



HAL
open science

Characterization of nuclear proteins that bind to the regulatory TGATTGGC motif in the human immunodeficiency virus type 1 long terminal repeat

Christian Schwartz, François Canonne-Hergaux, Dominique Aunis, Evelyne Schaeffer

► To cite this version:

Christian Schwartz, François Canonne-Hergaux, Dominique Aunis, Evelyne Schaeffer. Characterization of nuclear proteins that bind to the regulatory TGATTGGC motif in the human immunodeficiency virus type 1 long terminal repeat. *Nucleic Acids Research*, 1997, 25 (6), pp.1177-1184. 10.1093/nar/25.6.1177 . hal-02269563

HAL Id: hal-02269563

<https://hal.science/hal-02269563v1>

Submitted on 22 Aug 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Characterization of nuclear proteins that bind to the regulatory TGATTGGC motif in the human immunodeficiency virus type 1 long terminal repeat

Christian Schwartz, François Canonne-Hergaux, Dominique Aunis and Evelyne Schaeffer*

Unité 338 INSERM, Centre de Neurochimie, 5 rue Blaise Pascal, 67084 Strasbourg Cedex, France

Received December 4, 1996; Revised and Accepted January 30, 1997

ABSTRACT

We have recently elucidated the nature and function of transcription factors present in Jurkat, glial and neuronal cells that interact with modulatory region B, the nuclear receptor responsive element, in the long terminal repeat of human immunodeficiency virus type 1 (HIV-1). Considering the key role that the combination of host cell proteins plays in HIV-1 gene transcription, it appears essential to characterize proteins interacting with the adjacent region A. *In vitro* experiments revealed that the 5'-TGATTGGC-3' motif of region A is the target for at least three distinct proteins, one belonging to the nuclear factor I family, while two others are related to the cAMP response element binding (CREB) protein family. One of these proteins, present in DNA-protein complex C2, is formed by distinct polypeptides of relative molecular mass 43 000 and 50 000. We have purified the 43 kDa protein, which is distinct from CREB-43, and have shown that renatured p43 is able to specifically interact with site A. Transient expression experiments with vectors containing wild-type or mutant motif A revealed that basal HIV-1 gene transcription in Jurkat cells is regulated by antagonistic effects of the site A binding proteins.

INTRODUCTION

Replication and gene expression of human immunodeficiency virus type 1 (HIV-1) depend upon multiple sequences present in the long terminal repeat (LTR). Transcription is regulated by an interplay of viral and cellular host proteins with regulatory elements of the LTR. Key elements such as the TATA, Sp1 and NF- κ B region, present in the proximal part of the LTR (+1 to -112) have been extensively characterized (for reviews see 1,2). The role of the NF- κ B element especially has been investigated in a variety of cell types, including lymphocytes, monocytes and central nervous system-derived cells (3-7).

However, transcription control elements present in the -112 to -459 modulatory region of the LTR have been poorly characterized. This region includes regulatory elements that can mediate transcription in many cell types, under a variety of growth and differentiation conditions (8). Recently several reports have

described the nuclear receptor-responsive element (NRRE) located within the -356/-320 region as representing a site for complex regulatory interactions among hormone receptors and orphan receptors, as well as the AP-1 transcription factor (9-12). We have recently reported that AP-1 is unable to directly bind to the NRRE sequence but interacts indirectly via nuclear receptors belonging to the steroid-thyroid-retinoid nuclear receptor super-family (13-14). Moreover, we have recently reported the role of retinoic acid receptors and the orphan receptor COUP-TF on HIV-1 gene transcription in human central nervous system-derived cell lines (14). Since recent data have highlighted the importance of the U3 region of the LTR in determining the pathogenicity of HIV-1 (15), studies concerning the nature and role of cellular transcription factors that interact with various elements of the LTR appear essential.

In this study we have characterized the regulatory element spanning the -385 to -362 region of the LTR, which is directly adjacent to the NRRE. This region has previously been footprinted with nuclear extracts from Jurkat T cells and named site A (16-17). Our previous studies have revealed that this sequence is the binding site for nuclear proteins present in a great variety of cell lines, such as Jurkat, HeLa, astrocytoma, oligodendroglioma and neuronal cells (13). However, the nature and function of the nuclear proteins interacting with site A remain elusive. Here we demonstrate that the TGATTGGC core sequence is the binding site for at least three distinct proteins present in Jurkat and in HeLa cells. By *in vitro* experiments we have analyzed the nature of the proteins forming the different DNA-protein complexes. We have shown that one nuclear protein belongs to the nuclear factor I (NFI) family (18), while other factors are related to the CREB protein family (for a review see 19). We report here the purification of a nuclear factor of relative molecular mass 43 000, distinct from CREB-43, which specifically binds to site A. Moreover, we provide evidence, by transient expression experiments, that site A binding proteins play antagonistic roles in the control of basal transcription of the HIV-1 genome in Jurkat cells.

MATERIALS AND METHODS

Plasmid constructs

To construct the 1N/tk-CAT, 1L/tk-CAT, 1Lm/tk-CAT, 2 \times 1L/tk-CAT and 2 \times 1Lm/tk-CAT vectors, one copy of the 1N, 1L and

*To whom correspondence should be addressed. Tel: +33 3 88 45 67 18; Fax: +33 3 88 60 08 06; Email: schaeffer@neurochem.u-strasbg.fr

1Lm oligonucleotides and two copies of the 1L and 1Lm oligonucleotides respectively were subcloned in sense orientation in the blunt-ended *SalI* site of pBLCAT2, containing the herpes simplex virus thymidine kinase promoter in front of the *CAT* gene (20; a gift of C.Kedinger, Unit INSERM 184, Strasbourg, France). The 1L-LTR-CAT and 1Lm-LTR-CAT vectors were constructed by site-directed mutagenesis with oligonucleotides 1L and 1Lm (Transformer TM Site-directed Mutagenesis Kit; Clontech). Mutations were controlled by DNA sequencing.

Cell culture and nuclear extract preparation

Jurkat and HeLa cells were respectively grown in RPMI 1640 and DMEM medium supplemented with 10% fetal calf serum and 10 mM HEPES in the presence of penicillin and streptomycin (100 U/ml). Nuclear extracts were prepared as previously described (21). A phosphatase inhibitor, 0.5 mM β -glycerophosphate, was added during the preparation of nuclear extracts from Jurkat cells.

Transfections and CAT assays

Jurkat cells (5×10^6 – 10^7 cells/transfection) were transfected by the DEAE-dextran technique with 3 pmol plasmid DNA. After 48 h, cells were collected and chloramphenicol acetyltransferase (CAT) assays were performed as described previously (21) with 20 μ g protein extract and a 2 h incubation at 37°C.

Electrophoretic mobility shift assays

Protein-DNA binding reactions were performed with 5 or 10 μ g crude nuclear extract as previously described (13). For competition assays, unlabeled oligonucleotides were added at a 50- or 200-fold molar excess at the same time as the probe. When purification columns were analyzed, 10 μ l of the fractions were used. The sequences of the oligonucleotides used as probes or competitors are shown in Figure 1. In supershift assays performed as previously described (13), we used antibodies directed against CREB-1 and ATF-1 and reactive with ATF-1 p35, CREB-1 p43 and CREM-1 (Santa Cruz Biotechnology). Binding reactions were carried out with the protein CREB bZIP (Santa Cruz Biotechnology).

Methylation interference assays

Methylation interference assays were performed with oligonucleotide 1L as described previously (13).

UV cross-linking analysis

Crude nuclear extract (10 μ g) from Jurkat cells and partially purified proteins (10 μ g) of a 0.4 M KCl fraction from the heparin-Sepharose column were incubated for 15 min on ice with the 5'-end-labeled 1L probe substituted with bromodeoxyuridine under standard mobility shift assay conditions. Following electrophoresis, the gel was irradiated for 12 min using a UV transilluminator (254 nm; Bioblock Scientific). Radioactive bands corresponding to complexed and free DNA identified by autoradiography were excised and incubated in 40 μ l denaturing buffer (62 mM Tris-HCl, pH 6.7, 2% SDS, 10% glycerol, 50 mM dithiothreitol) and heated at 100°C for 5 min. The gel slices and the denaturing buffer were loaded on a 10% SDS-polyacrylamide gel. The gel was fixed, dried under vacuum and the proteins directly involved in the DNA interaction were identified by

autoradiography. We used prestained molecular weight markers (BioRad).

Southwestern blotting

Nuclear proteins partially purified by heparin-Sepharose (20 μ g) or by DNA affinity chromatography (5 μ g) were resolved on a 10% SDS-PAGE gel and electrotransferred onto nitrocellulose membranes. The membranes were washed three times for 1 h with renaturation buffer (10 mM Tris, pH 7.5, 150 mM NaCl, 10 mM dithiothreitol, 2.5% Nonidet P-40, 10% glycerol, 3% BSA) (22). The membranes were then placed in a heat-sealable pouch in binding buffer (10 mM Tris-HCl, pH 7.5, 40 mM NaCl, 1 mM DTT, 1 mM EDTA, 8% v/v glycerol, 3% BSA, 60 μ g poly(dI-dC), 5 mM MgCl₂, 5 μ g/ml ³²P-labeled oligonucleotide 1L or 1Lm). After 16 h constant agitation, membranes were washed for 2 h at room temperature with 10 mM Tris-HCl, 50 mM NaCl with several changes and autoradiographed on Kodak X-OMAT film.

Purification of a 43 kDa protein

Aliquots of 25 mg nuclear protein extract, obtained from 5×10^9 Jurkat cells were applied to a 5 ml heparin-Sepharose column equilibrated with buffer Z (20 mM HEPES, pH 7.9, 0.2 mM EDTA, 20% glycerol, 1 mM dithiothreitol) containing 50 mM KCl. Proteins were separated using a step gradient from 0.1 to 1 M KCl in buffer Z. Activity was detected in fractions eluted between 0.3 and 0.4 M KCl by gel mobility shift assay. These fractions were pooled and dialyzed against buffer Z containing 0.1 M KCl. The pool was mixed with salmon sperm DNA (25 μ g/ml), incubated on ice for 15 min and loaded onto a 2 ml oligonucleotide 1Lm affinity column equilibrated in buffer Z containing 0.1 M KCl. The column was washed with the same solution. The flow-through was loaded onto a 2 ml oligonucleotide 1L affinity column equilibrated in buffer Z containing 100 mM KCl. The column was washed with the same solution and eluted with a step gradient from 0.1 to 1 M KCl in buffer Z. Oligonucleotide 1L binding activity was detected in fractions at 0.5 M KCl. These fractions were pooled, dialyzed against buffer Z containing 0.1 M KCl and recycled on the same column. An aliquot (10 μ l) of each fraction was assayed by gel retardation assay using the oligonucleotide 1L as probe. The active fractions were also analyzed by 10% SDS-PAGE. The sequence-specific DNA affinity columns were prepared according to Kadonaga and Tjian (23) using CNBr-activated Sepharose 4B (Pharmacia).

Denaturation and renaturation experiments

Fractions (20 μ g) containing peak activity after heparin-Sepharose chromatography were subjected to a preparative SDS-10% PAGE gel. After electrophoresis, the gel was cut horizontally into 1 mm slices between 40 and 70 kDa. The proteins were eluted from these slices as described by Hager and Burgess (24). After elution, proteins were precipitated with acetone and resuspended in buffer H (50 mM HEPES, pH 7.9, 0.5 mM EDTA, 0.05% Nonidet P-40, 20% glycerol, 1 mM dithiothreitol, 0.5 mM phenylmethylsulfonyl fluoride, 0.1 M NaCl and 5 μ g/ml leupeptin and pepstatin) containing 6 M guanidine-HCl for 20 min. The denaturing agent was diluted 50-fold with buffer H and renaturation was allowed for 4 h at 4°C. Activity of renatured proteins was

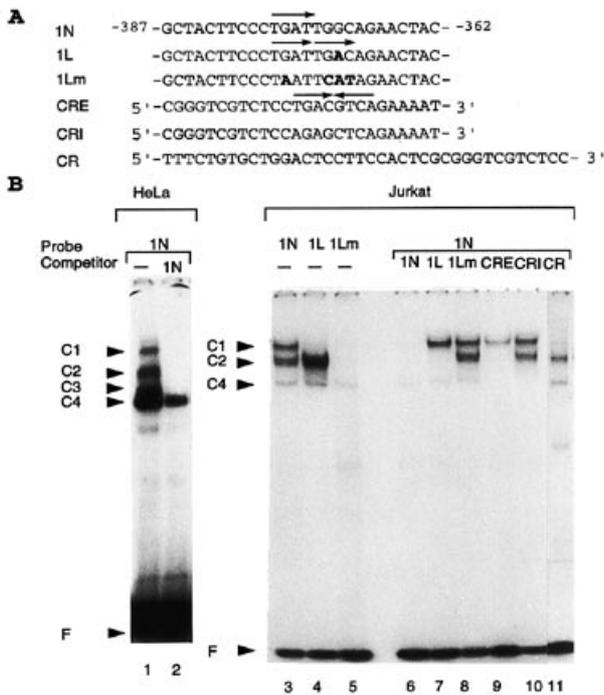


Figure 1. Gel mobility shift analysis of nuclear proteins from HeLa and Jurkat cells binding to region A of the HIV-1 LTR. (A) Sequences of wild-type and mutant oligonucleotides 1N, 1L and 1Lm, corresponding to the LTR region A, and of oligonucleotides CRE, CRI and CR. The bold type represents mutant oligonucleotides. (B) Gel retardation assays were carried out by incubating 10 μ g nuclear extract from HeLa or Jurkat cells with probes 1N, 1L and 1Lm. Binding reactions were performed either in the absence of competitor (lanes -) or in the presence of a 50-fold molar excess of unlabeled 1N, 1L or CR oligonucleotide or a 200-fold molar excess of unlabeled CRE or CRI oligonucleotide, as indicated on top. DNA-protein complexes C1-C4 and the free probe (F) are indicated.

tested by gel mobility shift assay in the absence of non-specific competitors.

RESULTS

Multiple host nuclear factors bind to the site A of the HIV-1 LTR

The sequence spanning the -385/-362 distal region of the HIV-1 LTR has previously been footprinted with nuclear extracts from different cell types and named site A (13,16,17). A number of studies have shown that the adjacent region B is the binding site of proteins belonging to the nuclear receptor superfamily (11-13). In order to analyze the nuclear proteins interacting with site A, we performed DNA mobility shift assays using oligonucleotides 1N, 1L and 1Lm (Fig. 1A). The 1N sequence, which corresponds to site A, gave rise to four DNA-protein complexes, C1-C4, with extracts from HeLa cells (Fig. 1B, lane 1) and to three complexes, C1, C2 and C4, with extracts from Jurkat cells (lane 3). Specificity of binding was ascertained by competition experiments using a 50-fold molar excess of unlabeled homologous 1N oligonucleotide. The formation of complexes C1, C2 and C3 was specifically competed by oligonucleotide 1N (lanes 2 and 6). Under similar conditions the formation of complex C4 appeared less specific, since half of the complex was still detected (lanes 2 and 6).

Analysis of complex C1

Identification of the proteins binding to site A was first achieved using competition gel shift experiments. A close inspection of the 1N oligonucleotide revealed the sequence TGGCA, which corresponds to the consensus binding site of proteins of the CTF/NFI family (18). We therefore used as a probe the mutant oligonucleotide 1L containing the TGACA 1 bp mutation; interestingly this mutation abolished formation of complex C1 (lane 4). When oligonucleotide 1L was used in a 50-fold molar excess as a competitor of the 1N probe, formation of complex C1 was unaffected, while formation of complexes C2-C4 was abolished (lane 7). To further assess that the protein forming complex C1 belongs to the NFI family, we used as a competitor oligonucleotide CR, previously shown to be the binding site of the NFI protein in the promoter of the human transferrin gene (25). As expected, in the presence of a 50-fold molar excess of CR competitor, formation of complex C1 was specifically prevented (lane 11). As a control, the mutant oligonucleotide 1Lm, which contains 4 bp mutations, was unable to give any specific complex when used as a probe (lane 5) or to compete for formation of any complex in competition experiments (lane 8). A recent report described that transcription factor YB-1, a member of the NFI family, is able to interact with site A of the HIV-1 LTR (26). However, supershift experiments with antibodies directed against YB-1 did not alter formation of complex C1 (results not shown). Taken together, these results suggest that complex C1 is formed by a member of the NFI protein family, distinct from YB-1.

Analysis of complexes C2-C4

Oligonucleotide 1L contains the TGAC half-site consensus sequence of the cAMP response element (CRE), in direct repeat with the sequence TGAT, which varies by one base from the CRE consensus (19). We therefore performed gel shift assays, using as a competitor in 200-fold excess a CRE oligonucleotide containing the consensus palindromic sequence TGACGTCA. Formation of complexes C2 and C4 was totally abolished (lane 9). As a control, a mutant CRI oligonucleotide containing mutations in the CRE site did not affect formation of complexes C1 and C2 (lane 10). Similarly, when oligonucleotide TRE containing the TGACTCA binding site for proteins of the AP-1 family was used as a competitor, the pattern was unaffected (results not shown). These results suggest that the protein forming complex C2 is related to the ATF/CRE binding protein family. We therefore performed shift assays in the presence of antibodies directed against CREB-1 (27) or ATF-1 (28). All the complexes were unaffected (results not shown), indicating that the CRE-binding protein(s) in complex C2 is (are) different from CREB-1 and ATF-1.

We next tested whether a truncated CREB-1 protein, containing the binding and dimerization domain, was able to bind to oligonucleotides 1L and 1N (Fig. 2). Interestingly, with the 1N probe CREB-1 was able to form a weak retarded complex. With the 1L probe, containing a TGAC site, and with the CRE probe, containing the palindromic TGACGTCA site, CREB-1 gave a more intense retarded complex. This result demonstrates that CREB proteins are able to interact with the 1L sequence and with a weaker affinity with the 1N sequence. Taken together, these results indicate that complex C2 is formed by a protein distinct from CREB-1 and ATF-1 and related to proteins of the CREB/ATF family. In HeLa cells, the additional complex C3 had a behavior similar to that of complex C2 (results not shown),

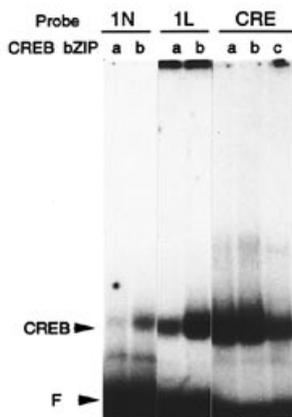


Figure 2. Binding activity of CREB-1 bZIP. Gel retardation assays were performed by incubating different amounts of protein CREB-1 bZIP (from Santa Cruz Biotechnology) with probes 1N, 1L and CRE. Lanes a, b and c correspond to 100, 500 and 10 ng protein respectively.

suggesting that proteins forming complex C3 are also related to the CREB/ATF family.

In order to precisely identify the binding site of proteins forming complex C2, compared with proteins forming complexes C3 and C4, we performed methylation interference experiments using the 1L probe and nuclear extracts from HeLa cells. Figure 3 shows that methylation at the guanine residues at positions -372 and -376 on the coding strand and at position -370 on the non-coding strand resulted in a strong interference with methylation. In addition to a strong interference with methylation of these guanine residues, methylation of the G residues at -378, -379, -380 and -383 of the non-coding strand slightly decreased binding in complex C3. In complex C4, only weak interference with methylation was observed. These results indicate that all the G residues in the TGATTGAC motif on both strands are critical in formation of complex C2. Similar results were obtained for complexes C2 and C4 formed with nuclear extracts from Jurkat cells (results not shown). The importance of these G residues is further confirmed by the results of the gel shift assays presented in Figure 1 (lane 5): with the 1Lm probe, which contains point mutations of the critical G residues, no complex C2 was detected.

Functional importance of the site A binding proteins in Jurkat cells

To evaluate the functional effect of the different proteins interacting with site A of the LTR, we constructed a series of pBLCAT2 reporter plasmids containing one copy of either oligonucleotide 1N, 1L or 1Lm or two copies of 1L or 1Lm in front of the herpes simplex virus thymidine kinase promoter upstream of the *CAT* gene. These vectors were transfected into Jurkat cells and *CAT* activities were determined (Fig. 4). Interestingly, the activity of all the vectors was higher than that of the control pBLCAT2 vector. Transfection of the 1Lm/tk-CAT and 2×1Lm/tk-CAT vectors resulted in 1.8- and 4.0-fold stimulation of *CAT* activity respectively, reflecting the positive action of the protein(s) forming complex C4. Transfection of the 1L/tk-CAT and 2×1L/tk-CAT constructs led to 3.2- and 6.2-fold transcriptional activation respectively, indicating that proteins forming complexes C2 and C4 are able to optimally stimulate transcription.

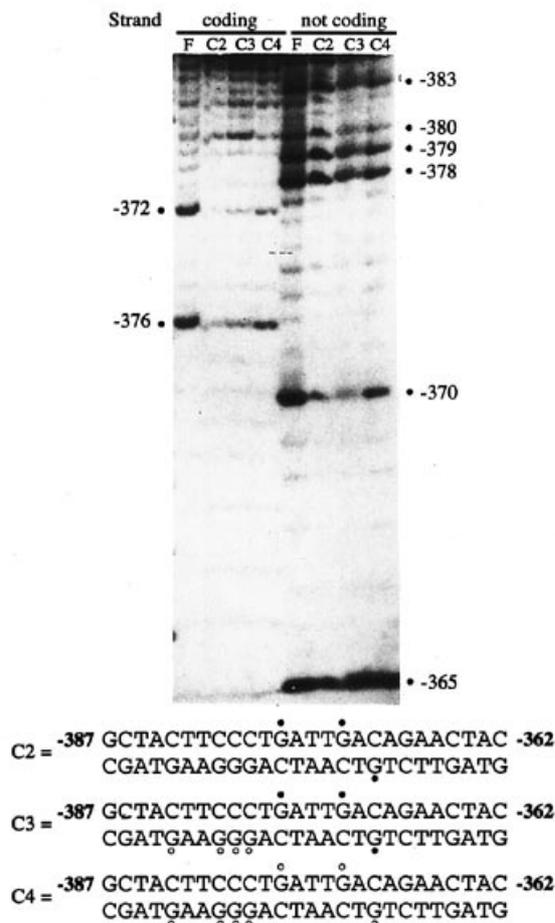


Figure 3. Methylation interference pattern of DNA-protein complexes C2-C4 formed with the 1L probe. Oligonucleotide 1L was labeled at the 5'-end of one strand. Guanine residues were partially methylated with dimethylsulfate. The oligonucleotide was used in a gel shift assay with nuclear proteins from HeLa cells. The DNA-protein complexes C2-C4 and free DNA were eluted from a band shift gel and cleaved at methylated guanine residues with piperidine. The protein-bound DNA was compared with free DNA on a sequencing gel. (Bottom) Summary of the pattern obtained with complexes C2-C4. Closed and open circles indicate strong and weak interference of binding respectively.

Interestingly, the *CAT* activity of the 1N/tk-CAT construct was reduced, when compared with that of 1L/tk-CAT, which could suggest mutual inhibition between proteins forming complexes C1 and C2. Another explanation could be that complex C1 contains a protein(s) that acts as an inhibitor of site A promoter activity and that is overwhelmed by the positive activity of proteins forming complexes C2 and C4.

To examine the effect of mutagenesis of site A in the context of the intact LTR, we have constructed 1L-LTR-CAT and 1Lm-LTR-CAT vectors, by site-directed mutagenesis with the 1L and the 1Lm oligonucleotides respectively (Fig. 5). With the 1L-LTR-CAT vector, where binding of the protein forming complex C1 is prevented, *CAT* activity was increased 1.4-fold, suggesting that this protein acts as a weak inhibitor of LTR-driven transcription. *CAT* activity of the 1Lm-LTR-CAT vector was stimulated 2-fold, suggesting that binding of proteins forming complexes C1 and C2 inhibits LTR transcription. This observation indicates that in the context of both the tk and LTR promoters the protein forming

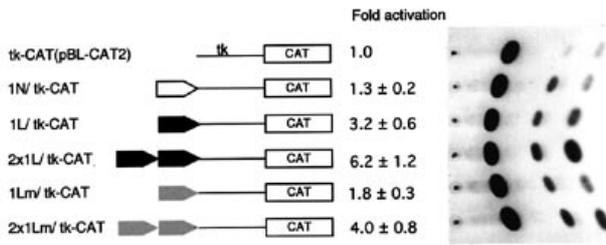


Figure 4. Functional effect of site A in the tk-CAT reporter vectors. Jurkat cells were transfected with pBLCAT2 reporter plasmids containing one or two copies of oligonucleotides 1N, 1L or 1Lm (shown in the right panel). After 48 h, CAT activities were determined as described in Materials and Methods. Values on the right are the mean \pm SE for at least three independent experiments performed in duplicate and are expressed relative to the CAT activity of pBLCAT2, set at 1. An autoradiogram of one typical CAT assay is shown on the right.

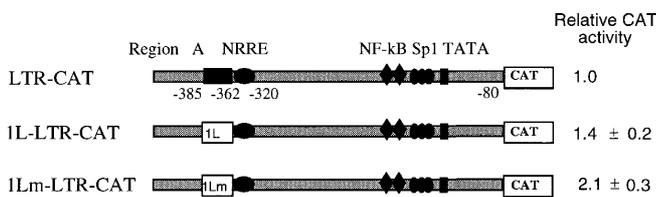


Figure 5. Functional significance of site A within the HIV-1 LTR. Jurkat cells were transfected with the wild-type LTR-CAT and mutant 1L-LTR-CAT and 1Lm-LTR-CAT vectors. CAT activities were determined as in Figure 4 and are expressed relative to that of LTR-CAT.

complex C4 acts as an activator, while the protein forming complex C1 acts as an inhibitor. Proteins forming complex C2 act as activators in the context of the tk promoter and function as inhibitors in the context of the LTR promoter. These data stress both the promoter-dependent activity and the antagonistic effects of the site A binding proteins.

Analysis and purification of a protein present in complex C2

To estimate the apparent molecular mass of the protein(s) present in the major complex C2, we performed UV cross-linking experiments. Gel shift assays were carried out with bromodeoxyuridine-substituted ³²P-labeled 1L probe in the presence of either crude nuclear extract from Jurkat cells or with partially purified proteins eluted from a heparin–Sepharose column. After electrophoresis, the native gel was UV irradiated and the band corresponding to complex C2 was excised and subjected to SDS–10% PAGE. As shown in Figure 6, two oligonucleotide–protein complexes were detected with apparent molecular masses of 67 and 60 kDa. As a control, similar experiments performed with the C4 complex revealed a unique band migrating in the 80 kDa range (results not shown). These results suggest that the C2 complex is composed of two proteins, with apparent molecular masses of ~43 (\pm 5) and 50 (\pm 5) kDa, while the C4 complex is formed of one protein of ~60–66 kDa.

The molecular mass of the protein(s) able to specifically bind to the 1L sequence was further confirmed with the Southwestern technique. This experiment was carried out with crude nuclear extracts and with partially purified proteins eluted at 0.4 M KCl from a heparin–Sepharose column (Fig. 7). When the renatured

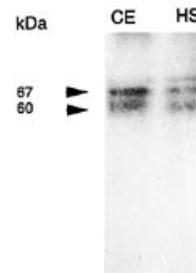


Figure 6. UV cross-linking analysis of Jurkat nuclear proteins forming complex C2. 10 μ g Jurkat crude nuclear extract (CE) and 8 μ g nuclear proteins eluted at 0.4 M KCl from a heparin–Sepharose column (HS) were incubated with a ³²P-labeled bromodeoxyuridine-substituted 1L probe. After separation on a 6% polyacrylamide gel and UV cross-linking, complex C2 was excised and analyzed on a SDS–10% polyacrylamide gel. Arrows indicate the two oligonucleotide–protein complexes of 67 and 60 kDa. Since the molecular mass of oligonucleotide 1L alone is 17 kDa, the molecular mass of the two proteins forming complex C2 is estimated as 50 (\pm 5) and 43 (\pm 5) kDa.

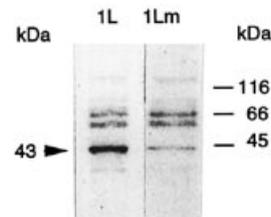


Figure 7. Southwestern blots of partially purified Jurkat proteins binding to site A of the LTR. Aliquots (20 μ g) of Jurkat proteins eluted from a heparin–Sepharose column were analyzed by Southwestern blotting as described in Materials and Methods. The blot was probed with ³²P-labeled oligonucleotides 1L or 1Lm, indicated on top. The labeled proteins were visualized by autoradiography. Molecular weight markers are shown on the right. The left arrow indicates the 43 kDa protein.

proteins were probed with radioactively labeled oligonucleotide 1L, three major protein bands of 43, 62 and 66 kDa were detected. In contrast, when we used the 1Lm probe, the 43 kDa protein was poorly detected, while the 62 and 66 kDa proteins were labeled with a similar intensity as with the 1L probe. A control TRE probe failed to bind to any of the three major species (results not shown). These data confirm that the 43 kDa protein is likely to correspond to one of the proteins forming complex C2. They also suggest that the 50 kDa protein is not easily renatured. Since complex C4 is detected with both the 1L and 1Lm probes in gel shift assays (Fig. 1), the 62 or the 66 kDa species may correspond to the protein forming complex C4.

In order to further characterize the proteins forming complex C2, we developed the following purification strategy. Throughout the purification procedure, 1L binding activity was monitored by gel shift assay (Fig. 8A). Crude nuclear extract from Jurkat cells was first loaded onto a heparin–Sepharose column and proteins were eluted with a step gradient of 0.1 to 1.0 M KCl in buffer Z. Complexes C2 and C4 were detected in fractions eluted at 0.4 M KCl. The pooled fractions were loaded onto a DNA affinity column containing multimers of oligonucleotide 1Lm, to remove most of the proteins forming complex C4. The flow-through fraction, which contained the proteins involved in formation of

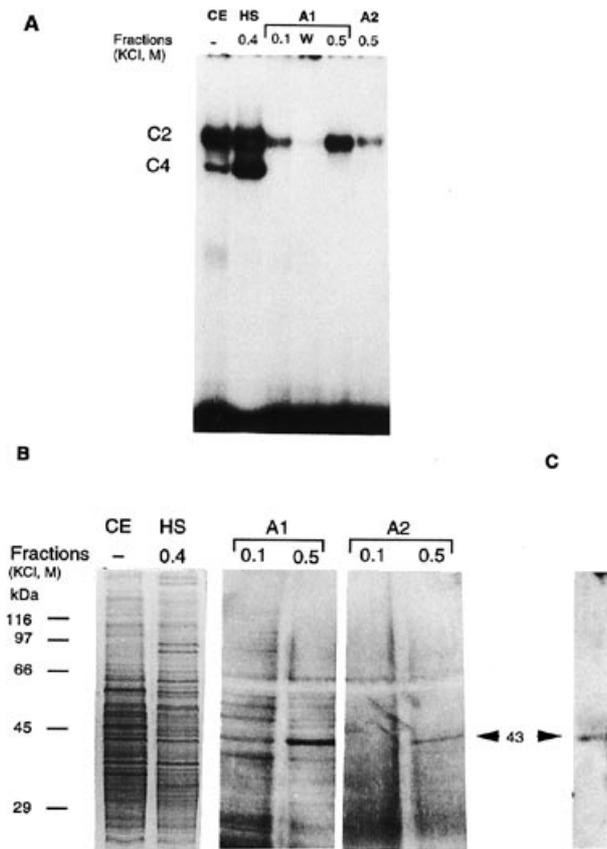


Figure 8. Analysis of the purification steps of p43 by gel shift assay, SDS-PAGE and Southwestern blotting. (A) Gel retardation analysis of crude nuclear extract from Jurkat cells (CE, 5 μ g) and fractions eluted from a heparin-Sepharose column (HS) and an oligonucleotide 1L affinity column (A1 and A2). The binding activity of fractions eluted at 0.1 M KCl, after washing with 0.1 M KCl (w) and at 0.5 M KCl is shown. Retarded complexes C2 and C4 are indicated. (B) SDS-10% PAGE analysis of crude nuclear extract (CE, 25 μ g) and partially purified proteins: 25 μ g 0.4 M KCl heparin-Sepharose fraction (HS), 3 μ g first 1L affinity eluate (A1) and 1 μ g second 1L affinity eluate (A2). Protein bands were stained with Coomassie blue (lanes CE and HS) or silver stained (lanes A1 and A2). The molecular weight markers are shown on the left. (C) Southwestern blot of the 1L affinity eluate. An aliquot (3 μ g) of the 1L affinity eluate (A1, 0.5 M KCl) was analyzed on 10% SDS-PAGE and processed for Southwestern blotting as described in Materials and Methods. The arrow shows the 43 kDa protein.

complex C2, was loaded onto an oligonucleotide 1L affinity column (A1). The 1L binding activity was eluted at 0.5 M KCl and these proteins were recycled on the same column (A2).

The fractions containing the 1L binding activity were analyzed by SDS-PAGE and proteins were visualized by Coomassie blue or silver staining (Fig. 8B). A silver stained gel of proteins present in the 0.5 M KCl fraction A1 showed an enriched band of 43 kDa, which became the predominant protein present in the 0.5 M KCl fraction A2. The 1L-specific binding activity of the purified protein was controlled by Southwestern blotting. Figure 8C shows that p43 present in the 0.5 M KCl A1 fraction was recognized by the 1L probe.

In order to further ascertain that p43 possessed 1L binding activity, proteins purified through heparin-Sepharose were separated by SDS-PAGE and 1 mm slices were excised. After elution, denaturation and renaturation of each sample, the binding

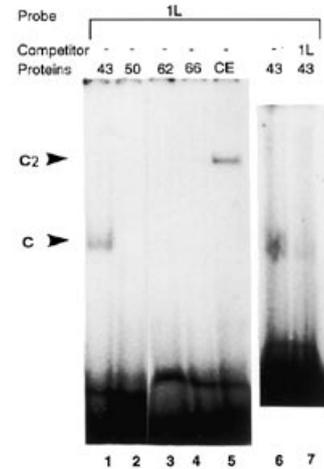


Figure 9. Renaturation of the SDS-PAGE gel-purified 43 kDa protein. An aliquot (20 μ g) of the 0.4 M KCl heparin-Sepharose fraction was fractionated by SDS-PAGE. After electrophoresis, half of the gel was silver stained to visualize the proteins. The other half was used to excise 1 mm slices; after elution, samples containing 43, 50, 62 and 66 kDa proteins were denatured, renatured and assayed for their ability to bind to the 1L probe in a gel shift assay. Complex C2, formed with crude nuclear proteins (CE), and complex C, formed after renaturation of p43, are indicated. The binding of renatured p43 was tested in the absence or presence of a 50-fold molar excess of 1L competitor oligonucleotide (lanes 6 and 7).

activity of the proteins was examined by gel shift assay (Fig. 9). Only the sample containing p43 was able to form a retarded complex C with the 1L probe (lanes 1-4). Interestingly, complex C migrated faster than complex C2 formed with crude nuclear proteins (lane 5), confirming that p43 is one component of the heterodimer present in complex C2. The binding specificity of purified p43 was further demonstrated by competition assay: with an excess of unlabeled oligonucleotide 1L, formation of complex C was completely abolished (lanes 6 and 7). A similar experiment performed with the wild-type 1N oligonucleotide showed that p43 is also able to form a retarded complex C with the 1N probe (results not shown). We tried, with no success, to restore formation of complex C2 by complementation experiments with renatured p43 and proteins eluted from gel slices.

DISCUSSION

Recent studies have revealed the increasing importance of the modulatory region of the LTR in regulation of HIV-1 gene expression. The -356/-320 NRRE is the site of interaction of transcription factors of the nuclear receptor superfamily which regulate transcription in a cell-type specific manner (10-14,16). However, the nature and role of nuclear host proteins which interact with the previously described (13,16,17) -385/-362 adjacent region A remain to be investigated.

Here we describe a preliminary characterization of the nature and functional role of nuclear proteins present in HeLa and Jurkat cells which interact with region A. Our studies reveal the complexity of the interactions, since multiple distinct protein species have the ability to interact with this distal region. In Jurkat and HeLa cells respectively three and four distinct DNA-protein

complexes were formed in gel mobility shift assays with oligonucleotide 1N, corresponding to site A.

Our gel retardation competition experiments suggest that the protein forming complex C1 belongs to the NFI family of proteins. Moreover, our data indicate that the proteins forming the specific complex C2 in HeLa and Jurkat cells and complex C3 in HeLa cells interact with the core sequence 5'-TGATTGGC-3'. The importance of this core sequence for binding of the distinct proteins was revealed both by methylation interference and by gel mobility shift data obtained with the 1 nt mutated oligonucleotide 1L, containing the modified sequence TGATTGAC. While this mutation abolished formation of complex C1, it increased the binding ability of proteins forming complexes C2 and C3. Interestingly, the core sequence corresponds to a direct repeat of a mutant TGAC half-site of the consensus CRE, TGACGTCA (19). Our competition data with the CRE and mutant oligonucleotides raise the possibility that proteins forming complexes C2 and C3 may be related to the CREB/ATF family. However, our data show that CREB-1 and ATF-1 are not present in complexes C2 and C3. It is known that many proteins are able to bind *in vitro* to CRE-related elements and that members of the ATF/CREB family form cross-family heterodimers with members of the AP-1 family (29). In that respect it is worth noting that proteins forming complexes C2 and C3 were not able to interact with the related TRE binding site of proteins of the AP-1 family. Interestingly, our data show that although the 1L sequence contains a mutated direct repeat of the CRE consensus sequence, it was able to bind the CREB-1 protein. As expected, the binding affinity of CREB-1 for the 1N sequence was lower than that for the 1L and CRE sequence. Interaction of ATF/CREB proteins with the LTR region of human T cell lymphotropic virus type I (HTLV-I) has been reported (30). However, binding to the HIV-1 LTR of a protein related to the CREB family has not yet been described.

The functional role of the site A binding proteins was first examined in Jurkat cells by transfecting tk-CAT reporter vectors containing the different 1N, 1L and 1Lm oligonucleotides. Transient expression data revealed that in the context of the tk promoter, all the nuclear proteins binding to the various 1L, 1N and 1Lm sequences, corresponding to the wild-type and mutated site A, were able, to a different extent, to activate transcription. Transcriptional activity was highest with the 1L sequence, forming the major complex C2 in gel shift assays. Activity was decreased with the 1N sequence, forming the two major specific complexes C1 and C2, suggesting either mutual inhibition between two proteins binding to the same site or an ability of the protein present in complex C1 to act as a repressor. In addition, the functional role of the various site A binding proteins was established in the context of the entire HIV-1 LTR region. Transcriptional activity was highest with the 1Lm-LTR-CAT vector, was decreased with 1L-LTR-CAT and was lowest with the wild-type LTR-CAT vector. These data reveal that in the context of the LTR, proteins forming complexes C1 and C2 are able to inhibit the positive action of protein(s) forming complex C4.

Since the protein(s) forming complex C2 functions as a modulator(s) of HIV-1 gene transcription, we deemed it essential to analyze these factors more closely. According to UV cross-linking data, complex C2 is formed by a heterodimer consisting of 43 and 50 kDa proteins. As suggested by our gel competition data, at least one member of this heterodimer has a DNA binding domain able to recognize the CRE sequence. It will be interesting to

determine in future which member of the dimer possesses the CRE binding activity.

DNA affinity chromatography, using concatemers of oligonucleotide 1L and 1Lm, was a convenient method to purify a protein of relative molecular mass 43 000 which was recognized by the 1L probe on a Southwestern blot. After elution and renaturation, p43 was able to specifically bind to the 1L probe and to form a DNA-protein complex C of faster mobility than complex C2 formed with crude nuclear proteins, which is consistent with our UV cross-linking data.

Protein p43 appears to be a novel nuclear factor, since a search for homologous N-terminal amino acid sequences in protein databanks did not reveal any homology with a known protein (M.H.Metz-Boutigue; C.Schwartz, D.Aunis and E.Schaeffer, unpublished results). To definitely resolve the question of whether p43 or p50 are related to the CREB/ATF family, it will be necessary to carry out comparative studies using cloned cDNAs. Therefore, cloning of the corresponding cDNA represents an essential step for complete identification of this protein and for elucidation of its role in modulation of HIV-1 gene transcription. Interestingly, region A is a binding site for proteins of lymphoid and non-lymphoid cells; future studies are necessary to investigate precisely the nature as well as the role of the cellular site A binding proteins in different cell types infected by HIV-1.

ACKNOWLEDGEMENTS

We thank H.D.Royer for YB-1 antibodies, B.E.Sawaya for helpful discussions and D.Filliol for excellent technical assistance. This work was supported by the Institut National de la Santé et de la Recherche Médicale (INSERM), the Agence Nationale des Recherches sur le SIDA, the Association 'Ensemble contre le SIDA', the Fondation pour la Recherche Médicale (FRM) and the Association pour la Recherche sur le Cancer. C.S. received a fellowship from INSERM and F.C.H. from FRM.

REFERENCES

- 1 Gaynor,R. (1992) *AIDS*, **6**, 347-363.
- 2 Garcia,J.A. and Gaynor,R.B. (1994) *Prog. Nucleic Acid Res. Mol. Biol.*, **49**, 157-196.
- 3 Nabel,G.J. and Baltimore,D. (1987) *Nature (Lond.)*, **326**, 711-713.
- 4 Griffin,G.E., Leung,K., Folks,T.M., Kunkel,S. and Nabel,G. (1989) *Nature (Lond.)*, **339**, 70-73.
- 5 Rattner,A., Korner,M., Walker,M.D. and Citri,Y. (1993) *EMBO J.*, **12**, 4261-4267.
- 6 Kaltschmidt,C., Kaltschmidt,B., Neumann,H., Wekerle,H. and Baeuerle,P.A. (1994) *Mol. Cell. Biol.*, **14**, 3981-3992.
- 7 Taylor,J.P., Pomerantz,R.J., Raj,G.V., Kashanchi,F., Brady,J.N., Amimi,S. and Khalili,K. (1994) *J. Virol.*, **68**, 3971-3981.
- 8 Zeichner,S.L., Hirka,G., Andrews,P.W. and Alwine,J.C. (1992) *J. Virol.*, **66**, 2268-2273.
- 9 Franza,B.R., Jr, Rauscher,F.J., Josephs,S.F. and Curran,T. (1988) *Science*, **239**, 1150-1153.
- 10 Cooney,A.J., Tsai,S.Y., O'Malley,B.W. and Tsai,M.-J. (1991) *J. Virol.*, **65**, 2853-2860.
- 11 Lee,M.I., Hobbs,P.D., Zhang,X.K., Dawson,M.I. and Pfahl,M. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 5632-5636.
- 12 Ladias,J.A.A. (1994) *J. Biol. Chem.*, **269**, 5944-5951.
- 13 Canonne-Hergaux,F., Aunis,D. and Schaeffer,E. (1995) *J. Virol.*, **69**, 6634-6642.
- 14 Sawaya,B.E., Rohr,O., Aunis,D. and Schaeffer,E. (1996) *J. Biol. Chem.*, **271**, 22895-22900.
- 15 Deacon,N.J., Tsykin,A., Solomon,A., Smith,K., Ludford-Menting,M., Hooker,D.J., McPhee,D.A., Greenway,A.L., Ellett,A., Chatfield,C., et al. (1995) *Science*, **270**, 988-991.

- 16 Orchard,K., Perkins,N., Chapman,C., Harris,J., Emery,V., Goodwin,G., Latchman,D. and Collins,M. (1990) *J. Virol.*, **64**, 3234–3239.
- 17 Demarchi,F., d'Agaro,P., Falaschi,A. and Giacca,M. (1992) *J. Virol.*, **66**, 2514–2518.
- 18 Rupp,R.A.W., Multhaup,G., Göbel,U., Beureuther,K. and Sippel,A.E. (1990) *Nucleic Acids Res.*, **18**, 2607–2616.
- 19 Lalli,E. and Sassone-Corsi,P. (1994) *J. Biol. Chem.*, **269**, 17359–17362.
- 20 Luckow,B. and Schütz,G. (1987) *Nucleic Acids Res.*, **15**, 5490.
- 21 Guillou,F., Zakin,M.M., Part,D., Boissier,F. and Schaeffer,E. (1991) *J. Biol. Chem.*, **266**, 9876–9884.
- 22 Marriott,S.J., Lindholm,P.F., Brown,K.M., Gitlin,S.D., Duvall,J.F., Radonovich,M.F. and Brady,J.N. (1990) *Mol. Cell. Biol.*, **10**, 4192–4201.
- 23 Kadonaga,J.T. and Tjian,R. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5889–5893.
- 24 Hager,D.A. and Burgess,R.B. (1980) *Anal. Biochem.*, **109**, 76–86.
- 25 Brunel,F., Ochoa,A., Schaeffer,E., Boissier,F., Guillou,Y., Cereghini,S., Cohen,G.N. and Zakin,M.M. (1988) *J. Biol. Chem.*, **263**, 10180–10185.
- 26 Kashanchi,F., Duvall,J.F., Dittmer,J., Mireskandari,A., Reid,R.L., Gitlin,S.D. and Brady,J.N. (1994) *J. Virol.*, **68**, 561–565.
- 27 Montminy,M.R. and Bilezikjian,L.M. (1987) *Nature (Lond.)*, **328**, 175–178.
- 28 Hai,T., Liu,F., Coukos,W.J. and Green,M.R. (1989) *Genes Dev.*, **3**, 2083–2090.
- 29 Hai,T. and Curran,T. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 3720–3724.
- 30 Zhao,L.J. and Giam,C.Z. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 7070–7074.