ISAS - INTERNATIONAL SCHOOL
FOR ADVANCED STUDIES

Molecular and functional regulation of HIV-1 expression by the Long Terminal Repeat

Thesis submitted for the Degree of

Doctor Philosophiae

Candidate: Fabrizio d'Adda di Fagagna

Supervisor: Dr. Mauro Giacca

Academic Year 1994/1995

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

TRIESTE

INDEX

SYNOPSIS

Like that of all retroviruses, the rate of transcription of the integrated proviral form of the human immunodeficiency virus type 1 (HIV-1) is the major determinant of the levels of viral gene expression and replication. Viral replication can be detected in all clinical stages of the infection, but, noteworthy, the levels of viral expression correlate with disease progression.

Once integrated, the proviral genome can be considered as a cellular gene, its transcription being tightly regulated by cellular factors and by the viral trans-activator Tat. A single viral transcript originates from the transcription start site at the Long Terminal Repeat (LTR), the unique transcriptional promoter of HIV-1, located at the 5' end of the provirus.

In this thesis, the results of a study concerning the molecular and functional interactions of human transcription factor USF with the LTR of HIV-1 and the production of biologically active recombinant Tat are reported.

We employed USF purified from HeLa cells and recombinant USF^{43} produced in our laboratory to show that they both bind to the HIV-1 LTR in a variety of in vitro assays. We provide evidence that binding is sensitive to the methylation state of the target sequence and that it induces a bent in the DNA template. By a variety of assays, we provide evidence that USF is a positive regulator of HIV-1 transcription. In in vitro transcription experiments, a mutation that impairs USF binding to the LTR reduces the transcriptional activity promoted by the LTR and the addition of a plasmid containing several binding sites for USF downregulates transcription by acting as a decoy. When added to an in vitro transcription mixture, increasing amounts of recombinant USF^{43} progressively enhance transcription. In vivo, when the LTR are transiently transfected in cultured cells, the overexpression of USF^{43} upregulates transcription driven by the LTR. Interestingly, USF^{43} is unable to transactivate the HIV-1 promoter in the presence of the Tat protein.

In the second part of this thesis, the issue of the production of a biologically active recombinant Tat protein is assessed. Two fusion proteins between glutathione-

S-transferase (GST) and one-exon Tat or two-exons Tat were obtained. A novel strategy was developed that allows the production of large quantities of soluble and highly active protein. Both one-exon and two-exons Tat are active in binding to their target RNA by gel retardation assay, promote in vitro transcription from the LTR promoter, and activate the in vivo expression of an LTR-reporter construct when they are given to responsive cells. Therefore, these proteins represent potential useful reagents to investigate the molecular mechanisms by which Tat exerts its known pleiotropic effects.

Finally, we also show that both the recombinant Tat and USF^{43} are in vitro modified by purified DNA dependent protein kinase (DNA-PK) and discuss the possible role of these events in the control of the functional regulation of the HIV-1 promoter.

Part of the results presented in this thesis have already been published in the following papers:

A human binding site for transcription factor USF/MLTF mimics the negative regulatory element of Human Immunodeficiency Virus type 1 - Giacca, M., Gutierrez, M. I., Menzo, S., d'Adda di Fagagna, F., Falaschi, A. 1992. *Virology (Vol. 186, p133-147).*

Stimulation of the adenovirus major late promoter in vitro by transcription factor USF is enhanced by the adenovirus DNA binding protein. Zijderveld, D., d'Adda di Fagagna, F., Giacca, M., Timmers, H. T. M., van der Vliet, P. 1994. *Journal of Virology. Dec 1994, vol. 68, No 12, p 8288- 8295.*

Molecular and functional interactions of transcription factor USF with the Long Terminal Repeat of Human Immunodeficiency Virus type 1. d'Adda di Fagagna, F., Marzio, G., Gutierrez, M.I., Kang, L.K., Falaschi, A. and Giacca, M. 1995. *Journal of Virology. May 1995, vol. 69, No 5, p 2765-2775.*

INTRODUCTION

HIV-1 genome organization and transcripts

The human immunodeficiency virus type 1 (HIV-1) is the etiologic agent of the Acquired ImmunoDeficiency Syndrome (AIDS) (Barre-Sinoussi *et al.*, 1983). The virus belongs to the Retroviridae family (lentivirus genus); its genome is a 9700 nucleotides long RNA template. Following cellular infection, this RNA molecule is reverse-transcribed into a double stranded DNA by the viral-encoded reverse-transcriptase and integrates into the host genome.

The proviral DNA is flanked by the two LTR (Long Terminal Repeats) sequences that are generated during the process of reverse transcription, and contains three major structural genes (Gag, Pol and Env), two necessary regulatory genes (Rev and Tat) and other auxiliary genes (Vip, Vpr, Vpu and Nef). For a schematic representation of the genomic organization of HIV-1, see Figure 1. Excellent reviews on the structure and function of these genes have been published during the last years (Subbramanian and Cohen, 1994; Trono, 1995).

A single transcript originates from the 5' LTR and finishes at the polyadenylation site at the 3' LTR. From this transcript, more than 30 different mRNAs are produced by extensive usage of alternative processing mechanisms (Schwartz *et al.*, 1990). Either during the infection of susceptible cell cultures (Kim *et al.*, 1989; Klotman *et al.*, 1991) or in cellular systems in which transcription is inducible (Michael *et al.*, 1991; Pomerantz *et al.*, 1990), a temporal regulation of the synthesis of these transcripts is detectable: initially, short (~ 2 kb), multi-spliced transcripts are prevalent, encoding for regulatory proteins; subsequently, intermediate (~4.5 kb), single-spliced mRNAs (coding for Env Vif, Vpr and Vpu) and full-length (~9 kb), unspliced mRNA (coding for Gag, Pol and representing the viral genome to be eventually assembled in the new virions) accumulate. This dual pattern for transcript processing is probably related to the presence of the viral Rev protein (Cullen and Greene, 1989; Pomerantz *et al.*, 1990), which is synthesized

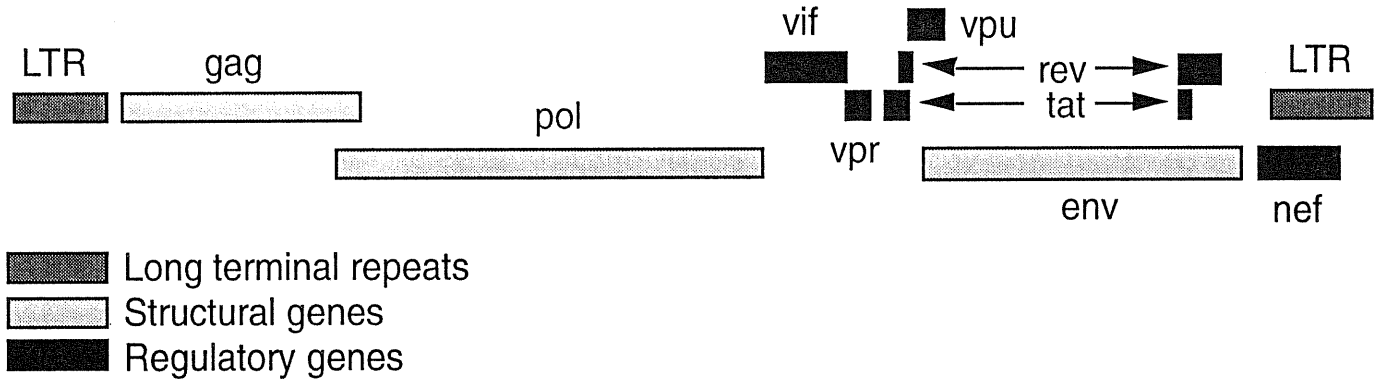


Fig. 1. HIV-1 gene organisation

from the short transcripts. In addition to this RNA splicing regulation, different proteins arise from the extensive use of other differential processing events, such as the utilization of alternative reading frames, ribosomal frameshifts, bicistronic mRNAs, and proteolitic cleavage of precursor proteins.

Two main viral proteins affect HIV expression: Tat and Rev. The Tat protein is the viral transactivator that binds to the 5' end of all nascent RNAs, augmenting the rate of transcription initiation and elongation (see below). The regulatory Rev protein controls the nuclear export of unspliced and partially spliced mRNAs by binding to a Rev Responsive Element (RRE), a highly structured RNA sequence located within the REV encoding sequence, thus favoring the translation of viral structural proteins. This post-transcriptional regulation ensures the switch from the early to the late phase of viral gene expression.

HIV-1 expression and disease development

The regulation of HIV-1 gene expression plays a key role in triggering viral replication, and, consequently, in disease development. Recent quantitative studies show that HIV-1 is expressed throughout all stages of the disease; nevertheless, during the course of the disease, a clear correlation exists between progression and the presence of increasing amounts of infectious virus, viral antigens and virus-specific nucleic acids. This notion derives from studies that measured the number of infected cells in peripheral blood, the amount of proviral DNA and viral genomic RNA in serum. The results obtained by these studies were more recently validated, also in our laboratory, by the introduction of quantitative methods for measuring the concentration of viral nucleic acids, based on the competitive PCR or the branched DNA techniques (Bagnarelli *et al.*, 1992; Comar *et al.*, 1995; Ho *et al.*, 1995; Menzo *et al.*, 1992; Piatak *et al.*, 1993; Scadden *et al.*, 1992; Wei *et al.*, 1995). In particular, some of these studies drew a correlation between the levels of viral expression (as measured from the ratio between viral transcripts and proviral DNA in infected cells) and the clinical course of the disease (Bagnarelli *et al.*, 1992; Gupta *et al.*, 1993;

Saksela *et al.*, 1994). These observations suggest that disease progression is paralleled by an increase in the transcriptional rate of the provirus, and indicate that the study of the molecular mechanisms controlling the rate of transcription initiation could offer useful insights into the mechanisms of disease development.

Molecular basis of transcriptional control of HIV-1

The regulation of HIV-1 transcription initiation is achieved through the recognition of the 5' LTR by human nuclear factors and through their interactions with the RNA polymerase II machinery. In this respect, the regulation of transcription of the integrated provirus does not differ from that of any other cellular gene. Only one viral protein (the product of the tat gene) cooperates with this machinery to increase the rate of transcription initiation and elongation.

On the basis of in vitro binding studies, analysis of deletion mutants viability, and transient transfection experiments, the LTR appears as a mosaic of binding sites for nuclear proteins (Gaynor, 1992). With the use of an in vivo footprinting technique in infected cell lines, our laboratory has shown that most of the LTR regions are indeed engaged in protein-DNA interactions also in vivo, although the pattern of these interactions is different in different cell lines (Demarchi *et al.*, 1992; Demarchi *et al.*, 1993).

From the functional standpoint, the 5' LTR can be divided in three main regions: the basal promoter region, encompassing the transcription start site, and supporting a positive basal effect on transcription; the enhancer region, that mediates transcriptional inducibility of the provirus upon a variety of stimuli which trigger cellular activation and proliferation, and the negative regulatory element (NRE), extending upstream of the enhancer site, whose overall function is to downregulate transcription. Figure 2 is the schematic representation of the HIV-1 LTR and of the human proteins that interact with it.

The basal promoter consists of three sites for factors of the Sp1 family, of a TATA box, and of an initiator element. These three elements are essential for

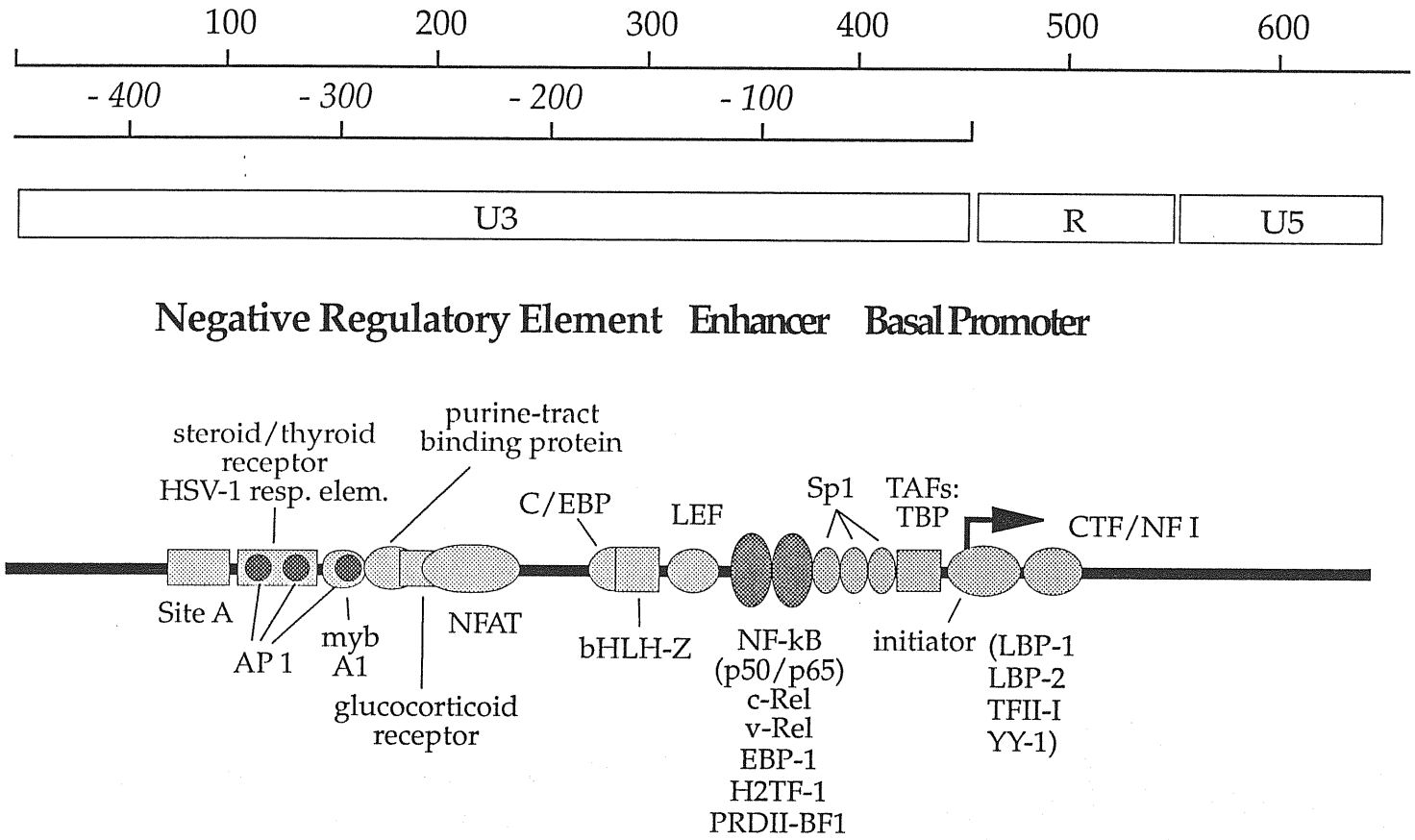


Fig. 2. Schematic representation of the HIV-1 Long Terminal Repeats and the human proteins reported to bind to it. Transcription start site is indicated by an arrow at the U3-R boundary; the lower scale indicates nucleotides upstream of transcription start site.

minimal promoter activity. The TATA element is recognized by the TATA binding protein (TBP) of the TFIID chromatographic fraction. In addition, it was shown that this particular TATA sequence also interacts with a peculiar TATA element modulatory factor (TMF) (Garcia *et al.*, 1989). The initiator element is located in correspondence of the transcription start site. Several proteins capable of binding to the initiator element in vitro have been identified, including LBP-1 (Jones *et al.*, 1988; Kato *et al.*, 1991), LBP-2 (Margolis *et al.*, 1993), TFII-I (Roy *et al.*, 1991), and YY-1 (Margolis *et al.*, 1994; Seto *et al.*, 1993; Seto *et al.*, 1991). The LBP-1 protein binds with higher affinity also to sequences in the LTR immediately downstream of the transcription start site (Jones *et al.*, 1988; Kato *et al.*, 1991). The overall functional relevance of the proteins binding to the initiator element to the process of transcriptional activation still needs to be clearly established.

Cellular transcription factor/nuclear factor I (CTF/NF-I) has also been demonstrated to bind downstream of the transcription start site (Jones *et al.*, 1988), although this observation has not been pursued any further.

The sequence partially overlapping with the transcriptional start site and immediately downstream of it (from nt -5 to +80) is capable of promoting the synthesis of short non processive transcripts (inducer of short transcripts, IST) (Ratnasabapathy *et al.*, 1990; Sheldon *et al.*, 1993). These ~60 nt RNA fragments accumulate in the cell in the absence of the viral Tat protein (Adams *et al.*, 1994; Kessler and Mathews, 1992; Laspia *et al.*, 1989).

The enhancer region of HIV-1 is composed of two direct repeats of the sequence GGGACTTTCC (κB sites) located between -104 and -81 (Nabel and Baltimore, 1987). It is active in both stimulated and unstimulated T-cells (Nabel and Baltimore, 1987; Siekevitz *et al.*, 1987), and in most of the cell lines tested so far, including HeLa cells (Dinter *et al.*, 1987; Garcia *et al.*, 1987; Rosen *et al.*, 1985). This region is able to support basal transcription from the downstream domains and mediates enhancement of viral expression after treatment with stimuli which activate T cells, such as phorbol 12-myristate 13-acetate (PMA), phytohemagglutinin (PHA)

(Kaufman *et al.*, 1987; Nabel and Baltimore, 1987; Tong-Starksen *et al.*, 1987), cytokines (Duh *et al.*, 1989; Lowenthal *et al.*, 1989; Osborn *et al.*, 1989; Siekevitz *et al.*, 1987), or monoclonal antibodies directed against T-cell membrane antigens (Tong-Starksen *et al.*, 1989). This genetic element is the target of the inducible transcription factor NF-κB, a heterodimer composed of two subunits (p50 and p65) (Kawakami *et al.*, 1988), that resides in the cytoplasm when complexed with the inhibitor IκB (Baeuerle and Baltimore, 1988). Activating agents, such as phorbol esters and tumor necrosis factor α, promote phosphorylation and dissociation of IκB, subsequent migration of NF-κB into the nucleus, and consequent transcriptional activation. Moreover, the HIV-1 protease can process the p50 precursor and increase the levels of active nuclear NF-κB complex (Riviere *et al.*, 1991). This model is probably an oversimplification, since, in addition to NF-κB, several other cellular and viral proteins have been reported to interact specifically with the κB recognition motif. These include p50 and p65 homodimers, other Rel family members such as v-Rel, c-Rel (Ballard *et al.*, 1990), and p52 (Schmid *et al.*, 1991) and other apparently unrelated factors such as H2TF1 (Baldwin and Sharp, 1988), PRDII-BF1(Baldwin *et al.*, 1990). The oncoprotein Bcl-3 was shown to be able to transactivate through κB motifs via association with DNA-binding p52 homodimers (Bours *et al.*, 1990). The physiological role of each of these factors, possibly also in concert with the others, still needs to be clarified.

Since the deletion of the region located upstream of the enhancer sequence increases gene expression from the downstream domains (Rosen *et al.*, 1985) and markedly augments viral replication in Jurkat (T-cell lymphoma), U-937 (monocytic) (Lu *et al.*, 1989), HeLa (epithelial) (Giacca *et al.*, 1992) cell lines, this region was named negative regulatory element (NRE). Several binding sites for human nuclear proteins were mapped within the NRE. The region between position -130 and -200 appears to be important for viral replication in peripheral blood lymphocytes and T cell lines (Kim *et al.*, 1993). This region contains the binding sites for a lymphocyte-specific high mobility group protein called LEF, found in immature B and T cells and

in mature T cells, which behaves as an activator of transcription (reviewed in (Jones and Peterlin, 1994)), and for transcription factor USF (see below). Several other sequences appear to be potentially engaged in protein-DNA interactions at the upstream region of the NRE. These include three potential sites for the AP-1 factor between nucleotides -349 and -343, -337 and -371, -291 and -299 (Franza *et al.*, 1988), a palindromic sequence conferring responsiveness to herpes simplex virus infection at position from -353 to -327 (Feng *et al.*, 1993), two sequences interacting with nuclear extracts from activated T cells from -379 to -361 and from -350 to -327 (Orchard *et al.*, 1990), a region binding to the nuclear factor of activated T cells (NFAT) between nucleotides -216 and -254 (Crabtree, 1989; Shaw *et al.*, 1988), a sequence binding the c-Myb transcription factor from nucleotide -314 to -293 (Dasgupta *et al.*, 1990), a sequence related to steroid/thyroid hormone receptor response elements between -357 and -325 (Cooney *et al.*, 1991; Orchard *et al.*, 1992), a glucocorticoid receptor-binding site centered around position -257 (Ghosh, 1992). Recombinant NF-IL6 binds to at least three different sites in the LTR (centered at position -169, -109 and -250 nucleotides upstream the start site). It has been claimed that NF-IL6 is a negative regulator of transcription (Tesmer *et al.*, 1993). T Cell Factor-1α (TCF-1α) occupies the region between nucleotides -123 and -138. This 55 kDa transcription factor is specific for T lymphocytes (Waterman *et al.*, 1991). The sequencing of its cDNA highlighted the similarity between its DNA binding domain and those of the HMG proteins. The function of TCF-1α in LTR-driven expression has not yet been demonstrated.

Finally, by in vivo footprinting experiments, our laboratory identified two additional purine-rich regions engaged in protein-DNA interactions in living cells, from nucleotides -260 to -275 (in H9/HIV_{IIIB} cells) and -205 to -216 (in H9/HIV_{IIIB} and HL3T1 cells) (Demarchi *et al.*, 1992). Both of these sites contain a purine-rich tract very similar to a protein binding site in the fibronectin promoter, in a region possibly involved in downregulation of transcription (Bernath *et al.*, 1990). The upstream site partially overlaps with the glucocorticoid responsive element (Ghosh, 1992), and

shares sequence similarity with a regulatory domain of the interleukin-2 promoter (Fujita *et al.*, 1986). In vitro, it is the target for different cellular proteins (Li *et al.*, 1991; West *et al.*, 1992).

All the protein binding sites of the LTR are recognized by nuclear proteins physiologically controlling the expression of a variety of cellular genes, indicating that the LTR represents a very interesting biological example of evolutionary tinkering.

Transcription factor USF

A sequence of the LTR located at the 3' end of the negative regulatory element of the LTR contains the hexanucleotide CAC(A/G)TG (E box), which is the consensus target sequence of proteins of the B class of the basic-Helix-Loop-Helix-Leucine Zipper (b-HLH-Zip) family (Dang *et al.*, 1992).

Upstream Stimulatory Factor (USF) or Major Late Transcription Factor (MLTF) is a nuclear protein initially characterized for its ability to bind to and activate from the upstream element of the major late promoter of Adenovirus type 2 (Sawadogo, 1988). After extensive purification from HeLa cells nuclear extracts, USF activity turned out to be composed of two equimolarly represented proteins of 43 and 44 kDa both able to bind independently to the consensus sequence CACPuTG. The two proteins are the products of two different genes. Cloning of the 43 kDa form provided evidence that it belongs to the basic (B)-helix-loop-helix (HLH)-leucine zipper (LZ) family of proteins (Pognonec and Roeder, 1991). Other members of this family are c-Myc (Murre *et al.*, 1989), Max (Blackwood and Eisenman, 1991), Mad (Ayer *et al.*, 1993), Mxi1 (Zervos *et al.*, 1993), TFEB (Carr and Sharp, 1990), TFE3 (Beckmann *et al.*, 1990), and USF^{44} (Sirito *et al.*, 1992). This family is characterized by the presence of a basic region directly involved in DNA binding, while both the helix-loop-helix and leucine zipper structures provide dimerization ability (see Figure 3). None of the other members of the family are able to heterodimerize with the 43-kDa form of USF. The only putative oligomerization

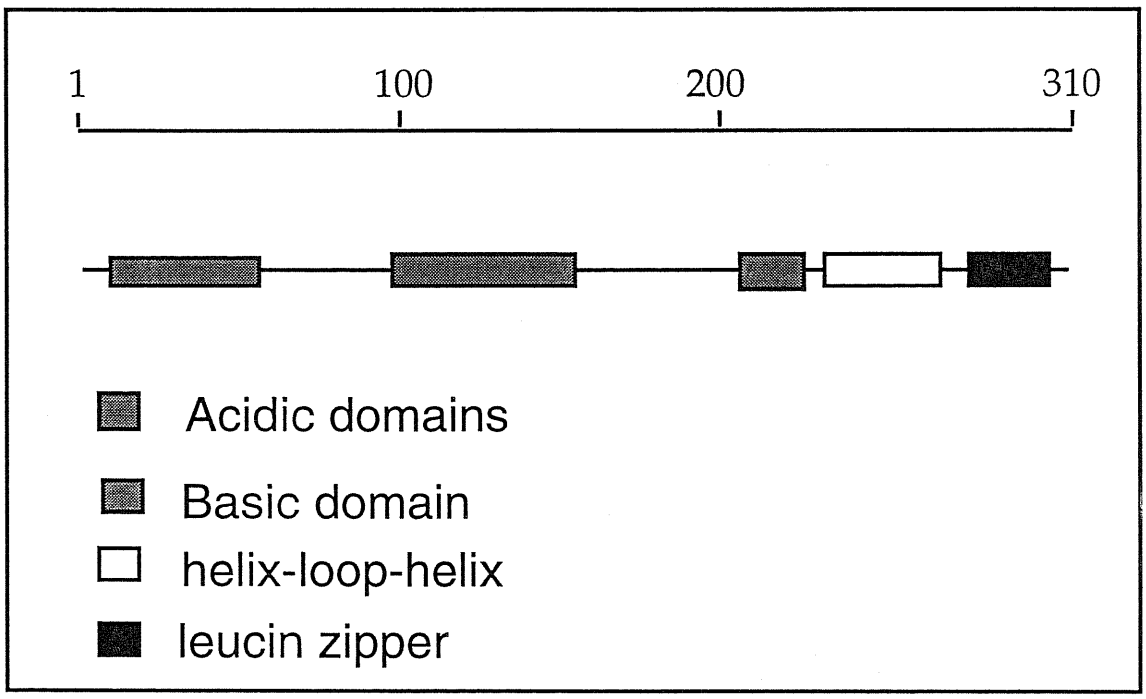


Fig. 3. Stucture of USF[43]

partner is the 44-kDa form of USF. In vitro experiments show that both USF[43] and USF[44] are able to homodimerize and heterodimerize.

Binding sites for b-HLH-Zip proteins and cognate binding activities are very conserved through evolution (Giacca *et al.*, 1989; Giacca *et al.*, 1992) and they control the expression of several unrelated genes. For example, in the human genome, binding sites have been described in the β-globin locus control region (Bresnick and Felsenfeld, 1993), in the promoter regions of the L-type pyruvate kinase gene (Vaulont and Kahn, 1994), heme oxygenase 1 gene (Sato *et al.*, 1990), class I alcohol dehydrogenase gene (Edenberg and Brown, 1992), type 1 plasminogen activator inhibitor gene (Riccio *et al.*, 1992), human growth hormone gene (Peritz *et al.*, 1988), CD2 gene (Outram and Owen, 1994), chicken αA-crystallin gene (Cvekl *et al.*, 1994), N-ras gene (Thorn *et al.*, 1991), cyclin B1 gene (Cogswell *et al.*, 1995), topoisomerase I gene (Heiland *et al.*, 1993), amyloid β-protein precursor gene (Hoffman and Chernak, 1995), in the enhancer of the insulin gene (Read *et al.*, 1993) and in a region that contains an origin of DNA replication (Biamonti *et al.*, 1992a; Csordás Tóth *et al.*, 1993; Giacca *et al.*, 1994). It was also reported that USF is also involved in the control of transcription by polymerase I (Datta *et al.*, 1995). Additionally, it has been recently shown that the E box of the LTR of HIV-1 is also essential for the transcriptional control of the negative strand of the provirus exerted by the LTR at the 3' end (Michael *et al.*, 1994).

Finally, a laboratory also reported that USF can activate transcription acting on initiator elements (Du *et al.*, 1993; Roy *et al.*, 1991), possibly in conjunction with transcription factor TFII-I. However, the ability of the protein to bind and activate transcription from two unrelated sequences, namely the upstream hexanucleotide element and the initiator element, is still controversial (see below).

The analysis of the LTR in HIV-1 infected H9 cells by in vivo footprinting has revealed that the E box is engaged in protein recognition also in infected living cells (Demarchi *et al.*, 1992).

The Tat protein of HIV-1

The Tat protein of HIV-1 is encoded by a two exon gene in the 3' portion of the viral genome. In most of the viral strains, the protein has 86 amino acids, 72 of which are encoded by the first exon. In some viral isolates, the presence of an extra C-terminal tail deriving from the absence of a canonical stop codon have been predicted from the analysis of the primary nucleotide sequence, with the production of a 103 amino acid protein.

Structure of Tat

Different structural and functional domains have been recognized in the protein (Carroll *et al.*, 1991; Kuppuswamy *et al.*, 1989; Tiley *et al.*, 1990). The N-terminal region (amino acids 1-9) forms an amphipatic α-helix, a feature frequently observed in the activation domain of other transcription factors (Rappaport *et al.*, 1989). The region between amino acids 22 and 37 contains seven cysteins, and it usually referred to as the cystein-rich domain. All but one of these cysteins are essential for Tat function (Kuppuswamy *et al.*, 1989; Rice and Carlotti, 1990a; Sadaie *et al.*, 1988). The function of the cysteine residues is unknown, but they may bind metal ions with specific coordinations which may be necessary for proper protein conformation (Rice and Carlotti, 1990b). Alternatively, or additionally, this cysteine-rich domain could mediate protein-protein interactions, with particular reference to the ability of the protein to dimerize and multimerize (see below). Amino acids 38 to 48 form the so-called core domain; amino acids from 49 to 57 are rich in ariginines and constitute the basic domain, containing the nuclear localization signal and the RNA-binding region (Hauber and Cullen, 1988; Kamine *et al.*, 1991; Subramanian *et al.*, 1991). According to Carroll et al., the cysteine-rich, core, and basic domains are evolutionary conserved within the lentiviridae genus (Carroll *et al.*, 1991). The C-terminal domain is constituted by the second exon (amino acids 73-86). The first exon only is usually required for full activity of the protein (Arya *et al.*, 1985; Muesing *et al.*, 1987); however, some particular functions have been more recently ascribed

exclusively to full length Tat (Kim and Panganiban, 1993; Tong-Starksen *et al.*, 1993).

Controversial results have been obtained about the ability of the protein to form dimers. Frankel et al. described the ability of the protein to form metal-linked dimers in a process mediated by the cysteine-rich domain (Frankel *et al.*, 1989). Accordingly, the protein was found to form homomultimers in the yeast two-hybrid system (Bogerd *et al.*, 1993). This function, however, was mediated by the core and not by the cysteine-rich motif. On the contrary, the protein appeared as a monomer in extracts of expressing cells (Rice and Carlotti, 1990b; Rice and Chan, 1991). These apparent discrepancies are partially due to the tendency of the protein to oxidize very easily during manipulation: in the oxidated state, the protein tends to form dimers and multimers (Koken *et al.*, 1994a) - see also the section on Tat purification.

Localization of Tat

The localization of the protein has been reported to be mainly nuclear (Hauber *et al.*, 1987; Koken *et al.*, 1994a), where it exerts most of its function. In addition, the protein appears to be secreted by producing cells, and can be taken up by neighboring cells, be localized into the nucleus, and transactivate HIV-1 LTR-directed gene expression (Barillari *et al.*, 1993; Bohan *et al.*, 1992; Ensoli *et al.*, 1993; Frankel *et al.*, 1989; Helland *et al.*, 1991; Mann and Frankel, 1991). Secreted Tat binds to cells, with $> 10^7$ sites per cell, it is then adsorbed by an endocytosis process and eventually it appears in endosomes/lysosomes (Mann and Frankel, 1991). The basic domain of the protein mediates binding to the cell surface; binding can be inhibited by polyanions, such as heparin and dextran sulfate (Mann and Frankel, 1991). In this respect, the protein seems to act similarly to some growth factors (e.g. basic and acidic FGF), that lack a recognizable signal sequence and yet can be secreted or, in some instances, are found in intracellular/nucleolar locations (ref. in (Fawell *et al.*, 1994)). In addition to being secreted in the medium, the protein can enter neighboring cells by cell-to-cell contact (Helland *et al.*, 1991).

Functions of Tat

The Tat protein of HIV-1 exerts a variety of pleiotropic effects on viral and cellular gene expression, and on cellular growth and metabolism. The most remarkable function of the protein is to act as an extremely potent activator of transcription from the LTR, by increasing expression of LTR-controlled genes by several thousand folds.

The protein binds to the 5' end of nascent HIV RNAs, in a region containing a stem-loop structure (position +1 to +60 downstream of the transcription start site) denominated TAR. The key determinant in HIV-1 TAR RNA for specific binding of Tat is a 3-nucleotide bulge in the stem (Cordingley *et al.*, 1990; Dingwall *et al.*, 1990; Harper and Logsdon, 1991; Roy *et al.*, 1990a). The basic region of the protein, located between residues 49 and 57, recognizes and directly binds to the bulge and to few surrounding base pairs (Calnan *et al.*, 1991; Weeks *et al.*, 1990). The TAR region of HIV-1 is essential for viral replication, both in terms of conservation of the nucleotide sequence of the bulge and of maintenance of base complementarity (regardless of sequence) in the stem region (Klaver and Berkhout, 1994).

The effect of Tat on LTR-controlled transcription is exerted both at the transcriptional initiation and elongation levels. Binding of the protein to the TAR structure increases the rate of transcriptional initiation (Laspia *et al.*, 1989; Muesing *et al.*, 1987; Peterlin *et al.*, 1986; Rice and Mathews, 1988). In this respect, the function of the protein is not conceptually different from that of canonical transcriptional factors, with the exception of RNA instead of DNA binding. The function of the protein strictly depends on the presence of functional upstream Sp1 and TFIID sites on the LTR (Lu *et al.*, 1993). Transactivation still occurs if the κB and LBP-1 sites are deleted (Lu *et al.*, 1993). Specific protein-protein interactions have been described between Tat and Sp1 (Jeang *et al.*, 1993) and TBP (Kashanchi *et al.*, 1994; Veschambre *et al.*, 1995). In addition to this function in the increase of the rate transcriptional initiation, Tat also acts as an anti-terminator protein, by increasing the

rate of elongation of initiated RNAs. In the absence of the protein, transcripts originating form the LTR and stalling at position ~+80 accumulate, while in the presence of Tat these transcripts are elongated (Kao *et al.*, 1987; Laspia *et al.*, 1989). In other terms, the protein is required for the assembly of elongation-competent transcription complexes from the LTR (Kato *et al.*, 1992). Therefore, two types of transcription complexes are formed at the LTR: processive complexes which require the presence of Sp1 and κB sequences (Lu *et al.*, 1993), and non-processive complexes depending on the presence of TATA and TAR (Ratnasabapathy *et al.*, 1990).

In addition to the described activities at the transcriptional level, the Tat protein acts also at the post-transcriptional level, by augmenting the efficiency of translation of HIV mRNAs mainly by relieving the inhibition of translation conferred by the Tar structure (SenGupta *et al.*, 1990). A clear evidence of how this activity is exerted is still lacking. A Tat-dependent, nucleus-specific covalent modification of RNA which allows translation has been described (Braddock *et al.*, 1990; Braddock *et al.*, 1991). Furthermore, it has been reported that the two-exon Tat protein acts post-transcriptionally in a TAR-independent manner, by binding to a sequence contained at the 5' end of the env gene (Kim and Panganiban, 1993).

In addition to these activities on gene expression, a variety of other miscellaneous activities have been ascribed to the Tat protein, most of which are consequences of its ability of being secreted and enter neighboring cells. The protein is able to inhibit lymphocyte proliferation (Benjouad *et al.*, 1993; Gutheil *et al.*, 1994; Viscidi *et al.*, 1989). This function, that is exerted on antigen- but not on mitogen-induced activation (Viscidi *et al.*, 1989) is exerted by a synthetic peptide corresponding to amino acids 1-58 (Chirmule *et al.*, 1995). In accordance to this finding, Tat protein was found able to bind to CD26 and inhibit its function (Gutheil *et al.*, 1994). The CD26 antigen is dipeptidyl aminopeptidase IV (DP IV), a serine protease being a marker of activation of memory T-cells responding to recall antigens. This protein has also been proposed as a co-receptor for HIV (Callebaut *et*

al., 1993), although this finding has been recently questioned (Broder *et al.*, 1994). Interestingly, the levels of CD26 have been found reduced in patients with HIV infection (Vanham *et al.*, 1993).

It is obviously of extreme importance to fully understand the meaning of these potential immunomodulatory activities of Tat in the context of the development of immunodeficiency in the course of HIV infection.

The Tat protein has also been found to be neurotoxic, by exerting a toxic effect on neural cells in vitro and neurotoxicity in mice in vivo (Sabatier *et al.*, 1991). Peptides corresponding to the conserved basic region 49-57 of exogenous recombinant Tat induce morphological changes in neurons and astrocytes in developing rodent brain cell cultures without causing cell death (Kolson *et al.*, 1993). Also in this case, it appears to be extremely important to understand the significance of these findings in the context of the development of neurological abnormalities in AIDS patients.

Controversial reports have been published about the role of the protein on cell growth. Initially, the protein was found to promote cell survival and proliferation (Zauli *et al.*, 1993). However, more recently it was reported that the protein, when given at high concentration in the medium or endogenously produced by permanently expressing cells, induces cell apopotosis (Li *et al.*, 1995). Again, the confirmation of the latter reports would be of extreme importance in the understanding the pathogenesis of cell depletion in AIDS.

Finally, Tat was suggested to play a role in the induction of Kaposi's sarcoma, a neoplasia commonly found in homosexual men with AIDS. In particular, Tat stimulates the growth of spindle cells of vascular origin derived from KS lesions of patients with AIDS (AIDS-KS cells) and that of normal vascular cells which have been exposed to inflammatory cytokines (Ensoli *et al.*, 1990; Ensoli *et al.*, 1993). In addition, Tat-transgenic mice develop dermal lesions resembling KS (Vogel *et al.*, 1988).

Action of Tat on cellular genes

Given the pleiotropic effects of Tat on several cellular functions, it is not surprising that the protein was reported to act as a transcriptional regulator of several cellular genes. The protein is able to activate expression of extracellular matrix proteins in glial cells (Taylor *et al.*, 1992), of TGF-β1 in lymphocytic cells, glial cells, and marrow macrophages (Cupp *et al.*, 1993), of Il-4 receptor in B-lymphoblasts (Puri and Aggarwal, 1992), of TNF-β in T lymphocytic and B-lymphoblastoid cells (Buonaguro *et al.*, 1992) and to inhibit the expression of Il-10 (Masood *et al.*, 1994). Clear evidence is still lacking of how the mechanism of transactivation of these genes is exerted, with particular reference to the understanding of whether these effects are due to a direct interaction of the protein at the promoter region of these genes, or - more likely - to indirect effects on other cellular functions. Furthermore, the protein has also been reported to transactivate the murine CMV major immediate-early promoter (Kim and Risser, 1993). In this case, a transactivation responsive region has been mapped at the DNA level, that acts also post-transcriptionally.

The Tat protein is also able to regulate negatively the expression of several other genes. Among these, there are the p68 kinase gene (Roy *et al.*, 1990b), an interferon-induced gene, the manganese superoxide dismutase (Mn-SOD) gene (Roy *et al.*, 1990b), where the protein was suggested to directly bind to the mRNA, and the MHC class I genes (Howcroft *et al.*, 1993), where the effect was ascribed to two-exon Tat but not to the one-exon protein.

Interactions of Tat with cellular factors: the search for the Tat cofactor

The effects played by the Tat protein by interaction with the HIV-1 TAR region are likely to occur in the context of a complex set of interactions between both Tat itself and TAR with a variety of cellular proteins. Genetic evidence suggests that Tat-mediated transactivation occurs only by interaction with specific human factors. For example, while Tat is very active in human cells, its transactivation ability is highly impaired in rodent cells, where it can be restored by the presence of human

chromosome 12 (Alonso *et al.*, 1992; Hart *et al.*, 1989; Newstein *et al.*, 1990). In other terms, human chromosome 12 is likely to encode for a host function that is essential for Tat activity. In the the search for the Tat cofactor, several authors have identified specific Tat and TAR binding proteins.

Among the Tat-binding proteins, TBP1 was cloned by the screening of a human cDNA library with a Tat probe (Nelbock *et al.*, 1990). An homologous protein (TBP7) was found by screening of a cDNA library with the TPB1 cDNA (Ohana *et al.*, 1993). Another Tat-binding protein that increases Tat-mediated transactivation is MSS1, a suppressor of a cold-sensitive CDC2-like kinase mutant (Shibuya *et al.*, 1992). These factors are different subunits of the ATP-dependent 26 S protease, the degradative component of the ubiquitin pathway. More recently, by the use of the yeast two hybrid system, a novel member of the C3HC4 or ring finger family of zinc finger proteins was identified, that specifically binds to the activation domain of HIV-1 Tat (Fridell *et al.*, 1995).

It is really difficult at this moment to understand what is the physiological role - if any - of the interaction of all these protein with Tat in the context of Tat function.

A similar problem applies to a variety of proteins that were found to interact with the TAR RNA region. Here, genetic evidence suggests that this region, with particular reference to the maintenance of the stem-loop structure is critical for Tat function, independently of actual Tat binding (that requires, as described above, only the region of the bulge). Different cellular proteins were found able to bind in vitro to TAR. TRBP is a 37-kDa RNA-binding protein binding to TAR, that activates LTR-mediated transcription and is synergistic with Tat function (Gatignol *et al.*, 1991; Gatignol *et al.*, 1989); p68 is a loop-binding protein capable of enhancing transcript elongation (Marciniak *et al.*, 1990). The TAR region between nucleotides +17 to +43 binds to a large heteromeric complex containing a 185 kDa RNA-binding protein, belonging to a family of proteins of 110- to 70-kDa (Sheline *et al.*, 1991; Wu *et al.*, 1991). Among the stem-binding proteins, Han et al. described a host cell factor that depends on its phosphorylation state for transactivation activity (Han *et al.*, 1992).

With similarity to the problem of the Tat-binding proteins, it is still unclear which of these interactions are critical for Tat activity.

In this context, it is also interesting to report the identification of a cellular protein (TAK) that specifically associate with Tat and acts as a serine/threonine kinase (Herrmann and Rice, 1995). Mutations in Tat that abolish transactivation activity in vivo abrogate the ability of the mutants to bind to the kinase in vitro (Herrmann and Rice, 1993). In the course of the work described in this thesis, we have found that another kinase, the DNA-dependent protein kinase (DNA-PK), is potentially involved in Tat phosphorylation (see below). Interestingly, it has also been reported that Ku, the DNA binding component of DNA-PK, can recognize the loop region of Tar (Kaczmarski and Khan, 1993).

Construction and purification of recombinant Tat proteins

A technical problem that has hindered a better understanding of Tat function is the discrepancy between the functional activity of the protein and its actual concentration in producing cells. In other terms, the protein is an extremely powerful activator of transcription, although it is barely detectable or usually undetectable by western blotting. Additionally, the protein is difficult to overexpress and purify due to its tendency to aggregate (Harper and Logsdon, 1991; Rhim *et al.*, 1993). This problem is most likely due to the presence of the cysteine-rich domain (that contains 7 cysteines within 16 residues), which renders the protein extremely sensitive to oxidation and consequent aggregation. Additionally, the protein is basic with a high degree of nonspecific affinity for nucleic acids which may contaminate protein preparations and promote aggregation.

Several authors have encountered these problems during their attempts of purification of Tat by different strategies, including the production of a His-tagged protein (Song *et al.*, 1994), direct purification from human cells (Rice and Chan, 1991), in vitro transcription and translation (Rice and Carlotti, 1990b; Rhim *et al.*, 1993), fusion to GST (Koken *et al.*, 1994), overexpression and purification from E.

coli cell extracts or inclusion bodies (Dingwall *et al.*, 1990; Fawell *et al.*, 1994; Frankel *et al.*, 1989).

Here, we describe a strategy for the purification of a recombinant Tat protein which is highly active both in vitro and in vivo, and is free of contaminating nucleic acids. This protein is highly amenable to a variety of studies addressing at the understanding of how Tat exerts its pleiotropic functions.

DNA-dependent protein kinase (DNA-PK)

DNA-PK is a nuclear serine/threonine kinase, found in humans and in a wide range of eukaryotic species. The unique feature that distinguishes this kinase is its need for DNA binding in order to be active (Gottlieb and Jackson, 1993). DNA recognition is achieved through a 70/80 kDa heterodimer complex known as autoimmunoantigen Ku, while the catalytic site lies in a larger subunit of 460 kDa of relative molecular weight named DNA-PKcs. The most unusual characteristic of Ku is its DNA binding ability: Ku recognizes free ends of linear DNA molecules.

DNA-PK has been implicated in several relevant cellular processes. It was shown to phosphorylate many transcription factors such as Sp1, SRF, Oct1, CTF/NF-I (Anderson, 1993; Gottlieb and Jackson, 1993; Liu *et al.*, 1993) and the product of the proto-oncogenes c-Jun (Bannister *et al.*, 1993), Fos (Anderson, 1993) and c-Myc (Iijima *et al.*, 1992), and the anti-oncogene p53 (Lees-Miller *et al.*, 1992). It has also been reported that DNA-PK phosphorylates the carboxyl-terminal domain (CTD) of RNA polymerase II (Dahmus, 1995); phosphorylation of CTD has been implicated in the switch of the RNA pol II from a non-processive to a more processive state (Dahmus, 1995). DNA-PK was also demonstrated to be an inhibitor of transcription by RNA polymerase I (Kuhn *et al.*, 1995) and possibly of RNA pol II and III (White, Kuhn, Smith and Jackson, unpublished data).

On the basis of DNA-PK ability to recognize discontinuities in the DNA linear structure, a proposed model for DNA-PK function is to inhibit transcription of those genes whose DNA is damaged (Gottlieb and Jackson, 1994). The same ability led

also to postulate a role for this protein in DNA recombination and DNA double-strand breaks repair. In fact, it was demonstrated that cell lines hypersensitive to radiation and impaired in V(D)J recombination, are defective in DNA-PK components (Finnie *et al.*, 1995; Taccioli *et al.*, 1993). Noteworthy, also the SCID mouse, the mouse bearing the severe combined immunodeficiency, lacks DNA-PK and its cells can be cured by the introduction of the gene of the catalytic subunit of this protein (Blunt *et al.*, 1995). The above reported ability of Ku to recognize Tar (Kaczmarski and Khan, 1993) and the availability of purified DNA-PK in our laboratory, prompted us to study the potential role of this kinase in the regulation of HIV-1 transcription.

MATERIALS AND METHODS

Purification of USF from HeLa cells

The procedure for the purification of USF from HeLa cells has already been reported (Csordás Tóth *et al.*, 1993). Briefly, a nuclear extract was precipitated by 35% ammonium sulfate; the precipitated protein fraction was subsequently purified through BioRex 70, HiLoad S-Sepharose and Mono Q columns. The active fractions from the Mono Q column were loaded on a specific DNA-Sepharose affinity column containing ligated concatenamers of the E box site. The purified binding activity consists of a protein doublet with a relative molecular weight of 43 and 44 kDa. The binding activity was monitored throughout the purification by gel retardation and south-western assays.

Cloning, expression and purification of USF^{43}

The plasmid pGST-USF^{43} was constructed by cloning the coding region of the 43-kDa form of USF (USF^{43}) in the commercial vector pGEX2T (Pharmacia, Uppsala, Sweden). This plasmid expresses the glutathione-S-transferase (GST) gene in E. coli, under the control of a promoter inducible by isopropyl-galacto-pyranoside (IPTG). The coding sequence of USF was rescued by polymerase chain reaction (PCR) amplification from the plasmid pBSA1MLTF (a kind gift of Dr. Giuseppe Biamonti), which contains the coding sequence of USF^{43} under the control of the human A1 gene promoter (unpublished). The primers for amplification (RUSF1, 5'-CGAGGATCCAAGGGGCAGCAGAAAACA-3' and RUSF2, 5'-GCTGAATTCCTTAGTTGCTGTCATTCTTGATGAG-3') were designed in order to generate a DNA fragment containing the restriction sites for *Eco*R I and *Bam*H I at the two extremities allowing oriented cloning in the vector. Amplification was carried out in 50 µl of a solution containing Tris 10 mM (pH 8.0), KCl 50 mM, $MgCl_2$ 1.5 mM, gelatin 0.01%, each dNTP 200 µM, the two primers 1 µM, 1 ng of template plasmid DNA and 2.5 units of Taq DNA polymerase (Perkin Elmer Cetus, Norwalk, CT), with

35 repetitions of the following cycle: 45 sec at 94°C, 45 sec at 52°C, and 45 sec at 72°C, in a Perkin-Elmer 480 Thermal Cycler. The sequence of the plasmid insert was verified by DNA sequencing.

A colony of the SF8 strain of E. coli transformed by pGST-USF^{43} was grown overnight in 100 ml of Terrific Broth containing 50 µg/ml ampicillin at 30°C. The culture was diluted by the addition of 900 ml of fresh medium and was allowed to grow until the optical density of 0.6-0.8 at 600 nm was reached; protein expression was then induced by the addition of IPTG (Sigma, St. Louis, MO, 1 mM final concentration). After an additional 3 to 5 hours of incubation, the culture was centrifuged at 5000 x g at 4°C, and the bacterial pellet was resuspended in 10 ml of cold phosphate buffered saline (PBS) containing 4 mM dithiothreitol (DTT). Cells were sonicated in ice by three pulses of 20 sec each. After centrifugation of the lysate, the supernatant was mixed with 1 ml of a 50% (v/v) slurry of glutathione cross-linked agarose beads (Sigma). The fusion protein was allowed to bind to the beads at 4°C on a rotating wheel for 1 hour. The suspension was then loaded on an empty plastic column (BioRad) letting the unbound proteins pass through, and the beads were subsequently washed with 50 ml of PBS containing 4 mM DTT. Finally, USF^{43} was eluted in 1 ml containing 100 mM Tris, 4 mM DTT and 20 mM free glutathione (Sigma). With this procedure, 2-3 mg of protein per liter of medium were usually obtained. The purity and integrity of the protein was checked by SDS-PAGE and Coomassie staining.

Cloning, expression and purification of recombinant Tat

The plasmids pGST-Tat1E and pGST-Tat2E were constructed by cloning either the coding region of the first exon of Tat (pGST-Tat1E) or both exons of Tat (pGST-Tat2E) in the commercial vector pGEX2T (Pharmacia, Uppsala, Sweden) between the *Eco*R I and *Bam*H I sites . The coding sequence of Tat was rescued by polymerase chain reaction (PCR) amplification from the plasmid pC132Tat (a kind gift of Dr. Battaglia, Istituto di Sanita', Rome) using oligonucleotide primers with modified 5' ends carrying the appropriate restriction enzyme sites. The sequence of the plasmid inserts were verified by DNA sequencing.

A colony of the SF8 strain of E. coli transformed by pGST-Tat1E or pGST-Tat2E was grown overnight in 100 ml of Terrific Broth containing 50 µg/ml ampicillin at 30°C. The culture was diluted by the addition of 900 ml of fresh medium and was allowed to grow until the optical density of 0.6-0.8 at 600 nm was reached; protein expression was then induced by the addition of IPTG (Sigma, St. Louis, MO, 1 mM final concentration). After an additional 3 to 5 hours of incubation, the culture was centrifuged at 5000 x g at 4°C, and the bacterial pellet was resuspended in 10 ml of buffer A (50 mM Tris pH 8.0, 5% glycerol, 2 mM DTT). Cells were sonicated in ice by three pulses of 20 sec each. After centrifugation of the lysate, the supernatant was mixed with 1 ml of a 50% (v/v) slurry of glutathione cross-linked agarose beads (Sigma). The fusion protein was allowed to bind to the beads at 4°C on a rotating wheel for 1 hour. The suspension was then loaded on an empty plastic column (BioRad) letting the unbound proteins pass through, and the beads were subsequently washed with 90 ml of buffer A. An additional high-salt wash (10 ml of buffer A 0.8 M NaCl) followed by a 50 ml wash of buffer A was introduced to free the fusion protein by contaminant bacterial nucleic acids. Finally, the fusion protein was eluted in 1 ml 100 mM Tris containing 2 mM TD and 20 mM free glutathione (Sigma). By this procedure, 1 mg of protein per liter of medium were usually obtained. The purity and integrity of the protein was checked by SDS-PAGE and staining.

Cleavage of GST-Tat fusion proteins was achieved by incubating 2 µg of thrombin (Sigma) with 100 µg of recombinant protein in a buffer containing 2.5 mM $CaCl_2$, 0.5 M Tris pH 7.5, and 0.15 M NaCl at 30°C for one hour.

DNA-and RNA-free production of recombinant USF and Tat.

When the production of recombinant factors absolutely free from any nucleic acid contamination was needed, bacterial cells were resuspended in the proper buffer containing 2.5 mM $CaCl_2$, 5 mM $MgCl_2$ and Dnase I and RNase A were added to final concentrations of 100 U/ml and 100 µg/ml respectively. The rest of the procedure was the same as reported above.

Western blots

Fifteen nanograms of GST-USF^{43} or 100 ng GST-Tat proteins were resolved on a 12% SDS-polyacrylamide gel and transferred to a nitrocellulose filter by electroblotting. The filter was then incubated in 10% milk-TBS buffer (10% w/v non-fat dried milk in 125 mM NaCl, 10 mM Tris, pH 7.4) for 1 hr at 37°C. Incubation with antibodies against full-length rUSF^{43} (kindly donated by Dr. R.G. Roeder) or monoclonal antibodies against Tat (MRC ADP352/NT3) was performed in 5% milk-TBS for 1 hr at room temperature. The filter was then washed in TBS-0.1% Tween 20 and incubated in 5% milk-TBS with alkaline phosphatase-conjugated swine anti-rabbit immunoglobulins for 1 hr at room temperature. After several washes, bound antibodies were revealed using BCIP/NBT color development solution (BioRad, Richmond CA, USA).

Purified DNA-PK, resolved on a 10% SDS-polyacrylamide gel and transferred to a nitrocellulose filter, was recognized by an immune serum raised against Ku and monoclonal antibodies (pk18-2) against DNA-PKcs (a kind gift of T. Shenk).

Gel retardation assays

All the oligonucleotides used in this work were synthesized by the ICGEB Oligonucleotide Synthesis Service on an Applied Biosystem 380B synthesizer using phosphoramidite chemistry. The sequence of the double-stranded oligonucleotides B48BS, AdMLP, HIV (Giacca *et al.*, 1992), and GAL2, Met I, and Met III (Giacca *et al.*, 1989) have already been described. Oligo Sp1 contains the sequence of the HIV-1 LTR from nucleotides -69 to -46 upstream of the transcription start site, containing two binding sites for transcription factor Sp1. Oligo MUT is a derivative of oligo HIV, where the core E box sequence (CACGTG) was mutated to CATATG.

The synthesized oligonucleotides were resolved by polyacrylamide gel electrophoresis, eluted from the gels, purified and annealed with the complementary strand and end-labeled with [γ^{32}P]-ATP (Amersham, UK; 3000 Ci/mmol; 10 mCi/ml) and T4 polynucleotide kinase.

Gel retardation assays with USF were carried out by the incubation of end-labeled DNA probes (10^4 cpm) with 1-5 ng of purified USF^{43} or 2 µl of purified HeLa USF (Mono Q fraction (Csordás Tóth *et al.*, 1993)) and 0.25 µg of poly[d(A-T)]:poly[d(A-T)] or 1.5 µg of poly[d(I-C)]:poly[d(I-C)], respectively, in binding buffer (20 mM Hepes pH 7.3, 50 mM NaCl, 4 mM DTT, 0.2 mM EDTA, 5% glycerol), in 10 µl final volume. The addition of 1 µl of heat-treated (95°C for 5 minutes) normal human serum to the binding reaction was used to stabilize the protein-DNA complex. After 20 minutes incubation at 30°C, samples were resolved by 5% polyacrylamide gel electrophoresis in 0.5 x TBE; the gel was then dried and exposed. Under these conditions, 1 ng of USF^{43} is able to generate a retarded complex which can be detected after an overnight exposure, with an estimated molar DNA/protein ratio of 1/5.6.

Competition experiments were carried out by mixing a 6 to 150 fold excess of cold oligonucleotides to the probe before the addition of the protein. Supershifting of

the retarded complex was obtained by the addition to the reaction mixture of 1 µl of a 1:10 dilution of immune serum, before the addition of the protein.

Gel retardation assays with Tat were carried out by the incubation of 1 ng of TAR probe (about 6 x 10^6 cpm/µg) with different amounts of thrombin-cleaved GST-Tat1E or GST-Tat2E in a buffer containing 25 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM $MgCl_2$, 0.5 mM DTT, 0.1 % Triton X-100, 10 U RNasin (Promega) and 10 µg of BSA and 0.5 µg of salmon sperm DNA. Binding was allowed at 25°C for 20 minutes and then protein-RNA complex was resolved on a 5% polyacrylamide gel/ 0.5x TBE buffer.

Circular permutation assay

A set of six probes (B1 to B6, Figure 9 panel A) for the circular permutation assay with USF was generated by PCR amplification using the pLTRCAT plasmid (Giacca *et al.*, 1992a) as a template. The localization on the LTR of the oligonucleotides utilized for the amplifications are indicated in Figure 9 panel A. Amplifications were carried out in 50 µl of a solution containing Tris 10 mM pH 8.0, KCl 50 mM, $MgCl_2$ 1.5 mM, gelatin 0.01%, each dNTP 200 µM, both primers 0.1 µM, one of which previously end-labeled with [γ^{32}P]-ATP and T4 polynucleotide kinase, 1 ng of template plasmid DNA and 2.5 units of Taq DNA polymerase, with 35 repetitions of the following cycle: 45 sec at 94°C, 45 sec at 60°C, and 45 sec at 72°C. PCR products were resolved on a 5% polyacrylamide native gel and eluted for a few hours at 65°C in Tris 10 mM (pH 8.0) and EDTA 0.1 mM. Curve fitting was performed using the Curve Fit program on a Macintosh computer by the application of a cosine function.

DNase I footprinting

The probes for DNase I footprinting were generated by PCR amplification of the HIV-1 LTR from the plasmid pLTRCAT. For the experiments with purified USF, the primers for amplifications were FOOTP (5'-GCAAGCTTGAAGAGGCCAAT-3')

and USF1 (5'-AGCAAGCTCGATGTCAGCAGTTCTT-3'); for the experiments with recombinant USF^{43}, the primers used were from position -256 to -220 and from -46 to -70 respectively, relative to the LTR sequence numbering of transcription start site. One of the two primers was end-labeled with [γ^{32}P]-ATP and T4 polynucleotide kinase before PCR amplification in order to generate an asymmetrically-labeled DNA fragment.

Plasmid pLTRΔUSF was obtained by PCR amplification from plasmid pLTRCAT with primers pLTR-GlessI (5'-GCGAATTCTAGGGCGAATTGGGTACC-3') and ΔUSF (5'-GCTCTCGGGGCGAATTCATGAAATGCTAGGCGGC-3'), the latter bearing the CACGTG sequence mutated into an *Eco*R I (GAATTC) restriction site. The amplification product was purified, cut with *Ava* I and *Kpn* I and ligated between the *Ava* I and *Kpn* I sites of pLTRCAT, therefore substituting for the wild type sequence.

In the experiments with USF purified from HeLa cells, about 10^5 cpm of the probe were incubated with 20 µl of purified protein (mono Q fraction), with 2 µg of poly[d(I-C)]:poly[d(I-C)], in the same buffer as that used for the gel retardation assays. After 20 minutes incubation at room temperature, the sample was mixed with an equal volume of a solution containing 5 mM $CaCl_2$ and 10 mM $MgCl_2$. DNase I (Boehringer Mannheim, Germany) was added to a final concentration of 6 ng/µl. After 1 minute, DNase I activity was stopped by the addition of SDS and EDTA (1% and 25 mM final concentrations respectively). Proteins were removed by phenol extraction and the DNA fragments were precipitated, dissolved, denatured and loaded on a 8% polyacrylamide sequencing gel. As a control, the probe was incubated in the same conditions as above without the addition of the protein. In this case, DNase I was used at a concentration of 2 ng/µl.

When testing the recombinant protein, 10 ng of USF^{43} or GST (in the control reaction) were mixed with $5x10^4$ cpm of the probe in the same conditions as above. DNase I was allowed to digest for 30 sec at a final concentration of 0.5 ng/µl.

A G+A ladder of the probe, obtained according to the Maxam and Gilbert chemical cleavage method (Maxam and Gilbert, 1980), was loaded alongside to align the DNase I digestion products.

Methylation protection assay

Methylation protection assays with dimetyl sulfate (DMS) were carried out essentially as described (Papavassiliou, 1993). Briefly, 5-fold scaled-up gel retardation assays were assembled; after electrophoresis, the gel was immersed in a 0.2% (v/v) DMS solution for 4 minutes; methylation reaction was stopped by soaking the gel in a 0.5 M β-mercaptoethanol solution. After autoradiography, the free and retarded bands were excised separately, DNA was eluted and cleaved by piperidine treatment (Maxam and Gilbert, 1980). The products were resolved on a sequencing gel alongside a G+A ladder.

In vitro transcription assays with G-less constructs

The LTR-containing G-less plasmids for in vitro transcription assays are derivatives of the pUGL400 plasmid, which contains a *Eco*R I-*Xho* I fragment from plasmid Syn-O-TG (Schorpp *et al.*, 1988), carrying a G-free cassette of ~380 bp (Sawadogo and Roeder, 1985a), cloned between the *Eco*R I-*Sal* I sites of pUC19. The HIV-1 promoter element to be cloned upstream of the G-less cassette was obtained by PCR amplification from plasmid pLTRCAT (Giacca *et al.*, 1992), with primers pLTR-GlessI, containing a *Eco*R I restriction site at the 5' end continuing with the HIV-1 sequence upstream of the LTR, and primer pLTR-GlessII (5'-CGGAGCTCAGGCAAAAAGCAGCTGCTTA-3', containing a *Sac* I site at the 5' end in the correspondence of the LTR transcription start site); the amplification product obtained was cloned between the *Eco*R I and *Sac* I sites of pUGL400 to obtain plasmid pGLA. Plasmid pGLE was obtained following the same strategy, with the exception that the plasmid pLTRΔUSF was the template DNA for PCR amplification.

As a consequence, the nucleotides at position -162 to -167 are GAATTC (containing a *Eco*R I site) instead of CACGTG.

Plasmid pFN2, containing the fibronectin promoter upstream of a ~200 nt G-less cassette, was used as an internal control for the in vitro transcription experiments. It is a derivative of plasmid pUGL200, which was obtained by cloning a PCR amplification product between the *Eco*R I and *Hind* III sites of pUC19. The amplification product derives from plasmid pUGL400 and was obtained by using the M13-universal primer and primer GL202 (5'-GGAAGCTTGGATCCCGGGATAAGATTG-3'). The latter contains at the 3' end a region complementary to the sequence from 187 to 108 of the ~380 nt G-less cassette (Sawadogo and Roeder, 1985a); PCR amplification was followed by digestion with *Eco*R I and *Hind* III. Plasmid pFN2 was obtained by cloning of a ~710 bp fragment obtained by partial digestion with *Aat* II and *Sac* I of plasmid p-220 (a kind gift of Dr. Alberto Kornblihtt), into the corresponding sites of pUGL200. The cloned fragment contains the human fibronectin promoter .

HeLa nuclear extracts for in vitro transcription assays were prepared according to the protocol of Hattori et al. (Hattori *et al.*, 1990) and modified as follows: cells were washed twice in PBS, once in RSB (10 mM Hepes, 10 mM KCl, 3 mM $MgCl_2$, 0.1 mM EDTA, 0.74 mM spermidine, 1 mM DTT, 0.5 mM PMSF, 2 µg/ml aprotinin (Sigma) and leupeptin (Sigma), 5 mM bestatin (Sigma)) and resuspended in RSB; following a 20 minutes incubation on ice, cells were lysed using a Dounce homogenizer.

Transcription reactions were carried out in 25-30 µl final volume containing 250 ng (in the experiments monitoring the effects of recombinant USF) or 750 ng (in the competition experiments with the E box decoys) of DNA template, 75 µg of nuclear extract, 500 µM ATP and CTP, 25 µM UTP, 10 µCi of [α-^{32}P]-UTP (Amersham, UK; 3000 Ci/mmol; 10 mCi/ml), 0.1 mM *o*-methylguanosin, 10 mM *N*-2-hydroxethylpiperazine-*N'*-2-ethanesulfonic acid (Hepes pH 7.4), 7.5 mM $MgCl_2$, 5 U RNase T_1 (Boehringer), 30 U RNasin (Promega, Madison, WI), 4 mM DTT. Nuclear

extract was equilibrated for 15 minutes at 30°C with the addition of the appropriate amounts of protein or decoy template. DNA was then added, preincubated for 10 min at 30°C and the reaction was initialized by the addition of the above described reaction mix. Transcription was allowed to run at 30°C for 45 min and stopped by adding 175 µl of a solution containing 300 mM Tris pH 7.4, 300 mM sodium acetate, 0.5% SDS, 2 mM EDTA, 3 µg/ml tRNA. Proteins were then extracted with 200 µl of PCI (Phenol/Chloroform/Isoamilic alcohol 25/24/1), RNA was precipitated with ethanol, dried and run at 12 W on a denaturing 8% polyacrylamide gel in 1xTBE and 0.1% SDS. Finally, the gel was dried and exposed for autoradiography.

Plasmid pUF128 (a gift of Dr. Fabio Cobianchi) is a pUC19 derivative containing 128 copies arranged in tandem of a 65 bp *Alu* I-*Fnu* D2 sequence from the insert of plasmid pL15 (Giacca *et al.*, 1992), containing the E box present in the human lamin B2 origin of DNA replication (Biamonti *et al.*, 1992b; Giacca *et al.*, 1994).

In vitro transcription assays by primer extension analysis

When assaying Tat mediated transactivation, 500 ng of pLTR-CAT or derivatives and 1.5 µg of pRSVΔCAT (an internal control plasmid carrying a deletion of 99 bases in the CAT gene generated by a recombinant PCR technique to be described) were assayed in a 20 µl final volume containing 10 µl of nuclear extract (about 10 µg/µl), 250 µM each rNTP, 7 mM $MgCl_2$, 300 ng poly [I-C], 50 ng poly [d(I-C)], 4 mM DTT and 28 U Rnasin. After one hour incubation at 30°C , reactions were stopped as described above, phenol extracted and, prior to ethanol precipitation, 50 fmoles of ^{32}P radiolabelled CAT oligonucleotide were added. This primer sequence matches at +400 downstream of the transcription start site of the CAT gene in pLTRCAT and at position +301 of pRSVΔCAT. The dried pellet was resuspended in TE and annealing between RNA and primer was allowed for 45 minutes at 56°C. Subsequently, a final volume of 20 µl was reached by adding dNTPs (1 mM final), 4 µl M-MLV-RT Buffer 5X (Promega), 28 U Rnasin and 1 µl of M-MLV Rev

transcriptase (Promega) and the mixture was incubated for one hour at 42°C. Elongated cDNAs were then precipitated, resuspended in formamide buffer, resolved on a 5% denaturing gel, and visualized by autoradiography.

Purification of DNA-PK

DNA-PK purification will be described elsewhere. Briefly, a HeLa nuclear extract obtained by the Dignam method (Csordás Tóth *et al.*, 1993) was precipitated with 35% w/v ammonium sulfate. The pellet was resuspended, loaded on a DEAE-Sephacel chromatography column and eluted with increasing salt concentrations. Active fractions were pooled and loaded on a phosphocellulose column. Subsequently, active fractions were loaded on a ds-DNA column. The purification was followed using a specific peptide substrate derived from p53. Coomasie and silver stained gels demonstrate that purified DNA-PK was ~80% pure.

Transfection assays

The following plasmids, purified twice on cesium chloride gradients, were employed in transfection assays:

pLTRCAT and pLTRCATΔUSF are pBlueScript KS (Stratagene, La Jolla CA, USA)-derivatives containing, upstream of the CAT gene, the whole HIV-1 LTR and an LTR in which the USF site was mutated into a *EcoR*I site respectively (Giacca *et al.*, 1992).

pUSFminCAT is a derivative of pG5BCAT (a kind gift of dr. S.J. Triezenberg, Michigan State University, USA) in which the CAT gene is under the control of a synthetic promoter constituted of a TATA box upstream of which four binding sites for USF (derived from the LTR of HIV-1) were cloned.

pCXUSF bears the coding region (930 nt) of USF[43] cloned downstream the cytomegalovirus early promoter. PCX vector is a gift of Dr. R. Eisenman (Seattle, USA).

pSV2-Tat contains the first exon of the HIV-1 tat gene under the control of the SV40 early promoter. It is a gift of Dr. A. Meyerhaus (Institut Pasteur, Paris).

Hela or HeLa-TAT cells (a cell line constitutively expressing the HIV-1 transactivator Tat, kindly provided by Dr. V. Ciminale, University of Padua) were transfected by a modification of the DEAE-dextran method (Sambrook *et al.*, 1989). $1.5x10^6$ cells were plated on a 100 mm Petri dish. Twentyfour hours later, a batch of Transfection Solution (TS) was freshly-prepared, containing the CAT reporter plasmid and 1 µg of an independent indicator plasmid, pGL2-control (Promega), expressing firefly luciferase under the control of the SV40 early promoter. The volume was taken up to 900 ml in sterile TBS (25 mM Tris-HCl, pH 7.4, 137 mM NaCl, 5 mM KCl, 0.7 mM $CaCl_2$, 0.5 mM $MgCl_2$, 0.6 mM Na_2HPO_4). After washing twice the plates with TBS, 300 µl sterile DEAE-dextran (1 mg/ml in TBS, Sigma) were added to TS and the resulting mix was added to each plate. Following a 30 minutes incubation at 37°C, TS was removed, plates were washed and the culture medium was supplied. 24-72 hours after transfection, cells were scraped off of the plates, pelleted and resuspended in 400 µl 250 mM Tris-HCl, pH 7.8. Two 200 µl aliquots were separately assayed for luciferase and CAT activity.

Luciferase assay

A modified version of Promega Luciferase Assay System protocol was followed. Cells were pelleted and resuspended in 50 µl lysis buffer (Promega). After 10 minutes at room temperature, the cell-lysate was centrifuged and 5 µl of cell-lysate supernatant were mixed with 60 µl luciferase assay buffer (Promega) containing the substrate luciferyl-CoA. Photon emission was measured using a LS 1801 liquid scintillation counter (Beckman Instruments) equipped with single photon-monitor software, at 12 second intervals from the second minute on; the mean value over a 3 minute interval was considered. Protein concentration of cell lysate was determined using Bradford assay (Bradford, 1976). The ratio between photon

emission and protein concentration was the estimate of transfection efficiency expressed as counts per Optical Density (O.D.) units.

CAT assay

According to a modified version of the Gorman protocol (Gorman *et al.*, 1982), protein extract was obtained by three cycles of freeze-and-thaw (3 minutes in dry ice-ethanol followed by 3 minutes at 37°C) and a further incubation at 60°C for 5 minutes to inhibit endogenous deacetylases. After pelleting debris, protein concentration of the supernatant was determined as above. The amount of cell extract to be assayed for CAT activity was a function of protein concentration (expressed as O.D.) over transfection efficiency. Samples were then incubated for times ranging from 30 minutes to 1 hour at 37°C with 2.5 µl ^{14}C-1-deoxychloramphenicol (50 mCi/mmol Amersham), 70 µl 1M Tris-HCl, pH 7.8, 20 µl 4 mM Acetyl-CoA in a final volume of 140 µl. The reaction was stopped by extracting chloramphenicol with 1 ml of ethyl acetate and, following lyophilization, samples were resuspended, spotted on TLC silica gel plates and developed by ascending chromatography in a 95% chloroform-5% methanol mixture. Autoradiography resulted in qualitative information, whereas quantitation was achieved by cutting the silica gel in pieces and counting them in the scintillation counter using a liquid scintillation cocktail.

Cat assays with recombinant Tat protein were performed using HL3T1 cells, a HeLa derivative cell line harboring the LTR-CAT construct integrated in its genome. Tat protein was added to the culture medium in the presence of 100 µM chloroquine. After 24 hours, cells were washed twice with PBS and fresh medium was added to the culture. When Tat was lipofected, the protein was diluted in 100 µl of Optimem medium and mixed with 50 µl of Lipofectin (Gibco BRL) diluted 1:1 with Optimem medium. After 15 minutes at room temperature, the mixture was added to the cells, which has been previously washed twice the with same medium. Five hours later,

cells were rinsed with PBS and grown in Dulbecco medium. After 48 hours, cells were scraped and treated as above for the CAT assay.

DNA-PK phosphorylation assays

DNA-PK phosphorylation assays were carried out by incubating the protein substrate with 1 µl of [γ^{32}P]-ATP (Amersham, UK; 3000 Ci/mmol; 10 mCi/ml), 0.5-1.5 µl of purified DNA-PK in the presence or in the absence of DNA, in the same buffer as described in (Gottlieb and Jackson, 1993). Reactions were run at 37°C for 10 minutes, stopped by adding SDS-denaturing loading buffer, heated and loaded on a SDS-polyacrylamide gel.

RESULTS

PART I - Interactions of transcription factor USF with the LTR

Purified USF and recombinant USF^{43} bind to the Long Terminal Repeats of HIV-1

The presence of a human nuclear factor binding to the HIV-1 LTR region centered around nucleotide -164 upstream of the transcription start site has been originally identified by DNase I footprinting using HeLa cell nuclear extracts (Garcia *et al.*, 1987). Subsequently, we have shown that the protein interacting with this sequence also binds to the upstream element of the adenovirus MLP and to several other sequences containing the hexanucleotide consensus sequence CAC(A/G)TG (Giacca *et al.*, 1992), which is the E box target sequence for members of the B class of the bHLH family of proteins (Dang *et al.*, 1992).

By means of a combination of ion-exchange and sequence-specific affinity chromatography techniques (Csordás Tóth *et al.*, 1993), we have purified to homogeneity a protein complex binding to the E box. This complex is composed of two polypeptides of 43- and 44-kDa; its size, heat stability, and target DNA sequence suggest that it corresponds to the transcription factor USF, a member of the b-HLH-Zip family (Gregor *et al.*, 1990; Sirito *et al.*, 1992); furthermore, the 43-kDa polypeptide is recognized by antibodies raised against 43 kDa-USF (Csordás Tóth *et al.*, 1993).

In order to study in further detail the interactions of this transcription factor with the HIV-1 LTR, we expressed and purified USF^{43} from bacteria as a recombinant protein fused to glutathione-S-transferase (GST). On the basis of the available cDNA sequence (Gregor *et al.*, 1990), the coding region of USF^{43} was cloned in plasmid pGEX2T, at the 3'-end of the GST gene maintaining the same open reading frame. The resulting plasmid, pGST-USF^{43}, expresses the USF^{43} protein as an extension of the C-terminus of GST under the control of a promoter

inducible by IPTG. In Figure 4 panel A the results of a single-step purification of USF^{43} from bacteria are shown. The identity of the purified protein was further confirmed by western blot analysis using antibodies raised against rUSF^{43} (Pognonec and Roeder, 1991) (Figure 4 panel B).

The interactions of USF purified from HeLa cells and of recombinant USF^{43} with the E box of the HIV-1 LTR and other E boxes are shown in Figure 5. A double-stranded oligonucleotide corresponding to the sequence from -174 to -151 of the LTR (oligo HIV) specifically binds to purified USF in a gel retardation assay resulting in the formation of a retarded complex (Figure 5 panel B, lane 2). This complex can be competed by the addition of an excess of cold oligonucleotides with the same sequence (lanes 9-11) or corresponding to other E box sequences (B48BS, contained in a human origin of DNA replication (Giacca *et al.*, 1994); AdMLP, the MLP upstream element; GAL2, a sequence upstream of the yeast GAL2 gene (Giacca *et al.*, 1989)). An oligonucleotide encompassing the two downstream-positioned Sp1 sites of the LTR (oligo SP1, lane 15) is ineffective in the competition assay.

Similar to the purified factor, the recombinant USF^{43} protein also binds to the HIV oligo in gel retardation assays (Figure 5 panel C, lane 2). The addition of anti-rUSF^{43} antibodies to the binding reaction causes a supershift of the protein-DNA complex, with analogy to the effect described for the purified protein (Csordás Tóth *et al.*, 1993) (Figure 5 panel D). Again, the specificity of binding was challenged by the addition of a 6 to 30 fold excess of other cold oligonucleotides to the binding reactions. Competition was obtained with the same HIV oligonucleotide (Figure 5 panel C, lanes 3 and 4) and with the related oligonucleotides AdMLP and B48BS (lanes 5-8), but not with the oligonucleotide containing the Sp1 sites (lanes 11-12).

An oligonucleotide carrying a TpA dinucleotide in the core position of the E box instead of CpG (CA<u>TA</u>TG, oligo MUT-III) is not able to compete for binding (lanes 9 and 10). This result indicates that these nucleotides are essential for the specificity of sequence recognition by the protein as already suggested (Bendall and

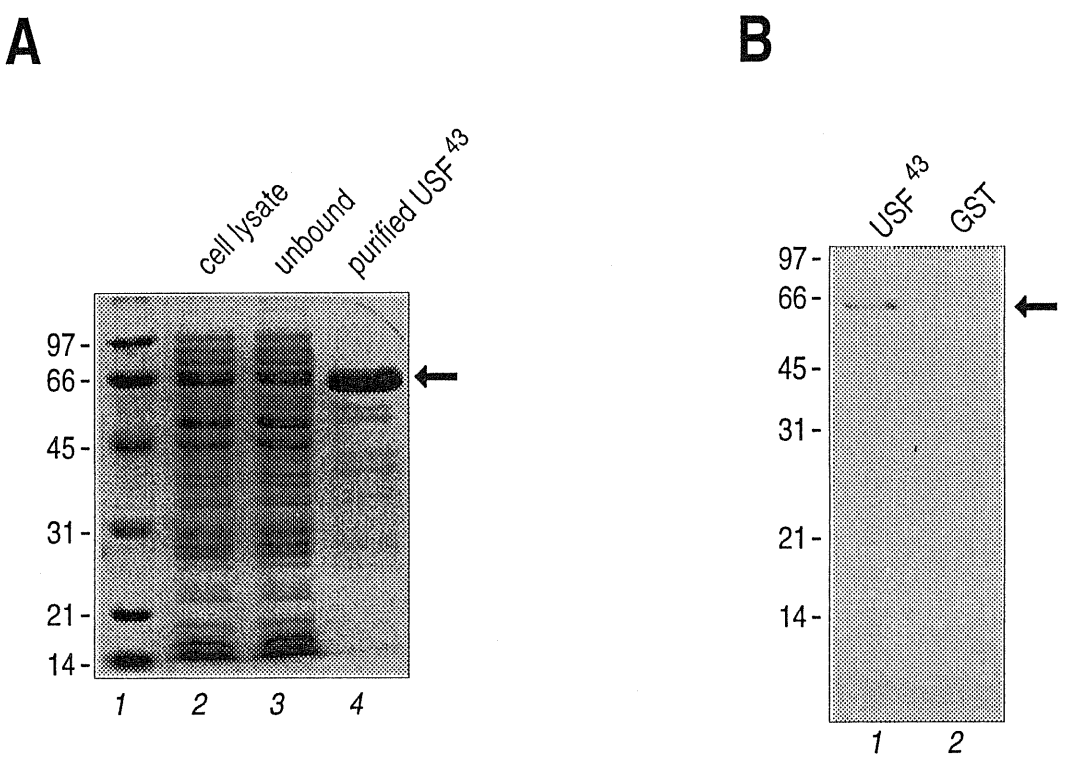


Fig. 4. Expression and purification of USF^{43}. Panel A. Single-step purification of recombinant USF^{43}. Protein samples were resolved on a 12% Laemmli-gel and stained with Coomassie blue. Lane 1: protein molecular weight markers; lane 2: 10 µl of the supernatant of a bacterial lysate of the SF8 E. coli strain transformed with pGST-USF^{43} after induction with IPTG; lane 3: 10 µl of the flow-through fraction from the agarose-glutathione column; lane 4: 40 µl of a fraction obtained by elution of the column with free glutathione. The arrow indicates the position of the GST-USF^{43} fusion protein. Panel B. Western blot analysis. Lane 1: purified USF^{43} (15 ng); lane 2: purified GST (15 ng). The arrow indicates the position of the GST-USF^{43} fusion protein. The positions of protein molecular weight markers are indicated.

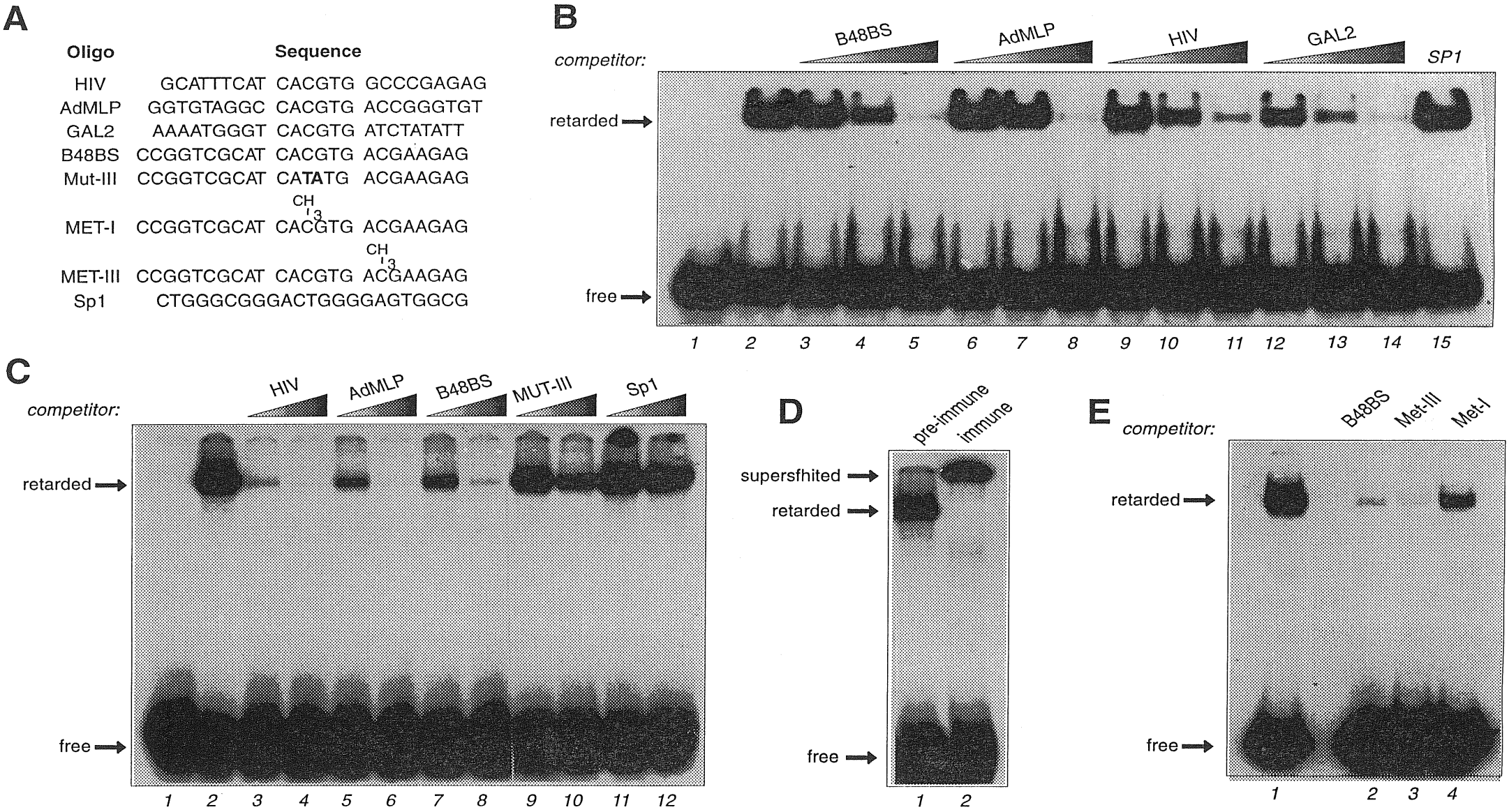


Fig. 5. Gel retardation and competition assays. Panel A. Sequences of the upper strands of the oligonucleotides used for the gel retardation and competition assays (see Materials and Methods for their localizations). The E box sequence is boxed. Mutations of this sequence in the oligo MUT-III are indicated by boldface. The localizations of the methyl groups in oligos Met-I and Met-III are indicated. Panel B. Gel retardation and competition assays with purified USF and oligo HIV as a probe. For the competition experiments, increasing amounts of competitor (6-, 20-, 60-fold molar excess with respect to the probe; 60-fold only for the Sp1 competitor) were added to the binding reactions as indicated on top of the Figure. The arrows indicate the positions of the free and retarded bands. Lane 1: probe without protein.
Panel C. Gel retardation and competition assays with recombinant USF^{43} and oligo HIV as a probe. Lane 1: probe without protein. Competitor molar excesses were 6 and 30 fold for each competitor. Panel D. Gel retardation-supershift analysis with anti-rUSF^{43} antiserum. The binding reaction between oligo HIV and USF^{43} was incubated with anti-rUSF^{43} antiserum (lane 2) or preimmune serum (lane 1). The arrows indicate the position of the free, retarded, and supershifted bands. Panel E. Effect of cytosine methylation on USF^{43} binding to oligo HIV. Lanes 1-4: oligo HIV plus USF^{43}; lanes 2, 3, and 4: competition with a 30-fold molar excess of the indicated oligonucleotides.

Molloy, 1994; Dang *et al.*, 1992). Since the CpG dinucleotide is the target for physiological cellular methylation, we investigated the role of cytosine methylation within the E box by competition experiments using methylated oligonucleotides with the same sequence of B48BS (Figure 5 panel E). The results indicated that, whilst methylation on both strands of a CpG outside of the E box consensus sequence (oligo Met-III) has no effect on the competing ability of the oligonucleotide with respect to the unmethylated sequence (compare lanes 2 and 3), methylation of the core CpG dinucleotide of the E box (oligo Met-I) greatly affects competition (lane 4). These results obtained with the recombinant USF^{43} protein reflect those described for crude nuclear extracts from HeLa cells (Giacca *et al.*, 1989).

Interactions of USF with the E box of the HIV-1 Long Terminal Repeats

Further details of the interactions between purified USF or recombinant USF^{43} and the HIV-1 LTR were obtained by a DNase I footprinting assay. As shown in Figure 6 panel A, the incubation of a LTR DNA probe labeled at the 5'-end of the non-coding strand with the purified factor (lane 3) prevents DNase I cleavage of a box centered around the CACGTG sequence (nucleotides from -173 to -157 upstream of the transcription start site). Lanes 4 and 5 show the results of DNase I digestion of a probe derived from plasmid pLTRΔUSF, in which the CACGTG motif was mutated into an *Eco*RI site (GAATTC). In this case, no difference can be detected in the DNase I digestion pattern obtained in the presence or absence of the protein, again indicating that the E box is absolutely required for protein binding. Protein binding causes the appearance of four strong hypersensitive sites at the 3' end of the recognized box, corresponding to nucleotides G_{-179}, A_{-178}, T_{-177}, and T_{-174} (arrows on the right side of Figure 6 panel A). A similar pattern of footprinting is produced by the binding of recombinant USF^{43} (Figure 6 panel C). On the upper, coding strand, the recombinant protein protects the E box form DNase I digestion, footprinting over nucleotides -157 to -172, again generating three strong hypersensitive sites at the 3' end (nucleotides C_{-152}, G_{-150}, and C_{-149}, Figure 6

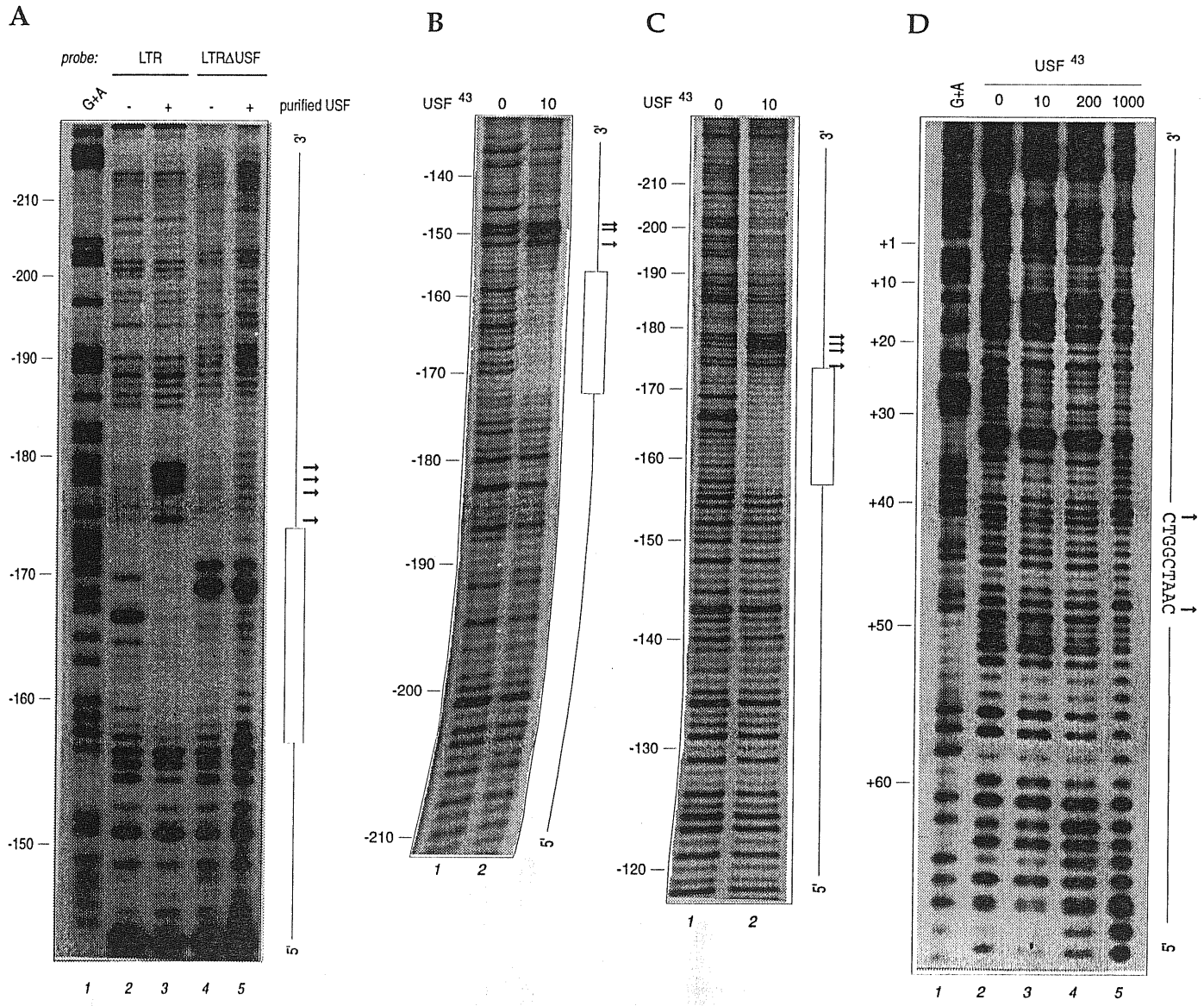


Fig. 6. DNase I footprinting assays. Panel A. DNase I footprinting using purified USF. Lane 1: G+A chemical cleavage ladder; lanes 2 and 3: DNA fragment pattern generated by DNase I treatment after incubation of a 5'-end labelled probe in the absence (lane 2) or in the presence (lane 3) of USF; lanes 4 and 5: same as lanes 2 and 3, but using a DNA probe with a mutated E box. The box and the arrows on the right side indicate the protected region and the hypersensitive nucleotides, respectively, of lane 3. Panel B. DNase I footprinting on the LTR coding strand using recombinant USF^{43}. Lane 1: 10 ng of GST added; lane 2: 10 ng of recombinant GST-USF^{43} fusion protein. The box and the arrows on the right side indicate the protected region and the hypersensitive nucleotides, respectively, of lane 2. Panel C. DNase I footprinting on the LTR non-coding strand with recombinant USF^{43}. Lane 1: 10 ng of GST; lane 2: 10 ng of recombinant GST-USF^{43} fusion protein. Panel D. DNase I footprinting on the region encompassing the transcription start site with recombinant USF^{43}. Lane 1: G+A chemical cleavage ladder; lane 2: 10 ng of GST; lanes 3-5: 10, 200, 1000 ng of recombinant GST-USF^{43} fusion protein. A sequence with altered DNase I sensitivity appearing using 1000 ng of protein, with the formation of two hypersensitive sites indicated by arrows, is shown on the right side.

panel B). The generation of these DNase I hypersensitive sites is most likely due to a structural alteration of the double helix upon protein binding resulting in a widened groove where nucleotides are more accessible to DNase I digestion.

It has been reported that USF[43], in addition to E box elements, can also bind to the initiator elements of the adenovirus MLP and of HIV-1 (Du *et al.*, 1993; Roy *et al.*, 1991). Therefore, we challenged a DNA probe encompassing the HIV-1 transcription start site in DNase I footprinting experiments with increasing amounts of highly purified (>95% homogeneity) USF[43] (from 10 ng to 1000 ng, Figure 6 panel D). No changes were observed using up to 200 ng of protein, in contrast with the clear protection of the E-box which is complete with even 10 ng, panels B and C). Only upon addition of 1 μg of protein, some changes in the footprinting pattern were detected in the region from nucleotides +42 to +47 relative to transcription start site, with the formation of two hypersensitive sites at nucleotides C_{40} and C_{48}. However, it should be considered that this sequence does not fully match to the previously reported Inr elements of the HIV-1 promoter (Du *et al.*, 1993), and that these changes can be non specific, due to the very high amount of protein used.

A further insight into the pattern of interaction of USF with the HIV-1 LTR was obtained by methylation protection experiments with dimethyl sulfate (DMS), a sensitive chemical method to probe DNA-protein contacts which allows the determination of the purines protected from methylation by DNA-bound proteins. The results of methylation protection experiments with USF[43] and the HIV-1 LTR as a probe are shown in Figure 7 panel A for the upper, coding strand and panel B for the lower, non-coding strand. Guanines at positions -162 and -164 on the coding strand, and -165, -167, and -170 on the non-coding strand are clearly protected from methylation, while guanine at position -161 on the coding strand is hypersensitive.

The results obtained by the DNase I and methylation protection experiments are summarized in Figure 8. It is evident that the E box consensus sequence CACGTG is centered at the 2-fold rotational axis of symmetry of protein-DNA interaction, with major contacts occurring at the 3' half of the DNA sequence on both

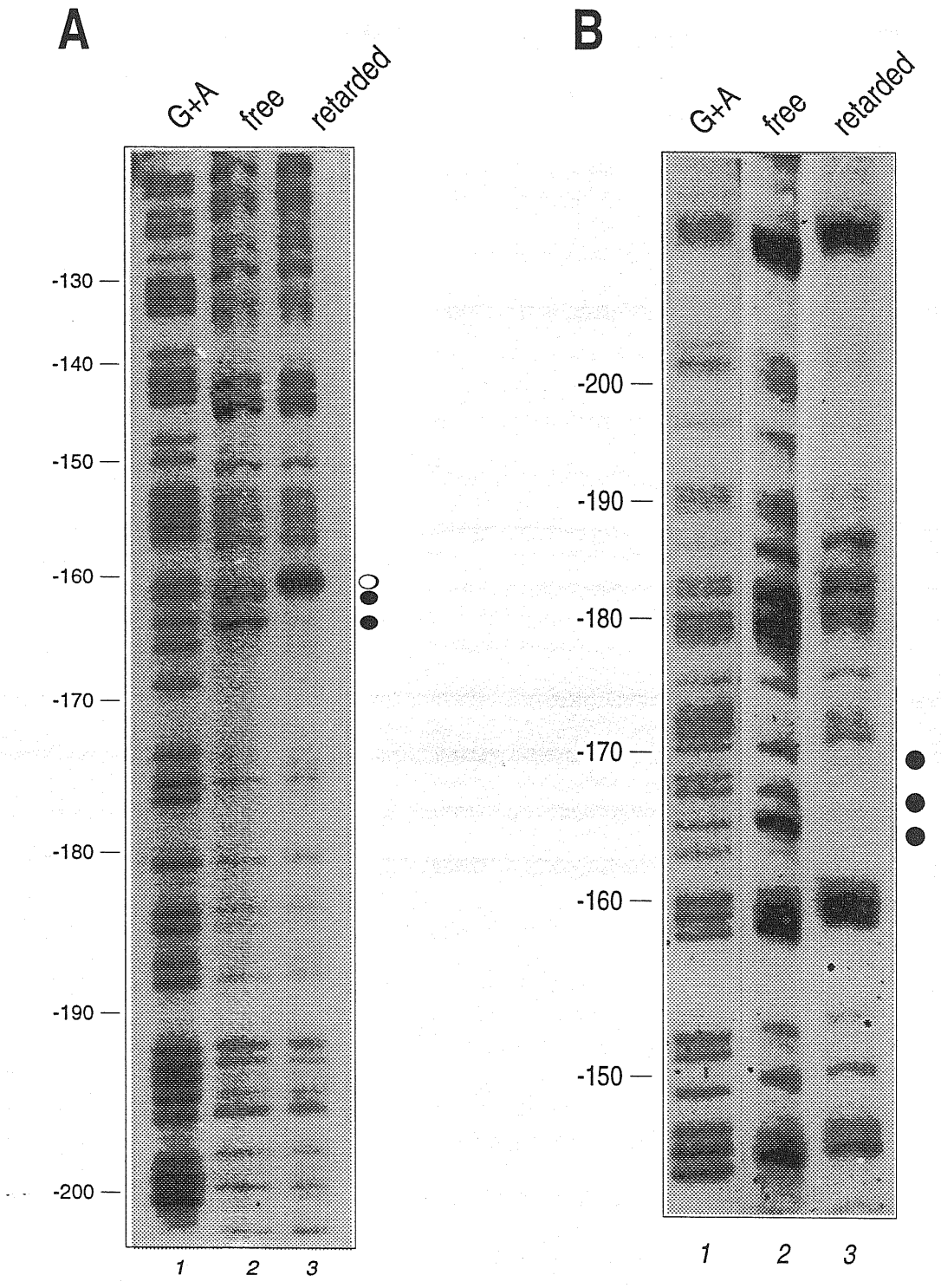


Figure 7. Methylation protection assays. Panel A. Methylation protection on the LTR coding strand using recombinant USF[43]. Lane 1: G+A chemical cleavage ladder; lane 2: methylation pattern of the free band recovered from the gel shift assay; lane 3: methylation pattern of the retarded band. The filled and empty circles on the right side indicate, respectively, protected and hypersensitive purines. Panel B. Methylation protection on the non-coding strand of the LTR using recombinant USF[43]. Lane 1: G+A chemical cleavage ladder; lane 2: methylation pattern of the free band recovered from the gel shift assay; lane 3: methylation pattern of the retarded band. The filled circles on the right side indicate protected purines.

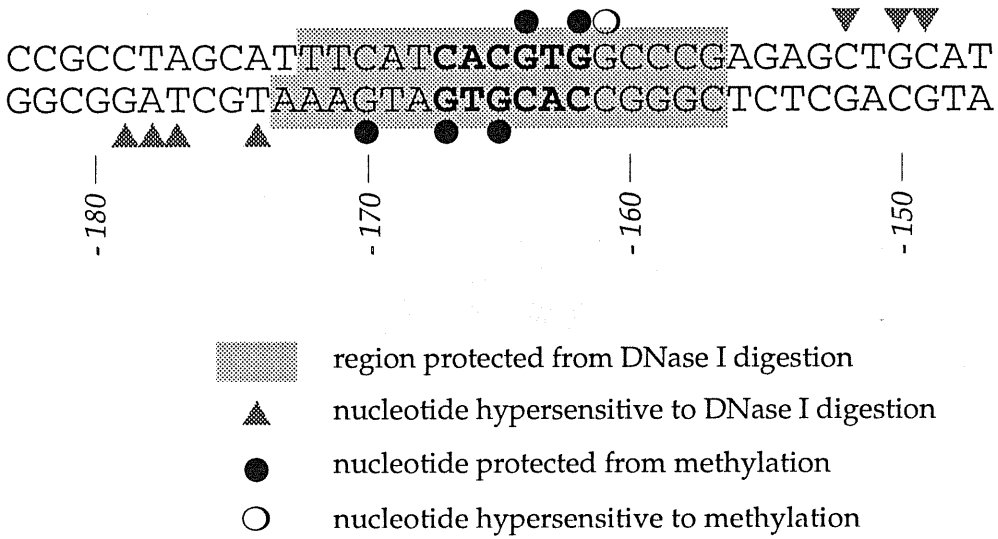


Fig. 8. Summary of the interactions of USF with the LTR E box obtained by DNase I footprinting and methylation protection experiments. The nucleotides of the E box motif are typed with bold characters.

strands. This pattern closely resembles the one produced by USF binding to the MLP upstream element (Miyamoto *et al.*, 1985; Sawadogo and Roeder, 1985b).

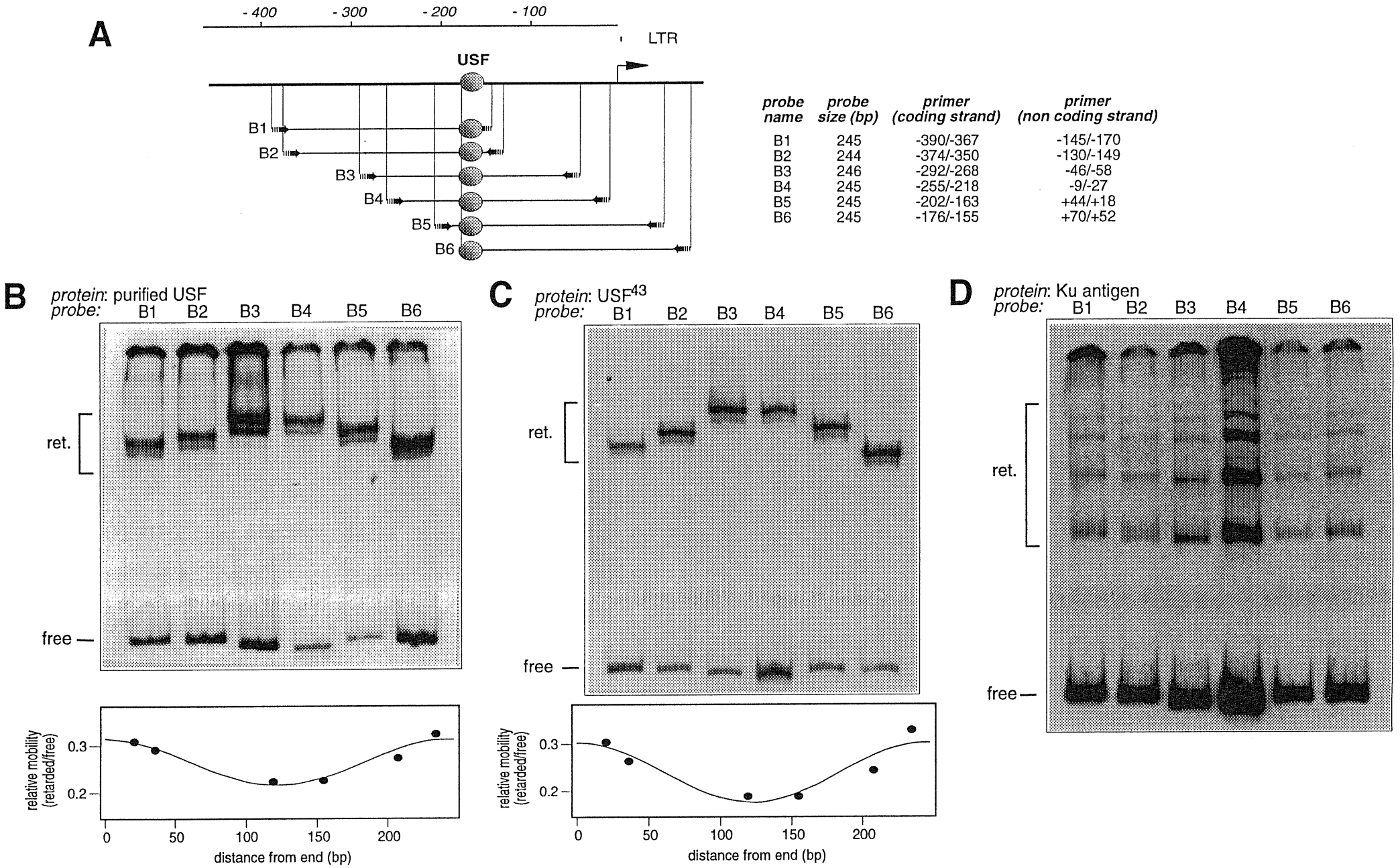
Template LTR bending upon USF binding

The generation of sites of increased sensitivity to DNase I symmetrically located on both strands and the presence of a site hypersensitive to methylation on the coding strand (see Figure 8), suggest that DNA undergoes a structural distortion upon protein binding.

The ability of USF to bend the LTR template upon binding was tested by circular permutation assay, a method based on the position-dependent effect of DNA bends on the electrophoretic mobility of DNA fragments (Wu and Crothers, 1984). To exploit this technique, a set of six probes (B1 to B6) was obtained by PCR amplification using the LTR as a template. All these probes have approximately the same length (from 244 to 246 bp) but differ in the position of the USF binding site (Figure 9 panel A). They were tested in gel retardation assays both with USF^{43} and with purified USF (Figure 9 panels B and C respectively). As expected, all the probes gave rise to a retarded complex upon incubation with both protein species. However, the relative mobilities of the retarded bands of the probes containing the binding site in the middle were lower than those of the probes containing the binding site at one extremity, suggesting that the template DNA becomes bent upon protein binding. The centre of the flexure was mapped by plotting the mobility as a function of the distance of the binding site from the end of the probe (Figure 9 panels B and C, lower graph) and turned out to correspond to the position of the CACGTG box.

Estimation of the bending angle according to Thompson and Landy (Thompson and Landy, 1988) indicates values of 90° of bending for purified USF and 110° for recombinant USF^{43}, as evaluated on the average values obtained from three independent experiments.

As a negative control, the Ku protein, which binds to DNA without an apparent sequence specificity giving rise to several retarded complexes due to



probe name	probe size (bp)	primer (coding strand)	primer (non coding strand)
B1	245	-390/-367	-145/-170
B2	244	-374/-350	-130/-149
B3	246	-292/-268	-46/-58
B4	245	-255/-218	-9/-27
B5	245	-202/-163	+44/+18
B6	245	-176/-155	+70/+52

Fig. 9. Circular permutation assay. Panel A. A set of six probes (B1 to B6) containing the USF site in different position was generated by PCR amplification using the primers localized as indicated in the Table on the right side of the panel (numbering is referred to transcription start site according to sequence hxb2cg of GenBank). The arrow on top indicates transcription start site. Panel B. Circular permutation assay using probes B1-B6 and purified USF. The localizations of the free and retarded complexes are indicated. The graph in the lower part of the panel shows the relative mobility of the retarded complexes (distance from the well of the retarded band/distance of the free band) plotted against the distance of the CACGTG box from the end of the fragment. Panel C. Circular permutation assay with recombinant USF^{43}. Panel D. Circular permutation assay with the Ku protein.

multiple protein dimers bound to the probe (Csordás Tóth *et al.*, 1993), was also tested in gel retardation assays with the same probes. As shown in Figure 9 panel D, the retarded bands generated by all the probes migrate with the same mobility.

In vitro transcription assays

The role of USF in the control of transcription from the HIV-1 LTR was determined by in vitro transcription assays. For this purpose, the LTR region upstream of nucleotide -1 was cloned upstream of a ~380 nt G-less cassette. Each in vitro transcription experiment was simultaneously performed by the addition to the same tube of a plasmid template containing a shorter G-less cassette (~200 nt) under the control of the fibronectin promoter, to be used as an internal control.

The addition of increasing amounts of highly purified recombinant USF[43] to HeLa cell nuclear extracts progressively upregulates transcription from the LTR in a concentration-dependent manner, while it does not affect the transcription driven by the fibronectin promoter (Figure 10 panel A, lanes 2-6). Quantification by scintillation counting and optical scanning of the autoradiograms indicated a reproducible three-to fourfold increase in LTR-driven signal intensity after normalization to fibronectin; the ratio between the intensity of the RNA bands for the two constructs for each experimental point is reported in the graph below the gel .

Mutation of the CACGTG hexanucleotide to an unrelated *Eco*R I restriction site abolishes responsiveness of the LTR construct to protein addition (Figure 10 panel B, lanes 2-6). Furthermore, the transcription efficiency of this mutated promoter is invariably lower than that of the wild type LTR (compare lanes 1 in panels A and C with lanes 1 in panels B and D).

The HeLa cell extract used in the in vitro transcription experiments contains, among other E box binding factors, an endogenous amount of USF (estimated in the range of 0.15-1.5 ng/μg of extract (Pognonec and Roeder, 1991)), which presumably contributes to the basal level of transcriptional activation from the LTR template. In order to saturate the endogenous E box binding proteins, increasing amounts of a

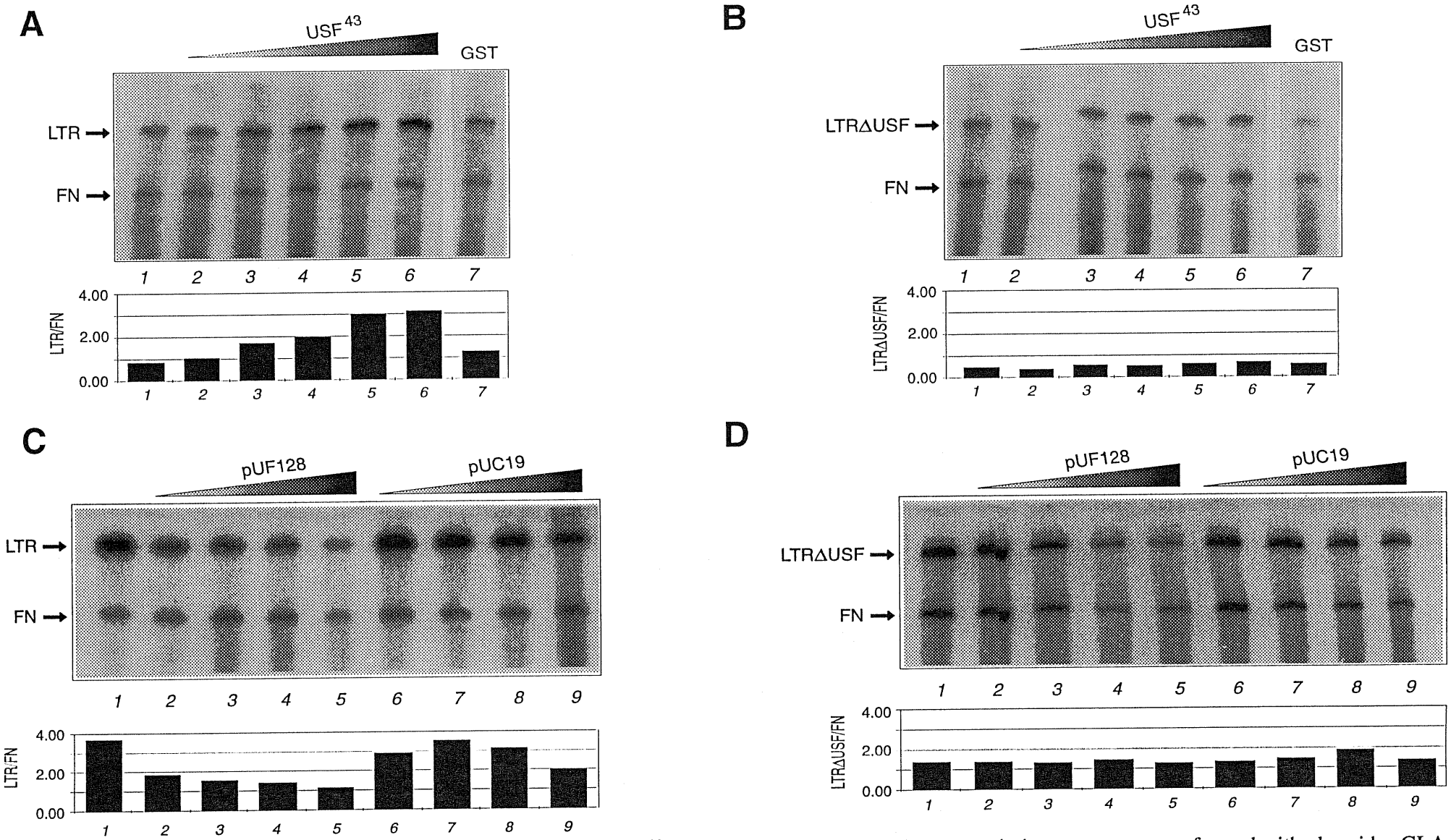


Fig. 10. In vitro transcription assays. Panel A. Effect of the addition of USF^{43} on the LTR template. In vitro transcription assays were performed with plasmids pGLA (LTR upstream of a 380 nt G-less cassette) and pFN2 (fibronectin promoter upstream of a 200 nt G-less cassette), and HeLa nuclear extracts. Lane 1: plasmid templates plus nuclear extract; lanes 2-6: plus the addition of increasing amounts of recombinant GST-USF^{43} (50, 100, 200, 500, 1000 ng, respectively); lane 7: plus the addition of 1000 ng of GST. The RNA bands corresponding to the two transcripts are indicated. The graph in the lower part of the panel shows the ratio between the amount of the two transcripts for each experimental point. Panel B. Effect of the addition of USF^{43} on the LTRΔUSF template. In vitro transcription assays were performed with plasmids pGLE (LTRΔUSF/G-less cassette) and pFN2, and HeLa nuclear extracts. The amounts of USF^{43} and GST are the same as in panel A. Panel C. Effect of the addition of E box decoys on the LTR template. In vitro transcription assays were performed with plasmids pGLA and pFN2, and HeLa nuclear extracts. Lane 1: plasmid templates plus nuclear extract; lanes 2-5: plus the addition of increasing amounts of plasmid pUF128, containing 128 copies of a CACGTG box (250, 500, 750, 1000 ng); lanes 6-9: plus the addition of the same molar excesses of vector pUC19. Panel C. Effect of the addition of E box decoys on the LTRΔUSF template. In vitro transcription assays were performed with plasmids pGLE and pFN2, and HeLa nuclear extracts. The amounts of pUF128 and pUC19 are the same as in panel C.

plasmid (pUF128) carrying 128 copies of a human E box containing sequence were added to the transcription reactions (Figure 10 panel C, lanes 2-5). The addition of this decoy progressively decreases the level of transcription from the LTR as compared to the fibronectin promoter. As a control, the same molar amounts of vector pUC19 are ineffective (lanes 6-9). Again, mutation of the E box of the LTR abolishes responsiveness of the LTR construct to the specific decoy addition (Figure 10 panel D, lanes 2-5).

Altogether, these results show that the E box in the context of the U3 LTR region is a positive cis-acting element in the control of transcription and that USF is most probably a trans-acting factor contributing to this function.

Function of USF^{43} in vivo

The above in vitro studies provided compelling evidences that USF specifically binds to the CACGTG hexanucleotide in the LTR and activates the transcription from this promoter in vitro.

To extend these observations, the question of the role of USF in vivo was addressed. The first goal was to verify whether USF could indeed bind to, and activate from, the CACGTG sequence of the LTR. To this purpose, we cloned four oligonucleotides, each one containing the sequence protected from DNase I digestion in the above described footprinting experiments, in a head-to-tail orientation. This USF responsive cassette was cloned upstream of a TATA box controlling the expression of the reporter gene chloramphenicol acetyl transferase (CAT). This construct acts as a synthetic minimal promoter specifically responsive to USF. We also constructed a vector expressing USF^{43} in eukaryotic cells under the control of the cytomegalovirus early promoter. These plasmids were transiently expressed by transfection in HeLa cells. As expected, the expression level of the transfected reported plasmid alone is low. Cotransfection of increasing amounts of the USF^{43} expressing plasmid progressively activates its expression (see Figure 11, panel A). These results demonstrate that the DNA fragment bound in vitro by USF

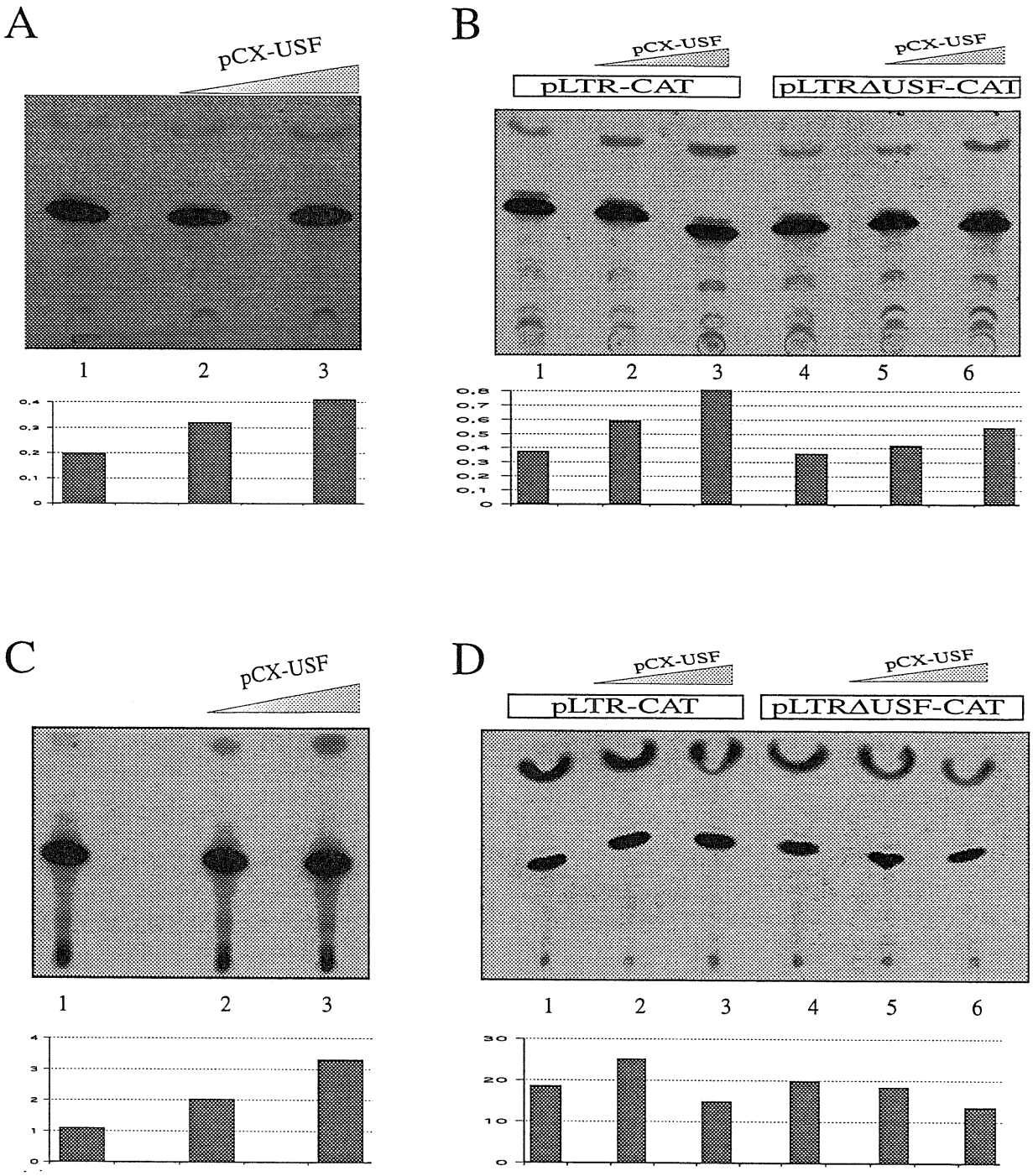


Fig. 11. Panel A. CAT activity of HeLa cells transfected with plasmid pUSFminCAT carrying four USF binding sites (whose sequence is derived from the LTR of HIV-1) and a TATA box upstream of the CAT gene. Basal CAT activity of 5 μg of pUSFminCAT (lane 1) is triggered by the cotransfection of 5 and 10 μg of pCX-USF (lanes 2 and 3 respectively), a plasmid expressing USF[43]. Panel B, CAT activity of HeLa cells transfected with the plasmid pLTRCAT and pLTRΔUSFCAT. In the two plasmids the transcription of the CAT gene is controlled by the wild-type LTR or by a mutant promoter carrying a mutation in the USF binding site respectively. Basal CAT activity of 5 μg of pLTRCAT (lane 1) is triggered by the cotransfection of 5 and 10 μg of pCX-USF (lanes 2 and 3 respectively). Lane 4 shows the activity of the LTRΔUSF promoter and lanes 5 and 6 show the effect of the cotransfection of 5 and 10 μg of PCX-USF. Panel C, CAT activity of HeLa-Tat cells, a HeLa cell line derivative constitutively expressing Tat, transfected with the plasmid pUSFminCAT. Basal CAT activity of 5 μg of pUSFminCAT (lane 1) is triggered by the cotransfection of 5 and 10 μg of pCX-USF (lanes 2 and 3 respectively), a plasmid expressing USF[43]. Panel D, CAT activity of HeLa-Tat cells transfected with the plasmid pLTRCAT and pLTRΔUSFCAT. The experimental design was the same as for panel B.

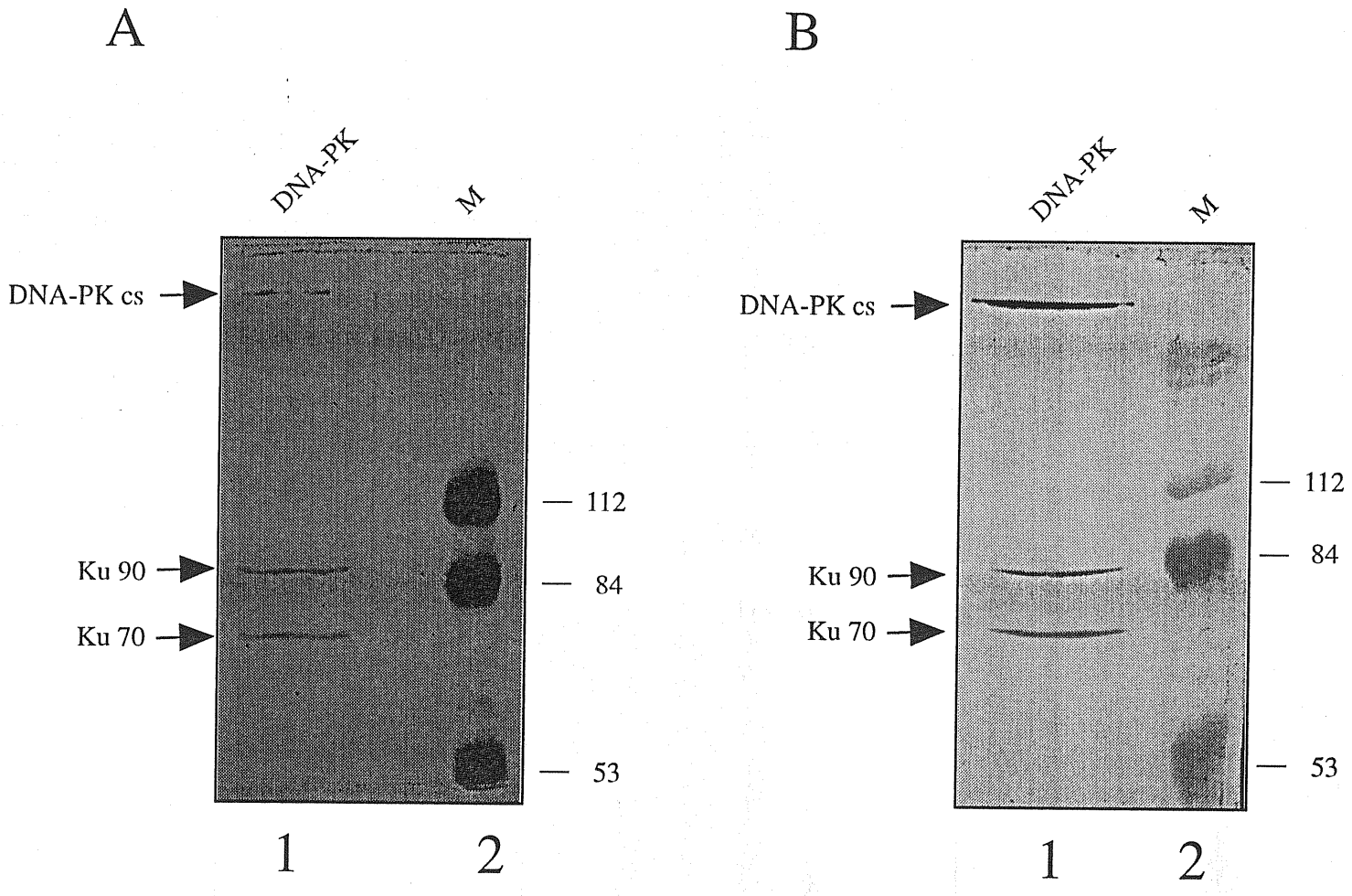


Fig. 12. Purification of DNA-PK. (A) Purified DNA-PK. Proteins were resolved on a 10 % Laemmli gel and stained with Coomassie blue. Lane 1, 25 µl of purified DNA-PK; lane 2, protein molecular weight markers. (B) Western blot analysis. Lane 1, 10 µl of purified DNA-PK; lane 2, protein molecular weight markers.

can also be recognized by USF in vivo and that the USF^{43} protein can activate transcription efficiently in vivo.

The next object was to study the effects of the overexpression of USF^{43} on the activity of the entire LTR. To this end, we cotransfected the LTR-CAT plasmid, in which CAT transcription is controlled by the HIV-1 LTR, with increasing amounts of the USF expressing plasmid. As shown in Figure 11, panel B, the USF vector progressively augmented LTR activity, indicating that USF is a positive regulator of LTR directed transcription also from the whole LTR. Interestingly, the mutation of the target sequence for USF into an unrelated EcoRI restriction site impaired USF transactivation, but not completely, suggesting that, in vivo, USF can weakly mediate LTR transactivation also independently of the E-box.

The same experiment was also performed by the cotransfection of a Tat-expressing plasmid in HeLa cells, or by using the HeLa-Tat cell line, stably expressing Tat. In the presence of Tat, USF^{43} activates the synthetic minimal promoter, as efficiently as when Tat is absent (Figure 11, panel C). However, when we tested the effect of USF^{43} overexpression on the activity of the whole LTR in the presence of Tat, USF overexpression did not significantly affect LTR driven CAT transcription (Figure 11, panel D). In the discussion section we will try to explain this unexpected result.

DNA-PK phosphorylates USF^{43}

In order to address the problem of the role of post-transcriptional modifications of factors involved in the control of transcription of HIV-1, and taking advantage of the availability of DNA-PK (the DNA-dependent protein kinase), purified in our laboratory, we tested whether DNA-PK could modify USF^{43} or Tat.

Figure 12, panel A, shows a Coomassie blue stained SDS-polyacrylamide gel in which the purified DNA-PK employed in these assays was loaded. From this gel we estimate that DNA-PK is 80% pure. To verify the genuinity of the purified polypeptides, they were probed in western blot assays using antibodies against

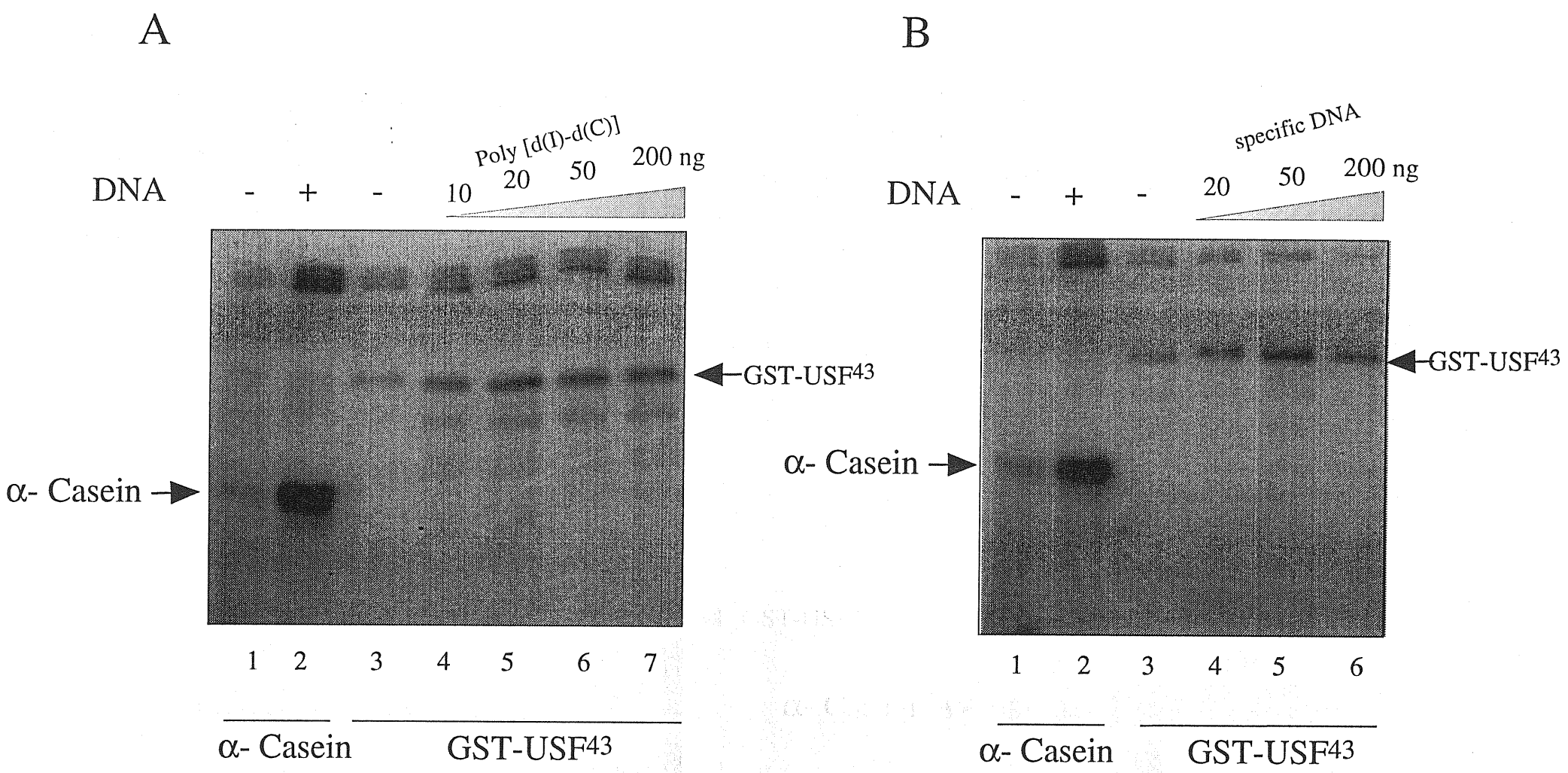


Fig. 13. Phosphorylation of GST-USF^{43} by purified DNA-PK. (A) Lanes 1 and 2, α-casein phosphorylation as control; lane 3-7, phosphorylation of GST-USF^{43} without or with 10, 20, 50, 200 ng of poly [d(I)-d(C)]. (B) Lanes 1 and 2, α-casein phosphorylation as control; lane 3-6, phosphorylation of GST-USF^{43}, without or with 20, 50, 200 ng of ligated double stranded oligonucleotides containing a binding site for USF. The upper band in each lane represents unresolved material at the boundary between stacking and running gels.

DNA-PKcs (a kind gift of T. Shenk) and our own made antibodies raised against Ku. As shown in Figure 12, panel B , the antibodies clearly recognize polypeptides of the expected sizes.

On the basis of the predicted amino acidic sequence of the 43 kDa form of USF, DNA-PK can phosphorylate this protein at threonine 107, 111, 162, and at serine 194. Threonine 111 and serine 194 are potential phosphorylation sites also for casein kinase II (CK II).

The first two phosphorylation sites lie in a region that was shown to be involved in transcriptional transactivation of USF in vitro (Kirschbaum *et al.*, 1992), while serine 194 flanks the DNA binding domain of USF. Consequently, phosphorylation of this residues could have a role in the control of transactivation or DNA binding.

Since DNA-PK is sensitive to the presence of DNA (being activated by it), and since bacterially-purified transcription factors are often contaminated by bacterial DNA and RNA, we extensively treated recombinant factors with RNAse A and DNase I before performing the in vitro phosphorylation experiments (see Materials and Methods).

Figure 13 shows the results of a phosphorylation assay using DNA-free USF^{43} and purified DNA-PK. These results show that the addition of increasing amounts of poly[d(I-C)]:poly[d(I-C)] (in a range from 10 to 200 ng) results in the progressive phosphorylation of USF, indicating that this factor is a substrate for this kinase.

We also employed ligated oligonucleotides bearing binding sites for USF as the activator nucleic acid in the phosphorylation assay. This oligonucleotide is active as well. However, higher amounts of this specific DNA inhibited the reaction.

To determine the precise position of the residue(s) phosphorylated by DNA-PK, we plan to make triptic digestion of the phosphorylated protein in order to determine the precise localization of the phosphorylated residue(s).

PART II - EXPRESSION AND PURIFICATION OF BIOLOGICALLY ACTIVE TAT

Production of recombinant Tat proteins

Two recombinant Tat proteins were obtained as fusion products between glutathione-S-transferase (GST) and the first exon (GST-Tat1E) or both exons (GST-Tat2E) of the HIV-1 Tat protein. The cloning strategy was based on PCR amplification of the respective coding sequences and oriented cloning in the polylinker of the pGEX2T vector. Figure 14 shows the results of a single-step purification procedure to obtain recombinant GST-Tat1E (panel A lane 2) and GST-Tat2E (panel B lane 2) by affinity chromatography using glutathione-conjugated agarose beads starting from bacterial lysates. With respect to similar procedures for the purification of other GST-fusion proteins, the protocol for purification of Tat had to be optimized to avoid precipitation of the protein in inclusion bodies and prevent extensive oxidation of the protein. The procedure adopted is extensively detailed in the Materials and Methods section. Furthermore, it was serendipitously observed that the protein obtained by this procedure has low biological activity because it is strongly contaminated by bacterial nucleic acids (mostly E. coli ribosomal RNA, not shown). Therefore, an additional high salt wash step was included in the purification procedure to get rid of these contaminating nucleic acids. An example of the different biological activity of salt-treated in comparison to untreated Tat is presented in Figure 16 panel A, where the differently purified proteins were added to LTR-CAT containing HL3T1 cells (see below): it is easily observed that there is an over ten-fold difference in the amount of protein necessary to obtain the same biological activity (compare lane 2, salt-washed Tat1E 100 ng and lane 4, untreated Tat1E 1 μg).

Both GST-Tat1E and GST-Tat2E are specifically recognized by anti-Tat monoclonal antibodies in Western blotting assays (shown for the latter protein in Figure 14 panel C).

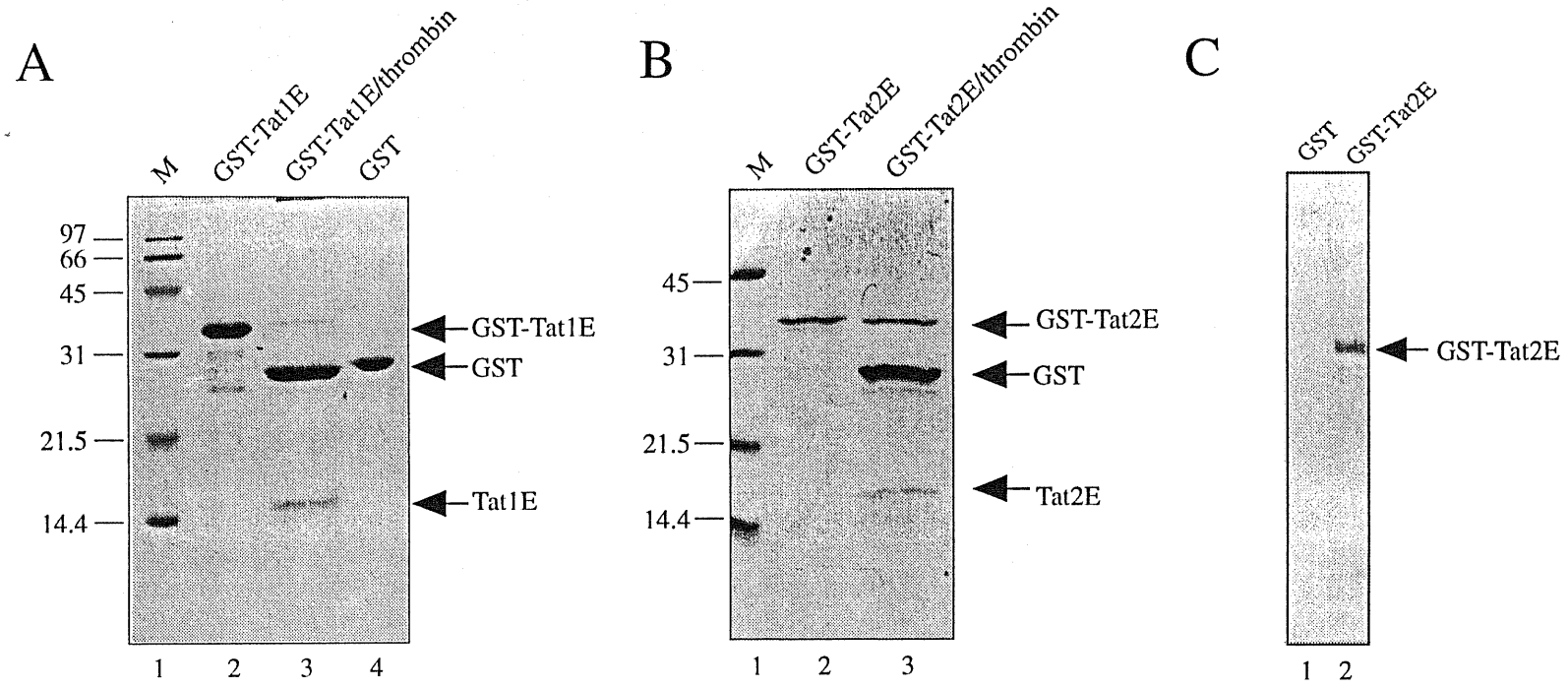


Fig. 14. Expression, purification and cleavage of Tat proteins. Proteins sample were resolved on a 12% Laemmli gel and stained with Coomassie blue. (A) Lane 1, protein molecular weight marker; lane 2, 20 µl of purified GST-Tat1E; lane 3, 20 µl of purified GST-Tat1E after cleavage with thrombin protease; lane 4, 5 µl of purified GST. (B) Lane 1, protein molecular weight marker; lane 2, 20 µl of purified GST-Tat2E; lane 3, 20 µl of purified GST-Tat2E after cleavage with thrombin protease (in this particular occasion, cleavage was only partial). (C) western blot analysis using antibodies against Tat. Lane 1, GST; lane 2, GST-Tat2E.

The two recombinant Tat proteins can be cleaved with thrombin to separate Tat from the GST moiety. Figure 14, panel A, lane three and panel B, lane 3 show Coomassie-stained SDS-polyacrylamide gels after enzymatic digestion. In should be noted, however, that in the course of our utilization of these Tat proteins in the experiments presented in this thesis as well as in other experiments carried out in the laboratory, no gross difference could be detected between thrombin-released Tat in comparison to Tat fusions. An exception to this finding are the gel retardation experiments, where only cleaved Tat preparations were effective in shifting the mobility of TAR RNA, while the fusion proteins caused precipitation of Tat-TAR complexes in the gel wells. An example of these experiments is presented in Figure 15, where increasing amounts of Tat2E (from 200 to 1600 ng) or Tat1E (from 500 to 1000 ng) were added to a radiolabelled TAR probe and resolved on a 5% polyacrylamide gel. The appearance of a retarded band is clearly visible.

Biological activities or recombinant Tat

The biological activity of recombinant Tat proteins was assessed by both in vivo and in vitro experiments. The in vivo results were obtained by using the HL3T1 cell line, a HeLa-derived cell line containing an integrated LTR-CAT construct, developed in G. Pavlakis' laboratory (Schwartz *et al.*, 1990). CAT activity in this cell line is almost undetectable in basal conditions, and can be strongly transactivated by a variety of stimuli, including transfection of Tat expression plasmids. Addition of purified Tat proteins to the medium of these cells results in a dose-dependent activation of CAT activity (Figure 16), that exploits the ability of the protein of freely cross cell membranes by a still uncharacterized mechanism (Frankel *et al.*, 1989; Frankel and Pabo, 1988).

We have explored different methods to optimize the efficiency of this treatment. In particular, to obtain the best transactivation is necessary to treat cells with chloroquine, a drug that stabilizes endolysosomal pH, thus preventing extensive protein degradation in endolysosomal vacuoles (Frankel and Pabo, 1988) (Figure 16

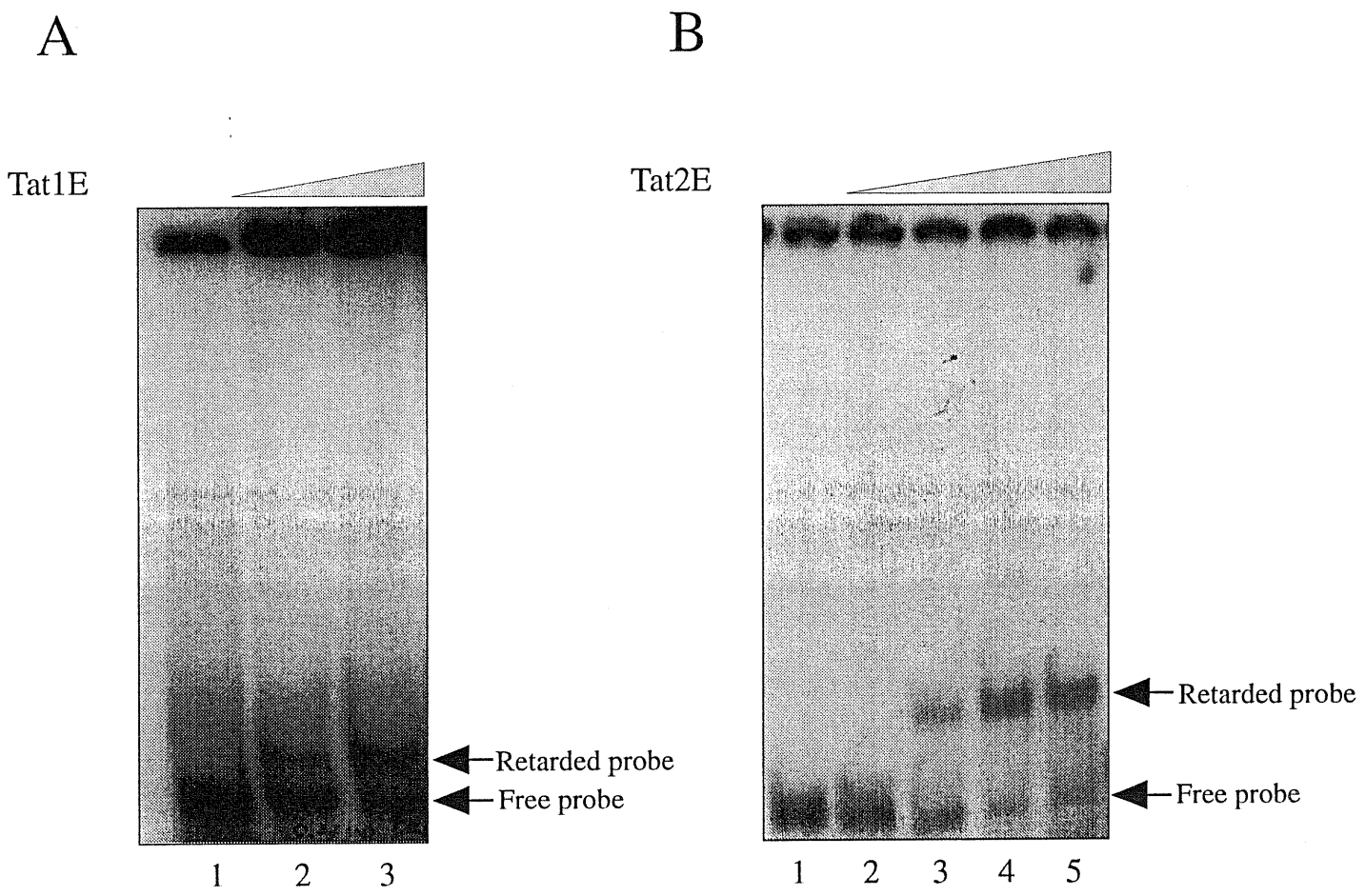


Fig. 15. Gel retardation assays with recombinant Tat. (A) Lane 1, TAR probe alone; lane 2 and 3, TAR plus 500 ng and 1000 ng of Tat1E respectively. (B) Lane 2, TAR probe alone; lane 2-5, TAR plus 20, 400, 1000 and 1600 ng of Tat2E.

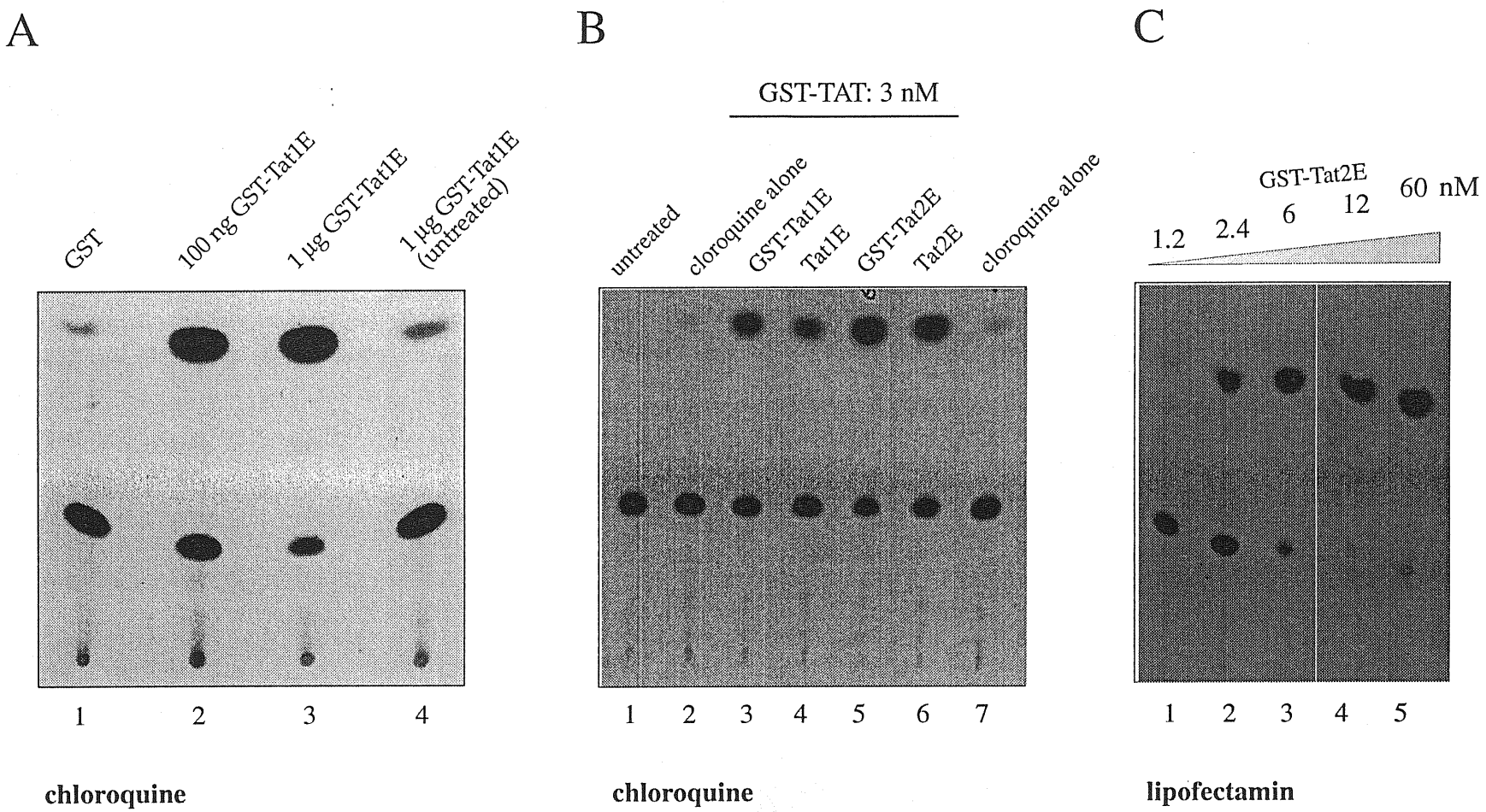


Fig. 16. CAT activity of HL3T1 cell extracts after cell treatment with recombinant Tat. (A) Induced CAT activity by differently purified recombinant Tat in the presence of chloroquine. Lane 1, effect of 1 µg of GST; lanes 2 and 3, effect of 100 ng and 1 µg of high salt washed GST-Tat1E respectively; lane 4, effect of 1 µg of untreated GST-Tat1E. (B) Induced CAT activity by 3 nM recombinant one-exon and two-exon Tat, prior or after cleavage with thrombin in the presence of chloroquine. Lane 1, CAT activity of untreated cells; lane 2, effect of chloroquine alone; lanes 3 and 4, effect of GST-Tat1E, prior and after cleavage with thrombin respectively; lanes 5 and 6, effect of GST-Tat2E, prior and after cleavage with thrombin respectively; lane 7, effect of chloroquine alone. (C) Induced CAT activity upon incubation with increasing amounts of GST-Tat2E in the presence of Lipofectamin. Lanes 1-5, effect of 1.2, 2.4, 6, 12, 60 nM GST-Tat2E.

panels A and B), or to use a lipofection procedure (panel C). The latter procedure is preferable, since cell treatment with chloroquine alone results in weak but detectable activation of CAT activity, related to the induction of the NF-κB factor (not shown). The relative transactivation activity of one-exon and two-exon Tat proteins was compared. No gross difference could be detected between the two proteins, neither in their uncleaved nor the in the cleaved forms (panel B lanes 3 to 6). These results are in agreement to former observations that indicated that the first exon of the protein is necessary and sufficient to obtain complete transactivation by reporter gene transfection assays (Arya *et al.*, 1985; Muesing *et al.*, 1987).

The transactivation ability of GST-Tat2E was also assessed by in vitro transcription assays. These experiments were performed by primer extension of RNA transcripts originating from the HIV-1 LTR using reverse transcriptase, using a radiolabeled oligonucleotide (see Materials and Methods). These experiments were carried out in the presence of an internal control, carrying an LTR derivative under the control of the RSV promoter. This derivative contains a 99 bp deletion in the CAT coding sequence, positioned downstream of transcription start site and upstream of the annealing site for the primer used for the detection of transcription product. As a consequence, this plasmid acts as an internal competitive template for transcript detection, thus taking into account all the experiments variables that could affect each in vitro transcription experiment. Figure 17 shows the results of the addition of recombinant GST-Tat2E to an in vitro transcription experiment. It is evident that the Tat protein is able to activate transcription from the LTR promoter but not from the RSV promoter.

The Tat protein is a substrate for DNA-PK phosphorylation

The analysis of the primary amino acidic sequence of Tat revealed three potential phosphorylation sites by DNA-PK. In fact, serine 16, 62, and 85 are followed by glutamine, and this is the minimal requirement for serine phosphorylation by DNA-PK (Bannister *et al.*, 1993). To test whether Tat protein is a substrate for DNA-PK in

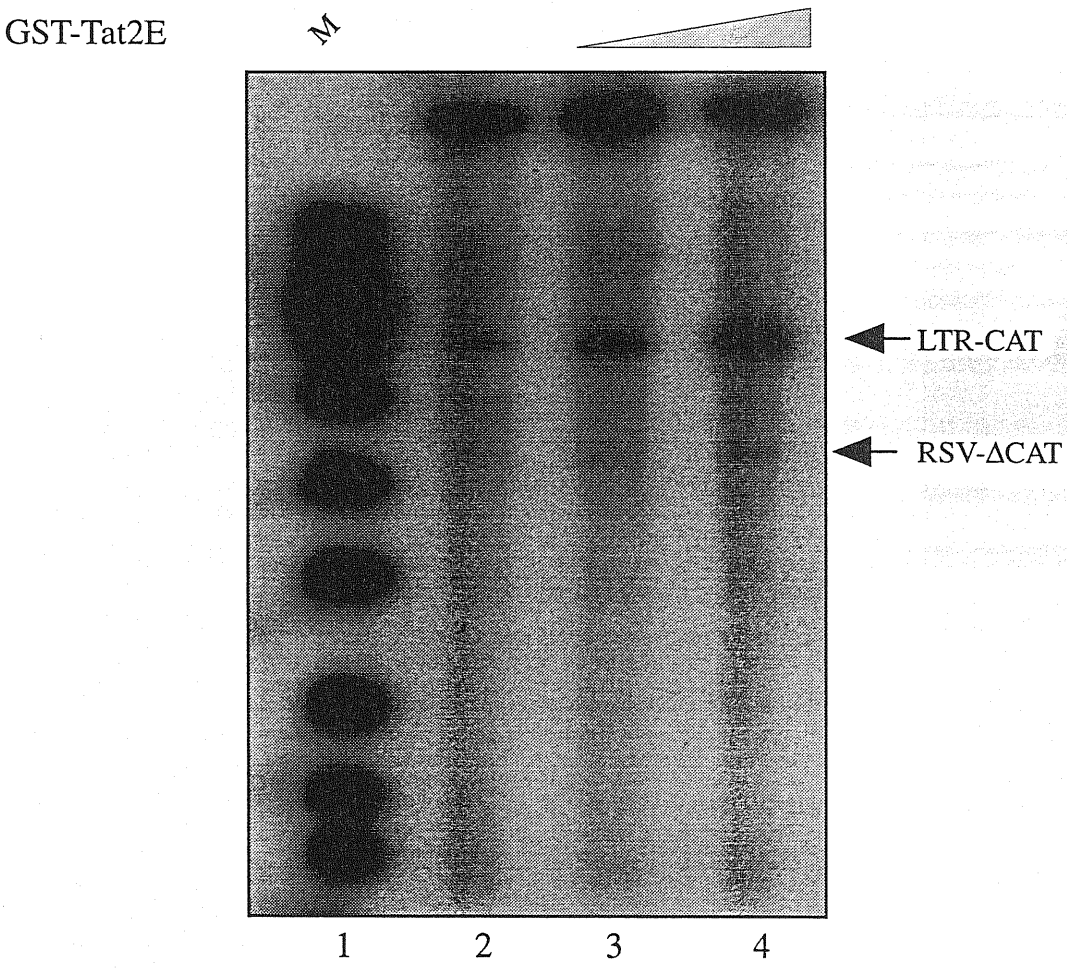


Fig. 17. In vitro transcription assay by primer extension analysis. In vitro transcription assays were performed with plasmids pLTR-CAT and pRSVΔCAT, bearing a deletion of 99 bases in the CAT gene and HeLa nuclear extracts. Lane 1, molecular weight marker; lane 2, plasmids and nuclear extract plus 1 μg of GST; lane 3-4 plasmids and nuclear extract plus 100 and 500 ng of GST-Tat2E.

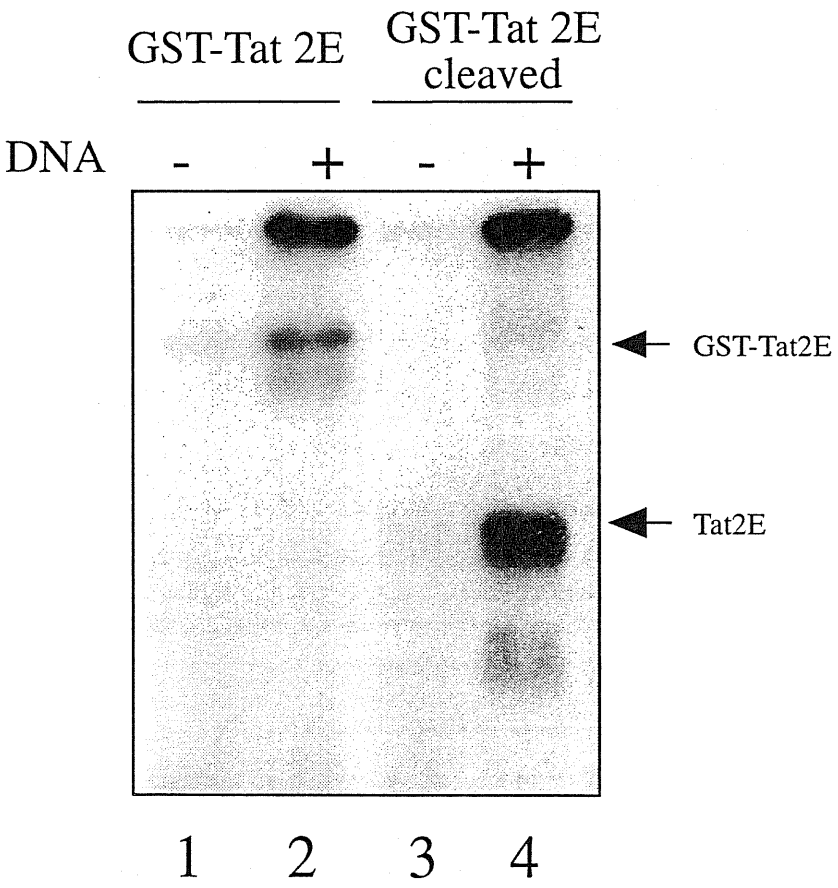


Fig. 18. Phosphorylation of Tat by purified DNA-PK. Lanes 1 and 2, phosphorylation of GST-Tat2E by DNA-PK in the absence or in the presence of DNA respectively. Lanes 3 and 4, phosphorylation of thrombin-cleaved GST-Tat2E by DNA-PK in the absence or in the presence of DNA respectively.

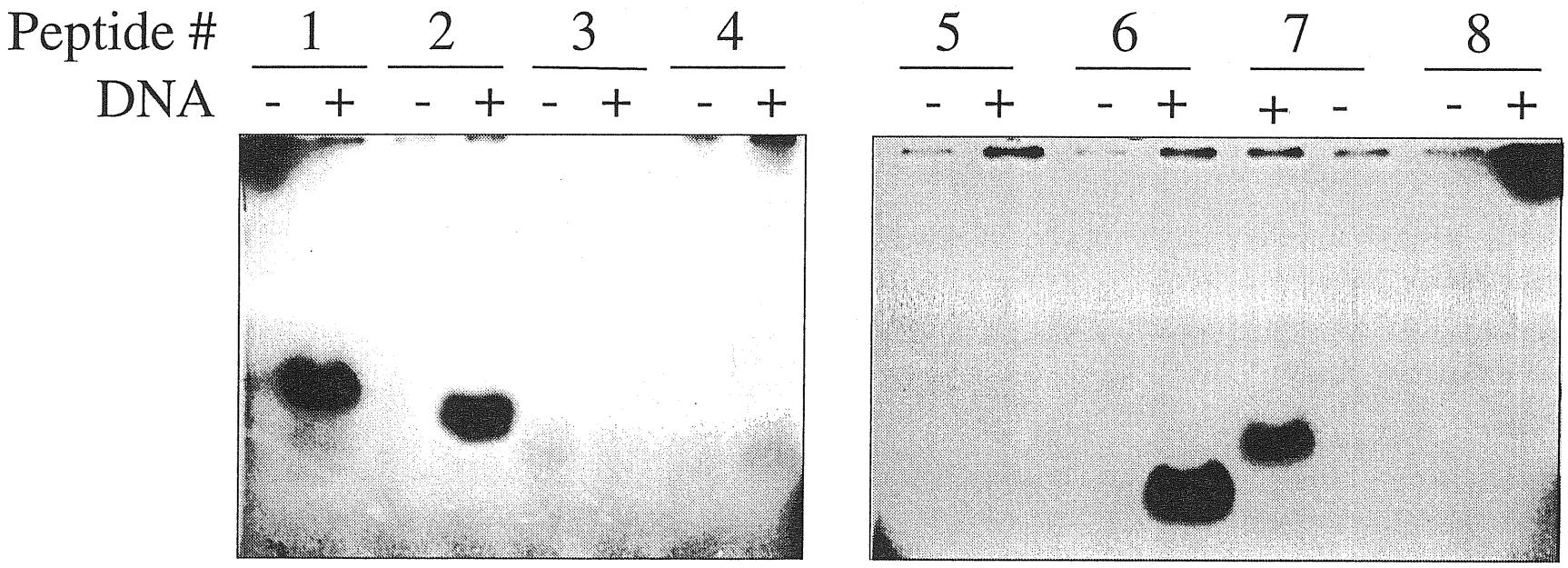


Fig. 19. Phosphorylation of a panel of 8 peptides spanning the sequence of Tat, by purified DNA-PK. The peptides are 20 amino acids long and each one has 10 amino acids in common with the preceding one.

vitro, we used DNase/RNAse treated Tat protein in our phosphorylation assays using purified DNA-PK. As shown in figure 18, the Tat protein is an efficient substrate for DNA-PK in the presence of DNA both when fused to the GST moiety and after its separation by proteolitic cleavage with thrombin; as expected, phosphorylation relies strictly on the presence of DNA, being the phosphorylation activity undetectable in the absence of it.

To precisely map the phosphorylation site(s), we employed a set of 8 peptides 20 amino acid long, spanning the entire Tat sequence. The results of phosphorylation assays using these peptides are shown in Figure 19. Each peptide was assayed in the presence and in the absence of DNA. From this experiment it was clear that serine 16 and serine 62 are phosphorylated by DNA-PK, while serine 85 is not. Probably, the arginine residue near the phosphorylation site inhibits the modification by DNA-PK. Indeed, it has already been reported that positively charged amino acids can negatively affect DNA-PK recognition on neighboring serines or threonines (Bannister *et al.*, 1993).

DISCUSSION

After the interaction of the HIV-1 virion with its cellular membrane receptor and the consequent internalization and decapsidation, the viral RNA genome is reverse transcribed into double strand DNA by the virus-encoded reverse transcriptase and integrated into the host genome. The proviral form is maintained and propagated indefinitely, and its transcription provides proper expression of structural and regulatory proteins. Since the retroviral particles bear a plus strand RNA template, full length transcription of the proviral genome provides at the same time also the replicative form of the virus. These notions highlight the fundamental role of transcription in the control of viral replication. Furthermore, several reports published in the last few years indicate that replication of HIV-1 is occurring during the whole course of infection (Bagnarelli *et al.*, 1992; Ho *et al.*, 1995; Michael *et al.*, 1992; Pantaleo *et al.*, 1993; Piatak *et al.*, 1993; Wei *et al.*, 1995). However, there are clear evidences indicating that progression of the disease is concomitant with an increase in viral load in patients. This conclusion derives from quantitative molecular studies on the amount of infectious virus, viral antigens and virus-specific nucleic acids in patients at different stages of disease (Bagnarelli *et al.*, 1992; Connor *et al.*, 1993; Gupta *et al.*, 1990; Gupta *et al.*, 1993; Ho *et al.*, 1989; Jurriaans *et al.*, 1992; Michael *et al.*, 1992; Schnittman *et al.*, 1991; Schnittman *et al.*, 1990; Simmonds *et al.*, 1990). In particular, the transcriptional rate, measured as the ratio between the amount of viral mRNAs and proviral DNA in peripheral blood mononuclear cells of infected individuals, is increased in patients at later stages of disease (Bagnarelli *et al.*, 1992). Additionally, there is a strong correlation between HIV replication in these cells and the future clinical course of HIV infection, often independently of CD4 counts and clinical staging (Saksela *et al.*, 1994). Furthermore, an extraordinarily large number of latently infected CD4+ lymphocytes and macrophages are detectable throughout the lymphoid system from early to late stages of infection (Embretson *et al.*, 1993). Altogether, these observations indicate that the study of the

molecular mechanisms controlling the rate of transcription from the HIV-1 provirus is not only relevant for understanding the biological mechanisms of transcriptional control of eukaryotic genes, but could also offer useful insights into the mechanisms of AIDS development.

Once integrated, the proviral genome can be considered as a cellular gene, its expression being tightly regulated by both cellular and viral factors.

As reported in the introduction section, more than 20 different cellular factors were shown to bind to the relatively short (~400 bp) LTR U3 region. Some of the binding sites for these factors, indeed, were shown to be engaged in protein-DNA interactions also in vivo, at least in the highly expressing H9/HIV_{IIIB} cell line (Demarchi *et al.*, 1992). Most of these protein binding sites of the LTR are recognized by nuclear proteins already known to physiologically control the expression of a variety of cellular genes, indicating that this viral promoter represents a very interesting biological example of evolutionary tinkering. Why such a complex regulation is needed in the HIV-1 life cycle is still not clear. In this respect, it should be observed that, with the exception of the arrangement of the basal promoter (initiator-TATA box-Sp1 sites) and the enhancer region, the contribution of any of the other sites to transcriptional activation is apparently weak and none of these sites appear to be necessarily required for the production of infectious virus. It is conceivable that the reason for such a redundancy of transcription factor binding sites reflects the need either for responsiveness to a variety of cellular environments, or for fine tuning of viral transcription in response to a variety of cellular events. Both of these demands are hardly reproducible in in vitro experiments.

The results of an in vivo footprinting assay, showed that a region at the proximal end of the NRE, centered on the E-box CACGTG palindromic sequence, is engaged in protein-DNA interaction (Demarchi *et al.*, 1992). This sequence appears to be strictly conserved, despite the overall HIV-1 sequence variability in different HIV-1 infected patients and during HIV infection progression over several years (Michael *et al.*, 1994). These observations, as well as the evolutionary conservation

of the E-box motif and of the cognate binding proteins in the upstream sequences of several cellular genes (Giacca *et al.*, 1989; Giacca *et al.*, 1992) suggest that this sequence is likely to play a role in the context of the control of HIV-1 expression.

Interactions of USF with the LTR E box target site

In the past we could demonstrate that an oligonucleotide containing the HIV-1 LTR E-box (Giacca *et al.*, 1992a) was able to bind specifically to a nuclear protein in gel retardation assays. In DNase I footprinting experiments, using crude nuclear extracts, the same region was protected. Several nuclear proteins are capable of specifically binding to the CACPuTG sequence and, in the last years, the cDNA sequences of some proteins of the b-HLH-Zip family interacting with this E box were cloned. After extensive purification, the major DNA-binding activity detectable by gel retardation assays turned out to be the USF transcription factor as demonstrated by its binding specificity, its biochemical properties, its microsequence and its immunological reactivity (Csordás Tóth *et al.*, 1993). USF is a constitutive transcription factor composed of two polypeptides of 43 and 44 kDa relative molecular weight, encoded by two different genes (Gregor *et al.*, 1990; Sirito *et al.*, 1992).

The 43- and 44-kDa polypeptides preferentially bind as a heterodimer to the target site (Ferré-D'Amaré *et al.*, 1994), even if each one is individually able to bind to DNA (Pognonec and Roeder, 1991; Sirito *et al.*, 1992), as demonstrated also by south-western analysis (Csordás Tóth *et al.*, 1993). We exploited the available cDNA sequence for the 43-kDa form of USF to produce a recombinant fusion USF^{43} protein in bacteria. Both the purified ($USF^{43/44}$) and the recombinant (USF^{43}) proteins show the same binding specificity for the HIV-1 LTR and other targets containing an E-box sequence. As expected, USF^{43} binding is affected in the same way by the mutations that abolish DNA binding of the eukaryotic protein.

With analogy to the interactions occurring at the MLP upstream element (Miyamoto *et al.*, 1985; Sawadogo and Roeder, 1985b), the G residues which exhibit

decreased methylation in the presence of USF are symmetrically situated in the two halves of the region of dyad symmetry. This symmetry is consistent with the proposed binding of the factor as a dimer to the target sequence, with each monomer contacting the DNA sequence on the 3' site of the dyad axis of the E box element (Ferré-D'Amaré *et al.*, 1994). It is likely that the protein dimer binds at the centre of the palindromic site with the basic domains extending symmetrically into the major grooves of each half site, as suggested (Fisher *et al.*, 1992).

Similar to the sensitivity of the native factor (Giacca *et al.*, 1989; Watt and Molloy, 1988), binding of recombinant USF^{43} to the LTR is abolished by methylation of the core CpG dinucleotide of the E box. This observation suggests that a possible mechanism of control for the function of USF (which is constitutively and ubiquitously expressed (Sirito *et al.*, 1994)) could be through epigenetic modification of the target site, namely by altering the methylation state of DNA. Since it has been shown that methylation of the LTR represses HIV-1 transcription (Bednarik *et al.*, 1990), it is likely that the E box could be one of the target sites mediating this effect. Finally it is interesting to observe that the mechanism for preventing binding by target site methylation is not restricted to USF, since binding by other members of the b-HLH-Zip family is also sensitive to methylation of the core CpG sequence of the E box (Prendergast and Ziff, 1991).

Role of template LTR bending upon USF binding

The circular permutation analyses performed, indicate that both USF purified from HeLa cells and recombinant USF^{43} bend DNA upon binding to the target site. It is likely that the sites of hypersensitivity to DNA methylation and DNase I digestion detected in the methylation protection and DNase I footprinting experiments are generated by an increase in the groove width, typical of DNA bending upon protein binding. Although circular permutation assay alone does not allow the distinction between DNA bending and increased DNA flexibility, the detected extent of bending (90°-110°) cannot be due only to a mere increase in flexibility. DNA bending by a

number of b-HLH-Zip proteins (including USF and Myc/Max) has been reported (Fisher *et al.*, 1992; Wechsler and Dang, 1992). However, crystal analysis of the complex of DNA with Max (Ferré-D'Amaré *et al.*, 1993) and with the USF b-HLH-Zip regions (Ferré-D'Amaré *et al.*, 1994) could not show any net bend in the double helical axis. The reasons for this discrepancy are likely to be related to the crystal packing mode of the two structures.

In our experiments, the calculated bending angles induced by HeLa-purified USF and recombinant USF^{43} were 90° and 110°, respectively. A trivial explanation for this difference could be that recombinant USF^{43}, bearing the GST extension, has a different shape and a greater size than HeLa-purified USF. However, it was demonstrated that there is no significant correlation between the molecular weight of the protein and the extent of the induced DNA bending (Kerppola and Curran, 1991). Therefore, our preferred interpretation is that the difference in the bending ability is due to the different molecular compositions of the dimers interacting with DNA: an obligate USF^{43} homodimer in the experiments with the recombinant protein and a preferred $USF^{43/44}$ heterodimer in the experiments with the purified factor. In the latter case, the bending angle results from the vectorial sum of the bending angles induced by the two different monomers. In this respect, there are also other examples of differential bending by monomeric or heteromeric forms such as Myc/Max and Fos/Jun heterodimers (Kerppola and Curran, 1991; Wechsler and Dang, 1992).

In the context of HIV-1 transcriptional regulation, the studies on the functional role of LTR DNA bending by transcriptional factors probably deserve further scrutiny. In fact, it should be considered that, in addition to USF, other transcription factors that also bind to the LTR (Jun/Fos (Kerppola and Curran, 1991), NF-κB (Schreck *et al.*, 1990), TBP (Horikoshi *et al.*, 1988), YY1 (Margolis *et al.*, 1994; Natesan and Gilman, 1993), Sp1 (Ikeda *et al.*, 1993)) are able to bend the target DNA sequence. As a consequence, the traditional picture of the LTR as a linear structure should be replaced by a more realistic 3D-view of a highly structured promoter where also non

adjacent proteins, by DNA bending and looping, can interact among themselves and with the basal transcriptional machinery. According to this view, it has been reported that Sp1 interacts with NF-κB (Perkins *et al.*, 1993), that both Sp1 and Tat bind to TBP (Emili *et al.*, 1994; Kashanchi *et al.*, 1994b), that YY1 physically interacts with Sp1 (Lee *et al.*, 1993), and that Fos is able to bind to the 44 kDa form of USF (Blanar and Rutter, 1992). The role of such interactions has also been demonstrated by showing that the binding of NF-kB is not by itself sufficient to induce HIV gene expression. Instead, an interaction between NF-κB and the transactivating domain of Sp1 bound to an adjacent site must occur (Majello *et al.*, 1994). In this light, small peptides (HMG proteins) have shown to work as adaptor that favors DNA bending and protein interactions (Grosschedl *et al.*, 1994). Preliminary data suggest that HMG proteins do indeed affect the transcriptional rate of HIV-1 both in vivo and in vitro (d'adda di Fagagna, unpublished results).

Probably, the more convincing proof that DNA bending has a role in transcription came from the work of Fisher and collaborators (Parvin *et al.*, 1995). These authors show that pre-bending of a promoter sequence enhances 100 folds the affinity of the TATA-binding factor for its target sequence. This suggests that transcription factors have a dual mechanism for transcriptional control, namely by recruiting the RNA polymerase II machinery and by increasing the affinity of the promoter for the essential TATA-binding factor.

In addition to the role of DNA bending in favoring these interactions, it has also been suggested that the energy stored in a protein-induced bend itself could be used to promote the formation of an open transcription complex (Van der Vliet and Verrijzer, 1993).

It is also interesting to observe that the b-HLH-Zip domain of USF was reported to exist as a bivalent tetramer, potentially able to bind simultaneously to two independent sites, with a possible role in DNA looping (Ferré-D'Amaré *et al.*, 1994). However, despite reported evidence showing interactions between USF and the initiator element of the HIV-1 LTR (Du *et al.*, 1993; Roy *et al.*, 1991), we were not

able to obtain a clear footprint on this region, even with the addition of an amount of protein as high as 100 fold as that needed to footprint over the E box. Accordingly, the analysis of the USF binding sites by the random oligonucleotide selection procedure showed that the E box motif is almost absolutely required for binding (Bendall and Molloy, 1994).

Functional role of USF binding to the LTR

In vitro transcription experiments, performed either by the addition of recombinant USF^{43}, or by the subtraction of endogenous E box binding proteins by E box decoys, indicate that USF acts as a positive regulator of transcription driven by the LTR promoter. The in vivo results mirror those obtained in vitro: a decoy plasmid for USF downregulates LTR-driven transcription (not shown); USF^{43} overexpression upregulates CAT transcription when controlled by a wild type LTR; the mutation of the CACGTG sequence of the LTR impairs USF transactivation, even if it does not completely abolish it. This residual effect could rely on some interactions of the initiator elements (Roy *et al.*, 1991) (in the previous in vitro experiments, these elements could not be taken into account, since the LTR region downstream of transcription start site has been replaced by a G-less cassette). Even if in our hands the initiator elements of HIV-1 could not be recognized by purified recombinant USF^{43}, it is still possible that in vivo USF^{43} can interact with some other factor in order to gain a higher affinity toward the initiator element. TFII-I (Roy *et al.*, 1991) is the most likely candidate for this function. In this regard, we have shown that also other proteins can modulate binding of USF to its target site (Zijderveld *et al.*, 1994).

The inability of USF^{43} to activate the LTR in the presence of the Tat protein was unexpected. This observation is in agreement with the data reported by Moses et al., showing that the mutation of the LTR E-box significantly reduces LTR activity in both stimulated and unstimulated primary human macrophages, but only in the absence of Tat (Moses *et al.*, 1994). Furthermore, a linker scanning mutagenesis analysis of the LTR in the Jurkat T-cell lines showed that the effect of the E-box

deletion is more pronounced in the absence of Tat (Zeichner *et al.*, 1991). Even if more data are needed for a better understanding of this effect, some hypotheses can be formulated. It is unlikely that these findings can be attributed to the inactivation of USF by the oxidative intracellular environment induced by the Tat protein by the repression of the expression of the manganese superoxide dismutase gene (Flores *et al.*, 1993). Although the factor is highly sensitive to its redox state (Pognonec *et al.*, 1992), it is still functional in Tat expressing cells, since it can positively affect expression from the minimal USF-responsive promoter used in our control experiments. It is more likely that the lack of transactivation of the LTR in the presence of Tat could be due to an alteration of the architecture of the LTR itself induced by Tat, that renders the promoter unresponsive to USF. The presence of the Tat protein is expected to induce several changes at the downstream LTR region (including the TATA box and the Sp1 sites) by establishing protein-protein interactions. Furthermore, Tat directly activates nuclear translocation of NF-κB (our unpublished observation). These changes could unfavour either the ability of USF to bind to the E-box, or its ability to activate transcriptions, or both. In particular, there are some indirect observations in the literature that suggest the possibility of structural interactions between the USF and κB sites (Garcia *et al.*, 1987; Maekawa *et al.*, 1991). In the future, we will try to validate this hypothesis by in vivo footprinting experiments of the LTR in the presence or absence of the Tat protein. A correlate of this hypothesis is the prediction that USF will not be able to transactivate the LTR when cells are activated by PMA (an inducer of NF-κB) and, conversely, that USF will still be active in the presence of Tat if the κB sites have been deleted.

It is also noteworthy that USF can play an additional role on the expression of HIV-1. Michael (Michael *et al.*, 1994) and colleagues surprisingly demonstrated that also the 3' LTR has a transcriptional role, promoting the synthesis of a negative strand RNA , antisense in respect to the canonical transcripts. This is very interesting because this transcript bears a open reading frame conserved among the strains and also provide an antisense regulation of the translation of the positive strand encoded

proteins. USF binding site has been demonstrated to be absolutely required for the transcription by this negative strand promoter. We are collaborating with this group to study the details of the role of USF in the control of this promoter.

Open questions about the function of Tat protein

In addition to the understanding of the role of USF in the control of HIV-1 mediated transcription, in this thesis we have started to address the problem of obtaining a highly active recombinant Tat protein. This reagent will be used to highlight some of the mechanisms of Tat function, with particular reference to the interplay of the protein with the cellular transcription factors at the LTR promoter. For this purpose, we have obtained two GST-fusion proteins, containing only the first or both Tat exons from strain HIV_{HXBII}. Despite the reported difficulty in maintaining the functional activity of chemically synthesized Tat and of Tat purified from E. coli due to its tendency to oxidate and form aggregates (the protein is difficult to overexpress and purify due to its tendency to aggregate (Rhim *et al.*, 1993; Fawell *et al.*, 1994; Frankel *et al.*, 1989; Harper and Logsdon, 1991; Koken *et al.*, 1994b; Rhim *et al.*, 1993; Rice and Carlotti, 1990b; Rice and Chan, 1991; Song *et al.*, 1994)), we have not experienced such problems with our fusion products. It is conceivable the GST moiety, while not interfering with the functional activity of protein, does protect Tat from being easily oxidized. Although some biological differences have been described between the function of one-exon or two-exon Tat, as already reported in the Introduction, in our experience the two proteins perform identically in their ability of transactivation of the LTR promoter both in vitro and in vivo.

The development of these recombinant proteins is of particular interest given the property of Tat of crossing cell membranes. As a consequence, these proteins can be used by addition to the cell medium in the presence of chloroquine or by lipofection and avoid the use Tat expressing plasmids.

There are a variety of open questions that still need to be resolved about Tat function, in which these recombinant proteins will be hopefully useful. In particular:

1) While it is conceivable that the activation domain of Tat could act similarly to the activation domains of other transcription factors in augmenting the rate of transcription initiation by the interaction with the transcriptional machinery, Tat remains unique since it is the only factor brought to the promoter by RNA-protein interaction instead of DNA-protein interaction. Whether it represents the prototype of a novel type of transcription factors, it still remains to be understood. In addition, there is a very striking difference between the transactivation efficiency of the protein in in vivo and in vitro experiments: while Tat is an extremely effective activator in vivo, where it increases LTR-mediated expression of several thousand folds, it is relatively weak in vitro, where it augments expression less then ten-fold. Conversely, the LTR promoter in the absence of Tat is extremely weak in vivo, while it behaves as a strong promoter in in vitro transcription assays. The molecular correlates of these observations still need to be clarified.

2) Several lines of evidence clearly demonstrated that one of the major actions of Tat on HIV gene expression is to increase the rate of elongation of nascent RNAs. However, the actual molecular mechanisms of how this effect is exerted still remain to be understood. It has been proposed that Tat mimics a cellular protein that, by associating with a transcription complex, increases its ability to elongate (Lu *et al.*, 1993). Since phosphorylation of the carboxyl-terminal domain of RNA-polymerase II has been proposed to trigger the transition from initiation to active elongation and also to influence later stages during elongation (Dahmus, 1995), particular importance could be attributed to the identification of protein kinases as Tat cofactors. Among the potential candidates for this activity, both the TAK kinase (Herrmann and Rice, 1995) and DNA-PK (Dahmus, 1995) are able to phosphorylate the CTD domain of RNA pol II. Whether this occurs also in vivo, it remains to be established. Moreover, the effects of Tat on the transcriptional machinery could be much more complex than the phosphorylation of the C-terminal domains of RNA polymerase II, including the potential interactions with basal transcription factors such as TFIIE, TFIIF, TFIIH, and TFIIJ (Flores *et al.*, 1992) (Weinmann, 1992).

3) Several in vitro studies showed that Tat is released from expressing cells and can be taken up by neighboring cells, where it can enter the nucleus and trigger LTR-driven gene expression. There are still lack of molecular details on how this activity occurs. In particular, it is not clear whether the function of Tat on adjacent cells is due to the actual localization of Tat in the nucleus with consequent transcriptional activation, or can be mediated by the interactions with membrane receptor. Low affinity receptors of the integrin-type have been shown to bind to Tat and trigger activation of a signal transduction pathway (Barillari *et al.*, 1993; Brake *et al.*, 1990). Whether high affinity receptors exist as suggested (Weeks *et al.*, 1993), they still need to be fully characterized.

4) Due to the very low levels of Tat production in infected cells, and to the overall limited amount of these cells in infected individuals, the question whether Tat release could be actually involved in the pathogenesis of HIV disease is still completely unaddressed in vivo. There are several effects where Tat release could play a major pathogenetic role that goes beyond the virus infected cell. Among these effects, there are the induction of immunological anergy (Viscidi *et al.*, 1989) or the triggering of apoptosis (Li *et al.*, 1995) in uninfected bystander cells of the immune system or of the nervous system, with the induction of immunosuppression and of neurological disturbances respectively. Both of these events are hallmarks of HIV disease.

Phosphorylation of USF[43] and Tat by DNA-PK

The DNA dependent protein kinase (DNA-PK) is a protein kinase which becomes activated once bound to DNA (Gottlieb and Jackson, 1993). DNA binding relies on the ability of the Ku heterodimeric protein, a component of DNA-PK, to recognize discontinuities in the DNA double helix. Ku recognizes mainly DNA ends (double strand breaks in the genome) but also nicks, small gaps and dumpbell structures. Following DNA binding, Ku triggers the kinase activity of the catalytic subunit of DNA-PK (DNA-PKcs). This, in turn, can phosphorylate serines and threonines followed by glutamines, being Ser/Thr-Gln its consensus for

phosphorylation (Bannister *et al.*, 1993). Proteins binding in cis on the same DNA molecule are preferentially phosphorylated. Many DNA binding factors and non-DNA binding proteins have been demonstrated in vitro to be modified by DNA-PK (Anderson, 1993). Usually, phosphorylation sites lies in the transactivating domain of transcription factors. To date a clear function of such a modification has not been described, since the mutation of modified residues to a different amino acid (usually alanine) has not resulted in dramatic effects. In addition to these specific observations, a more general role for DNA-PK in the control of transcription has been described. In fact, in in vitro transcription experiments using a linearized template, DNA-PK has been shown to inhibit transcription by RNA polymerase I (Kuhn *et al.*, 1995) and RNA pol II and III (S. Jackson, personal communication). This activity relies on the phosphorylating ability of DNA-PK, since DNA-PK inhibitors relieve the inhibition.

On the basis of the primary amino acidic sequence, both USF^{43} and Tat are potential substrates for DNA-PK phosphorylation. Our results show that, at least in vitro, these proteins are indeed modified by DNA-PK.

Tat phosphorylation sites were precisely mapped by the use of a panel of 20 aa peptides. Two of the three potential phosphorylation sites are phosphorylated in vitro. These sites are located in the transactivating domain (Ser 16) and basic domain (Ser 62); both of these regions are essential for Tat functions and very conserved. Mutational analysis is currently under way in order to detect the phenotype of Ser->Ala 16 and Ser->Ala 62 Tat mutants in vivo.

On the basis of its amino acidic sequence, also USF^{43} phosphorylation by DNA-PK could modify amino acids involved either in DNA binding or transactivation. We are currently addressing the problem of the recognition of the actual phosphorylation sites by triptic mapping. Subsequently, mutagenesis of these residues will demonstrate the effective role of such a modification, in vivo.

More generally, the role of USF phosphorylation has not been studied thoroughly so far. Maekawa and collaborators reported that phosphatase treatment of

HeLa purified USF abolishes DNA binding (Maekawa *et al.* 1991) and we have repeated the same experiment obtaining identical results (data not shown). Further work is obviously needed for a better understanding of these results. Finally, the observation that Ku specifically recognizes stem-loop structures and that, in particular, it binds to TAR RNA (Kaczmarski and Khan, 1993) raises the puzzling possibility that DNA-PK could be recruited specifically at the LTR promoter. As a matter of fact, several transcription factors engaged in protein-DNA interactions at the LTR are substrates for DNA-PK phosphorylation, (Anderson, 1993). Furthermore, DNA-PK phosphorylates the CTD domain of RNA pol II (Dvir *et al.*, 1993), an event resulting in increased processivity of the polymerase (which, as noted above, is one of the major functions of Tat activity).

Altogether, these observations are intriguing. Even if much speculation should be avoided before further data are collected, the possibility that this kinase could play a role in the control of LTR-directed gene expression currently appears to be an exciting topic of investigation.

ACKNOWLEDGMENTS

This is to acknowledge with thanks the collaboration of my colleagues at ICGEB and, in particular, Dr. P. Taneja, Dr. M.I. Gutierrez and Dr. G. Marzio with whom I worked closely. I am indebted to Dr. M. Giacca for the encouragement and fruitful discussions. I would also like to thank Prof. A. Falaschi for having provided me with the opportunity to work at ICGEB and for having set up such a stimulating environment.

REFERENCES

Adams, M., Sharmeen, L., Kimpton, J., Romeo, J.M., Garcia, J.V., Peterlin, B.M., Groudine, M. and Emerman, M. (1994) Cellular latency in human immunodeficiency virus-infected individuals with high CD4 levels can be detected by the presence of promoter-proximal transcripts. *Proc. Natl. Acad. Sci. USA*, **91**, 3862-3866.

Alonso, A., Derse, D. and Peterlin, B.M. (1992) Human chromosome 12 is required for optimal interactions between Tat and TAR of human immunodeficiency virus type 1 in rodent cells. *J Virol*, **66**, 4617-21.

Anderson, C.W. (1993) DNA damage and the DNA-activated protein kinase. *Trends Biochem Sci*, **18**, 433-7.

Arya, S.K., Guo, C., Josephs, S.F. and Wong-Staal, F. (1985) *Trans*-Activator Gene of Human T-Lymphotropic Virus Type III (HTLV-III). *Science*, **229**, 69-73.

Ayer, D.E., Kretzner, L. and Eisenman, R.N. (1993) Mad: a heterodimeric partner for Max that antagonizes Myc transcriptional activity. *Cell*, **72**, 211-222.

Baeuerle, P.A. and Baltimore, D. (1988) IκB: a specific inhibitor of the NF-κB transcription factor. *Science*, **242**, 540-546.

Bagnarelli, P., Menzo, S., Valenza, A., Manzin, A., Giacca, M., Ancarani, F., Scalise, G., Varaldo, P.E. and Clementi, M. (1992) Molecular profile of human immunodeficiency virus infection in asymptomatic subjects and AIDS patients. *J. Virol.*, **66**, 7328-7335.

Baldwin, A.S., Jr., LeClaire, K.P., Singh, H. and Sharp, P.A. (1990) A large protein containing zinc finger domains binds to related sequence elements in the enhancers of the class I major histocompatibility complex and κ immunoglobulin genes. *Mol. Cell. Biol.*, **10**, 1406-1414.

Baldwin, A.S.J. and Sharp, P.A. (1988) Two transcriptions factors, NF-κB and H2TF1, interact with a single regulatory sequence in the class I major histocompatibility complex promoter. *Proc. Natl. Acad. Sci. USA*, **85**, 723-727.

Ballard, D.W., Walker, W.H., Doerre, S., Sista, P., Molitar, J.A., Dixon, E.P., Paffer, N.J., Hannink, M. and Greene, W.C. (1990) The v-rel oncogene encodes a κB enhancer binding protein that inhibits NF-κB function. *Cell*, **63**, 803-814.

Bannister, A.J., Gottlieb, T.M., Kouzarides, T. and Jackson , S.P. (1993) c-Jun is phosphorylated by the DNA-dependent protein kinase in vitro, definition of the minimal kinase recognition motif. *Nucl. Acids. Res.*, **21**, 1289-1295.

Barillari, G., Gendelman, R., Gallo, R.C. and Ensoli, B. (1993) The Tat protein of human immunodeficiency virus type 1, a growth factor for AIDS Kaposi sarcoma and cytokine-activated vascular cells, induces adhesion of the same cell types by using integrin receptors recognizing the RGD amino acid sequence. *Proc Natl Acad Sci U S A*, **90**, 7941-5.

Barre-Sinoussi, T., Chermann, J.C., Reye, F., Nugeure, M.T., Chamaret, S., Gruest, J., Dauquet, C., Alxer-Blin, C., Vezinet-Brun, F., Rouzious, C., ., Rosenbaum, W. and Montagnier, L. (1983) Isolation of a T-lymphotrophic retrovirus from a patient at risk for acquired immunodeficiency syndrome. *Science*, **220**, 868-871.

Beckmann, H., Su, L.-K. and Kadesch, T. (1990) TFE3: a helix-loop-helix protein that activates transcription through the immunoglobulin enhancer μE3 motif. *Genes Dev.*, **4**, 167-179.

Bednarik, D.P., Cook, J.A. and Pitha, P.M. (1990) Inactivation of the HIV LTR by DNA CpG methylation: evidence for a role in latency. *EMBO J.*, **9**, 1157-1164.

Bendall, A.J. and Molloy, P.L. (1994) Base preferences for DNA binding by the bHLH-Zip protein USF: effects of $MgCl_2$ on specificity and comparison with binding of Myc family members. *Nucleic Acids Res.*, **22**, 2801-2810.

Benjouad, A., Mabrouk, K., Moulard, M., Gluckman, J.C., Rochat, H., Van Rietschoten, J. and S abatier, J.M. (1993) Cytotoxic effect on lymphocytes of Tat from human immunodeficiency virus (HIV-1). *FEBS Lett*, **319**, 119-24.

Bernath, V.A., Muro, A.F., Vitullo, A.D., Bley, M.A., Baranao, J.L. and Kornblihtt, A.R. (1990) Cyclic AMP inhibits fibronectin gene expression in a newly developed granulosa cell line by a mechanisms that suppresses c-AMP-responsive element-dependent transcriptional activation. *J. Biol. Chem.*, **265**, 18219-18226.

Biamonti, G., Giacca, M., Perini, G., Contreas, G., Zentilin, L., Weighardt, F., Guerra, M., Della Valle, G., Saccone, S., Riva, S. and Falaschi, A. (1992a) The gene for a novel human lamin maps at a highly transcribed locus of chromosome 19 which replicates at the onset of S-phase. *Mol. Cell. Biol.*, **12**, 3499-3506.

Biamonti, G., Perini, G., Weighardt, F., Riva, S., Giacca, M., Norio, P., Zentilin, L., Diviacco, S., Dimitrova, D. and Falaschi, A. (1992b) A human DNA replication origin: localization and transcriptional characterization. *Chromosoma*, **102**, S24-S31.

Blackwood, E.M. and Eisenman, R.N. (1991) Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA binding complex with Myc. *Science*, **251**, 1211-1217.

Blanar, M.A. and Rutter, W.J. (1992) Interaction cloning - identification of a Helix-Loop-Helix zipper protein that interacts with c-Fos. *Science*, **256**, 1014-1018.

Blunt, T., Finnie, N.J., Taccioli, G.E., Smith, G.C., Demengeot, J., Gottlieb, T.M., Mizuta, R., Varghese, A.J., Alt, F.W., Jeggo, P.A. and Jackson, S.P. (1995) Defective DNA-dependent protein kinase activity is linked to V(D)J recombination and DNA repair defects associated with the murine scid mutation. *Cell*, **80**, 813-823.

Bogerd, H.P., Fridell, R.A., Blair, W.S. and Cullen, B.R. (1993) Genetic evidence that the Tat proteins of human immunodeficiency virus types 1 and 2 can multimerize in the eukaryotic cell nucleus. *J Virol*, **67**, 5030-4.

Bohan, C.A., Kashanchi, F., Ensoli, B., Buonaguro, L., Boris-Lawrie, K.A. and Brady, J.N. (1992) Analysis of Tat transactivation of human immunodeficiency virus transcription in vitro. *Gene Expr*, **2**, 391-407.

Bours, V., Villalobos, J., Burd, P.R., Kelly, K. and Siebenlist, U. (1990) Cloning of a mitogen-inducible gene encoding a kappa B DNA-binding protein with homology to the rel oncogene and to cell-cycle motifs. *Nature*, **348**, 76-80.

Braddock, M., Thorburn, A.M., Chambers, A., Elliott, G.D., Anderson, G.J., Kingsman, A.J. and K ingsman, S.M. (1990) A nuclear translational block imposed by the HIV-1 U3 region is relieved by the Tat-TAR interaction. *Cell*, **62**, 1123-33.

Braddock, M., Thorburn, A.M., Kingsman, A.J. and Kingsman, S.M. (1991) Blocking of Tat-dependent HIV-1 RNA modification by an inhibitor of RNA polymerase II processivity. *Nature*, **350**, 439-41.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantyties of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*, **72**, 248-254.

Brake, D.A., Debouck, C. and Biesecker, G. (1990) Identification of an Arg-Gly-Asp (RGD) cell adhesion site in human immunodeficiency virus type 1 transactivation protein, tat. *J Cell Biol*, **111**, 1275-81.

Bresnick, E.H. and Felsenfeld, G. (1993) Evidence that the transcription factor USF is a component of the human beta- globin locus control region heteromeric protein complex. *J. Biol. Chem.*, **268**, 18824-18834.

Broder, C.C., Nussbaum, O., Gutheil, W.G., Bachovchin, W.W. and Berger, E.A. (1994) CD26 antigen and HIV fusion? [letter; comment]. *Science*, **264**, 1156-9; discussion 1162-5.

Buonaguro, L., Barillari, G., Chang, H., Bohan, C.A., Kao, V., Morgan, R., Gallo, R.C. and Ensoli, B. (1992) Effects of human immunodeficiency virus type 1 Tat protein on the expression of inflammatory cytokines. *J. Virol.*, **66**, 7159-7167.

Callebaut, C., Krust, B., Jacotot, E. and Hovanessian, A.G. (1993) T cell activation antigen, CD26, as a cofactor for entry of HIV in CD4+ cells [see comments]. *Science*, **262**, 2045-50.

Calnan, B.J., Biancalana, S., Hudson, D. and Frankel, A.D. (1991) Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA-protein recognition. *Genes Dev*, **5**, 201-10.

Carr, C.S. and Sharp, P.A. (1990) A helix-loop-helix protein related to the immunoglobulin E box-binding proteins. *Mol. Cell. Biol.*, **10**, 4384-4388.

Carroll, R., Martarano, L. and Derse, D. (1991) Identification of lentivirus tat functional domains through generation of equine infectious anemia virus/human immunodeficiency virus type 1 tat gene chimeras. *J Virol*, **65**, 3460-7.

Chirmule, N., Than, S., Khan, S.A. and Pahwa, S. (1995) Human immunodeficiency virus Tat induces functional unresponsiveness in T cells. *J Virol*, **69**, 492-8.

Cogswell, J.P., Godlevski, M.M., Bonham, M., Bisi, J. and Babiss, L. (1995) Upstream stimulatory factor regulates expression of the cell cycle- dependent cyclin B1 gene promoter. *Mol Cell Biol*, **15**, 2782-90.

Comar, M., Marzio, G., D'Agaro, P. and Giacca, M. (1995) Quantitative dynamics of HIV-1 expression. *AIDS Res. Hum. Retrov.*, in press.

Connor, R.I., Mohri, H., Cao, Y. and Ho, D.D. (1993) Increased viral burden and cytopathicity correlate temporally with CD4+ T- lymphocyte decline and clinical progression in human immunodeficiency virus type 1-infected individuals. *J. Virol.*, **67**, 1772-1777.

Cooney, A.J., Tsai, S.Y., O'Malley, B.W. and Tsai, M.J. (1991) Chicken ovalbumin upstream promoter transcription factor binds to a negative regulatory region in the human immunodeficiency virus type 1 long terminal repeat. *J. Virol.*, **65**, 2853-2860.

Cordingley, M.G., LaFemina, R.L., Callahan, P.L., Condra, J.H., Sardana, V.V., Graham, D.J., Nguyen, T.M., LeGrow, K., Gotlib, L., Schlabach, A.J. and al., e. (1990) Sequence-specific interaction of Tat protein and Tat peptides with the transactivation-responsive sequence element of human immunodeficiency virus type 1 in vitro. *Proc Natl Acad Sci U S A*, **87**, 8985-9.

Crabtree, G.R. (1989) Contingent genetic regulatory events in T lymphocyte activation. *Science*, **243**, 355-361.

Csordás Tóth, E., Marusic, L., Ochem, A., Patthy, A., Pongor, S., Giacca, M. and Falaschi, A. (1993) Interactions of USF and Ku antigen with a human DNA region containing a replication origin. *Nucleic Acids Res.*, **21**, 3257-3263.

Cullen, B.R. and Greene, W.C. (1989) Regulatory pathways governing HIV-1 replication. *Cell*, **58**, 423-426.

Cupp, C., Taylor, J.P., Khalili, K. and Amini, S. (1993) Evidence for stimulation of the transforming growth factor beta 1 promoter by HIV-1 Tat in cells derived from CNS. *Oncogene*, **8**, 2231-6.

Cvekl, A., Sax, C.M., Bresnick, E.H. and Piatigorsky, J. (1994) A complex array of positive and negative elements regulates the chicken alpha A-crystallin gene: involvement of Pax-6, USF, CREB and/or CREM, and AP-1 proteins. *Mol Cell Biol*, **14**, 7363-76.

Dahmus, M.E. (1995) Phosphorylation of the C-terminal domain of RNA polymerase II. *Biochim Biophys Acta*, **1261**, 171-82.

Dang, C.V., Dolde, C., Gillison, M.L. and Kato, G.J. (1992) Discrimination between related DNA sites by a single amino acid residue of Myc-related basic-helix-loop-helix proteins. *Proc. Natl. Acad. Sci. USA*, **89**, 599-602.

Dasgupta, P., Saikumar, P., Reddy, C.D. and Reddy, E.P. (1990) Myb protein binds to human immunodeficiency virus 1 long terminal repeat (LTR) sequences and transactivates LTR-mediated transcription. *Proc. Natl. Acad. Sci. USA*, **87**, 8090-8094.

Datta, P.K., Ghosh, A.K. and Jacob, S.T. (1995) The RNA polymerase I promoter-activating factor CPBF is functionally and immunologically related to the basic helix-loop-helix-zipper DNA-binding protein USF. *J Biol Chem*, **270**, 8637-41.

Demarchi, F., D'Agaro, P., Falaschi, A. and Giacca, M. (1992) Probing protein-DNA interactions at the long terminal repeat of human immunodeficiency virus type 1 by in vivo footprinting. *J. Virol.*, **66**, 2514-2518.

Demarchi, F., D'Agaro, P., Falaschi, A. and Giacca, M. (1993) In vivo footprinting analysis of constitutive and inducible protein-DNA interactions at the long terminal repeat of human immunodeficiency virus type 1. *J. Virol.*, **67**, 7450-7460.

Dingwall, C., Ernberg, I., Gait, M.J., Green, S.M., Heaphy, S., Karn, J., Lowe, A.D., Singh, M. and S kinner, M.A. (1990) HIV-1 tat protein stimulates transcription by binding to a U-rich bulge in the stem of the TAR RNA structure. *Embo J*, **9**, 4145-53.

Dinter, H., Chiu, R., Imagawa, M., Karin, M. and Jones, K.A. (1987) *In vitro* activation of the HIV-1 enhancer in extracts from cells treated with a phorbol ester tumor promoter. *EMBO J.*, **6**, 4067-4071.

Du, H., Roy, A.L. and Roeder, R.G. (1993) Human transcription factor USF stimulates transcription through the initiator elements of the HIV-1 and the Ad-ML promoters. *EMBO J.*, **12**, 501-511.

Duh, E.J., Maury, W.J., Folks, T.M., Fauci, A.S. and Rabson, A.B. (1989) Tumor necrosis factor α activates human immunodeficiency virus type 1 through induction of nuclear factor binding to the NF-κB sites in the long terminal repeat. *Proc. Natl. Acad. Sci. USA*, **86**, 5974-5978.

Dvir, A., Stein, L.Y., Calore, B.L. and Dynan, W.S. (1993) Purification and characterization of a template-associated protein kinase that phosphorylates RNA polymerase II. *J. Biol. Chem.*, **268**, 10440-10447.

Edenberg, H.J. and Brown, C.J. (1992) Regulation of human alcohol dehydrogenase genes. *Pharmacogenetics*, **2**, 185-196.

Embretson, J., Zupancic, M., Ribas, J.L., Burke, A., Racz, P., Tenner-Racz, K. and Haase, A.T. (1993) Massive covert infection of helper T lymphocytes and macrophages by HIV during the incubation period of AIDS. *Nature*, **362**, 359-362.

Emili, A., Greenblatt, J. and Ingles, C.J. (1994) Species-specific interaction of the glutamine-rich activation domains of Sp1 with the TATA box-binding protein. *Mol. Cell. Biol.*, **14**, 1582-1593.

Ensoli, B., Barillari, G., Salahuddin, S.Z., Gallo, R.C. and Wong-Staal, F. (1990) Tat protein of HIV-1 stimulates growth of cells derived from Kaposi's sarcoma lesions of AIDS patients. *Nature*, **345**, 84-.

Ensoli, B., Buonaguro, L., Barillari, G., Fiorelli, V., Gendelman, R., Morgan, R.A., Wingfield, P. and Gallo, R.C. (1993) Release, uptake, and effects of extracellular human immunodeficiency virus type 1 Tat protein on cell growth and viral transactivation. *J Virol*, **67**, 277-87.

Fawell, S., Seery, J., Daikh, Y., Moore, C., Chen, L.L., Pepinsky, B. and Barsoum, J. (1994) Tat-mediated delivery of heterologous proteins into cells. *Proc Natl Acad Sci U S A*, **91**, 664-8.

Feng, C.P., Kulka, M. and Aurelian, L. (1993) The GGTCA palindrome and cognate cellular factors in trans-regulation of human immunodeficiency virus long terminal repeat by herpes simplex virus. *J. Gen. Virol.*, **74**, 715-723.

Ferré-D'Amaré, A.R., Pognonec, P., Roeder, R.G. and Burley, S.K. (1994) Structure and function of the b/HLH/Z domain of USF. *EMBO J.*, **13**, 180-189.

Ferré-D'Amaré, A.R., Prendergast, G.C., Ziff, E.B. and Burley, S.K. (1993) Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature*, **363**, 38-45.

Finnie, N.J., Gottlieb, T.M., Blunt, T., Jeggo, P.A. and Jackson, S.P. (1995) DNA-dependent protein kinase activity is absent in xrs-6 cells: implications for site specific recombinations and DNA double strand repair. *Proc. Natl. Acad. Sci.*, **92**, 320-324.

Fisher, D.E., Parent, L.A. and Sharp, P.A. (1992) Myc/Max and other helix-loop-helix/leucine zipper proteins bend DNA toward the minor groove. *Proc. Natl. Acad. Sci. USA*, **89**, 11779-11783.

Flores, O., Lu, H. and Reinberg, D. (1992) Factors involved in specific transcription by mammalian RNA polymerase II. Identification and characterization of factor IIH. *J Biol Chem*, **267**, 2786-93.

Flores, S.C., Marecki, J.C., Harper, K.P., Bose, S.K., Nelson, S.K. and McCord, J.M. (1993) Tat protein of human immunodeficiency virus type 1 represses expression of manganese superoxide dismutase in HeLa cells. *Proc Natl Acad Sci U S A*, **90**, 7632-6.

Frankel, A.D., Biancalana, S. and Hudson, D. (1989) Activity of synthetic peptides from the Tat protein of human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A*, **86**, 7397-401.

Frankel, A.D. and Pabo, C.O. (1988) Cellular uptake of the Tat protein from Human Immunodeficiency Virus. *Cell*, **55**, 1189-1193.

Franza, B.R.J., Rauscher, F.J.I., Josephs, S.F. and Curran, T. (1988) The Fos complex and Fos-related antigens recognize sequence elements that contain AP-1 binding sites. *Science*, **239**, 1150-1153.

Fridell, R.A., Harding, L.S., Bogerd, H.P. and Cullen, B.R. (1995) Identification of a novel human zinc finger protein that specifically interacts with the activation domain of lentiviral Tat proteins. *Virology*, **209**, 347-57.

Fujita, T., Shibuya, H., Ohashi, T., Yamanishi, K. and Taniguchi, T. (1986) Regulation of human interleukin-2 gene: functional DNA sequences in the 5' flanking region for the gene expression in activate T lymphocytes. *Cell*, **49**, 401-409.

Garcia, J.A., Harrich, D., Soultanakis, E., Wu, F., Mitsuyasu, R. and Gaynor, R.B. (1989) Human immunodeficiency virus type 1 LTR TATA and TAR region sequences required for transcriptional regulation. *Embo J*, **8**, 765-78.

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

Gatignol, A., Buckler-White, A., Berkhout, B. and Jeang, K.T. (1991) Characterization of a human TAR RNA-binding protein that activates the HIV- 1 LTR. *Science*, **251**, 1597-600.

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

Gaynor, G. (1992) Cellular transcription factors involved in the regulation of HIV-1 gene expression. *AIDS*, **6**, 347-363.

Ghosh, D. (1992) Glucocorticoid receptor-binding site in the human immunodeficiency virus long terminal repeat. *J. Virol.*, **66**, 586-590.

Giacca, M., Gutierrez, M.I., Diviacco, S., De Marchi, F., 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. Comm.*, **165**, 956-965.

Giacca, M., Gutierrez, M.I., Menzo, S., d'Adda di Fagagna, F. and Falaschi, A. (1992) A human binding site for transcription factor USF/MLTF mimics the negative regulatory element of human immunodeficiency virus type 1. *Virology*, **186**, 133-147.

Giacca, M., Zentilin, L., Norio, P., Diviacco, S., Dimitrova, D., Contreas, G., Biamonti, G., Perini, G., Weighardt, F., Riva, S. and Falaschi, A. (1994) Fine mapping of a replication origin of human DNA. *Proc. Natl. Acad. Sci. USA*, **91**, 7119-7123.

Gorman, C.M., Moffat, L.F. and Howard, B.H. (1982) Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.*, **2**, 1044-1051.

Gottlieb, T.M. and Jackson, S.P. (1993) The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell*, **72**, 131-142.

Gottlieb, T.M. and Jackson, S.P. (1994) Protein kinases and DNA damage. *TIBS*, **19**, 500-503.

Gregor, P.D., Sawadogo, M. and Roeder, R.G. (1990) The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer. *Genes Dev.*, **4**, 1730-1740.

Grosschedl, R., Giese, K. and Pagel, J. (1994) HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. *Trends Genet*, **10**, 94-100.

Gupta, P., Enrico, A., Armstrong, J., Doerr, M., Ho, M. and Rinaldo, C. (1990) A semiquantitative microassay for measurement of relative number of blood

mononuclear cells infected with human immunodeficiency virus. *AIDS Res. Hum. Retrov.*, **6**, 1193-1196.

Gupta, P., Kingsley, L., Armstrong, J., Ding, M., Cottrill, M. and Rinaldo, C. (1993) Enhanced expression of human immunodeficiency virus type 1 correlates with development of AIDS. *Virology*, **196**, 586-595.

Gutheil, W.G., Subramanyam, M., Flentke, G.R., Sanford, D.G., Munoz, E., Huber, B.T. and Bachovchin, W.W. (1994) Human immunodeficiency virus 1 Tat binds to dipeptidyl aminopeptidase IV (CD26): A possible mechanism for Tat's immunosuppressive activity. *Proc. Natl. Acad. Sci. USA*, **91**, 6594-6598.

Han, X.M., Laras, A., Rounseville, M.P., Kumar, A. and Shank, P.R. (1992) Human immunodeficiency virus type 1 Tat-mediated trans activation correlates with the phosphorylation state of a cellular TAR RNA stem- binding factor. *J Virol*, **66**, 4065-72.

Harper, J.W. and Logsdon, N.J. (1991) Refolded HIV-1 tat protein protects both bulge and loop nucleotides in TAR RNA from ribonucleolytic cleavage. *Biochemistry*, **30**, 8060-6.

Hart, C.E., Ou, C.Y., Galphin, J.C., Moore, J., Bacheler, L.T., Wasmuth, J.J., Petteway, S.R., Jr. and S chochetman, G. (1989) Human chromosome 12 is required for elevated HIV-1 expression in human- hamster hybrid cells. *Science*, **246**, 488-91.

Hattori, M., Tugores, A., Veloz, L., Karin, M. and Brenner, D.A. (1990) A simplified method for the preparation of transcriptionally active liver nuclear extracts. *DNA and Cell Biology*, **9**, 777-781.

Hauber, J. and Cullen, B.R. (1988) Mutational Analysis of the *trans*-Activation -Responsive Region of the Human Immunodeficiency Virus Type I Long Terminal Repeat. , **62**, 673-679.

Hauber, J., Perkins, A., Heimer, E.P. and Cullen, B.R. (1987) Trans-activation of human immunodeficiency virus gene expression is mediated by nuclear events. *Proc. Natl. Acad. Sci.*, **84**, 6364-6368.

Heiland, S., Knippers, R. and Kunze, N. (1993) The promoter region of the human type-I-DNA-topoisomerase gene - Protein-binding and sequences involved in transcriptional regulation. *Eur. J. Biochem.*, **217**, 813-822.

Helland, D.E., Welles, J.L., Caputo, A. and Haseltine, W.A. (1991) Transcellular transactivation by the human immunodeficiency virus type 1 tat protein. *J Virol*, **65**, 4547-9.

Herrmann, C.H. and Rice, A.P. (1993) Specific interaction of the human immunodeficiency virus Tat proteins with a cellular protein kinase. *Virology*, **197**, 601-8.

Herrmann, C.H. and Rice, A.P. (1995) Lentivirus Tat proteins specifically associate with a cellular protein kinase, TAK, that hyperphosphorylates the carboxyl-terminal domain of the large subunit of RNA polymerase II: candidate for a Tat cofactor. *J Virol*, **69**, 1612-20.

Ho, D.D., Neumann, A.U., Perelson, A.S., Chen, W., Leonard, J.M. and Markowitz, M. (1995) Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature*, **373**, 123-126.

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. and Pease, L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, **77**, 51-59.

Hoffman, P.W. and Chernak, J.M. (1995) DNA binding and regulatory effects of transcription factors SP1 and USF at the rat amyloid precursor protein gene promoter. *Nucleic Acids Res*, **23**, 2229-35.

Horikoshi, M., Carey, M.F., Kakidani, H. and Roeder, R.G. (1988) Mechanism of activation of a yeast activator: direct effect of GAL4 derivatives on mammalian TFIID-promoter interactions. *Cell*, **54**, 665-669.

Howcroft, K., Strebel, K., Martin, M.A. and Singer, D.S. (1993) Repression of MHC class I gene promoter activity by two-exon Tat of HIV. *Science*, **260**, 1320-1322.

Iijima, S., Teraoka, H., Date, T. and Tsukada, K. (1992) DNA-activated protein kinase in Raji Burkitt's lymphoma cells. Phosphorylation of c-Myc oncoprotein. *Eur. J. Biochem.*, **206**, 595-603.

Ikeda, K., Nagano, K. and Kawakami, K. (1993) Possible implications of Sp1-induced bending of DNA on synergistic activation of transcription. *Gene*, **136**, 341-343.

Jeang, K.T., Chun, R., Lin, N.H., Gatignol, A., Glabe, C.G. and Fan, H. (1993) In vitro and in vivo binding of human immunodeficiency virus type 1 Tat p rotein and Sp1 transcription factor. *J Virol*, **67**, 6224-33.

Jones, K.A., Luciw, P.A. and Duchange, N. (1988) Structural arrangements of transcription control domains within the 5'-untranslated leader regions of the HIV-1 and HIV-2 promoters. *Genes Dev.*, **2**, 1101-1114.

Jones, K.A. and Peterlin, B.M. (1994) Control of RNA initiation and elongation at the HIV-1 promoter. *Annu. Rev. Biochem.*, **63**, 717-743.

Jurriaans, S., Dekker, J.T. and de Ronde, A. (1992) HIV-1 viral DNA load in peripheral blood mononuclear cells from seroconverters and long-term infected individuals. *AIDS*, **6**, 635-641.

Kaczmarski, W. and Khan, S.A. (1993) Lupus autoantigen Ku protein binds HIV-1 TAR RNA in vitro. *Biochem. Biophys. Res. Commun.*, **196**, 935-942.

Kamine, J., Loewenstein, P. and Green, M. (1991) Mapping of HIV-1 Tat protein sequences required for binding to Tar RNA. *Virology*, **182**, 570-7.

Kao, S.-Y., Calman, A.F., Luciw, P.A. and Peterlin, B.M. (1987) Anti-termination of transcription within the long terminal repeat of HIV-1 by *tat* gene product. , **330**, 489-493.

Kashanchi, F., Piras, G., Radonovich, M.F., Duvall, J.F., Fattaey, A., Chiang, C.M., Roeder, R.G. and B rady, J.N. (1994) Direct interaction of human TFIID with the HIV-1 transactivator tat. *Nature*, **367**, 295-9.

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

Kato, H., Sumimoto, H., Pognonec, P., Chen, C.H., Rosen, C.A. and Roeder, R.G. (1992) HIV-1 Tat acts as a processivity factor in vitro in conjunction with cellular elongation factors. *Genes Dev.*, **6**, 655-666.

Kaufman, J.D., Valandra, G., Roderiquez, G., Bushar, G., Giri, C. and Norcross, M.A. (1987) Phorbol ester enhances human immunodeficiency virus-promoted gene expression and acts on a repeated 10-base-paier functional enhancer element. *Mol. Cell. Biol.*, **7**, 3759.

Kawakami, K., Scheidereit, C. and Roeder, R.G. (1988) Identification and purification of a human immunoglobulin-enhancer-binding protein (NF-κB) that activates transcription from a human immunodeficiency virus type 1 promoter in vitro. *Proc. Natl. Acad. Sci. USA*, **85**, 4700-4704.

Kerppola, T.K. and Curran, T. (1991) Fos-Jun heterodimers and Jun homodimers bend DNA in opposite orientations: implications for transcription factor cooperativity. *Cell*, **66**, 317-326.

Kessler, M. and Mathews, M.B. (1992) Premature termination and processing of human immunodeficiency virus type 1- promoted transcripts. *J. Virol.*, **66**, 4488-4496.

Kim, J.Y., Gonzalez-Scarano, F., Zeichner, S.L. and Alwine, J.C. (1993) Replication of type 1 human immunodeficiency viruses containing linker substitution mutations in the -201 to -130 region of the long terminal repeat. *J. Virol.*, **67**, 1658-1662.

Kim, S., Byrn, R., Groopman, J. and 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, Y.S. and Panganiban, A.T. (1993) The full-length Tat protein is required for TAR-independent, posttranscriptional trans activation of human immunodeficiency virus type 1 env gene expression. *J Virol*, **67**, 3739-47.

Kim, Y.S. and Risser, R. (1993) TAR-independent transactivation of the murine cytomegalovirus major immediate-early promoter by the Tat protein. *J Virol*, **67**, 239-48.

Kirschbaum, B.J., Pognonec, P. and Roeder, R. (1992) Definition of the transcriptional activation domain of recombinant 43-kilodalton USF. *Molecular and cellular biology*, **12**, 5094-5101.

Klaver, B. and Berkhout, B. (1994) Evolution of a disrupted TAR RNA hairpin structure in the HIV-1 virus. *Embo J*, **13**, 2650-9.

Klotman, M.E., Kim, S., Buchbinder, A., DeRossi, A., Baltimore, D. and Wong-Staal, F. (1991) Kinetics of expression of multiply spliced RNA in early human immunodeficiency virus type 1 infection of lymphocytes and monocytes. *Proc. Natl. Acad. Sci. USA*, **88**, 5011-5015.

Koken, S.E., Greijer, A.E., Verhoef, K., van Wamel, J., Bukrinskaya, A.G. and Berkhout, B. (1994) Intracellular analysis of in vitro modified HIV Tat protein. *J Biol Chem*, **269**, 8366-75.

Kolson, D.L., Buchhalter, J., Collman, R., Hellmig, B., Farrell, C.F., Debouck, C. and Gonzalez- Scarano, F. (1993) HIV-1 Tat alters normal organization of neurons and

astrocytes in primary rodent brain cell cultures: RGD sequence dependence. *AIDS Res Hum Retroviruses*, **9**, 677-85.

Kuhn, A., Gottlieb, T.M., Jackson, S.P. and Grummt, I. (1995) DNA-dependent protein kinase: a potent inhibitor of transcription by RNA polymerase I. *Genes and Development*, **9**, 193-203.

Kuppuswamy, M., Subramanian, T., Srinivasan, A. and Chinnadurai, G. (1989) Multiple functional domains of Tat, the trans-activator of HIV-1, defined by mutational analysis. *Nucleic Acids Res*, **17**, 3551-61.

Laspia, M.F., Rice, A.P. and Mathews, M.B. (1989) HIV-1 Tat protein increases transcriptional initiation and stabilizes elongation. *Cell*, **59**, 283-292.

Lee, J.S., Galvin, K.M. and Shi, Y. (1993) Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. *Proc. Natl. Acad. Sci. USA*, **90**, 6145-6149.

Lees-Miller, S.P., Sakaguchi, K., Ullrich, S.J., E., A. and Anderson, C.W. (1992) Human DNA-activated protein kinase phosphorilates serines 15 and 37 in the amino-terminal domain of human p53. *Mol. Cell. Biol.*, **12**, 5041-5049.

Li, C., Lai, C., Singman, D.S. and Gaynor, R.B. (1991) Cloning of a cellular factor, interleukin binding factor, that binds to a NFAT-like motifs in the human immunodeficiency virus long terminal repeat. *Proc. Natl. Acad. Sci. USA*, **88**, 7739-7743.

Li, C.J., Friedman, D.J., Wang, C., Metelev, V. and Pardee, A.B. (1995) Induction of apoptosis in uninfected lymphocytes by HIV-1 Tat protein. *Science*, **268**, 429-31.

Liu, S.H., Ma, J.T., Yueh, A.Y., Lees-Miller, S.P., Anderson, C.W. and Ng, S.Y. (1993) The carboxyl-terminal transactivating domain of human serum response factor contains DNA-activated protein kinase phosphorylation sites. *J. Biol. Chem.*, **268**, 21147-21154.

Lowenthal, J.W., Ballard, D.W., Bohnlein, E. and Greene, W.C. (1989) Tumor necrosis factor α induces proteins that bind specifically to κB-like enhancer elements and regulate interleukin 2 receptor α-chain gene expression in primary human T-lymphocytes. *Proc. Natl. Acad. Sci. USA*, **86**, 2331-2335.

Lu, X., Welsh, T.M. and Peterlin, B.M. (1993) The human immunodeficiency virus type 1 long terminal repeat specifies two different transcription complexes, only one of which is regulated by Tat. *J Virol*, **67**, 1752-60.

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.

Maekawa, T., Sudo, T., Kurimoto, M. and Ishii, S. (1991) USF-related transcription factor, HIV-TF1, stimulates transcription of human immunodeficiency virus-1. *Nucleic Acids Res.*, **19**, 4689-4694.

Majello, B., De Luca, P., Hagen, G., Suske, G. and Lania, L. (1994) Different members of the Sp1 multigene family exert opposite transcriptional regulation of the long terminal repeat of HIV-1. *Nucleic Acids Res*, **22**, 4914-21.

Mann, D.A. and Frankel, A.D. (1991) Endocytosis and targeting of exogenous HIV-1 Tat protein. *Embo J*, **10**, 1733-9.

Marciniak, R.A., Garcia-Blanco, M.A. and Sharp, P.A. (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.

Margolis, D.M., Ostrove, J.M. and Straus, S.E. (1993) HSV-1 activation of HIV-1 transcription is augmented by a cellular protein that binds near the initiator element. *Virology*, **192**, 370-374.

Margolis, D.M., Somasundaran, M. and Green, M.R. (1994) Human transcription factor YY1 represses human immunodeficiency virus type 1 transcription and virion production. *J. Virol.*, **68**, 905-910.

Masood, R., Lunardi-Iskandar, Y., Moudgil, T., Zhang, Y., Law, R.E., Huang, C.L., Puri, R.K., L evine, A.M. and Gill, P.S. (1994) IL-10 inhibits HIV-1 replication and is induced by tat. *Biochem Biophys Res Commun*, **202**, 374-83.

Maxam, A.M. and Gilbert, W. (1980) In Grossman, L. and Moldave, K. (eds.), *Sequencing end-labeled DNA with base-specific chemical cleavages*, Academic Press, New York, Vol. 65, pp. 499-560.

Menzo, S., Bagnarelli, P., Giacca, M., Manzin, A., Varaldo, P.E. and Clementi, M. (1992) Absolute quantitation of viremia in HIV-infected asymptomatic subjects by

competitive reverse-transcription and polymerase chain reaction. *J. Clin. Microbiol.*, **30**, 1752-1757.

Michael, N.L., D'Arcy, L., Ehrenberg, P.K. and Redfield, R.R. (1994a) Naturally occurring genotypes of the human immunodeficiency virus type 1 long terminal repeat display a wide range of basal and Tat-induced transcriptional activities. *J. Virol.*, **68**, 3163-3174.

Michael, N.L., Morrow, P., Mosca, J., Vahey, M., Burke, D.S. and Redfield, R.R. (1991) Induction of human immunodeficiency virus type 1 expression in chronically infected cells is associated primarily with a shift in RNA splicing patterns. *J. Virol.*, **65**, 1291-1303.

Michael, N.L., Vahey, M., Burke, D.S. and Redfield, R.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.

Michael, N.L., Vahey, M.T., d'Arcy, L., Ehrenberg, P.K., Mosca, J.D., Rappaport, J. and Redfield, R.R. (1994) Negative-strand RNA transcripts are produced in human immunodeficiency virus type 1-infected cells and patients by a novel promoter downregulated by Tat. *J Virol*, **68**, 979-87.

Miyamoto, N.G., Moncollin, V., Egly, J.M. and Chambon, P. (1985) Specific interaction between a transcrition factor and the upstream element of the adenovirus-2 major late promoter. *EMBO J.*, **4**, 3563-3570.

Moses, A.V., Ibanez, C., Gaynor, R., Ghazal, P. and Nelson, J.A. (1994) Differential role of long terminal repeat control elements for the regulation of basal and Tat-mediated transcription of the human immunodeficiency virus in stimulated and unstimulated primary human macrophages. *J Virol*, **68**, 298-307.

Muesing, M.A., Smith, D.H. and Capon, D.J. (1987) Regulation of mRNA accumulation by a human immunodeficiency virus *trans*-activator protein. *Cell*, **48**, 691-701.

Murre, C., McCaw, P.S., Vaessin, H., Caudy, M., Jan, L.Y., Jan, Y.N., Cabrera, C.V., Buskin, J.N., Hauschka, S.D., Lassar, A.B., Weintraub, H. and Baltimore, D. (1989) Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell*, **58**, 537-544.

Nabel, G. and Baltimore, D. (1987) An inducible transcription factor activates expression of human immunodeficiency virus in T cells. *Nature*, **326**, 711-713.

Natesan, S. and Gilman, M.Z. (1993) DNA bending and orientation-dependent function of YY1 in the c-fos promoter. *Genes Dev.*, **7**, 2497-2509.

Nelbock, P., Dillon, P.J., Perkins, A. and Rosen, C.A. (1990) A cDNA for a protein that interacts with the human immunodeficiency virus Tat transactivator. *Science*, **248**, 1650-1653.

Newstein, M., Stanbridge, E.J., Casey, G. and Shank, P.R. (1990) Human chromosome 12 encodes a species-specific factor which increases human immunodeficiency virus type 1 tat-mediated trans activation in rodent cells. *J Virol*, **64**, 4565-7.

Ohana, B., Moore, P.A., Ruben, S.M., Southgate, C.D., Green, M.R. and Rosen, C.A. (1993) The type 1 human immunodeficiency virus Tat binding protein is a transcriptional activator belonging to an additional family of evolutionarily conserved genes. *Proc Natl Acad Sci U S A*, **90**, 138-42.

Orchard, K., Lang, G., Collins, M. and Latchman, D. (1992) Characterization of a novel T lymphocyte protein which binds to a site related to steroid/thyroid hormone receptor response elements in the negative regulatory sequence of the human immunodeficiency virus long terminal repeat. *Nucleic Acids Res.*, **20**, 5429-5434.

Orchard, K., Perkins, N., Chapman, C., Harris, J., Emery, V., Goodwin, G., Latchman, D. and Collins, M. (1990) A novel T-cell protein which recognizes a palindromic sequence in the Negative Regulatory Element of the human immunodeficiency virus Long Terminal Repeat. *J. Virol.*, **64**, 3234-3239.

Osborn, L., Kunkel, S. and Nabel, G.J. (1989) Tumor necrosis factor alpha and interleukin 1 stimulate the human immunodeficiency virus enhancer by activation of the nuclear factor kappa B. *Proc. Natl. Acad. Sci. USA*, **86**, 2336-2340.

Outram, S.V. and Owen, M.J. (1994) The helix-loop-helix containing transcription factor USF activates the promoter of the CD2 gene. *J Biol Chem*, **269**, 26525-30.

Pantaleo, G., Graziosi, C., Demarest, J.F., Butini, L., Montroni, M., Fox, C.H., Orenstein, J.M., Kotler, D.P. and Fauci, A.S. (1993) HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease. *Nature*, **362**, 355-358.

Papavassiliou, A.G. (1993) Localization of DNA-protein contact points by DMS resistance of complexes resolved in gel retardation assays. *Nucleic Acids Res.*, **21**, 757-758.

Parvin, J.D., McCormick, R.J., Sharp, P.A. and Fisher, D.E. (1995) Pre-bending of a promoter sequence enhances affinity for the TATA-binding factor. *Nature*, **373**, 724-7.

Peritz, L.N., Fodor, E.J.B., Silversides, D.W., Cattini, P.A., Baxter, J.D. and Eberhardt, N.L. (1988) The human growth hormone gene contains both positive and negative control elements. *J. Biol. Chem.*, **263**, 5005-5007.

Perkins, N.D., Edwards, N.L., Duckett, C.S., Agranoff, A.B., Schmid, R.M. and Nabel, G.J. (1993) A cooperative interaction between NF-kappa B and Sp1 is required for HIV-1 enhancer activation. *EMBO J.*, **12**, 3551-3558.

Peterlin, B.M., Luciw, P.A., Barr, P.J. and Walker, M.D. (1986) Elevated levels of mRNA can accaunt for the trans-activation of human immunodeficiency virus. *Proc. Natl. Acad. Sci.*, **83**, 9734-9738.

Piatak, M., Jr., Saag, M.S., Yang, L.C., Clark, S.J., Kappes, J.C., Luk, K.-C., Hahn, B.H., Shaw, G.M. and Lifson, J.D. (1993) High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. *Science*, **259**, 1749-1754.

Pognonec, P., Kato, H. and Roeder, R.G. (1992) The helix-loop-helix/leucine repeat transcription factor USF can be functionally regulated in a redox-dependent manner. *J Biol Chem*, **267**, 24563-7.

Pognonec, P. and Roeder, R.G. (1991) Recombinant 43-kDa USF binds to DNA and activates transcription in a manner indistinguishable from that of natural 43/44-kDa USF. *Mol. Cell. Biol.*, **11**, 5125-5136.

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

Prendergast, G.C. and Ziff, E.B. (1991) Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. *Science*, **251**, 186-189.

Puri, R.K. and Aggarwal, B.B. (1992) Human immunodeficiency virus type 1 tat gene up-regulates interleukin 4 receptors on a human B-lymphoblastoid cell line. *Cancer Res*, **52**, 3787-90.

Rappaport, J., Lee, S.J., Khalili, K. and Wong-Staal, F. (1989) The acidic amino-terminal region of the HIV-1 Tat protein constitutes an essential activating domain. *New Biol*, **1**, 101-10.

Ratnasabapathy, R., Sheldon, M., Johal, L. and Hernandez, N. (1990) The HIV-1 long terminal repeat contains an unusual element that induces the synthesis of short RNAs from various mRNA and snRNA promoters. *Genes Dev.*, **4**, 2061-2074.

Read, M.L., Clark, A.R. and Docherty, K. (1993) The helix-loop-helix transcription factor USF (upstream stimulating factor) binds to a regulatory sequence of the human insulin gene enhancer. *Biochem. J.*, **295**, 233-237.

Rhim, H., Chan, F., Echetebu, C.O. and Rice, A.P. (1993) Rapid purification of monomer HIV-2 Tat protein expressed in Escherichia coli. *Protein Expr Purif*, **4**, 24-31.

Riccio, A., Pedone, P.V., Lund, L.R., Olesen, T., Olsen, H.S. and Andreasen, P.A. (1992) Transforming growth factor beta 1-responsive element: closely associated binding sites for USF and CCAAT-binding transcription factor-nuclear factor I in the type 1 plasminogen activator inhibitor gene. *Mol. Cell. Biol.*, **12**, 1846-1855.

Rice, A.P. and Carlotti, F. (1990a) Mutational analysis of the conserved cysteine-rich region of the human immunodeficiency virus type 1 Tat protein. *J Virol*, **64**, 1864-8.

Rice, A.P. and Carlotti, F. (1990b) Structural analysis of wild-type and mutant human immunodeficiency virus type 1 Tat proteins. *J Virol*, **64**, 6018-26.

Rice, A.P. and Chan, F. (1991) Tat protein of human immunodeficiency virus type 1 is a monomer when expressed in mammalian cells. *Virology*, **185**, 451-4.

Rice, A.P. and Mathews, M.B. (1988) Transcriptional but not translational regulation of HIV-1 by the *tat* gene product. , **332**, 551-553.

Riviere, Y., Blank, V., Kourilsky, P. and Israel, A. (1991) Processing of the precursor of NF-kappa B by the HIV-1 protease during acute infection. *Nature*, **350**, 625-6.

Rosen, C.A., Sodroski, J.G. and Haseltine, W.A. (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.

Roy, A.L., Meisterernst, M., Pognonec, P. and Roeder, R.G. (1991) Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF. *Nature*, **354**, 245-248.

Roy, S., Delling, U., Chen, C.H., Rosen, C.A. and Sonenberg, N. (1990a) A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-m ediated trans-activation. *Genes Dev*, **4**, 1365-73.

Roy, S., Katze, M.G., Parkin, N.T., Edery, I., Hovanessian, A.G. and Sonenberg, N. (1990b) Control of the interferon-induced 68-kilodalton protein kinase by the HIV-1 tat gene product. *Science*, **247**, 1216-9.

Sabatier, J.M., Vives, E., Mabrouk, K., Benjouad, A., Rochat, H., Duval, A., Hue, B. and Bahraoui, E. (1991) Evidence for neurotoxic activity of tat from human immunodeficiency virus type 1. *J Virol*, **65**, 961-7.

Sadaie, M.R., Benter, T. and Wong-Staal, F. (1988) Site-directed mutagenesis of two trans-regulatory genes (*tat*-III, *trs*) of HIV-1. , **239**, 910-913.

Saksela, K., Stevens, C., Rubinstein, P. and Baltimore, D. (1994) Human immunodeficiency virus type 1 mRNA expression in peripheral blood cells predicts disease progression independently of the numbers of CD4+ lymphocytes. *Proc. Natl. Acad. Sci. USA*, **91**, 1104-1108.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual/II Edition.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sato, M., Ishizawa, S., Yoshida, T. and Shibahara, S. (1990) Interaction of upstream factor with the human heme oxygenase gene promoter. *Eur. J. Biochem.*, **188**, 231-237.

Sawadogo, M. (1988) Multiple forms of the human gene-specific transcription factor USF. II. DNA binding properties and transcriptional activity of the purified HeLa USF. *J. Biol. Chem.*, **263**, 11994-12001.

Sawadogo, M. and Roeder, R.G. (1985a) Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative in vitro assay. *Proc. Natl. Acad. Sci. USA*, **82**, 4394-4398.

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

Scadden, D.T., Wang, Z. and Groopman, J.E. (1992) Quantitation of plasma human immunodeficiency virus type 1 RNA by competitive polymerase chain reaction. *J. Infect. Dis.*, **165**, 1119-1123.

Schmid, R.M., Perkins, N.D., Duckett, C.S., P.C., A. and G.J., N. (1991) Cloning of an NF-κB subunit which stimulates HIV transcription in synergy with p65. *Nature*, **352**, 733-736.

Schnittman, S.M., Greenhouse, J.J., Lane, H.C., Pierce, P.F. and Fauci, A.S. (1991) Frequent detection of HIV-1-specific mRNAs in infected individuals suggests ongoing active viral expression in all stages of disease. *AIDS Res Hum Retroviruses*, **7**, 361-7.

Schnittman, S.M., Greenhouse, J.J., Psallidopoulos, M.C., Baseler, M., Salzman, N.P., Fauci, A.S. and Lane, H.C. (1990) Increasing viral burden in CD4+ T cells from patients with human immunodeficiency virus (HIV) infection reflects rapidly progressive immunosuppression and clinical disease. *Ann. Intern. Med.*, **113**, 438-443.

Schorpp, M., Kugler, W., Wagner, U. and Ryffel, G.U. (1988) Hepatocyte-specific promoter element HP1 of the *Xenopus* albumin gene interacts with transcriptional factors of mammalian hepatocytes. *J. Mol. Biol.*, **202**, 307-320.

Schreck, R., Zorbas, H., Winnacker, E.L. and Baeuerle, P.A. (1990) The NF-κB transcription factor induces DNA bending which is modulated by its 65-kD subunit. *Nucleic Acids Res.*, **18**, 6497-6502.

Schwartz, S., Felber, B.K., Benko, D.M., Fenyo, E.M. and Pavlakis, G.N. (1990) Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. *J. Virol.*, **64**, 2519-2529.

SenGupta, D.N., Berkhout, B., Gatignol, A., Zhou, A.M. and Silverman, R.H. (1990) Direct evidence for translational regulation by leader RNA and Tat protein of human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A*, **87**, 7492-6.

Seto, E., Lewis, B. and Shenk, T. (1993) Interaction between transcription factors Sp1 and YY1. *Nature*, **365**, 462-464.

Seto, E., Shi, Y. and Shenk, T. (1991) YY1 is an initiator sequence-binding protein that directs and activates transcription in vitro. *Nature*, **354**, 241-245.

Shaw, J.P., Utz, P.J., Durand, D.B., Toole, S.S., Emmer, E.A. and D.R., C. (1988) Identification of a putative regulator of early T cell activation genes. *Science*, **241**, 202-205.

Sheldon, M., Ratnasabapathy, R. and Hernandez, N. (1993) Characterization of the inducer of short transcripts, a human immunodeficiency virus type 1 transcriptional element that activates the synthesis of short RNAs. *Mol. Cell. Biol.*, **13**, 1251-1263.

Sheline, C.T., Milocco, L.H. and Jones, K.A. (1991) Two distinct nuclear transcription factors recognize loop and bulge residues of the HIV-1 TAR RNA hairpin. *Genes Dev*, **5**, 2508-20.

Shibuya, H., Irie, K., Ninomiya-Tsuji, J., Goebl, M., Taniguchi, T. and Matsumoto, K. (1992) New human gene encoding a positive modulator of HIV Tat-mediated transactivation. *Nature*, **357**, 700-2.

Siekevitz, M., Josephs, S.F., Dukovich, M., Peffer, N., Wong-Staal, F. and Greene, W.C. (1987) Activation of the HIV-1 LTR by T cell mitogens and the trans-activator protein of HTLV-I. *Science*, **238**, 1575-1578.

Simmonds, P., Balfe, P., Peutherer, J.F., Ludlam, C.A., Bishop, J.O. and Brown, A.J. (1990) Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. *J. Virol.*, **64**, 864-872.

Sirito, M., Lin, Q., Maity, T. and Sawadogo, M. (1994) Ubiquitous expression of the 43- and 44-kDa forms of transcription factor USF in mammalian cells. *Nucleic Acids Res.*, **22**, 427-433.

Sirito, M., Walker, S., Lin, Q., Kozlowsi, M.T., Klein, W.H. and Sawadogo, M. (1992) Members of the USF family of helix-loop-helix proteins bind DNA as homo- as well as heterodimers. *Gene Expression*, **2**, 231-240.

Song, C.Z., Loewenstein, P.M. and Green, M. (1994) Transcriptional activation in vitro by the human immunodeficiency virus type 1 Tat protein: evidence for specific interaction with a coactivator(s). *Proc Natl Acad Sci U S A*, **91**, 9357-61.

Subbramanian, R.A. and Cohen, E.A. (1994) Molecular biology of the human immunodeficiency virus accessory proteins. *J Virol*, **68**, 6831-5.

Subramanian, T., Govindarajan, R. and Chinnadurai, G. (1991) Heterologous basic domain substitutions in the HIV-1 Tat protein reveal an arginine-rich motif required for transactivation. *Embo J*, **10**, 2311-8.

Taccioli, G.E., Rathbun, G., Oltz, E., Stamato, T., Jeggo, P.A. and Alt, F.W. (1993) Impairement of V(D)J recombination in double-strand break repair mutants. *Science*, **260**, 207-210.

Taylor, J.P., Cupp, C., Diaz, A., Chowdhury, M., Khalili, K., Jimenez, S.A. and Amini, S. (1992) Activation of expression of genes coding for extracellular matrix proteins in Tat-producing glioblastoma cells. *Proc. Natl. Acad. Sci. USA*, **89**, 9617-9621.

Tesmer, V.M., Rajadhyaksha, A., Babin, J. and Bina, M. (1993) NF-IL6-mediated transcriptional activation of the long terminal repeat of the human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. USA*, **90**, 7298-7302.

Thompson, J.F. and Landy, A. (1988) Empirical estimation of protein-induced DNA bending angles: application to site-specific recombination complexes. *Nucleic Acids Res.*, **16**, 9687-9705.

Thorn, J.T., Todd, A.V., Warrilow, D., Watt, F., Molloy, P.L. and Iland, H.J. (1991) Characterization of the human N-ras promoter region. *Oncogene*, **6**, 1843-1850.

Tiley, L.S., Brown, P.H. and Cullen, B.R. (1990) Does the human immunodeficiency virus Tat trans-activator contain a discrete activation domain? *Virology*, **178**, 560-7.

Tong-Starksen, S.E., Baur, A., Lu, X.B., Peck, E. and Peterlin, B.M. (1993) Second exon of Tat of HIV-2 is required for optimal trans-activation of HIV- 1 and HIV-2 LTRs. *Virology*, **195**, 826-30.

Tong-Starksen, S.E., Luciw, P.A. and Peterlin, B.M. (1987) Human immunodeficiency virus long terminal repeat responds to T-cell activation signals. *Proc. Natl. Acad. Sci. USA*, **84**, 6845-6849.

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

Trono, D. (1995) HIV-1 accessory genes. *Cell*, **82**, 189-192.

Van der Vliet, P.C. and Verrijzer, C.P. (1993) Bending of DNA by transcription factors. *Bioessays*, **15**, 25-32.

Vanham, G., Kestens, L., De Meester, I., Vingerhoets, J., Penne, G., Vanhoof, G., Scharpe, S., H eyligen, H., Bosmans, E., Ceuppens, J.L. and et al. (1993) Decreased expression of the memory marker CD26 on both CD4+ and CD8+ T lymphocytes of HIV-infected subjects. *J Acquir Immune Defic Syndr*, **6**, 749-57.

Vaulont, S. and Kahn, A. (1994) Transcriptional control of metabolic regulation genes by carbohydrates. *Faseb J.*, **8**, 28-35.

Veschambre, P., Simard, P. and Jalinot, P. (1995) Evidence for functional interaction between the HIV-1 Tat transactivator and the TATA box binding protein in vivo. *J Mol Biol*, **250**, 169-80.

Viscidi, R.P., Mayur, K., Lederman, H.M. and Frankel, A.D. (1989) Inhibition of antigen-induced lymphocyte proliferation by Tat protein from HIV-1. *Science*, **246**, 1606-8.

Vogel, J., Hinrichs, S.H., Reynolds, R.K., Luciw, P.A. and Jay, G. (1988) The HIV *tat* gene induces dermal lesions resembling Kaposi's sarcoma in transgenic mice. *Nature*, **335**, 606-.

Waterman, M.L., Fischer, W.H. and Jones, K.A. (1991) A thymus-specific member of the HMG protein family regulates the human T cell receptor C alpha enhancer. *Genes Dev.*, **5**, 656-669.

Watt, F. and Molloy, P.L. (1988) Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev.*, **2**, 1136-1143.

Wechsler, D.S. and Dang, C.V. (1992) Opposite orientations of DNA bending by c-Myc and Max. *Proc. Natl. Acad. Sci. USA*, **89**, 7635-7639.

Weeks, B.S., Desai, K., Loewenstein, P.M., Klotman, M.E., Klotman, P.E., Green, M. and Kleinman, H.K. (1993) Identification of a novel cell attachment domain in the HIV-1 Tat protein and its 90-kDa cell surface binding protein. *J Biol Chem*, **268**, 5279-84.

Weeks, K.M., Ampe, C., Schultz, S.C., Steitz, T.A. and Crothers, D.M. (1990) Fragments of the HIV-1 Tat protein specifically bind TAR RNA. *Science*, **249**, 1281-5.

Wei, X., Ghosh, S.K., Taylor, M.E., Johnson, V.A., Emini, E.A., Deutsch, P., Lifson, J.D., Bonhoeffer, S., Nowak, M.A., Hahn, B.H., Saag, M.S. and Shaw, G.M. (1995)

Viral dynamics in human immunodeficiency virus type 1 infection. *Nature*, **373**, 117-122.

Weinmann, R. (1992) The basic RNA polymerase II transcriptional machinery. *Gene Expr*, **2**, 81-91.

West, M., Mikovits, J., Princler, G., Liu, Y.-L., Ruscetti, F.W., Kung, H.-F. and Raziuddin (1992) Characterization and purification of a novel transcriptional repressor from HeLa cell nuclear extracts recognizing the negative regulatory element region of human immunodeficiency virus-1 long terminal repeat. *J. Biol. Chem.*, **35**, 24948-24952.

Wu, F., Garcia, J., Sigman, D. and Gaynor, R. (1991) tat regulates binding of the human immunodeficiency virus trans-activating region RNA loop-binding protein TRP-185. *Genes Dev*, **5**, 2128-40.

Wu, H.M. and Crothers, D.M. (1984) The locus of sequence-directed and protein-induced DNA bending. *Nature*, **308**, 509-513.

Zauli, G., Gibellini, D., Milani, D., Mazzoni, M., Borgatti, P., La Placa, M. and Capitani, S. (1993) Human immunodeficiency virus type 1 Tat protein protects lymphoid, epithelial, and neuronal cell lines from death by apoptosis. *Cancer Res*, **53**, 4481-5.

Zeichner, S.L., Kim, J.Y. and Alwine, J.C. (1991) Linker-scanning mutational analysis of the transcriptional activity of the human immunodeficiency virus type 1 long terminal repeat. *J Virol*, **65**, 2436-44.

Zervos, A.S., Gyuris, J. and Brent, R. (1993) Mxi1, a protein that specifically interacts with Max to bind Myc-Max recognition sites. *Cell*, **72**, 223-232.

Zijderveld, D., d'Adda di Fagagna, F., Giacca, M., Timmers, H.T.M. and van der Vliet, P.C. (1994) Stimulation of the adenovirus major late promoter in vitro by transcription factor USF is enhanced by the adenovirus DNA binding protein. *J. Virol.*, **68**, 8288-8295.