ISAS - INTERNATIONAL SCHOOL FOR ADVANCED STUDIES

Interactions between the transcription factor USF/MLTF and the Long Terminal Repeat of the Human Immunodeficiency Virus type I

Thesis submitted for the Degree

of *Magister Philosophiae*

Candidate: Fabrizio d'Adda di Fagagna

Supervisors: Prof. Arturo Falaschi

Dr. Mauro Giacca

Academic Year 1992/1993

TRIESTE

SISSA - SCUOLA INTERNAZIONALE SUPERIORE DI STUDI AVANZATI

TRIESTE

Strada Costiera 11

Interactions between the transcription factor USF/MLTF and the Long Terminal Repeat of the Human Immunodeficiency Virus type I

Thesis submitted for the Degree

of *Magister Philosophiae*

Candidate: Fabrizio d'Adda di Fagagna

Supervisors: Prof. Arturo Falaschi

Dr. Mauro Giacca

Academic Year 1992/1993

Contents

1.0 INTRODUCTION

1.1 HIV-1 genome organization

The human immunodeficiency virus type 1 (HIV-1) is the etiologic agent of the Acquired Immunodeficiency Syndrome (AIDS) [1]. The virus belongs to the Retroviridae family (lentivirus genus); its genome is a 9700 nucleotides long RNA template. Following cellular infection, this RNA is reverse-transcribed into a double strand DNA by the viral-encoded reverse-transcriptase and integrates into the host genome. The proviral DNA is delimited by the Long Terminal Repeats (LTR) flanking the genes encoding for the three major structural proteins (Gag, Pol and Env) and other regulatory proteins: Tat, which triggers transcription of all viral mRNAs; Rev, which positively regulates the expression of genes encoding for virion proteins, and other polipeptides whose function is less well defined, such as Nef, Vpr, Vpu and Vif. For a schematic representation of the genomic organization of HIV-1, see Figure 1.

All the transcripts originate from the 5' LTR and end within the 3' LTR. Basically HIV-1 produces three classes of mRNAs in infected cells: full length unspliced mRNAs 9 kb long, intermediate single-spliced mRNAs 4.5 kb long and small multi-spliced mRNAs 2 kb long. Moreover, different proteins arise from the use of different reading frames, ribosomal frameshifts and bicistronic mRNAs. The Tat protein, the viral transactivator, binds to the nascent RNA whose sequence is shared by all the transcripts, augmenting the rate of transcription; probably it works as a classical transcription factor with the difference that it binds RNA instead to DNA [2]. 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 [3]; these transcripts encode for the viral structural proteins.

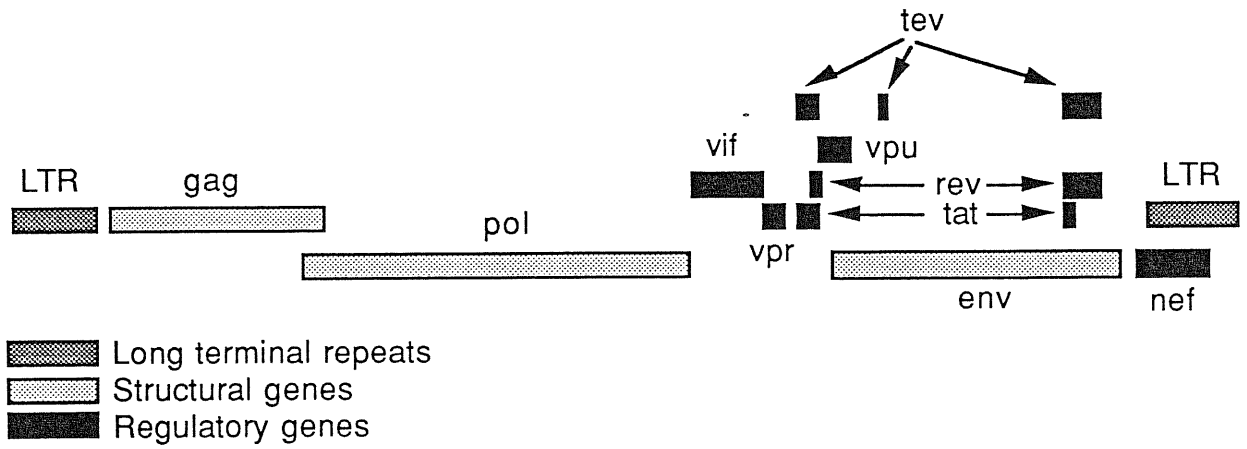


Figure 1: HIV-1 genes organization

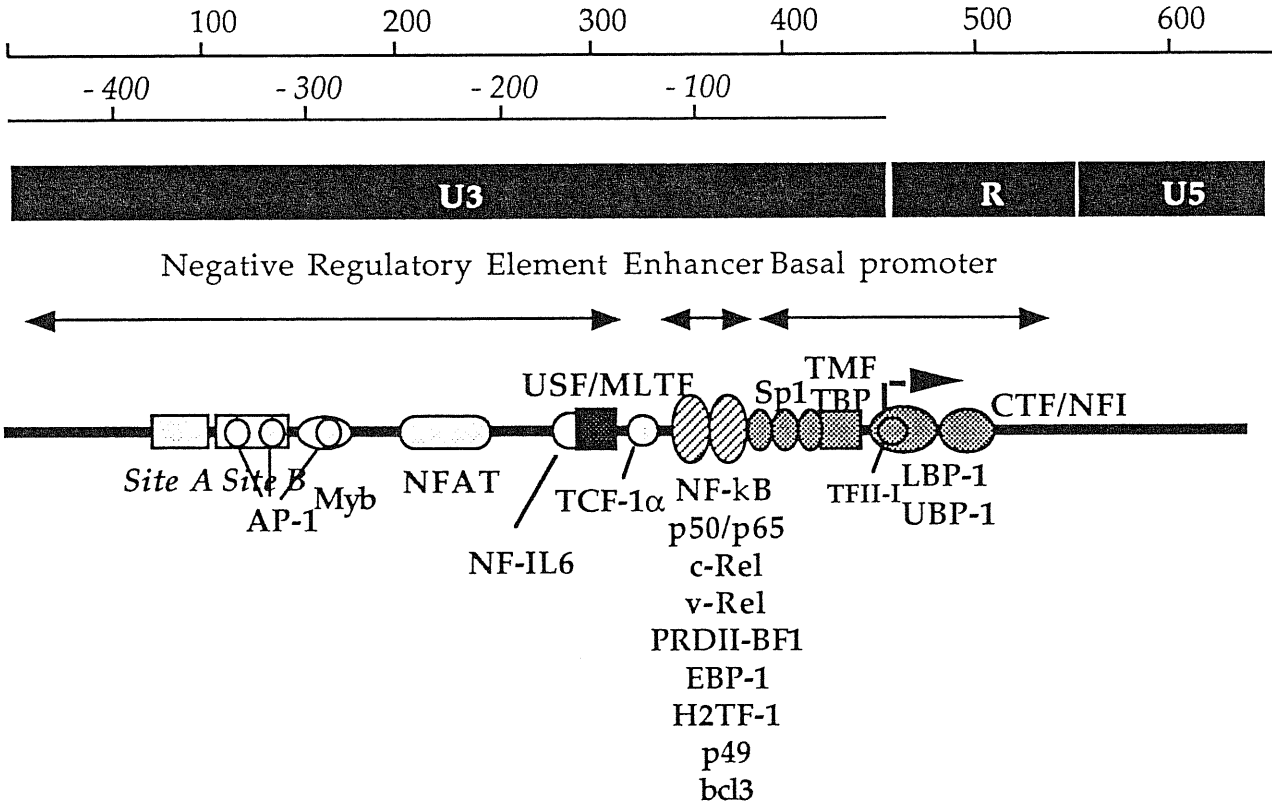


Figure 2a: Schematic representation of the HIV-1 Long Terminal Repeat and the human proteins that interact with it. Transcription start site is indicated by an arrow at the U3-R boundary; the lower scale indicates nucleotides upstream of transcription start site.

1.2 Transcriptional control of HIV-1

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, the levels of HIV-1 RNA expression, measured as RNA/DNA ratio, directly correlate with disease stage, suggesting the involvement of viral transcriptional activation in disease progression [4][5][6][7][8]. The regulation of transcription of HIV-1 is modulated by the synergetic action of the viral protein Tat and cellular proteins both interacting with cis-acting viral regulatory elements in the Long Terminal Repeat (LTR) at the 5' end of the provirus. 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 [9]. From the functional point of view, the LTR can be divided in three main regions: the basal promoter region, encompassing the transcription start site, which exerts a positive basal effect on transcription; the enhancer region, which mediates the transcriptional inducibility of the provirus upon a variety of stimuli which trigger cellular activation and proliferation [10][11][12], and the negative regulatory element (NRE), extending upstream of the enhancer site, whose overall function is to downregulate transcription. Figure 2a is the schematic representation of HIV-1 LTR and the human proteins that interact with it. Figure 2b summarizes the proteins which bind to the LTR, and some of their properties.

The **basal promoter** consists of three binding sites for the constitutive factor Sp1, the deletion of which impairs HIV-1 transcription and replication, and a TATA element, absolutely necessary for transcription, recognized by the TATA binding protein (TBP) of the TFIID chromatographic fraction; interestingly, it was shown that this particular TATA sequence interacts with a novel TATA element modulatory factor (TMF), which exhibts an inhibitory function [13]. Downstream of the TATA box a binding site for the Leader-Binding Protein-1 (LBP-1) is located, probably related to the

	Protein	MW	Region bound	Consensus	Comments	Bending ability	References
BASAL PROMOTER	CTF/NF-I	52-66 Kd	+32 +52	AGCCAG		negative	Jones et al. 1988
	LBP-1/UBP-1	61-63 Kd	-13 +27	(T/A)CTGG			Jones et al. 1988, Wu et al. 1988
	TFII-I	120 Kd	-1 + 11 +33 + 45	YAYTCYYY			Roy et al. 1991
	TFIID	38 Kd	-42 -13	TATAA		positive	Starcich et al. 1985, Wu et al. 1988, Garcia et al. 1988
	Sp-1	100 Kd	-76 -67 -66 -57 -56 -45	G/TGGGCGG PuPuPy		positive	Jones et al. 1986, Harrich et al. 1989
ENHANCER	NF-κB	50 Kd + 65 Kd	-81 -90 -95 -104	GGGACTTTCC	constitutive in B cells, inducible in the others	positive	Lenardo et al. 1988, Kawakami et al. 1988
	EBP-1	57 Kd			active form of NF-κB		Clark et al. 1989, Clark et al.1990
	PRDII-BF1						Baldwin et al. 1990
	c-Rel, v-Rel						Wu et al. 1988
	H2TF-1				constitutive		Baldwin et al. 1988
	bcl 3						Bours et al. 1993
	p49	49 Kd					Schmid et al. 1991
NEGATIVE REGULATORY ELEMENT	TCF-1α	55 Kd	-123 -138				Waterman et al. 1990
	USF/MLTF	43 and 44 kD	-174 -152	CACPuTG	constitutive repressor	positive	Garcia et al. 1987, Giacca et al. 1992
	NFAT-1		-254 -216		inducible		Crabtree et al. 1989, Boise et al. 1993, Thompson et al.1992
	Myb		-312 -290	YAACG/TG	constitutive in T-cells, inducible in others		Dasgupta et al. 1990
	A1		-315 -265		negatively regulated by nef		Guy et al. 1990
	AP-1		-342 -349 -330 -338 -291 -299	G/CTGACTC/AA		positive	Franza et al. 1988
	site A protein		-379 -361				Orchard et al. 1990
	site B protein		-350 327	GGTCA(N)TGACC	constitutive, homologous to steroid receptor, repressor		Orchard et al. 1990

Figure 2b: human proteins which bind to the Long Terminal Repeats of HIV-1

upstream binding protein-1 (UBP-1) characterized in a different laboratory [14][15]; mutations of this element diminishes promoter strength. In addition, Kato and collaborators [16] reported a second binding site for LBP-1/UBP-1 overlapping the TATA box which, actually, represses transcription: therefore, the role of this factor on HIV-1 transcription is still unclear. Cellular transcription factor/nuclear factor I (CTF/NF-I) has also been demonstrated to bind downstream of the transcription start site [17].

The **enhancer** region mediates the transcriptional inducibility of the provirus upon a variety of stimuli which trigger cellular activation and proliferation [10][11][12]. This genetic element is composed of two repeats (κB sites), which are the targets of the inducible transcription factor NF-κB, a heterodimer composed of two subunits (p50 and p65) [18], that resides in the cytoplasm when complexed with the inhibitor IκB [19]. 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 p50 precursor and increase levels of active nuclear NF-κB complex [20]. 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 NF-κB recognition motif. These include p50 and p65 homodimers, other Rel family members such as v-Rel, c-Rel [21], and p49 [22], and other apparently unrelated factors such as H2TF1 [23], EBP-1 [24], PRDII-BF1[25]. Recently, the oncoprotein Bcl-3 was shown to be able to transactivate through κB motifs via association with DNA-binding p52 homodimers (a protein closely related to p50) [26]. The physiological role of each of these factors, possibly also in concert with the others, still needs to be clarified.

The **Negative Regulatory Element,** extending upstream of the enhancer site, is a cis-acting element whose overall function is to downregulate transcription, since its deletion results in a higher transcriptional rate [27][28]. Several cellular factors are reported to bind to this region.

Two sites from nucleotides -379 to -327, called site A and site B, are occupied in DNase I footprinting experiments using Jurkat nuclear extracts. Site B resembles the steroid/thyroid hormone response element. Mutations, that abolish binding to site B, resulted in an increased expression of HIV-1 LTR [29].

AP-1 is an inducible transcriptional factor composed of two subunits: Fos and Jun [30]. Three binding sites are present within the LTR but their role is still unclear.

Myb protein is a transcription factor inducible in mitogen stimulated peripheral blood lymphocytes and constitutively expressed in several CD4 T-cell lines. Introduction of purified Myb in HeLa cells harboring LTR-CAT, increases LTR mediated expression [31].

Nuclear factor of activated T cells (NFAT) binds to the NRE upon stimulation with mitogens [11]. It was recently demonstrated that it contains Fra-1 and JunB proteins in a heterodimeric form [32]. Anyway, by mutation analysis, this binding site seems to be dispensable both in response to mitogens and as a negative regulatory element [33].

Recombinant NF-IL6 binds at least in three different sites in the LTR (roughly around 169, 109 and 250 nucleotides upstream of the start site). It has been claimed that NF-IL6 is a negative regulator of transcription [34].

T Cell Factor-1α (TCF-1α) occupies the region between nucleotides -123 and -138. This 55 kDa transcription factor is specific for T lymphocytes [35]. The sequencing of its cDNA highlighted the similarity between its DNA binding domain and that of the HMG proteins. The function of TCF-1α in LTR-driven expression has not been demonstrated.

1.3 Transcription factor USF/MLTF

Upstream Stimulatory Factor (USF) or Major Late Transcription Factor (MLTF) is a nuclear protein initially characterized for its ability to bind to and activate the upstream element of the major late promoter of Adenovirus type 2 [36]. After extensive purification of nuclear extracts from HeLa cells, USF/MLTF activity turned out to be composed of two equimolar proteins of 43 and 44 kDa both able to bind independently to the consensus sequence CACGTG. The two proteins are probably the products of two different genes, since they are immunologically unrelated. 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 [37]. The basic region is believed to be directly involved in DNA binding, while both the helix-loop-helix and leucine zipper structures provide dimerization ability. These two dimerization moieties, are common to other transcriptional factors. The leucine zipper motif allows dimerization of many transcriptional factors (among which: GCN4, Fos and Jun proto-oncogenes) through the formation of a coiled coil between two parallel α helices in which the leucine residues, aligning on the same face of helix, provide a hydrophobic interface. The helix-loop-helix element is composed of two amphipatic helices, connected by a loop of variable length. It is present in transcription factors as TFEB [64], TFE3 [63], Max and Myc [65]. Anyway, in *in vitro* experiments, none of the above are able to dimerize with the 43 kDa form of USF/MLTF (USF^{43}). The only putative oligomerization partner is the 44 kDa form of USF/MLTF (USF^{44}) which has been recently cloned [38], and demonstrated to belong to the B-HLH-LZ family as well. *In vitro* experiments show that it is able to homodimerize with itself and to heterodimerize with USF^{43}.

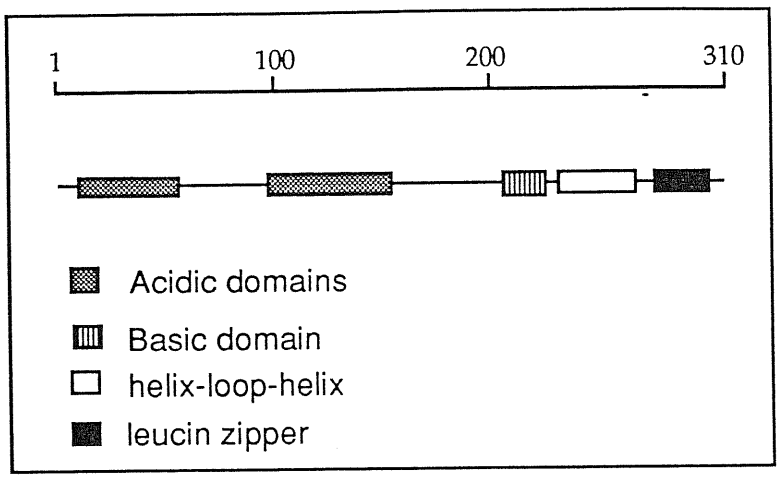


Figure 3: Stucture of USF [43]

Binding sites for b-HLH-Zip proteins and cognate binding activities are very conserved through evolution [39], both in viral promoters (Adenovirus Major Late Promoter upstream element, Human Cytomegalovirus 2.7 Kb RNA promoter) and cellular promoters (human growth hormone gene promoter, human heme oxygenase promoter, mouse metallothionein I gene promoter and many others) and we have described the presence of one of such sites in an origin of DNA replication of human chromosomes [40][41].

Recently, a laboratory reported that USF/MLTF is also able to activate transcription acting on initiator elements [42][43]. The ability of a single protein to bind and activate transcription from two unrelated sequences, namely the upstream hexanucleotide element and the initiator elements, is still controversial.

Our laboratory has previously reported that the CACGTG dyad symmetry motif, located between nucleotides -152 and -174 upstream of HIV-1 transcription start site, is recognized by a nuclear activity constitutively present, with the characteristics of binding specificity and thermostability reminiscent of transcription factor USF/MLTF. An *in vivo* footprinting assay set up in our lab has confirmed that this binding site is indeed occupied in living cells [44]. In transient transfection studies the substitution of the whole

Negative Regulatory Element of the HIV-1 LTR with an oligonucleotide containing a CACGTG target, still results in downregulation of transcription. This evidence suggests that USF/MLTF acts as a negative regulator of HIV-1 transcription; this is intriguing, since in all the promoters so far reported to be affected by USF/MLTF, it works as a positive regulator of transcription.

The work exposed in this thesis aims to demonstrate that both USF/MLTF purified from HeLa cells and recombinant USF^{43} bind to the CACGTG motif within the NRE of the LTR and, upon binding, bend DNA; moreover, some evidences will be presented about the role of USF^{43} in HIV-1 transcriptional control.

2.0 MATERIALS AND METHODS

2.1 Purification of USF/MLTF from HeLa cells

The procedure for the purification of USF/MLTF from HeLa cells has already been reported [45]. Briefly, a nuclear extract was precipitated by ammonium sulfate; the precipitated fraction of proteins was subsequently purified through a 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 USF/MLTF binding 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 all the purification by gel retardation and south-western assays. In the DNA binding experiments presented in this thesis the active fractions after the MonoQ column were used.

2.2 Purification of Ku antigen from HeLa cells

Ku antigen was purified via a combination of ion-exchange and affinity chromatographic techniques; after the last fractionation, Ku turned out to be purified to homogeneity as demonstrated by silver staining of SDS-PAGE gels [Gutierrez et al. manuscript in preparation]. Ku protein is composed of two subunits of 70 and 90 kDa and its binding specificity is very loose. The protein used in circular permutation assays was the one eluted from the affinity chromatography step.

2.3 Cloning, expression and purification of GST-USF^{43}

The plasmid pGST-USF^{43} was constructed by cloning the coding region of the 43-kDa form of USF/MLTF (USF^{43}) in the commercial vector pGEX2T (Pharmacia). This plasmid expresses the glutathione-S-transferase (GST) in prokaryotes, under the control of a promoter inducible by isopropyl-galacto-pyranoside (IPTG). The coding

sequence of USF/MLTF was rescued by PCR from the plasmid pBSA1USF^{43}, containing the coding sequence of USF^{43} under the control of the human A1 gene promoter. The plasmid was a kind gift of Giuseppe Biamonti. The primers for amplification (RUSF1 and RUSF2, see Table I) were designed in order to generate a DNA fragment ending with the restriction sites for *Eco*R I and *Bam*H I to allow oriented cloning in the vector. Amplification was carried out in 50 µl of a solution containing 10 mM Tris (pH 8.0), 50 mM KCl , 1.5 mM $MgCl_2$, 0.01% gelatin, 200 µM each dNTP, 1 µM each primer, 1 ng of template plasmid DNA and 2.5 units of *Taq* DNA polymerase (Perkin Elmer), 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 recombinant plasmid carries a continuous open reading frame from GST to USF^{43} in order to express the fusion protein GST-USF^{43} (see Figure 4)

A colony of the SF8 strain of *E. coli* transformed by pGST-USF^{43} was grown overnight in 100 ml of Terrific Broth [46] containing 50 µg/ml of ampicillin at 30°C. After an overnight incubation, 900 ml of the same medium were added and the culture was allowed to grow until it reached the optical density of 0.6-0.8 at 600 nm; then, protein expression was induced by adding IPTG (Sigma, 1 mM final concentration). After additional 3 to 5 hours incubation, the culture was centrifuged at 5000 x g at 4°C, and the bacterial pellet was resuspended in 10 ml of cold PBS containing 4 mM DTT. Proteins were extracted by 3 cycles of sonication of 20 sec in ice. 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 in the cold room on a rotating wheel for 1 hour. After this period, the suspension was loaded on an empty plastic column (BioRad) to let unbound proteins pass through, and the beads were subsequently washed with 50 ml of PBS containing 4 mM DTT. Finally, the polypeptide was eluted in 1 ml PBS containing 4 mM DTT and 20 mM free glutathione (Sigma). Usually, the yield was that of 2-3 mg of protein per liter of medium.

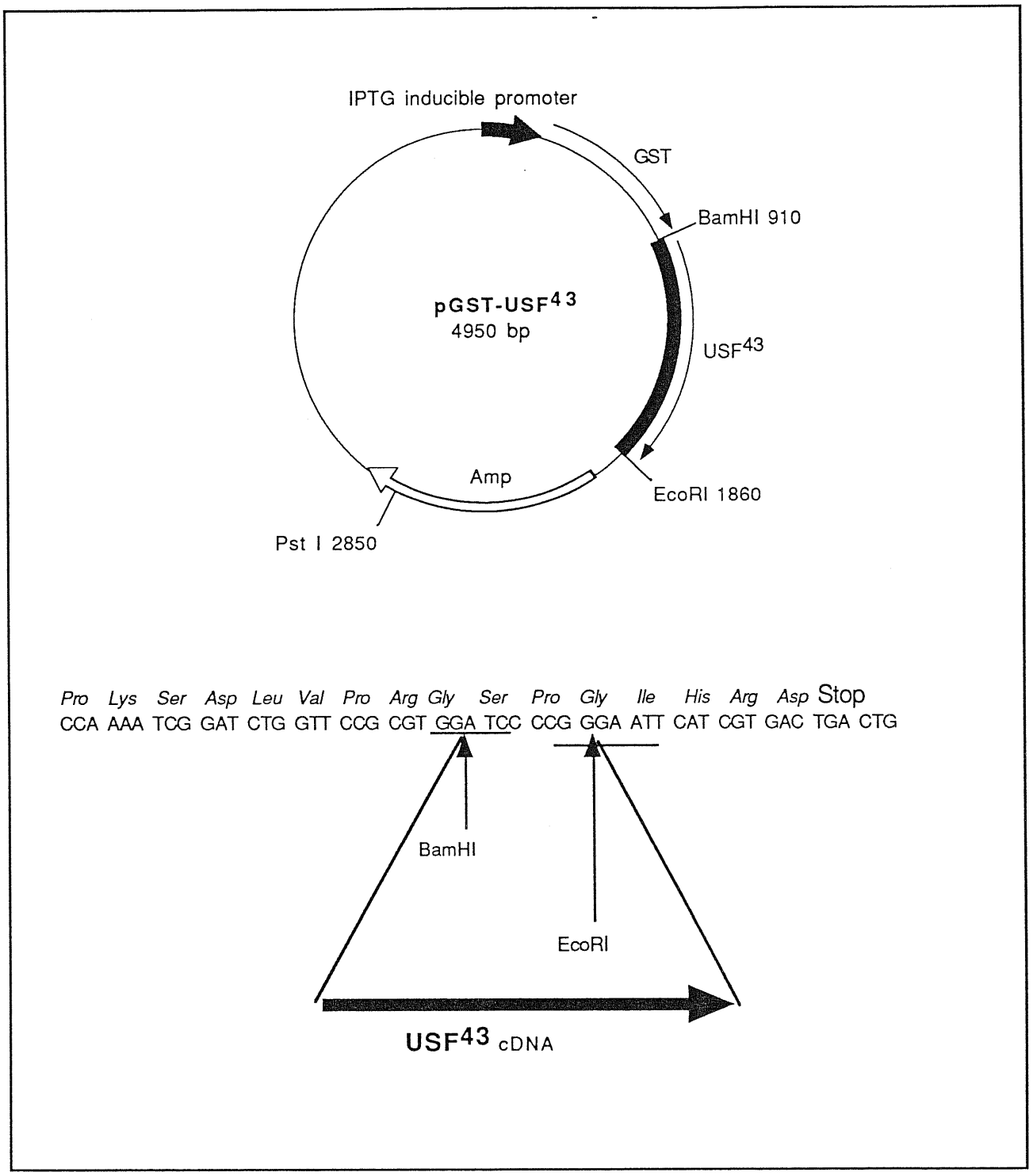


Figure 4: Cloning of the USF^{43} cDNA coding region downstream of the GST gene.

The purity and integrity of the protein was checked by SDS-PAGE and Coomassie staining.

2.4 Western blotting

15 ng of protein were resolved on a 12% SDS-PAGE and transferred to a nitrocellulose filter by electroblotting. The filter was incubated in 10% milk-TBS buffer (10% w/v non-fat dried milk in 125 mM NaCl, 10 mM Tris-HCl, pH 7.4) for 1 hr at 37°C. Incubation with antibodies against full-length rUSF^{43} (kindly supplied by R.G. Roeder) 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 diluted 1:1000, for 1 hr at room temperature. After several washes, bound antibodies were revealed using BCIP/NBT color development solution (BioRad).

2.5 Gel retardation assays and competition analysis

Gel retardation assays were carried out by the incubation of end-labelled DNA probes (10^4 cpm) with 1-5 ng of purified GST-USF^{43} protein, or 2 µl of the purified HeLa USF/MLTF (Mono Q fractions [45]) 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 a final volume of 10 µl. After 20 minutes of incubation at 30° C, samples were resolved by 5% polyacrylamide gel electrophoresis in 0.5 x TBE. Under these conditions, 1 ng of GST-USF^{43} was enough to generate a clear retarded complex visible after an overnight exposure, with an estimated molar DNA:protein ratio of 1:5.6.

Competition experiments were carried out by mixing a 30 fold excess of cold oligonucleotides to the probe before the addition of the protein. The sequence of the oligonucleotides utilized, is reported in Table I.

Source	Competitor sequence	Name
Plasmid pB48 (nucleotides 781-804), including B48 binding site	5'-CCGGTCGCATCACGTGACGAAG-3' 3'-CCAGCGTAGTGCACTGCTTCTC-5'	B48BS
Ad2 Major Late Promoter upstream element	5'-GGTGTAGGCCACGTGACCGGGT-3' 3'-ACATCCGGTGCACTGGCCCACA-5'	AdMLP
HIV-1 LTR, from nucleotides -174 to -151 upstream of transcription start site	5'-GCATTTCATCACGTGGCCCGAG-3' 3'-TAAAGTAGTGCACCGGGCTCTC-5'	HIV
Same as B48BS, with the CACGTG site mutated in CATATG	5'-CCGGTCGCATCA**TA**TGACGAAG-3' 3'-CCAGCGTAGT**AT**ACTGCTTCTC-5'	MUT-III
Same as B48BS, with a methyl group on the indicated citosine	5'-CCGGTCGCATCA^{CH_3}CGTGACGAAG-3' 3'-CCAGCGTAGTGCACTGCTTCTC-5' $_{CH_3}$	MET-I
Same as B48BS, with a methyl group on the indicated citosine	5'-CCGGTCGCATCACGTGA^{CH_3}CGAAG-3' 3'-CCAGCGTAGTGCACTGCTTCTC-5' $_{CH_3}$	MET-III
Two HIV-1 Sp1 binding sites	5'-CTGGGCGGGACTGGGGAGTGGCG-3' 3'-GACCCGCCCTGACCCCTCACCGC-5'	SP1
Aim	**Oligonucleotide sequence**	**Name**
PCR amplification of the 200 N-ter aa of USF^{43}, from pBSA1USF^{43} and cloning into pM1	5'-CGGGAATTCAAGGGGCAGCAG-3'	Gusf 1
	5'-CGGAAGCTTCTCTCATCCCGAG-3'	Gusf 2
Costruction and cloning of four Gal4 binding sites upstream of the NRE in pLTRΔNRE-CAT	5'-CGGATCC(CGGAAGACTCTCCTCCG)x2 G-3'	GAL4-US
	5'-AATTC(CGGAAGACTCTCCTCCG)x2 GGTAC-3'	GAL4-UD
	5'-AATTC(CGGAGGAGGAGAGTCTTCCG)X2 GGATCCGGTAC-3'	GAL4-LS
	5'-C(CGGAGGAGAGTCTTCCG)X2 G-3'	GAL4-LD
PCR amplification of USF^{43}, from pBSA1USF^{43} and cloning into pGEX2T	5'CGAGGATCCAAGGGGCAGCAGAAAACA-3'	RUSF1
	5'-GCTGAATTCCTTAGTTGCTGTCATTCTTGATGAC-3'	RUSF2

Table I: sequences of the oligonucleotides used in gel retardation assays and PCR amplifications

Super-shifting of the retarded complex was obtained by adding 1 µl of a 1:10 dilution of immune serum, before the addition of the protein to the test tube.

2.6 DNase I footprinting

The probe for DNase I footprinting was generated by PCR amplification of the region of the HIV-1 LTR between nucleotides -285 and -112, encompassing the binding site for USF/MLTF, from plasmid pLTRCAT [47]. Primers for amplifications were FOOTP (Table II) and USF1 (5'-AGCAAGCTCGATGTCAGCAGTTCTT-3'); the latter was previously end-labeled with [γ^{32}P]-ATP and T4 polynucleotide kinase, in order to generate a DNA fragment asymmetrically labelled on one strand. The conditions for amplification were the same as reported [47].

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 used for the gel retardation assay. After 20 min 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 was added to a final concentration of 6 ng/ml . After 1 min, 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, redissolved, denatured and loaded on a 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/ml.

A G+A ladder of the probe, produced according to the Maxam and Gilbert chemical cleavage method [48], was loaded on the same gel to align the DNase I digestion products.

2.7 Methylation protection

Methylation protection by dimetyl sulfate (DMS) was carried out essentially as described by A. G. Papavassiliou [49]. Briefly, following electrophoresis of a 5-fold

scaled-up binding reaction, 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 [48]. The products were resolved on a sequencing gel in parallel with a G+A ladder.

2.8 Preparation of probes for gel retardation and circular permutation assays

All the oligonucleotides were synthesized by the ICGEB Oligonucleotide Synthesis Service on an Applied Biosystem 380B synthesizer; their sequences are reported in Tables I and II.

For the gel retardation assays, the oligonucleotides were purified from denaturing polyacrylamide gels and annealed. End-labeling with [γ^{32}P]-ATP and T4 polynucleotide kinase was carried out following standard procedures [46].

For the circular permutation assays, a set of six probes (B1 to B6, Figure 9) was generated by PCR using the pLTRCAT plasmid [47] as template. The sequence of the oligonucleotides employed in the amplification (C1 and USF III for probe B1, C2 and BD1 for probe B2, FOOTP and SP2 for probe B3, NFAT1 and NEWPA for probe B4, NRE2A and A3 for probe B5, and HIV1 and BD2 for probe B6) are shown in Table II. Amplifications were carried out in 50 µl of a solution containing 10 mM Tris (pH 8.0), 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin, 200 µM each dNTP, 0.1 µM both primers, 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 (Perkin Elmer), with 35 repetitions of the following cycle: 45 sec at 94°C, 45 sec at 60°C, and 45 sec at 72°C, in a Perkin-Elmer 480 Thermal Cycler. PCR products were resolved on a 5% polyacrylamide native gel and eluted for few hours at 65°C in Tris 10 mM, EDTA 0.1 mM.

Name	Oligonucleotide Sequence	Probe
C1 USF III	5'-AGGCTACTTCCCTGATTAGCAGAA-3' 5'-GGATGCAGCTCTCGGCCACGTGATG-3'	B1
C2 BD1	5'-TAGCAGAACTACACACCAGGGCCAG-3' 5'-TTCTTGAAGTACTCCGGATG-3'	B2
FOOTP SP2	5'-GCAAGCTTGAAGAGGCCAAT-3' 5'-CGCCACTCCCCAGTCCCGCCCCAG-3'	B3
NFAT1 NEWPA	5'-CTTGTTACACCTGTGAGCCTGCATGGATGGATGACC-3' 5'-CAAAAAGCAGCTGCTTAT-3'	B4
NRE2A A3	5'-TAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATGGATC-3' 5'-GCCAGAGAGCTCCCAGGCTCAGATCTG-3'	B5
HIV1 BD2	5'-GCATTTCATCACGTGGCCCGAG-3' 5'-GGCTTAAGCAGTGGGTTCC-3'	B6

Table II: Sequence of the oligonucleotides used for the PCR amplification of the probes for circular permutation assays

2.9 Calculation of the bending angle

Although a clear theory for calculating protein-induced bend angles is lacking, the bend angle α can be estimated from gel mobilities determined during circular permutation analysis. The experimental equation $\mu_M/\mu_E=\cos(\alpha/2)$ is broadly used [50] where μ_M is the mobility of the complex with the protein bound in the middle of the fragment and μ_E is the mobility with the protein bound at the end. The value calculated in this way is that of the angle from which the segment of DNA depart from linearity.

2.10 Plasmid constructs

pLTRCAT and pLTRCATΔNRE are pBlueScript KS (Stratagene, La Jolla CA, USA)-derivatives containing the whole HIV-1 LTR and the LTR deleted upstream of position -159, respectively. These constructs were cloned upstream of the CAT gene and its polyadenilation sites in plasmid pULB3574, a gift of Dr. Pierre Spegelaere (ULB, Bruxelles).

pBSA1USF^{43} bears the coding region (930 nt) of USF^{43} cloned downstream the A1 promoter, a constitutively strong human promoter. It is a gift of Dr. Giuseppe Biamonti (IGBE, Pavia).

pM1 contains the sequence coding for the 147 amino-terminal aminoacids of the yeast protein Gal4, responsible for DNA binding, under the control of the early SV40 promoter; downstream of Gal4 sequence, a polylinker site was cloned in order to allow in-frame cloning. It is a gift of Dr. I. Sadowsky (University of British Columbia, Vancouver).

pGal4-USF^{43} was constructed by cloning the DNA sequence coding for the 200 N-terminal aminoacids of USF^{43} between *Eco*RI and *Hind*III sites of plasmid pM1. The sequence of USF^{43} was amplified by PCR from pBSA1USF^{43}. The oligonucleotides used in the amplification (Gusf1 and Gusf2, see Table I) contain an *Eco*RI e *Hind*III site, respectively, at their 5' ends, to allow an oriented and in-frame cloning. The final construct was checked by the determination of its sequence.

pGAL4-VP16 contains the sequence coding for the 147 amino-terminal aminoacids of the yeast protein Gal4 responsible for DNA binding, fused to the sequence encoding for the 78 carboxy-terminal aminoacids of VP16, a strong acidic transcriptional activator, under the control of the early SV40 promoter. It is a kind gift of Dr. S.J. Triezemberg (Michigan State University).

pRSV-LUC is a commercial plasmid (Promega) which expresses the firefly luciferase enzyme under the control of the Rous Sarcoma Virus promoter

pLTRΔNRE-Gal4-CAT is a derivative of pLTRΔNRE in which a 86 bp-long oligonucleotide, containing four binding sites for Gal4, was cloned upstream the LTR in a *Kpn*I site.

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

2.11 Cell cultures

HeLa and HL3T1 (a HeLa derivative in which the LTR-CAT construct was integrated into the host genome) cells , were cultured in Dulbecco's modification of Eagle's minimal essential medium (DMEM, GIBCO) in monolayer cultures. All the cultures were additioned of 50 μg/ml gentamicin, 10% fetal calf serum, following scomplementation in the case of HL3T1 cells, and 2 mM glutamine. The HL3T1 cell line is a kind gift of Prof. P. Amati.

2.12 Transient transfection assays: transfection, standardization by luciferase assay and CAT assay

Transfection: Hela cells were transfected by a modification of the DEAE-dextran method. 1.5 million cells were plated in a 100 mm Petri dish; 24 hours later, Transfection Solution (TS) was freshly-prepared, containing 2 μg of pLTRCAT plasmid,

or its derivatives, 5 μg of pSV2-tat, and, eventually, the indicated amount of plasmid acting in trans on the LTR. To normalize between groups, co-transfection was carried out with 500 ng of a second independent indicator plasmid: pRSV-LUC, expressing the firefly luciferase and a non-coding plasmid was used to even the amount of DNA in each transfection. The volume was taken up to 900 μl 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). Therefore, after having washed the plates twice in 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°, TS was removed and the culture medium was added to the cells. 48 hours after transfection, cells were scraped off of the plates, pelleted and carefully resuspended in 400 μl 250 mM Tris-HCl, pH 7.8. Two 200 μl aliquots were separately assayed for lucyferase and CAT activity. This is to overcome the incompatibility between the types of lysis required for the two assays (in fact, the SDS, which is present in the lysis buffer used in the luciferase assay, inhibits 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) and, after centrifugation, 10 μl of cell lysate supernatant were mixed with 50 μl of luciferase assay buffer (Promega) containing the substrate luciferyl-CoA. Photon emission was measured starting from the second minute every 12 seconds using a LS 1801 liquid scintillation counter (Beckman) equipped with single photon monitor software; the mean value over a 3 minute interval was considered. In parallel, protein concentrations of the cell lysate were determined using a Bradford assay [46]. The ratio between photon emission and optical density (O.D.), was an estimate of transfection efficiency expressed as counts per O.D. units.

CAT assay: According to a modified version of the Gorman protocol [51], protein extract is obtained by three cycles of freeze-and-thaw (3 minutes in dry ice-ethanol followed by 3 minutes at 37°). After pelleting debris, the O.D. of the supernatant

was determined as above. The amount of cell extract to be assayed for CAT activity was a function of O.D. times transfection efficiency. Then, samples were incubated for times ranging from30 minutes to 4 hours at 37° with 2.5 µl ^{14}C-1-deoxychloramphenicol (50 mCi/mmol Amersham), 70 µl 1M Tris-HCl pH 7.8 and 20 µl 4 mM Acetyl-CoA in a final volume of 140 µl. The reaction was stopped extracting chloramphenicol with 1 ml of ethylacetate and, following lyophilization, samples were resuspended in 15 µl ethylacetate, spotted on TLC silica gel plates and developed by thin layer 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 Cerenkov counting in the scintillation counter.

Transfection experiments were repeated at least three times and the mean value considered.

Standardization was skipped when transfecting HL3T1 cells; in this case Tat protein was provided to enhance LTR activity by supplementing the medium containing 100 µM chloroquine (Sigma) with 2 µg of recombinant GST-Tat expressed and purified in our lab (manuscript in preparation). In these conditions, GST-Tat protein was able to enter into the cells and activate the LTR-driven transcription.

3.0 RESULTS

3.1 Cloning, expression and purification of recombinant USF^{43}

The USF/MLTF binding activity purified from HeLa cells is composed of two proteins of 43 and 44 kDa, both able to bind DNA [52][45]. The 43 kDa form of USF/MLTF (USF^{43}) was cloned and sequenced [53], and shown to belong to the basic-helix-loop-helix-leucine zipper class of transcription factors.

In order to study in further details 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 GST. The coding region of USF^{43} was cloned in plasmid pGEX2T, downstream of the GST gene, maintaining the same open reading frame (see Figure 4). The resulting plasmid, pGST-USF^{43}, expresses the USF^{43} protein as an extension of the carboxy -terminus of GST under the control of a promoter inducible by IPTG. Panel A of Figure 5 shows the results of a single-step purification procedure for GST-USF^{43}.

3.2 Western blot analysis

The identity of the purified protein was further confirmed by western blot analysis using antibodies raised against recombinant USF^{43} (a kind gift of R. Roeder). As shown in panel B of Figure 5, lane 1, 15 ng of pure GST-USF^{43} can be easily detected by antibodies, while the same amount of GST protein is not (lane 2).

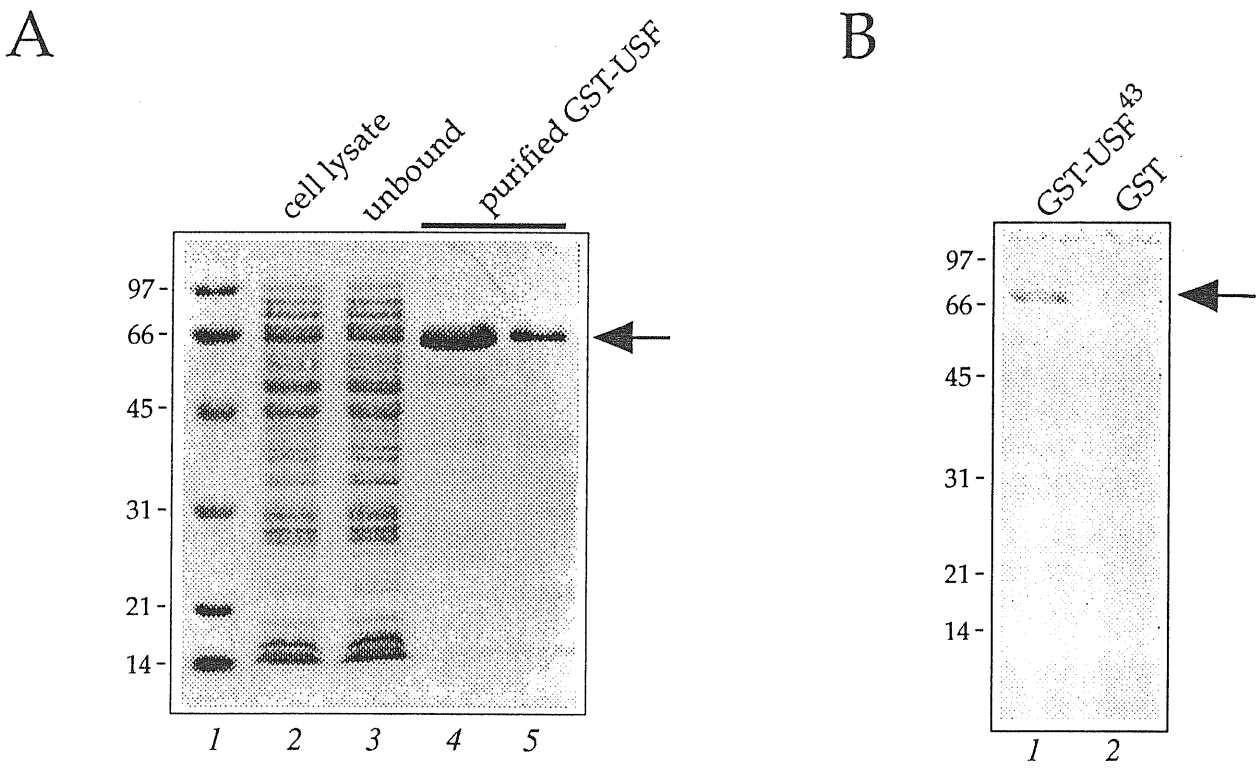


Figure 5: Panel A: purification of GST-USF resolved by a SDS PAGE. Lane 1: protein molecular weight markers. Lane 2: crude cell lysate. Lane 3: flow through of the glutathione cross-linked agarose beads column. Lane 4 and 5: purified GST-USF eluted from the column.
Panel B: Western Blot analysis of purified GST-USF using antibodies raised against USF43. Lane 1: GST-USF. Lane 2: GST.

3.3 Binding activity of recombinant and purified USF/MLTF in gel retardation assay

Our laboratory reported that in crude nuclear extracts of human cells, a binding activity exists for a sequence of the Negative Regulatory Element of the HIV-1 LTR; this same binding activity also interacts with several other targets, including a human DNA origin of replication, the Adenovirus Major Late Promoter Upstream Element (MLP-UE), a yeast centromere, and the promoters of other mammalian genes [39][47]. In this new set of experiments, we have now challenged some of these sequences for binding to purified USF/MLTF and recombinant GST-USF^{43}.

An end-labeled oligonucleotide containing the CACGTG box of the HIV-1 LTR (oligo HIV, see Table I) forms a specific complex with GST-USF^{43} in a gel retardation assay (Figure 6, panel A, lane 1). The specificity of this complex was challenged by the addition of a 30 fold excess of several other cold oligonucleotides to the binding reaction. Competition was obtained with the same cold HIV oligonucleotide (lane 2), and with oligonucleotides corresponding to the Adenovirus MLP-UE (lane 3) and to the human DNA origin B48 (lane 4), all of which contain the same CACGTG core sequence, but not with an oligonucleotide corresponding to the Sp1 region of the LTR (lane 7). Mutation of the core CG dinucleotide to TA completely abolishes binding to the competitor (lane 6). Cytosine methylation on both strands of a distal CG dinucleotide of the binding site does not affect binding of GST-USF^{43}. These data fully agree to those reported about the behavior of HeLa USF/MLTF[39].

Binding specificity was further tested by the addition to the binding reactions of serum obtained before and after immunization against recombinant USF^{43}. Pre-immune serum was not effective (lane 8), while the immune one caused the retarded band to supershift to the well (lane 9).

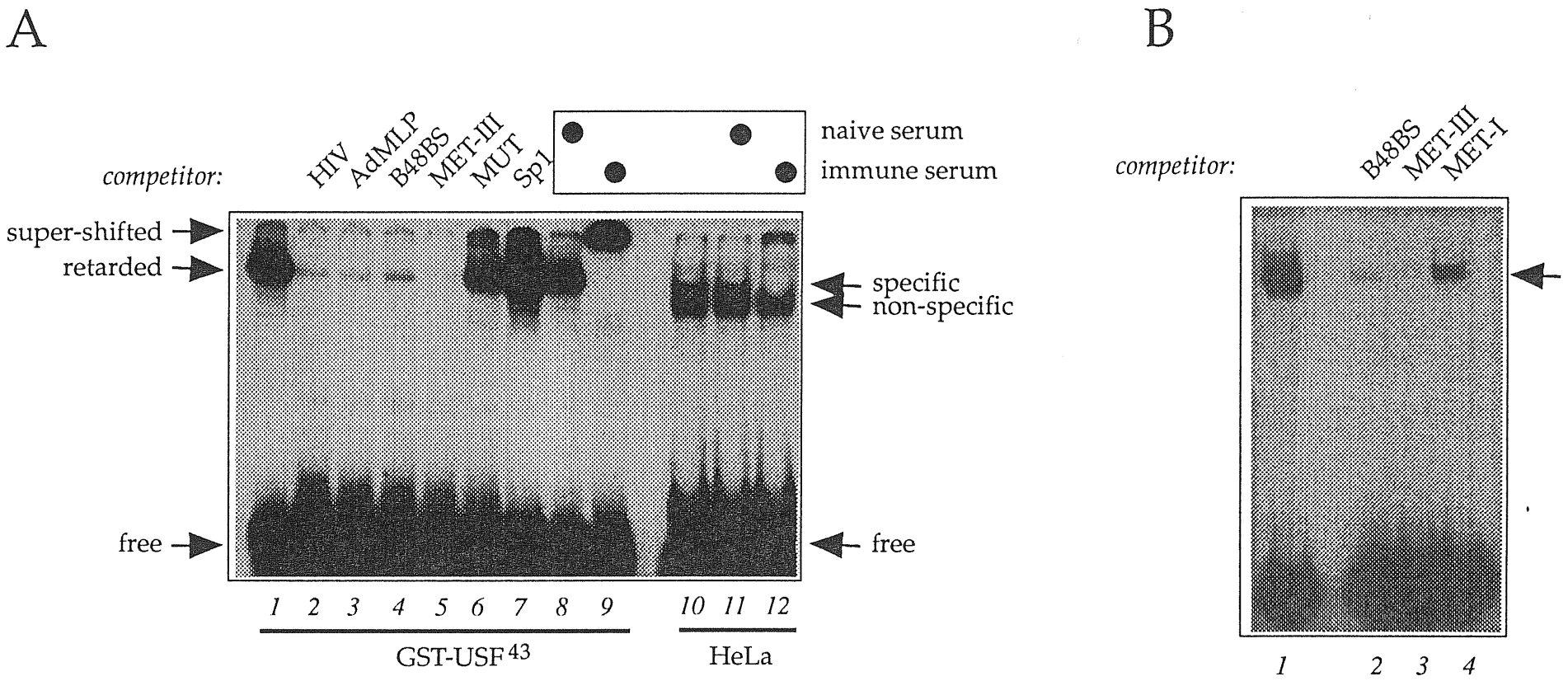


Figure 6: panels A and B: gel retardation assays and competitions using the HIV oligonucleotide probe (see sequences in Table I) with GST-USF (lanes 1-9 of panel A and 1-4 of panel B) or crude HeLa nuclear extract (lanes 10-12 of panel A); the experiments are detailed in paragraph 3.3.

Binding of Hela USF/MTF to the HIV-1 LTR template was also tested by gel retardation assays (Figure 6, panel B). In this case, two retarded complexes are detected (lane 10), but only the upper one can be supershifted by the immune serum, while the lower is unaffected, suggesting that the antibody reactivity is specific, as expected, and that, using crude nuclear extracts, only the upper band correspond to the USF/MLTF complex.

We already reported that methylation of the cytosine of the central CpG dinucleotide in the CACGTG binding motif inhibits the binding of HeLa USF/MLTF [39]. In order to check if GST-USF^{43} behaves in the same manner, we challenged the DNA/GST-USF^{43} complex (Figure 6, panel B, lane 1) with a 30-fold excess of the B48 oligonucleotide methylated at different positions (see the sequences in Table 1). The B48 oligonucleotide in its unmethylated form (lane 2) competes efficiently, resulting in the disappearance of the retarded band; methylation on both strands of a CpG dinucleotide five base pairs downstream the axis of symmetry of the binding site (lane 3) does not seem to affect binding, while methylation in the core of the CACGTG motif significantly affects the challenging ability of the competitor (lane 4).

3.4 DNase I footprinting

Further details of the interactions of purified USF/MLTF and the HIV-1 LTR were obtained by a DNase I footprinting assay. As shown in Figure 7, the incubation of the purified factor with a LTR probe in which the lower strand was 5' labeled (lane 3) prevents DNase I cleavage of a box centered around the CACGTG sequence (nucleotides from -173 to -158 upstream of the transcription start site) (see in comparison the pattern produced by DNase I digestion in the absence of protein in lane 2) . Upon protein-DNA binding, DNase I hypersensitive nucleotides appear at the 3'-end of the protected area. More precisely purins G_{-175} and A_{-178}, G_{-179}, G_{-180} are dramatically affected by protein binding. These DNase I hypersensitive sites are

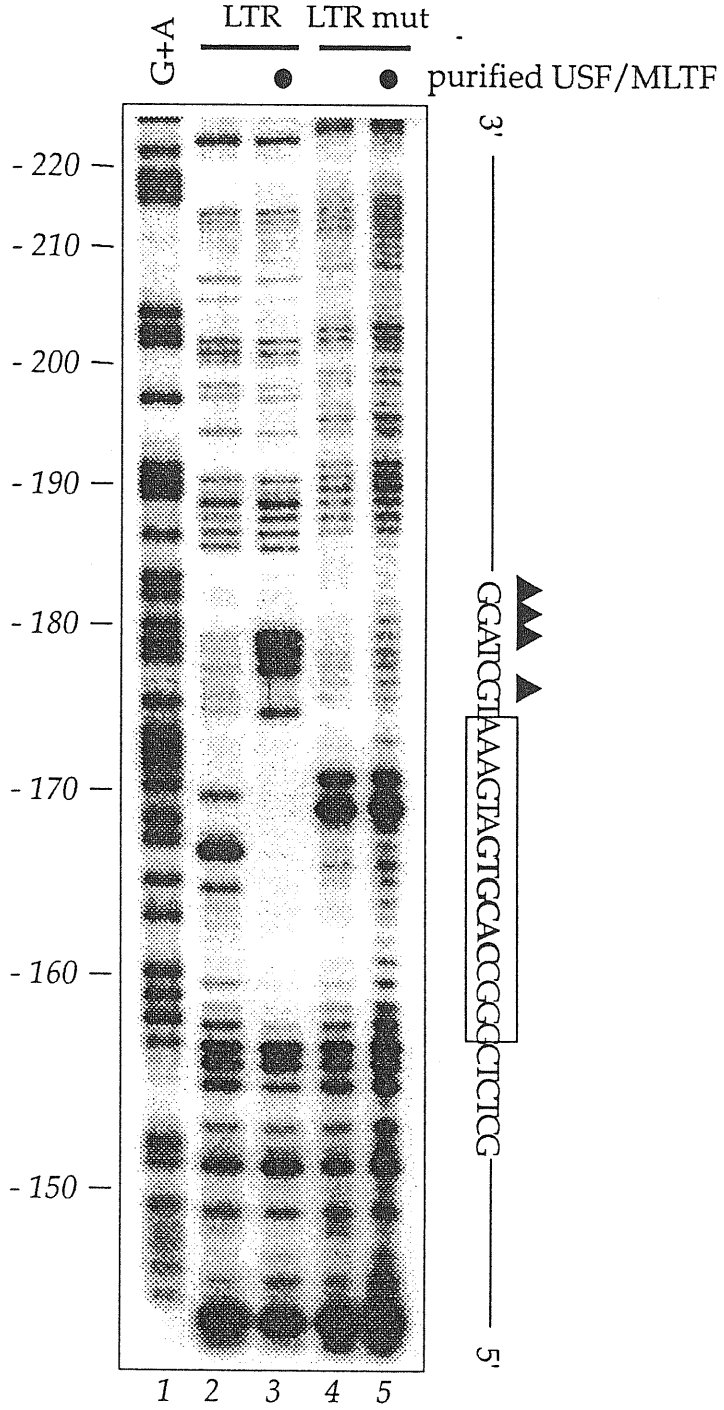


Figure 7: DNase I footprinting experiment using USF/MLTF purified from HeLa cells. Lane 1: G+A ladder. Lanes 2 and 3: DNA fragment pattern generated by DNase I treatment after incubation of an end-labelled probe in absence (lane 2) or in presence (lane 3) of USF/MLTF. Lane 4 and 5: the same as lanes 2 and 3, but using a DNA probe in which the binding site for USF/MLTF was mutated into a EcoRI site.

probably due to a structural alteration of the double helix upon protein binding, resulting in a widened DNA groove where nucleotides are more accessible to DNase I digestion.

Lane 4 and 5 show the results of DNase I digestion using a probe in which the CACGTG motif was mutated into an *Eco*RI site (GAATTC): protein binding does not occur anymore and DNA fragments pattern is identical in both lanes.

3.5 Methylation protection

Methylation protection by dimethyl sulfate (DMS) is a powerful chemical method to probe DNA-protein contacts. It allows the determination of purins protected from methylation by the DNA bound protein and indicates a close proximity of proteins to the DNA. We utilized this technique to investigate GST-USF[43] protein interactions with HIV-1 LTR. Panel A of figure 8 shows the results for the lower strand: purins A_{-164}, G_{-166} and G_{-168} are protected by DMS alkylation, suggesting that these bases are tightly in contact with GST-USF[43]; no hypersensitive sites are detected. Panel B is dedicated to the investigation of the upper strand: guanines G_{-162} and G_{-164} are clearly protected, while G_{-161} is a clearly hypersensitive to DMS.

In summary, DNase I footprinting and methylation protection assays demonstrate that the area centered on the CACGTG motif is clearly protected; hypersensitive sites have been detected by both techniques, indicating that DNA undergoes a structural distortion upon protein binding.

3.6 Recombinant and purified USF/MLTF bend the Long Terminal Repeats of HIV-1

Several transcription factors are known to bend DNA upon binding to their target sequences. Circular permutation assay is the simplest approach to detect DNA bending caused by protein interaction. In order to exploit this technique, a set of six probes (B1 to B6) was generated by PCR using the LTR as template. These probes have all

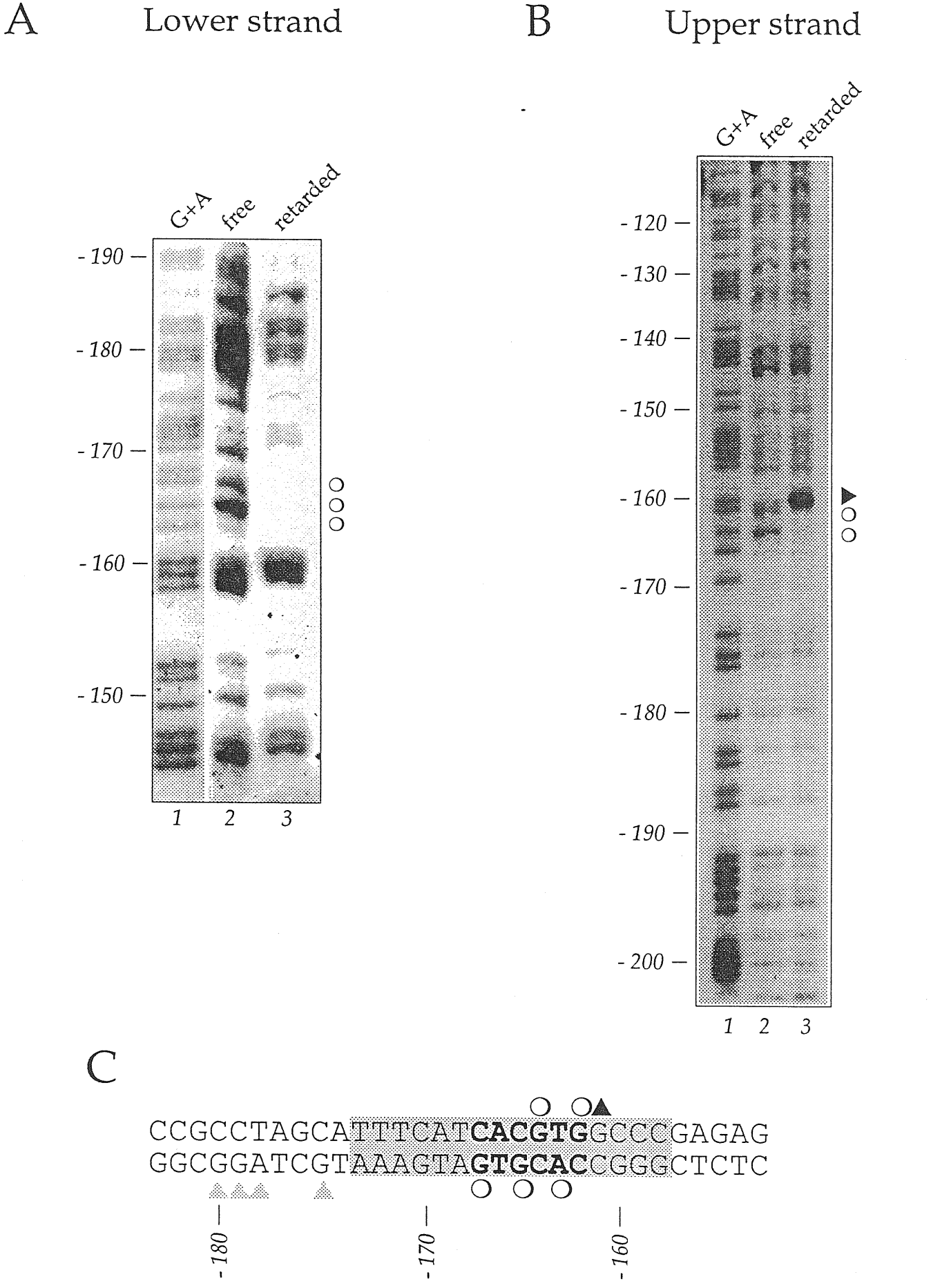
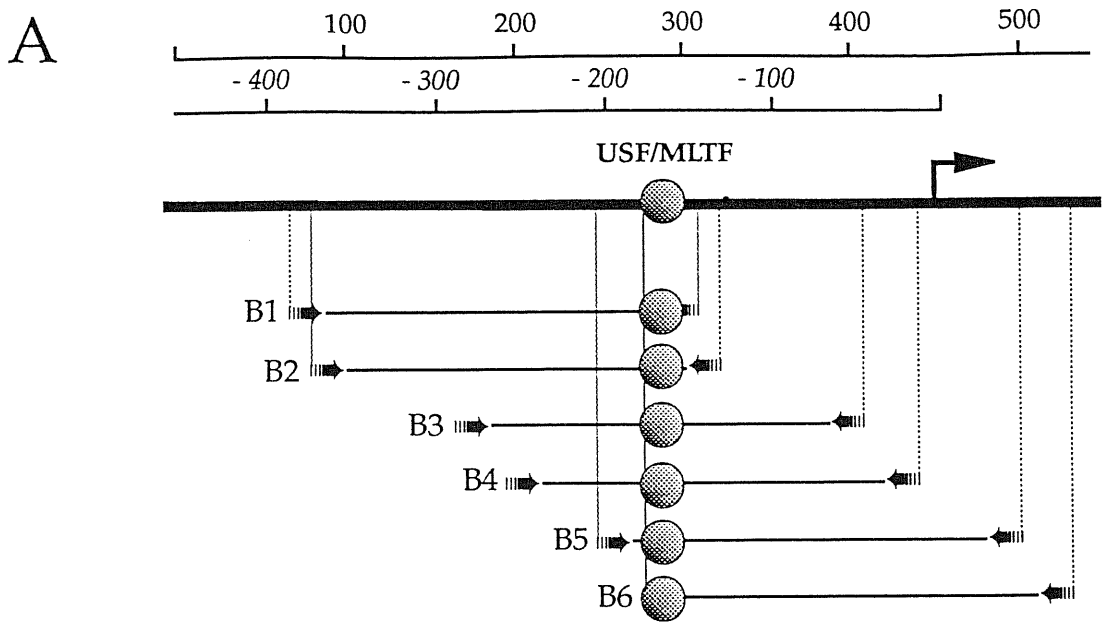


Figure 8: methylation protection assays of the lower (panel A) or the upper (panel B) strand (see paragraph 3.5 for the details). Panel C summarizes the data from DNase I footprinting and methylation protection experiments; empty circles indicate protected nucleotides, triangles indicate hypersensitive bases, the shaded area represents the region of DNA which results protected in the DNase I footprinting experiment.



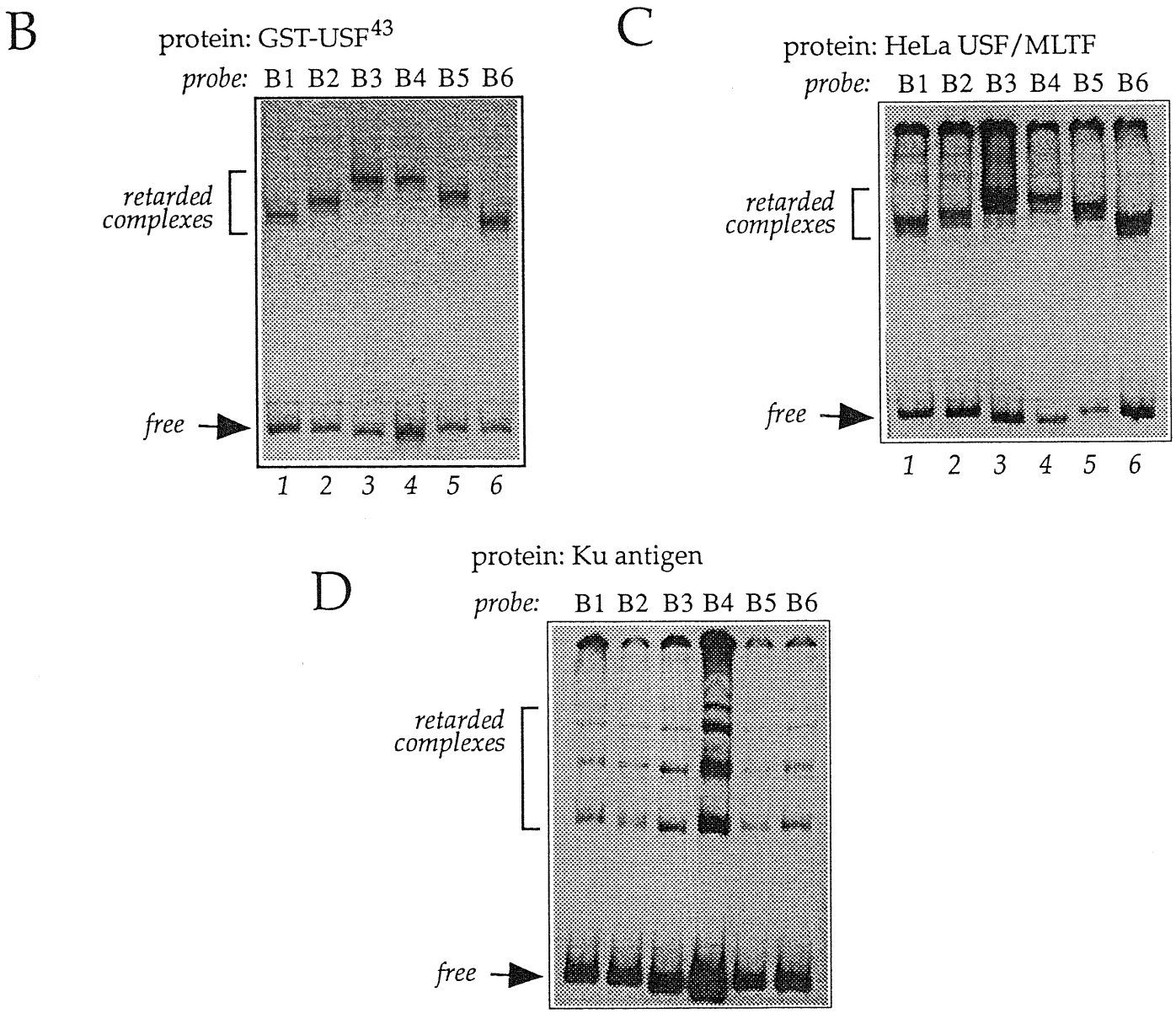


Figure 9: Panel A: position of the binding site for USF/MLTF within the probes used in the circular permutation assays. Panels B, C and D circular permutation assays using GST-USF, purified USF/MLTF and KU respectively.

approximately the same length (from 245 to 247 bp), but differ for the position of the binding site for USF/MLTF (Figure 9, panel A); they were tested in gel retardation assays both with GST-USF^{43} and with purified USF/MLTF (panel B and C, respectively). As expected, all the probes give 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 are more influenced than those of the probes containing the binding site at one extremity. This indicates that the overall distortion of the linear structure of the DNA when the protein binds near one end, is lower than that obtained by protein binding at the center of the molecule.

Estimation of the bending angle according to Thompson and Landy [54] indicates values of 90 degrees of bending for the purified protein and 110 degrees for the recombinant one.

The DNA binding protein Ku was also tested for its ability to bend DNA in a circular permutation assay with the same probes: as shown in panel D, all the six probes employed produce the same pattern of retarded bands demonstrating the inability of Ku antigen to bend DNA and providing a useful negative control for the circular permutation assay.

3.7 Function of USF^{43} *in vivo*

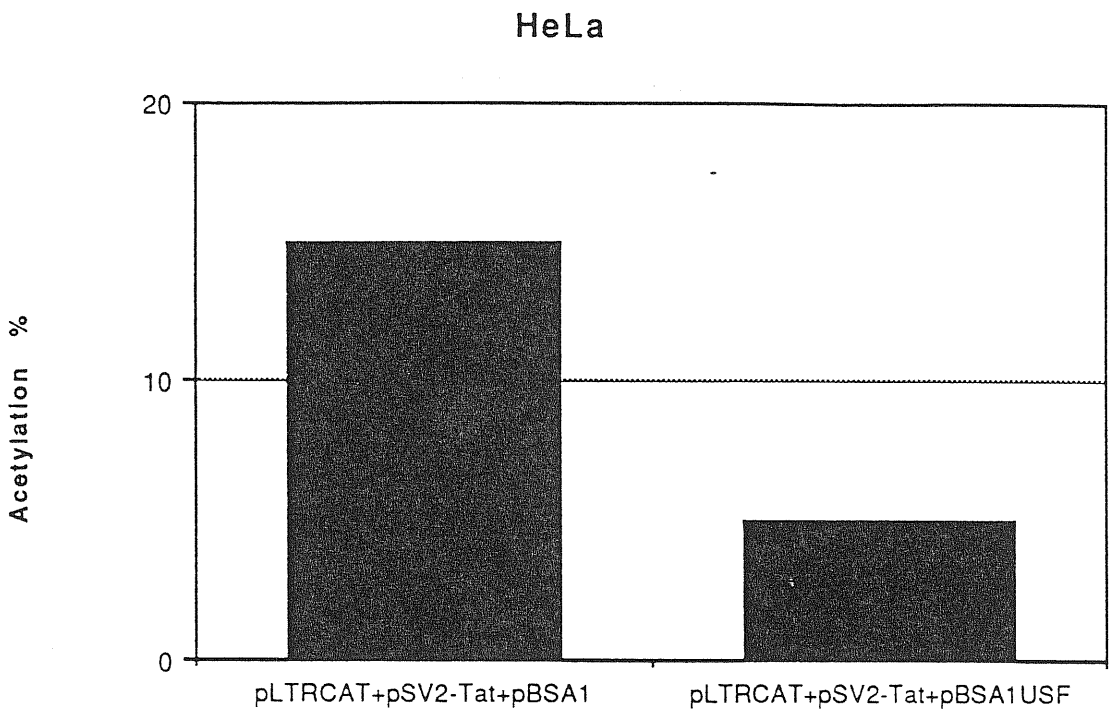
The above *in vitro* studies provided compelling evidences that USF/MLTF binds specifically to the CACGTG hexanucleotide in the Negative Regulatory Element (NRE) of the LTR. Previously, we already investigated the role of this element in the transcriptional control of HIV-1 [47]. In transient transfection experiments, the cloning of a binding site for USF/MLTF in place of the NRE, resulted in decreased transcriptional strength of LTR-driven expression. This evidence suggested that a nuclear factor was able to recognize this element and to exert its negative activity when bound to the LTR. In this regard, we reasoned that

the USF/MLTF transcription factor could be a good candidate. In order to verify this hypothesis, we overexpressed the 43 kDa form of USF/MLTF by transfection of a plasmid carrying the USF^{43} coding region under the control of a strong promoter in HL3T1 cells. This cell line is a HeLa-derivative in which the LTRCAT construct is integrated into the host genome. The results of this experiment are shown in the lower panel of Figure 10: LTR activity is reduced to a half when USF^{43} is overexpressed (in CAT assays, the percentage of the acetylated form of ^{14}C chloroamphenicol was decreased from 77% to 40%).

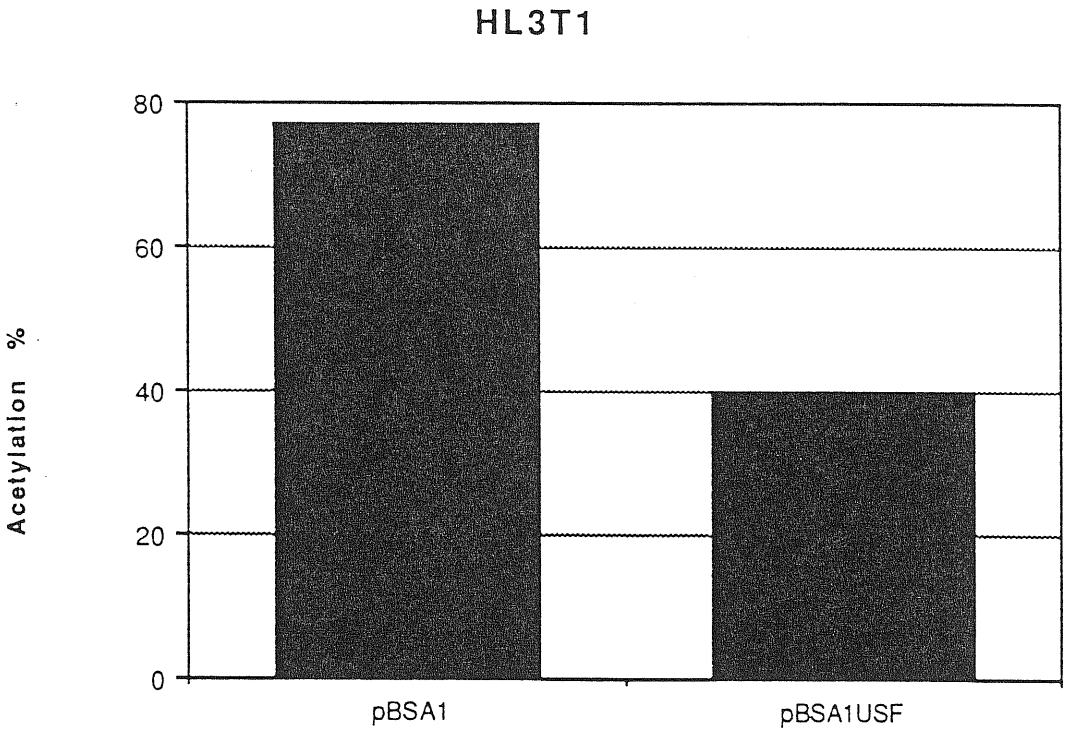
The same results were obtained when both the LTRCAT plasmid and pBSA1USF^{43} were transiently introduced in HeLa cells. In this case, LTR transcriptional activity was reduced as well, being the percentage of acetylated chloramphenicol decreased from 15% to 5%. It should be stressed that since the values obtained in the CAT assays are not dramatically different for the constructs utilized, we developed a very reproducible method for monitoring the efficacy of transfection, by co-transfecting a plasmid expressing the luciferase gene and quantifying the amount produced enzyme by a luminescence assay.

These data lead us thinking that the transcriptional inhibitory activity of the CACGTG element of the LTR was indeed due to USF/MLTF or, at least, to its 43 kDa form. This is surprising, since USF/MLTF was reported to activate transcription from several other cellular and viral promoters.

We made the hypothesis that this effect could be due to the localization of the site within the LTR and the resulting interactions between USF/MLTF and the proteins binding to the promoter. In order to verify this hypothesis, we tried to dissect the transcription activating domain of USF^{43} and define the domains which are involved in transcriptional activation and those involved in transcriptional inhibition. To this purpose, we constructed a plasmid expressing the DNA binding domain of



Panel A



Panel B

Figure 10: histograms heights represent the percentage of acetylated chloramphenicol in CAT assays. Upper panel: pSV2-Tat and pLTRCAT reporter plasmid transfected in HeLa cells with plasmid pBSA1 or PBSA1-USF^{43} expressing the 43 kDa form of USF/MLTF. Lower panel: pBSA1 or pBSA1-USF^{43} transfected in HL3T1 cells, a cell line bearing the LTRCAT construct integrated in the genome.

the yeast transcription factor Gal4 fused to the first 200 amino-terminal aminoacids of USF^{43} (pGal4-USF^{43}). This portion of the protein does not contain the DNA binding domain or the dimerization element, therefore it is likely that in it could reside the transactivating ability of USF^{43}. The activity of this chimeric protein was assayed on an LTR plasmid (named pLTRΔNRE-Gal4-CAT) in which the Negative Regulatory Element, containing the USF/MLTF binding site, was substituted by four Gal4 binding sites. Previously, the transcriptional activity of Gal4-USF^{43} fusion protein was checked on a synthetic minimal promoter, constituted by five Gal4 binding sites and a TATA box cloned upstream the CAT gene. The reporter plasmid was not transcriptionally active when transfected alone, while cotransfection of the plasmid expressing Gal4-USF^{43} significantly triggered CAT activity, demonstrating that the transactivating domain of USF^{43} effectively resides in the amino-terminal portion and that, at least in this simple promoter, it acts as a positive regulator of transcription (data not shown). As a positive control, in this experiments we used a fusion protein between Gal4 and VP16, the herpes virus strong acidic trans-activator.

Then, we transfected pLTRΔNRE-Gal4-CAT in the absence or presence of pGal4-USF^{43}. The transcriptional efficiency of this mutated LTR was assayed in the CAT assay depicted in Figure 11 (in all the experiments an equal amout of pSV2-Tat was used): Lanes 1 and 2 show the basal transcriptional rate of pLTRΔNRE-Gal4-CAT plus pSV2-tat. As expected, when Gal4-VP16 was provided in trans, it greatly upregulated transcription (lane 3). Interestingly, Gal4-USF^{43} activated transcription as well (lane 4). This observation is in apparent contrast with the data reported before, in which USF^{43} is a downregulator of HIV-1 transcription and with the ones, previously reported, showing that a binding site for USF/MLTF mimics the Negative Regulatory Element [47]. I will try to reconcile these conflicting data in the Discussion section.

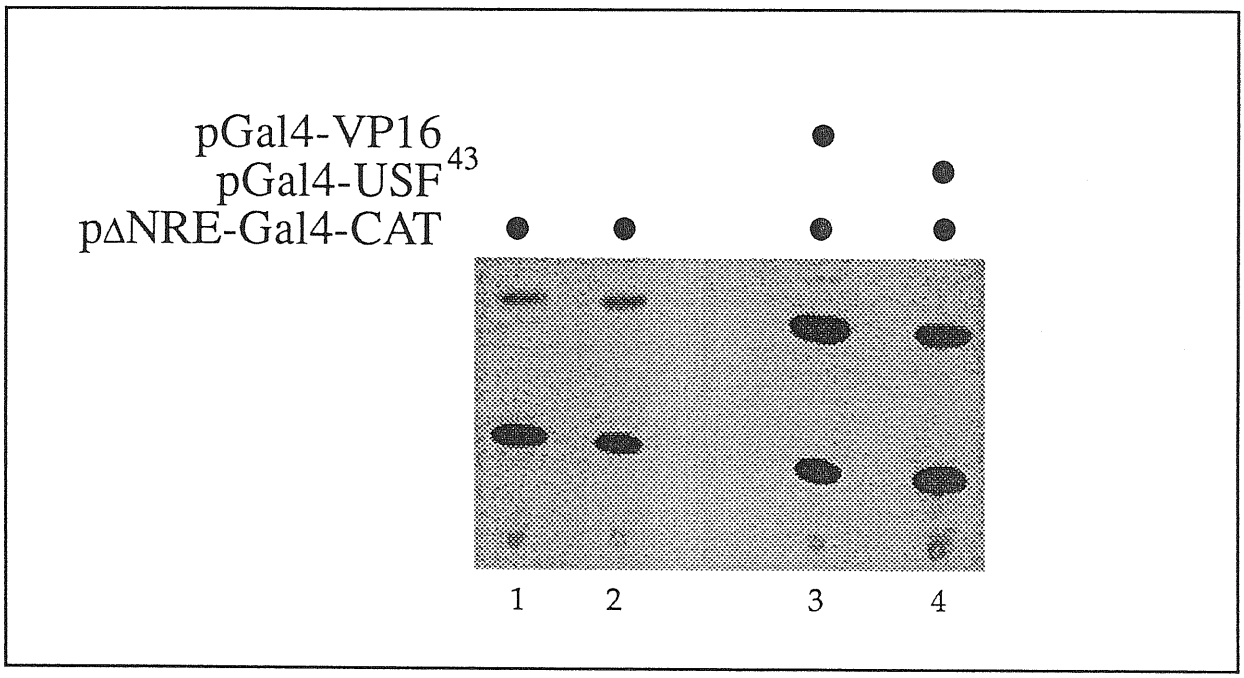


Figure 11: CAT assay. Lanes 1 and 2: basal level of transcription of pΔNRE-Gal4-CAT transfected alone; lane 3: the reporter plasmid was co-transfected with pGal4-VP16; lane 4 it was co-transfected with pGal4-USF43.

4.0 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 evidences highlight the fundamental role of transcription in the control of viral replication. Moreover, in HIV-1 infected patients, disease progression correlates with a progressive activation of viral transcription as measured by the ratio between the copies of viral RNA and DNA molecules [4], stressing the importance of a systematic and thoroughly study of the mechanisms controlling viral transcription.

Once integrated, the proviral genome can be considered as a cellular gene, its expression being tightly regulated by both cellular and viral factors. The molecular dissection of the 5' Long Terminal Repeat, the promoter of HIV-1 transcription, evidences three functional cis-acting elements: a basal promoter centered on the TATA box and strongly dependent from the integrity of the three Sp1 sites; an enhancer region sensitive to the stimuli coming from the environment outside the cell; and a Negative Regulatory Element (NRE), whose progressive deletion strengthens LTR-driven transcription.

In the recent literature, there is a great amount of proteins reported to bind and to exert some influence on this promoter. Although most of these results come from *in vitro* studies which could not necessarily reflect the situation *in vivo*, the appearance is that of a extremely complex promoter. The evolutionary advantage for the virus, resulting from such a complicated mosaic of binding sites, is probably related to the possibility to finely

tune its expression in different cells in different tissues, or to respond to different stimuli, or intriguingly, to maintain a latency state until transcription massively turned on.

In order to better understand which are the sequences effectively bound by transcription factors, an *in vivo* footprinting assay was developed in our laboratory. In lymphoid cells, the most common cellular target for HIV-1, few binding sites are occupied by proteins *in vivo*: the TATA box, the Sp1 sites, the enhancer element, the NFAT site, a purin rich region and a region at the proximal end of the NRE, centered on the CACGTG palindromic sequence [44]. This element had already been characterized in our laboratory [47]. We were able to demonstrate that an oligonucleotide containing this sequence 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. Furthermore, in south-western experiments, in which a labeled oligonucleotide was used to probe a filter where nuclear proteins, resolved according to their size, were blotted, demonstrated that a ~44 kDa protein was specifically recognized by the probe. After extensive purification, this polipeptide turned out to be the USF/MLTF transcription factor as demonstrated by its binding specificity, its biochemical properties and its immunological reactivity [45]. The purified activity was composed by two immunologically unrelated polipeptides of 43 and 44 kDa relative molecular weight, both able to bind to the same target sequence with the same specificity and independently one from the other, probably in a dimeric form [52]. The *in vivo* role of this binding site was investigated by cloning a homologous binding site for USF/MLTF, localized in a DNA region coinciding to a DNA origin of replication [40], in place of the Negative Regulatory Element of the LTR. While the deletion of the NRE increased LTR transcriptional strength as expected, its substitution with an oligonucleotide corresponding to the binding site for USF/MLTF in both orientations restored the negative effect, suggesting that this binding site works as a negative regulator of HIV-1 transcription. This is puzzling, since USF/MLTF is a well known activator of transcription as demonstrated in the case of the Adenoviral Major Late

Promoter and in other cellular promoters such as the rat γ–fibrinogen, the human growth hormone and the human heme oxygenase one.

These data prompted us to investigate in further details the activity of this transcriptional factor and in particular of its 43 kDa form, whose cDNA was cloned and characterized [53].

4.1 *In vitro* studies

We have express the 43 kDa form of USF/MLTF (USF^{43}) in bacteria as a GST-USF^{43} fusion protein. This technology was employed for its broad use, reliability and for the speed of purification that it allows. The recombinant protein was specifically recognized by an antiserum raised against purified USF/MLTF.

First, the activity of GST-USF^{43} was detected by gel retardation assay. All the sequences used, including the HIV-1 LTR, the Adenovirus Major Late Promoter and the B48 DNA origin of replication that we previously reported to interact with HeLa USF/MLTF, turned out to be capable to bind GST-USF^{43}. Coherently, GST-USF^{43} appeared to be sensitive to the same mutations that abolish DNA binding of the eukaryotic protein.

Since HeLa USF/MLTF is sensitive to cytosine methylation at position 5 of the CpG dinucleotide within the core of the CACGTG binding site, we challenged the GST-USF^{43}/DNA complex with oligonucleotides methylated at different positions: DNA methylation sensitivity mirrored the one of HeLa USF/MLTF; in fact, only methylation of the core CpG affected binding, while methylation of a distal CpG did not.

Purification of USF/MLTF from HeLa cells provided the way to demonstrate that the protected area over the CACGTG sequence that we detected in DNase I footprinting assay using crude total nuclear extract, was indeed due to USF/MLTF binding. In fact, DNase I footprinting experiments using HeLa-purified USF/MLTF clearly demonstrated that the factor binds to the core sequence, and that binding is specific, since mutation of the CACGTG sequence abolishes binding. Interestingly the appearance of a clear

hypersensitive region at the 3' boundary of the lower strand, suggests that the structure of the DNA helix is altered upon binding. To obtain further details on the interaction between GST-USF^{43} and DNA, methylation protection assay was carried out. This assay allows to determine which purins are directly involved in protein binding. Purins from base -167 to base -162 in respect to the transcriptional start site are engaged in binding; furthermore a dramatic hypersensitive site is evident at position -161, suggesting that it is particularly exposed upon binding. This observation is very interesting since the same signal is detected in living cells by means of *in vivo* footprinting with DMS [44].

The data from the DNase I footprinting and methylation protection assays are summarized in Figure 8. Although these data are still preliminary, (DNase I footprinting experiments using GST-USF^{43} are in progress) they suggest that, upon binding of the protein to the DNA, a local distortion occurs in such a way that both DNase I and DMS have an easier access to nucleotides, even if at different positions.

The above-mentioned considerations leaded us to consider the possibility that USF/MLTF could bend DNA upon binding. Several transcriptional factors are reported to have this property. In this regard, it is noteworthy to distinguish intrinsically bent DNA, which is an intimate property of a DNA sequence with a certain stretch of bases (usually adenines) arranged in a cyclic fashion (every 10-11 bases), from DNA which becomes bent only upon protein binding. A role for the bending ability of transcriptional factors *in vivo* has been demonstrated in prokaryotic cells, where a binding site for the CAP protein (a factor which activates transcription and bends DNA of 90°) can be substituted with an intrinsically bent DNA fragment, suggesting that bending per se can be sufficient for transcriptional activation [55]. In eukaryots such a clear evidences is still missing, although it was suggested that DNA bending by transcriptional factors can bring distantly bound proteins closer together by facilitating DNA-looping or it can mediate the interaction between transcription factors and the general transcription machinery by formation of large nucleoprotein structures in which the DNA is wrapped around the

protein complex. Alternatively, the energy stored in a protein-induced bend could be used to favor the formation of an open transcription complex [56].

Circular permutation analysis is the easiest way to investigate the bending ability of transcription factors. Bent DNA can be detected by its anomalous migration during electrophoresis. Reduction of migration is maximal when the bend is located at the centre of a DNA molecule and minimal when it is located near one end. Therefore, we used several probes in which the binding site for USF/MLTF is moved from one end through the centre to the other end, to determine the presence of a bent DNA structure.

The results obtained, indicate that both HeLa USF/MLTF and GST-USF^{43} bend DNA, although at a different extent; the bending angle for the two complexes being respectively 90° and 110°. 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 the HeLa purified USF/MLTF; however, recently Kerppola and Curran [57] have demonstrated that there is no significant correlation between the molecular weight and the DNA bending ability of a protein. Our preferred interpretation, albeit highly speculative, relies on the evidences that USF^{43} is able to form heterodimers or heteromultimers with the 44 form of USF/MLTF [58]. Assuming that different complexes also have different bending properties, it seems sound to consider 90° the bending angle of the USF^{43} homodimer and 110° the bending angle resulting from the vectorial sum of the two different bending angles of USF^{43} and USF^{44}. In this regard, there are several examples of differential bending by monomeric or heteromeric forms, such is the case of Myc/MAX and Fos/Jun heterodimers [59][60].

The studies on the functional role of DNA bending by transcriptional factors should probably deserve more scrutiny. In fact, it should be considered that at least four other transcription factors that bind to the LTR (Jun/Fos , NF-κB , Sp1, TBP (the TATA binding protein [60][61][62]) in addition to USF/MLTF, are able to bend DNA upon binding. As a consequence, the traditional picture of a linear promoter where

transcription factors interact individually with the DNA, should be replaced by a more realistic view of a highly structured promoter where proteins, by bending and looping, can interact among them and, as a final result, control the rate of transcription initiation.

4.2 *In vivo* studies

Previous *in vivo* experiments suggested that the binding site for USF/MLTF transcription factor was a cis-acting negative regulator of HIV-1 transcription [47]. The possibility of transiently overexpressing USF^{43} in living cells provided a mean to verify the activity of USF^{43}. We transfected a plasmid encoding USF^{43} in HL3T1 cells, a HeLa derivative in which an LTRCAT construct is integrated in the host genome. Overexpression of USF^{43} resulted in the inhibition of transcription. The same effect was also obtained by cotransfecting a pLTRCAT reporter plasmid together with the vector expressing USF^{43} in HeLa cells.

These results are puzzling, since USF/MLTF usually behaves as a transcriptional activator from all the promoters where a binding site has been identified. The negative role of the factor on HIV-1 transcription has been independently supported by a different laboratory, which showed that a virus with a mutated USF/MLTF sequence replicates more efficiently as compared to the wild type [28]. A possible explanation for this apparent discrepancy in the behavior of this factor, can derive from the localization of its binding site far away from the transcription start site in the HIV-1 LTR, while it is closer in other promoters (for example, at position -55 in the Adenovirus MLP). To test this hypothesis, we replaced the NRE with a binding site for transcription factor GAL4, and tested the function of a protein containing the GAL4 binding domain fused to the USF^{43} activation domain on this promoter. The results clearly indicate that the fusion protein is a positive effector in these conditions, suggesting that the negative role of the USF/MLTF factor is an intrinsic property of the intact protein.

A likely explanation for this effect can derive from the observation that the fusion protein lacks the USF^{43} dimerization moiety, constituted by the leucine zipper and the helix-loop-helix motif. As a consequence, this protein is unable to dimerized with other partners, which could be needed for its negative function. A likely candidate for dimerization is USF^{44}, which copurifies in equimolar amount with USF^{43}, and was shown to be able to form heterodimers with USF^{43} in *in vitro* studies. Other putative partners for dimers formation are TFII-I, which binds to the initiatior elements of the HIV-1 LTR and cooperates with USF^{43} in binding to this element. Cloning and expression of these proteins and their utilization in vitro transcription experiments will allow to solve this problem.

An other possible explanation to clarify the discrepancy between the behavior of USF^{43} and that of Gal4-USF^{43} is the lack of DNA bending ability of the latter. As suggested above, DNA bending could alter the promoter architecture in such a way to influence transcription in a positive or a negative way.

Finally, it should be noted that other proteins of the HLH family are potentially able to bind to the CACGTG core sequence of the HIV-1 LTR, among which TFE3 [63], TFEB [64], Myc/Max heterodimers [65], MAD [66] and Mxi1[67]. We are presently using the HL3T1 cell line to test the ability of these factors to affect the LTR-driven expression.

5.0 REFERENCES

1. 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., Montagnier, L. 1983. Science 220, 868-871

2. Malim, M.H., Bohnlein, S., Hauber, J., Cullen, B.R. 1989. Cell 58, 205-214

3. Olsen, H.S., Cochrane, A.W., Dillon, P.J., Nalin, C.M., Rosen, C.A. 1990. Genes Develop. 4, 1357-1364

4. Bagnarelli, P., Menzo, S., Valenza, A., Manzin, A., Giacca, M., Ancarani, F., Scalise, G., Varaldo, P.E., Clementi, M. 1992. J. Virol. 66, 7328-7335

5. Michael, N.L., Vahey, M., Burke, D.S., Redfield, R.R. 1992. J. Virol. 66, 310-316

6. Pantaleo, G., Graziosi, C., Demarest, J.F., Butini, L., Montroni, M., Fox, C.H., Orenstein, J.M., Kotler, D.P., Fauci, A.S. 1993. Nature 362, 355-358

7. Embretson, J., Zupancic, M., Ribas, J.L., Burke, A., Racz, P., Tenner-Racz, K., Haase, A.T. 1993. Nature 362, 359-362

8. Piatak, M., Jr., Saag, M.S., Yang, L.C., Clark, S.J., Kappes, J.C., Luk, K.-C., Hahn, B.H., Shaw, G.M., Lifson, J.D. 1993. Science 259, 1749-1754

9. Gaynor, R. 1992. AIDS 6, 347-363

10. Siekevitz, M., Josephs, S.F., Dukovich, M., Peffer, N., Wong-Staal, F., Greene, W.C. 1987. Science 238, 1575-1578

11. Crabtree, G.R. 1989. Science 243, 355-361

12. Chinnadurai, G. 1991. Gene 101, 165-170

13. Garcia, J.A., Harrich, D., Soultanakis, E., Wu, F., Mitsuyasu, R., Gaynor, R.B. 1989. EMBO J. 8, 765-778

14. Garcia, J.A., Wu, F.K., Mitsuyasu, R., Gaynor, R.B. 1987. EMBO J. 6, 3761-3770

15. Wu, F.K., Garcia, J.A., Harrich, D., Gaynor, R.B. 1988. EMBO J. 7, 2117-2129

16. Kato, H., Horikoshi, M., Roeder, R.G. 1991. Science 251, 1476

17. Toohey, m.G., Jones, K.A. 1989. Genes Develop. 3, 265

18. Kawakami, K., Scheidereit, C., Roeder, R.G. 1988. Proc. Natl. Acad. Sci. USA 85, 4700-4704

19. Baeuerle, P.A., Baltimore, D. 1988. Science 242, 540-546

20. Riviere, Y., Blank, V., Kourlisky, P., Israel, A. 1991. Nature 350, 625-627

21. Ballard, D.W., Walker, W.H., Doerre, S., Sista, P., Molitar, J.A., Dixon, E.P., Paffer, N.J., Hannink, M., Greene, W.C. 1990. Cell 63, 803-814

22. Schmid, R.M., Perkins, N.D., Duckett, C.S., P.C., A., G.J., N. 1991. Nature 352, 733-736

23. Baldwin, A.S.J., Sharp, P.A. 1988. Proc. Natl. Acad. Sci. USA 85, 723-727

24. Clark, L., Pollock, R.M., Hay, R.T. 1988. Genes Dev. 2, 991-1002

25. Baldwin, A.S., Jr., LeClaire, K.P., Singh, H., Sharp, P.A. 1990. Mol. Cell. Biol. 10, 1406-1414

26. Bours, V., Villalobos, J., Burd, J., Kelly, K., Siebenlist, U. 1990. Nature 348, 76

27. Rosen, C.A., Sodroski, J.G., Haseltine, W.A. 1985. Cell 41, 813-823

28. Lu, Y., Stenzel, M., Sodroski, J.G., Haseltine, W.A. 1989. J. Virol. 63, 4115-4119

29. Orchard, K., Perkins, N., Chapman, C., Harris, J., Emery, V., Goodwin, G., Latchman, D., Collins, M. 1990. J. Virol. 64, 3234-3239

30. Landshulz, W.H., Johnston, P.F., McKnight, S.L. 1988. Science 240, 1259-1276

31. Dasgupta, P., Sakumar, P., Reddy, D., Reddy, P. 1990. Proc Nat Acad Sci 87, 8090-8094

32. Boise, L.H., Petryniak, B., Mao, X., June, C.H., Wang, C., Lindsten, T., Bravo, R., Kovary, K., Leiden, J.M., Thompson, C.B. 1993. Mol. Cell. Biol. 13, 1911-1919

33. Markovitz, D.M., Hannibal, M.C., Smith, M.J., Cossman, R., Nabel, G.J. 1992. J Virol 66, 3961-3965

34. Tesmer, V., Rajadhyaksha, A., Babin, J., Bina, M. 1993. Proc Natl Sci 90, 7298-7302

35. Waterman, M.L., Fisher, W.H., Jones, K.A. 1990. New Biol. 2, 221-236

36. Sawadogo, M. 1988. J. Biol. Chem. 263, 11994-12001

37. Pognonec, P., Roeder, R.G. 1991. Mol.Cell.Biol. 11, 5125

38. Sirito, M., Walker, S., Lin, Q., Kozlowsi, M.T., Klein, W.H., Sawadogo, M. 1992. Gene Expression 2, 231-240

39. Giacca, M., Gutierrez, M.I., Diviacco, S., Demarchi, F., Biamonti, G., Riva, S., Falaschi, A. 1989. Biochem. Biophys. Res. Comm. 165, 956-965

40. Tribioli, C., Biamonti, G., Giacca, M., Colonna, M., Riva, S., Falaschi, A. 1987. Nucleic Acids Res. 15, 10211-10218

41. Biamonti, G., Giacca, M., Perini, G., Contreas, G., Zentilin, L., Weighart, F., Saccone, S., Valle, G.D., Riva, S., Falaschi, A. 1992. Mol. Cell. Biol. 12, 3499-3506

42. Roy, A.L., Meisterernts, M., Pognonec, P., Roeder, R.G. 1991. Nature 354, 245-248

43. Du, H., Roy, A.L., Roeder, R.G. 1993. Embo J. 12, 501-511

44. Demarchi, F., D'Agaro, P., Falaschi, A., Giacca, M. 1992. J. Virol. 66, 2514-2518

45. Csordas Toth, E., Marusic, L., Ochem, A., Patthy, A., Pongor, S., Giacca, M., Falaschi, A. 1993. Nucleic Acids Res. in press

46. Sambrook, J. Fritsch, E.F and Maniatis, T. 1989 Molecular Cloning - A laboratory manual/II edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

47. Giacca, M., Gutierrez, M.I., Menzo, S., d'Adda di Fagagna, F., Falaschi, A. 1992. Virology 186, 133-147

48. Maxam, A.M., Gilbert, W. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. In: Methods in Enzymology. L. Grossman, K. Moldave Editore, New York, 499-560

49. Papavassiliou, A.G. 1993. Nucl. Ac. Res. 21, 757-758

50. Zinkel, Crothers. 1987. Nature 328, 178-181

51. Gorman, C.M., Moffat, L.F., Howard, B.H. 1982. Mol.Cell.Biol. 2, 1044-1051

52. Sawadogo, M., Van Dike, M.W., Gregor, P.D., Roeder, R.G. 1988. J. Biol. Chem. 263, 11985-11993

53. Gregor, P.D., Sawadogo, M., Roeder, R.G. 1990. Genes Devel. 4, 1730-1740

54. Thompson, J.F., Landy, A. 1988. Nucleic Acids Res. 16, 9687-9705

55. Bracco, L., Kotlarz, D., Kolb, A., Diekmann, S., Buc, H. 1989. EMBO J. 8, 4289-4296

56. van der Vliet, P.C., C.P., V. 1993. BioEssay 15, 25-32

57. Kerppola, T.K., Curran, T. 1993. Mol. Ce. Biol. 13, 5479-5489

58. Sirito, M., Walker, S., Lin, Q., Koxlowski, M.T., Klein, W., Sawadogo, M. 1992. Gene Expression 2, 231-240

59. Wechsler, D., Dang, V.C. 1992. Proc. Natl. Acad. Sci. 89, 7635-7639

60. Kerppola, T.K., Curran, T. 1991. Cell 66, 317-326

61. Schreck, R., Zorbas, H., Winnacker, E.L., Baeuerle, P.A. 1990. Nucleic Acids Res. 18, 6497-6502

62. Horikoshi, M., Carey, M.F., Kakidani, H., Roeder, R.G. 1988. Cell 54, 665-669

63. Beckmann, H., Su, L.-K., Kadesch, T. 1990. Genes & Development 4, 167-179

64. Carr, C.S., Sharp, P.A. 1990. Molecular and Cellular Biology 10, 4384-4388

65. Blackwood, E.M., Eisenman, R.N. 1991. Science 251, 1211-1217

66. Ayer, D.E., Kretzner, L., Eisenaman, R.N. 1993. Cell 72, 211-222

67. Zervos, A.S., Gyuris, J., Brent, R. 1993. Cell 72, 223-232

ACKNOWLEDGEMENTS

This is to acknowledge with thanks the collaboration of my colleagues at ICGEB and, in particular, 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.