag of the fusion protein. As can be seen in figure 4.14, several bands turned up when the Western blot was developed. Since lane 1 in figure 4.13.B showed that no bacterial proteins reacted with the primary antibody 1D9D5 and the secondary anti-mouse antibody, the more powerful bands in the Western blot in figure 4.14 were thought to be Tat protein. The anticipated Tat band at approximately 16.5 kD is there, and any bands larger than this are possibly multimeric forms of Tat. All the lanes showed a band that was a little smaller than 16.5 kD. Seeing how this band also was present in the lane where the anti-his antibody had been used, the degradation of the protein could not have been N-terminal.

In some, but not all the lanes a third band, a little smaller than these first two was present. The first lane, where the anti-his antibody had been used, does not show the band in this figure,

but when the Western blot was overexposed (data not shown) this band was discernible also in this lane, showing that the putative degradation of the protein could only to a small extent have been N-terminal, and that most, if not all, of the amino acids missing (which would explain the small size of the band), were from the C-terminus. The lanes where the antibodies 1D9D5, 4B4C4 and 1D2F11 were used, show this band clearly, while the lane where 5G7D8 has been used shows this band a little bit less clearly. However, the lane where the antibody 2D9E7 has been used does not show this band. This would indicate that 2D9E7 binds to an epitope closer to the C-terminus of the Tat protein than the others. The fact that this band turns up in some but not all of the lanes excludes the possibility that the band is caused by the secondary antibody binding to a bacterial protein, and the fact that primary antibodies shown to bind different epitopes indicate the same band show that the protein is probably a modified version of Tat.

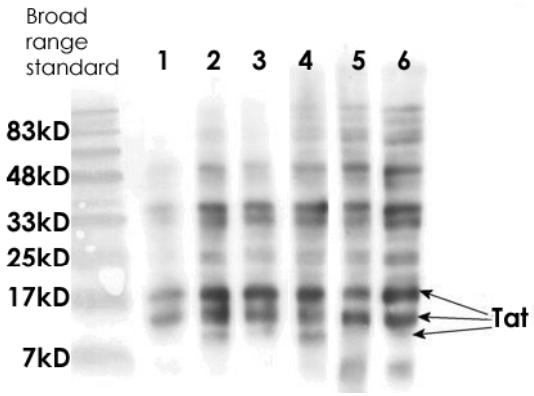


Figure 4.14: This Western blot was performed to see if the epitopes to which the different antibodies bind could be more clearly defined. In each well there was applied the supernatant after sonication of *E.coli* transfected with the pRSET-Tat101 plasmid, as seen in lane 4 of figure 4.14.B. The negative control in lane 1 of that same figure was thought appropriate for use in this figure as well. In lanes 1-6 monoclonal antibodies recognizing different epitopes on the 6xhis-Tat101 protein have been used. The Broad Range Standard (NEB) is used to show the size of the bands, and the relative sizes are indicated to the left. The primary antibody used in the different lanes: Lane 1: Anti-his-tag antibody, lane 2: 1D9D5, lane 3: 4B4C4, lane 4: 1D2F11, lane 5: 2D9E7, lane 6: 5G7D8. The secondary antibody was anti-mouse conjugated to POD, and the Western blot was developed by ECL.

Calculations were made to determine the sizes of the proteins seen in these bands in kD and the number of amino acids, to be able to define the location of the epitopes these antibodies bind to (data not shown). From smallest to largest, the three bands recognized as 6xhis-Tat were 16.5, 10 and 8 kD, corresponding to Tat proteins consisting of 96, 55 and 28 amino acids when the N-terminal 39 amino acids of the fusion partner has been subtracted. This would then indicate that all the mAbs to be tested but 2D9E7 bind an epitope between amino acid 1-28, while 2D9E7 bind an epitope between 1 and 55.

4.4 Studies of Tat function and cellular localization studies of Tat

4.4.1 Expression of the CAT protein in COS-7 cells to show transcriptional transactivation and transcellular transcriptional transactivation

The transcriptional transactivation activities of the purified Tat protein and the EGFP-Tat fusion protein were examined. COS-7 cells transfected with the LTR-CAT construct were used in an assay to investigate this (section 3.4.5). Purified Tat protein was added to such cells and other cells were co-transfected with the LTR-CAT plasmid and the EGFP-Tat plasmids (section 3.4.4). The next day the cells were lysed and CAT expression was detected by Western blot (figure 4.15). A band of approximately 25 kD, which is the expected size of CAT is visible in lanes 3-8 in that figure. There is no such band in lane 2, where the COS-7 cells have been transfected with the LTR-CAT vector, but there is no Tat protein present. The CAT-band in lane 7, where the sample comes from cells that have had Tat86 protein added to the media in which they grow, is very weak, but still discernible. Since the CAT protein was detected this shows that the EGFP-Tat115 protein (lane 3), the EGFP-Tat101 protein (lane 4) and the 6xhis-Tat101 protein (lane 6) all transactivated the LTR promoter. This Western blot also showed that the EGFP-Tat86 protein (lane 5) and purified 6Xhis-Tat86 protein (lane 7) transactivates the LTR promoter as expected. Since the CAT protein was expressed when the purified Tat101 protein was added to the media (lane 6), this indicated that the protein had been taken up by the cells and that it had a transcellular activity.

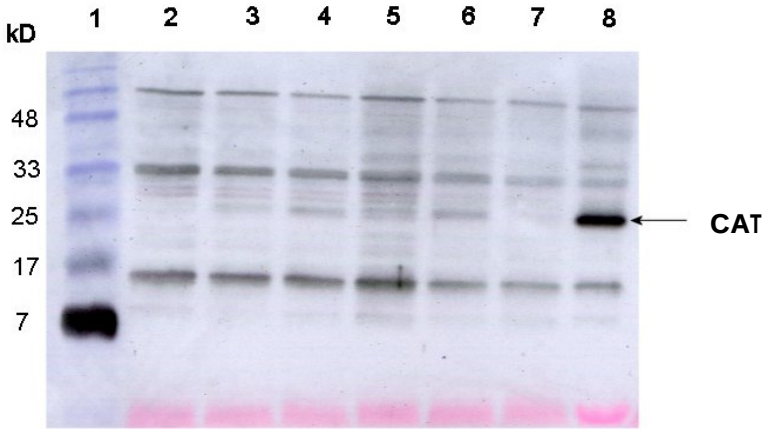


Figure 4.15: Western blot analysing the functionality of the Tat101 and Tat115 proteins on the HIV-1 promoter. The first lane contains the NEB prestained protein standard; the sizes of the proteins are indicated to the left. In lanes 2-8 lysate of COS-7 cells transfected with the LTR-CAT vector have been applied. The cells used in lanes 3-8 were either co-transfected with a Tat expressing plasmid, or Tat protein was added to the media. The cells were lysed 48 hours after transfection. Lane 2: COS-7 cells transfected with LTR-CAT, lane 3: COS-7 cells transfected with LTR-CAT and EGFP-Tat115, lane 4: COS-cells transfected with LTR-CAT and EGFP-Tat101, lane 5: COS-7 cells transfected with LTR-CAT and EGFP-Tat86, lane 6: COS-7 cells transfected with LTR-CAT and with 7.5 µg 6Xhis Tat101 protein added to the medium after 24 hours, lane 7: COS-7 cells transfected with LTR-CAT and with 7.5 µg 6Xhis Tat86 protein added to the medium after 24 hours, lane 8: COS-7 cells transfected med LTR-CAT and EGFP-Tat101 (from Vibeke Andresen). The cells in lane 8 were transfected with FuGene, while the cells used in the other lanes were transfected by the calcium sulphate precipitation method. The primary antibody used was the anti-CAT antibody, the secondary antibody used was anti-rabbit conjugated to POD, and the Western blot was developed by ECL.

4.4.2 Intracellular localization of the Tat proteins.

The main reason for constructing the EGFP-Tat plasmids was to be able to examine the cellular distribution of the Tat101 and Tat115 proteins. The plasmids expressing these EGFP fusion proteins were transfected into COS-7 cells either by the calcium sulphate precipitation method or by electroporation (sections 3.4.3 and 3.4.4). The cells were grown in 24 well plates on coverslips and fixed with paraformaldehyde and methanol after approximately 48 hours in culture (section 3.4.8). The Tat protein was also detected by immunofluorescence using the monoclonal antibody 1D9D5, which binds to the N-terminal part of Tat. Tat115 could however not be detected by immunofluorescence, most likely due to difference in sequence in the N-terminal part of Tat. The cells were then examined by fluorescence microscopy and images were captured (section 3.4.9). The same experiment was also conducted with plasmids expressing EGFP, EGFP-Tat72 and EGFP-Tat86, allowing the intracellular distribution of these proteins to be compared with that of EGFP-Tat101 and EGFP-Tat115 (figures 4.16 - 4.20).

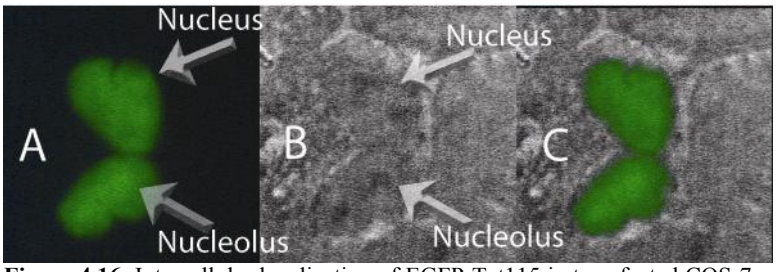


Figure 4.16: Intracellular localization of EGFP-Tat115 in transfected COS-7 cells. A: The fluorescence of the EGFP moiety of EGFP-Tat115, B: The cells visualized by phase contrast image, C: The fluorescence of EGFP overlaid on the phase contrast image of the cells.

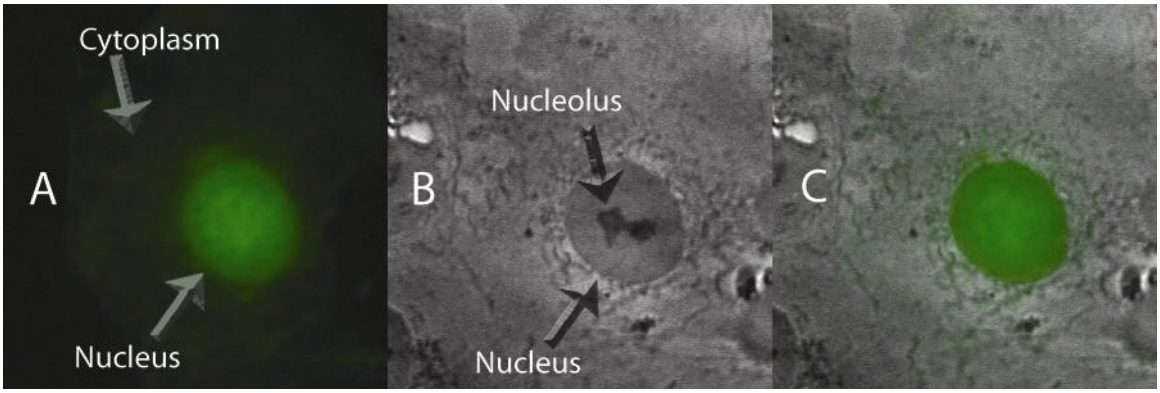


Figure 4.17: Intracellular localization of EGFP in trasfected COS-7 cells. A: The fluorescence of the EGFP, B: The cells visualized by phase contrast image, C: The fluorescence of EGFP overlaid on the phase contrast image of the cells.

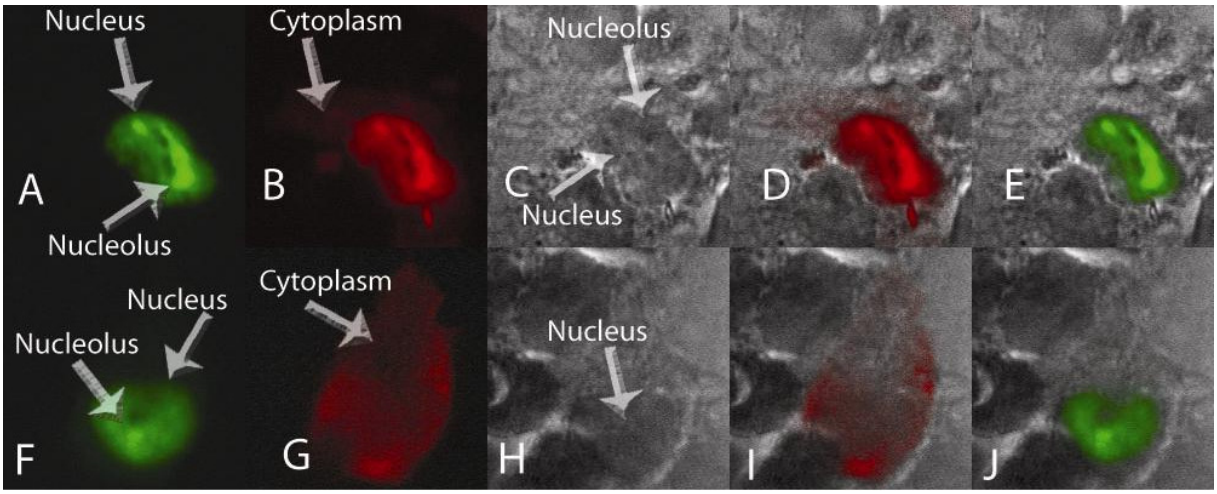


Figure 4.18: The intracellular localization of EGFP-Tat101 in transfected COS-7 cells. A: The fluorescence of the EGFP moiety of EGFP-Tat101, B: Immunofluorescence visualizing Tat in red, C: The cells visualized by phase contrast image, D: The immunofluorescence overlaid on the phase contrast image of the cells, E: The fluorescence of EGFP overlaid on the phase contrast image of the cells. F: The fluorescence of the EGFP moiety of EGFP-Tat101, G: Immunofluorescence visualizing Tat in red, H: The cells visualized by phase contrast image, I: The immunofluorescence overlaid on the phase contrast image of the cells, J: The fluorescence of EGFP overlaid on the phase contrast image of the cells.

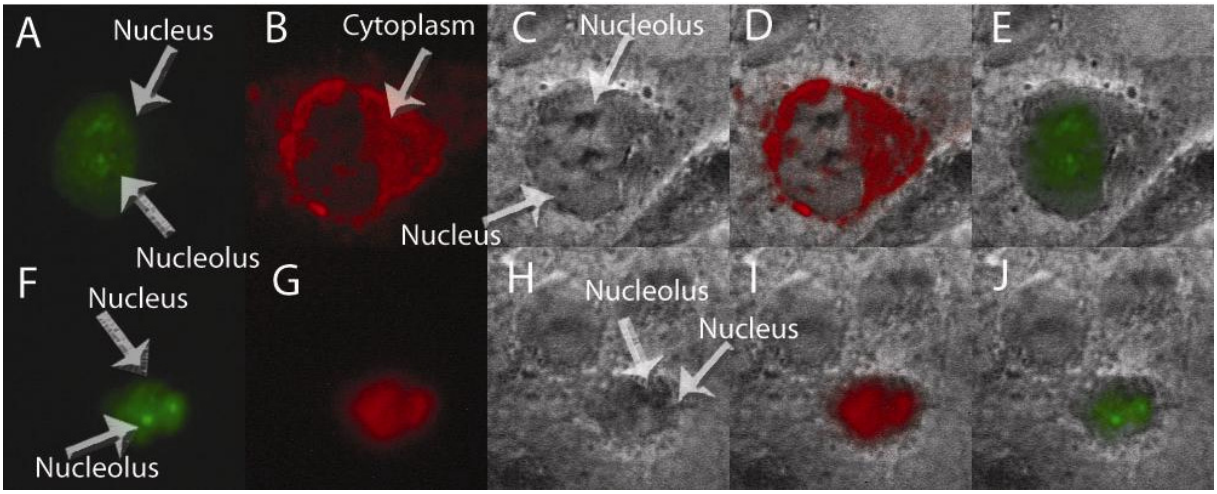


Figure 4.19: The intracellular localization of EGFP-Tat72 in transfected COS-7 cells. A: The fluorescence of the EGFP moiety of EGFP-Tat72, B: Immunofluorescence visualizing Tat in red, C: The cells visualized by phase contrast image, D: The immunofluorescence overlaid on the phase contrast image of the cells, E: The fluorescence of EGFP overlaid on the phase contrast image of the cells. F: The fluorescence of the EGFP moiety of EGFP-Tat72, G: Immunofluorescence visualizing Tat in red, H: The cells visualized by phase contrast image, I: The immunofluorescence overlaid on the phase contrast image of the cells, J: The fluorescence of EGFP overlaid on the phase contrast image of the cells.

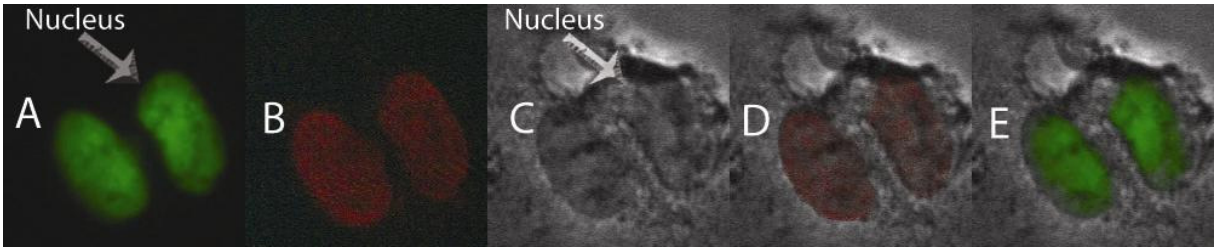


Figure 4.20: The intracellular localization of EGFP-Tat86 in transfected COS-7 cells. A: The fluorescence of the EGFP moiety of EGFP-Tat86, B: Immunofluorescence visualizing Tat in red, C: The cells visualized by phase contrast image, D: The immunofluorescence overlaid on the phase contrast image of the cells, E: The fluorescence of EGFP overlaid on the phase contrast image of the cells.

The images showed the location of Tat101, Tat86 and Tat72 by its EGFP moiety (figures 4.18 (A, E, F, J), 4.19 (A, E, F, J), and 4.20 (A, E)); the proteins were seen mainly in the nucleus

and especially the nucleoli. There is a weak fluorescence in the cytoplasm that cannot be observed in these images. The results of the immunofluorescence were somewhat contradictory when it came to the Tat72 protein and the Tat101 protein as can be seen in the figures 4.18 and 4.19. In most cells the results are similar, but with a stronger fluorescence in the cytoplasm. However, in some cells the immunofluorescence showed Tat mainly being in the cytoplasm, and almost excluded from the nucleus, but in the same cells the pattern is different when looking at the fluorescence from the EGFP moiety (figures 4.18.F, 4.18.G, 4.19.F, and 4.19.G). Since the localization of EGFP is naturally in the nucleus and the cytoplasm (figure 4.17), and differs from that of the EGFP-Tat proteins, the intracellular localization of the EGFP-Tat proteins is due to properties of the Tat proteins.

The localization of Tat115 could only be detected through its EGFP moiety, and the images show that it differs slightly from that of the others. It was also located mainly in the nucleus and the nucleoli and very little in the cytoplasm. But when compared to the other Tat proteins in this study it would seem that it is possibly not as concentrated in the nucleoli.

5 Discussion

The Tat protein of HIV-1 has been extensively studied and a plethora of attributes has been assigned to it, and to various parts of it. Yet the last part of the second exon encoded region has been ignored. This is because the first subtypes of HIV-1 to be studied in detail seem to have acquired a point mutation in the 87th codon, inserting a stop there (Jeang et al., 1999). However, most wild type variants of the HIV-1 Tat protein have 101 amino acids.

In this study genes for two variants of the HIV-1 protein Tat were constructed by PCR based methods. These genes were called *tat101* and *tat115*. Following construction of these two different *tat* genes, they were inserted into vectors allowing study of the intracellular localization of both proteins and purification of the Tat101 protein. When the Tat101 protein was purified several subspecies of the protein were detected on Western blot. This was thought to indicate that the smaller subspecies had lost C-terminal amino acids. Valvatne *et al.* had previously epitope mapped some monoclonal anti Tat antibodies (1996), and the detected Tat subspecies were attempted used to further map the epitopes to which these Anti-Tat mAbs bind. The trans-activating function of both Tat proteins was verified, and the trans-cellular activity of the purified Tat101 protein was demonstrated in this work. Computer analyses were performed for the two proteins, and a new domain potentially important for the function of the protein, as well as a possible novel NLS for Tat101 were identified.

5.1 Computer based predictions of Tat protein properties

Bioinformatical analyses of the two different Tat proteins were conducted. Based on these analyses a new domain was identified and named the EK domain. Furthermore, a potential novel NLS was found for Tat101. Comparison of the two Tat proteins established that several domains of Tat101 had their counterparts in Tat115 and that one domain was absent in the Tat115 domain, indicating which domains might be essential for Tat function during the HIV life cycle.

Pairwise blast used to align Tat101 with Tat115 showed that amino acids 85-94 of Tat115 and 86-95 of Tat101 had a strikingly similar motif, E(S,Q)(E,K)EKKVE(R,K)ET. This novel domain was named the EK domain to reflect the high content of E and K. Evaluation of the sequences of Tat variants present in the Los Alamos catalogue of HIV-1 sequences (section A.1.3) indicates that EK is a conserved domain. Motifs similar to the this domain are found in several different proteins, including RNA binding proteins (section A.2.10). This domain can be part of a larger polar area of the protein. However, all the three reading frames in this part

of the HIV-1 genome are used, as part of the *tat*, *rev* and *env* genes. If any one of these sequences has an important function, less genetic variation would be found. The nucleotide sequence could also be conserved if the RNA sequence forms a secondary structure, or the sequence in itself has a function that is important in the HIV life cycle. Examples of this would be TAR, exon splicing enhancers, or exon splicing silencers. However, large parts of the sequence of the *tat* gene have not been highly conserved. For instance it has been shown that an important RNA element called exon splicing silencer 2 (ESS2) located in the first coding exon of the *tat* gene in strains from HIV-1 group M (these nucleotides encode amino acids 6-9 of Tat) is not found in strains from group O (Bilodeau et al., 1999).

Further investigations to determine whether the corresponding amino acid sequences of Rev and Env, and to which extent the nucleotide sequence has been conserved, should be performed. If these sequences have been conserved this could indicate that the EK domain might just be an artefact with no important function. If these investigations turn out negative, it would be prudent to compare the functions of the other proteins that have EK domains.

Unlike the Tat101 protein, Tat115 has no glutamine rich region. This would indicate that the specific amino acid sequence, particularly the high frequency of glutamines, of this region is not of critical importance with regards to the HIV life cycle. This assumption is corroborated by the fact that no essential functions have yet been designated to this region.

The RGD sequence found has previously been suggested to confer binding to integrins on the cell surface, and it was long thought to be important for the transcellular properties of Tat (Valvatne *et al.,* 1996). However, this motif is not highly conserved in Tat (section A.1.3) indicating that it is not important for Tat function.

In Tat 101, the sequence PKESKKK, located at amino acids 84-90, was predicted to be an NLS. This sequence is very similar to a motif from the ribosomal L5 protein that confers nuclear and nucleolar localization (Rosorius *et al.,* 2000). This sequence is well conserved among different sequenced HIV genomes (http://hiv-web.lanl.gov, section A.1.3) indicating some biological importance. Why a protein would need to have two NLS is puzzling, if it indeed is a NLS. Experiments conducted in order to verify the predicted function of this amino acid sequence has not led to conclusive results (Andresen, personal communication). The two NLS are located in separate exons; therefore the one-exon variant of Tat (Tat72), which is expressed as an HIV late protein, contains only one NLS. That the two-exon variant

of Tat is much more rare in the later stages of the HIV life cycle might reflect a need to shut off the activities of the second exon, possibly attributes from the NLS and the EK domain. Both this potential NLS and the EK domain are absent in the Tat86 protein. These motifs may have been detrimental to some HIV strains. Possibly this has caused the mutation that led to an 86 amino acid Tat protein, that is encoded by several strains of HIV from group M subtypes B and D and H.

The PSORT II program reports the NLS of Tat115 to possibly be bipartite (Appendix, section A.2.5, Results figure 4.1) whether it is, should be an important question to address.

The predicted phosphorylation sites differ between the two Tat proteins. Péloponèse *et al.* have suggested that differences in phosphorylation might influence the virulence of different HIV strains. The electric charge of the phosphate groups could influence interactions with other proteins and cellular factors important to its functions. The two proteins might have some different functions since the phosphorylation sites have different locations, or they accomplish the same functions through different structures.

Secondary structure computer analysis predicts that Tat101 and Tat115 have similar and mostly coiled structure. The two proteins have similar amino acid sequence from 1-54, and the regions predicted to have secondary structure up to that are the same. After amino acid 54 the only secondary structure predicted lies in the EK domain, but in the two Tat variants this domain is predicted to have different structures. In Tat115 seven of these amino acids are predicted to form a β strand structure, while in Tat101 all ten amino acids are predicted to form a helical structure. However, in both proteins the predicted structure has a low confidence value for this domain (section A.2.3). The predicted structure does not correlate too well with the structure found by Bayer *et al.* in 1995. Using NMR they found that a Tat86 from group M subtype D strain did not have any secondary structure at all.

5.2 Construction of the tat *genes*

The *tat101* and *tat115* genes were constructed through PCR reactions (section 4.2.1). This was done in order to study variants of the HIV-1 Tat protein previously not characterized, which could lead to new insights relating to the role of Tat in the HIV-1 life cycle. The sequence of *tat* from HXB3 was used as a basis for the sequence of *tat101*. In the *tat101* sequence the stop codon found at codon 87 of HXB3 has been replaced by a codon for serine, which is a common amino acid in that position for other group M subtype B sequences. This

strategy required only one nucleotide substitution, indicating that this could be the amino acid originally present prior to the mutation that caused the protein to be truncated. The *tat115* gene was constructed based on the sequence from strain MVP5180 of the HIV-1 group O.

It turned out, however, that the sequence of *tat115* had a point mutation compared to the original sequence. This was due to an error in the design of the oligonucleotides used to construct the gene. The mutation lay in the primer named Tatmvp6 (section 2.8). The mutation that occurred in codon 97 was considered to be of less importance as the amino acid substitution it caused was from a serine to a threonine, and these two amino acids have similar chemical properties.

5.3 Purification of Tat proteins

The procedure used to purify the Tat proteins was successful for the Tat101 protein, and this protein was used in the further experiments. However, only a very small amount of the Tat115 protein was purified, and this amount was considered too small to be used in the transcellular activity assay.

The his-tagged protein was released from the column using a buffer with low pH. This change of pH might cause a change in the protein structure, denaturing the protein. To minimize any degradation of the protein, 1 M Tris-Cl pH 8.0 was added to increase the pH of the samples containing the Tat protein immediately following purification.

There were two clear bands in the Western blots of the purified Tat101 protein (figure 4.13). This indicated that despite the several protease inhibitors used in the lysis buffer, some degradation of the protein had occurred. It was thought to be due to proteolytic cleavage near the C-terminus because the protein would have been washed away with the flowthrough if the his-tag located very close to the N terminus had been cut off. That neither of the two bands were visible in the lane where bacterial lysate from bacteria transformed with pRSET-C had been applied indicated that both of these bands were Tat products.

One could hypothesize that the low yield of the Tat115 protein was due to problems with the codon usage between *E.coli* and HIV. However, the bacteria used, *E. coli* BL21 CodonPlus (DE3)-RIL, are designed to express proteins with human codon usage. Still, it is possible that a different strain of bacteria could have given a higher yield of protein. Another possible variation would be differerent growth rates, which can affect the yield of protein. This was not

attempted in this work, but several alterations were done in the procedure in unsuccessful attempts to express Tat115.

An explanation for the low yield of Tat115 protein could be the formation of inclusion bodies. An attempt to purify more protein by dissolving the bacteria in a buffer containing 8 M urea did yield more protein. This suggests that Tat115 was expressed. These samples were not studied further. Other variations of the procedure included altering the temperature in which the bacteria were incubated, changing the concentration of IPTG and altering the incubation time. However, the protein could never be detected against the background in the bacterial lysates by Coomassie blue staining of SDS-PAGE. Because of that, these data were not included, nor were they followed up. However, as Western blot is a more sensitive way of detecting protein, this procedure could perhaps have shown variations that Coomassie blue staining of SDS-PAGE could not. In future attempts to purify Tat115, such tests should be performed in order to see if variations of procedure to express this protein have any effect.

5.4 Attempted epitope mapping of antibodies

In the Western blot in figure 4.13 it can be seen that there are two bands in the lanes where the Tat protein has been purified. Neither of these two bands is present in the lane where lysate of the bacteria transformed with the empty pRSET-C vector have been applied. This means that both bands occured as a result of the expression of Tat in the bacteria, and that they likely are both the Tat protein, and not the result of either antibody (1D9D5 and the anti-mouse antibody) binding to a bacterial protein. This must mean that a fraction of the Tat protein in some way had been modified so that it migrated faster through the gel. This could be explained by the net charge of the protein having become more negative through chemical modifications, or that the physical size of the protein had been decreased, either by loss of amino acid residues through proteolytic cleavage, or through different folding of the protein.

The two subspecies of Tat on the Western blot (figure 4.13) were assumed to be caused by one of them having been proteolytically cleaved near the C-terminus. This phenomenon was then used to investigate some mAbs for Tat that had been mapped to the area 49-86 (Valvatne *et al.*, 1996). The antibodies are thought to be dependent on the folded conformation of the protein to be able to bind. On the Western blot (figure 4.14) two clear bands were visible in each lane, regardless of which anti-Tat antibody was used. However, a third band was visible in all lanes but one. That the same band was visible in lanes where antibodies binding to different epitopes were used, excluded the possibility of this band being the result of the

antibodies binding to a bacterial protein found in the lysate. This indicated that this third band either was a third Tat subspecies present, one in which more of the protein had been cleaved off, or that this band was a result of unspecific binding by the secondary antibody. That it was not present in all lanes excluded the possibility of the latter. The presence of this band in the 1D9D5 lane, as well as in the anti-his lane when overexposed, indicated that the amino acids lost were C-terminal. Since this band was not seen in the lane where the antibody 2D9E7 had been applied this indicated that this mAb binds to an epitope closer to the C terminus than the others.

However, when the sizes of these bands were estimated, the uppermost of the three bands showed approximately the estimated size of the 6Xhis-Tat101 protein, while the two smaller bands seemed to be too small to contain the epitope to which the mAbs bind. Valvatne *et al.* showed these mAbs to bind a peptide encompassing amino acids 49-86 of Tat, and hypothesized that the entire peptide was needed for the correct folding of the epitope they bind to. This was because the mAbs did not bind to overlapping 20-mer peptides covering this region, and did not bind to a full-length inactive recombinant Tat protein. Since these bands seem too small to contain the epitope, the hypothesis that the bands on the Western blot are due to proteolytic cleavage of the 6Xhis-Tat101 protein does not agree with the results of Valvatne *et al*. (1996). Then it becomes more likely that the faster migration through the gel of these bands is due to chemical modification of the Tat protein seen in these bands, giving them a more negative charge. This could be due to deamidation of glutamines and asparagines, by being exposed to high temperature during the boiling of the samples in the SDS-PAGE procedure, or there could be differences in the charge of the protein species due to covalently attached non-protein moieties, like phosphate.

5.5 Cellular uptake of Tat and transactivation

By detecting the expression of the reporter protein CAT on Western blot, it was shown that the EGFP-Tat fusion proteins transactivate the LTR promoter. This means that the EGFP moiety of the fusion proteins does not have an effect on the conformation of Tat in any manner that would affect its transactivating activity.

The LTR used was from HXB3. Tat115 is from the MVP5180 strain, from HIV-1 group O, and is a somewhat distant relative. As the *tat* gene evolved in MVP5180 as did the TAR element. Any mutations in either Tat or TAR that decreases binding affinity between them would be responded with a mutation in the other adjusting it, for the virus to compete in its

host. Tat protein from HIV-1 subtype B has been shown to have different transactivating properties on LTRs from different subtypes within group M (Montano et al., 1997; Jeeninga et al., 2000). One would expect from this that the Tat115 protein might not efficiently transactivate this promoter, however it did, and seemingly slightly more efficiently than the EGFP-Tat86 protein.

The purified 6Xhis-Tat101 protein was also shown to transactivate the LTR promoter in these cells when added to the media, demonstrating that it was taken up by the cells. The fact that there was a CAT band in the lane with cells where his-tagged Tat protein had been added, shows that this protein was internalised into the cells and that it managed to transactivate the promoter. This indicates that the additional 15 amino acids, 87-101, have no clear detrimental effect on the cellular uptake of Tat.

As a positive control cells co-transfected by the EGFP-Tat101 plasmid and the LTR-CAT plasmid were used. Lane 8 in figure 4.15 shows a much stronger CAT band. It should be noted that Vibeke Andresen provided this sample and that the more efficient FuGene procedure had been used for transfection. In the lane showing the lysates of cells transfected with the LTR-CAT plasmid alone there is no clear band around 25 kD, one such can be seen very clearly in the positive control. The his-tagged Tat86 protein was used as a positive control, but does not seem to have been functional, indicating that something could have been wrong with the sample.

5.6 Intracellular localization of the different Tat variants

By fluorescence microscopy the EGFP-Tat101 and EGFP-Tat115 proteins were shown to be located in the nucleus and the nucleolus of COS-7 cells (figures 4.16 and 4.18). These results are likely to show the *in vivo* location of Tat, seeing how the COS-7 cells used are primate cells, and should contain a minimum of cellular factors not found in humans that would interfere with the intracellular localization. The results are also in accordance with what has previously been reported for the intracellular localization of the Tat86 protein (Stauber and Pavlakis, 1998). The Tat101 and Tat115 proteins were not shown to have differences in the cellular localization compared to the 72 and 86 amino acid versions of the HXB3 strain.

Stauber and Pavlakis have suggested that the nucleolar localization of Tat is due to overexpression of Tat in the cells (1998), and this hypothesis seems to agree with the results in this study. The cells that expressed Tat in moderate amounts did not have such a distinct

accumulation of the Tat protein in the nucleoli. When the Tat proteins were examined by observing the fluorescence from EGFP, they all showed mainly nuclear, and in particular nucleolar, localization. The images taken with the camera could not confer that a weak shimmer of fluorescence can be seen in the cytoplasm. The Tat115 protein could not be detected when the Tat proteins were inspected by immunofluorescence. This was most likely due to differences in the amino acid sequence of the epitope the monoclonal antibody that was used, 1D9D5, binds to. This epitope has been mapped to amino acids 2-21 of Tat (Valvatne *et al.*, 1996). And there are some differences in the sequence of this region between Tat101, which has a sequence similar to the protein that was used when the antibody was made; and Tat115 (figure 4.1).

When the intracellular distribution of the Tat proteins was examined by immunofluorescence, there were some discrepancies. In some cells the protein was detected only in the cytoplasm by immunofluorescence, while the fluorescence of the EGFP moiety indicated nuclear and nucleolar localization. There presence of EGFP in the nucleus should exclude the possibility of Tat not being there. This indicates that possibly something hinders the antibodies, one or both, from reaching the nucleus. This could be caused by the paraformaldehyde fixation causing chemical crosslinking within the cells (Anne Marie Szilvay, personal communication).

Methanol fixation have been reported to cause an artificial nuclear localization of some proteins, and one should possibly have studied the localization of Tat in live cells, as suggested by Richard *et al.* (2002). However, Stauber and Pavlakis do show that Tat-86-EGFP does localize in the nucleus and the nucleolus both in living and in fixed cells (1998). This indicates that the localization of the Tat protein is undisturbed by the fixation process. However, any final conclusion regarding the intracellular distribution of the Tat101 and Tat115 proteins based on this should not be drawn.

5.7 Future perspectives

Several experiments could be conducted to see whether the results of the bioinformatical analyses accurately have predicted the properties of the Tat101 and Tat115 proteins. The constructed genes allow further studies of these proteins. With a Tat protein with a complete exon 2 available, it is be possible to conduct co-localization studies of the Tat72 and Tat101 proteins, in order to get a final clarification on the role of the second exon-encoded parts of

the protein on its intracellular distribution. The *tat* genes could also be used in the yeast two hybrid system to investigate which proteins in a library the Tat proteins interact with.

There should be further analyses of what happens during expression of the 6Xhis-Tat101 protein that lead to several subspecies of the protein. Further attempts to purify the Tat115 protein in order to determine whether it has trans-cellular activity should also be performed. The expression of this protein could be done in a different system.

The constructed EGFP-Tat vectors can be further used in co-localization studies with other proteins. Or they could be used in assays determining their shuttling capabilities, like the heterokaryon assay used by Stauber and Pavlakis (1998), or the microinjection assay used by Kalland *et al.* (1994).

References

Amara, R. R., and H. L. Robinson. 2002. A new generation of HIV vaccines. Trends Mol. Med. **8:**489-495.

Amersham Pharmacia Biotech. Affinity Chromatography – Principles and Methods. Code No. 18-1022-29.

Bartz, S. R., and M. Emerman. 1999. Human Immunodeficiency Virus Type 1 Tat induces Apoptosis and Increases Sensitivity to Apoptotic Signals by Up-Regulating FLICE/Caspase-8. J. Virol. **73:**1956-1963.

Bayer, P., M. Kraft, A. Ejchart, M. Westendorp, R. Frank, and P. Rösch. 1995. Structural Studies of HIV-1 Tat Protein. J. Mol. Biol. **247:**529-535.

Benelli, R., A. Barbero, S. Ferrini, P. Scapini, M. Cassatella, F. Bussolini, C. Tacchetti, D. M. Noonan, and A. Albini. 2000. Human Immunodeficiency Virus Transactivator Protein (Tat) Stimulates Chemotaxis, Calcium Mobilization, and Activation of Human Polymorphonuclear Leukocytes: Implications for Tat-Mediated Pathogenesis. J. Inf. Dis. **182:**1643-51

Benkirane, M., R. F. Chun, H. Xiao, V. V. Ogryzko, B. H. Howard, Y. Nakatani, and K.-T. Jeang. 1998. Activation of Integrated Provirus Requires Histone Acetyltransferase. J. Biol. Chem. **273:**24898-24905.

Bilodeau, P. S., J. K. Domsic, and M. Stoltzfus. 1999. Splicing Regulatory Elements within *tat* Exon 2 of Human Immunodeficiency Virus Type 1 (HIV-1) Are Characteristic of Group M but Not Group O HIV-1 Strains. J. Virol. **73:**9764-9772.

Blattner, W. A.. Human Retroviruses: Their role in Cancer. 1999. Proc. Ass. Am. Phys.. **111:**563-572.

Col, E., B. Gilquin, C. Caron, and S. Khochbin. 2002. Tat-controlled Protein Acetylation. J. Biol. Chem. **277:**37955-37960.

Conant, K., A. Garzino-Demo, A. Nath, J. C. McArthur, W. Halliday, C. Power, R. C. Gallo, and E. O. Major. 1998. Induction of monocyte chemoattractant protein-1 in HIV-1 Tat-stimulated astrocytes and elevation in AIDS dementia. Proc. Natl. Acad. Sci USA. **95:**3117-3121.

Dang, C. V., and W. M. Lee. 1989. Nuclear and Nucleolar Targeting Sequences of c-*erb*-A, c-*myb,* N-*myc*, p53, HSP70, and HIV *tat* Proteins. J. Biol. Chem. **264:**18019-18023.

Dillon, P. J., and C. A. Rosen. 1990. A Rapid Method for the Construction of Synthetic Genes Using the Polymerase Chain Reaction. Biotechniques, **9:**298-300.

Ensoli, B., L. Buonaguro, G. Barillari, V. Fiorelli, R. Gendelman, R. A. Morgan, P. Wingfield, and R. C. Gallo. 1993. Release, Uptake, and Effects of Extracellular Human Immunodeficiency Virus Type 1 Tat Protein on Cell Growth and Viral Transactivation. J. Virol. **67:**277-287.

Frankel, A. D., J. A. T. Young. 1998. HIV-1: Fifteen Proteins and an RNA. Annu. Rev. Biochem. **67:**1-25.

Gougeon, M.-L., and L. Montagnier. 1993. Apoptosis in AIDS. Science **260:**1269-1270.

Hahn, B. H., G. M. Shaw, K. M. De Cock, P. M. Sharp. 2000. AIDS as a Zoonosis: Scientific and Public Health Implications. Science **287:**607-614.

Helland, D. E., J. L. Welles, A. Caputo, and W. A. Haseltine. 1991. Transcellular Transactivation by the Human Immunodeficiency Virus Type 1 *tat* Protein. J. Virol. **65:**4547-4549.

Howcroft, T. K., K. Strebel, M. A. Martin, D. S. Singer. 1993. Repression of MHC Class I Gene Promoter Activity by Two-Exon Tat of HIV. Science **260:**1320-1322.

Jeang, K.-T., H. Xiao, and E. A. Rich. 1999. Multifaceted Activities of the HIV-1 Transactivator of Transcription, Tat. J. Biol. Chem. **274:**28837-28840.

Jeeninga, R. E., M. Hoogenkamp, M. Armand-Ugon, M. De Baar, K. Verhoef, and B. Berkhout. 2000. Functional Differences between the Long Terminal Repeat Transcriptional Promoters of Human Immunodeficiency Virus Type 1 Subtypes A through G. J. Virol. **74:**3740-3751.

Kalland, K.-H., A. M. Szilvay, K. A. Brokstad, W. Sætrevik, and G. Haukanes. 1994. The Human Immunodeficiency Virus Type 1 Rev Protein Shuttles between the Cytoplasm and Nuclear Compartments. Mol. Cell. Biol. **14:**7436-7444.

Karn, J. 1999. Tackling Tat. J. Mol. Biol. **293:**235-254.

Keen, N. J., M. J. Churcher, and J. Karn. 1997. Transfer of Tat and release of TAR RNA during the activation of the human immunodeficiency virus type-1 transcription elongation complex. EMBO J. **16:**5260-5272.

Kiernan, R. E., C. Vanhulle, L. Schiltz, E. Adam, H. Xiao, F. Maudoux, C. Calomme, A. Burny, Y. Nakatani, K.-T. Jeang, M. Benkirane, and C. Van Lint. 1999. HIV-1 Tat transcriptional activity is regulated by acetylation. EMBO J. **18:**6106-6118.

Kleivbo, S. M. 2001. Sequence analysis of Tat and V3 from R5 and X4 type HIV-1. Cand. Scient thesis, UiB.

Kobor, M. S., J. Greenblatt. 2002. Regulation of transcription elongation by phosphorylation. Biochim. Biophys. Acta **1577:**261-275.

Korber, B., M. Muldoon, J.T heiler, F. Gao, R. Gupta, A. Lapedes, B. H. Hahn, S. Wolinsky, and T. Bhattacharya. 2000. Timing the Ancestor of the HIV-1 Pandemic Strains. Science **288 :**1789-1796.

Louie, J. K., L. C. Hsu, D. H. Osmond, M. H. Katz, and S. K. Schwarcz. 2002. Trends in Causes of Death among Persons with Acquired Immunodeficiency Syndrome in the Era of Highly Active Antiretroviral Therapy, San Francisco, 1994-1998. J. Inf. Dis. **186:**1023-1027.

Mann, D. A., A. D. Frankel. Endocytosis and targeting of exogenous HIV-1 Tat protein. 1991. EMBO J. **10:**1733-1739

Mendonca, N. F. F. de, and N. M. P. de B. Lavoura. 1998. Expression, Purification and transcellular *trans*-activation properties of the recombinant 6Xhis-Tat protein. Report, UiB.

Montano, M. A., V. Novitsky, J. T. Blackard, N. L. Cho, D. A. Katzenstein, and M. Essex. 1997. Divergent Transcriptional Regulation among Expanding Human Immunodeficiency Virus Type 1 Subtypes. J. Virol. **71:**8657-8665.

Nath, A., K. Conant, P. Chen, C. Scott, and E. O. Major. 1999. Transient Exposure to HIV-1 Tat Protein Results in Cytokine Production in Macrophages and Astrocytes. J. Biol. Chem. **274:**17098-17102.

Neuveut, C., and K.-T. Jeang. 1996. Recombinant Human Immunodeficiency Virus Type 1 Genomes with *tat* unconstrained by Overlapping Reading Frames Reveal Residues in Tat Important for Replication in Tissue Culture. J. Virol. **70:**5572-5581.

Ott, D. E., L. V. Coren, D. G. Johnson, B. P. Kane, R. C. Sowder II, Y. D. Kim, R. J. Fisher, X. Z. Zhou, K. P. Lu and L. E. Henderson. 2000. Actin-Binding Cellular Proteins inside Human Immunodeficiency Virus Type 1. Virology **266:**42-51.

Ott, M., S. Emiliani, C. Van Lint, G. Herbein, J. Lovett, N. Chirmule, T. McCloskey, S. Pahwa, E. Verdin. 1997. Immune Hyperactivation of HIV-1 Infected T Cells Mediated by Tat and the CD28 Pathway. Science **275:**1481-1485.

Péloponèse Jr., J.-M., Y. Collette, C. Grégoire, C. Bailly, D. Campèse, E. F. Meurs, D. Olive, and E. P. Loret. 1999. Full Peptide Synthesis, Purification, and Characterization of Six Tat Variants. J. Biol. Chem. **274:**11473-11478.

Ping, Y.-H., and T. M. Rana. 2001. DSIF and NELF Interact with RNA Polymerase II Elongation Complex and HIV-1 Tat Stimulates P-TEFb-mediated Phosphorylation of RNA Polymerase II and DSIF during Transcription Elongation. J. Biol. Chem. **276:**12951-12958.

Piot, P., M. Bartos, P. D. Ghys, N. Walker, and B. Schwartländer. 2001. The global impact of HIV/AIDS. Nature **410:**968-973.

Preston, B. D., and J. P. Dougherty. 1996. Mechanisms of retroviral mutation. Trends Microbiol. **4:**16-21.

Price, D. H. 2000. P-TEFb, a Cyclin-Dependent Kinase Controlling Elongation by RNA Polymerase II. Mol. Cell. Biol. **20:**2629-2634.

Prodromou, C., and L. H. Pearl. 1992. Recursive PCR: a novel technique for total gene synthesis. *Prot. Engineering,* **5:**827-829.

Purcell, D. F. J., and M. A. Martin. 1993. Alternative Splicing of Human Immunodeficiency Virus Type 1 mRNA Modulates Viral Protein Expression, Replication, and Infectivity. J. Virol. **67:**6365-6378.

Reeves, J. D., and R. W. Doms. 2002. Human immunodeficiency virus type 2. J. Gen. Virol. Online: http://www.socgenmicrobiol.org.uk/JGVDirect/18253/18253ft.htm

Richard, J. P., K. Melikov, E. Vives, C. Ramos, B. Verbeure, M. J. Gait, L.V. Chernomordik, and B. Lebleu. 2002. Cell-penetrating peptides: a re-evaluation of the mechanism of cellular uptake. J. Biol. Chem. Manuscript M209548200.

Rosorius, O., B. Fries, R. H. Stauber, N. Hirschmann, D. Bevec, and J. Hauber. 2000. Human Ribosomal Protein L5 Contains Defined Nuclear Localization and Export Signals. J. Biol. Chem. **275:**12061-12068.

Rubartelli, A., A. Poggi, R. Sitia and M. R. Zocchi. 1998. HIV-1 Tat: a polypeptide for all seasons. Immunol. Today **19:**543-545.

Rusnati, M., G. Taraboletti, C. Urbinati, G. Tulipano, R. Giuliani, M. P. Molinari-Tosatti, B. Sennino, M. Giacca, M. Tyagi, A. Albini, D. Noonan, R. Giavazzi, and M. Presta. 2000. Thrombospondin-1/HIV-1 Tat protein interaction: modulation of the biological activity of extracellular Tat. FASEB J. **14:1**917-1930.

Sandhu, G. S., R. A. Aleff, and B. C. Kline. 1992. Dual Asymmetric PCR: One-Step Construction of Synthetic Genes. Biotechniques, **12:**14-16.

Staalesen, V. 1999. Cellular uptake and biological activities of purified HIV-1 Tat-exon 1 protein. Cand. Scient thesis, UiB.

Stauber, R. H., and G. N. Pavlakis. 1998. Intracellular Trafficking and Interactions of the HIV-1 Tat Protein. Virology **252:**126-136.

Stemmer, W. P. C., A. Crameri, K. D. Ha, T. M. Brennan, and H. L. Heyneker. 1995. Single-Step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. Gene **164:**49-53.

Szilvay, A. M., K. A. Brokstad, R. Kopperud, G. Haukenes, and K.-H. Kalland. 1995. Nuclear Export of the Human Immunodeficiency Virus Type 1 Nucleocytoplasmic Shuttle Protein Rev is Mediated by Its Activation Domain and Is Blocked by Transdominant Negative Mutants. J. Virol. **69:**3315-3323.

Tirelli, U., D. Bernardi, M. Spina, E. Vaccher. 2002. AIDS-related tumors: integrating antiviral and anticancer therapy. Crit. Rev. Onc. Hem. **41:** 299-315.

Tyagi, M., M. Rusnati, M. Presta, and M. Giacca. 2001. Internalization of HIV-1 Tat requires Cell Surface Heparan Sulfate Proteoglycans. J. Biol. Chem. **276:**3254-3261.

Valvatne, H., A. M. Szilvay, and D. E. Helland. 1996. A Monoclonal Antibody Defines a Novel HIV Type 1 Tat Domain invoved in *trans*-Cellular *trans*-activation. AIDS Res. Hum. Retroviruses **12:**611-619.

Verhoef, K., A. Klein and B. Berkhout. 1996. Paracrine activation of the HIV-1 LTR Promoter by the Viral Tat Protein is Mechanistically Similar to *Trans*-Activation within a Cell. Virology **225:**316-327.

Verhoef, K., M. Bauer, A. Meyerhans, and B. Berkhout. 1998. On the Role of the Second Coding Exon of the HIV-1 tat Protein in Virus Replication and MHC Class I Downreglation. AIDS Res. Hum. Retroviruses **14:**1553-1559.

Wei, X., J. M. Decker, S. Wang, H. Hui, J. C. Kappes, X. Wu, J. F. Salazar-Gonzalez, M. G. Salazar, J. M. Kilby, M. S. Saag, N. M. Komarova, M. A. Nowak, B. H. Hahn, P. D. Kwong and G. M. Shaw. 2003. Antibody neutralization and escape by HIV-1. Nature **422:**307-312.

Weiss, R. A. 2000. Getting to know HIV. Trop. Med. Int. Health **5:**A10-A15.

Westendorp, M. O., R. Franck, C. Ochsenbauer, K. Stricker, J. Dhein, H. Walczak, K.-M. Debatin, and P. H. Krammer. 1995. Sensitization of T cells to CD95-mediated apoptosis by HIV-1 Tat and gp120. Nature **375:**497-500.

Zhu, T., B. T. Korber, A. J. Nahmias, E. Hooper, P. . Sharp and D. D. Ho. 1998. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. Nature **391:**584-597.

Internet addresses:

Blast from NCBI: http://www.ncbi.nlm.nih.gov/

Folkehelseinstituttet's homepage: www.fhi.no

HIV Medicine homepage: http://hivmedicine.com/index.htm

Los Alamos HIV sequence database: http://hiv-web.lanl.gov

The modified Alkaline lysis/PEG precipitation method of plasmid purification:
http://www.biotech.iastate.edu/Facilities/DSSF/plasmid_prep.html

The Predict Protein tool at the PHD server at Columbia: http://dodo.cpmc.columbia.edu.

The ProtParam tool: http://us.expasy.org/tools/protparam.html

The PSIPRED protein structure prediction server: http://bioinf.cs.ucl.ac.uk/psiform.html

PSORT II: http://psort.nibb.ac.jp/form2.html

ScanProsite: http://us.expasy.org/prosite

World Health Organization homepage: http://www.who.int

UNAIDS homepage: http://www.unaids.org

A Appendix

A.1 Sequence data

A.1.1 *Tat115* sequence data

The gene referred to as *tat115* was based on the *tat* gene from the HIV-1 subtype O strain MVP5180 and has the following DNA sequence:

ATG GAT CCA GTA GAT CCT GAG ATG CCC CCT TGG CAT CAC CCT GGG AGC AAG CCC CAA ACC CCT TGT AAT AAT TGC TAT TGC AAA AGA TGC TGC TAT CAT TGC TAT GTT TGT TTC ACA AAG AAG GGT TTG GGA ATC TCC CAT GGC AGG AAG AAG CGA AGA AGA CCA GCA GCT GCT GCA AGC TAT CCA GAT AAT AAA GAT CCT GTA CCA GAG CAA TCC CTG TCC CAC ACC GGC AGG AAG CAG AAA CGC CAG GAA GAA CAG GAG AAG AAG GTG GAG AAG GAG ACA GGC CCA A**C**T GGA CAG CCT TGC CAC CAG GAT TCT TGC AAC AGT TGT ACA CGG ATC TCA GGA CAA *TAA*

The nucleotide shown in bold represents a deviation from the sequence of the gene as it is found in the MVP5180 strain (a guanine has been replaced by a cytosine). Marked in italics are the start and stop codons.

This renders the following amino acid sequence for the Tat115 protein (one letter amino acid code):

M D P V D P E M P P W H H P G S K P Q T P C N N C Y C K R C C Y H C Y V C F T K K G L G I S H G R K K R R R P A A A A S Y P D N K D P V P E Q S L S H T G R K Q K R Q E E Q E K K V E K E T G P **T** G Q P C H Q D S C N S C T R I S G Q

The amino acid shown in bold marks the mutation that arose due to the nucleotide change mentioned above (serine97 has been replaced by a threonine).

A.1.2 *Tat101* sequence data

The DNA sequence of the *tat101* gene was based on the sequence of the *tat* gene from the HXB3 strain:

ATG GAG CCA GTA GAT CCT AGA CTA GAG CCC TGG AAG CAT CCA GGA AGT CAG CCT AAA ACT GCT TGT ACC AAT TGC TAT TGT AAA AAG TGT TGC TTT CAT TGC CAA GTT TGT TTC ATA ACA AAA GCC TTA GGC ATC TCC TAT GGC AGG AAG AAG CGG AGA CAG CGA CGA AGA CCT CCT CAA GGC AGT CAG ACT CAT CAA GTT TCT CTA TCA AAG CAA CCC ACC TCC CAA TCC CGA GGG GAC CCG ACA GGC CCG AAG GAA T**C**G AAG AAG AAG GTG GAG AGA GAG ACA GAG ACA GAT CCA TTC GAT *TAG*

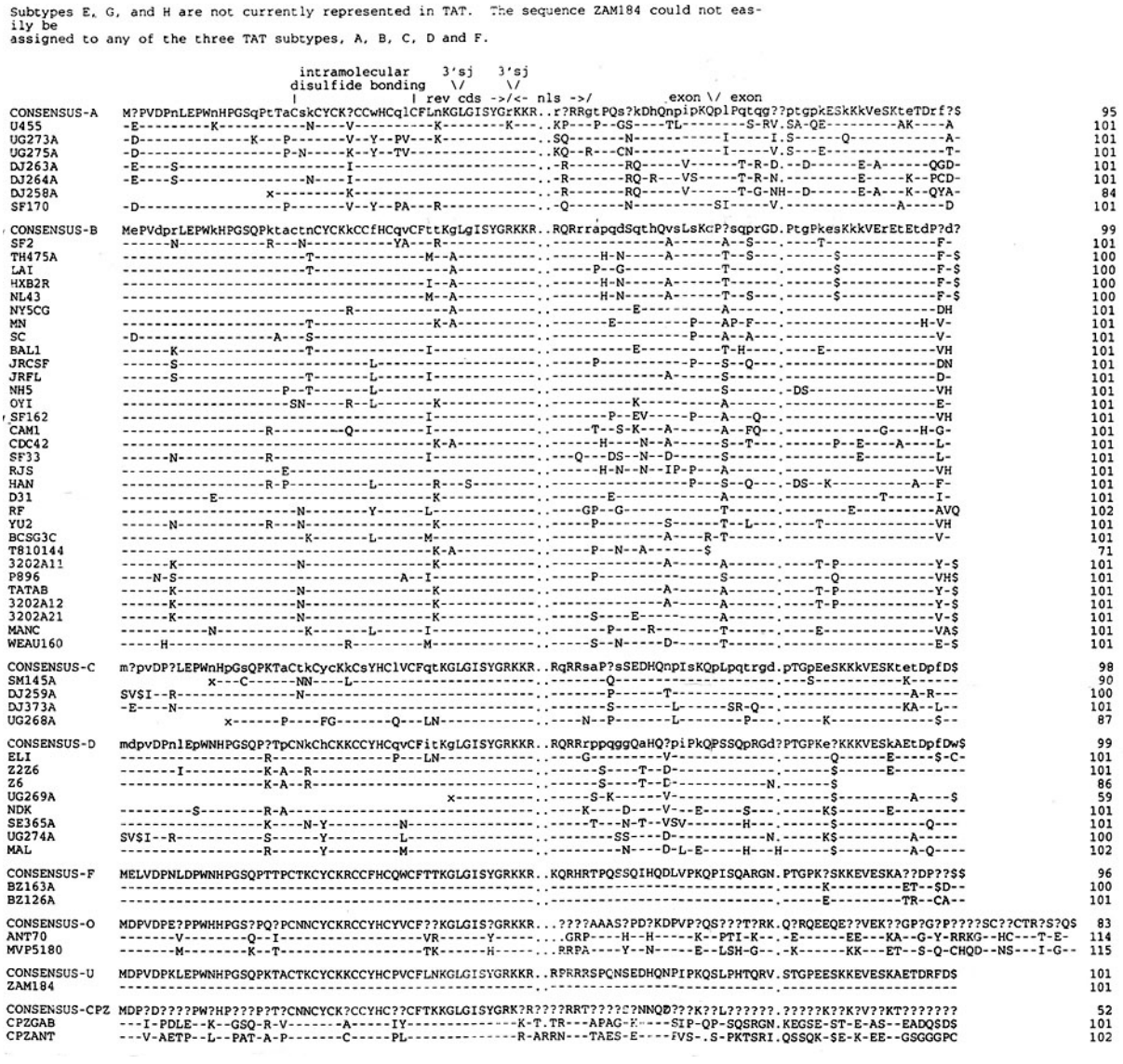
The nucleotide shown in bold represents a deviation from the sequence of the gene as it is found in the HXB3 strain (an adenine has been replaced by a cytosine). Marked in italics are the start and stop codons.

This renders the following amino acid sequence for the Tat101 protein (one letter amino acid code):

M E P V D P R L E P W K H P G S Q P K T A C T N C Y C K K C C F H C Q V C F I T K
A L G I S Y G R K K R R Q R R R P P Q G S Q T H Q V S L S K Q P T S Q S R G D P T G
P K E **S** K K K V E R E T E T D P F D G S T G S R

HXB3 has a stop codon at position 87. In *tat101* one nucleotide is changed from A to C. This means that the stop codon is replaced by a serine, and 14 additional amino acids are encoded by the gene, which is common among other strains from HIV-1 group M.

A.1.3 Amino acid sequences of several types of Tat (from the Los Alamos database)



A.2 Computer results

A.2.1 Protparam results for Tat101

Number of amino acids: 101

Molecular weight: 11575.2

Theoretical pI: 9.64

Amino acid composition:

Ala (A)	2	2.0%
Arg (R)	9	8.9%
Asn (N)	1	1.0%
Asp (D)	4	4.0%
Cys (C)	7	6.9%
Gln (Q)	8	7.9%
Glu (E)	6	5.9%
Gly (G)	6	5.9%
His (H)	3	3.0%
Ile (I)	2	2.0%
Leu (L)	3	3.0%
Lys (K)	12	11.9%
Met (M)	1	1.0%
Phe (F)	3	3.0%
Pro (P)	11	10.9%
Ser (S)	8	7.9%
Thr (T)	8	7.9%
Trp (W)	1	1.0%
Tyr (Y)	2	2.0%
Val (V)	4	4.0%
Asx (B)	0	0.0%
Glx (Z)	0	0.0%
Xaa (X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 10
Total number of positively charged residues (Arg + Lys): 21

Extinction coefficients:

Conditions: 6.0 M guanidium hydrochloride
0.02 M phosphate buffer
pH 6.5

Extinction coefficients are in units of M^{-1} cm^{-1} .

The first table lists values computed assuming ALL Cys residues appear as half cystines, whereas the second table assumes that NONE do.

	276 nm	278 nm	279 nm	280 nm	282 nm
Ext. coefficient	8735	8781	8710	8610	8360
Abs 0.1% (=1 g/l)	0.755	0.759	0.752	0.744	0.722

	276	278	279	280	282

	nm	nm	nm	nm	nm
Ext. coefficient	8300	8400	8350	8250	8000
Abs 0.1% (=1 g/l)	0.717	0.726	0.721	0.713	0.691

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 65.13
This classifies the protein as unstable.

Aliphatic index: 32.77

Grand average of hydropathicity (GRAVY): -1.290

With his-tag:

Number of amino acids: 140

Molecular weight: 15890.1

Theoretical pI: 10.14

Amino acid composition:

Ala (A)	10	7.1%
Arg (R)	14	10.0%
Asn (N)	1	0.7%
Asp (D)	4	2.9%
Cys (C)	7	5.0%
Gln (Q)	10	7.1%
Glu (E)	6	4.3%
Gly (G)	11	7.9%
His (H)	9	6.4%
Ile (I)	3	2.1%
Leu (L)	4	2.9%
Lys (K)	13	9.3%
Met (M)	5	3.6%
Phe (F)	3	2.1%
Pro (P)	12	8.6%
Ser (S)	10	7.1%
Thr (T)	9	6.4%
Trp (W)	2	1.4%
Tyr (Y)	3	2.1%
Val (V)	4	2.9%
Asx (B)	0	0.0%
Glx (Z)	0	0.0%
Xaa (X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 10
Total number of positively charged residues (Arg + Lys): 27

Atomic composition:

Carbon	C	678
Hydrogen	H	1084
Nitrogen	N	226
Oxygen	O	194
Sulfur	S	12

Formula: $C_{678}H_{1084}N_{226}O_{194}S_{12}$
Total number of atoms: 2194

Extinction coefficients:

Conditions: 6.0 M guanidium hydrochloride
0.02 M phosphate buffer
pH 6.5

Extinction coefficients are in units of M^{-1} cm^{-1} .

The first table lists values computed assuming ALL Cys residues appear as half cystines, whereas the second table assumes that NONE do.

	276 nm	278 nm	279 nm	280 nm	282 nm
Ext. coefficient	15585	15781	15715	15580	15160
Abs 0.1% (=1 g/l)	0.981	0.993	0.989	0.980	0.954

User-provided sequence:

```
      1          11         21         31         41         51
      |          |          |          |          |          |
    1 MEPVDPRLEP WKHPGSQPKT ACTNCYCKKC CFHCQVCFIT KALGISYGRK KRRQRRRPPQ
60
   61 GSQTHQVSLS KQPTSQSRGD PTGPKESKKK VERETETDPF D
```

A.2.2 Protparam results for Tat115

Number of amino acids: 115

Molecular weight: 13024.7

Theoretical pI: 9.06

Amino acid composition:

Ala (A)	4	3.5%
Arg (R)	8	7.0%
Asn (N)	4	3.5%
Asp (D)	5	4.3%
Cys (C)	10	8.7%
Gln (Q)	8	7.0%
Glu (E)	7	6.1%
Gly (G)	8	7.0%
His (H)	6	5.2%
Ile (I)	2	1.7%
Leu (L)	2	1.7%
Lys (K)	12	10.4%
Met (M)	2	1.7%

Phe (F)	1	0.9%
Pro (P)	13	11.3%
Ser (S)	8	7.0%
Thr (T)	6	5.2%
Trp (W)	1	0.9%
Tyr (Y)	4	3.5%
Val (V)	4	3.5%
Asx (B)	0	0.0%
Glx (Z)	0	0.0%
Xaa (X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 12
Total number of positively charged residues (Arg + Lys): 20

Extinction coefficients:

Conditions: 6.0 M guanidium hydrochloride
0.02 M phosphate buffer
pH 6.5

Extinction coefficients are in units of M^{-1} cm^{-1} .

The first table lists values computed assuming ALL Cys residues appear as half cystines, whereas the second table assumes that NONE do.

	276 nm	278 nm	279 nm	280 nm	282 nm
Ext. coefficient	11925	11835	11640	11410	11000
Abs 0.1% (=1 g/l)	0.916	0.909	0.894	0.876	0.845

	276 nm	278 nm	279 nm	280 nm	282 nm
Ext. coefficient	11200	11200	11040	10810	10400
Abs 0.1% (=1 g/l)	0.860	0.860	0.848	0.830	0.798

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 75.99
This classifies the protein as unstable.

Aliphatic index: 27.13

Grand average of hydropathicity (GRAVY): -1.343

A.2.3 Psipred results

Key

Conf: Confidence (0=low, 9=high)
Pred: Predicted secondary structure (H=helix, E=strand, C=coil)
AA: Target sequence

Tat101

```
Conf: 988888877543578887788875200123466467774235233147787666687888
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHCCCEECCCCCCCCCCCCCC
  AA: MEPVDPRLEPWKHPGSQPKTACTNCYCKKCCFHCQVCFITKALGISYGRKKRRQRRRPPQ
              10        20        30        40        50        60

Conf: 78888887778888888777667501233333331267889
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHCCCCCC
  AA: GSQTHQVSLSKQPTSQSRGDPTGPKESKKKVERETETDPFD
              70        80        90       100
```

Tat115

```
Conf: 9888888774214688877888753000112452566642252340367876787887 7
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHCCCEEECCCCCCCCCCCCCC
  AA: MDPVDPEMPPWHHPGSKPQTPCNNCYCKRCCYHCYVCFTKKGLGISHGRKKRRRPAAAAS
              10        20        30        40        50        60

Conf: 88776655456887887777685210134211104688886465443223456 89
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCEEEEEEECCCCCCCCCCCCCCCCCCCC
  AA: YPDNKDPVPEQSLSHTGRKQKRQEEQEKKVEKETGPTGQPCHQDSCNSCTRISGQ
              70        80        90       100       110
```

A.2.4 PSORT II results for Tat101

Results of Subprograms

```
PSG:  a new signal peptide prediction method
      N-region:  length 9;  pos.chg 1;  neg.chg 3
      H-region:  length 2;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -4.52
      possible cleavage site: between 44 and 45

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   0
      number of TMS(s) .. fixed
      PERIPHERAL  Likelihood =  2.65 (at 30)
      ALOM score:   2.65  (number of TMSs: 0)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  2.80
      Hyd Moment(95):  7.47    G content:       0
      D/E content:     2       S/T content:     0
      Score: -7.35

Gavel: prediction of cleavage sites for mitochondrial preseq
```

```
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: RKKR (5) at   49
      pat4: KKRR (5) at   50
      pat4: RRRP (4) at   55
      pat7: PKESKKK (3) at   84
      bipartite: none
      content of basic residues:  20.8%
      NLS Score:  1.61

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

SKL2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytplasmic/Nuclear discrimination
      Prediction: nuclear
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues
```

Results of the k-NN Prediction

$k = 9/23$

```
        78.3 %: nuclear
        13.0 %: cytoskeletal
         4.3 %: cytoplasmic
         4.3 %: vesicles of secretory system

>> prediction for QUERY is nuc (k=23)
```

A.2.5 PSORT II results for Tat115

Results of Subprograms

```
PSG:  a new signal peptide prediction method
      N-region:  length 7;  pos.chg 0;  neg.chg 3
      H-region:  length 9;  peak value   0.00
      PSG score:  -4.40

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1): -10.54
      possible cleavage site: between 48 and 49

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   0
      number of TMS(s) .. fixed
      PERIPHERAL  Likelihood =  9.28 (at 30)
      ALOM score:   9.28  (number of TMSs: 0)

MITDISC: discrimination of mitochondrial targeting seq
      R content:       0       Hyd Moment(75):  3.13
      Hyd Moment(95):  7.96    G content:       0
      D/E content:     2       S/T content:     0
      Score: -7.24

Gavel: prediction of cleavage sites for mitochondrial preseq
      cleavage site motif not found

NUCDISC: discrimination of nuclear localization signals
      pat4: RKKR (5) at   49
      pat4: KKRR (5) at   50
      pat4: KRRR (5) at   51
      pat4: RRRP (4) at   52
      pat7: none
      bipartite: KKGLGISHGRKKRRRPA at   40
      content of basic residues:  17.4%
      NLS Score:  1.21

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

SKL2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: none

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none
```

```
Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs:  none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs:  none

NNCN: Reinhardt's method for Cytplasmic/Nuclear discrimination
      Prediction: nuclear
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
      total: 0 residues
```

Results of the* k-*NN Prediction

k = 9/23

```
         69.6 %: nuclear
         13.0 %: cytoskeletal
          4.3 %: Golgi
          4.3 %: plasma membrane
          4.3 %: vesicles of secretory system
          4.3 %: cytoplasmic

>> prediction for QUERY is nuc (k=23)
```

A.2.6 ScanProsite results for Tat101

PROSITE Release 17.22, of 11-Oct-2002

>PDOC00005 PS00005 **PKC_PHOSPHO_SITE** Protein kinase C phosphorylation site [pattern] [Warning: pattern with a high probability of occurrence].

```
       87 - 89  SkK
```

>PDOC00006 PS00006 **CK2_PHOSPHO_SITE** Casein kinase II phosphorylation site [pattern] [Warning: pattern with a high probability of occurrence].

```
       77 - 80  SrgD
       95 - 98  TetD
```

>PDOC00008 PS00008 **MYRISTYL** N-myristoylation site [pattern] [Warning: pattern with a high probability of occurrence].

```
       44 - 49  GIsyGR
```

>PDOC00009 PS00009 **AMIDATION** Amidation site [pattern] [Warning: pattern with a high probability of occurrence].

```
       47 - 50  yGRK
```

>PDOC00016 PS00016 **RGD** Cell attachment sequence [pattern] [Warning: pattern with a high probability of occurrence].

```
       78 - 80  RGD
```

>PDOC50099 PS50311 **CYS_RICH** Cysteine-rich region [profile].

```
       22 - 37  CtncyckkccfhcqvC
```

A.2.7 ScanProsite results for Tat115

PROSITE Release 17.22, of 11-Oct-2002

>PDOC00005 PS00005 **PKC_PHOSPHO_SITE** Protein kinase C phosphorylation site [pattern] [Warning: pattern with a high probability of occurrence].

```
        39 - 41  TkK
        76 - 78  TgR
```

>PDOC00006 PS00006 **CK2_PHOSPHO_SITE** Casein kinase II phosphorylation site [pattern] [Warning: pattern with a high probability of occurrence].

```
        60 - 63  SypD
```

>PDOC00008 PS00008 **MYRISTYL** N-myristoylation site [pattern] [Warning: pattern with a high probability of occurrence].

```
        42 - 47  GLgiSH
        44 - 49  GIshGR
```

>PDOC00009 PS00009 **AMIDATION** Amidation site [pattern] [Warning: pattern with a high probability of occurrence].

```
        47 - 50  hGRK
        76 - 79  tGRK
```

>PDOC00015 PS00015 **NUCLEAR** Bipartite nuclear targeting sequence [rule] [Warning: rule with a high probability of occurrence].

```
        40 - 56  KKglgishgrkkrrrpa
```

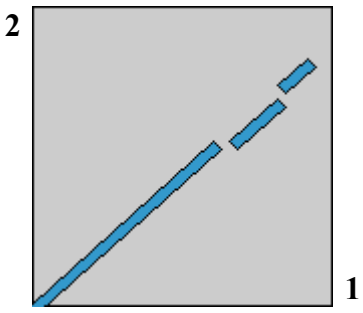
>PDOC50099 PS50311 **CYS_RICH** Cysteine-rich region [profile].

```
        22 - 37  CnncyckrccyhcyvC
```

A.2.8 Blast 2 Sequences (Tat101 and Tat115)

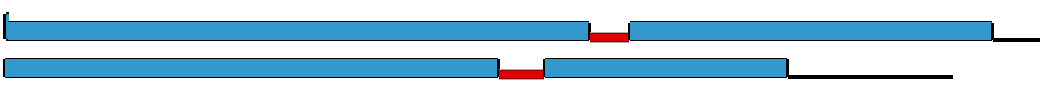
Sequence 1 lcl|seq_1 **Length** 101 (1 .. 101)

Sequence 2 lcl|seq_2 **Length** 115 (1 .. 115)



NOTE:The statistics (bitscore and expect value) is calculated based on the size of nr database

```
Score = 92.8 bits (227), Expect = 6e-19
Identities = 49/100 (49%), Positives = 61/100 (61%), Gaps = 11/100 (11%)
```



```
Query: 1   MEPVDPRLEPWKHPGSQPKTACTNCYCKKCCFHCQVCFITKALGISYXXXXXXXXXXXXX 60
           M+PVDP + PW HPGS+P+T C NCYCK+CC+HC VCF  K LGIS+
Sbjct: 1   MDPVDPEMPPWHHPGSKPQTPCNNCYCKRCCYHCYVCFTKKGLGISHGRKKRRRPAAAAS 60
```

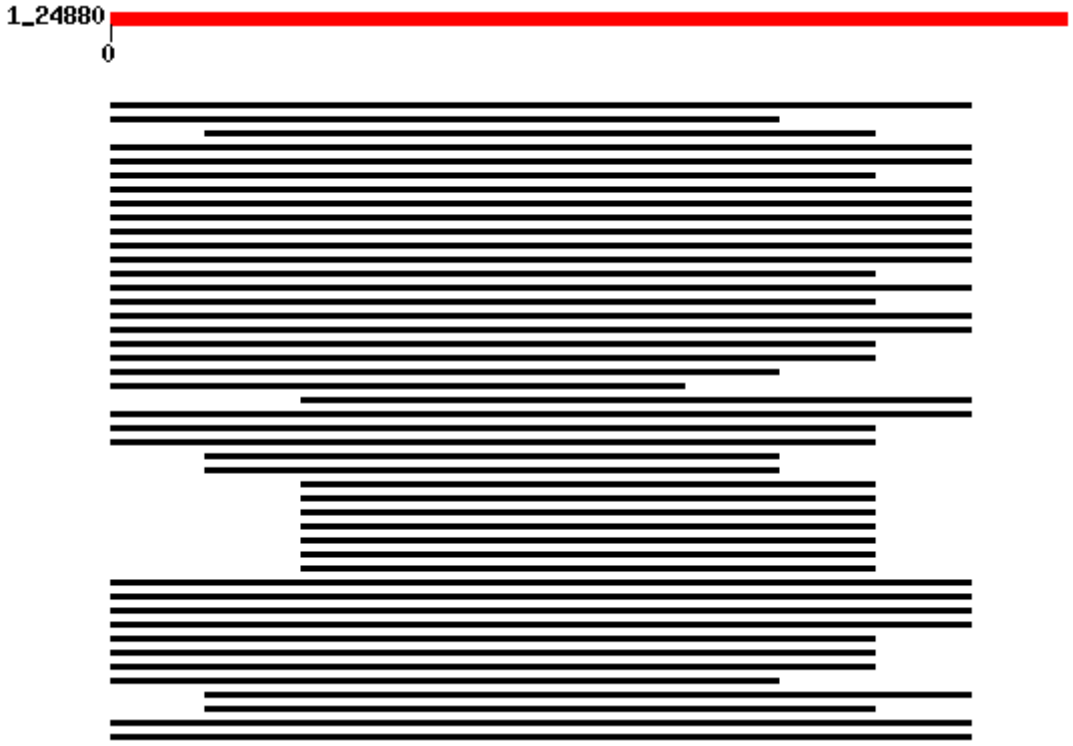
```
Query: 61 XSQTHQVSLSKQPTSQSRGDPTGPK-----ESKKKVERET 95
                  +K P  +     TG K     E +KKVE+ET
Sbjct: 61 YPD------NKDPVPEQSLSHTGRKQKRQEEQEKKVEKET 94
CPU time:     0.08 user secs.      0.01 sys. secs         0.09 total secs.

Gapped
Lambda     K      H
   0.315    0.131    0.429

Gapped
Lambda     K      H
   0.270   0.0470    0.230


Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Hits to DB: 184
Number of Sequences: 0
Number of extensions: 19
Number of successful extensions: 1
Number of sequences better than 10.0: 1
Number of HSP's better than 10.0 without gapping: 1
Number of HSP's successfully gapped in prelim test: 0
Number of HSP's that attempted gapping in prelim test: 0
Number of HSP's gapped (non-prelim): 1
length of query: 101
length of database: 181,542,687
effective HSP length: 46
effective length of query: 55
effective length of database: 98,859,895
effective search space: 5437294225
effective search space used: 5437294225
T: 9
A: 40
X1: 16 ( 7.3 bits)
X2: 128 (49.9 bits)
X3: 128 (49.9 bits)
S1: 41 (21.6 bits)
S2: 64 (29.3 bits)
```

A.2.9 Blast results for the EK domain of Tat115



Sequences producing significant alignments:	Score (bits)	E Value
gi|14041640|emb|CAC38424.1| (AJ302647) tat protein [Human i...	30	2.6
gi|20853427|ref|XP_136788.1| (XM_136788) similar to gag pro...	29	8.3
gi|21402726|ref|NP_658711.1| (NC_003995) hypothetical prote...	29	8.3
gi|23510183|ref|NP_702849.1| (NC_004318) hypothetical prote...	27	20
gi|7243125|dbj|BAA92610.1| (AB037793) KIAA1372 protein [Hom...	27	20
L		
gi|20483065|ref|XP_166244.1| (XM_166244) similar to KIAA137...	27	20
L		
gi|2108169|gb|AAC57424.1| (U89134) tat protein [Simian-Huma...	27	27
gi|9629920|ref|NP_046127.1| (NC_001870) tat [Simian-Human i...	27	27
gi|1055034|gb|AAA81040.1| (U39362) TAT protein [Human immun...	27	27
gi|1469311|gb|AAB05047.1| (U04908) tat protein [Human immun...	27	27
gi|23508435|ref|NP_701104.1| (NC_004315) hypothetical prote...	27	36
gi|16272277|ref|NP_438489.1| (NC_000907) ribonuclease T (rn...	27	36
gi|11497667|ref|NP_068888.1| (NC_000917) hypothetical prote...	27	36
gi|6321540|ref|NP_011617.1| (NC_001139) Nucleolar protein p...	26	49
gi|2497693|sp|Q63805|A1AI_MOUSE Alpha-1-acid glycoprotein 3...	26	49
gi|7490230|pir||T41163 cholinephosphate cytidylyltransferas...	26	65
gi|18376116|emb|CAD21182.1| (AL669998) conserved hypothetic...	26	65
gi|19076048|ref|NP_588548.1| (NC_003421) putative cholineph...	26	65
gi|6503222|gb|AAF14644.1|AF163772_4 (AF163772) 7138.4 [Leis...	26	65
gi|15227342|ref|NP_179292.1| (NM_127253) hypothetical prote...	26	65
gi|17558460|ref|NP_503869.1| (NM_071468) C36C5.12.p [Caenor...	26	65
L		
gi|19386842|dbj|BAB86220.1| (AP003922) contains ESTs AU0939...	26	65
gi|10383785|ref|NP_009943.2| (NC_001135) Cwh43p [Saccharomy...	25	87
gi|23482863|gb|EAA18721.1| (AABL01002201) glutamine-asparag...	25	87
gi|10733345|gb|AAG21972.2|AF257652_1 (AF257652) hexokinase-...	25	87

gi|140464|sp|P25618|YCQ7_YEAST HYPOTHETICAL 107.9 KD PROTEI... 25 87
gi|21593329|gb|AAM65278.1| (AY087741) unknown [Arabidopsis ... 25 87
gi|15238732|ref|NP_200152.1| (NM_124719) putative protein; ... 25 87
gi|20808632|ref|NP_623803.1| (NC_003869) DNA-directed RNA p... 25 87
gi|15241804|ref|NP_198192.1| (NM_122723) putative protein; ... 25 87
gi|23484411|gb|EAA19749.1| (AABL01002670) hypothetical prot... 25 87
gi|23488770|gb|EAA21392.1| (AABL01000545) hypothetical prot... 25 87
gi|23480318|gb|EAA16908.1| (AABL01001547) Drosophila melano... 25 87
gi|15233209|ref|NP_191737.1| (NM_116043) putative protein; ... 25 87
gi|17233362|ref|NP_490452.1| (NC_003276) unknown protein [N... 25 117
gi|21298777|gb|EAA10922.1| (AAAB01008960) agCP6519 [Anophel... 25 117
gi|19114260|ref|NP_593348.1| (NC_003424) putative ATP-depen... 25 117
gi|21038850|emb|CAD31825.1| (AL132948) Y39B6A.30 [Caenorhab... 25 117
gi|7513208|pir||T08880 NMDA receptor-binding protein yotiao... 25 117
gi|17565568|ref|NP_507714.1| (NM_075313) Y39B6A.v.p [Caenor... 25 117
L
gi|22538389|ref|NP_671695.1| (NM_147166) A kinase anchor pr... 25 117
L
gi|4584423|emb|CAB40713.1| (AJ131693) AKAP450 protein [Homo... 25 117
L
gi|20128861|ref|NP_569906.1| (NM_130550) EG:8D8.4 gene prod... 25 117
L
gi|7416518|dbj|BAA93904.1| (AB034447) Tat protein [Human im... 25 117
gi|22538391|ref|NP_671700.1| (NM_147171) A kinase anchor pr... 25 117
L
gi|21038841|emb|CAD31816.1| (AL132948) Y39B6A.18 [Caenorhab... 25 117
gi|23094539|emb|CAD45758.1| (AL766843) Unknown [Streptococc... 25 117
gi|23026768|gb|ZP_00065241.1| (NZ_AAAT01000002) hypothetica... 25 117
gi|22536299|ref|NP_687150.1| (NC_004116) ribose ABC transpo... 25 117
gi|22538393|ref|NP_671714.1| (NM_147185) A kinase anchor pr... 25 117
L
gi|4558862|gb|AAD22767.1|AF083037_1 (AF083037) A-kinase anc... 25 117
L
gi|15792868|ref|NP_282691.1| (NC_002163) putative transcrip... 25 117
gi|3645944|gb|AAC60380.1| (AC000066) yotiao [Homo sapiens] 25 117
L
gi|17539938|ref|NP_501328.1| (NM_068927) F20D12.2.p [Caenor... 25 117
L
gi|5051743|dbj|BAA78718.1| (AB019691) Centrosome- and Golgi... 25 117
L
gi|21483398|gb|AAM52674.1| (AY122162) LD23562p [Drosophila ... 25 117
L
gi|17565546|ref|NP_507706.1| (NM_075305) Y39B6A.n.p [Caenor... 25 117
L
gi|135171|sp|P27058|SYST_LYCES Systemin precursor >gi|74890... 25 117
gi|10834807|gb|AAG23833.1|AF290425_1 (AF290425) NADPH cytoc... 25 117
gi|7489266|pir||T06993 prosystemin PRO1 - potato >gi|291128... 25 117
gi|22538387|ref|NP_005742.4| (NM_005751) A kinase anchor pr... 25 117
L
gi|23490500|gb|EAA22262.1| (AABL01000770) hypothetical prot... 24 157
gi|21281669|ref|NP_060899.1| (NM_018429) transcription fact... 24 157
L
gi|18583964|ref|XP_101952.1| (XM_101952) hypothetical prote... 24 157
L
gi|19113217|ref|NP_596425.1| (NC_003423) hexose transporter... 24 157
gi|18405336|ref|NP_566816.1| (NM_113650) expressed protein;... 24 157
gi|15792002|ref|NP_281825.1| (NC_002163) putative DNA repai... 24 157
gi|11277423|pir||T43533 hexose transport protein Ght2 - fis... 24 157
gi|11096171|gb|AAG30220.1|AF298151_1 (AF298151) RNA polymer... 24 157
L

gi|17564292|ref|NP_507103.1| (NM_074702) Zinc finger, C4 ty... 24 157
L
gi|1352419|sp|P20810|ICAL_HUMAN Calpain inhibitor (Calpasta... 24 157
gi|480379|pir||S36779 ribosome-binding protein p34 - rat >g... 24 157
gi|12056962|gb|AAG48151.1|AF327443_1 (AF327443) calpastatin... 24 157
L
gi|17559086|ref|NP_505097.1| (NM_072696) D1014.5.p [Caenorh... 24 157
L
gi|16553841|dbj|BAB71602.1| (AK057871) unnamed protein prod... 24 157
gi|12545390|gb|AAA52759.2| (M86257) calpastatin [Homo sapiens] 24 157
gi|17530149|gb|AAL40720.1|AF402771_1 (AF402771) calreticuli... 24 157
gi|23478719|gb|EAA15730.1| (AABL01001174) hypothetical prot... 24 157
gi|5053101|gb|AAD38850.1|AF156027_1 (AF156027) calcium/calm... 24 157
gi|21361567|ref|NP_001741.2| (NM_001750) calpastatin [Homo ... 24 157
L
gi|21554138|gb|AAM63218.1| (AY086009) unknown [Arabidopsis ... 24 157
gi|951315|gb|AAB60371.1| (U31345) calpastatin [Homo sapiens] 24 157
gi|5819094|gb|AAC50136.2| (U26724) calpastatin [Homo sapiens] 24 157
gi|23480647|gb|EAA17150.1| (AABL01000162) hypothetical prot... 24 157
gi|180975|gb|AAA52066.1| (M28230) calpastatin [Homo sapiens] 24 157
gi|7305547|ref|NP_038950.1| (NM_013922) zinc finger protein... 24 157
L
gi|226919|prf||1611327A calpastatin [Homo sapiens] 24 157
gi|2754612|dbj|BAA24152.1| (AB010030) outer arm dynein ligh... 24 157
gi|12232589|emb|CAC21448.1| (AJ279120) TFNR [Homo sapiens] 24 157
gi|11096173|gb|AAG30221.1|AF298152_1 (AF298152) RNA polymer... 24 157
L
gi|2134859|pir||A38091 calpastatin, long form - human (frag... 24 157
gi|20095055|ref|NP_614902.1| (NC_003551) DNA-binding protei... 24 211
gi|23594238|ref|XP_177346.1| (XM_177346) similar to Fragile... 24 211
L
gi|23501918|ref|NP_698045.1| (NC_004310) sensory box histid... 24 211
gi|17987230|ref|NP_539864.1| (NC_003317) Sensory Transducti... 24 211
gi|541239|pir||S42875 dihydrolipoamide S-succinyltransferas... 24 211
gi|727184|gb|AAA64261.1| (L40990) tat [Simian immunodeficie... 24 211
gi|23509084|ref|NP_701752.1| (NC_004316) erythrocyte membra... 24 211
gi|23041665|gb|ZP_00073108.1| (NZ_AAAU01000025) hypothetica... 24 211
gi|14041630|emb|CAC38433.1| (AJ302646) tat protein [Human i... 24 211

A.3 Plasmid maps

pEGFP-C1, provided by Clontech (GenBank Accession #: U55763)

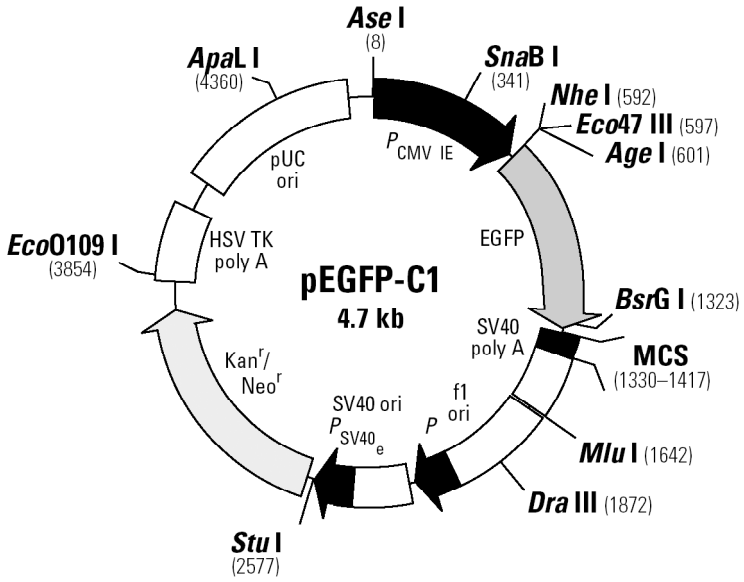


Figure 2.1: An overview of the most important features of the pEGFP-C1 plasmid.

This vector encodes the enhanced green fluorescent protein, EGFP, and neomycin/kanamycin resistance for selection in eukaryote cells and bacteria. The vector also contains SV40 and pUC origins of replication, as well as a multiple cloning site, MCS, after the EGFP coding region. When a gene is cloned into the MCS and it is in frame with the EGFP gene then the vector will express a fusion protein when in eukaryote cells (figure 2.1).

pEGFP-Tat72, kindly provided by Vibeke Andresen, UiB, Norway.

The pEGFP-C1 vector with the first exon of *tat* from the HXB3 clone cloned in frame with EGFP. Expresses a fusion protein, EGFP-Tat72, when in eukaryote cell lines. (An overview of the most important features of this plasmid can be seen in figure 4.11)

pEGFP-Tat86, kindly provided by Vibeke Andresen, UiB, Norway.

The pEGFP-C1 vector with the first 86 codons of *tat* from the HXB3 clone cloned in frame with EGFP. The 87th codon of *tat* from this clone is a stop codon. Expresses a fusion protein, EGFP-Tat86, when in eukaryote cell lines. (An overview of the most important features of this plasmid can be seen in figure 4.11)

pCR-Blunt II-TOPO, provided by Invitrogen.

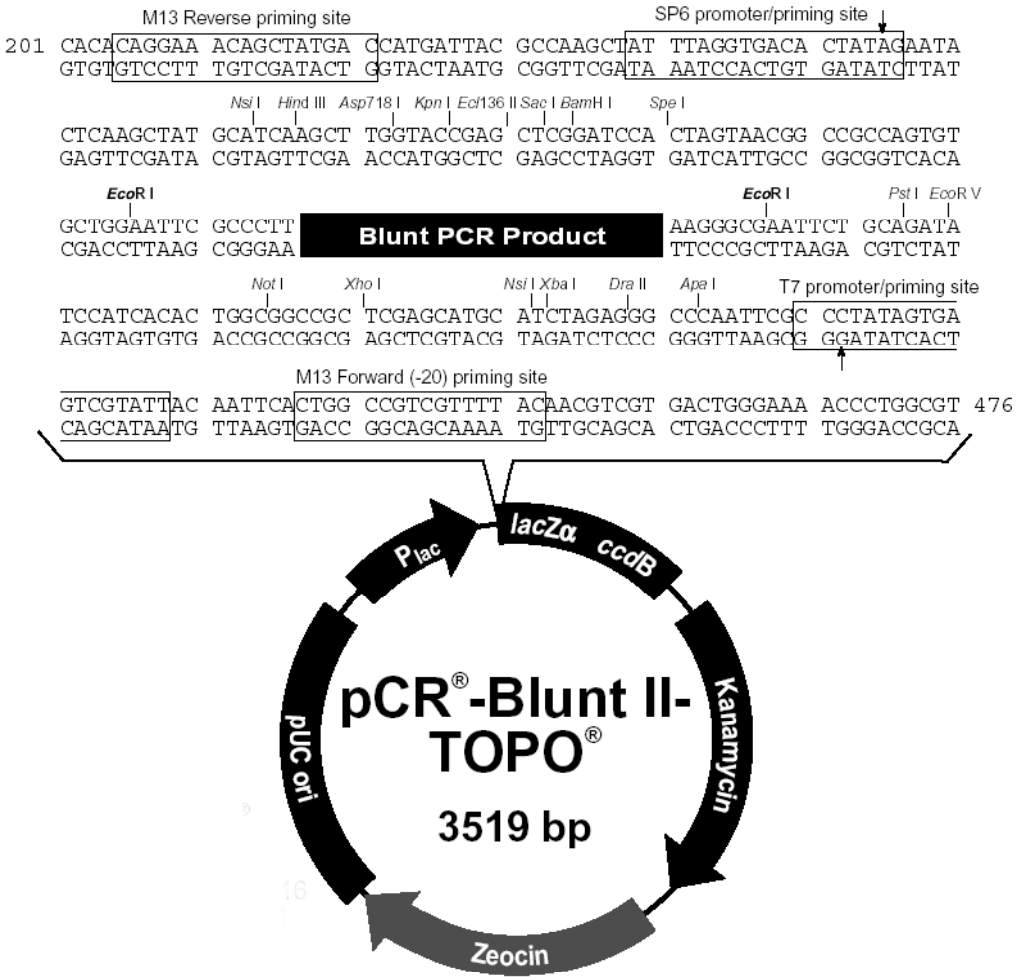


Figure 2.2: An overview of the most important features of the pCR-Blunt II-TOPO plasmid, as well as an overview of the cloning site.

This vector is part of the Zero Blunt™ TOPO™ PCR Cloning kit from Invitrogen. It is used to clone blunt-end PCR products. The vector is supplied linearized with vaccinia virus topoisomerase I covalently bound to the 3' end, which means that ligation with a blunt-end PCR product will occur spontaneously (figure 2.2).

pRSET-C, provided by Invitrogen.

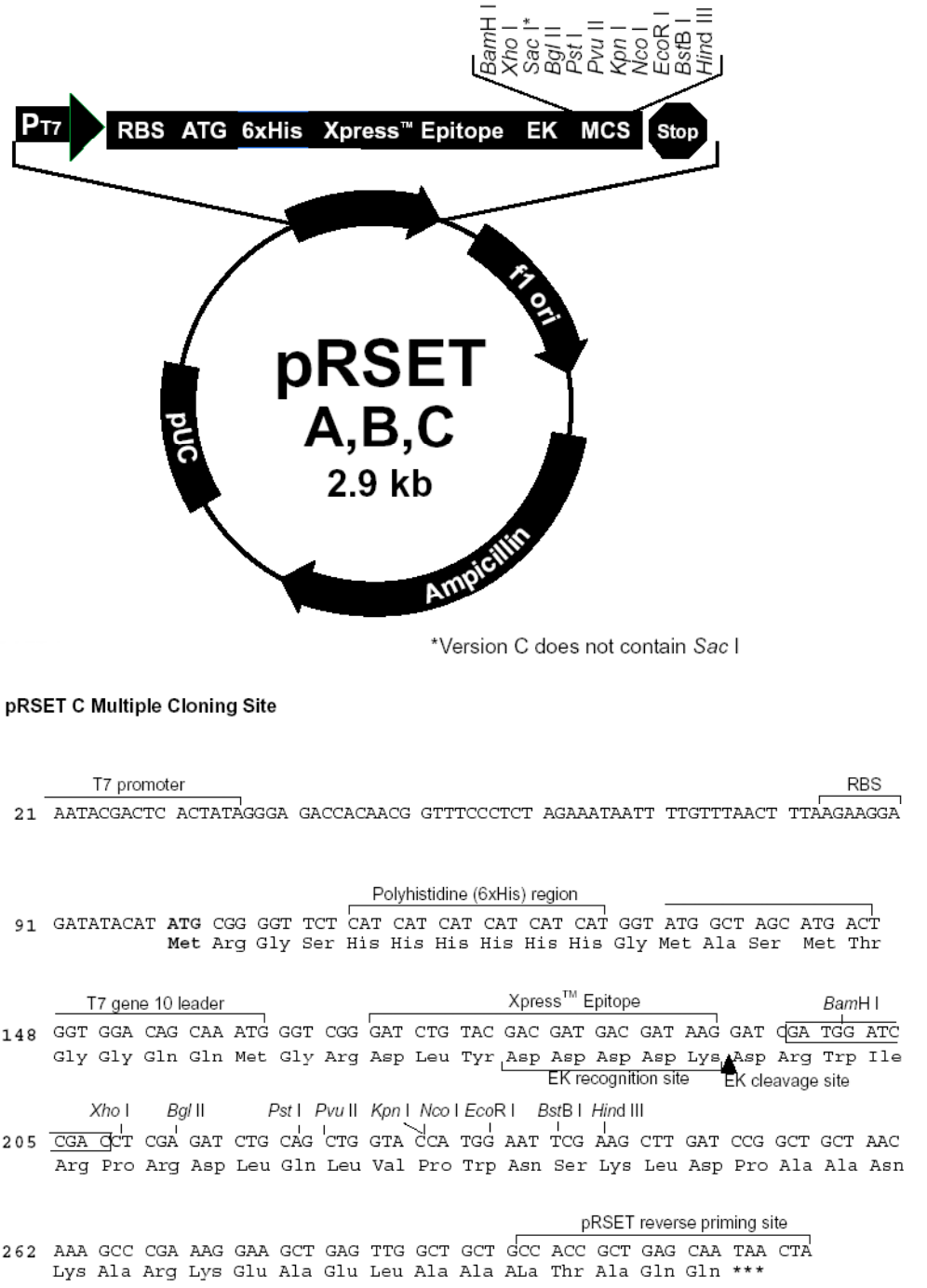


Figure 2.3: A map of the pRSET expression vector and of its multiple cloning site. There are three versions of this expression vector; the difference between them consists of a variable sequence-segment. In the different vectors the restriction sites are found in different reading frames due to this.

pRSET C is a bacterial expression vector (figure 2.3). A protein expressed from a gene inserted into the MCS will have an additional 30 something residues in the N-terminus. The exact number of additional amino acids depends on which restriction sites have been used when inserting the gene. A 6xhis-tag and an Enterokinase cleavage site is added to the N-

terminus of proteins expressed from this vector. The his-tag makes purification of the expressed protein possible by affinity chromatography, and the Enterokinase cleavage site makes it possible to remove most of the residues that have been added to the N-terminus of the protein. This can be useful seeing how extra residues may impair the function of expressed proteins. However these additional residues have previously been shown not to abolish transcriptional activity of either the Tat86 protein (Erik Stensrud, unpublished results) or the Tat72 protein (Staalesen, 2000).

LTR-CAT, kindly provided by Vibeke Andresen, UiB, Norway.

LTR-CAT is a reporter plasmid. It has been made using the psDred-N1 vector (Clontech). The CMV promoter has been replaced by the HIV-1 LTR. The HIV-1 LTR promoter (nucleotide –167 to +80) comes from the IIIextatIII plasmid (Dana-Farber). The chloramphenicol transferase (CAT) is not in reading frame with the gene encoding RFP. This results in the Tat dependent construct LTR-CAT. Since the LTR promoter is dependent on Tat to function properly, CAT will only be expressed when Tat protein is present. This plasmid can then be used to test the function of the various Tat proteins, or to examine whether extracellular Tat has been taken up by cells.