ISAS - INTERNATIONAL SCHOOL
FOR ADVANCED STUDIES

MOLECULAR DISSECTION OF HUMAN IMMUNODEFICIENCY VIRUS TYPE-1 TRANSCRIPTION REGULATION

Thesis submitted for the Degree of
Doctor Philosophiae

Candidate: Francesca Demarchi

Supervisor: Prof. Arturo Falaschi
Dr. Mauro Giacca

Academic Year 1991/1992

TRIESTE

SISSA - SCUOLA
INTERNAZIONALE
SUPERIORE
DI STUDI AVANZATI
TRIESTE
Strada Costiera 11

MOLECULAR DISSECTION OF HUMAN IMMUNODEFICIENCY VIRUS TYPE-1 TRANSCRIPTION REGULATION

Thesis submitted for the Degree of
Doctor Philosophiae

Candidate: Francesca Demarchi

Supervisor: Prof. Arturo Falaschi
Dr. Mauro Giacca

Academic Year 1991/1992

CONTENTS

INTRODUCTION

RESULTS

INTRODUCTION

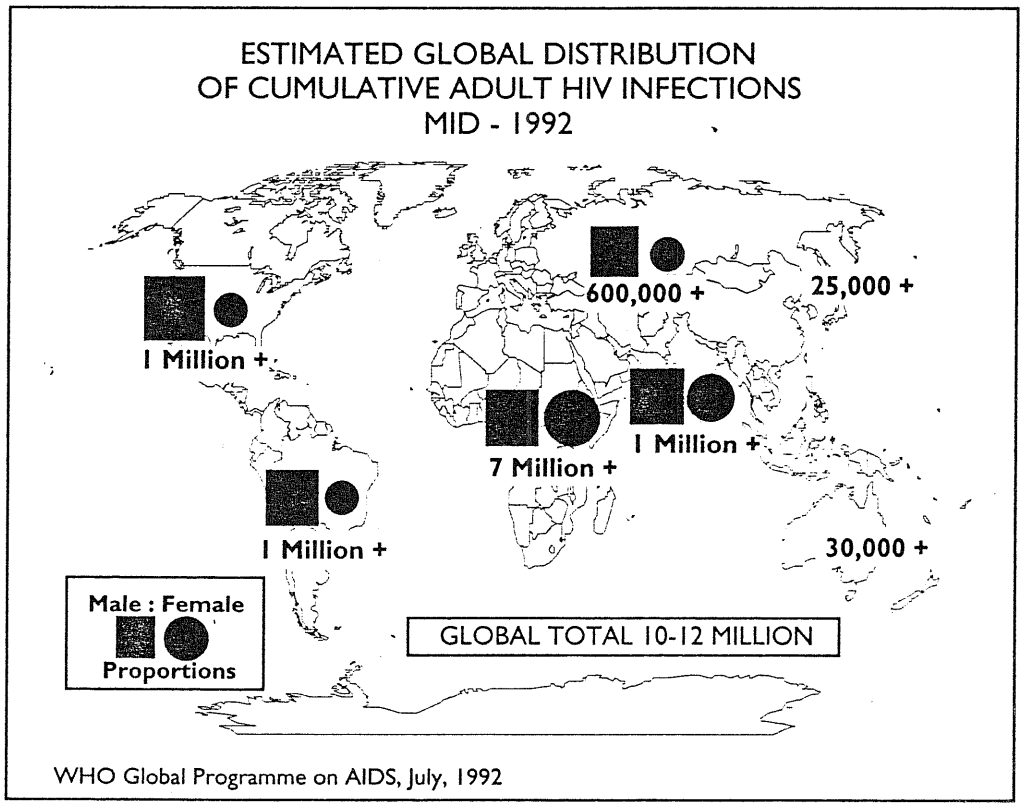
HISTORY AND CLASSIFICATION

Ten years after its definition in 1982, by the U.S. Center for Disease Control (CDC), the acquired immunodeficiency syndrome (AIDS) represents today a global health crisis, with its 418,404 AIDS cases reported to the WHO in October 1991.

Human immunodeficiency virus type 1 (HIV-1), isolated for the first time in 1983 in the laboratory of Luc Montaigner, is the etiologic agent of AIDS [Barre-Sinoussi 1983]. Since its discovery, the HIV-1 virus has been thoroughly studied, at a pace that far exceeds that of any other pathogen. Nevertheless, it is still not clear how infection with HIV-1 results in AIDS, and an efficient therapy is not yet available.

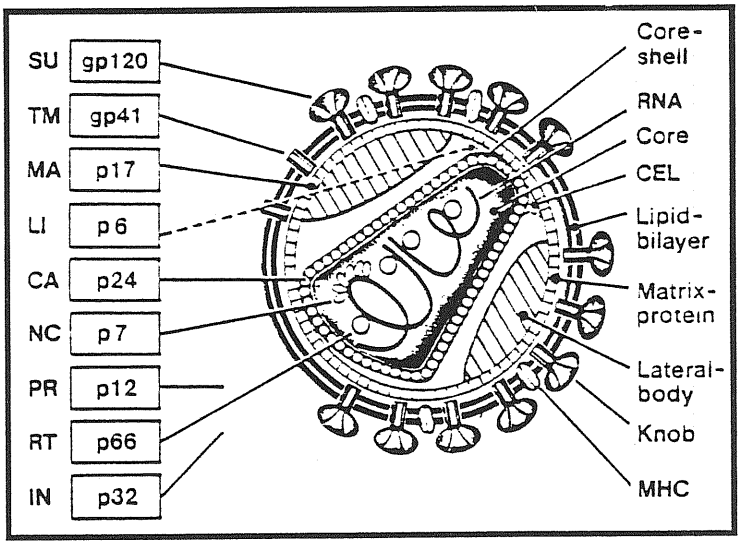
HIV-1 is the prototype of human lentiviruses, a group of retroviruses that induce persistent, slowly progressive, often fatal diseases in a number of mammalian species. Historically, the first retrovirus identified as an infectious agent, described already in 1904, was a lentivirus, equine infectious anemia virus (EIAV) [Vallee 1904]. Studying maedi-visna virus (MVV) in sheep, Sigurdsson coined the unifying term "slow virus diseases" for the clinically defined complex of animal diseases [Sigurdsson 1954]. In the mid-1900s, the development of inbred mice strains allowed novel approaches for murine leukemia viruses research, and the discovery of reverse transcriptase in 1970 [Baltimore 1970],

[Temin 1970] opened a new era of molecular virology and oncogene research. The rapid characterization of HIV-1 as the etiologic agent of AIDS is due not only to the ample application of molecular genetic techniques, but has considerably gained from the knowledge accumulated during many years of oncovirus research.



VIRUS STRUCTURE

Like all retroviruses, HIV-1 is a single positive strand RNA virus with particles approximately 100 nanometers in diameter. By electron microscopy, virions have a characteristic dense, cylindrical protein core that encases the duplex RNA genome, approximately 9700 nucleotides long, and viral enzymes required for the first steps of the virus life cycle: reverse transcriptase, integrase and ribonuclease. Surrounding the core is the membranous lipid envelope, whose ultrastructure with the cylindrical core is indistinguishable from that of the type-C retroviruses such as HTLVs. A matrix protein, p17, lines the inner surface of the membrane; the outer surface of the particle comprises an envelope protein organized into spikes, which is embedded in the lipid membrane [Haseltine 1991].



THE NUCLEOCAPSID

The nucleocapsid proteins are produced from a precursor peptide encoded by the late gene gag (described at pag.16). The major capsid protein is known as p24 since it has a molecular weight of 24 Kd in the mature virus particle [Schupbach 1984]. The purified p24 protein will self-assemble into tube-like structures that have the approximate diameter of the central core particle. Besides its structural role, the nucleocapsid is also directly involved in the assembly of the virus particle. This regulatory function is exerted by a region that is rich in histidines and cysteines and can bind zinc by formation of a zinc finger [Green 1989]. Mutations studies demonstrate that each of the cysteines and the histidines of the proposed finger are required for the binding of RNA and the formation of an infectious particle [Gorelick 1988]. A specific region of the viral RNA genome, which is located near the 5' terminus of the molecule, is required for encapsidation of the viral nucleic acid. Indeed, deletions of this region prevents entry of the viral RNA into the capsid protein [Lever 1989] . Alltogether mutation studies on the nucleocapsid protein and viral RNA molecule indicate that the RNA is captured into the virus particle via a specific nuclear protein-RNA interaction. The capture of viral RNA occurs when the nucleocapsid protein is part of a much larger precursor polyprotein, which is cleaved to release the small nucleocapsid protein only after budding [Haseltine 1991].

THE ENVELOPE

The external surface of the human immunodeficiency virus type 1 contains an external spike glycoprotein designated gp120, that is non-covalently bound to the transmembrane glycoprotein gp41. Both are made from the same gene as a gp160 precursor that undergoes cleavage by host enzymes at a highly conserved site; this cleavage is required for viral infectivity, indicating a key role of these proteins in the infectious cycle [Kowalski 1987].

The gp160 env product is made from a spliced message (env) that may or may not also include the vpu gene. The initial env precursor protein contains about 856 aminoacids. The first 30-32 are cleaved as a signal peptide, while the next 480 ultimately become gp120 and the carboxy terminal 344 become gp41. More than half of the molecular mass of gp120 is represented by carbohydrates, whereas only a quarter of gp41 is made of carbohydrates.

Gp120 is the most external glycoprotein of the HIV virus and contains active sites for virus neutralization and attachment to a host cell membrane receptor, namely the CD4 protein of CD4+ T lymphocytes and macrophages. Major domains for binding to CD4 have been identified both in the region of the carboxy terminus and in the aminoterminus of the

gp120 protein, thus indicating that a conformational structure linking together the ends of the molecule is necessary for the optimal effect of CD4 binding [Leonard 1990]. The gp160 molecule of the prototype HIV-1 strain, HXB2, has 31 glycosylation sites: 24 in the gp120 portion and 7 in the transmembrane portion. The envelope carbohydrates of gp160 include high mannose, complex, and hybrid forms [Leonard 1990]. All the glycosylation sites are of the N-linked type, where N-acetylglucosamine is linked to the amide nitrogen of asparagine and they are highly conserved in position between different types and isolates, presumably because they provide some benefit for evolutionary survival [Leonard 1990]. When simultaneous point mutations are made in more than two N-linked glycosylation sites in the amino terminus of the molecule, the ability of the virus to be infectious or to form syncytia is essentially eliminated. Conversely, the removal of up to six sites at the carboxy terminus of gp120 has not major effects on infectivity [Lee 1991]. These results suggest that conserved N-linked carbohydrates of the HIV envelope maintain the molecule in a configuration that may interfere with the generation of protective immunity.

The HIV-1 envelope glycoproteins constitute important targets for elicitation of neutralizing antibodies. Identification of the envelope domains involved in antibody neutralization of HIV-1 is a critical step in developing a vaccine for AIDS. Several continuous epitopes on glycoprotein gp120 recognized by neutralizing antibodies have been determined, including the loop structure in the V3 region which is believed to be the principal neutralization determinant [Olshevski 1990],

HIV-1 GENOME ORGANIZATION

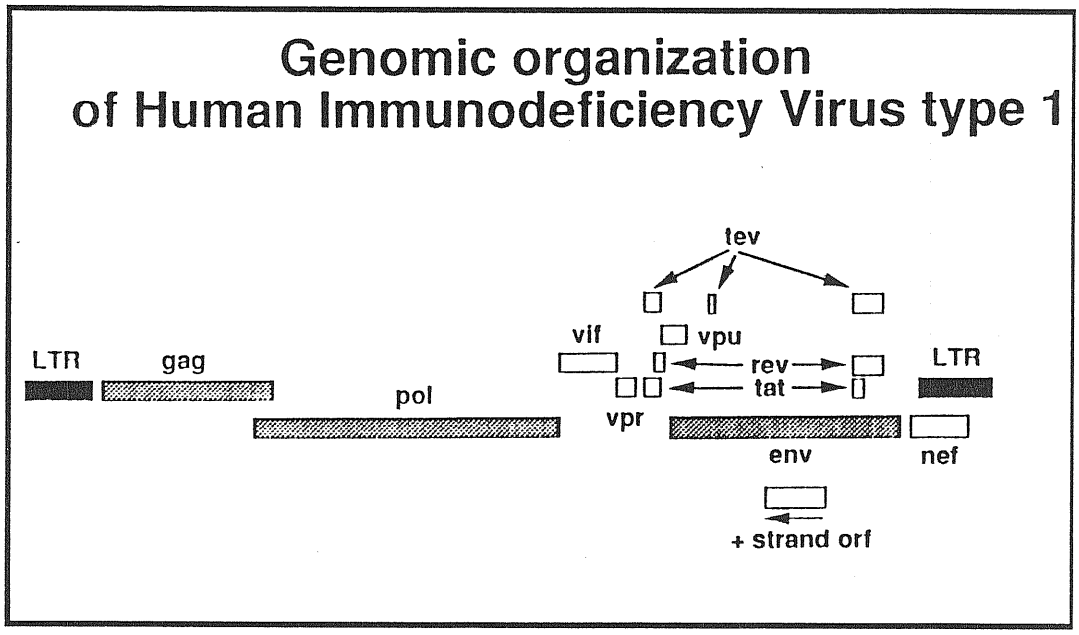
The HIV-1 genome is 9.7 Kb in length and has the same general organization as other retroviruses; the integrated proviral DNA has long terminal repeat regions flanking the genes coding for the major structural proteins, gag pol and env. However, as it is the case for the other lentiviruses, in addition to Gag, Gag-Pol and Env polyproteins, HIV-1 produces several other proteins, most of which are involved in the regulation of the virus life cycle.

The reverse transcriptase of HIV-1 and of all other known retroviruses is prone to error, whose frequency in vitro is one base misincorporation per genome per replication cycle [Dougherty 1988], [Leider 1988]. Moreover, deletions, insertions, duplications, and recombination events accompany the process of reverse transcription [Dorner 1985], thus producing an heterogeneous population of progeny viruses. The highly variable nature of HIV-1 virus has been demonstrated to occur *in vivo* by sequencing a number of different isolates in many patients [Hahn 1986]. As a consequence of this variability, each infectious inoculum *in vivo* is heterogeneous and unique and it is subject to continuous modifications. The presence of different variants of HIV-1 plays an important role in determining the length of the interval from infection to conclamated disease, together with the immunologic status of the host and exogenous agents which affect the virus, the host or both. The multitude of HIV-1 quasispecies can be divided in two major distinct biological variants: the syncytium inducing variants which determine a

[Cordell 1991], [Lasky 1987]. However, due to its sequence heterogeneity among isolates, the V3 loop induces type-specific neutralizing antibodies, which do not account for the broad virus-neutralizing activity detected in the sera of most infected persons. Steimer et al [Steimer 1991] have recently shown that such broadly neutralizing human antibodies are directed against discontinuous epitopes on gp120.

rapid CD4 decline and are associated with a rapid development of AIDS, and the variants non inducing syncytia, associated with a slow progression to the disease.

HIV-1 genome variability complicates a lot the choice of an efficient therapy oriented to inhibit specific features of the virus. As a consequence of HIV heterogeneity and high mutation rate, the reverse transcriptase inhibitors such as dideoxinosine and AZT, that have been largely used as antiviral agents in HIV-1 patients, have proven to have different effects on different patients and the positive effect, if present, has always been only transient.



LONG TERMINAL REPEATS

At both ends of the genome there are two identical 634-bp noncoding sequences divided into discrete functional units designated U3, R, and U5 going from the 5' to the 3' end. These functional units are critical for the integration of the virus into the host genome and contain promoter and enhancer elements as well as signals for transcription termination. The LTRs are flanked by short internal repeats with the conserved dinucleotide TG [Starcich 1985]. Immediately downstream of the 5' LTR there is the tRNA primer binding site with 18 base pairs complementary to tRNA Lys, which initiates minus strand DNA synthesis (see below at pag,). Immediately upstream of the 3' LTR there is a perfect 15-16 nucleotides polypurine tract important in initiation of plus strand DNA synthesis [Ratner 1985].

The U3 region is 453-456 base pairs long and terminates at the RNA transcription start site. This region of the genome contains an array of binding sites for cellular proteins involved in transcription initiation and regulation and will be discussed in detail in a subsequent section on trascription regulation. At the junction of the U3 region with the R region (-17 to +24) there are the sequences required for transactivation by the viral protein Tat (see below).

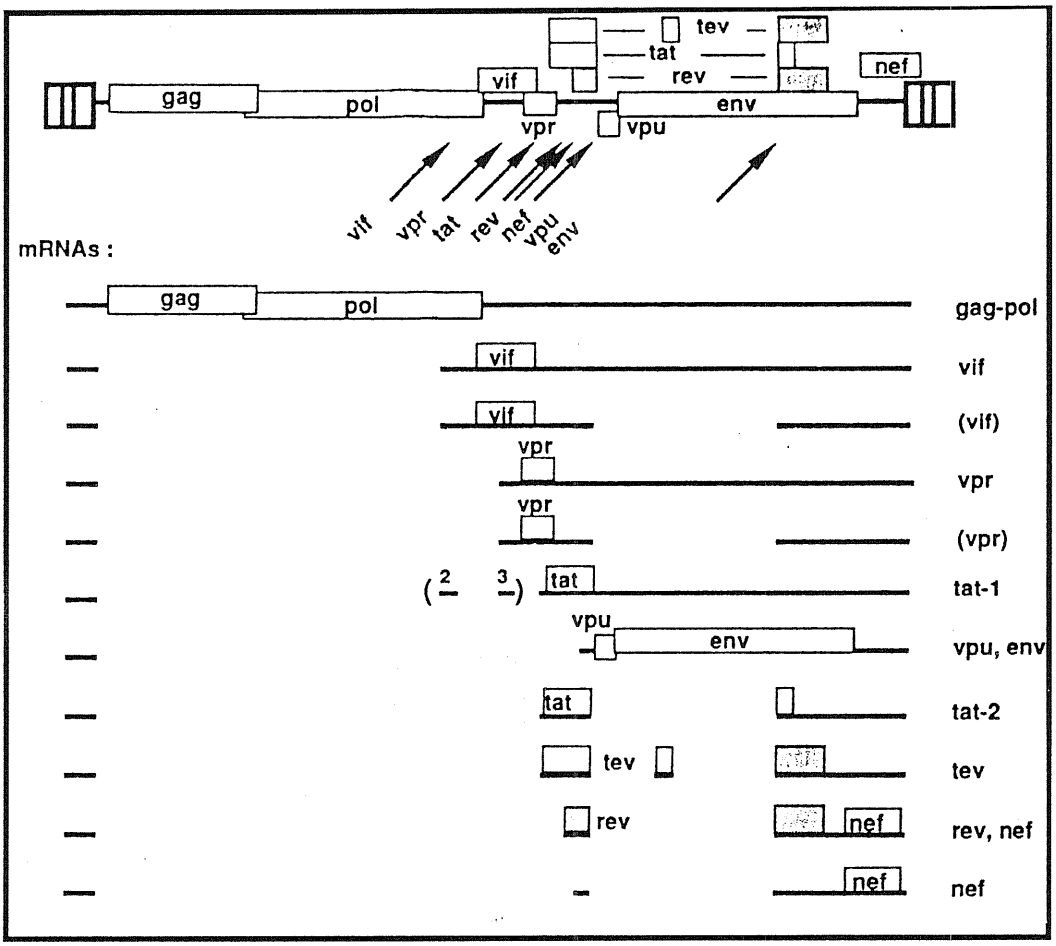
The R region is 98 nucleotides long and is located between the mRNA initiation site and the polyadenylation site with the terminal CA being the site of endonucleolitic cleavage [Bohnlein 1989]. The U5 region is 3' to the transcription unit and it is 83-85 bp long. This region contains

the polyadenylation signal AATAAA, 19 bases downstream to the R-U junction and GT clusters required and sufficient for proper and efficient polyadenylation [Bohnlein 1989].

CODING REGION

Three different mechanisms are utilized by HIV-1 to produce all its proteins starting from a relative short genome: ribosomal frame shifting for Gag-Pol production, alternative splicing to produce all the viral proteins except Env, and the use of bicistronic mRNA for Env expression.

HIV-1 produces three classes of mRNA in infected cells: full length unspliced which are 9 kb long, intermediate singly spliced of 4.5 Kb and the 2 Kb small doubly spliced RNAs. By cloning and sequencing PCR amplified HIV-1 cDNA more than 20 small RNAs and at least 14 intermediate-size RNA have been found in the HIV-1 infected cells [Schwartz 1990a], [Schwartz 1990b]. The number and types of the mRNA vary according to the viral strain, contributing to the *in vivo* biological variability of HIV-1. Moreover, also the relative quantities of the different mRNA species varies extensively in different isolates [Felber 1990]. With the exception of Env, which is always present as the second or third open reading frame (ORF), each translated ORF is preceded by a splice acceptor. This results in mRNAs that contain the initiator AUG of the expressed ORF as the first AUG [Pavlakis 1991].



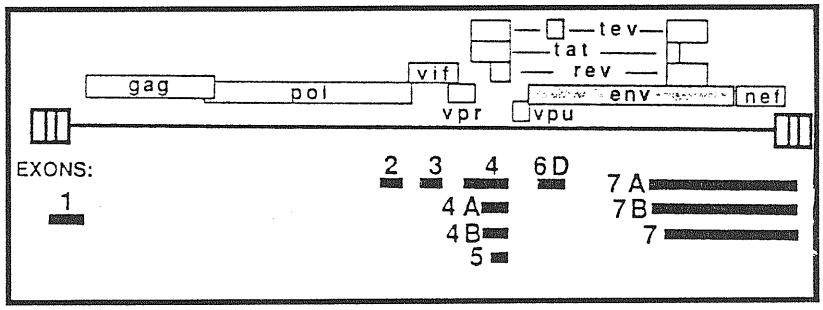
mRNA SPECIES OF HIV-1. Arrows indicate splice acceptor sites. Reproduced from Pavlakis 1991

The small multiply-spliced RNAs produce the regulatory proteins Tat, Rev, and Nef, as well as other variant proteins such as Tev and 6Drev [Benko1990], [Schwartz 90]. For each product, multiple mRNAs have been isolated, namely three *tat* mRNA containing exon 4, three *rev* mRNA containing exon 4A, three *rev* mRNA containing exon 4B and 3 *nef* mRNA containing exon 5. In addition, each member of these groups can contain one or both non coding exons 2 or 3, located in the 5' part of the virus,

contributing in this way to increase the number of mRNA species produced by HIV-1 virus [Haseltine 1991].

Exon 6D is present in the env region of some HIV-1 strains and it is used for the expression of Tev and 6Drev. Tev contains the first exon of Tat, 38 aminoacids of Env and the second exon of Rev. 6Drev starts at one of the two methionine initiation codons within exon 6D and contains the second exon of Rev. Tev accumulates primarily in the nucleolus and has both Tat and Rev activities, while 6Drev is found both in the cytoplasm and nucleolus and doesn' t have any Rev activity [Benko 1990].

In the last exon, exon 7, two alternative splice acceptor sites (7A and 7B) have been found upstream to the splice acceptor 7 [Schwartz1990]. Although the biological role of the mRNAs produced by the utilization of the acceptor sites 7A and 7B is not yet clear, it is interesting to note that they are conserved among HIV-1 isolates as well as in HIV-2 and SIV.



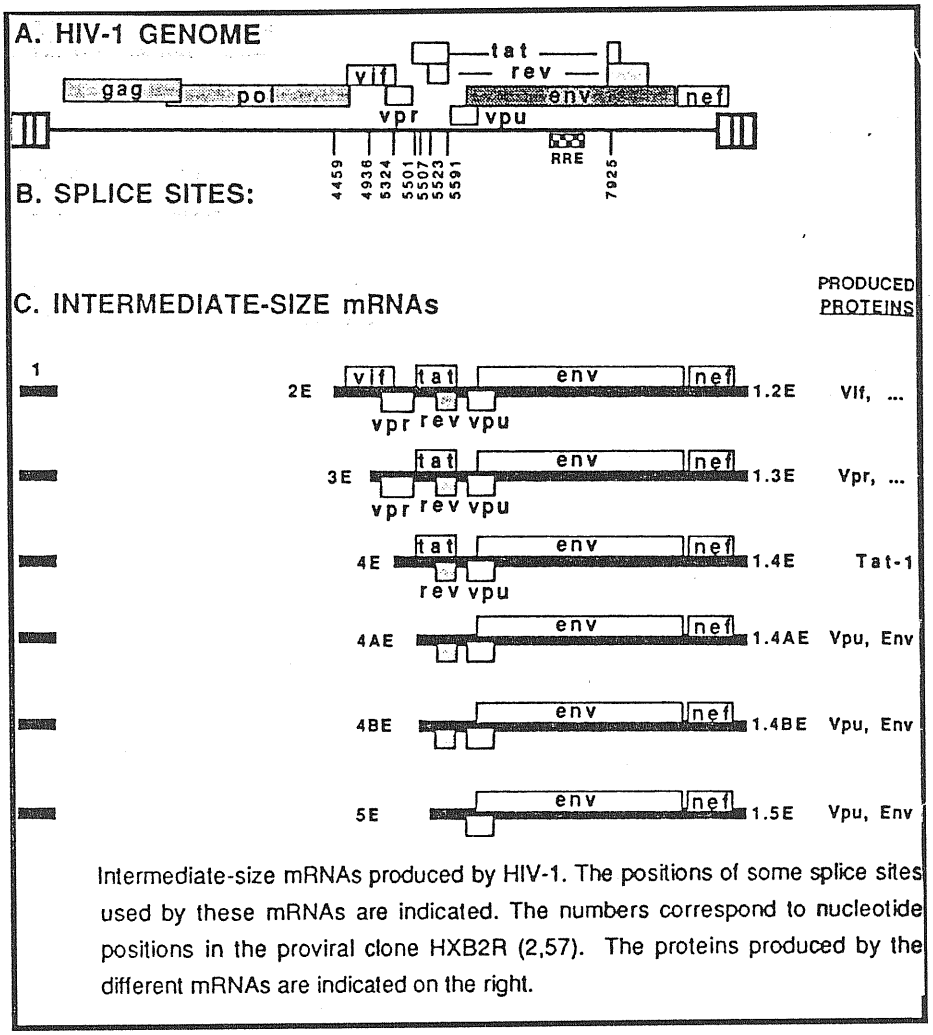
Schematic rapresentation of HIV-1 exons

Intermediate size mRNAs produced in HIV-1 infected cells encoding one exon Tat (Tat1), Vpu, Env, Vif and Vpr proteins can be distinguished respectively by the use of splice acceptor 4, 4A, 4B, 5,2 and 3 [Schwartz 1990]. As it is the case for the small RNAs, the intermediate - size mRNA can contain additional non coding exons such as 2 and 3. All intermediate size mRNAs contain the Rev responsive element (RRE) and are Rev dependent for their expression. Therefore, the proteins encoded by these messengers are expressed at a late stage in the viral life cycle. In the absence of Rev, further splicing occurs and small multiply-spliced RNAs are generated.

The redundancy in the produced mRNAs might be a reflection of the organizational principles of the lentiviral genome, it may also allow the rapid change of the virus by mutations affecting some splice sites. Alternative splicing may offer an evolutionary advantage to the virus, since proteins are produced from more than one mRNA; moreover, the plasticity of the genome may allow new combinations to arise and enable viral adaptation to a new environment.

Functional analysis of the small multiply spliced and intermediate-size mRNAs revealed that HIV-1 produces both monocistronic and bicistronic mRNAs [Schwartz 1990]. For example, the rev mRNAs encoded by exons 1. 4A. 7 and 1.4B.7 produce both Rev and Nef proteins and the env mRNA encoded by exons 1.5E, 1. 4AE, and 1. 4BE express both Vpu and Env proteins. On the contrary, all mRNAs that contain *tat* AUG as the first AUG produce only one product, despite the fact that

downstream ORFs were present on these mRNA. It has been shown that the HIV-1 mRNAs are translated by the scanning mechanism, and leaky scanning causes expression of more than one protein from some mRNA. In particular, the *rev* and *vpu* AUGs do not contain a signal for efficient initiation of translation, thus allowing high expression of downstream ORFs. On the other hand, *tat* AUG is a very strong initiator, resulting in the absence of expression from downstream ORFs.



Intermediate-size mRNAs produced by HIV-1. The positions of some splice sites used by these mRNAs are indicated. The numbers correspond to nucleotide positions in the proviral clone HXB2R (2,57). The proteins produced by the different mRNAs are indicated on the right.

STRUCTURAL GENES

gag. The first 5' open reading frame codes for the precursor of the internal structural proteins of the virus and extends from the ATG initiation codon at nucleotides 334-337 to nucleotide 1837 relative to the cap site. A 55 KDa precursor Gag peptide is synthesized from the unspliced genomic-lenght mRNA. The myristylated 55 KDa Gag protein is processed to a p41-kDa intermediate and, ultimately, to at least four smaller peptides. The N terminal protein, p17, is the major matrix protein and is myristylated as well as phosphorylated. The other three peptides constitute the viral core, surrounding the viral genome and replicative enzymes. The major phosphorylated core protein p24, initiates at the carboxy end of p17 at codon 133. A precursor peptide, p15, is cleaved at the C terminus of p24 at residue 378. This is further processed between aminoacid 447 and 448 to a 7-kDa N-terminal protein and a 6 KDa C terminal protein [Veronese 1987]. The p7 regions has two runs of repeated cysteine residues corresponding to nucleic acid binding regions in other retrovirus proteins [Mervis 1988]. The processing of the p55 precursor is mediated by the specific virus-coded protease gene in the adjacent and overlapping region. Indeed, inclusion of this overlapping protease reading frame has allowed the successful expression and processing of the Gag proteins in yeast, Escherichia coli, and vaccinia vectors [Kramer 1986], [Gowda 1989].

pol. The second open reading frame of the HIV-1 genome

overlaps the gag reading frame by 241 nucleotides with pol in the -1 phase with respect to gag [Ratner 1985]. Pol is expressed as a 90-92 KDa Gag/Pol fusion protein from a genomic length mRNA. To produce both the Gag and the out-of-phase Gag/Pol fusion proteins, there is a ribosomal frame shift during translation that occurs at a UUA leucine codon with an efficiency of 11%. This type of frame shifting is also seen in RSV and MMTV. A minimum of 26 nucleotides in this region with a run of six uracil residues appears to be sufficient for frame shifting [Wilson 1988]. This suggests that a predicted stem loop structure immediately downstream from UUA may not be required for frame shifting [Wilson 1988]. Early sequence similarities with other retroviruses allowed tentative identification of functional discrete proteins coded within the pol open reading frame. The long pol open reading frame encodes for at least four polypeptides with four associated enzymatic functions, the 10 KDa fully processed protease, reverse transcriptase/RNaseH (p66 and p55), and the 34 KDa endonuclease/integrase. At the N-terminus there are sequences similar to the proteases of a number of retroviruses including BLV, HTLV-I, HTLV-II, Mo-MULV, RSV, and visna virus. The location of the protease at the 5' end of the pol open reading frame is structurally similar to Moloney murine leukemia and visna virus in contrast to HTLV-I, where such activity is located between the pol and gag genes. This gene product responsible for enzymatic cleavage of gag and gag/pol precursors is released from the gag/pol precursor in an E. coli system by a two step cleavage first at the Phe-Pro bond between protease and reverse transcriptase resulting in an 18 KDa product and then a subsequent internal cleavage at the Phe-Pro bond releasing the stable active 10 KDa

protein [Mous 1988]. The protease belongs to the class of aspartic proteases found in Mo-MuLV, bovine leukemia virus, and HTLV-I , since it is partially inhibited by pepstatin A and an internal mutation of the asparagine to threonine or alanine causes loss of activity [Seelmeier 1988], [Le Grice 1988].

The reverse transcriptase protein is encoded adjacent to the C-terminus of protease and exists in two forms of 66 KDa and 55 KDa respectively. Reverse transcriptase activity is localized in the N terminus of the protein of p66 and of p55, which share 156 aminoacids downstream of the start site. When the reverse transcriptase gene is expressed in bacterial vectors, besides the reverse transcriptase activity, also an RNAase activity is detected. This activity is localized in the terminus of p66 as well as in the p15 polypeptide which appears to be the C terminus cleavage product obtained from a further processing of p66 in p55.

Mutational analysis of the reverse transcriptase gene suggests two areas that might be good candidates for triphosphate-binding sites, namely the region between aminoacid 257 and 266 and the region between aminoacid 109 and 116. Substitutions in these areas result in a significant decline in azidothymidine triphosphate sensitivity [Larder 1987], and their natural occurrence in vivo is the cause of failure of antiviral therapy with nucleotides analogues, such as AZT, ddI and ddC.

At the 3' end of the pol open reading frame there is a region encoding for a 34 kDa protein with endonuclease-like activity similar to the integrase of other retroviruses [Ratner 1985].

env. The large open reading frame coding for the envelope glycoprotein is situated at the 3' end of the the genome, extending from nucleotide 5781 to 8369 with the ATG at nucleotide 5802 [Ratner 1985]. The env gene codes for a large 854-873 aminoacid precursor protein which is processed by endoproteolytic cleavage to form the N terminal exterior gp120 and the C-terminal transmembrane protein, gp41[Veronese 1985]. The 4.2-kb env message is a singly spliced polycistronic message with the splice donor at nucleotide 287 and the splice acceptor at 5358 [Muesing 1985], or 5557 [Saldaie 1988]. The env coding sequence partially overlaps coding sequences for regulatory proteins such as tat, nef, rev and vpu. Furthermore some cis acting negative regulatory sequences (CRS) are located within the env gene. When present in the RNA, the CRS prevent the message from being used as a substrate for protein synthesis, probably by retaining the RNA in the nuclear compartment. The negative effect of CRS is relieved through the interaction of the Rev protein and the rev responsive element (RRE), which is also present within the env open reading frame [Rosen 1986]. The interaction of Rev with its target site overrides the repressive effect of the CRS, by permitting the transport of the messenger in the cytoplasm [See also pag.26].

The precursor Env polypeptide has three stretches of hydrophobic residues corresponding to an N-terminal signal peptide that is cleaved prior to further processing, an internal arginine rich site of proteolytic cleavage and the membrane-spanning region of the transmembrane protein. After cleavage of the leader signal sequence at amino acid 37, the N-terminus of the precursor polypeptide codes for the 480-amino acid

hydrophilic peptide, gp120 [Starcich 1986]. The 345 amino acid transmembrane protein, gp41 (516-863) at the C-terminus of the precursor polypeptide is cleaved from gp120 at a site located in an arginine-rich region adjacent to the hydrophobic stretch of gp41. After cleavage of gp120 from gp41, the two molecules continue to be noncovalently linked via a number of sites at the N-termini of both proteins [Kowalski 1987].

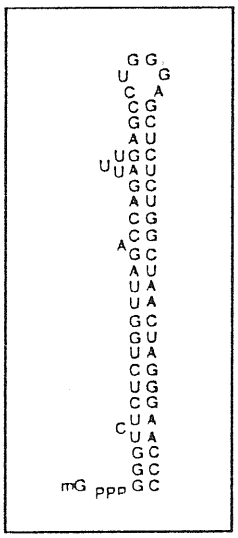
REGULATORY GENES

tat. The transactivator gene Tat is an essential gene, since *tat* mutants are not able to replicate [Dayton 1986], [Fisher 1986]. The product of the tat gene functions as a positive regulator by increasing both its own expression and the expression of all viral proteins [Sodroski 1985].

Transactivation by Tat requires a specific recognition of this effector protein with the target sequence TAR, contained in the R region of the viral LTR [Rosen 1985]. TAR has been precisely mapped between nucleotides +1 and + 42 with respect to the initiation site of viral transcription, by means of extensive mutagenesis [Hauber 1988], [Jakobovitz 1988]. It has also been shown that the position and orientation of the TAR element with respect to the initiation start site is critical for its function. Indeed, inversion of the sequence or an increased spacing from the initiation start site diminishes Tat transactivation activity [Muesing 1987], [Shelby 1989].

The RNA encoded by TAR, between +1 and +111, has the potential

to form one or more stem and loop secondary structures which are potentially present in all newly transcribed mRNAs. The secondary structure typical of the TAR is necessary for its function. Indeed, it has been shown that mutations altering this structure significantly impair Tat mediated transactivation, whereas mutations preserving the hairpin and the bulge do not have any effect [Berkhout 1989], [Feng 1988].



TRANSACTIVATION RESPONSIVE REGION (TAR)

Purified Tat can interact directly with TAR RNA; but not with TAR DNA. The binding appears to occur specifically at the three-nucleotide bulge structure present in the stem of TAR RNA (+23, +25), as it is completely abolished by a single uridine nucleotide change within this structure [Roy 1990 a]. Tat-TAR interaction is certainly required for

transactivation, since a single point mutation which prevents this interaction *in vitro* abolishes also transactivation *in vivo* [Roy 1990 a]. However, it is not sufficient, but requires also additional cellular factors, since some specific mutations of the loop which do not alter *in vitro* binding, can diminish tat transactivation *in vivo* [Roy 1990 b]. Several groups have isolated cellular proteins which are able to bind to TAR RNA [Gatignol 1989], [Gaynor 1989], [Marciniak 1990]. Moreover, host proteins which can specifically bind purified Tat forming a protein-protein complex have been identified. One of these proteins, TBP-1, suppresses Tat-mediated trans-activation *in vivo* [Nelbock 1990]; another one, MSS1, appears to play a key role in Tat mediated activation of HIV genes [Shibuya 1992].

Tat transactivation is bimodal, working both at the transcriptional and post-transcriptional level [Cullen 1986] and the contribution of each component may depend on the nature of the target cell. It has been shown that in the absence of Tat, the viral transcripts are prematurely terminated approximately 60 nucleotides after the start of mRNA transcription, suggesting that Tat could act primarily as an antiterminator of RNA elongation [Kao 1987]. A subsequent study showed that Tat both increases the frequency of initiation and stabilizes transcriptional elongation from the TAR element of the LTR [Laspia 1989]. To explain the dual function of Tat it has been proposed that its binding on TAR RNA attracts transcription factors to the adjacent promoter region [Sharp 1989]. This hypothesis is supported by several lines of evidence. For example, a Tat-Jun hybrid protein is able to activate a LTR containing four AP1 target sites inserted into the TAR element. This activation, however, represents

only a fraction of the normal tat activity and it is even lower if NF-kB and SP1 sites are deleted, demonstrating an important role for this factors in the activation process [Berkhout 1990]. Another evidence supporting Tat function in increasing transcription initiation is that a chimeric protein containing Tat and the RNA binding region of the Rev protein is able to activate transcription from an LTR in which the TAR has been replaced with the specific sequence recognized by Rev [Southgate 1990]. Thus, the ability of Tat to activate transcription in the absence of TAR indicates that Tat alone, without any additional TAR binding proteins, is able to activate transcription.

Besides its role in increasing the frequency of initiation and stabilizing transcription elongation, there is a strong indication that Tat can work also at the post-transcriptional level. When TAR containing chloramphenicol acetyltransferase (CAT) RNA was microinjected into the nucleus of Xenopus oocytes along with purified Tat, expression of the CAT reporter gene occurred even in the presence of transcription inhibitors. On the other hand, if the injected RNA didn't contain the TAR element, CAT gene expression was not detected, indicating that Tat is able to increase either the stability or the translatability of TAR containing RNA [Braddock 1989].

It has been proposed that, besides its key role in HIV-1 regulation, Tat may affect also the activation of cellular genes of both infected and uninfected cells and modulate cell proliferation contributing to the progression of at least one malignancy, the Kaposi sarcoma. Kaposi sarcoma cells injected in nude mice produce an increase in vascular

permeability and angiogenesis, but do not induce neoplasia. The same effect is obtained by inoculating the growth medium of Kaposi cells grown with HIV-1 infected cells: the responsible factors are: tat, IL1, IL8, IL2, TNF, corticosteroids, bFGF. Tat is produced by adjacent HIV-1 infected cells, upon stimulation by cytokines such as IL1, IL6, PDGF, GMCSF. Tat protein is present in the medium of infected cells and its action on Kaposi cells can be explained only assuming that Tat is secreted and taken up by adjacent cells. An evidence supporting this hypothesis is that Tat is present in the medium of infected cells at a concentration which does not correlate directly with the amount of cell death. Moreover, the Tat protein is found in the medium of cells transfected with the tat gene [Ensoli 1990]. Together these findings suggest that Tat can be actively secreted, although the mechanism is still unknown. It has also been shown that Tat is actively taken up by cells [Frankel 1988], and it is targeted to the nucleus, where trans-activates LTR-directed gene expression [Gentz 1989]. Furthermore, cocultivation of Tat expressing cells with cells containing an integrated copy of an HIV-1 LTR reporter gene results in a significant increase in the activity of the reporter gene confirming that Tat can be taken up by cells. Tat secretion and uptake by adjacent cells, besides playing a role in Kaposi malignancy, probably through the activation or repression of cellular genes (see Discussion for further details) can result in transactivation of HIV-1 LTR in infected cells in which the virus is silent.

The function of Tat as a very efficient transactivator of virus production makes it a suitable target for antiviral therapy. Pharmaceutical industries are putting a lot of effort in designing efficient Tat antagonists

and some of these products, such as for example RO5 335 from La Roche have already been used in pilots studies in patients with encouraging results. However, the growing body of evidence suggesting a role for Tat in the activation of cellular genes indicates that a Tat- antagonist might have some effects also on the host cell, and such effects should be carefully evaluated.

nef. The HIV-1 nef gene can be deleted without abrogating the ability of the virus to replicate [Luciw 1987], [Terwillenger 1986], [Ratner 1985]. However, the conservation of this gene in HIV-2 and simian immunodeficiency virus from macaque argues for an important role in the virus life cycle.

The term nef is an acronym for negative factor. Early reports [Terwilligen 1986], [Luciw 1987] showed a small negative influence of the nef gene on virus replication. Subsequent studies reported even more dramatic down-regulatory effects of nef [Ahmod Venkatesan 1989]. However, other authors found no effect of nef on HIV-1 replication nor on HIV-1 long terminal repeats driven CAT expression [Kim 1989], [Hammes 1989].

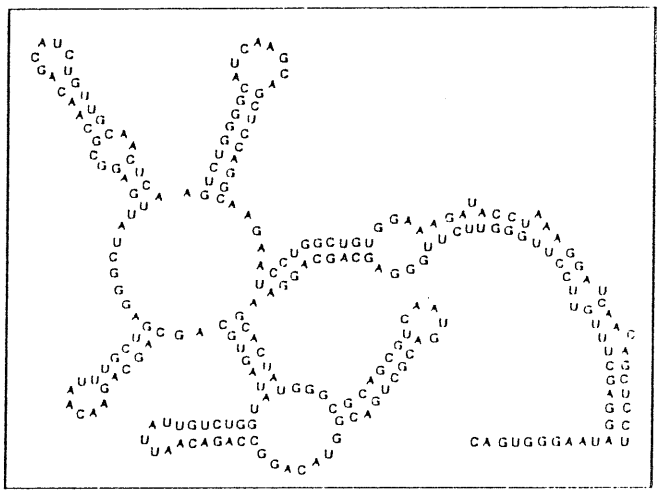
A possible role for nef emerges from studies of SIV mac 239 infected rhesus monkey. In this system nef is required for mantaining high virus loads during the course of persistent infection *in vivo* and for full pathologic potential [Kelleher 1991].

As it is the case of the other early expressed genes, nef is a good candidate for vaccine development. Eight or more epitopes in the central

region of nef are target of immune response during the first phase of infection; in later phases such response decreases progressively as a result of the establishment of escaping mutants, of the suppression of cells involved in the cell mediated response and of the drastic decrease of these cells due to apoptosis and necrosis induced by HIV-1. Encouraging results have been obtained in the lab of J.P. Levy by using recombinant lipopeptides containing the central region of nef to vaccinate macaques.

rev. The regulator of virion protein expression, rev, is a gene that positively regulates expression of genes encoding for virion proteins, but negatively regulates the expression of regulatory genes [Sodroski 1986], [Terwilliger 1988]. The Rev protein is specified by two coding exons that overlap the tat coding exons. The rev initiation codon lies 3' to that of tat and the same splice acceptor sites are used for both tat and rev second coding exons.

In the absence of Rev, the unspliced and partially spliced RNAs are neither exported from the nucleus, nor translated efficiently. Rev increases the stability and the transport of these messengers from the nucleus to the cytoplasm, where they are efficiently translated into structural proteins. Rev function is exerted through its specific interaction with the Rev Responsive element (RRE).



REV RESPONSIVE ELEMENT
(RRE)

Measurements of the level of cytoplasmic RRE-containing RNA resulted in a 3 to 10 fold increase in presence of Rev. Interestingly, this increase is accompanied by a several hundred fold increase in proteins produced from these RNAs [Felber 1989]. This discrepancy led to the suggestion that Rev may increase the translation of the messengers containing the REV responsive element. In accordance with this hypothesis, it has been shown that the small portion of the messengers containing the Rev responsive element that escape the nucleus in the absence of Rev are not translated efficiently [Felber 1989]. Moreover, studies on truncated gag mRNA showed that in the absence of Rev the cytoplasmic species do not associate with the polysomal fraction. These data explain the extremely low level of structural proteins produced in the absence of Rev and demonstrate the involvement of Rev in several post-transriptional steps, including transport, stability and utilization of the RRE containing messengers [Malim 1989].

Recently, a human gene, RBP927,encoding for a protein able to interact with the Rev Responsive Element has been isolated in the laboratory of Pavlakis. This gene is induced by interferons and its product can antagonize the effect of Rev [Communicated by Pavlakis at the VIII International Conference on AIDS - Amsterdam 1992].

Vpr The vpr coding region overlaps the 3' end of the vif gene and terminates before the initial methionine of Tat. The vpr open reading frame, which is highly conserved among HIV-1 and HIV-2 isolates as well as other lentiviruses like visna virus, codes for either a 77 or a 95-96 aminoacid protein that appears to be immunogenic. Two variants of the C-

terminus have been described; in HIV/HTLV-IIIB derived clones there is a termination codon at aminoacid 77, while in other isolates, a frame shift at this aminoacid causes the production of 18-19 additional aminoacids.

The Vpr gene product is dispensable for HIV-1 replication in culture; indeed, several isolates lacking a functional vpr are replication competent, even if their replication efficiency and their cytopathicity are slightly reduced [Ogawa 1989]. It has also been shown that vpr is able to trans-activate HIV-1 LTR specific gene expression by approximately two to three fold in transient expression assays, thus perhaps explaining the effect on viral replication [Cohen 1990].

vif. The VIF protein of human immunodeficiency virus type 1 is an important determinant of viral infectivity, it can increase the infectivity of HIV particles as much as 100- 1000- fold and may also enhance cell to cell virus transmission [Fisher 1987], [Sakai 1991]. Several lines of evidence suggest that vif plays an essential role during natural infections. First of all, vif antibodies are found in sera of patients at all stages of HIV-1 infection. Moreover, a vif open reading frame is found in other lentiviruses, including HIV-2, simian immunodeficiency virus, visna virus, caprine arthritis-encephalitis virus, and feline immunodeficiency virus [Chakrabarti 1987], [Myers 1988], [Talbott 1989]. Recently, it has been shown that the requirement for vif differs among established CD4 cell lines. In the absence of vif, the onset of virus replication in SupT1, C8166, and Jurkat cells was delayed, but peak levels of viral replication were similar. In contrast, virus replication was nearly undetectable in H9 and CEM cultures infected with a vif defective virus. These data suggest that vif can

compensate for cellular factors required for production of viral infectious particles that are present in some cell lines such as SupT1, C8166 and Jurkat, but are absent in others, such as CEM and H9 [Gabuzda 1992].

The 23 Kd vif protein is encoded by a singly spliced 5-Kb transcript which requires rev for expression and therefore appears late in the replication cycle, together with the gag, pol and env mRNAs. Since it is present in the endoplasmic reticulum and Golgi of infected cells, but it is not associated with the mature virus particle, [Arya 1986], [Kan 1986], [Lee 1986], [Sakai 1991], it has been suggested that vif may play a role in the processing or conformation of the HIV-1 envelope glycoproteins [Guy 1991].

Vpu. Vpu is encoded by a small open reading frame located at the 3' end of the genome, downstream of the first coding exons for tat and rev and codes for a 16 KDa protein, which does not have any equivalent in the other related lentiviruses. The Vpu protein is phosphorylated *in vivo* and appears to be located in the cytoplasmic membranes of expressing cells [Strebel 1989]. The vpu product is not essential; however, the loss of a functional vpu results in a 5 to 10 fold reduction of viral production [Strebel 1988]. The precise function of vpu product is still not clear, but the accumulation of cell-associated virions in vpu-defective strains suggests that it may have a role in enhancing viral release from the cell [Terwillinger 1989].

The functions attributed to HIV-1 regulatory gene products are summarized on the following page.

PROTEIN	FUNCTION
Tat	_ increase of transcription initiation frequency _ stabilization of transcription elongation _ stabilization or increase in translatability of mRNA _ transcription activation of cellular genes (?)
Nef	_ negative effect on viral replication (?) _ maintenance of high virus loads during infection (?)
Rev	_ increase of mRNA stability _ increase of mRNA transport andutilization
Vpr	_ increase of viral replication
Vif	_ processing and conformation of envelope protein
Vpu	_ enhance of viral release

HIV-1 INFECTION AND LIFE CYCLE

To initiate the infectious process, HIV-1 must first be delivered to a susceptible host. The routes for transmission demonstrated so far are sexual intercourse, exposure to infected blood or blood products, and transmission from infected mothers to their infants perinatally or from lactation after birth [Curran 1988]. As it is the case for other lentiviral infections, the ease of transmission may also be influenced by the physical state of the virus (free or cell-bound) and the general conditions of the host. The most efficient route of transmission is cell-cell contact; however, in an experimental animal system, inoculation of cell-free HIV-1 onto apparently unabraded vaginal mucosa did lead to infection of chimpanzees [Fultz 1986].

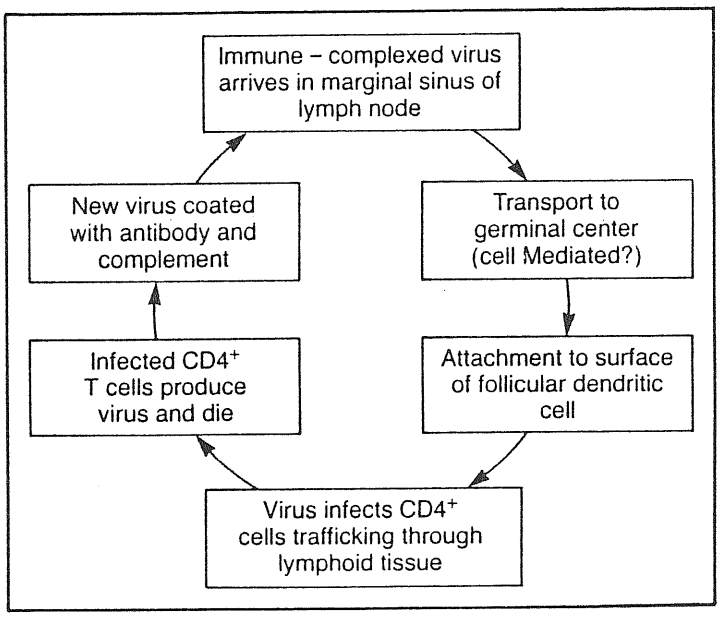
After transmission the virus is disseminated internally by host-dependent forces, which remove the virus from the interstitial spaces and deliver it to the lumen of blood or lymphatic vessels. During this process, specific interactions between the viral protein gp120 and the CD4 receptor present on MHC class II+ host cells, such as CD4+ T cells, monocytes, macrophages, microglial cells and Langherans cells, take place. The movement of antigen presenting cells in the lymph nodes is an important component of an effective immune response [Knipe 1990]. Entering via afferent lymphatics, they become part of the antigen presenting population of cells in the paracortical T cell zone [Razzecca 1986] and play a role in moving antigen and immune complexes from the marginal sinus to the follicles of lymph nodes (B cell zones). Infected cells may present viral antigens to initiate immune response against HIV-1 and

simultaneously spread the virus to uninfected cells by cell-cell transmission. When antigen-activated T cells reach the primary follicle, a rapid oligoclonal B response occurs and a germinal center is formed. The germinal center is composed of CD4 T+ cells, a large number of dividing B cells and a few antigen bearing dendritic cells . The immediate result of this process is the massive follicular hyperplasia, which results in lymphadenopathy, typically seen early in infection. A plateau of follicular hyperactivity may continue throughout infection or may decrease after some weeks. Antigen-containing complexes remain on follicular dendritic cells for prolonged periods of time and are believed to be involved in the generation and maintenance of both antibody-secreting plasma cells and B memory cells from the germinal center of B cell blasts [Klaus 1980].

The active germinal center and surrounding T cell rich cortex are sites of growth and activation of T cells, increasing the likelihood of integration of HIV provirus followed by viral expression. When a successful infection takes place, the cell may die or leave the germinal center or the lymph node to continue the spreading of the virus. The fate of the infected cell is influenced by the viral strain, the type of the cell and whether or not the cell becomes activated. As a consequence of these initial events, the first phase of infection is characterized by a drop in CD4+ T cells and a burst in HIV-1 replication. Following suppression of the HIV viremia, CD4+T cells increase a little, but usually remain lower than pre-infection levels, and the individual enters the asymptomatic phase of infection. The asymptomatic phase can last for many years during which time there is a progressive decline in CD4+ T cells.

The events that occur in the follicles are not completely understood,

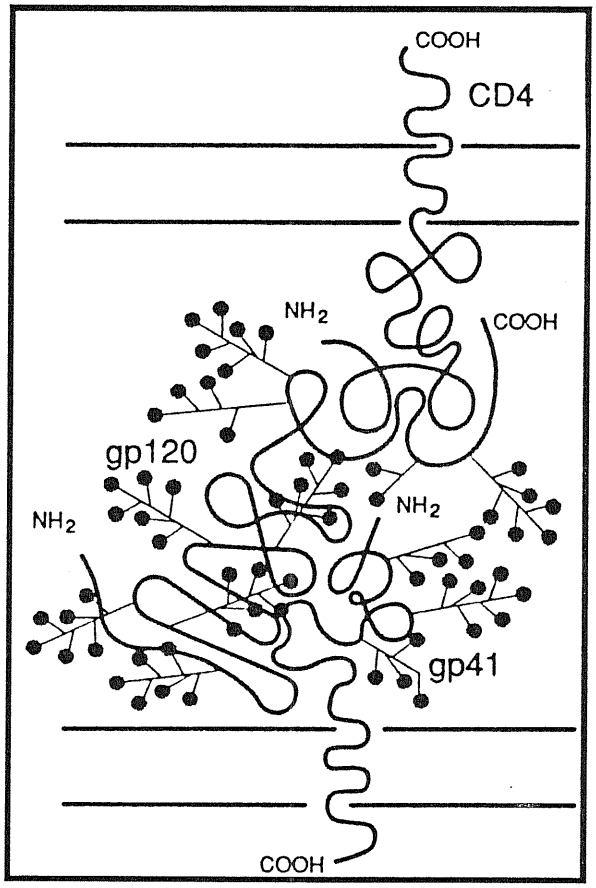
but the pathological appearance of a carefully collected set of lymphoid tissues suggest that at some point in the disease the cellular balance of the germinal center is altered. One key event is that follicular dendritic cells (FDCs) become infected with HIV and begin to die. When FDCs begin to decrease there is a concomitant rise in the level of circulating infectious virus and a decrease in the amount of detectable viral RNA in the germinal centers. Involution of the germinal centers is accompanied by dramatic drop in CD4+ T cells and, at the clinical level, this causes appearance of opportunistic infections and cancers.



Virus entry into the cell

The first step required for HIV-1 infection is the virus binding to the CD4 receptor; this event however is not sufficient for internalization, since murine cells transfected with CD4 cDNA remain resistant to infection [Maddon 1986]. Additional intermolecular events are required for entry

into permissive cells such as CD4 phosphorylation [Fields 1988], interaction with a cell surface, triptase related protease [Hattori 1989], and pH-independent fusion between the lipid bilayer of the virus and the plasma membrane of the target cell [Stein 1987], mediated by the hydrophobic N-terminus of gp41 [McCune 1988]. The virus can enter nonproliferating cells in vitro [Zack 1990] and the RNA genome is converted into a DNA form. Incomplete reverse transcription yields a labile replicative intermediate with an in vitro half-life of approximately one day. Since CD4+ cells in the body are usually resting, it is possible that most infections with HIV-1 will proceed to this point and then recede.



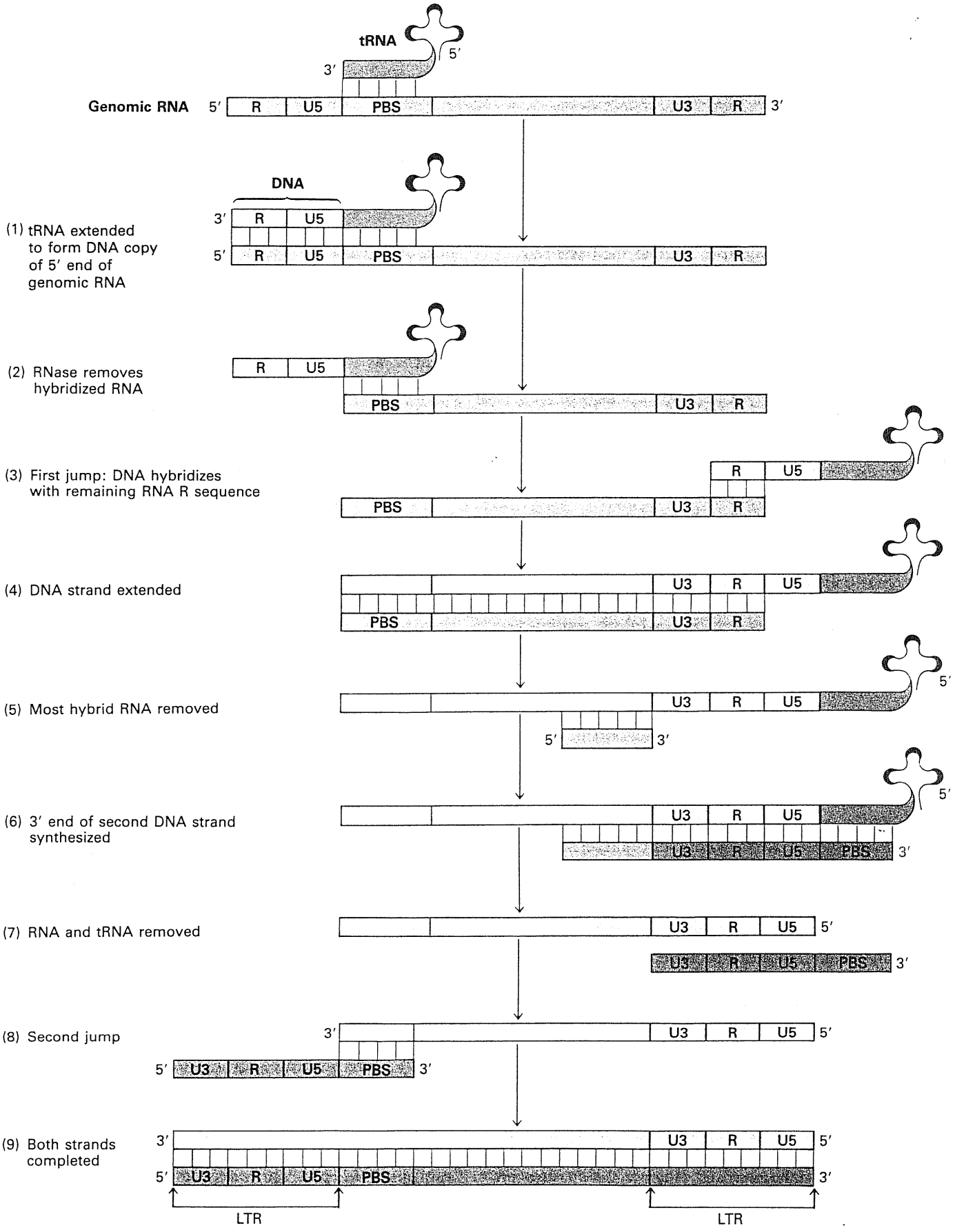
CD4-gp120 interaction

Reverse transcription

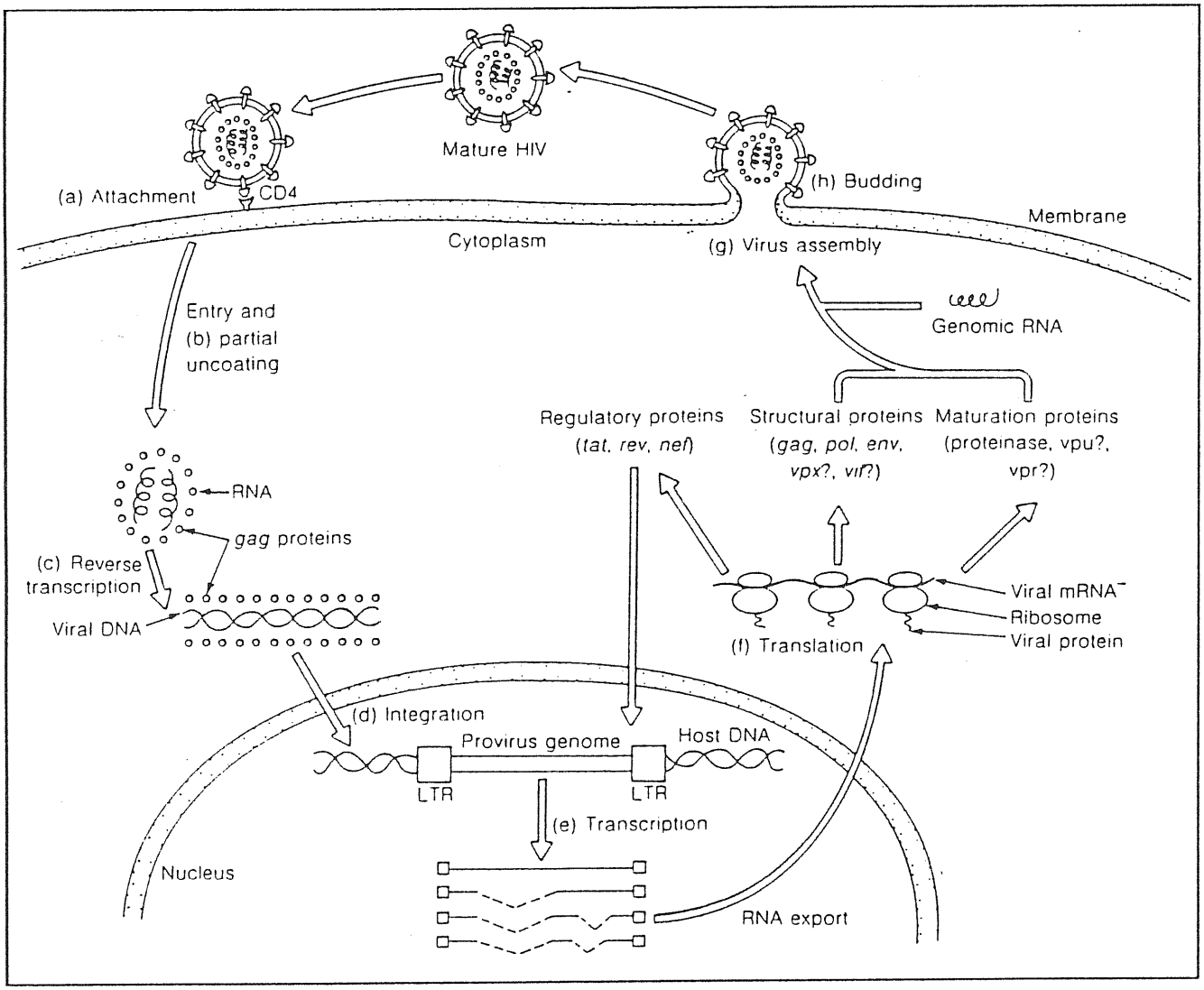
Reverse transcription of viral RNA occurs in the cytoplasm of the infected cell shortly after infection. Initiation of DNA synthesis begins as the viral polymerase adds nucleotides to the 3' end of a tRNA lysine that is hydrogen-bonded near the 5' end of the genome. An initial short DNA/RNA hybrid is formed as a product of the initial elongation reaction. The 3' portion of the RNA of the initial DNA/RNA hybrid is degraded via the action of ribonuclease H and the 3' end of the newly sinthesized DNA is free to hybridize with the 3' end of the other RNA molecule. Elongation continues until a complete copy of the genome is obtained. Second strand synthesis begins near the 3' end of the genome, at a specific sequence rich in purines that is recognized by an endonuclease activity associated with ribonuclease H. The RNA fragment produced by this cleavage is utilized as primer for the synthesis of the second DNA strand. Then RNA and tRNA are removed and a second jump brings the second strand fragment just synthesized to the 5' end where it will be used as primer to complete second strand synthesis.

The newly synthesized proviral DNA migrates to the nucleus as a nucleoprotein complex. Within the nucleus, circularization may occur either via blunt end joining of the full-lenght linear provirus, yielding a circular molecule with two adjacent long terminal repeats, or by reciprocal recombination within LTR originating a molecule with a single LTR. The substrate for the integration into the host genome is, however, the linear viral DNA. The integration process is not completely known, but it certainly requires intact terminal repeats and a functional integrase protein, as well

as cell activation. *In vitro* integration occurs within 24 hours after infection of growing T cells [Farnet 1990].



HIV-1
Replication cycle



Regulation of viral expression

Activation of transcription of the integrated provirus requires an active cell [McDougal 1985]. The balance between a latent and a productive viral infection is regulated by a complex interplay between exogenous signals, host transcriptional proteins, HIV-1 regulatory proteins, and cis-acting sequences in the HIV-1 genome, as well as by viral burden and tissue distribution. The external signals which are able to induce cell activation are the same that induce viral production.These include phorbol esters, UV irradiation, antigens, anti-CD3 antibodies, mitogens such as phytohemagglutinin and cytokines such as tumor necrosis factor, IL6, and granulocyte-macrophage colony-stimulating factor. Most of the signal transduction pathways triggered by the stimuli mentioned above converge upon the transcriptional activator NF-kB. A detailed description of transcription regulation is presented in a subsequent chapter (pag. 47).

A lot of effort has been put in order to understand the role of cofactors in the development of the AIDS disease, among these are mycoplasma, herpes simplex virus, human herpes virus 6, cytomegalovirus, Epstein-Barr virus, papovaviruses, hepatitis B viruses and human T cell leukemia virus type 1 and 2.

Their role in AIDS pathogenesis is articulated and not always clear, in the case of mycoplasma (fermentans, pirum, and penetrans) for example, the induction of tumor necrosis factor α secretion has been reported. TNFα induces viral expression and promotes cell death by two

distinct mechanisms: cell mediated killing by induction of cytotoxic CD8 lymphocytes and apoptosis by superantigens release and consequent deletion of matching vβ T cells. In other cases, such as Herpes viruses and Hepatitis B virus infection, a transactivation mechanism of HIV-1 expression, mediated by the factors interacting with the LTR has been suggested (see Discussion).

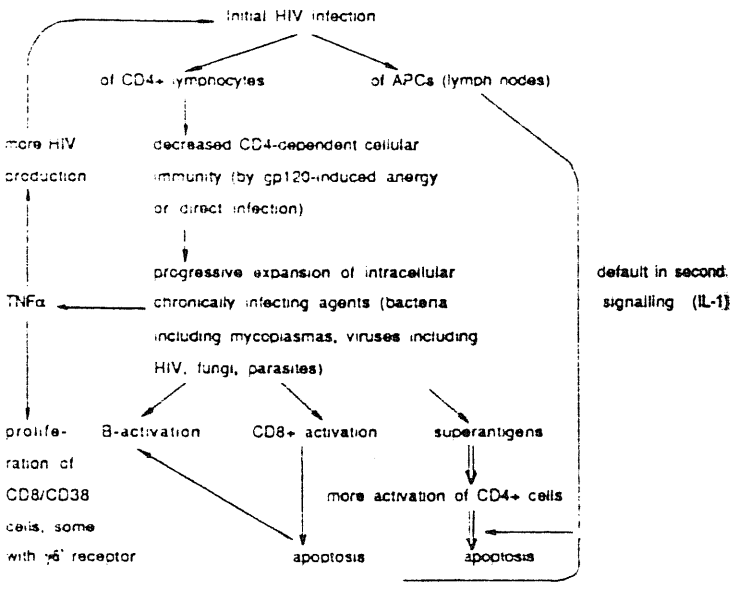
After activation, early HIV-1 gene expression is marked by the cytoplasmic expression of doubly spliced mRNA-encoding viral regulatory gene products. Among these, are tat, which functions as a strong positive transactivator of all viral genes, and rev, which plays a pivotal role in the expression of late gene functions and downregulates the expression of the early gene products. Rev activity determines a high production of viral progeny, the product of the env and gag gene are translated and provide a structural framework for encapsidation of two genomic RNA molecules [Kim 1989]. After encapsidation, the nucleocapsid protein promotes annealing of a tRNA to the primer binding site of the genomic RNA, and the virus buds from the viral membrane.

In vivo, the activated state of the T cell is transient: most cells will eventually cycle back to a resting phase and will express viral genes at a very low level, insufficient to produce infective particles. After activation, triggered by the signals mentioned above, the virus will be actively produced, and/or will induce necrosis or apoptosis of the infected cell.

Apoptosis is a physiological phenomenon of programmed cell death which plays an important role during embryogenesis and differentiation, particularly in the selective elimination in the thymus of auto-reactive clones of T lymphocytes and in the shaping of the nervous

which cleave the DNA at the internucleosomal sites, yielding DNA fragments of 200 base pairs or multiples of this unit. The apoptosis process accompanying the HIV-1 infection is certainly influenced by the environment where the infection takes place, indeed it is a typical phenomenon of infected humans and macaques, whereas HIV-1 infected mangabeys and chimpanzees, which do not develop the disease, are not interested by an induction in programmed cell death. Possible signals inducing apoptosis are: virus infection, abnormal activation via CD4 by the viral glycoprotein and subsequent activation via the T cell receptor, defaults in second signaling by antigen presenting cells, defaults in intrathymic maturations, antigens, superantigens [Newell 1990].

Superantigens, as conventional antigens, induce in vitro proliferation of normal mature CD4+ T cells, and apoptosis in normal immature thymocytes. Superantigens bound to the MHC-II molecule have specificity for the vβ part of the T cell receptor and by engaging vβ can stimulate up to 30% of the CD4+ T cells.



REGULATION OF TRANSCRIPTION IN THE EUKARYOTES

Initiation of messenger RNA transcription is a primary control point in the regulation of differential gene expression in response to developmental and environmental cues. The modulation of the extent of transcription, as well as the turning on or off of specific genes, is determined by the result of the multiple interactions between specific DNA binding proteins and the *cis* regulatory elements present both upstream and downstream of the RNA start site of a gene [Polyanovsky 1990].

In contrast to the paradigm of prokaryotic cis elements at -10 and -35 nucleotides relative to the initiation start site, in the eukaryotes there is a promoter region composed of an array of *cis* elements in the immediate vicinity of the start site, and enhancer sequences which regulate gene expression from a distance up to 30 kb in an orientation- independent manner. The promoter region is highly conserved in many cellular genes transcribed by RNA polymerase II, and typically contains elements such as TATA, GC and CCAAt boxes. Selective initiation of transcription in eukaryotes requires the specific binding of transcription factor IID to the TATA box, located at a short distance upstream of the initiation site. The binding of TFIID to its target site triggers the ordered assembly of RNA polymerase II and other general initiation factors such as TFIIA, TFFIIB, TFIIE, TFIIF on the promoter. This multiprotein complex is the final target of regulatory processes leading to transcription activation or repression [Buratowski 1989], [Greenblat 1991].

A large number of less conserved binding sites for nuclear factors are located both in the 5' untranslated region and in the introns of genes,

and their effect on transcription may be positive or negative according to the cell type, or, in a given cellular environment, according to their localization with respect to other binding sites. Specific factors allow transcriptional activation or repression as a function of a large variety of stimuli, such as heat shock, hormones and growth factors and it is their relative abundance or availability in a specific cell type which determine the differential turning on and off of genes in any given cell at any given time. Competition between factors may occur for the same binding site or for partially overlapping targets; cross-talk between bound factors may also occur with the result of a synergistic activation or the silencing of gene expression due to the block of activity of cis acting enhancers. The interplay between the various factors involved in the regulation of a specific gene confers to each gene an individualized spatial and temporal transcriptional program [Mitchell 1989].

Transcription factors are generally composed of two different and separable domains: the DNA binding domain and the transcriptional activation domain which is involved in protein-protein interactions. The specific DNA binding activities of several mammalian factors have been localized to relatively small subregions consisting of 60 to 100 aminoacids. The DNA binding domains of the known factors can be grouped in a limited number of binding motifs: the zinc fingers, the OCT-2 homeodomain, the Jun DNA binding domain, the CTF DNA binding domain. Three different types of protein domains responsible for transcriptional activation are commonly found: the GAL4 acidic domain, the Sp1 glutamine-rich domain, the CTF proline-rich domain.

Some factors bear only one of the two functions required for

transcription activation and therefore can work only in cells bearing molecules that provide the missing function. For example, the herpes virus protein VP16 bears an acidic activating region, but it is unable to interact directly with DNA. On the other hand the mammalian protein Oct1 specifically binds DNA, but it lacks an activating region. VP16 binding of Oct1 complexed with DNA triggers transcription activation, thus VP16 activates transcription only in the cells containing Oct-1 [Berger 1990].

Besides transcription activators or repressors and DNA binding activities, another activity, "the mediator", further complicates the regulation of transcription. This type of factor could function as a bridge between activators and the initiator factors at the promoter [Flanagam 1991], allowing an additional level of regulation. Recently, it was suggested that several activator proteins, including Sp1, USF, and GAL4-VP16, require a coactivator or mediator in order to stimulate transcription *in vitro*. In the case of VP16, for example, two models have been suggested [Kelleher 1990], [Berger 1990]. According to the first model, the mediator is an adaptor that couples VP16 to a general initiation factor like the TFIIB subunit of the TATA box binding factor TFIID. According to the second model, the mediator is a bridge that holds two VP16 activation domains at the correct distance apart to bind TFIIB and TBP (another TFIID subunit) simultaneously.

By competing for the same adaptor molecules, activators can inhibit transcription stimulated by a second activator. This process is known as squelching and represents one of the proposed mechanisms for transcriptional repression by a potential activator molecule. Other models to explain the inhibitory function of transcription factors have

been proposed, such as direct repression, in which the negative control factor directly blocks the activity of the basal transcription complex, competition for the same binding site or for an overlapping one, and 'quenching', that is interaction between adjacent factors which prevent productive contacts with the transcription complex [Levine 1989]. It is interesting to note that all the eukaryotic transcriptional repressor studied so far can function also as activators. Their positive or negative role may depend on their abundance relative to the general transcription machinery or to other specific factors, the localization or orientation of their binding sites respect to other transcription elements, and on the tertiary structure of DNA in their target sites [Jones 1991].

An intriguing chapter of the eukaryotic transcription factors story is their involvement in the regulation of DNA replication. Activation of DNA replication by enhancers and other transcription elements appears to be a quite general feature of the replication of viruses infecting eukaryotic cells [Guo 1992]. Moreover, a transcription activator of many yeast genes, OBF1, has been shown to be also an enhancer of DNA replication in *S. cerevisiae* [Walker 1990]. Interestingly, many mammalian factors function both in transcription activation and in DNA replication initiation (see table at page 46). However, little is known about the mechanism of activation of DNA replication origins by transcription factors, and even less about their putative role in the regulation of DNA replication in higher eukaryotes. Some hints on transcription-replication coupling come from studying the relation between the time of replication and gene expression. It has been shown, both in yeast [Rivier 1992] and higher eukaryotes [Hatton1988],

that actively expressed genes are replicated early in the S phase. On this basis it has been suggested that the presence of specific transcription factors might be responsible for origin activation, or alternatively, activation of specific origins may allow preferential capture of limiting transcription factors by newly synthesized DNA and thus trigger expression of specific genes [Riggs1992].

Some of the work presented in the experimental part of this thesis deals with the characterization of transcription signals present in a human origin of DNA replication.

MAMMALIAN PROTEIN FACTORS AFFECTING REGULATION OF TRANSCRIPTION AND OF DNA REPLICATION

FACTOR	*SEQUENCE MOTIF BOUND*	*TRANSCRIPTION*		*REPLICATION INITIATION*	
		GENE	FUNCTION	GENOME	FUNCTION
Sp1	CCGCCC	late genes of papova many housekeeping genes	upstream element binding	Papovavirus	stimulation of origin activation
T Antigen	GAGGC	early genes of papova	regulation of transcription	Papovavirus	essential for origin activation
CTF/NF-I	GCCAAT	globin genes, genes stimulated by TGF-beta and many RNA pol II - transcribed genes	upstream element binding	Adenovirus 2 and 4	essential for initiation
NF-III	ATGCAAAT	H2B histone genes Immunoglobulin genes U1 and U2 snRNA genes SV40	upstream element binding upstream element and enhancer binding enhancer binding	Adenovirus 2 and 4	essential for initiation
MLTF/USF	PuPyCACGTGPuPy	Adenovirus late genes mouse metallothionein gene rat gamma-fibrinogen gene HIV-I LTR	upstream element binding negative regulation (?)	Human (HL60)	(?)

HIV-1 TRANSCRIPTION REGULATION

Like all viruses, HIV is an intracellular parasite. Beginning with its entry into its host cell by way of the CD4 glycoprotein, the virus interacts with many cellular proteins to complete its life cycle. The importance of cellular host-cell factors is particularly apparent in the process of transcription, which is regulated by the synergic action of the viral protein Tat and of a number of nuclear factors interacting with the Long Terminal Repeats of the virus. These cellular factors play a crucial role in the decision of the virus fate after integration. At this time the cell lacks any Tat activity and therefore relies completely on cellular factors for transcription activation. The Long Terminal Repeats are completely occupied by target sites for specific binding proteins, which have been identified by *in vitro* binding studies and characterized by transient transfection experiments.

The **TATA box** homology, functioning to correctly initiate the start of transcription by serving as the binding site for the TATA box DNA binding protein TFIID, is located at position -22 with respect to the start site of transcription. Very recently, the cloning of a gene encoding for a TATA element modulatory factor (TMF) has been reported. Interestingly TMF binds *in vitro* to the HIV-1 TATA box, and inhibits transcription of the viral LTR as shown by in *in vitro* transcription assays [Garcia 1992].

Two sites surrounding the transcriptional initiation site are the target for the cellular protein **LBP-1** [Jones 1988], [Kato 1991]. LBP-1 binding to the downstream site (-16,+27) enhances transcription, whereas

interaction with the upstream site, which overlaps the TATA box, causes transcriptional repression *in vitro* [Kato 1991].

Immediately upstream from the TATA box are three tandem, closely spaced G+C rich rich sequences (-46 to -68) which bind the purified cellular transcription factor **SP1** *in vitro,* with variable affinity. Point mutations within the SP1 binding sites, particularly at site two and three, greatly reduce the rate of transcription both *in vitro* and *in vivo*, indicating a role for these binding sites in HIV-1 transcription regulation. [Jones 1986]. Moreover, multiple SP1 binding sites confer enhancer-independent activation on the HIV-1 promoter [Proudfoot 1992].

The **enhancer** region, extending between -108 and -80 in the HIV LTR, has been the subject of intense investigation. This region contains two NF-kB motifs, which are also found in a number of other viral and cellular enhancer elements, including the cytomegalovirus, simian virus 40, Immunoglobulin K chain, and IL-2 enhancers. Mutagenesis of the HIV enhancer region results in decrease in both basal and Tat-induced gene expression from the HIV LTR in a variety of different cell lines [Nabel 1987]. Nevertheless, NF-kB motif mutations inserted into proviral constructs do not result in defects in viral growth in many cell types [Leonard 1989], except that in activated T cells [Harrich 1990]. These findings suggest that viral regulatory proteins or /and other *cis* acting regulatory elements in the wild type provirus may substitute for the enhancer element in the regulation of HIV gene expression.

A number of different cellular factors, including the prototype NF-kB [Sen

1986] [Kretzschmar 1992], EBP-1 [Clark 1988], H2TF1 [Baldwin 1988], KBF1 [Israel 1987], and HIVen 86 [Franza 1987] have been demonstrated to bind to the NF-kB sites. Recent studies indicate that NF-kB proteins, including HIVEN86, comprise a family of proteins, related to the proto-oncogene c-rel [Ballard 1990], [Kieran 1990], [Nolan 1991]. This family of proteins is evolutionarily conserved and a rel-related protein known as dorsal is also found in Drosophila cells [Steward 1987]. The genes encoding several of these rel-related proteins, including p50, p65, and p80, have been cloned and shown to be highly conserved in their amino-terminal DNA binding and dimerization domains. Another cellular factor known as PROII-BF1/ MBP-1 also binds to NF-kB motifs [Baldwin 1990], [Fan 1990]. This 300-kDa protein contains two widely separated zinc finger domains each of which bind NF-kB motifs. Like the rel family members, factors related to PRDII-BF1 have been identified in both human and rat tissues. However, PRDII-BFI derived cDNAs does not result in stimulation of either basal or tat-induced activated gene expression, suggesting that this factor could bind constitutively to the NF-kB sites, and then be replaced by c-rel family members following activation [Muchardt 1992].

Homodimers and heterodimers between different members of the rel family can be formed. In the original model, NF-kB is an heterodimer between p50 and p65 and resides in an inactive form complexed with the IkB inhibitor, via the p65 subunit, in the cytoplasm of infected T cells and some monocytes. Cellular activation by mitogens or tumor necrosis factor α releases the active DNA-binding complex, by disactivating the Ik-B inhibitor through its phosphorylation [Bauerle 1988], [Ruben 1992].

Recent data have complicated a lot this apparently simple regulatory scheme. Homodimers of p50 (equivalent to KBF1 [Kieran 1990]), p65, p49 and c-rel (equivalent to HIVEN-86 [Franza 1987]) have been shown to bind DNA and activate transcription *in vitro*. However, *in vivo,* only the p65, p49, and c-rel homodimers are able to transactivate in a fashion similar to the p50-p65 heterodimer [Fujita 1992]. The p50 homodimer, which is abundant in the nucleus of resting T cells and of immature monocytes, on the other hand, inhibits transactivation by displacing the activating heterodimer. Furthermore, stimulation by cytokines and phorbol esters does not only have a post-transcriptional activation effect, but also induces the expression of the gene encoding for p50, as well the expression of other members of the rel family, such as for example the inhibitor I-rel. This complex regulatory network reflects a need for the cell to control the action of NF-kB and related factors, given their pleiotropic functions [Lenardo 1989].

A number of disparate signals besides cytokines and TPA can trigger NF-kB activation and viral expression. These include UV, peroxides [Schreck 1991], and HIV-1 infection itself [Bachelerie 1991]. In the U937 promonocytic cell line, infection by HIV-1 has recently been shown to be a sufficient signal for NF-kB stimulation, through the action of the viral protease on p50 precursor, p105. The active NF-kB will in turn promote viral expression generating an activation loop which allows viral perpetuation.

To complicate further the scene, an unrelated protein, Bcl3, has been shown to participate in the NF-kB story [Franzoso 1992]. Bcl3 can bind p50 and has diametrically opposite effects according to its relative

concentration with respect to p50. If present in large amount it can compete with p65 and inhibits p50/p65 mediated transactivation. On the other hand, a small amount of Bcl3 works as antirepressor by blocking p50 mediated inhibition.

The region upstream of the enhancer has been named **NRE**, negative regulatory element, because its overall effect is to downregulate viral transcription both *in vivo* and *in vitro*. It has been demonstrated that the deletion of the region upstream of nucleotide -167 results in an increase in gene expression promoted by the downstream domains in CAT assay experiments [Rosen 1985]. Deletion of the same region markedly increases viral replication both in the Jurkat T cell line and in the monocytic U937 cell line [Lu 1989].

In the NRE a number of target sites for nuclear factors has been found. Some of the factors have been shown to have an activating function, others a negative function, and interestingly, a third group seems to have either positive or negative function according to the cellular environment. A compilation of the known NRE binding factors, including the location of the target site respect to the transcription start site, the cell line used for characterization studies, the putative function, and the bibliographic reference, is shown on the table reported at page 53. Looking at the table, two remarkable features immediately emerge: the abundance of identified factors target sites present in a relatively short region of about 250 base pairs, and the large number of questions marks in the column indicating the function. On the basis of these data two major questions immediately arise:

1) Which binding sites are actually occupied in the living infected cell?

2) Which factors are directly involved in the activation or repression of HIV-1 transcription?

These two open questions are the reason of a large part of the work presented in this thesis.

PROTEIN	TARGET SITE	SYSTEM	FUNCTION	REFERENCE
USF/MLTF	-174 -152	Jurkat	activator	Maekawa 1991
		HeLa, Jurkat	repressor	Lu 1990, Giacca 1992
NFAT-1	-254 -216	activatedT cells	?	Shaw 1988
site B protein	-350 -327	Jurkat	repressor	Orchard 1990, 1992
site A protein	-379 -361	Jurkat	?	Orchard 1990
AP1	-342 -349 -330-338 -291 -299	HeLa	?	Franza 1988 Yamamoto 1991
A1	-315 -265	HeLa	?	Guy 1990
ILF	-283 -295	HeLa	?	Li 1991
GRE	-276 -235	vitro	?	Ghosh 1992
Myb	-151 -156	HeLa	activator	Dasgupta 1990
URE binding factor	-157 -122	MOLT-4	?	Nakanishi 1991
		Jurkat	repressor	
		MT4	activator	
NRE binding factor	-320 -295	H9	repressor	Yamakoto 1991
Sp-50	-400 -368	YT-1	?	Smith 1989
GRE	-235 -276	purified factor	?	Ghosh 1992
IBF	-283 -195	HeLa	?	Li 1991

The outcome of the multiple interactions between the LTR and the above mentioned factors and the interplay among different proteins is the fine modulation of the HIV-1 expression.

The accurate determination of the expression of the HIV-1 genome in the human host is essential for understanding the natural history of HIV-1 disease progression in individuals, and assessing the efficacy of antiviral intervention strategies. HIV-1 expression and replication occur during the entire course of natural infection, therefore the period of clinical latency is not correlated with a true latency at the molecular level. However, HIV-1 expression at the level of viral DNA and RNA increases with the decline in CD4+ T cell count and the progression of disease *in vivo* [Schnittman 1989], [Michael 1992]. The ratio between the RNA copy number of a specific gene and the viral DNA copy number increases 100-1000 fold during the course of natural infection [Michael 1992]. The central aim of the experimental work presented in this thesis is to contribute in unraveling the molecular mechanisms underlying the dramatic transcription increase, which is observed in the course of HIV-1 disease. Understanding the regulatory processes involved in the progression of HIV-1 disease from a phase of clinical latency to AIDS might be important for future development of prophylactic or therapeutic intervention strategies.

RESULTS

PROBING PROTEIN-DNA INTERACTIONS AT THE HIV-1 LONG TERMINAL REPEAT BY *IN VIVO* FOOTPRINTING IN HIV-1 INFECTED H9 CELLS

After the integration of the HIV-1 virus in the genome of the host cell, the first control point of viral production is the regulation of transcription. In the first phase of the viral cycle, before the production of the viral transactivator Tat, the regulation of transcription relies completely on nuclear factors of the host cell. The region of HIV-1 genome involved in provirus transcription regulation is the Long Terminal Repeat located at the 5' end of the HIV-1 genome. As it is shown in the scheme below, the LTR is a mosaic of binding sites for nuclear proteins, most of which are able to bind nuclear proteins *in vitro.* Their occupancy *in vivo* however, is affected also by other parameters, such as chromatin structure, DNA conformation, and interactions with other proteins.

To identify which are the DNA sites actually occupied *in vivo* in a living chronically infected HIV-1 producing cell line, we analyzed H9 cells infected with HTLVIIIB by *in vivo* dimethylsulfate (DMS) footprinting using the ligation-mediated polymerase chain reaction (LM-PCR) technique [Mueller 1989]. This technique, in combination with *in vitro* studies, and transient transfection assays, allows the understanding of the complex mechanisms of transcription regulation. A flow-chart of DMS in vivo footprinting is outlined in the following page. Figure 2 is a cartoon of the

Long Terminal Repeat with the binding factors described in the literature, and shows as well the sequence and location of the primers used in this study. The patterns of DMS methylation of upper, coding strand (obtained using primer sets A and B) and lower, non coding strand (primer sets C and D) are shown in Figure 3 and 4, respectively, and the results summarized along the LTR sequence in Figure 5. The guanines with altered DMS sensitivity indicated in the Figures are those which have been reproducibly detected in at least three independent experiments. In Figures 3, 4 and 5, the regions of the LTR recognized as binding sites for nuclear proteins by *in vitro* experiments are also shown alongside.

In H9/HIV-$1_{HTLVIIIB}$ cells, supporting a highly productive viral infection, major footprints appear at the basal promoter and enhancer elements; in particular, proteins appear to sit over the TATA box, the three Sp1 sites and the two repeats of the enhancer region (primer sets A and D). Most of the bases protected or hypersensitive to DMS methylation in this region correspond to those detected by *in vitro* methylation interference studies [Jones 1988]], [Kawakami 1988]. No sites in close proximity or downstream of transcription start site appear to contact specific proteins by this technique, on the contrary of what reported on the basis of studies *in vitro* [Kato 1991]. In the Negative Regulatory Element, protections appear over the USF/MLTF site on both strands (primer sets A and C), and over the binding site for NFAT-1, mainly on the coding strand (primer set B); in the same region, the two adenines at positions -247 and -249 appear clearly hypersensitive. Although the piperidine treatment used mainly favors cleavage at alkylated guanines, it may also reveal, with reduced sensitivity, adenine residues that are particularly reactive

with DMS [Mueller 1989]. Few protections are also visible in the region from nucleotides -285 to -315 (primer sets B and C), where the binding sites for several proteins are located, including the AP-1 factor, a factor named A1 and where purified c-myb protein was shown to interact *in vitro*. Finally, two protected guanines are evident at positions -331 and -338 on the non coding strand (primer set C), in a region containing a palindromic sequence with homology to the steroid/thyroid hormone response element.

Very interestingly, two novel sites, located between nucleotides from -260 to -275 and from -204 to -216, respectively, appear to be involved in protein-DNA interactions *in vivo* (primer sets B and C). Both sites are purine-rich and share the common sequence 5'-GGAGAGA-3' (positions -269 and -216, respectively, on the coding strand), with the first two guanines protected from DMS methylation. Adenines at position -274, -262 and -260, surrounding the consensus sequence of the upstream site, are hypersensitive. Some preliminary studies on the characterization of the upstream site, which we called SISSA, are reported in a following chapter.

Finally, the guanine at position -132 on the coding strand (primer set A) is clearly hypersensitive, possibly as a consequence of interactions of factors bound on the LTR both upstream and downstream of this location.

The study presented in this chapter has already been published [Demarchi 1992].

LIGATION-MEDIATED POLYMERASE CHAIN REACTION

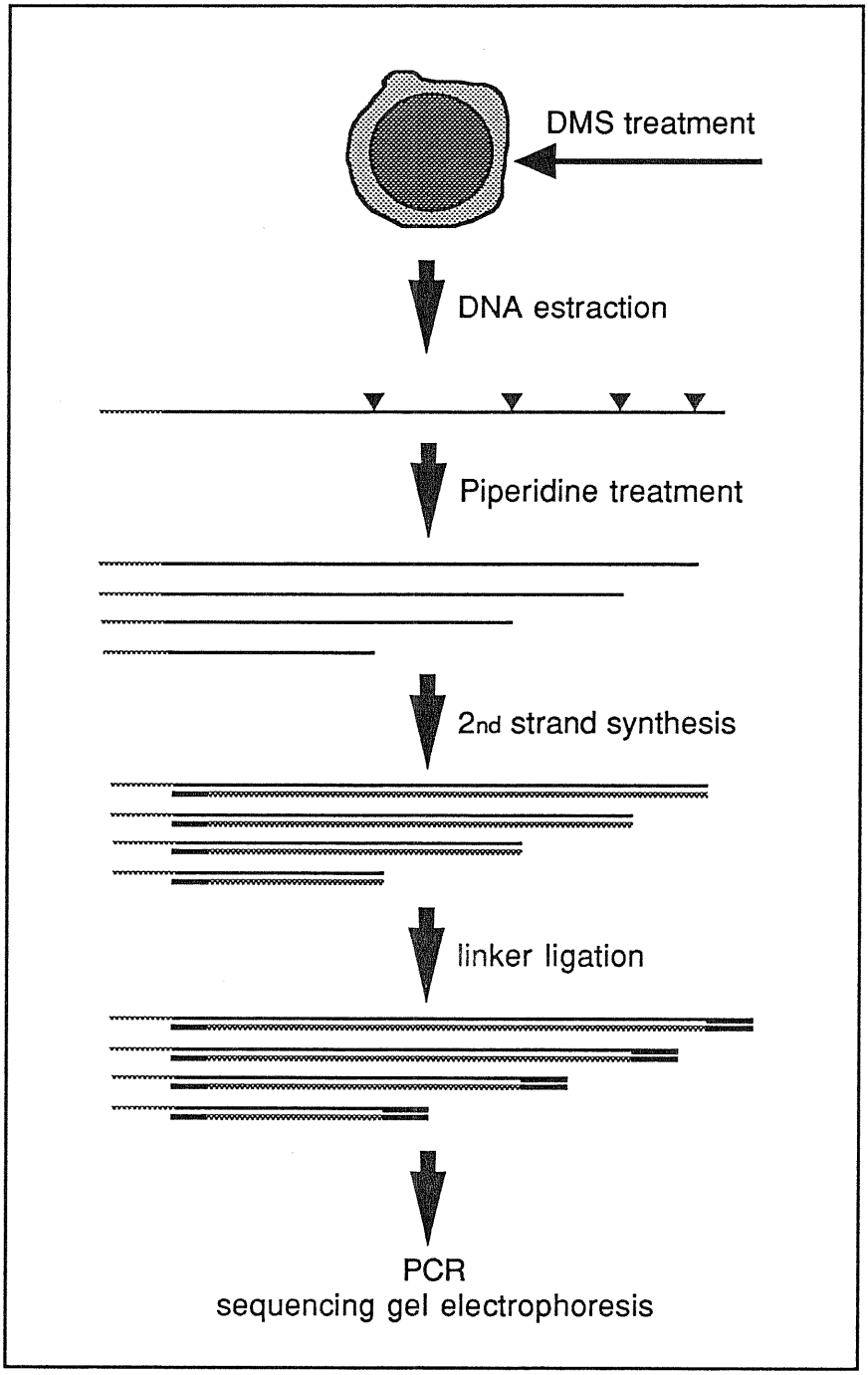
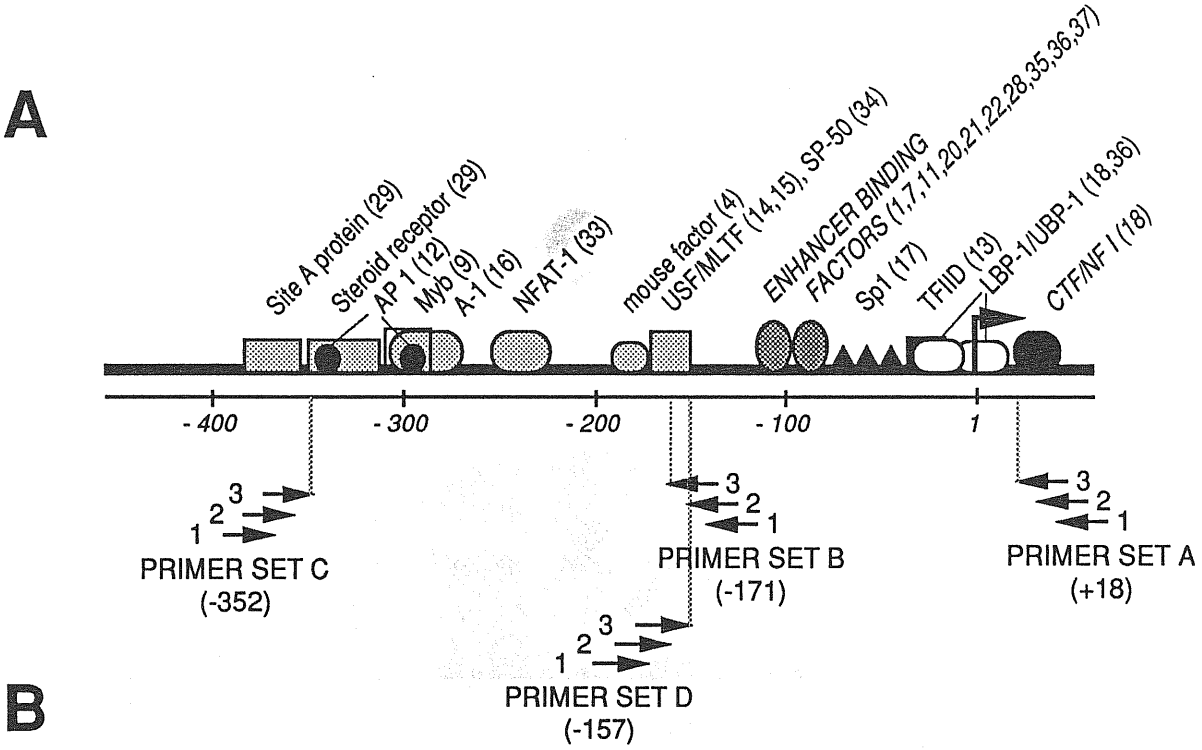


FIGURE 1



PRIMER SET A

1. 5'-TATTGAGGCTTAAGCAGTGGGTTCC-3'
2. 5'-CCCTAGTTAGCCAGAGAGCTCCCAG-3'
3. 5'-GCCAGAGAGCTCCCAGGCTCAGATCTG-3'

PRIMER SET B

1. 5'-AGCAAGCTCGATGTCAGCAGTTCTT-3'
2. 5'-GAAGTACTCCGGATGCAGCTCTCGG-3'
3. 5'-GGATGCAGCTCTCGGGCCACGTGATGA-3'

PRIMER SET C

1. 5'-AGGCTACTTCCCTGATTAGCAGAAC-3'
2. 5'-TAGCAGAACTACACACCAGGGCCAG-3'
3. 5'-TAGCAGAACTACACACCAGGGCCAGGG-3'

PRIMER SET D

1. 5'-AGTGTTAGAGTGGAGGTTTGACAGC-3'
2. 5'-GCCGCCTAGCATTTCATCACGTGGC-3'
3. 5'-CCGCCTAGCATTTCATCACGTGGCCCG-3'

FIGURE 2 Schematic representation of the HIV-1 LTR and sequence of the oligonucleotide primers utilized for this study.

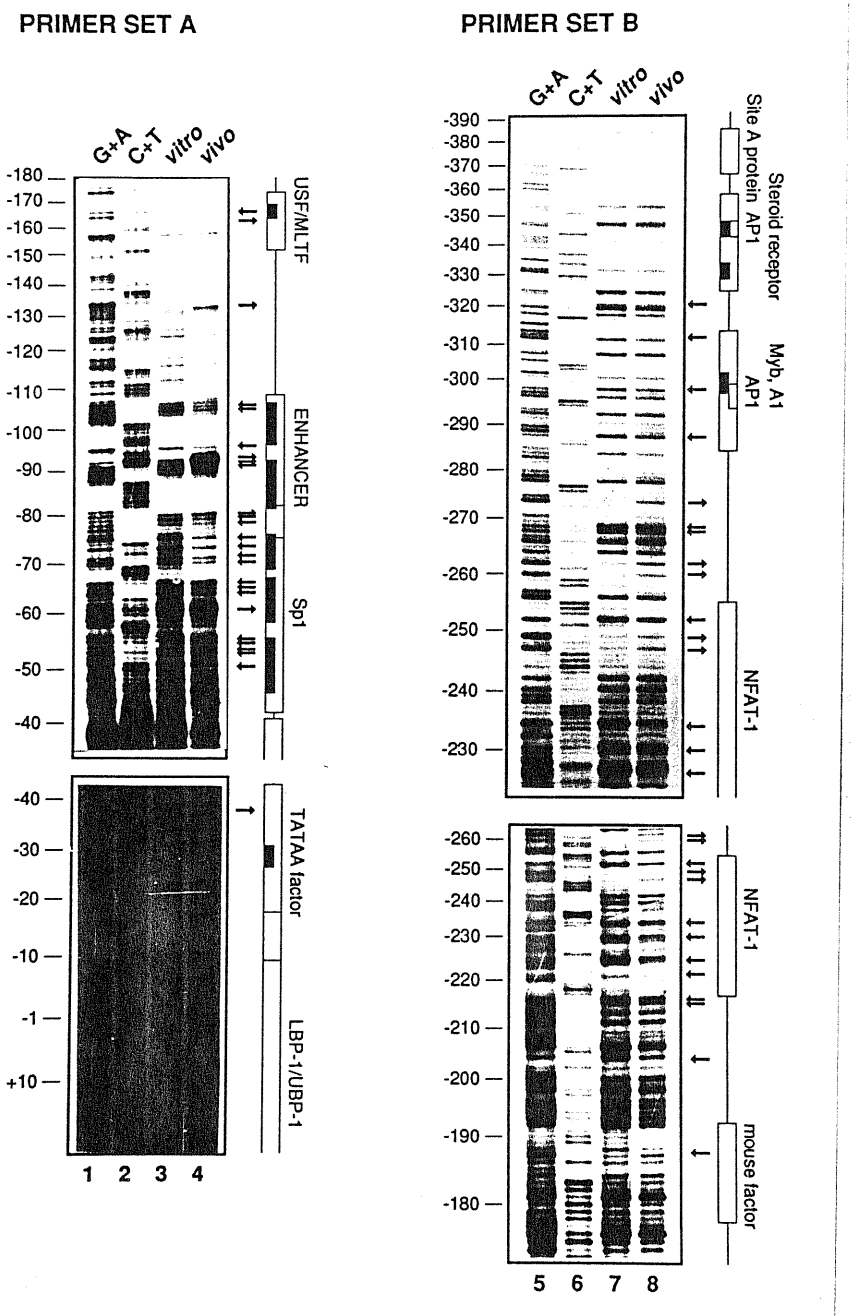


FIGURE 3: Genomic footprinting of HIV-1 LTR in HIV infected H9 cells. Coding strand

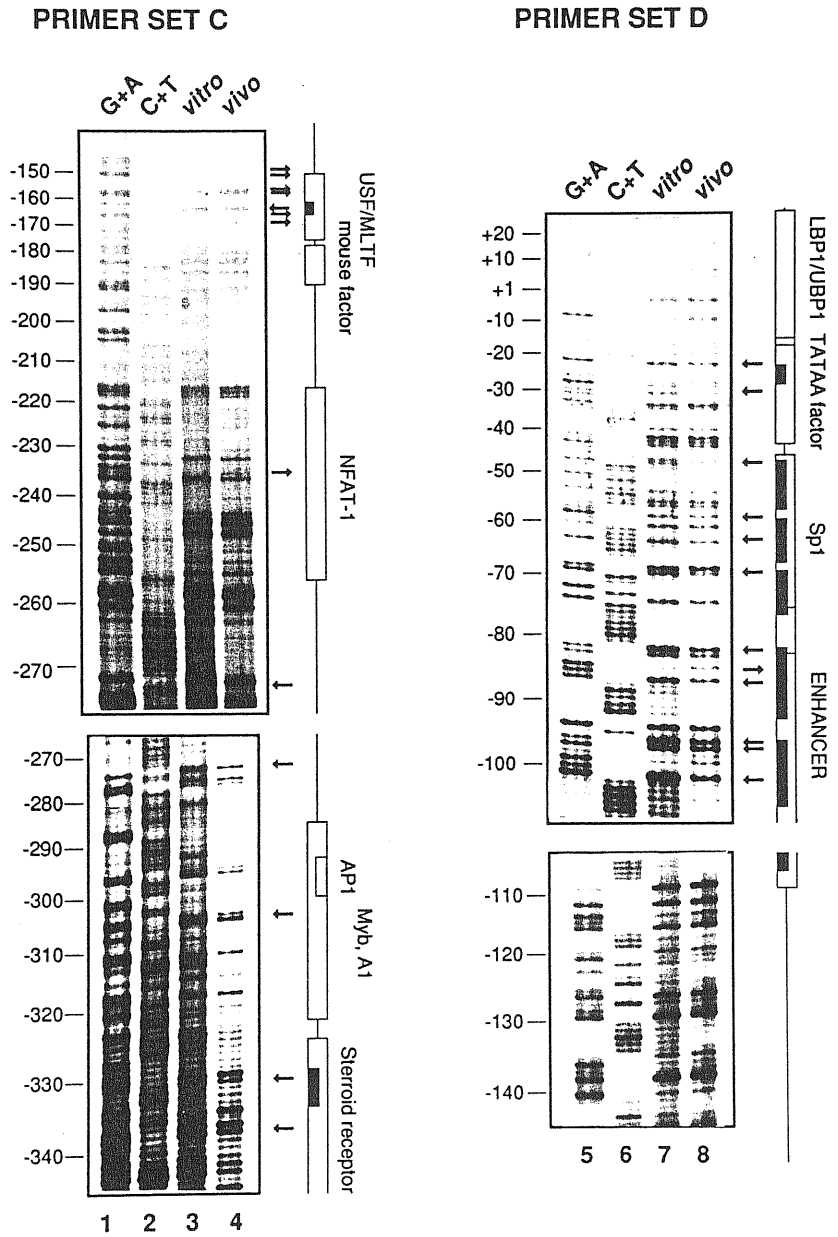
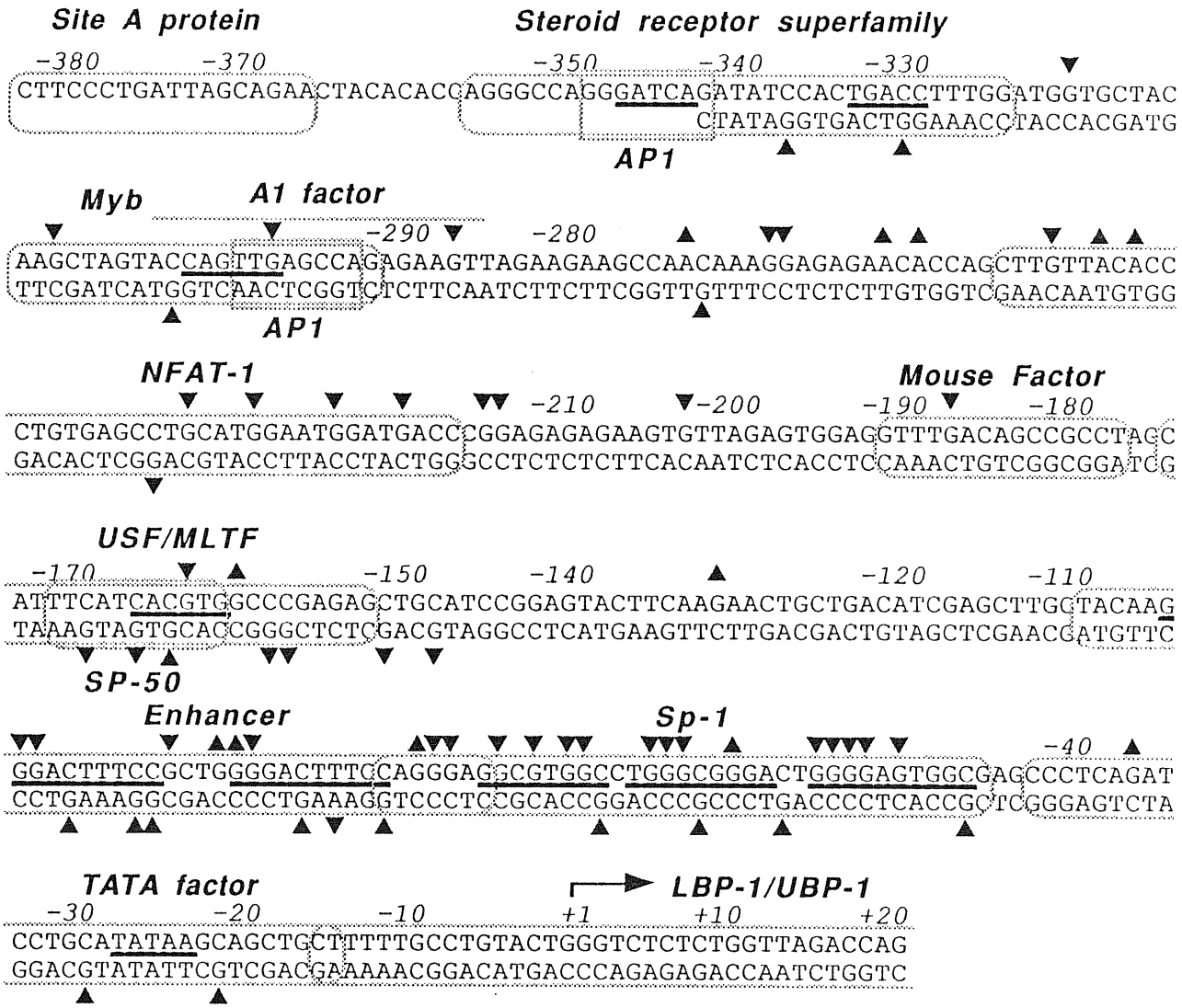


FIGURE 4: Genomic footprinting of HIV-1 LTR in HIV infected H9 cells. Non coding strand

-390
AGGCTA


Site A protein
Steroid receptor superfamily
-380
-370
-350
-340
-330
CTTCCCTGATTAGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGCTAC
CTATAGGTGACTGGAAACCTACCACGATG
AP1
Myb
A1 factor
-290
-280
AAGCTAGTACCAGTTGAGCCAGAGAAGTTAGAAGAAGCCAACAAAGGAGAGAACACCAGCTTGTTACACC
TTCGATCATGGTCAACTCGGTCTCTTCAATCTTCTTCGGTTGTTTCCTCTCTTGTGGTCGAACAATGTGG
AP1
NFAT-1
Mouse Factor
-210
-200
-190
-180
CTGTGAGCCTGCATGGAATGGATGACCCGGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGC
GACACTCGGACGTACCTTACCTACTGGGCCTCTCTCTTCACAATCTCACCTCCAAACTGTCGGCGGATCG
USF/MLTF
-170
-150
-140
-120
-110
ATTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAG
TAAAGTAGTGCACCGGGCTCTCGACGTAGGCCTCATGAAGTTCTTGACGACTGTAGCTCGAACGATGTTC
SP-50
Enhancer
Sp-1
-40
GGACTTTCCGCTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGAT
CCTGAAAGGCGACCCCTGAAAGGTCCCTCCGCACCGGACCCGCCCTGACCCCTCACCGCTCGGGAGTCTA
TATA factor
LBP-1/UBP-1
-30
-20
-10
+1
+10
+20
CCTGCATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAG
GGACGTATATTCGTCGACGAAAAACGGACATGACCCAGAGAGACCAATCTGGTC

ACTIVATION OF HIV-1 TRANSCRIPTION IN U1 CELLS

The high complexity of the HIV-1 disease complicates a lot the search for suitable models of HIV-1 clinical latency, yet there are cultured cell lines that harbor the viral genome, but produce only small numbers of viral progeny [Folks 1987], [Clouse 1989]. These cell lines can be induced by cytokines or phorbol esters to produce much larger quantities of virus and, therefore, have been proposed as models of clinical latency [Pomerantz 1990]. One of such cell lines, the U1, is a derivative of the U937 monocyte/macrophage cell line, and has been shown to be inducible by 13-phorbol-12-myristate acetate (PMA, TPA) [Folks 1987]. The induction of HIV-1 production has been shown to be preceded by a dramatic increase in viral mRNA production [Pomerantz 1990]. In order to define the molecular mechanisms responsible of this up-regulation, we decided to look for differences in protein-DNA interactions at the HIV-1 LTR before and after induction of viral expression. To this end we decided to analyze the LTR of U1 cells in a time course experiment of stimulation with TPA. U1 cells treated with TPA were harvested at subsequent time points, (0, 1, 2, 3, 6, 24 hours) after addition of the stimulus. From each sample an aliquot was used for RNA extraction, and an aliquot was used for *in vivo* DMS footprinting. Furthermore, the supernatants were utilized to determine the amount of the viral protein p24. P24 in the supernatant was measured also three days after TPA addition. As indicated in the Table below (fig.6), p24 antigen increases exponentially during the first 24 hours, and then at a linear rate in the following hours.

To examine the kinetics of modification of the HIV-1 RNA

expression pattern over time, total RNA from TPA stimulated cells was analyzed by Northern blot with a probe specific for the LTR, which is contained in all the different classes of viral transcripts. As it is evident in figure 6, all the three classes of viral mRNAs are already detectable, although at a very low level, in samples extracted from unstimulated cells (T0). The concentration of the 2Kb RNA species gradually increases over time, while the 4 kb class starts to increase only after 3 hours and the 9 kb class only after 6 hours. Interestingly, the relative abundance of the different classes undergoes two subsequent shifts. In the beginning, the most abundant class is the 2kb, at 6 hours the 9 kb becomes the most abundant, and at 24 hours the 4 kb becomes the second most abundant, while the 2 kb class, even if still increasing, is become the less abundant of all three classes. Most interestingly, this pattern of RNA expression is analogous to the the RNA expression described during a one step growth of HIV-1 [Kim 1989]. During the first phase of infection the 2kb. class is predominant; later, in presence of a sufficient quantity of the Rev protein, the unspliced (9kb), and singly spliced (4kb.) messengers start increasing and become predominant in the last phase of infection.

ACTIVATION OF HIV-1 TRANSCRIPTION BY TPA IN U1 CELLS

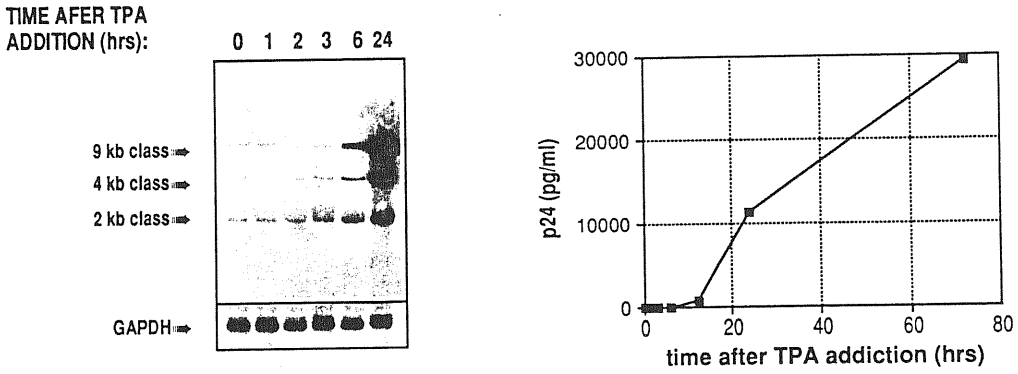


FIGURE 6

CONSTITUTIVE AND INDUCIBLE PROTEIN-DNA INTERACTIONS AT THE HIV-1 LTR ENHANCER EXAMINED BY *IN VIVO* FOOTPRINTING

In order to identify which are the DNA sites actually occupied *in vivo* in the U1 cell line and to discover which are the sites involved in transcription activation, we analyzed U1 cells by *in vivo* dimethylsulphate footprinting in a time course experiment of stimulation with TPA (see previous chapter). For a description of the tecnique and of the primer used for this study see fig. 1 and 2.

The patterns of DMS methylation of the upper, coding strand (obtained by using primer sets A and B) and lower, non coding strand (primer set C) are shown in fig. 7, 8 and fig.9 respectively, and the results are summarized along with the LTR sequence in fig. 10. The locations of the bands with respect to the transcription start site are indicated at the left of the photograph. Protected guanines are shown by arrows pointing leftward and hypersensitive guanines are indicated by arrows pointing rightward. The first two lanes are sequencing reactions (G+A and C+T) which allow the localization of the bands along the LTR. The third lane is the naked DNA control, that is *in vitro* treated DNA from U1 cells. The next following lanes are *in vivo* treated samples from cells stimulated with TPA for 0, 1, 3, 6, and 24 hours respectively and correspond to the same time points used for the Northern blot and the p24 quantification described in the previous chapter.

The first interesting result shown in fig. 7 is that also in untreated cells, where the mRNA is expressed at basal levels (see previous

chapter), major footprints appear over the three Sp1 sites, and the two repeats of the enhancer region. Upon TPA stimulation, besides the constitutive footprints, a striking enhancement of hypersensitivity at a site located at the border of the enhancer site adjacent to the Sp1 boxes occurs gradually during the course of stimulation.

Figure 8 shows the results obtained studying the coding strand of the negative regulatory element. Protections appear over the NFAT-1 site and over the SISSA binding site, and the pattern remains unchanged also after stimulation with TPA.

In the non coding strand of the negative regulatory element, shown on fig. 9 , there are not any significant alterations in the sensitivity to DMS in the *in vivo* treated samples as compared to the *in vitro* controls. A constitutively protected region, visible in the upper part of the gel, is present in a sequence adjacent to the enhancer. Interestingly, the sequence corresponding to the core of this site -TGACATCGA-, resembles the binding site for the well known CREB nuclear factor, belonging to the ATF family of transcription factors. The results of this study are summarized along the LTR sequence on fig. 10.

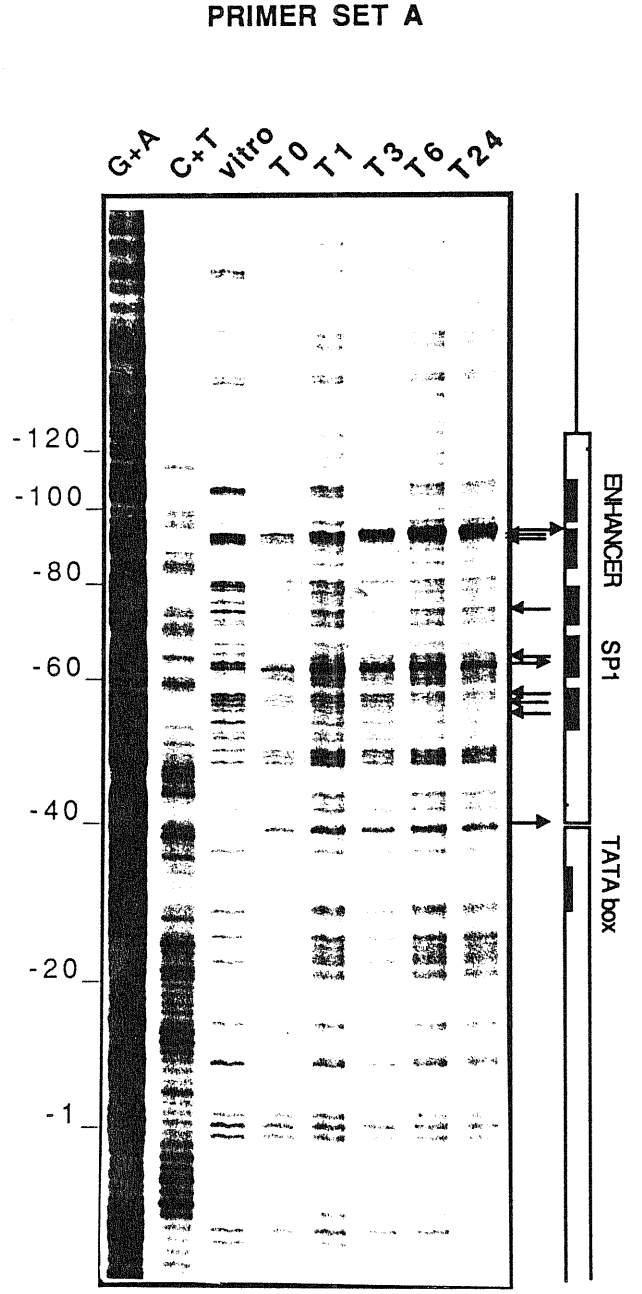


FIGURE 7: Time course activation of U1 cells by TPA analyzed by in vivo DMS footprinting. Coding strand of HIV-1 LTR

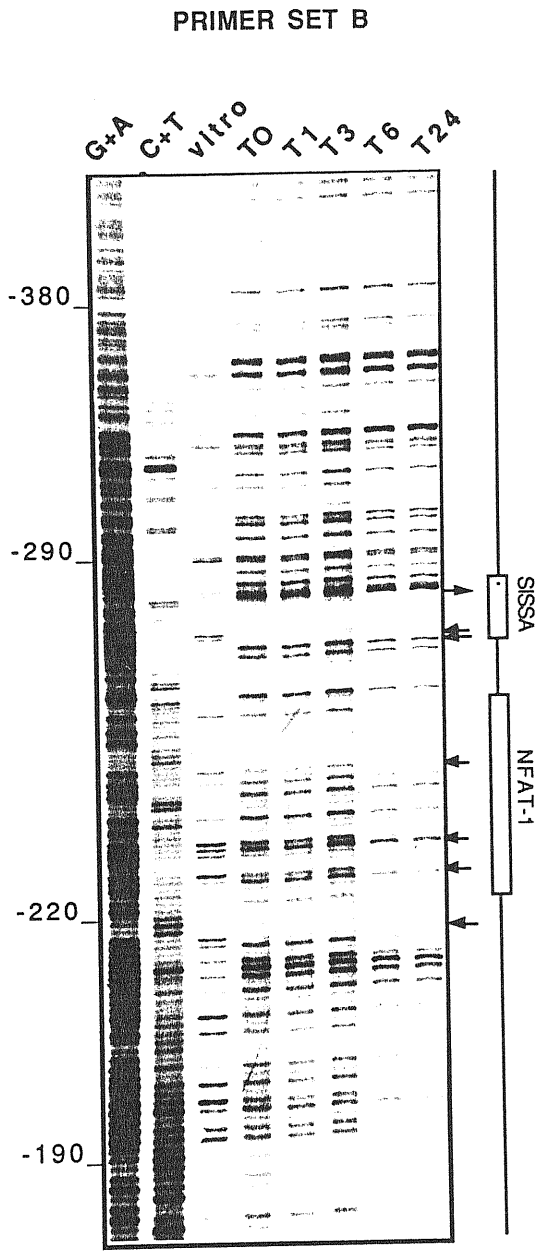


FIGURE 8: Time course activation of U1 cells by TPA analyzed by in vivo DMS footprinting. Coding strand of HIV-1 LTR

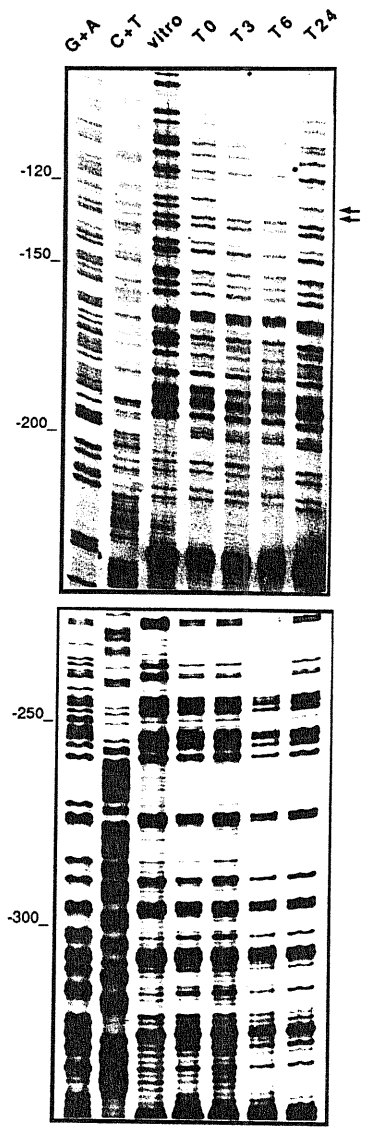


FIGURE 9 : Time course activation of U1 cells by TPA analyzed by in vivo DMS footprinting. Non coding strand of HIV-1 LTR

IN VIVO FOOTPRINTING OF HIV-1 LTR IN U1 CELLS
summary of results

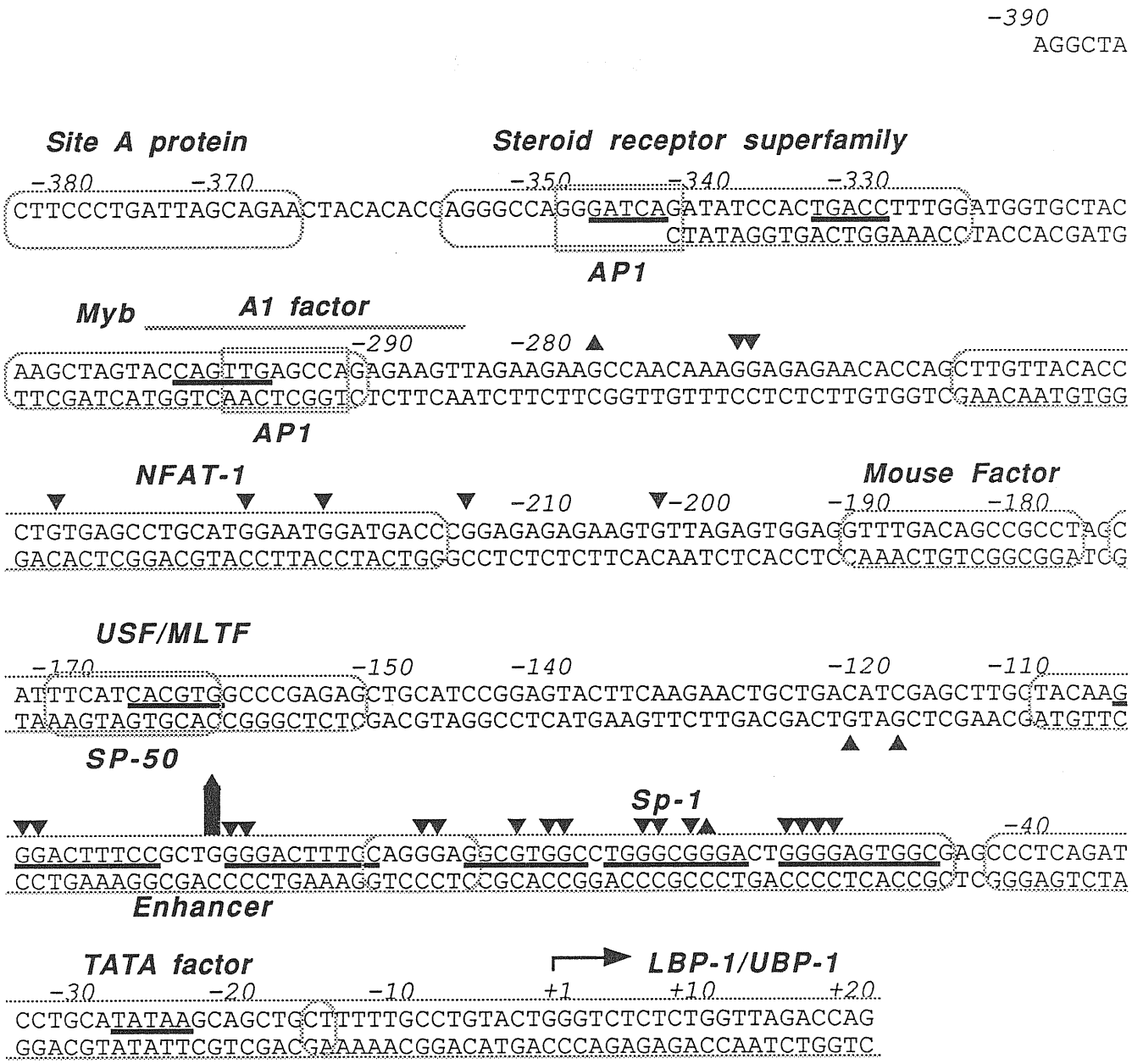


FIGURE 10

INDUCTION OF LTR DRIVEN TRANSCRIPTION BY HERPES SIMPLEX VIRUS AND BY HUMAN HERPES VIRUS-6 ANALYZED BY IN VIVO FOOTPRINTING

Herpes simplex virus and human herpes virus-6 have been suggested as potential cofactors in AIDS, due to their capability to transactivate LTR dependent transcription. In the H3T3 cell line (HeLa-derived cell line containing an integrated LTR-CAT) LTR driven transcription is upregulated after infection with HSV and with HHV6, as monitored by CAT assay (Di Luca 1991).

In order to unravel the molecular events associated to this induction at the level of protein DNA- interaction, we have analyzed by high-resolution *in vivo* dimethyl sulfate footprinting the LTR of H3T3 cells before and after infection with each of the two viruses. Figure 11 reports the results obtained with primer set A (see the first chapter of this section) which is able to evidentiate the region between the transcription initiation site and the negative regulatory element of the coding strand. The first two lanes are sequencing reactions (G+A and C+T) which allow the localization of the bands along the LTR. The third lane is the naked DNA control, that is *in vitro* treated DNA from H3T3 cells. The next following lanes are *in vivo* treated samples from control cells and from cells infected eight hours before with HSV and HHV6 respectively .

Interestingly, also in the untreated cells, where CAT expression is at basal levels, major footprints appear over the three Sp1 sites, and the two repeats of the enhancer region. Upon infection with both viruses, besides the constitutive footprints, a striking enhancement of

hypersensitivity at a site located at the border of the enhancer site closer to the Sp1 boxes occurs. It is worthwhile noting that the pattern of protein-DNA interaction on the HIV-1 LTR in this system before and after transcription induction by viral infection appears to be the same as observed in the U1 cells upon stimulation with TPA.

The *in vivo* footprinting technique constitute the only means to determine the occupancy of a protein binding site *in vivo.* This technique, in combination with *in vitro* studies and transient transfection assays allows the understanding of the complex mechanisms of transcription regulation. The first logical approach to understand the high degree of complexity of the HIV-1 promoter is the dissection of the mosaic of binding sites and the analysis of each specific region one by one. In the following chapters I will describe our studies on the characterization of two specific regions which we found involved in protein-DNA interactions *in vivo* in H9 cells, namely the USF binding site and the SISSA binding site. As far as it concerns the USF binding site, I will describe, besides our data on HIV-1 transcription regulation, also some results obtained studying this binding site in the context of a human origin of DNA replication isolated in the laboratory of prof. Falaschi some years ago [Tribioli 1987].

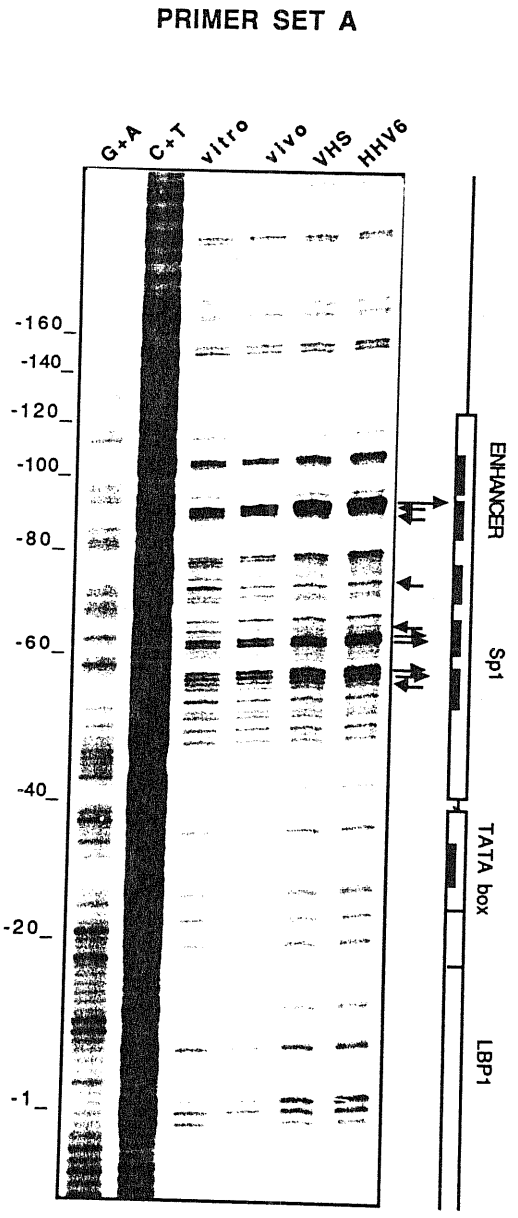


FIGURE 11: Genomic footprinting of HIV-1 LTR in H3T3 cells before and after infection with HSV or HHV-6. Coding strand

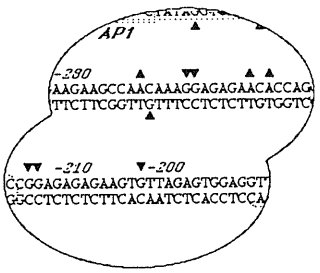
CHARACTERIZATION OF A NOVEL NUCLEAR FACTOR BINDING SITE PRESENT IN THE HIV-1 LONG TERMINAL REPEAT

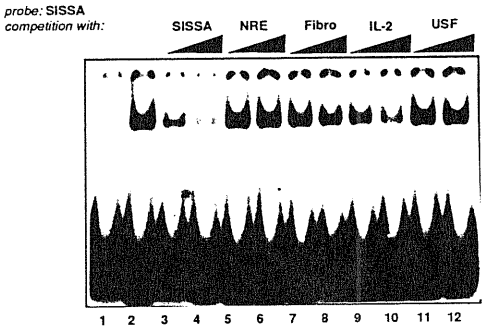
By *in vivo* footprinting experiments in HTLVIIIB infected H9 cells [Demarchi 1992] we have detected the occupancy of a previously unrecognized site in the HIV-1 LTR, and named it SISSA binding site. This site is present in the negative regulatory element, in a purine-rich rich stretch between position -260 and -275 respect to the transcription start site. In order to determine if the protein binding *in vivo* to the SISSA binding site is able to bind to a nuclear protein *in vitro*, we performed gel retardation experiments using HeLa cells nuclear extracts. A double-stranded oligonucleotide encompassing the SISSA site is specifically recognized by a HeLa cell nuclear factor, generating a retarded complex (fig 12). This complex is specific, since it can be competed by the addition of increasing amounts of the same cold oligonucleotide, but not by an oligonucleotide with an unrelated sequence used as negative control.

Furthermore, we show that an oligonucleotide containing a similar sequence, corresponding to the binding site for the nuclear factor NFAT present in the IL2 promoter, is able to compete for the binding to the same nuclear factor (fig 12). Another oligonucleotide sharing a similar sequence, corresonding to a protein binding site present in the fibronectin promoter is able to compete, even if with very low efficiency (fig.12). On the other hand, an oligonucleotide (NRE) containing a sequence similar to SISSA present in the negative regulatory element (-204, -216), does not compete, as it is the case for a negative control oligonucleotide (USF)

(fig.12).

A NOVEL BINDING SITE IN THE NEGATIVE REGULATORY ELEMENT OF THE HIV-1 LTR





OLIGO	SOURCE	from	HOMOLOGY	to
SISSA	HIV-1 LTR	-275	CAACAAAGGAGAGAACA	-266
NRE	HIV-1 LTR	-196	ATGACCCGGAGAGAGAA	-213
Fibro	Human fibronectin promoter	-429	CCCGAAGAGAGGTGAC	-414
IL-2	Human IL-2 promoter	-291	AAAGAAAGGAGGAAAA	-277

FIGURE 12: Band shift and competition experiments

CONSTRUCTION OF HIV-1 LTR-CAT MUTATED CONSTRUCT

We have shown that an oligonucleotide encompassing the SISSA binding site is able to bind a nuclear protein *in vitro*. A substitution of only eight central nucleotides (oligonucleotide FREE) involved in protein-DNA interaction *in vivo* abolishes binding, as assessed by gel retardation experiments (data non shown). The same modification was introduced in LTR-CAT plasmid by a combination of conventional molecular biology techniques and overlap extension mutagenesis using PCR.

The mutated insert was constructed as follows:

1) PCR amplification of a plasmid containing the LTR using two different sets of primers, namely :

_ an oligonucleotide corresponding to the lower strand, encompassing the site to be mutated (FREE1, see above) and an oligonucleotide corresponding to the upper strand corresponding to the 5' end of the LTR and encompassing a KpnI restriction site (KPN).

_ an oligonucleotide corresponding to the upper strand, encompassing the site to be mutated (FREE2, see above) and an oligo corresponding to the lower strand in a region located downstream to an Ava I restriction site (USF1).

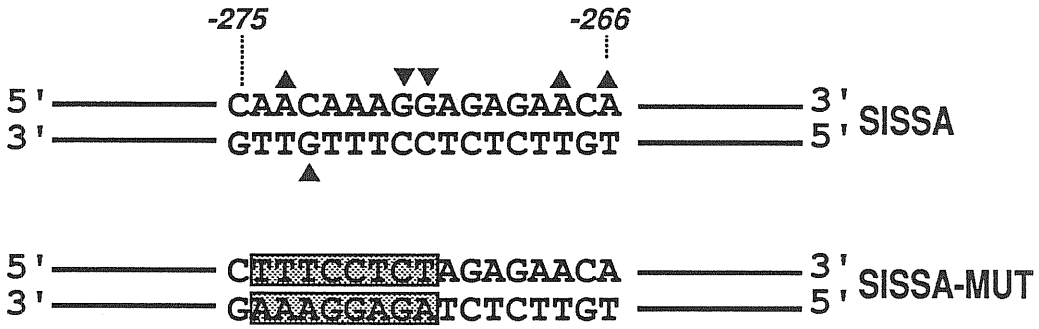
2) PCR of a mixture of the two different PCR products, after analysis on agarose gel, and dilution, using the external primers of the above reactions, namely KPN and USF1.

3) Restriction digestion of the PCR product with AvaI and KpnI.

The mutated insert was purified and ligated to an LTR-CAT plasmid previously digested with KpnI and AvaI and dephosphorylated.

The construct was checked by sequencing using the conventional Sanger method, the sequence of the mutated region is shown in the following page (fig 13).

SEQUENCE OF THE MUTATED CONSTRUCT



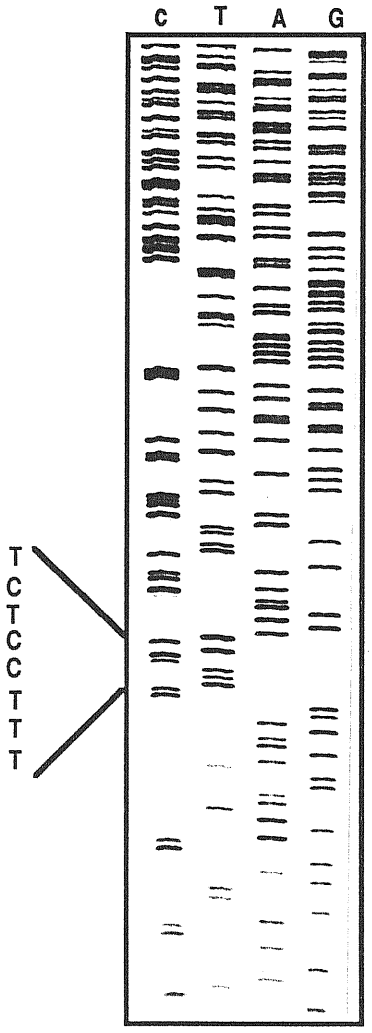


FIGURE 13

THE SISSA BINDING SITE IS AN UP-REGULATOR OF VIRAL TRANSCRIPTION IN H9 CELLS

In order to investigate the biological function of the SISSA binding site in the context of the HIV-1 LTR we decided to compare the promoter strength of a wild type LTR with the LTR mutated in the SISSA binding site. To this end, a number of transient transfections experiments of an LTR-CAT plasmid and of a mutated plasmid, constructed as described in the previous chapter, were performed in the H9 cell line. H9 cells were transfected either with or without cotransfection of an expression plasmid for the HIV-1 tat gene. Each experiment was performed at least three times and with different plasmid preparations in order to take into account differences in transfection efficiencies. In the absence of tat we didn't find any difference in the CAT activity using extracts from cells transfected with the two different constructs (data not shown). On the contrary, we detected a clear difference in the promoter strength in H9 cells upon cotransfection with an expression plasmid for the tat gene (fig.14). The results show that a specific mutation in the SISSA binding site results in a decrease in CAT activity, in presence of tat transactivation, suggeting a positive role for SISSA in the promotion of transcription.

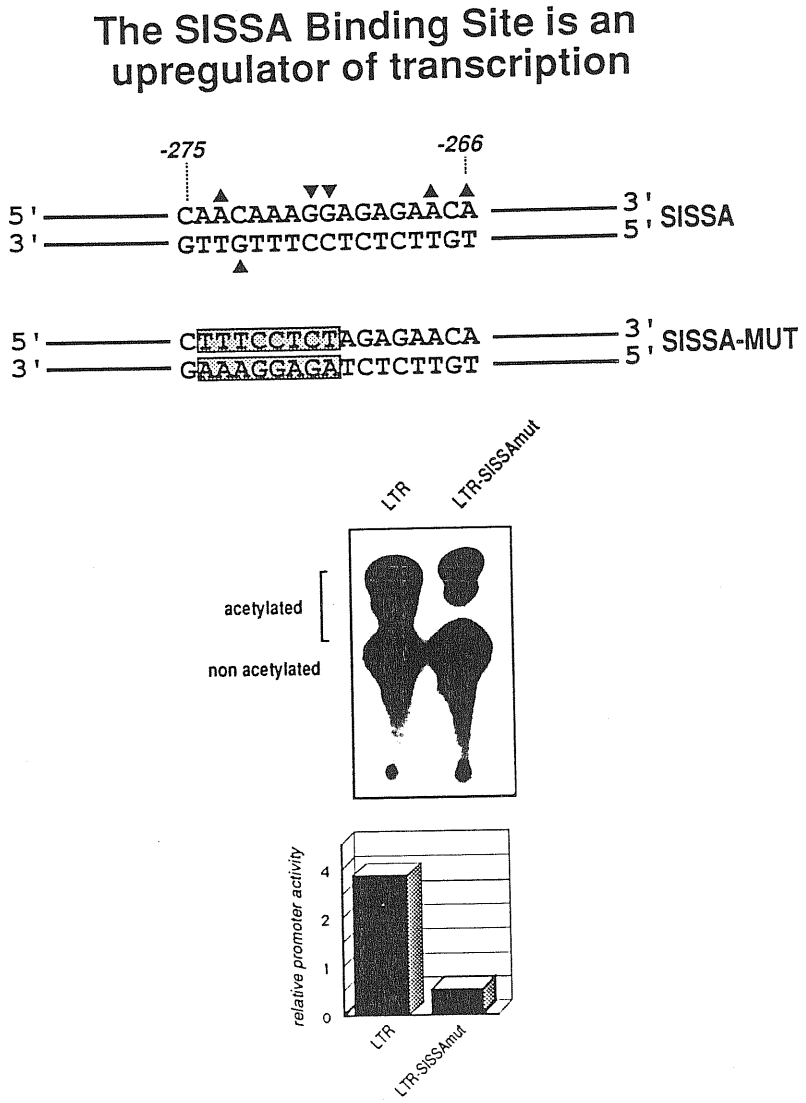


FIGURE 14: Cat activity of an HIV-1 LTR derivative bearing a specific substitution at the SISSA binding site

CHARACTERIZATION OF AN EVOLUTIONARILY CONSERVED PROTEIN BINDING SITE

In this chapter I describe a region located at position from -159 to -173 upstream of transcription start site of HIV-1. It has been demonstrated by *in vitro* DNase I protection studies [Garcia 1987] that this domain represents a binding site for host cellular factors. It has been recently demonstrated that deletion of this site results in a marked increase of both viral replication and reporter gene transcription as compared to the wild type [Lu 1990]. This site is homologous to the sequence corresponding to the upstream element of the Major Late Promoter (MLP-UE) of Adenovirus 2, located between nucleotides -63 and -52 upstream of the start site of the late transcripts. This element is the target for a nuclear factor, called by different authors MLTF [Carthew 1985], USF [Sawadogo 1985] or UEF [Moncollin 1986], present in several cell types and tissues with an abundance of about 10,000-20,000 molecules/cell [Chodosh 1986] [Sawadogo 1988]. The binding motif for this factor is present in many different systems, including higher and lower eukaryotes, viruses and plants, as it is illustrated in the table shown on the following page.

Table I. Sequences carrying the conserved CACPuTG dyad symmetry binding motif

Source	*Sequence*	*Position*	*Factors identified*	*References*
MAMMALS				
pB48 Binding Site	CGT**CACGTG**ATG	-84	three proteins of 44, 70 and 110 KDa	Tribioli *et al.*, 1987; Giacca *et al.*, 1989
Human growth hormone gene promoter	ACC**CACGTG**ACC	-267		Lemaigre *et al.*, 1989
Human heme oxygenase gene promoter	GCC**CACGTG**ACC	-47	USF	Sato *et al.*, 1990
Human U1 snRNA gene promoter	GGG**CAAGTG**AC	-68		Gunderson *et al.*, 1988
Rat γ-fibrinogen gene promoter	GAC**CCCGTG**ACC	-84	MLTF	Chodosh *et al.*, 1987
Rat heme oxygenase gene promoter	CAC**CACGTG**ACT	-50	HOTF (40 KDa)	Sato *et al.*, 1989
Rat pyruvate kinase gene promoter	GCG**CACGGGGCA** ACT**CCCGTG**GTT	-168 -157	UEF; two proteins of 52 and 115 KDa	Vaulont *et al.*, 1989
Mouse Hox 1.1 gene upstream region	GGT**CACGTG**CCG	-128		Giacca *et al.* 1989
Mouse IgH enhancer - site μE3	GGT**CATGTG**GCA		NF-μE3, C2-binding protein (42-45 KDa), TFE3 (59 KDa), TFEB (~60 KDa); a yeast homologue identified (YEB3, 33-41 KDa)	Peterson and Calame, 1988; Sen and Baltimore, 1988; Beckmann and Kadesh, 1989; Beckmann *et al.*, 1990; Carr and Sharp, 1990
Mouse metallothionein I gene promoter	GGG**CGCGTG**ACT	-104	MLTF	Carthew *et al.*, 1987
BIRDS				
Duck histone H5 gene promoter	GTC**CACGTG**ACC	+505	USF	Düring *et al*, 1990
AMPHIBIANS				
X. laevis TFIIIA gene promoter	TAT**CACGTG**CTCC	-272	USF	Scotto *et al.*, 1989
X. laevis Hox 1.1 gene upstream region	GAT**CACGTG**GCC	-145		Giacca *et al.*,1989
YEASTS				
S. cerevisiae centromere DNA element I (CDEI)	A**PuTCACPuTG**ATA		CP1 (58 KDa), CDEI-binding proteins (64 and 37 KDa), CBF1 (39KDa)	Bram and Kornberg, 1987; Baker *et al.*, 1989; Jiang and Philippsen, 1989; Cai and Davis, 1990
S. cerevisiae GAL2 gene promoter	GGT**CACGTG**ATC		CP1 (57-64 KDa)	Bram and Kornberg, 1987
S. cerevisiae MET25, MET2, SAM2, MET3 genes upstream regions	**CACPuTG**	~-200 ~-300		Thomas and Kornberg, 1989
S. cerevisiae PHO5 gene upstream region	TAG**CACGTT**TTC TCA**CACGTG**GGA	-368 -256	PHO4	Vogeland *et al.*, 1989
S. cerevisiae QH2: cytochrome *c* oxydoreductase subunits VI, VIII and FeS genes upstream regions	**PuTCACGTG**		GFII	Dorsman *et al.*, 1988
VIRUSES				
Adenovirus Major Late Promoter upstream element	GGC**CACGTG**ACC	-63	USF, MLTF, UEF (43-46KDa); TFEB (~60 KDa); yeast homologues identified: yUEF (60 KDa), MRF	Chodosh *et al.*, 1989; Chodosh *et al.*, 1986; Sawadogo and Roeder, 1988; Moncollin *et al.*, 1987; Moncollin *et al.*, 1990b
Adenovirus IVa2 gene promoter	AGA**CACATG**TCG	-116	UEF	Moncollin *et al.*, 1990
Adeno-associated virus P5 promoter	GGT**CACGTG**AGT	-83	MLTF	Chang *et al.*, 1989
Human cytomegalovirus 2.7 Kb RNA promoter	CGT**CACGTG**AAA	-114	USF/MLTF	Klucher and Spector, 1990
HIV-1 Long Terminal Repeats	CAT**CACGTG**GCC	-169	three factors of 44, 70 and 110 KDa	this paper
PLANTS				
Arabidopsis thaliana and tomato RBCS genes promoters	C/A-**CACGTG**GCA	~-250	GBF; a yeast homologue identified (yGBF)	Giuliano *et al.*, 1988; Donald *et al.*, 1990
Petroselinum crispum chalcone synthase gene promoter	TT**CCACGTG**CCA	-170		Schulze-Lefer *et al.*, 1989
Antirrhinum majus chalcone synthase gene promoter	TGT**CACGTG**CCA	-136	CG-1	Staiger *et al.*, 1989
Nicotiana plumbaginifolia cab-E gene promoter	TCAG**ACGTG**GCA	-245	GBF	Schindler and Cashmore, 1990
Arabidopsis thaliana adh gene promoter	C**CACGTG**GGA	-210		McKendree *et al.*, 1990

Only sequences for which a nuclear binding activity was identified are listed and the cognate binding factors are indicated. Matches to the CACGTG motif are shown by bold typed letters; the positions upstream of transcription start sites (where available and appropriate) are indicated.

Our study of USF/MLTF began when we found a binding site for this factor in a human DNA region replicated at the onset of S phase in synchronized HL60 cells and probably involved in the activation of an origin of DNA replication (plasmid pB48) [Tribioli 1987], [Falaschi 1988] [Giacca 1989]. In order to determine if the protein binding to pB48 is able to bind to other promoter elements, some of the binding sites indicated in the table above where analyzed for their ability to compete in binding competition experiments.

Band shift competition experiments were performed by incubating a ^{32}P labeled 24-mer oligonucleotide encompassing the binding site of pB48 with an excess of cold competitor DNA. Molar excess of 10, 20, 40 fold were used when the competitor was a restriction fragment , while higher molar excess (20, 60, 180 fold) were used with cold oligonucleotides competitors. The affinity of the nuclear factor , indeed, was lower for oligonucleotides than for longer DNA fragments, in agreement with what recently observed for USF [Sawadogo 1988], probably due to the involvement of surrounding non specific DNA in protein-DNA interactions.

The results of competition experiments with a variety of sequences are shown in fig. 16. The strength of competition is indicated by (+/-) symbols when the competition assay was performed with a purified cold restriction fragment, or is visualized by the intensity of the retarded band in the case of oligonucleotides. Competition was observed with a fragment of HIV-LTR as well as with an oligonucleotide encompassing nucleotides 174 to 151 upstream of the transcription start site of viral DNA.

An insert containing CDEI and CDEIII of yeast CEN VI competed poorly for binding, while a clear competition is observable when an oligonucleotide for region CDEI is used as a probe. This discrepancy is probably due to the fact that the centromer fragment used contains binding sites for other human proteins which can interfere with the binding activity under study.

The data here presented have already been published [Giacca 1989].

Source	Sequence	Competition with pB48 Bindig Site	
		as plasmid insert	as oligo
			20x 60x 180x
pB48 Binding Site	TTCGTCACGTGATGCGA	++++	
Human Alu (Blur 8)	CAGATCACCTGAAGTCA	-	nd
Human 0-LTR Family	CCAGACATGTGGAAGTA	-	nd
Human Hox (HHO.c1.95)	GGAATCTCGTAAAACCG	-	
Mouse Hox 1.1	GCGGTCACGTGCCGCGG	nd	
Xenopus Hox 1.1	CAGATCACGTGGCCCAG	nd	
Yeast TRP1-ARS1	TTGAGCACGTGAGTATA	-	nd
Yeast Centromere (chr. VI)	TTCATCACGTGCTATAA	+	
Yeast GAL2 promoter	TGGGTCACGTGATCTAT	nd	
Adenovirus MLP	TAGGCCACGTGACCGGG	++++	
HIV-1 LTR	TTCATCACGTGGCCCGA	+++	
Human growth hormone	GCACCCACGTGACCCT	nd	nd
Rat gamma-fibrinogen	GACCCCGTGACC	nd	nd
Mouse methallotionein I	CGGGGCGCGTGACTATA	nd	nd

FIGURE 16: Band Shift competition experiments

A HUMAN USF BINDING SITE DOWNREGULATES TRANSCRIPTION OF HIV-1 LTR

The role of the USF/MLTF binding site on the HIV-1 LTR was investigated with respect to its ability to modulate transcription within the cell. Since this site is located at the 3' boundary of a large region known to exert a negative regulation on transcription (NRE) (Rosen et al., 1985) (Garcia et al., 1987), the question whether the USF/MLTF site could account for this negative regulation was addressed. For this reason, two series of plasmids were constructed; a schematic representation of these plasmids is reported in figure 17, Panel F. The first set derives from plasmid pC15CAT (Arya et al., 1985), which contains most of the LTR and the nef open reading frame cloned upstream of the CAT gene in plasmid pSV0CAT. Three derivatives were constructed, containing an AvaI - AvaI deletion which removes the whole NRE upstream of position -159, with simple ligation (pEM0), or by cloning a 100 bp fragment of human origin containing the pB48 Binding Site in the two possible orientations (pEM1 and pEM2). The activities of the four plasmids were tested after transfection in HeLa cells without or with cotransfection of an expression plasmid for the HIV-1 tat protein (figure 17, Panels A and B, respectively). The experiments were performed with cotransfection of another independent plasmid (pCH110), coding for the ß-galactosidase gene under the control of the SV40 early promoter, in order to take into account differences in transfection efficiencies. The results show that the deletion of the NRE results in an increase of CAT activity, which is more evident in presence of tat transactivation (compare figure 17, Panel B, lanes 5 and 4), confirming the expected negative role that this region exerts on

transcription. This negative effect can be, to a great extent, replaced by the 100 bp human fragment containing the USF/MLTF binding site, when cloned in one orientation (pEM1, Panel A, lane 4 and Panel B, lane 3), but not in the opposite one (pEM2, Panel A, lane 5 and Panel B, lane 2).

Since the USF/MLTF site is placed asymmetrically within the 100 bp fragment which functionally replaces the NRE, it is not possible from these data to argue whether it is active only in one orientation due to a real polarity of the binding site, despite of its sequence symmetry, or to a position effect of the site itself. For this reason, another set of plasmids was constructed, derived from another vector for cloning opportunities. These include a plasmid containing the KpnI - HindIII fragment of pC15CAT placed upstream of the CAT gene (pLTRCAT), its Δ-159 derivative (pLTRDCAT), and its Δ-159 derivative with an oligonucleotide corresponding to pB48 Binding Site cloned upstream in the two orientations (pLTRDCATBS1 and pLTRDCATBS2, see Materials and Methods for plasmid constructions). These constructs were always tested after cotransfection with a tat-expressing plasmid. The results presented in figure 17 (panels C-E) are representative of 3-5 different experiments. Differences in transfection efficiencies between different samples within the same experiment were standardized by measuring the protein concentration of the cell extracts used for the CAT assays, since this measurements were previously found to correlate very closely with the ß-galactosidase activity (see Panels A and B, lower parts). Deletion of the NRE upstream of position -159 (pLTRDCAT) results in a 5-6 times increase of CAT activity after transfection of HeLa (Panel C) or COS-1

able to replace most of the negative function of the NRE when cloned upstream of the deletion in both cell lines (plasmids pLTRDCATBS1 and pLTRDCATBS2). The same result was obtained in HeLa cells after treatment with PMA (panel E), although the overall effect of the deletion of the NRE resulted lower in comparison to untreated cells.

The results presented in this chapter have already been published [Giacca 1992]

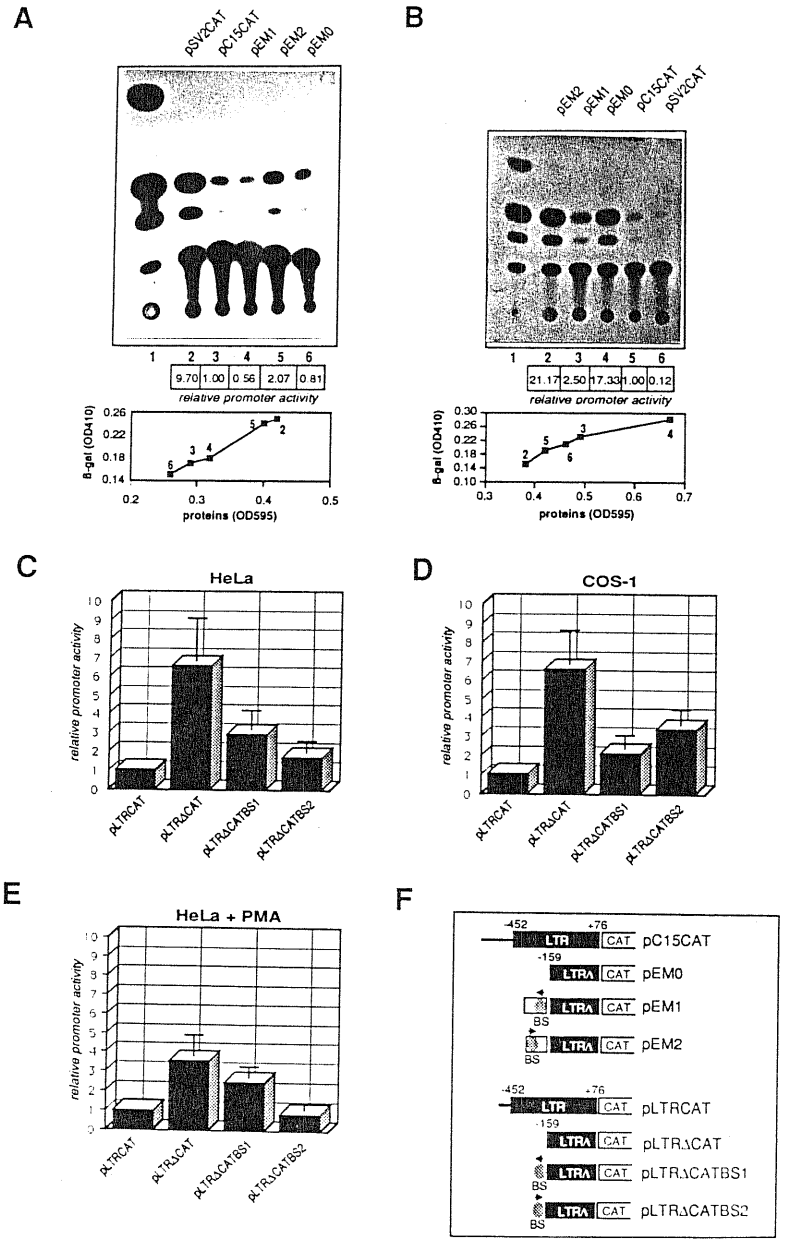


FIGURE 17: CAT activity of some deletion/reconstitution derivatives of HIV-1 LTR

THE USF BINDING SITE PRESENT IN A HUMAN ORIGIN OF DNA REPLICATION FUNCTIONS AS AN ACTIVATOR OF TRANSCRIPTION

As already mentioned above, we began our study on the USF/MLTF binding site in the course of the characterization of a human origin of DNA replication (pB48) isolated some years ago in the laboratory of prof. Falaschi [Tribioli 1987]. PB48 was isolated for its early replicating property in HL60 cells, and it was later shown to encompass the 3' non translated region of a lamin gene and the 5' end of a yet unknown gene. It contains a 600 base pairs HTF (Hpa Tiny Fragments) island typically present upstream of house-keeping genes, and a number of putative binding sites for transcription factors. Moreover, as already described at pag. 84, it contains a 17 base pairs sequence similar to the core consensus for USF/MLTF that binds to a nuclear factor, as shown by band shift and *in vitro* footprinting experiments. On the basis of the observed abundance of transcription signals, we decided to search for the possible presence of an active promoter in pB48 by CAT assay experiments. We already showed that a 900 bp pB48 insert in the appropriate orientation can promote transcription from a downstream reporter gene in a transient transfection assay [Falaschi 1988]. To map this promoter to a higher resolution, we cloned in the assay plasmid, pCAT0, a pB48 fragment of 100 base pairs encompassing a TATA box and a USF/MLTF binding site, in both orientations with respect to the CAT reporter gene.

As it is shown in figure 18, the 100 base pair fragment is sufficient to drive transcription of the downstream reporter gene, when present in the appropriate orientation, even if at a lower extent than the longer

fragment.

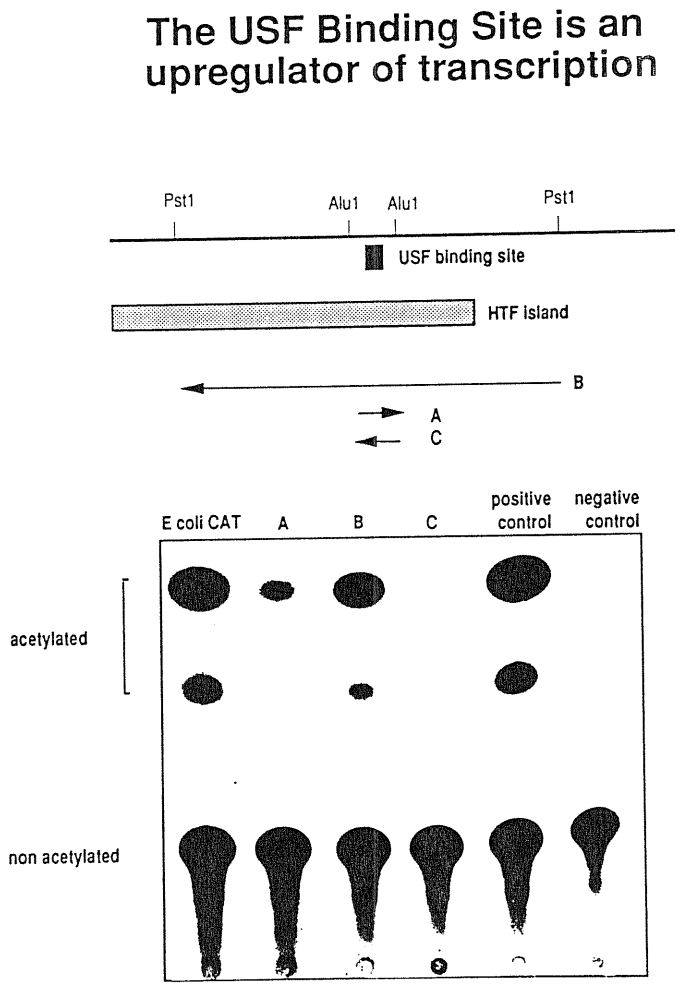


FIGURE 18: CAT activity of specific DNA segments of pB48

GENOMIC FOOTPRINTING OF A PUTATIVE HUMAN ORIGIN OF DNA REPLICATION (PB48) ENCOMPASSING A USF BINDING MOTIF

In vitro binding studies and transient transfection assays have shown the presence of a putative USF/MLTF binding site in pB48 and its involvement in promoting transcription of a reporter gene. However, as already mentioned above, the interaction of specific DNA binding proteins with their corresponding sites may not always occur in the living cell [Becker 1987], [Mueller 1989], where it is affected by several other parameters, such as chromatin structure, interactions with other proteins, conformational changes, epigenetic modifications of DNA like methylation, etc.

To determine whether the USF binding site present in the pB48 promoter is actually occupied *in vivo*, we analyzed HeLa cells by *in vivo* dimethylsulfate (DMS) footprinting using the ligation-mediated polymerase chain reaction (LM-PCR) technique [Mueller 1989].

The patterns of DMS methylation are shown in figure 19 and the results summarized along the pB48 sequence in figure 20. Figure 20 reports also some of the salient features of pB48. In Figures 19 the region of the pB48 recognized as binding site for nuclear proteins by *in vitro* experiments is also shown alongside.

In HeLa cells, a clear footprint appears at the core of the USF/MLTF target site (CACGTG), located from nucleotides 791 to 796. Although the piperidine treatment used mainly favors cleavage at alkylated guanines, it may also reveal, with reduced sensitivity, adenine residues that are particularly reactive with DMS [Mueller 1989], as it is the

case of adenine at position 797, adjacent to the core consensus of USF/MLTF.

Very interestingly, three novel sites, located respectively between nucleotides: 625 - 640, 723 - 731, and 758 - 773, appear to be involved in protein-DNA interactions in vivo. All of these three DNA segments are particularly GC-rich.

Finally, the guanine at position 763 is clearly hypersensitive, possibly as a consequence of interactions of factors bound on pB48 both upstream and downstream of this location. These results indicate that the pB48 region is crowded with nuclear factors in the living cell, suggesting a role for the control of transcription, and (or) origin activation.

Genomic footprinting of pB48 in HeLa cells

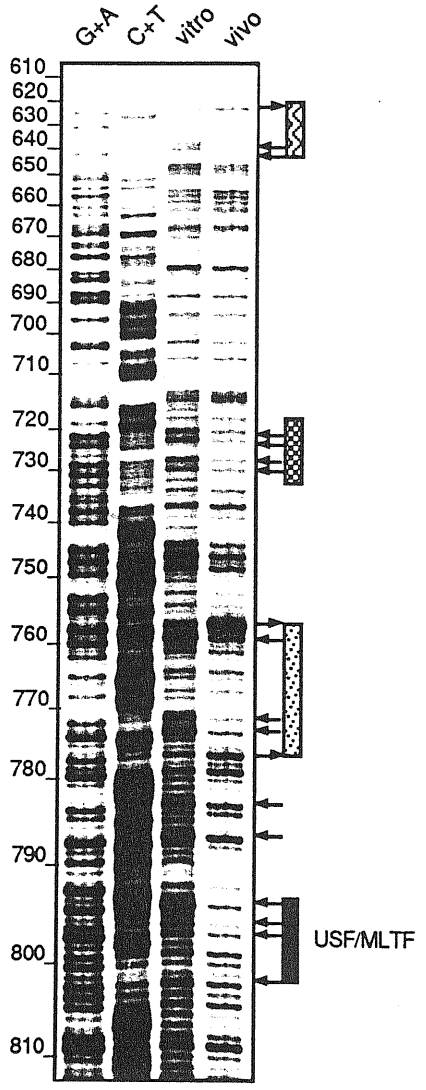
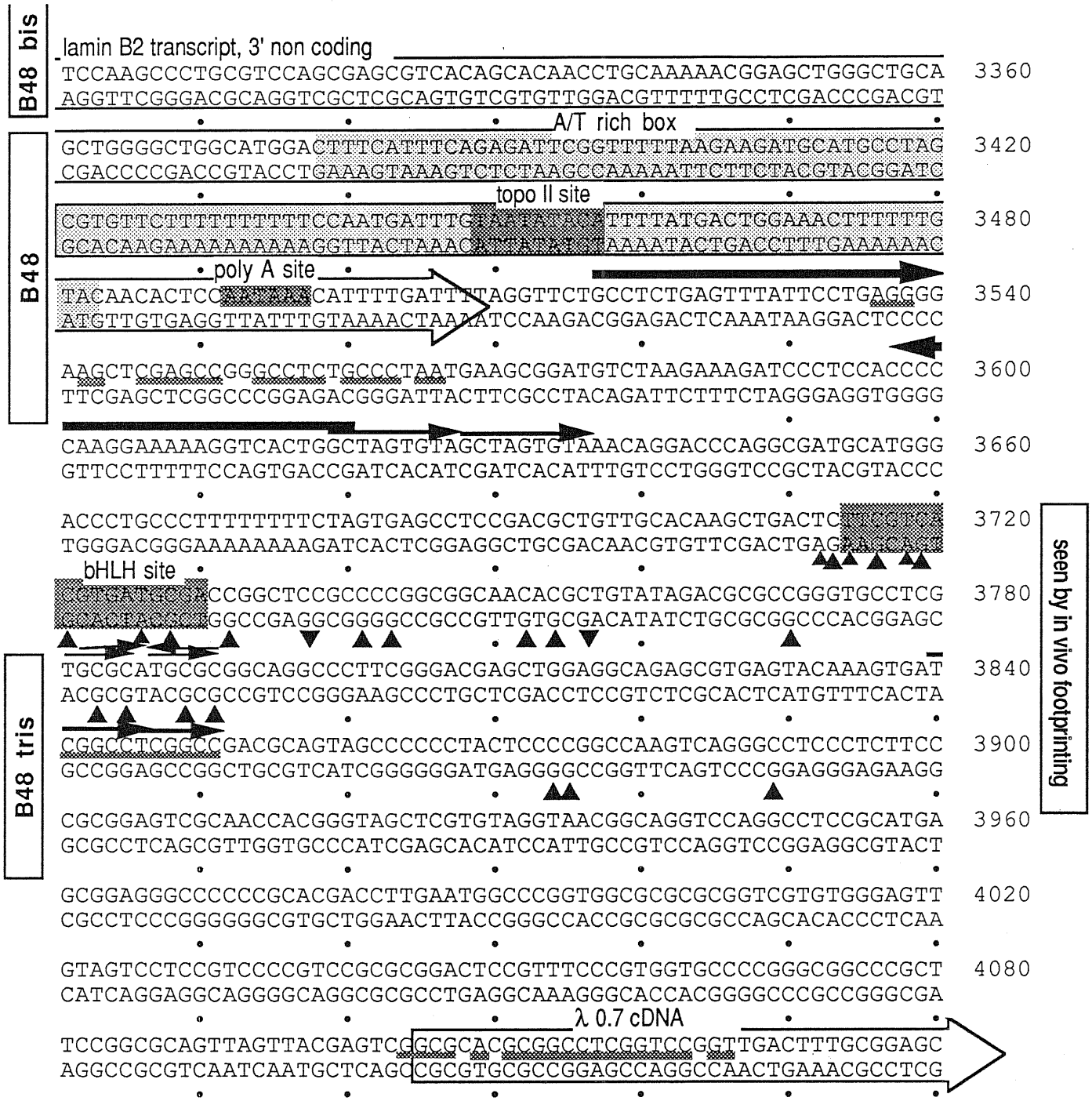


FIGURE 19

SEQUENCE AND FEATURES OF THE pB48 ORIGIN REGION



JCV/BKV homologies

left part: quantitative PCR amplified region

filled triangles: guanines protected or hypersensitive in *in vivo* footprinting experiments

FIGURE 20

DISCUSSION

WHICH BINDING SITES OF THE LTR ARE ACTUALLY OCCUPIED IN A LIVING INFECTED CELL?

The regulation of HIV-1 gene expression plays a key role in triggering the replication of the integrated provirus, and consequently in disease development. Transcription regulation of HIV-1 is modulated by the synergic action of the viral protein Tat and nuclear proteins interacting with *cis* acting viral regulatory elements in the Long Terminal Repeat at the 5' end of the HIV-1 provirus. In the early phase of the viral life cycle, in the absence of the potent viral transactivator Tat, the regulation is exclusively dependent on cellular factors, as it is the case for cellular genes. Therefore, determining the protein binding sites of the LTR actually occupied in a living infected cell represents an important facet of studying HIV-1 gene control.

THE EXPERIMENTAL APPROACH

The direct analysis of proteins bound to target sequences within the living cells is of particular importance in the study of transcriptional regulation of mammalian genes, since DNA is compacted into a highly organized chromatin structure within the eukaryotic nucleus. Protein-DNA interactions are therefore determined not only by the presence of regulatory proteins, but also by the accessibility of target sequences to such proteins. Thus, the detection of proteins in nuclear extracts capable of binding to specific DNA sequences *in vitro,* does not necessarily mean

that the binding sites of such proteins are occupied *in vivo*. The use of genomic footprinting has shown for example that the activation of the glucocorticoid inducible rat tyrosine aminotransferase gene is dependent on an ubiquitous factor, which binds on its target site only in expressing cells [Becker 1987]. Another interesting result obtained by genomic footprinting is the detection of protein-DNA contacts *in vivo*, which cannot be recreated *in vitro*, as it is the case for the interaction occurring at the enhancer of the Immunoglobulin heavy chain enhancer [Gerster 1987].

Finally, monitoring changes in the occupancy of specific protein targets on promoters within cells, it has been shown that transcription activation may occur without any change at the level of protein-DNA interactions. A constant pattern, despite activation, is observed for example at the histone H3 and H4 promoters during the whole cell cycle [Pauli 1989], at the interferon gene promoter both before and after stimulation by viruses [Palmieri 1990], and at the Serum Responsive Element present in the promoter of the c-fos gene independently of serum addition [Konig 1991]. In these cases, transcription activation might be mediated by alternative mechanisms such as alterations in protein-protein contacts, perhaps through the intervention of protein modifications, intervention of auxiliary factors, or modifications at the level of the general transcription machinery.

Few of the probing reagents commonly used for *in vitro* footprinting studies are suitable for *in vivo* footprinting experiments, because of the relative impenetrability of cellular membranes to the majority of high molecular weight enzymatic nucleases and low molecular weight, but frequently charged, chemical probing reagents.

We decided to start our analysis using dimethyl sulphate as a probe, because it appears the most convenient available reagent due to the following characteristics: free membrane permeability, high and rather well understood reactivity, small size, and therefore, low probability to induce structural perturbations. The predominant site of N- methylation on DNA occurs at the N7 position of guanine (accessible via the major groove); substantial less reaction occurs at the N-3 position of adenine (accessible through the minor groove). Altered reactivity of guanine bases is taken as an indication of intimate contact with protein side chains through the major groove.

Most frequently, reduced or abolished reactivity at a given position is detected if steric inhibition of the approach of DMS is occurring as a result of protein-DNA interactions. However, hypersensitivity is quite frequently seen and it has been interpreted as a consequence of the action of a protein side chain- mediated hydrophobic pocket in the vicinity of the N-7, an environment particularly favorable for DMS reaction. One drawback of the use of DMS is that it might disturb some protein-DNA interactions; as a consequence, some specific contacts might not be detected. To overcome this inconvenient the use of an alternative probing reagent is advisable. For this reason experiments using DNase-I as a probe are in progress.

In this thesis I present the data obtained by carrying out high resolution mapping of sites of DNA-protein interactions at the HIV-1 LTR by DMS footprinting in two different infected cell lines, namely the H9 human neoplastic T cell line and the promonocytic U1 cell line, and in a cell line containing only an integrated copy of the HIV-1 LTR.

PROTEIN-DNA INTERACTIONS AT THE HIV-1 LTR IN THE H9 CELL LINE

In the H9 cell line, which continuously produces large amounts of the HIV-1 virus, we have observed that the enhancer and Sp1 sites on the LTR are fully engaged in protein-DNA interactions, as expected from the well demonstrated role of these regions on the promotion of transcription [Jones 1986] [Chinnanduray 1991]. Somehow surprisingly, however, clear interactions are also present in the Negative Regulatory Element that has been shown to have a repressive function on HIV-1 transcription and viral replication [Rosen 1985], [Lu 1989]. We have observed alterations of DMS sensitivity in the following regions of the NRE: the USF/MLTF and NFAT-1 sites, a region where binding sites for several proteins including the AP1 factor [Franza 1988], and a factor named A1 [Guy 1990] are located and where purified Myb protein was shown to interact *in vitro* [Dasgupa 1990], and at a region containing a palindromic sequence with homology to the steroid\thyroid hormone responsive element [Orchard 1990].

The protein occupancy of a region involved in repression of transcription in a cellular environment where expression is at maximal levels is apparently paradoxical. One possible explanation is that at least some of the factors, despite their sitting on DNA, might not work as repressors in chronically infected H9 cells, because they are in an inactive form or because they do not interact with ancillary factors required for function. Alternatively, their sites could be occupied by different proteins that mask the repressor binding sites and prevent

negative regulation.

The binding site for transcription factor USF/MLTF has been shown to play different roles in different contexts, that is repression in the LTR of HIV-1 [Giacca 1992] and activation in the major late promoter of Adenovirus [Sawadogo 1988]. The potential to exert opposite functions according to the context in which the target site is present, and according to the cellular milieu, is a well established feature of several mammalian transcription factors. As a consequence, it becomes necessary not only to isolate and characterize the transcription factors, but also to reconstruct a three dimensional picture of proteins sitting on DNA, taking into account DNA tertiary structure, interactions with other specific factors, with nucleosomes and with the nuclear scaffold. To this end, genomic footprinting certainly represents a valuable tool.

Altered DMS sensitivity at positions from -260 to -275 and from -204 to -216 suggests that also these previously unrecognized sites are occupied by proteins *in vivo*. Both these sites contain a purine-rich tract very similar to the sequence of a protein binding site present in the fibronectin promoter (from -414 to -429) [Dean 1989], and in the interleukin 2 promoter, in a region which binds the NFAT factor *in vitro* [Show 1988]. The biological significance of these interactions is currently under investigation, and some preliminary results for the upper site (SISSA site) are discussed in a following chapter.

The results here discussed represent a reliable, static picture of the contact points of nuclear factors sitting on DNA in chronically infected H9/HIV-$1_{HTLVIIIB}$ cells. In order to obtain a dynamic picture of the traffic of factors on the LTR occurring upon transcription activation, we decided to

extend this type of analysis to other HIV-1 infected cell lines suitable for induction by chemical and biochemical agents.

CONSTITUTIVE AND INDUCIBLE PROTEIN-DNA INTERACTIONS AT THE HIV-1 LTR IN U1 CELLS EXAMINED BY IN VIVO FOOTPRINTING

HIV-1 is expressed at all stages of natural infection and therefore, the period of clinical latency is not due to a true latency at the molecular level. However, the level of HIV-1 RNA expression, measured as ratio of genomic RNA and viral DNA, is directly correlated with HIV disease stage. It has been shown that early disease is associated with a ratio between copies of genomic RNA and copies of viral DNA ranging from one to seven. This ratio increases to over 1000 late in infection [Michael 1992].

As already describe above, the regulation of transcription is exerted by the synergic action of the viral protein Tat interacting with its binding site and cellular transcription factors interacting with the LTR. In the early phase of the cell cycle the virus is in an environment free of Tat and, therefore, it relies exclusively on cellular factors for transcription activation.

The model system we utilized for studying the activation process is the promonocytic U1 cell line [Folks 1988], generally considered a suitable experimental model to study HIV-1 transcription activation [Pomerantz 1990]. In this cell line HIV-1 transcription is very low, but can be dramatically induced upon addition of specific stimuli, such as cytokines like TNFα, or tumor promoting agents, like PMA [Folks 1988]. Interestingly, the kinetics of activation of viral transcription in the U1 cell line faithfully recapitulates the activation of viral transcription associated

with the development of the AIDS pathogenesis [Kim 1986].

We have carried out high resolution *in vivo* DMS footprinting of the LTR in the U1 cell line before and after stimulation with TPA to address the two following questions:

1) Which are the protein target sites actually involved in protein-DNA interaction in the living promonocytic U1 cell line?

2) Which protein-DNA interactions are perturbed in the process of transcription activation?

We have found protein occupancy at the three Sp1 binding sites, the two enhancer boxes (also known as NF-kB sites), the NFAT site, the SISSA binding site, both before and after transcription activation, and at site adjacent to the second NF-kB site. Together with our previous data on the proteins involved in DNA interactions in the H9 T cell line (discussed in the previous paragraph), this study in the U1 cell line gives us the opportunity to compare one important facet of transcription regulation in cell types which are known to behave quite differently with respect to HIV-1 infection. It is well known, indeed, that, while infected lymphocytes produce a large amount of viral particles and die, macrophages are less susceptible to viral infection and generally survive infection. Interestingly, in chimpanzees, who do not develop illness although infected, these cells are not infected by the virus.

We found a considerable lower number of perturbed sites in the U1 cell line respect to the H9 cell line. This result could reflect also a difference in the monocytes and lymphocytes of HIV-1 infected patients. If each factor binding on the LTR represents the last effector of a signal transduction pathway, then it might be reasonable to think that a higher

number of binding factors could mean that a larger number of stimuli are able to trigger the activation process, thus increasing the probability of viral activation. If this is the case, transcription regulation would represent a further level of discrimination between lymphocytes and monocytes, and would favor the idea that monocytes represent an important reservoir for HIV-1.

A key result of this work is that, also before TPA addition, when the levels of transcription are extremely low, the sites which have been shown to be involved in the activation process (enhancer, Sp1 boxes, NFAT site) are clearly occupied. These data suggest the existence of further levels of regulation which may directly act on transcription factors, like for example modifications of the factors sitting on the promoter, or alternatively may involve the entry of novel factors functioning as bridges between binding activities and activators, or directly linking the basic transcription machinery to potential activators. Another possibility can be considered in light of recent data on the involvement of some components of the general transcription machinery, possibly TFIID-associated proteins, in the processes of transcription activation [Sakaguchi 1991] and repression [Garcia 1992].

Another interesting point emerging from this study is that transcription activation is accompanied by a dramatic increase in hypersensitivity of a guanine present in the enhancer binding site adjacent to the Sp1 boxes. This result can indicate different situations:

1) Increase in protein occupancy by a factor that is already present in the uninduced state (or increase in the number of cells where the site is occupied).

2) Substitution of a constitutive factor with an inducible factor.

3) Formation of a highly complex nucleoprotein structure involving factors bound both upstream and downstream to the hypersensitive site. As already described in the introduction, a large number of factors, most of which belong to the Rel family of transcription factors, can possibly interact with the enhancer elements. They can bind as homodimers or heterodimers, exerting either a positive or a negative function. Rel-regulated transcription appears quite complex; in particular, upon stimulation with TPA, at least four rel-related DNA binding proteins are induced with differential kinetics [Gerondokis 1991]. It has also been reported that induction of HIV enhancer by mitogens and cytokines such as tumor necrosis factor alpha is, at least in part, mediated by NF-kB binding activities in many cell types, including the promonocytic U1 cell line [Griffin 1989]. Our data confirm the previously reported indirect data on the involvement of the enhancer site in transcription activation. Given the large number of factors which have been shown to bind to the enhancer *in vitro*, and to the complex transcriptional and post-transcriptional regulation of each factor, our approach to the problem does not allow to discriminate which are the specific factors binding to the NF-kB sites. However, we may say that, according to our results, the activation process is not accompanied by a rapid change in site occupancy which would occur in case of a simple post-transcriptional activation mechanism, as described for NF-kB [Bauerle 1988]. In fact, we observed protein occupancy also when the transcription level was very low, and the detected increase in hypersensitivity occurs gradually, with the same kinetics of viral mRNA expression after induction. The change at

the level of protein DNA interaction appears concomitant and does not precede the increase of viral mRNA expression.

Interestingly, it has been suggested that the activation mediated by the enhancer and by the Sp1 boxes may be enhanced by the viral transactivator Tat [Liu 1992], [Southgate 1991]. As already described in the Introduction, Tat is an early gene product, which activates viral gene expression by inducing efficient elongation [Kato 1992], and by increasing transcription initiation [Laspia 1990] through a yet undefined mechanism. Very recently, it has been suggested that Tat mechanism of action might be related to those used by VP16 and other conventional DNA sequence specific transactivators [Tiley 1992]. Indeed, It was shown that the acidic activation domain of VP16, when targeted to an RNA sequence by a specific RNA binding domain, can activate transcription directed by the HIV-1 LTR promoter. On the basis of these data, it is tempting to speculate that the gradually increasing perturbation detectable over the NF-kB site after stimulation by TPA might be related to the gradual increase in Tat production. This is of course only a working hypothesis, and we are testing it at the moment.

It has been reported that, in the U1 cells, the HIV promoter can be activated through kB-independent mechanisms [Griffin 1989], induced by GM-CSF in combination with TNF-α. In our study, we didn't detect any modification in the pattern of protein-DNA interaction after TPA stimulation, besides the one involving the enhancer elements. Our results however, do not rule out the possibility of some changes in DNA sites occupancy, during the first hour of TPA addition, since the first time point analyzed corresponds to one hour after TPA addition (given that viral

mRNA level before, and one hour after TPA addition, appear equivalent). Moreover, as already stated above, transcription activation mediated by specific sites may occur without altering a preexisting footprint. An increasing body of evidence supports a model of transcription activation where a key role is played by ancillary proteins, known as adaptors or coactivators [Berger 1990], [Hoffman 1990], [Flanagam 1991]. These factors mediate the interaction between activation domains and the general transcription machinery. Importantly, coactivator proteins appear dispensable for basal transcription, but critical for activated transcription [Berger 1990].

This study has allowed us to answer to the initial questions concerning which were the sites occupied in a living infected cell and which protein-DNA interaction were perturbed upon transcription activation. The results here discussed open a number of new questions on the fine mechanisms of regulation of the factors involved and on the cross-talk between viral and cellular factors, and suggest that transcription regulation directed by the LTR *in vivo* is even more complex than prospected on the basis of *in vitro* studies.

HIV-1 TRANSCRIPTION ACTIVATION BY VIRAL COFACTORS

Symptomatic immunodeficiency does not usually occur until several years after infection with human immunodeficiency virus-1, although the length of this time interval is subject to considerable individual variation. The interval at which AIDS develops, can be influenced by many variables such as age [Galdert 1989], genetic background [Steel 1981], continuing drug abuse [Weber 1990], infection with other viruses [Webster 1992].

Epidemiological studies have not satisfactorily confirmed the clinical relevance of the infection with other viruses and the progression to AIDS; the controversial results can be partly explained by the complexity of the AIDS pathogenesis. On the other hand, a number of studies at the molecular level strongly argue in favor of the importance of viral cofactors in HIV-1 pathogenesis. In particular, herpes simplex type 1 [Mosca 1987], CMV [Davis 1987] EBV [Kenney 1988], papovaviruses [Gendelman 1986], adenoviruses [Nabel 1988], HBV [Seto 1988], and HHV-6 [Ensoli 1988] have been suggested as potential cofactors in AIDS, due to their capability to trans-activate, *in vitro,* HIV-1 LTR directed gene expression by Tat independent mechanisms.

Experimental work has been focused on two major problems: 1) which are the viral factors involved in transactivation, 2) which are the *cis*-acting HIV-1 sequences required for activation by cofactors. In the case of HSV-1, the point of the viral proteins involved in transactivation of LTR driven expression is a little controversial. Cotransfection experiments of single genes of HSV-1 with an LTR-CAT construct have shown that the

immediate early gene product ICP0 alone can account for transactivation [Mosca 1987]. On the other hand, by using viruses with specific mutations in the genes specifying the immediate early transcription regulatory proteins ICP0, ICP4 and ICP27, it was shown that ICP4 and not ICP0 is essential for transactivation [Alrecht 1989].

The *cis* acting sites on the LTR required for transactivation have been originally identified in the enhancer and Sp1 sites by transient transfection experiments with deletion mutants [Mosca 1987]. Moreover, by gel retardation assays, it was demonstrated that HSV-1 infection results in the induction of a nuclear factor binding to the enhancer [Gimble 1988]. On the other hand, it was reported that the viral protein ICP0 is active also when the entire enhancer, exept the TATA box and eight base pairs adjacent upstream sequences, were removed [Nabel 1988]. It was concluded that ICP0 action is independent of any single defined site, and may not require a cis-acting element at all. A recent report [Margolis 1992] contributes in putting together the apparent disparate results reported above. Margolis demonstrates that ICP0 and ICP4 transactivation of the LTR is largely dependent on the presence of NF-kB and Sp1. However in Jurkat CD4 positive cells, the transactivation mediated by the virus or by ICP0 and ICP4 is enhancer independent, suggesting that immediate early genes of HSV1 can transactivate HIV-1 LTR both through the induction of NF-kB and Sp1 and through another as yet undefined cellular factor.

In the case of HHV6, a gene encoding for an early protein (B701), which does not share any sequence similarity with the genes for ICP0 and ICP4 of herpes simplex, is responsible for activation [Geng 1992]. It has been demonstrated that B701 function is dependent on the enhancer site

[Geng 1992], which has previously been shown to be important to confer HHV-6 inducibility to a heterologous promoter [Ensoli 1988]. However, also the Sp1 sites are likely to play a role in the activation process, since low levels of trans -activation are still detectable after deletion of the HIV-1 enhancer, while a complete loss of activity is observed after deletion of the Sp1 binding region. Nevertheless, no binding to this region by nuclear extracts from HHV-6 infected cells was observed [Ensoli 1988].

On the basis of these data on HSV-1 and HHV-6 mediated transcription activation we thought that analyzing *in vivo* the traffic of proteins on the LTR induced upon infection with HSV-1 and HHV-6 would have been helpful in gaining some more pieces of information to complete the puzzle. The model system we have used is the H3T3 cell line, which is a HeLa-derivative cell line, containing an integrated copy of the LTR-CAT construct [DiLuca 1991]. In this cell line the CAT gene is expressed at very low levels and can be sharply induced upon infection with Herpes Simplex-1 or HHV-6 [Di Luca 1991].

We have analyzed protein -DNA contacts at the HIV-1 LTR in the H3T3 cell line by *in vivo* footprinting and found that also before transcription activation the Sp1 boxes and the enhancer motifs are occupied. The same result was obtained in the U1 cell line before induction of viral transcription by TPA (see previous chapter). Eight hours after infection we observed a clear change in the footprint over the enhancer site adjacent to the Sp1 boxes. In particular, we found an increase of DMS hypersensitivity at the same guanine which is affected in the U1 cells upon activation with TPA. (see previous chapter).

Given that the same results were obtained in the two different

systems analyzed, it is reasonable to think that Tat and one or more early gene products of HSV and HHV-6 may affect transcription directed by the LTR through a similar mechanism (see previous chapter), interacting either with NF-kB or with Sp1 proteins bound on their sites. Alternatively, specific early gene products of HSV and HHV-6 may determine directly or indirectly the substitution of a repressing factor with another factor which triggers transcription activation.

Altogether, the results obtained studying the LTR by *in vivo* footprinting show that at the level of protein-DNA contacts, the same events accompany activation by TPA (which acts through the same pathway activated by antigens and TNFα) and by herpes viruses, suggesting that HSV and HHV-6 infections may mimic normal signals involved in T-cell activation such as T-cell receptor triggering and protein kinase C activation. It will be very interesting to test this hypothesis.

Elucidating the mechanisms of HIV-1 activation mediated by viral cofactors is particularly important not only because it can contribute to understand the biology of HIV-1 virus and of transcription regulation in general, but also in light of possible future therapeutical trials on HIV-1 patients. In practical terms, identification of viral cofactors is extremely important, since abrogation of their effects, even if it extends the disease-free interval by 1 to 2 years, may offer a worthwhile benefit to the patient.

THE SISSA BINDING SITE UPREGULATES TRANSCRIPTION OF HIV-1 LONG TERMINAL REPEAT IN H9 CELLS

By *in vivo* dimethylsulfate footprinting we have detected the occupancy of a previously unrecognized purine rich sequence between nucleotides -260 and -275 of the HIV-1 LTR in living HTLVIII-B infected H9 cells. Moreover, we observed protein-DNA interaction at the same site, which we named SISSA, also in the living U1 cells, both before and after induction of LTR driven transcription by TPA.

The SISSA binding site shares a strong sequence similarity with a sequence present in the interleukin-2 promoter that has been shown to bind *in vitro* to the NFAT-1 factor, present only in activated T cells [Show 1988] and it is also similar to a not well characterized protein binding site of the fibronectin promoter. To investigate the biological function of this factor in the context of the LTR, we have started by analyzing nuclear extracts for the presence of SISSA binding activity by gel retardation. We found a specific activity binding to an oligonucleotide encompassing the SISSA site and, although with lower affinity, to an oligonucleotide corresponding to the NFAT binding site of the IL2 promoter and to an oligonucleotide encompassing the similar target present in the fibronectin promoter.

Interestingly, after we obtained these first results, the cloning of a gene encoding for an ubiquitous factor which binds to a purine rich region of the HIV-1 LTR, encompassing the SISSA site was reported [Li 1991]. Furthermore, a recent report shows that the glucocorticoid receptor DNA-binding domain can bind *in vitro* to a region of 44 nucleotides

encompassing the SISSA binding site [Ghosh 1992].

We focused our attention on the biological function of the SISSA element, and constructed a LTR-CAT construct with specific substitutions in the nucleotides involved in protein-DNA interactions *in vivo,* as detected by genomic footprinting. Given the high complexity of the LTR promoter, and the high density of target sites in a relative short DNA region, we found more convenient to modify the target, instead of constructing deletion mutants. We compared the promoter efficiency of the wild type LTR and the mutated LTR by transient transfection assays, and found that the mutation doesn't affect promoter efficiency in H9 cells free of Tat. However, upon cotransfection with a tat expression plasmid, the promoter strength of the mutated LTR is clearly lower respect to the wild type LTR.

This result suggests that the factor binding to SISSA might be important for Tat mediated transactivation. It is well established that Tat can enhance transcription of the LTR, as well as of cellular genes such as TNFα [Sastry 1990], interleukin 6, and fibronectin [Taylor 1992]. The mechanism of activation is not yet clear, but it has been suggested that Tat can stabilize the transcription complex interacting directly or indirectly with specific elements.

We have shown that mutations in the SISSA binding site reduce Tat mediated transcription activation; on the basis of this result, it is tempting to speculate that Tat function can affect the function exerted by the protein binding at the SISSA element. Interestingly, Tat can activate also transcription directed by the fibronectin promoter, which, by the way, contains a sequence similar to the SISSA binding site.

We have shown that the SISSA binding site can act as a positive regulator of viral transcription in Tat expressing H9 cells. At the moment we are trying to address the following questions:

1) Does the glucocorticoid receptor, which has been shown to interact *in vitro* with the region we are interested in, bind also *in vivo* to the same region, and if it does, what is its function?

2) What is the role (if any) of Tat in the activation mediated by SISSA?

3) What is the meaning of a site involved in activation of transcription in a region which is known to have an overall negative regulatory function?

CHARACTERIZATION OF AN EVOLUTIONARILY CONSERVED BINDING SITE PRESENT IN THE HIV-1 LONG TERMINAL REPEAT

We have identified and chacterized two binding sites for the mammalian transcription factor USF/MLTF in two diverse contexts, namely a human origin of DNA replication [Tribioli 1987], [Falaschi 1988], [Giacca 1989], and the promoter of the HIV-1 virus [Giacca 1992]. Sequences identical or very similar to the core CACGTG motif have been identified also in the upstream regions of several genes of mammals, birds, amphibians and plants. In most of these cases constitutive and ubiquitous nuclear binding activities interacting with these sequences have been recognized. The table at page 83 reports the binding sites containing this motif for which a binding activity was identified. The presented data suggest a strict conservation of the dyad symmetry DNA target sequence through evolution, albeit in different genetic contexts and argue in favor of the presence of a family of mammalian nuclear proteins with the same binding specificity. The cDNAs for three members of this family were recently cloned. They include the cDNAs for two factors: TFE3 [Beckmann 1990] and TFEB [Carr 1990], binding to the μE3 site of the immunoglobulin heavy chain enhancer and to the MLP-UE, and the cDNA encoding for USF [Gregor 1990]. Also the data obtained from the purification [Bram 1987] [Baker 1989] [Jiang and 1989] and cloning [Cai 1990] of the DNA binding proteins interacting with the CDEI sequence of yeast centromeres

suggest the existence of multiple proteins. These proteins belong to the c-myc-related family of DNA-binding proteins, structurally defined by the presence of a helix-loop-helix domain and a basic region required for binding to the DNA target sequence [Murre 1989]. The members of this family have been shown to interact with the CAXXTG motif. However, as observed by several authors and reported in Gregor et al., 1990, a subset of these members, including USF/MLTF and TFE3, seems to be specific for the CACPuTG motif, present in the MLP upstream element and in the μE3 site, while the same proteins do not bind to the slightly different kE2 site found within the immunoglobulin light chain enhancer. On the contrary, other HLH proteins like E12 and E47, bind to the kE2 site and to a closely related sequence in the muscle-specific creatin-kinase promoter but not to μE3 [Murre 1989].

Altogether, these data suggest the existence of a large family of nuclear proteins, probably structurally related, with slightly different DNA binding specificities, as it is the case for the CREB/ATF family of DNA-binding proteins [Hai 1988]. The conservation of the binding domains of these proteins and of their cognate DNA target sequences, although in the context of different functions, represents an intriguing example of evolutionary tinkering.

A HUMAN BINDING SITE DOWNREGULATES TRANSCRIPTION OF HIV-LTR

Genetic and clinical data suggest that the overall rate of virus expression is regulated also by negative elements; in particular, the rate of transcription initiation is determined also by negative functions provided by domains present within the LTRs. A negative regulatory element (NRE) has been originally defined upstream of nucleotide -167 with respect to transcription start site, since the deletion of nucleotides -423 to -167 results in an increase in gene expression promoted by the downstream domains in CAT assay experiments [Rosen 1985]; however, the exact boundaries of this element have not been defined precisely so far [Siekevitz 1987]. Recently, in an *in vitro* transcription system, a gradual enhancement of basal transcriptional level was demonstrated with several deletions from the 5' LTR, with a maximum effect obtained by deletion up to -117 [Okamoto 1990]. This probably suggests that the overall negative effect cannot be ascribed to the action of a single factor.

Several binding sites for human nuclear proteins, with either positive or negative function, have been located within the NRE (see table at page 53), suggesting a functional heterogeneity of this region and a potentially complex mechanism of regulation, involving a balance between positive and negative cellular factors.

In particular, it has been demonstrated that deletion of the sequence between -173 and -159 results in an increased rate of

expression of an heterologous gene driven by the viral LTR and in more rapid viral replication compared to the parental strain [Lu 1990]. In this region, we have identified a binding site for a HeLa cell nuclear factor between nucleotides -174 and -152 by DNase I footprinting experiments, also in accordance to previously published results [Garcia 1987]. The center of the 23 bp protected region contains the dyad symmetry element CACGTG (position -166), identical to the sequence present in the upstream element of the Major Late Promoter of Adenovirus (MLP-UE), located between nucleotides -63 and -52 upstream of the start site of the late transcripts. The MLP-UE is a target for the human transcription factor USF/MLTF [Sawadogo 1985] [Carthew 1985], a 43-44 KDa ubiquitous protein [Sawadogo 1988]. We have described the presence of a binding site for this factor in an early-replicated human DNA sequence (pB48 Binding Site), possibly involved in the control of initiation of DNA replication [Tribioli 1987], [Falaschi 1988], [Giacca 1989] (see next chapter). Gel retardation and competition experiments indicate that a common binding activity is present in human cells for oligonucleotides corresponding to the HIV-1 LTR sequence between - 174 and -151, to pB48 Binding Site and to the MLP-UE, but not for an oligonucleotide with the same sequence of pB48 Binding Site except for a double point mutation of the core CpG dinucleotide [Giacca 1992].

The functional role of the USF/MLTF binding site within the HIV-1 LTR was investigated with relation to the efficiency of promotion of transcription by deletion/replacement mutants. A 100 bp fragment of

human origin, containing pB48 Binding Site at one extremity, is able to replace most of the negative function exerted by the NRE when cloned upstream of deletion -159 after transfection in HeLa and COS-1 cells. This effect is entirely due to the USF/MLTF binding site, since it can be obtained also by a 17 bp oligonucleotide corresponding exclusively to pB48 Binding Site. The downregulation of transcription is observed in presence or absence of tat transactivation and after HeLa cell stimulation with PMA, an activator of the protein kinase C pathway, which is able to increase viral expression through the NF-kB and NFAT-1 binding sites [Siekevitz 1987] [Crabtree 1989].

The pB48 Binding Site oligonucleotide is effective when cloned in both orientations, which is expected since the core of the recognized DNA sequence is the dyad symmetry element CACGTG. A similar situation is observed for the adenoviral genome, where the late transcription units and the IVa2 gene are transcribed in opposite directions, and, share overlapping control regions including two USF/MLTF binding sites with stimulatory activity in both directions [Moncollin 1990]. On the contrary, the human 100 bp fragment is active only in one orientation, namely the one containing the binding site close to the downstream domains. The most likely explanation for this result is that the protein interacting with the USF/MLTF site acts through direct protein-protein interaction with other proteins bound downstream. This hypothesis is in agreement with an observation made by Garcia et al. [Garcia 1987], from DNase I footprinting data, showing that the deletion of the NRE (including the USF/MLTF site)

results in decreased protection over the enhancer region.

WHAT IS THE MEANING OF *CIS* ACTING TRANSCRIPTION SIGNALS NEARBY AN ORIGIN OF DNA REPLICATION ?

One of the most challenging problems of eukaryotic molecular biology is the regulation of the initiation of DNA replication. In mammals this field of research has been hampered by the lack of an assay system for testing directly origin function. Such an assay is available for yeast, and therefore, research in this system is far ahead, to the point that an origin recognition complex (ORC) composed of six proteins has been identified. This complex specifically interacts with the ARS element and is required for origin function, since point mutations which prevent the binding also eliminate origin function [Bell 1992]. Interestingly, *in vivo* footprinting experiments have revealed a footprint over the ARS element essentially identical to the one obtained *in vitro* using the purified ORC [Diffley 1992].

As already described in the Introduction (page 44), a common feature of DNA replication origins of yeast and viruses infecting eukaryotes is the presence of transcription signals. We have previously described the presence of active transcriptional signals in a human DNA sequence (pB48) synthesized by HL60 promyelocytic cells at the onset of S phase [Tribioli 1987], [Falaschi 1988]. Recently, we have shown that pB48 encompasses a bona fide origin of replication, by fine mapping of the replication fork movement [Norio 1992. Master thesis]. We have shown that this sequence is located between two transcription units and contains an active promoter [Falaschi 1988]. Furthermore, by gel retardation and *in vitro* footprinting, we have demonstrated that a nuclear

factor is able to bind to a specific segment of pB48 with a sequence corresponding to the consensus for the USF/MLTF transcription factor belonging to the helix-loop-helix family of proteins [Tribioli 1987].

One possible function of a transcriptional promoter nearby an origin of DNA replication is suggested by the transcriptional activation model proposed by Kornberg. Activation of an origin of DNA replication by transcription from an upstream promoter has been observed in λ phage and E. coli [Baker 1988], and it has been suggested to be due to the alteration of topological state of the origin. Transcription activation does not seem to occur in the pB48 human replicon; indeed, we showed by Northern blotting and run on experiments that its transcription is constant throughout the cell cycle (unpublished results, Demarchi, master thesis).

To address the question of the role of transcription signals nearby this origin of DNA replication, we focused our attention on the binding site for USF/MLTF mentioned above. We have shown that the target for this factor is contained in the minimal region required to promote transcription as assessed by CAT assay. To test directly whether this USF binding site could play a role in the initiation of DNA replication, we are planning to specifically mutagenize this region and to compare the replication fork movement on pB48 area in a cell line bearing the mutation and in the original wild type cell line.

Parallel work is in progress to study the traffic of proteins on the origin region by *in vivo* footprinting. By means of this technique we have shown so far that in living HeLa cells a clear footprint appears at the USF/MLTF target site (see previous chapter) confirming the results obtained by *in vitro* studies [Tribioli 1987]. Moreover, other three

previously unrecognized target sites located in its proximity, are fully engaged in protein-DNA interactions, as expected from the well demonstrated role of this region on the promotion of transcription [Falaschi 1988]. These results offer a reliable, static picture of the contact points of nuclear factors sitting on DNA at the pB48 region in HeLa cells. Studies are under way to extend this type of analysis to the HL60 cell line, where pB48 acts as an origin, in a time course experiment, after synchronization with aphidicolin. By this mean, we should be able to reconstruct also a dynamic picture of the traffic of factors on pB48 throughout the cell cycle, and ultimately to contribute to unraveling the mechanism of initiation of DNA replication in mammalian cells and to understand the role of transcription factors in mammalian replication origins.

A possible function for controlling DNA replication by transcription elements could be the regulation of tissue-specific DNA replication [Hatton 1988]. This could be an explanation for the relationship of the timing of DNA synthesis during early versus late S phase to its transcriptional activity. DNA replicated during early S phase appears to be correlated to highly transcribed DNA in specific tissues [Hatton 1988]. Furthermore, the frequency at which eukaryotic replicons show an association with transcriptional elements is consistent with the general involvement of cell-specific transcription factors in DNA replication.

Little is known about the mechanism by which cis-acting elements can affect the initiation of DNA replication. It is assumed that the binding of transcriptional factors must affect the accessibility of the origin, or the initiation activity of specific replication proteins. It has been suggested

that transcriptional enhancer elements present at viral origins of replication stimulate replication by perturbing the distribution of nucleosomes in the adjacent sequences, thus leaving the origin exposed for interaction with the replication machinery [Chang 1989]. It is expected however, that such a stimulatory mechanism will be sensitive to the distance of the enhancer from the origin core sequences. Therefore, this mechanism seems appropriate only for those enhancer sequences which require to be located in the proximity of the origin core. The yeast ARS enhancer, for example, is able to exert an activating function on DNA replication even when present at a distance from the core origin. In this case, it seems more likely that in order to stimulate the origin of replication, the ARS enhancer and its cognate protein have to interact with either the 3' auxiliary domain or the core origin itself. This could be accomplished by DNA looping that would place the protein bound to the enhancer in proximity to a target site, enabling a direct interaction of the regulatory protein with a component of the replication apparatus. Such a mechanism has been proposed for the action of some enhancer sequences and their cognate proteins in the regulation of transcriptional activation and repression [Ptashne 1986].

The involvement of transcriptional factors in cellular DNA replication suggests additional mechanisms by which transcriptional factors may also alter growth or transform cells by directly regulating DNA replication. The final elucidation of the mechanism of the enhancer function in replication, however, will require the identification and characterization of the remaining components of the replication initiation apparatus and ultimately the reconstruction of the regulatory apparatus *in*

vitro. The results obtained by *in vivo* footprinting studies should be a very valuable guide towards a physiologically appropriate reconstruction.

MATERIALS AND METHODS

Materials

Restriction and modification enzymes were purchased from New England Biolabs, Beverly, MA (U.S.A.) or Promega Corporation, Madison, WI (U.S.A.). [γ32P]-ATP, [α32P]-dNTPs (all at 110 TBq/mmol) and [14C]-chloramphenicol (at 2 GBq/mmol) were obtained from Amersham Corp. (UK).

Cell cultures

HeLa and COS-1 cells were cultured in Dulbecco's modification of Eagle's minimal essential medium (DMEM, GIBCO) in monolayer cultures; H9, H9/HIV-1HTLVIIIB, and U1 cells were grown in suspension in RPMI 1640 (GIBCO). All the cultures were additioned of 50 μg/ml gentamicin, 10% fetal calf serum and 2 mM glutamine.

Plasmid constructions

For the construction of LTR-CAT construct mutated in the SISSA site see Results section (pag.77).

pC15CAT contains a XhoI-HindIII fragment of HIV-1 cDNA clone c15 (Arya et al., 1985), encompassing a portion of the nef gene, the whole U3 and most of the R regions of the 3' LTR, cloned in the Hind III site of the pSV0CAT vector, upstream of the chloramphenicol-acetyltransferase (CAT) gene. It was a gift of Dr. Robert Gallo, NIH Bethesda. For competition experiments, a band corresponding to the LTR was excised

by Hind III digestion and purified from agarose gel.

pEM0 was constructed by deletion of the fragment AvaI (-159 upstream of transcription start site)-AvaI (-424) of plasmid pC15CAT, Klenow filling and ligation. The deleted fragment includes completely the Negative Regulatory Element, as defined by several authors [Rosen 1985] [Garcia 1987] [Lu 1989] [Tong-Starksen 1989].

pL15 is a pUC18-based plasmid containing a 106 bp AluI-AluI fragment from plasmid pB48, representing an early replicated fragment of human DNA [Tribioli 1987], from nucleotides 703 to 808, cloned by blunt-end ligation in the SmaI site of the vector. It contains a binding site (5'-TTCGTCACGTGATGCGA-3') for the same factors interacting with the MLP-UE [Giacca 1989].

pEM1 and pEM2 were constructed by blunt-end ligation of the insert of pL15, recovered by AccI/EcoRI digestion of the vector polylinker, to replace the AvaI-AvaI fragment of pC15CAT. All the extremities were filled with the Klenow fragment of E. coli polymerase I before ligation. The two plasmids differ for the relative orientation of the human insert, which was determined by restriction enzyme digestion and confirmed by sequencing with the dideoxynucleotide method [Sanger 1977].

pLTRCAT, pLTRΔCAT, pLTRΔBS1CAT and pLTRΔBS2CAT are pBlueScript KS (Stratagene, La Jolla CA, USA)-derivatives containing, respectively, the whole HIV-1 LTR, the LTR deleted upstream of position -159, the (-159) deleted LTR with an upstream cloned oligonucleotide corresponding to pB48 Binding Site in both orientations. All these constructs were placed upstream of the CAT gene and its polyadenylation sites of plasmid pULB3574, a gift of Dr. Pierre Spegelaere (ULB,

Bruxelles), obtained by cloning between the BamHI-HindIII sites of the vector a 1600 bp PCR amplification product of the CAT-SV40 region of plasmid pSV2CAT.

pSV2-tat contains the first exon of the HIV-1 tat protein under the control of the SV40 early promoter. It was obtained by cloning of the HindIII-KpnI fragment of HIV-1 (nucleotides 6026 to 6350 of clone HXB2) in the vector pSV2-gpt, to replace the first part of the gpt gene. It was a gift of Dr. A. Meyerhaus (Institut Pasteur, Paris).

CAT ASSAY

HeLa cells were transfected with recombinant plasmids by the calcium phosphate precipitation technique [Graham1973]. 5 X 105 cells per dish were plated on the previous day in DMEM. 10 mg of high molecular wheight calf thymus DNA as carrier (Boehringer Manheim) were used for each transfection. H9 cells were transfected by DEAE-Dextran mediated transfection, according to standard procedures . 48 h after transfection, cells were harvested and extracts were prepared and heated at 60^0 C for 10 min. CAT assays were performed according to the method of Gorman [Gorman 1982]. Reactions were carried out at 37^0 C for 15 min with 0,5 mCi of [14C] chloramphenicol.

GEL RETARDATION ASSAY AND COMPETITION ANALYSIS

DNA fragments for gel retardation assays were isolated by electrophoresis on agarose gels and extracted using DEAE membranes. Oligonucleotide were purified from denaturing acrylamide gels, and

annealed. The concentration of purified fragments and oligonucleotides were determined by ethidium bromide staining [Maniatis 1982]. Plasmid inserts and annealed oligonucleotide to be used as probes in band shift experiments, were end-labelled either with T4 polynucleotides kinase or Klenow fragment of E. coli DNA polymerase I as previously described [Falaschi 1988]. Binding reactions were carried out by incubating 10^4 cpm of end labeled probe with 4 µg of HeLa cell nuclear extract prepared by the method of Dignam [Dignam 1983] and 3 mg of poly[d(I-C):poly[d(I-C) (Boehringer Manheim) in 20mM Hepes, pH 7.3, 50 mM NaCl, 4 mM MgCl2, 2 mM DTT, 0,2 mM EDTA, 4 mM spermidine, 5% glycerol (final volume 20 µl). After 20 minutes incubation at room temperature, the protein-DNA complexes were resolved on a low-ionic strength 5% polyacrilamide gel.

Competition experiments were carried out by mixing a 10 to 40 fold molar excess of cold plasmid inserts or 20 to 180 fold molar excess of cold oligonucleotides to the probe before incubation with the nuclear extract

RNA PREPARATION AND NORTHERN BLOT ANALYSIS

Total RNA was prepared using the guanidinium thiocyanate procedure [Maniatis 1989]. After gel electrophoresis (1% agarose) in 0,2 M formaldheyde [Lehrach 1977], the RNA (10 µg) was blotted to nitrocellulose membrane (Gene Screen) in 20 X SSC and cross linked by UV treatment. Hybridization with [α^{32}P]dCTP-labeled probes (10^6 cpm/ml) prepared by nick translation [Rigby 1977] was performed in a solution of

45% (v/v) formamide, 0,2 M sodium phosphate (pH 7,2), 1 mM EDTA, 7% SDS, and 100 µg/ml denatured salmon sperm DNA at 450C for 16 hr [Joyner 1985]. The membranes were subsequently washed in 40 mM sodium phosphate (pH7,2), 1% SDS, at 55^0 C for 1 hour [Church 1984].

IN VIVO FOOTPRINTING IN H9-HTLVIII INFECTED CELLS

Ten millions of exponentially growing cells were treated by adding to the medium freshly prepared DMS (0.1% final concentration) for 5 minutes. Under these conditions, partial methylation of guanine residues at position N-7 is obtained [Ephrussi 1985]. The reaction was stopped by the addition of 10 volumes of ice-cold PBS; cells were collected by centrifugation and washed twice with PBS. Nuclei were isolated by lysis with 0.1% NP40 and DNA extracted following standard procedures. In parallel, 200 µg of naked genomic DNA (1 µg/µl in water) extracted from untreated cells were reacted in vitro with 1 µl DMS for 30 seconds at room temperature. The reaction was stopped by the addition of b-mercaptoethanol (0.2 M final) and DNA precipitated twice with sodium acetate. G+A and C+T reactions were performed on naked genomic DNA according to the sequencing protocol of Maxam and Gilbert [Maxam 1980]. Restriction enzyme digestion of the DNA preparations, strand scission with piperidine and precipitation were performed according to Pfeifer and Riggs [Pfeifer 1991]. LM-PCR was performed essentially as described by Mueller and Wold [Mueller 1989] with four sets of oligonucleotides able to evidentiate both coding and non coding strand of the U3 region of HIV-1 LTR (Figure 2, Results session). For each set,

primer 1 was annealed to DNA and extended with Sequenase to generate molecules that have blunt ends on one side. A common linker (same as described by Mueller and Wold [Mueller 1989]) was ligated to the blunt ends and PCR of the ligated fragments was performed using the longer oligonucleotide of the linker (linker primer [Mueller 1989]) and primer 2 of each set. After 18 cycles, the amplification products were labeled by 9 cycles of primer extension with 32P-end labeled primer 3 and visualized by sequencing gel electrophoresis. To overcome problems due to high G+C content, the nucleotide analog 7-deaza-2'-deoxyguanosine 5'-triphosphate (dc7GTP) was included in the primer extension and the PCR reaction steps [McConlgue 1988].

TPA ACTIVATION OF U1 CELLS

3 µl of 5mM TPA dissolved in DMSO and 13 µl of DMSO are well mixed and added to 284 ul di complete medium.This solution (500X) must be used immediately and cannot be kept on ice, or frozen.The working solution is obtained by mixing 300 µl of 500X solution and 149 ml e 700 µl of complete medium. The final concentration of TPA is 10^{-7} M.

IN VIVO FOOTPRINTING IN U1 CELLS

The conditions used were the same as describe above for in vivo footprinting in H9 cells, exept that *in vivo* DMS-treated samples were previously incubated with TPA. Cells were resuspended in TPA containing medium at a concentration of 0,5 $X10^6$ cells/ml and incubated at 37°C for the different time periods (see Results section, pag.63).

REFERENCES

_ Arya S.K. and Gallo R.C. 1986. Three novel genes of human T-lymphotropic virus type III: immune reactivity of their products with sera from acquired immunodeficiency syndrome patients. Proc. Natl. Acad. Sci. USA 83: 2209- 2213.

_ Bachelerie F., Alcami J., Arenzana-Seisdedos F., Virelizier J.L. HIV enhancer activity perpetuated by NF-kB induction of infection of monocytes. Nature 350: 709-712

_ Baldwin A.S., Sharp P.A. 1988. Two transcription factors, NF-kB and H2TF1, interact with a single regulatory sequence in the class I major histocompatibility complex promoter. Proc. Natl. Acad. Sci. USA 85: 723-727

_ Baldwin A.S., Le Clair K.P., Harinder S., Sharp P.A. 1990. A large protein containing zinc finger domains binds to related sequences elements in the enhancers of class I major histocompatibility complex and kappa immunoglobulin genes. Mol. Cell. Biol. 10: 1406-1414

_ Ballard D.W., Walker W.H., Doerre S., Siste P., Molitor J.A., Dixon E.P., Peffer N.J., Hannink M., Greene W.C. 1990. The v-rel oncogene encodes a kB enhancer binding protein that inhibits NF-kB function. Cell 63: 803-814

_ Baltimore D. 1970. RNA dependent DNA polymerase in virions of RNA tumor viruses. Nature 226: 1209-1211

_ Barre-Sinoussi F., Chermann .C., Rey F., Nugeryre M.T., Chamaret S., Gruest J., Dauguet J., Axler-Blin C., Vezinet W., Rouzioux C., Rozenbaum W., Montagnier L. 1983. Isolation of T lymphotropic retrovirus from a patient at risk for acquired immunodeficiency syndrome AIDS. Science 220: 866-870

_ Bauerle P.A., Baltimore D. 1988. IkB: a specific inhibitor of the NF-kB transcription factor. Science 242: 540-546

_ Becker P.B., Ruppert S., Schutz G. 1987. Genomic footprinting reveals cell type-specific DNA binding of ubiquitous factors. Cell 51: 453-443

_ Benko D.S., Schwartz G., Pavlakis G., Felber B. 1990. A novel human immunodeficiency virus type 1 protein tev, shares sequences with tat and rev proteins. J. Virol.64: 2505-2518

_ Berger S. L., Cress W. D., Cress A., Triezenberg S. J., Guarente L. 1990. Selective inhibition of activated byt not basal transcription by the acidic activation domain of VP16: evidence for transcriptional adaptors. Cell 61: 1199-1208

_ Berkout B., Jeang K.T. 1989. Trans- activation of human immunodeficiency virus type 1 is sequence specific for both the single stranded bulge and loop of the trans-acting responsive hairpin: a quantitative analysis. J. Virol. 63:5501-5504

_ Berkout B., Gatignol A., Rabson A., Jeang K.T. 1990. TAR independent activation of the HIV-1 LTR: evidence that tat requires specific regions of the promoter. Cell 62: 757-767

_ Bohnlein S., Hauber J., and Cullen B.R. 1989. Identification of a U5-specific sequence J. Virol. 63: 421-424

_ Braddock M., Chambers A., Wilson W., Esnouf M., Adams S., Kingsman S., Kingsman A. 1989. HIV-tat "activates" presinthetized RNA in the nucleus. Cell 58: 269-279

_ Buratowski S., Hahn S., Guarente L., Sharp P.A. 1989. Five intermediate complexes in transcription initiation by RNA polymerase II. Cell 56: 546-561

_ Carthew, R.W., Chodosh, L.A. and Sharp, P.A. 1987. The major late transcription factor binds to and activates the mouse metallothionein I promoter. Genes Dev. 1: 973-980

_ Chakrabarti L., Guyader M., Alizon M., Daniel M., Desrosiers R.C., Tsiollais P., Sonigo P. 1987. Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. Nature 328: 543-547

_ Chodosh, L.A., Carthew, R.W. and Sharp, P.A. 1986. A single polypeptide possesses the binding and transcription activities of the adenovirus major late transcription factor. Mol. Cell. Biol. 6: 4723-4733

_ Church 1984. Proc. Natl. Acad. Sci. USA 81: 1991

_ Clark L., Pollak M., Hay R.T. 1988. Identification and purification of EBP1: a HeLa cell protein that binds to a region overlapping the core of the SV40 enhancer. Genes Dev. 2: 991-1002

_ Clouse K., Powell D., Washington I., Poli G., Strebel K., Farrar W., Barstad P., Kovacs J., Fauci A., Folks T. 1989. Monokine regulation of human immunodeficiency virus-1 expression in a chronically infected human T cell clone. J. Immunol. 142: 431-438

_ Cohen E. A., Terwilliger E. F., Jalinoos Y., Proulx J., Sodroski J.G., Haseltine W.A. 1990. Identification of the HIV-1 vpr product and function. J. Acquir. Immune Defic. Syndrome 3: 11-18

_ Cullen B. 1986. Trans-activation of human immunodeficiency virus occurs via a bimodal mechanism. Cell 46: 973-982

_ Cullen B., Hauber J., Campbell K., Sodroski J., Haseltine W., Rosen C. 1988. Subcellular localization of the human immunodeficiency virus 1 trans-acting art gene product. J.Virol. 62: 2498-2501

_ Curran J., Jaffe H.W., Hardy A., Morgan W., Selik R., Dondero T. 1988. Epidemiology of HIV infection and AIDS in the United States. Science 239: 610-616

_ Dayton A.I., Sodroski J.G., Haseltine W.A. 1986. The trans-activator gene of the T cell lymphotropic virus type III is required for replication. Cell 44: 941-947

_ Davis M., Kenney S. C. Kamine J., Pagano J.S., Huang E. 1987. Immediate-early gene region of human cytomegalovirus trans-activates the promoter of human immunodeficiency virus .Proc. Natl. Acad. Sci. USA. 84: 8642-8646

_ Dean D., Blakeley M. Newby R., Ghazal P. Hennighausen L., Bourgeois S. 1989. Forskolin inducibility and tissue-specific expression of the fibronectin promoter. Mol. Cell. Biol. 4: 1498-1506

_ Demarchi F., D'Agaro P., Falaschi A., Giacca M. 1992. Probing protein-DNA interactions at the Long Terminal Repeat of human immunodeficiency virus ttype-1 by in vivo footprinting. J. Virol. 4: 2514-2518

_ Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucl. Acids Res.11: 1475-1491

_ Di Luca D., Secchiero P., Bovenzi P., Rotola A., Caputo A., MOnini P., Cassai E. 1991. Reciprocal in vitro interaction between human herpesvirus-6 and HIV-1 Tat. AIDS 9: 1095-1098

_ Dougherty J., Temin H. 1988. Determination of the rate of base pair substitution and insertion mutations in retrovirus replication. J. Virol. 62: 2817-2822

_ Ensoli B., Lusso P., Schachter F. 1989. Human herpes virus-6 increases HIV-1 expression in coinfected T cells via nuclear factors binding to the HIV-1 enhancer. EMBO J. 8: 3019-3027

_ Ensoli B., Barillari G., Salahuddin S., Gallo R., Wong-Staal F. 1990. Tat protein of HIV-1 stimulates growth of cells derived from Kaposi sarcoma lesions of AIDS patients. Nature 345: 84-86

_ Ephrussi, A., G.M. Church, S. Tonegawa, and W. Gilbert. 1985. B lineage-specific interactions of an immunoglobulin enhancer with cellular factors in vivo. Science 227: 134-140

_ Falaschi, A., Biamonti, G., Cobianchi, F., Csordas-Toth, E., Faulkner, G., Giacca, M., Pedacchia, D., Perini, G., Riva, S. and Tribioli, C. 1988. Presence of transcription signals in two putative DNA replication origins of human cells. Biochem. Biophys. Acta 951: 430-442

_ Fan C.M., Maniatis T. 1990. A DNA binding protein containing two widely separeted zinc finger motifs that recognize the same DNA sequence. Genes.Dev. 4: 29-42

_ Felber B., Hadzopoulou-Cladaras M., Cladaras C., Copeland T., Pavlakis G. 1989. Rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. Proc. Natl. Acad. Sci. USA 86: 1495-4199

_ Felber B.K., Drysdale C.M., Pavlakis G.N. 1990. Feedback regulation of human immunodeficiency virus type 1 expression by Rev protein. J. Virol. 64: 5448-5456

_ Feng S., Holland E. 1988. HIV-1 Tat trans-activation requires the loop sequence within Tar. Nature 334: 165-167

_ Fields A., Bednarik D., Hess A., May W. 1988. Human immunodeficiency virus induces phosphorylation of its cell surface receptor. Nature 333: 278-280

_ Fisher A.G. Feinber M.B., Josephs F.S. 1986. Nature 320: 367-371

_ Fisher A.G., Ensoli B., Ivanoff L., Chamberlain M., Petteway S., Ratner L., Gallo R.C., Wong-Staal F. 1987 The sor gene of HIV-1is required for efficient viral transmission in vitro. Science237: 888-893

_ Flanagam P.M., Kelleher R.J., Sayre M.H., Tschochner H., Kornberg R.D. 1991. A mediator required for activation of RNA polymerase II transcription *in vitro*. Nature 350: 436-438

_ Folks T., Justement J., Kinter A., Schnittman S., Orenstein J., Poli G., Fauci A. 1988. Charachterization of a promonocyte clone chronically infected with HIV and inducible by 13-phorbol-12-myristate acetate. J. of Immunol.: 1117-1122

_ Franza B., Josephs S.F., Gillman M., Ryan W., Clarkson B. 1987. Characterization of cellular proteins recognizing the HIV enhancer using a microscale DNA-affinity precipitation assay. Nature 330: 391-395

_ Franza, B.R., Jr., F.J. Rauscher III, S.F. Josephs, and T. Curran. 1988. The fos complex and fos-related antigens recognize sequence elements that contain AP1 binding sites. Science 239: 1150-1153

_ Franzoso G., Bours V., Park S., Tomita M., KellyK., Siebenlist U. 1992. The candidate oncoprotein Bcl-3 ia an antagonist of p50/NF-kB-mediated inhibition. Nature 359: 339-342

_ Frankel A., Pabo C. 1988. Cellular uptake of the Tat protein from human immunodeficiency virus. Cell 55: 1189-1193

_ Fujita T., Nolan G.P., Ghosh S., Baltimore D. 1992. Independent modes of transcriptional activation by the p50 and p65 subunits of NF-kB. Genes Dev. 6: 775-787

_ Fultz P., McClure H., Daugharty H., Brodie A., McGrath C., Swenson B., Francis D. Vaginal transmission of human immunodeficiency virus (HIV) to chimpanzee. J. Infect. Dis. 154: 896-900

_ Garcia, J.A., F.K. Wu, R. Mitsuyasu, and R.B. Gaynor. 1987. Interactions of cellular proteins involved in the transcriptional regulation of the human immunodeficiency virus. EMBO J. 6: 3761-3770

_ Gatignol A., Kumar A., Rabson A., Jeang K.T. 1989. Identification of cellular proteins that bind to the human immunodeficiency virus type 1 trans -activation TAR responsive element RNA. Proc. Natl. Acad. Sci. USA 86: 7828-7832.

_ Gabuzda D.H., Lawrence K., Langhoff E., Terwilliger E., Dorfman T., Haseltine W., Sodroski J. 1992. Role of vif in Replication of Human Immunodeficiency virus type-1 in CD4 T lymphocytes. J. Virol. 66: 6489-6495

_ Garcia, J.A., D. Harrich, E. Soultanakis, F. Wu, R. Mitsuyasu, and R.B. Gaynor. 1989. Human immunodeficiency virus type 1 LTR TATA and TAR region sequences required for transcriptional regulation. EMBO J. 8:765-778

_ Garcia J., Ou S.H., Wu F., Lusis A., Sparkes S., Gaynor R. 1992. Cloning and chromosomal mapping of a human immunodeficiency virus 1 TATA element modulatory factor. Proc. Natl. Acad. Sci. 89: 9372- 9376

_ Gaynor R., Soultanakis E., Kuwabara M., Garcia J., Sigmon D. 1989. Specific binding of a HeLa nuclear protein to RNA sequences in the human immunodeficiencyvirus transactivation region. Proc. Natl. Acad Sci. USA 86: 4858-4862

_ Gendelman H.E., Phelps W., Feigenbaum L., 1986. Transactivation of the human immunodeficiency virus long terminal repeat sequence by DNA viruses. Proc. Natl. Acad. Sci. USA 83: 9759-9763

_ Gentz R., Chen C.H. Rosen C. 1989. Bioessay for trans-activation using purified human immunodeficiency virus tat-encoded protein: trans-activation requires mRNA synthesis. Proc. Natl. Acad. Sci. USA 86: 821-824

_ Gerondakis S. 1991. Dangerous liasons. Current Biology 2: 103-105

_ Giacca, M., Gutierrez, M.I., Demarchi, F., Diviacco, S., Biamonti, G., Riva, S. and Falaschi, A. 1989. A protein target site in an early replicated human DNA sequence: a highly conserved binding motif. Biochem. Biophys. Res. Commun., 165:956-965

_ Giacca, M., M.I. Gutierrez, S. Menzo, F. d'Adda di Fagagna, and A. Falaschi. 1991. A human binding site for transcription factor USF/MLTF mimics the negative regulatory element of human

immunodeficiency virus type 1. Virology 186: 133-147

_ Goedert J, Kessler C., Aledort L. 1989. A prospective study of human immunodeficiency virus type 1 infection and the development of AIDS in people with hemophilia. New England J. Med. 321: 1121-1148

_ Gorelick R.J., Henderson, L.E., Hanser, J.P. and Rein A. 1988. Point mutants of Moloney murine leukemia virus that fail to package viral RNA: evidence for a specific RNA recognition by a 'zinc finger-like' protein sequence. Proc. Natl. Acad. Sci. USA 85: 8420-8424

_ Gorman C. M., Moffat L. Howard B. 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. 2: 1044-1051

_ Gosh D. 1992. Glucocorticoid receptor binding site in the human immunodeficiency virus long terminal repeat. J. Virol. 66: 586-590

_ Gowda S., Stein B.S,, Steimer K.S. and EnglemanE. G. 1989. Expression and processing of the human immunodeficiency virus type 1 gag and pol genes by cells infected with a reconbinant vaccinia virus. J. Virol. 63: 1451-1454

_ Graham F., Van der Eb A.J. 1973. A new tecnique for the assay of infectivity of human adenovirus 5 DNA. Virology 52: 456-467

_ Green L.M. and Berg J.M. 1989. A retroviral Cys-Xaa2-Cys-Xaa4-His -Xaa4-Cys peptide binds metal ions: Spectroscopic studies and a proposed three-dimensional structure. Proc. Natl. Acad. Sci. USA 86: 4047-4051

_ Greenblatt J. 1991. Roles of TFIID in transcription initiation by RNA polymerase II. Cell 66: 1067-1070

_ Griffin G., Leung K., Folks T., Kunkel S., Nabel G. Activation of HIV gene expression during monocyte differentiation by induction of NF-kB. Nature 339: 70-73

_ Guy B., R.B. Acres, M.P. Kieny, and J.P. Lecocq. 1990. DNA binding factors that bind to the negative regulatory element of the human immunodeficiency virus-1; regulation by nef. J. Acquired Immune Deficiency Syndromes 3: 797-809.

_ Guy B., Geist M., Dott K., Spehner D., Kieni M.P., Lecocq J.P.1991.

A specific inhibitor of cysteine proteases impairs a Vif dependent modification of human immunodeficiency virus type 1 Env protein. J. Virol. 65: 11325-1331

_ Guo Z.S. De Phamphilis M.L. 1992. Mol. Cell. Biol. 12: 2514-2524

_ Hahn B.H., Shaw G.M., Taylor M., 1986. Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk of AIDS. Science 232: 1548-1553

_ Hammes S.R., Dixon E.P., Malim M.H., Cullen B.R., Greene W.C. 1989. Nef protein of human immunodeficiency virus type-1: evidence against its role as a transcriptional inhibitor. Proc. Natl. Acad. Sci. USA 86: 9549-9553

_ Harrich D., Garcia A., Mitsuyasu R., Gaynor R.B. 1990. TAR-independent activation of the human immunodeficiency virus in phorbol ester stimulated limphocytes and macrophages. EMBO J. 9: 4417-4423

_ Haseltine W.A. 1991. Human immunodeficiency virus (HIV0 gene expression and function. The human retroviruses. Academic Press, Inc.

_ Hatton K.S., Dhar V., Brown E., Iqbal M. A, Stuart S. Didamo V., Schildkraut C.L. 1988. Replication program of active and inactive multigene families in mammalian cells. Mol. Cell. Biol. 8: 2149-2158

_ Hauber J., Bouvier M., Malim M.H., Cullen B. 1988. Phosphorilation of the Rev gene product of human immunodeficiency virus type 1. J. Virol. 62: 4801-4804.

_ Israel A., Kimura A., Kieran M., Yano D., Karrellopaulos J., Le Bail O. Kourilsky 1987. A common positive trans-acting factor binds to enhancer sequences in the promoters of the mouse H2 and B2-microglobulin genes. Proc. Natl. Acad. Sci. USA. 84: 2653-2657

_ Jakobovitz A., Smith D., Jakobovitz E., Capon D. 1988. A discrete element 3' of human immunodeficiency virus 1 and HIV-2 mRNA initiation sites mediates transcriptional activatioon by an HIV trans activation. Mol. Cell. Biol. 8: 2555-2561

_ Joiner 1985. Cell 43: 29

_ Jones, K., J. Kadonaga, P. Luciw, and R. Tjian. 1986. Activation of the AIDS retrovirus promoter by the cellular transcription factor SP1. Science 232: 755-759.

_ Jones, K., P. Luciw, and N. Duchange. 1988. Structural arrangement of transcription control domain within the 5'-untranslated leader regions of HIV-1 and HIV-2 promoters. Genes Dev. 2: 1101-1114.

_ Jones N. 1991. Complex inhibitions. Current Biology 4: 224-226

_ Kan N.C., Franchini G., Wong Staal F., DuBois G., Robey W., LautenbergerJ., Papas T. 1986. Identification of HTLV III Lav sor gene product and detections of antibodies in human sera. Science 231: 1546-1549

_ Kao S., Calman A., Luciw P., Peterin B. 1987. Antitermination of transcription within the long terminal repeat of HIV by tat gene product. Nature 330: 489-483

_ Kato H., Horikoshi M., Roeder R.G. 1991. Repression of HIV-1 transcription by a cellular protein. Science 251: 1476-1479

_ Kelleeher R.J., Flanagan, Kestler-III H.W., Ringler D.J., Mori K., Panicalli D.L., Sehgal P.K., Daniel M.D., Desrosier R.C. 1991. Cell 65: 651-662

_ Kenney S., Kamine J., Markovitz D., Fenrick R. Pagano J. 1988. An Epstein Barr virus immediate early gene product trans-activates gene expression form the human immunodeficiency long terminal repeat. Proc. Natl. Acad. Sci. USA. 85: 1652-1656

_ Kieran M., Blank V., Logeat F., Vandekerckerckhove J., Lottspeich F., Le Bail O., Urban M.B., Kenrilsky P., Bauerle P.A., Israel A. 1990. The DNA binding subunit of NF-kB is identical to KBF-1 and homologous to the rel oncogene product. Cell 62: 1007-1018

_ Kim S., Birn J., Groopman J., Baltimore D. 1989. Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. J. Virol. 63: 3708-3713

_ Kim S., Ikenchi K., Byrn R., Groopman J., Baltimore D. 1989. Lack of negative influence on viral growht by the nef gene of human immunodeficiency vierus type-1. Proc. Natl. Acad. Sci. USA 86: 9544-

9548

_ Klaus G., Humphrey J., Kunkl A., Dongworth D. 1980. The follicular dendritic cell: its role in antigen presentation in the generation of immunological memory. Immunol. Rev. 53: 3-15

_ Kowalski M., Potz J., Basiripour L., Dorfmant T., Goh W. C., Terwilliger A., Dayton A., Rosen C., Haseltine W.,Sodroski J.1987 Science 237: 1351-1355

_ Kramer R.A, Schaber M.D.., Skalka A.M.., Ganguli K., Wong-Staal F., Reddy E.P. 1986. Science 231: 1580-1584

_ Larder B. a., Purifoy D.M., Powell K.L., Darby G 1987. Site-specific mutagenesis of AIDS virus reverse transcriptase.Nature 327: 716-717.

_ Lasky L.A., Nakamura G. M., Smith D.H., Fennie C.,ShimasakiC., Patzer E., Berman P.W., Gregory T., Capon D.J. 1987 Delineaiton of a region of the human immunodeficiency virus type I gp 120 glycoprotein critical for the interaction with the CD4 receptor. Cell 50: 975-985

_ Laspia M., Rice A., Matthews M. 1989. Hiv-1 tat protein increases transcriptional initiation and stabilizes elongation. Cell 59: 283-292

_ Lee T.H., Coligan J.E., Allan J.S., McLane M.F., Groopman J.E., Essex M. 1986. A new HTLV-III Lav protein encoded by a gene foud in cytopathic retroviruses. Science 231: 1546-1549

_ Lee W.R., Syu W. J., Du B., Matsuda M., Tan S. C. , Wolf A., Essex M., Lee T. H. 1991 J. Virol. 65: 6349-6352

_ Le Grice S.F.J., Millis J., Mous J. 1988. Active site mutagenesis of the AIDS protease and its alleviation by *trans* complementatiion. EMBO J. 7: 2547-2553

_ Lehrach 1977 Biochemestry 16: 4743

_ Leider J., Palese P., Smith F. 1988. Determination of the mutation rate of a retrovirus. J. Virol. 62: 3084-3091

_ Leonard J., Parot C., Buckler -White A.J., Turner W., Ross E.K., Martin M.A., Robson A.B. 1989. The NFkB binding sitess in the human immunodeficiency virus type 1 long terminal repeat are not required

for virus infectivity. J. Virol. 63: 4919-4924

_ Leonard C. K., Spellman M. W., Riddle L., Harris R. J., Thomas j.N. Gregory J. 1990. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type I recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in chinese hamster ovary cells. J. Biological Chemestry 265: 10373-10382

_ Lenardo, M.J., A. Kuang, A. Giffird, D. and Baltimore, D. 1988. NF-kB protein purification from bovine spleen: Nucleotide stimulation and binding site specificity. Proc. Natl. Acad. Sci. USA 85:8825-8829

_ Lenardo M. J., Baltimore D. 1989. NF-kB: a pleiotropic mediator of inducible and tissue-specific gene control. Cell 58: 227-229

_ Leonard, J., C. Parrot, A.J. Buckler-white, W. Turner, E.K. Ross, M.A. Martin, and A.B. Rabson. 1989. The NF-kB binding sites in the human immunodeficiency virus type 1 long terminal repeat are not required for virus infectivity. J. Virol. 63: 4919-4924

_ Lever A., Gottlinger H, Haseltine W. and Sodroski 1989. Identification of a sequence for efficient packaging of HIV-1 RNA into virions. J. Virol.63: 4085-4087

_ Levine M., Manley J.L. 1989. Transcriptional repression of eukaryotic promoters. Cell 59: 405-408

_ Li C., Lai C., SigmanD.m Gaynor R. 1992. Cloning of acellular factor, interleukin binding factor, that binds to NFAT-like motifs in the human immunodeficiency virus long terminal repeat. Proc.Natl. Acad. Sci. USA 88: 7739-7743

_ Liu J., Perkins N., Schmid R., Nabel G. 1992. Specific NF-kB subunits act in concert with Tat to stimulate human immunodeficiency virus type 1 transcription. J. of Virol. 6: 3883-3887

_ Lu, Y., N. Touzjian, M. Stenzel, T. Dorfman, J. G. Sodroski, and W.A. Haseltine. 1990. Identification of cis-acting repressive sequences within the negative regulatory element of human immunodeficiency virus type 1. J. Virol. 64: 5226-5229

_ Lu, Y., Stenzel, M., Sodroski, J.G. and Haseltine, W.A. 1989. Effects of Long Terminal Repeat Mutations on Human Immunodeficiency Virus Type 1 Replication. J. Virol. 63: 4115-4119

_ Luciw P.A., Cheng-Mayer C., Levy J.A. 1987. Mutational analysis of the human immunodeficiency virus: the orf B region downregulates virus replication. Proc. Natl. Acad. Sci. USA 84: 1434-1438

_ Maddon P., Dalgleish A., McDougal J., Clapham P., Weiss R., Axel R. 1986. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. Cell 47: 333-348

_ Maekawa T., Kurimoto M., Ishii S. 1991. USF-related transcription factor HIV-TF1 stimulates transcription of human immunodeficiency virus 1. Nucl. Acids Res. 19: 4689-4694

_ Malim M.H., Bohnlein S., Hauber J., Cullen B.R. 1989.Functional dissection of the HIV-1 Rev Trans-Activator. Derivation of a Trans-Dominant Repressor of Rev function. Cell 58: 205-214

_ Maniatis et al. 1989. Molecular Cloning: A Laboratory Manual CSH Laborotory, CSH, N.Y.

_ Marciniak R., Garcia - Bianco M., Sharp P. 1990. Identification and characterization of a HeLa nuclear protein that specifically binds to the trans-activation response (TAR) element of human immunodeficiency virus. Proc. Natl. Acad. Sci. USA. 87: 3624-3628

_ Maxam, A.M. and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. 65, 499-560

_ McConlogue L., M.A.D. Brow, and A. Innis. 1988. Structure-independent DNA amplification by PCR using 7-deaza-2'-deoxyguanosine. Nucleic Acids Res. 16: 9869

_ McCune J.M. 1991: HIV-1: the infectivity process in vivo. Cell 64: 351-363

_ McCune J., RabinL., Feinberg M., Lieberman M., Kosek J., Reyes G., Weissman I. 19988. Endoproteolytic cleavage of gp160 is required for the activation of human immunodeficiency virus. Cell 53: 55-67

_ Mc Dougal J.S., MawleA., Cort S.P., Nicholson J.K.A., Cross J.D., Hicks D., Sligh J. 1985. J. Immunol. 135: 3151

_ Mervis R.J., Ahmad N., Lillehoj E.P., Raum M.G., Salazar F.K.R., Chan H.W., Venkatesan S. 1988. J. Virol. 62: 3993-4002

_ Michael N.L., Vahey M., Burke D.S., Redfield R. 1992. Viral DNA and mRNA expression correlate with the stage of human immunodeficiency virus (HIV) type 1 infection in humans: evidence for viral replication in all stages of HIV disease. J. Virol. 66: 310-316

_ Mitchell P.J., Tjian R. 1989. Transcriptional regulation in mammalian cells by sequence specific DNA binding proteins. Science 245: 371-376

_ Mitsuya H., Yarchoan R., Broder S. 1990. Molecular targets for AIDS therapy. Science 249: 1533-1544

_ Muchardt C., Scaler J.S., Nirula A., Shurland D.L., Gaynor R.B. 1992. Regulation of human immunodeficiency virus enhancer function by PRDII-BFI and crel gene products. J.Virol. 1: 244-250

_ Mous J., Heimer E.P., Le Grice S.F.J. 1988. Processing protease and reverse tanscriptase from human immunodeficiency virus type 1 polyprotein in E. coli. J. Virol. 62: 1433-1436

_ Mueller, P.R. and B. Wold. 1989. In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. Science 246: 780-786

_ Muesing M., Smith D., Capon D. 1987. Regulation of mRNA accumulation by a HIV trans-activator protein. Cell 48: 691-701

_ Myers G., Joseph F.S., Rabson A. B., Smith T.I., Wong-Staal F 1988. Human Retroviruses and AIDS. Los Alamos Laboratory, Los Alamos, N.Mex.

_ Nabel, G., and D. Baltimore. 1987. An inducible factor activates expression of human immunodeficiency virus in T cells. Nature (London) 326: 711-71

_ Nabel G., Rice S., Knipe D. Baltimore D. 1988. Alternative mechanisms for activation of human immunodeficiency virus enhancer in T cells. Science 239: 1299-1302

_ Nakanishi Y., Masamune Y., Kobayashi N. 1991. A novel cis acting element that controls transcription of human immunodeficiency virus type 1 DNA, depending on cell type. J. of Virol. 65: 6334-6338

_ Nelbock P., Dilion P., Perkins A., Rosen C.1990. A cDNA for a protein that interacts with the immunodeficiency tat transactivator.

Science 248: 1650- 1653

_ Nishizuka Y. 1986. Studies and perspectives of protein kinase C. Science 233:305-312

_ Newell M.K., Haughn L.J., Maroun C.R., Julius M.H. 1990. Death of mature T cells by separate ligation of CD4 and the TCR for antigen. Nature 347: 286-289

_ Nolan G.P., Ghosh S., Hsiou-Chi L., Tempst P., D. Baltimore. 1991. DNA binding and IkB inhibition of the cloned p65 subunit of NFkB, a rel-related polypeptide. Cell 64: 961-969

_ Ogawa K., Shibata R., Kiyomasu T., Higuchi I., Kishida Y., Ishimoto A., Adachi A. 1989. Mutational analysis of the human immunodeficiency virus vpr open reading frame. J. Virol. 63: 4410-414

_ Olshevsky U., Helseth M., Furman C., Li J., Haseltine W., Sodroski J. 1990. Identification of individual human immunodeficiency virus type I gp120 amino acids important for CD4 binding . J. Virol. 64: 5701-5707

_ Orchard, K., N. Perkins, C. Chapman, J. Harris, V. Emery, G. Goodwin, D. Latchman, and M. Collins. 1990. A novel T-cell protein which recognizes a palindromic sequence in the immunodeficiency virus long terminal repeat. J.Virol. 64: 3234-3239

_ Pauli U., Chrysogelos G., Stein G., Stein J., Nick H. 1987. Science 236: 1308-1311

_ Pavlakis G., Schwartz S., Benko D., Drysdale M., Solomin L.,Ciminale V., Harrison J., Campell M., Felber B. 1991. Genetic structure and regulation of HIV. edited by William A. Haseltine and Fossie Wong-Staal. Raven Press Ltd. New York

_ Pfeifer, G.P., and A. Riggs. 1991. Chromatin differences between active and inactive X chromosome revealed by genomic footprinting of permeabilized cells using DNase I and ligation-mediated PCR. Genes Dev. 5:1102-1113

_ Pomerantz, R.J., D. Trono, M.B. Feinberg, and D. Baltimore. 1990. Cells nonproductively infected with HIV-1 exhibit an aberrant pattern of viral RNA expression: a molecular model for latency. Cell 61: 1271-1276

_ Polyanovsky O., Stepchencko A.G. 1990. Eukariotic transcription factors. Bio Essay 12: 205-210

_ Proudfoot N.J., Lee B.A., Monks J. 1992. Multiple sP1 binding sites confer enhancer-independent, replication-activated transcription of HIV-1 and globin promoters. The New biologist 4: 369-381

_ Ratner L., Haseltine W., Patarca R., Livak K.J., Starchich B., Josephs S.F., Doran E.R., Rafalski J.A., Whitehorn E.A., Baumeister K.,Ivanoff L., Petteway S.R., Pearson M.L., Lautenberger J.A., Papas T.S., Ghrayeb J., Chang N.T., Gallo R.C. Wong-Staal F. 1985. Nature 313: 277-283

_ Ratner L., Starcich B., Josephs S.F., Hahn B. h., Reddy E.P., Livak K.J., Petteway S. R., Pearson H.L. Haseltine W.A. ,Arya S.K., Wong Staal F. 1985. Polymorphism of the 3' open reading frame of the virus associated with the acquired immunodeficiency syndrome, human T-limphotropic virus type III. Nucl. Acids Res. 13: 8219-8229

_ Razzecca K., Pillemer E., Weissman I., Rouse R. 1986. In situ identification of idiotype-positive cells participitating in the immune response to phosphorylcholine. Eur. J. Immunol. 16: 393-399

_ Rigby 1977. J. Mol. Biol. 5: 2705

_ Riggs A. D., Pfeifer G.P., 1992. X-chromosome inactivatioon and cell memory. Trends in Genetics 8: 169-174.

_ Rivier D.H., Rine J. 1992. An origin of DNA replicatiion and a transcriptioon silencer require a common element. Science 256: 659-663

_ Rosen, C. A., J.G. Sodrosky, and W. A. Haseltine. 1985. The location of cis-acting regulatory sequences in the human T cell lymphotropic virus type III (HTLV-III/LAV) long terminal repeat. Cell 41: 813-823

_ Rosen C., Sodroski J., and Haseltine W. 1985. The location of cis-acting regulatory seuqences in the HTLVII-LAV LTR. Cell 41: 813-823

_ Roy S., Delling U., Chen C.H., Rosen C., and N. Sonenberg 1990. A bulge structure in HIV-TAR RNA is required for tat binding and tat mediated transactivation. Genes. Dev. 4: 1365-1373

_ Roy S. , Parkin N., Rosen C., Itovich J., Sonenberg N. 1990. Structural requirements for trans-activation of human immunodeficiency type 1 long terminal repeat -directed gene expression by tat; importance of base pairing, loop sequence, and bulges in the tat - responsive sequence. J. Virol. 64: 1402-1406

_ Ruben S. M., Klement J., Coleman T., Maher M., Chen C., Rosen C. 1992. I-Rel: a novel rel-related protein that inhibits NF-kB transcriptional activity. Genes De. 6: 745-760

_ Sakai K., Ma X., Gordienko I., Volski D. 1991. Recombinational analysis of a natural noncitopathic human immunodeficiency virus type 1 (HIV1) isolate: role of the vif gene in HIV-1 infection kinetics and cytopathicity. J. Virol. 65: 5765-5773

_ Saldaie M.R., Rappaport J., Benter J., Josephs S.F., WillisR., Wong-Staal F.. 1988. Missense mutations in an infectious HIV genome: functional mapping of tat anddemonstration of a novel rev splice acceptor. Proc. Natl. Acad. Sci. USA 85: 9224-9228

_ Sanger F., Miklen S., Coulson A.R. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463-5467

_ Sawadogo M., Roeder R.G. 1985. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell 43: 165-175

_ Sawadogo M., Van Dike M.W., Gregor P.D.,Roeder R.G. 1988. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. J. Biol. Chem. 263: 11985-11993

_ Shibuia H., Irie K., Ninomiya J., Goebl M., Taniguchi T., Matsumoto K. 1992. New human gene encoding a positive modulator of HIV Tat-mediated transactivation. Nature 357: 700-702

_ Schnittman S., Psallidopoulos C., Lane H., Thompson L., Baseler M., Massari F., Fox C., Salzman P., Fauci A. 1989. The reservoir for HIV-1 in human peripheral blood is a T cell that maintains expression of CD4. Science : 305- 308

_ Schreck R., Rieber P., Bauerle A. 1991. Reactive oxygen intermediates as apparently widely used messangers in the activation

of the NF-kB transcription factor HIV-1. EMBO J. 10: 2247-2258

_ Schupbach J., Papovic M., Gilden R.V., Gonda M.A., Sarngadharan M.G., and Gallo R. C.1984. Science : 503-505

_ Schwartz S, Felber B.K., Benko D.M., Fenyo E.M., Pavlakis G.N. 1990. Env and Vpu rotein of the human immunodeficiency virus type 1 are produced from multiple bicistronic mRNA. J. Virol. 64: 2519-2529

_ Schwartz S., Felber B.K., Fenyo E.N., Pavlakis G.N. 1990. Env and Vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs. J. Virol.64: 5448-5456

_ Seelmeier R., Schmidt H., Turk V., von der Helm K. 1988 Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A. Proc. Natl. Acad. Sci. U.S.A. 85: 6612-6616

_ Selby M., Bain E., Luciw P., Peterlin B. 1989. Strcture, sequencem and position of the stem-loop in tar determine transcriptional elongation by tat through the HIV-1 LTR. Genes Dev. 3: 547-588

_ Sen R., Baltimore D. 1986. Inducibility of K immunoglobulin enhancer binding protein NF-kB by a post-translational mechanism. Cell 47: 921-928

_ Seto E., Yen T., Peterlin B., Ou J.H. 1988. Trans-activatioon of the human immunodeficiency virus long terminal repeat by the hepatitis B virus X protein. Proc. Natl. Acad. Sci. USA. 85:8286-8290

_ Sharp P., Marciniak. 1989. HIV TAR: an RNA enhancer? Cell 59: 229-230

_ Shaw, J.P., P.J. Utz, D.B. Durand, J.J. Toole, E.A. Emmel, and G.R. Crabtree. 1988. Identification of a putative regulator of early T cell activation genes. Science 241: 202-205

_ Sigurdsson B. 1954 Observations on three slow infections of sheep. Br. Vet. J. 110: 255-270Sigurdsson B.1954 Observations on three slow infections of sheep. Br. Vet. J. 110: 255-270

_ Smith, M.R., and W.C.Greene. 1989. The same 50 kDa cellular protein binds to the negative regulatory elements of the interleukin-2 receptor a-chain gene and the HIV-1 LTR. Proc. Natl. Acad. Sci. USA 86: 8526-8530

_ Sodroski J., Rosen C., Wong-Staal F., Popovic M., Arya S., Gallo R., HaseltineW. 1985 Trans-acting transcriptional regulation of human T-cell lymphotropic virus type III long terminal repeat.Science 227: 171-173

_ Southgate C., Zapp M., Green M. 1990. Activation of transcription by HIV-1 Tat protein tethered to nascent RNA through another protein. Nature 345: 640-642.

_ Starchich B., Ratner L., Josephs S.F., Okamoto T., Gallo R.C., Wong-Staal F. 1985. Charachterization of LTR sequences of HTLV-III. Science 227: 538-540

_ Steel C. M. Ludlam C.A. Beatson D. 1988. HLA haplotype A1 B8 DR3 as a risk factor for HIV-related disease. Lancet i: 1185-1188

_ Steimer K. S., Scandella C. J., Skiles P.V. Halgwood N.L. 1991. Science 253: 1-4

_ Stein B., Gowda D., Lifson J., Penhallow R., Beusch K., Engleman E. 1987. pH-independent HIV entry into CD4-positive T-cells via virus envelope fusion to the plasma membrane. Cell 49: 659-668

_ Steward R. 1987. Dorsal, an embryonic polarity gene in Drosophila is homologous to the vertabrate proto oncogene c-rel. Science 238: 692-694.

_ Strebel K., Klimkait T., Maldarelli F., Martin M.A. 1989. Molecular and biochemical analyses of human immunodeficiency virus type 1 vpu protein. J. Virol. 63: 3784-3791.

_ Talbot R.L., Sparger E.E., Lovelace K.M. Fitch W.M., Pedersen N.C., Luciw P.A., Helder J.K. 1989. Nucleotide sequence and genomic organization of feline immunodeficiency virus. Proc. Natl. Acad. Sci. USA 86: 5743-5747

_ Temin H., Mitzutani S. 1970. RNA-directed DNA polymerase in virions of Rous sarcoma virus.Nature 226: 1211-1213

_ Terwilliger E., Sodroski J. G., Rosen C.A., Haseltine W.A. 1986. Effect of mutations within the 3' orf open reading frame region of human T-cell lymphotropic virus type III (HTLV-III/LAV) on replication and cytopathogenicity. J. Virol. 60: 754-760

_ Terwillinger E.F., Sodroski J.C., Haseltine W., Rosen C.A. 1988. The art gene product of human immunodeficiency virus type 1 is required for replication. J. Virol. 62: 655-658

_ Tiley L., Madore S., Malim M., Cullen B. 1992. The VP16 transcription activation domain is functional when targeted to a promoter-poximal RNA sequence. Genes Dev. 6: 2077-2087

_ Tong-Starksen S. E., Luciw P.A., Peterlin B.M. 1989. Signaling through T lymphocyte surface proteins, TCR/CD3 and CD28, activates the HIV-1 long terminal repeat. J. of immunol. 142: 702-707

_ Tribioli, C., Biamonti, G., Giacca, M., Colonna, M., Riva, S. and Falaschi, A. 1987. Characterization of human DNA sequences synthesized at the onset of S-phase. Nucl. Acids Res. 15: 10211-10232

_ Vallee M., Carre' M. 1904 Sur la nature infectieuse de l'anemie du cheval.Compte Rendu de l'Academie de Science 139: 331-333

_ Venkatesan A., Cheng-Mayer C., Iannello P., Show K., Luciw P.A., Levy J.A. 1989. Differential effects of nef on HIV replication: implications for viral pathogenesis in the host. Science 246: 1629-1632

_ Veronese F.D.M., Rahmaan R., Copeland T.D., Oroszlan S. Gallo R.C., Sarngadharan M.G. 1987. AIDS Res. Hum. Retroviruses 3: 253-264

_ Weber R., Ledergerber B., Opravil M., Sieghenthaler W., Luthy R. 1990. Progression of HIV infection in misurers of injected drugs who stop injecting or follow a program of maintenance tratment with methadone. Br. Med. J. 301: 2-5

_ Webster A. 1992. Viral cofactors in acquired immunedeficiency syndrome. Reviews in Medical Virol. 2: 29-34

_ Wilson W., Braddock M., Adams S.E., Rathjen P.D., Kingsman S., King A. 1988. HIV expression strategies: ribosomal frameshifting is directed by a short sequence in both mammalian and yeast systems. Cell 55: 1159-1169

_ Wu, F., J. Garcia, R. Mitsuyasu, and R. Gaynor. 1988. Alterations in binding characteristics of human immunodeficiency virus enhancer factor. J. Virol. 62: 218-225

_ Yamamoto K., Mori S., Okamoto T., Shimotohno K., Kyogoku Y. 1991. Identification of transcriptional suppressor proteins that bind to the negative regulatory element of the human immunodeficiency virus type 1. Nucl. Acids Res. 19: 6107-6112

_ Zack J.S., Arrigo S., Wetsman S., Go A., Haislip A., Chen I. 1990. HIV-1 entry into quiescent primary lymphocytes: molecular analysis reveals a labile, latent viral structure. Cell 61: 213-222

ABSTRACT

The regulation of HIV-1 gene expression plays a key role in triggering the replication of the integrated provirus,, and consequently in disease development. A large number of cis-acting viral regulatory elements has been identified in the LTR of HIV-1 by *in vitro* binding studies. Aim of this study was to determine which are the DNA sites actually occupied in HIV-1 infected living cells, and which sites are involved in the process of transcription activation.

We analyzed protein-DNA interactions at the LTR by *in vivo* dmethylsulphate footprinting in the chronically infected H9 limphocytic T cell line, in which transcription is constitutively high, and in two systems where LTR-directed transcription is very low and can be dramatically induced by specific stimuli: the promonocytic U1 cell line before and after induction by TPA, and a HeLa-derivative cell line, bearing an integrated LTR, before and after infection with HSV-1 and HHV-6.

In all the systems analyzed major footprints appear at the three Sp1 sites and the two repeats of the enhancer independently from the level of transcription, indicating the existence of further levels of regulation probably involving ancillary factors and the basic transcription machinery. Upon transcription activation by TPA or by infection with HSV and HHV-6 the same perturbation is observed at the enhancer site adjacent to the Sp1 boxes, showing that at the level of protein-DNA contacts the same events accompany activation by TPA (which acts through the same pathway activated by antigens and TNFα) and by herpes viruses. Interestingly, the change in the footprint pattern is concomitant to the increase of viral expression, suggesting that a viral factor might be involved in such a change.

In H9 cells supporting a highly productive viral infection, clear interactions are also present at several sites in the NRE, including the USF/MLTF, which we showed to behave as a downregulator of transcription, the NFAT-1 site and two novel, purine-rich sites that share a common sequence motif. The upstream site was mutated and found to behave as a positive regulator of transcription.

In the U1 cells a lower number of sites are occupied respect to the H9 cells showing that transcription regulation represents another level which determines the different outcome of HIV-1 infection in monocytes and lymphocytes, and might be involved in determining the potential role of monocytes as reservoir for HIV-1.

ACKNOWLEDGMENTS

I am grateful to Prof. Arturo Falaschi and my colleagues at I.C.G.E.B. for helpful discussion and encouragement.