

Unbiased proteomic analysis of proteins interacting with the HIV-1 5'LTR sequence: role of the transcription factor Meis

A. Tacheny, S. Michel, M. Dieu, L. Payen, T. Arnould and P. Renard*

Laboratory of Biochemistry and Cell Biology (URBC), NAMur Research Institute for Life Sciences (NARILIS), University of Namur, 61 rue de Bruxelles, 5000 Namur, Belgium

Received January 23, 2012; Revised June 28, 2012; Accepted July 10, 2012

ABSTRACT

To depict the largest picture of a core promoter interactome, we developed a one-step DNA-affinity capture method coupled with an improved mass spectrometry analysis process focused on the identification of low abundance proteins. As a proof of concept, this method was developed through the analysis of 230 bp contained in the 5' long terminal repeat (LTR) of the human immunodeficiency virus 1 (HIV-1). Beside many expected interactions, many new transcriptional regulators were identified, either transcription factors (TFs) or co-regulators, which interact directly or indirectly with the HIV-1 5'LTR. Among them, the homeodomain-containing TF myeloid ectopic viral integration site was confirmed to functionally interact with a specific binding site in the HIV-1 5'LTR and to act as a transcriptional repressor, probably through recruitment of the repressive Sin3A complex. This powerful and validated DNA-affinity approach could also be used as an efficient screening tool to identify a large set of proteins that physically interact, directly or indirectly, with a DNA sequence of interest. Combined with an *in silico* analysis of the DNA sequence of interest, this approach provides a powerful approach to select the interacting candidates to validate functionally by classical approaches.

INTRODUCTION

The real crossing points between the genome and the proteome of an organism are transcription factors (TFs). They are much more than simple sequence-specific proteins bound or not on conserved *cis*-regulatory sequences in gene promoters/enhancers (1). Indeed, these factors are responsible for the fine coordination of gene

expression by modulating chromatin accessibility, general transcriptional machinery recruitment and also for coordinating interplays between transcription and other nuclear processes such as DNA repair or RNA processing and stability (2). However, despite the importance of deciphering such complex regulatory mechanisms, systematic identification of DNA-protein interactions occurring on regulatory regions of interest is still challenging.

In spite of recent progresses due to the development of chromatin immunoprecipitation (ChIP)-based methods such as ChIP-Seq, allowing genome-wide analysis of all the DNA sequences bound by a protein of interest (3,4), methods identifying proteins that interact with a sequence of interest are still poorly developed. Indeed, *in silico* analysis of the sequence of interest allows the prediction of putative binding sites for TF based on the comparison with consensus-binding sites contained in TF databases such as TRANSFAC (5) or JASPAR (6). However, the results are limited to database-contained TF, highly depend on the algorithm used, do not take into account binding site context like flanking sequences and chromatin organization (7). Thus, they generate a very large number of candidates among which many false positives occur (8). The selection of relevant candidates to validate by classical and time-consuming approaches, including DNA-binding assays, ChIP or reporter-based assays, is therefore uncertain.

The past decades have witnessed the progressive development of DNA-affinity approaches combining the capture of DNA-binding proteins on oligonucleotide probes fixed on a chromatographic support followed by the identification of captured proteins by mass spectrometry (MS) (9–11). Such approaches, although quite simple in their principle, are really challenging for essentially two reasons. First, most transcriptional regulators are of low abundance when compared with the bulk of other nuclear proteins. This problem of dynamic range makes critical the efficiency of the capture and the sensitivity of the MS-based identification process. Second, number of

*To whom correspondence should be addressed. Tel: +32 81 72 41 28; Fax: +32 81 72 41 35; Email: patsy.renard@fundp.ac.be

proteins, some of them of high abundance, is unspecifically captured by the negatively charged oligonucleotide probe and/or by the chromatographic support. Although different strategies have been proposed to improve the specificity of the DNA-affinity capture, such as prefractionation of nuclear extracts (NE) on successive columns prior to the DNA-affinity purification (12–14) or use of DNA competitors or detergents added before and during binding step (9,15), the major drawback of such strategies is the non-negligible risk to lose weak specific interactions (16). Separation of DNA/proteins complexes from the solid support before protein identification is another way to limit contamination of the results by proteins trapped by the solid support (17). Despite such improvements, relevant identified proteins are still embedded in large amounts of unspecifically bound proteins. Therefore, current DNA-affinity methods are mostly used to compare proteins captured by a short wild-type DNA sequence to those captured by the same sequence in which the binding site of interest has been mutated, providing a list of proteins to subtract (11,18). This principle has been successfully implemented using quantitative proteomics based on isotope-coded affinity tag (ICAT) (19–22) or stable isotope labeling by amino acids in cell culture (SILAC) (23) methodologies, a powerful strategy to identify the TF selectively captured by a specific binding site. Although efficient, this strategy is not fully compatible with an unbiased identification of the entire set of proteins interacting with a relatively large DNA sequence like a core promoter.

In this article, we describe an improved DNA-affinity method allowing the one-step identification of an unbiased large set of transcriptional regulators interacting with a relatively long capture probe (over 200 bp), that could correspond to a core promoter interactome. This was made possible by an efficient separation procedure of the protein/DNA complexes from the solid support, by an adapted chromatographic separation of the complex peptide mixture and by a specific MS analysis focused on the identification of low abundant proteins. The proof of concept of this method was made for the analysis of a fragment of the human immunodeficiency virus (HIV)-1 5' long terminal repeat (LTR), a DNA sequence that contains numerous TF-binding sites. Although this is probably one of the best studied regulatory DNA sequence (24), using this approach, several TF/regulators that were not known to interact with the HIV-1 5'LTR were identified. Among them, we identify myeloid ectopic viral integration site (Meis) that functionally interacts with and down-regulates the transcription of a HIV-1 5'LTR-luciferase construct.

MATERIALS AND METHODS

Cell culture and nuclear extract preparation

HeLa cells (human epithelial cell line, ATCC CCL-2) were grown in Dulbecco's modified eagle medium (DMEM) high glucose (Gibco) supplemented with 10% foetal bovine serum (FBS). The T-lymphoid Jurkat cell line (ATCC TIB-152) was cultured in Roswell Park Memorial Institute (RPMI) 1640 medium (Gibco)

supplemented with 10% FBS. Nuclear protein extracts were obtained from unstimulated or interleukin (IL)-1 β -stimulated (5 ng/ml during 45 min) HeLa cells were isolated as previously described (25). Protein concentrations were determined using the Pierce 660 nm protein assay (Thermo).

Plasmids

The pLTRwt plasmid containing a fragment of the HIV-1 5'LTR (corresponding to nt 1–789 where nt 1 is the start of the 5'LTR U3 region) upstream of the *Firefly* luciferase gene and the pLTR* κ B mutated for two NF κ B-binding sites were previously described (26) and generously given by Prof. C. Van Lint (IBMM, ULB, Belgium). The pLTRwt was used as a template to generate the pLTR*Meis mutant variant using the QuikChange site-directed mutagenesis kit (Stratagene) following the recommendations of the supplier. The point mutation was introduced in the Meis putative binding site with the following pair of primers: Fw: 5'-GTGTTAGAGTG GAGGTTTCACAGCCGCTAGCATTTC-3' and Rv: 5'-GAAATGCTAGGCGGCTGTGAAACCTCCACTC TAACAC-3' in which the Meis putative binding site is underlined and the introduced mutation is highlighted in boldface. The mutation was confirmed by sequencing.

Capture probe production and DNA-affinity approach

A 226-bp-long desthiobiotinylated double-stranded oligonucleotide corresponding to a fragment of the HIV-1 5'LTR (nt 229–455, where nt 1 is the start of the 5'LTR U3 region) was produced by polymerase chain reaction (PCR) using pLTR plasmid as template and the following pair of primers: Fw: 5-TGGATGACCCTGAGAGAG AA-3' and Rv: 5'-CCAGTACAGGCAAAAAGCAG-3'. These primers were modified to allow a reversible immobilization of the capture probe to streptavidin-coated magnetic beads (by adding a 5'-desthiobiotin moiety on the forward primer) and to estimate the probe binding efficiency on beads (Cy3-labelled reverse primer). To eliminate excess of free desthiobiotinylated primers after oligonucleotide amplification, the PCR product was purified using the Wizard SV Gel and PCR clean-up system (Promega). Using the pLTR* κ B as a template, a 226-bp-long capture probe mutated for two NF κ B-binding sites was also produced. In order to study specific protein interactions at the level of the Meis-1 putative binding site, two 101-bp-long capture probes centered on this site (nt 229–330) were also produced using the following pair of primers: Fw: 5'-TGGATGACCCTGAGAGAGAA-3' and Rv: 5'-GCAGTTCTTGA AGTACTCCG-3' and pLTR or pLTR*Meis as template for a wild type or a mutated version of the Meis-centred capture probe, respectively.

20 pmoles of capture probe in a final volume of 100 μ l of phosphate buffered saline (PBS)₅₀ (10 mM PO₄ pH 7.4, 50 mM NaCl) were incubated for 1 h at 21°C on a rotary wheel with 1 mg of streptavidin-coupled magnetic beads (Dynabeads M-280 Streptavidin, Dynal) that have been prior equilibrated with six successive washes (200 μ l of PBS₅₀). Three washes with PBS₅₀ were then performed

in order to eliminate unbound desthiobiotinylated oligonucleotides prior to the incubation with proteins. An equivalent of 1 mg of NE was pre-incubated for 15 min on ice with 1.5× volume of binding buffer [4 mM Hepes, pH 7.5, 120 mM KCl, 8% glycerol, 2 μM dithiothreitol (DTT), salmon sperm DNA (0.166 μg/μl), PolydIdC (0.166 μg/μl)] and then incubated with beads for 1 h at 21°C on a rotary wheel. Magnetic beads were then extensively washed: once with 500 μl of binding buffer, three times with 1 ml of PBS₅₀ + 0.1% Tween20 and twice with 50 mM NH₄HCO₃. In order to separate the DNA–protein complexes from beads, beads were re-suspended in a biotin-containing solution (30 μl of 5 mM biotin in 50 mM NH₄HCO₃) and incubated for 2 h at 21°C on a rotary wheel. Supernatant was containing the DNA–proteins complexes were then collected. Eventually, beads were washed with 10 μl of 5 mM biotin and both supernatants were combined.

Proteins bound to capture probes were next prepared for tryptic digestion. Proteins were boiled 5 min with PPS (3-[3-(1,1-bisalkoxyethyl)pyridin-1-yl]propane-1-sulfonate) Silent Surfactant (Protein Discovery; 0.8 % final concentration), reduced for 30 min at 50°C with DTT (Sigma; 5 mM final concentration) and then alkylated for 30 min in dark using iodoacetamide (15 mM final concentration). Samples were then digested overnight at 37°C with 1.8 μl of trypsin (1 μg/μl in 50 mM NH₄HCO₃, 1 mM CaCl₂; Trypsin Gold, Mass Spectrometry Grade; Promega). After digestion, samples were incubated for 30 min at 21°C on a rotary wheel in the presence of fresh streptavidin-coupled magnetic beads (600 μg) that have been prior equilibrated with four successive washes with PBS₅₀ followed by two washes of 50 mM NH₄HCO₃. Once the excess of free biotin has been captured and prior to MS run, samples were acidified by adding 1 μl of 12 N HCl and PPS detergent was hydrolysed by a 45-min incubation at 37°C followed by 10 min of centrifugation (13 000 RPM) at 4°C.

Liquid chromatography/tandem mass spectrometry analysis and protein identification

Peptide analyses were performed on a nano-liquid chromatography (LC) system Ultimate 3000 (Dionex) directly coupled to a maXis 4G electrospray Ultra-High Resolution Q-TOF mass spectrometer (Bruker). Peptides were separated at 60°C by reverse-phase LC using a 75 μm × 500 mm C18 Dionex column (Acclaim PepMap 100 C18) in the Ultimate 3000 LC system. Mobile phase A was 0.1% formic acid in water. Mobile phase B was 0.1% formic acid in 80% acetonitrile. For each run, 15 μl of the digest product was injected and the organic content of the mobile phase was linearly increased from 4% B to 40% during 150 min, from 40% B to 90% B for the next 5 min, maintained at 90% B for 5 min and then decreased at 4% B for the last 25 min. The column effluent was connected to an electrospray ionization (ESI) nano Sprayer (Bruker). As the peptide mixture was complex and presented a high dynamic range, the complete analysis was composed of two successive runs. For the first one, a scheduled precursor list (SPL) was generated

on the basis of the most abundant ions sequenced. During the second run, SPL ions were excluded to allow the mass spectrometer to be more focused on less abundant ions. Peak lists from both runs were generated using DataAnalysis 4.0 (Bruker) and saved for use with ProteinScape 2.1 (Bruker) with Mascot 2.2 as search engine (Matrix Science). The peak lists were searched against a decoy database obtained from the mammalian National Center of Biotechnology Information (NCBI) database. For each protein identification, a false-positive rate (FPR) was determined, corresponding to the number of decoy entries/total number of protein entries in the list. Peptide sequences were accepted if the peptide Mascot score was >20 (15 for the differential DNA-affinity capture with the Meis-centred capture probe). Protein identification was accepted if the FPR was <1% otherwise if the protein Mascot score was >60. Protein identification was further analysed and manually verified by using ProteinScape 2.1.

Cell transfection and luciferase reporter assays

HeLa cells were transiently co-transfected with 0.9 μg of *Firefly* luciferase reporter plasmid (pLTRwt or pLTR*Meis) and 0.1 μg pCMV-βgal (for transfection efficiency normalization) using Superfect (Qiagen) following the manufacturer's instructions. Cells were seeded 1 day before transfection at 70 000 cells/well in 12-well plates. The DNA/Superfect ratio was 1/2 (μg/μl). Twenty-four hours post-transfection, cells were stimulated or not with IL-1β (5 ng/ml) for an extra 24 h. Luciferase reporter assay (Promega) was performed, as well as β-galactosidase activities to normalize luciferase activities.

For Meis-1 knockdown assays, small interfering RNA (siRNA) transfection of HeLa cells was performed with Oligofectamine (Invitrogen) following the manufacturer's instructions 24 h before the DNA transfection. ON-TARGET plus SMARTpool siRNA against human Meis-1 (Thermo, L-011726-00) and ON-TARGET plus non-targeting pool (Thermo) were used at the final concentration of 50 nM. Efficient knockdown after siRNA transfection was confirmed after 24 h at the protein level by western blot using goat polyclonal antibodies against Meis-1/2 (Santacruz).

Jurkat cells were transiently co-transfected with 0.45 μg of the *Firefly* luciferase reporter plasmid (pLTRwt or pLTR*Meis) and 0.05 μg p*Renilla*Luc-TK (for transfection efficiency normalization) using Jurkat Trans-IT (Mirus) according to the manufacturer's instructions. Cells were seeded the day of transfection at 300 000 cells/well in 24-well plates. The DNA/Jurkat Trans-IT agent ratio was 1/2.5 (μg/μl). Twenty-four hours post-transfection, cells were stimulated or not with IL-1β (5 ng/ml), TNFα (10 ng/ml) or PMA (25 ng/ml) + ionomycin (500 nM). After 24 h of treatment, transfected cells were lysed and luciferase activities were measured using the dual luciferase reporter assay kit (Promega). *Firefly* luciferase activities were normalized for the *Renilla* luciferase activities. All assays were performed in three independent experiments containing triplicates and results were presented as means ± SD, and differences were statistically analysed

by a Student's *t*-test (*P* values are indicated in the figure legends).

RESULTS

Unbiased proteomic analysis of proteins interacting with a DNA sequence: proof of concept

DNA-affinity capture

As a proof of concept for this method, a 226-bp-long sequence of the HIV-1 5'LTR was used as a model capture probe. This sequence covers the core promoter, the enhancer and a short fragment of the modulatory region of the HIV-1 5'LTR. This well-described promoter can be activated or repressed by the binding of several cellular TF (27,24).

Streptavidin-coated magnetic beads were first chosen to immobilize the biotinylated DNA capture probe. Nevertheless, the identification of proteins captured by this complex revealed exclusively unspecifically bound proteins (mainly bovine serum albumin and actin; data not shown), demonstrating the necessity to dissociate the DNA-protein complex from the solid support as suggested by Praseuth and co-workers (17). We therefore took advantage of the reversibility of the desthiobiotin/streptavidin interaction (28). The same DNA sequence, amplified by PCR using a desthiobiotinylated forward primer and a Cy3-labelled reverse primer, was immobilized on streptavidin-coated beads and then displaced from the beads by an excess of biotin. The yields of the oligonucleotide captured by magnetic beads and of the recovery in the presence of biotin were quantified by measuring the Cy3-associated fluorescence. The binding efficiency in optimized conditions represents 60% and this yield could not be increased by higher amounts of oligonucleotide. However, the recovery yield after displacement in the presence of a free biotin excess is total, allowing an efficient recovery of the DNA-protein complexes before MS analysis (data not shown). Recovery using restriction enzyme or photocleavable biotin-based methods was also tested but gave lower recovery yields (data not shown).

Protein separation and identification

Identification of proteins interacting with a relatively long DNA sequence is of great biological interest but represents a technical challenge, as it requires the analysis of a complex mixture of peptides generated by the digestion of many proteins for which abundance might be distributed on a high dynamic range. Current progress in nano-LC-tandem mass spectrometry (MS/MS) encouraged us to choose a gel-free approach rather than a one- or two-dimensional electrophoresis protein separation that might suffer from poor detection of low abundant proteins (29). Proteins purified by DNA affinity were digested with trypsin. Of note, in-solution tryptic digestion had to be improved as regular digestion protocols generated incompletely digested proteins (data not shown). The complex peptide mixture obtained underwent a long reverse-phase chromatography separation before being sequenced by MS (Figure 1). To cope with

the problematic large dynamic range between highly abundant non-specifically bound proteins and potentially low abundant transcriptional regulators, MS/MS analysis was adapted to proceed in two successive runs for each sample. Raw data from the first run, largely composed of peptides from abundant bound proteins, were used to generate a SPL. This list contained the most abundant peptides sequenced during the first run and was then used as a peptide exclusion list during the second run to specifically focus on the sequencing of low abundant peptides. Peptides sequenced from both runs were then merged to proceed to protein identification.

Identification of the TF NFκB captured by the HIV-1 5'LTR sequence

The process described above and illustrated in Figure 1 was first applied, as a proof of concept, to identify a TF previously reported to bind the HIV-1 5'LTR sequence. We chose NFκB as the HIV-1 5'LTR sequence contains two distinct κB-binding sites essential for efficient HIV-1 transcription (30). Although the NFκB family contains several members (p105/p50, p100/p52, c-Rel, RelB and RelA) (31), the p50/RelA heterodimer is the most common transcriptionally active form described to interact with the HIV-1 5'LTR (32). NFκB is a TF sequestered in the cytoplasm in most resting cell types, which can be activated and translocated in the nucleus after cell stimulation with IL-1β, among others (33). Thus, comparing NE prepared from unstimulated and IL-1β-stimulated cells provides conditions for differential binding of NFκB to the HIV-1 5'LTR.

The NFκB family members identified in these conditions are listed in Table 1. In NE from unstimulated cells, p50 is the only NFκB member captured by the probe and identified. This is consistent with the repressive role of p50 homodimer, due to the lack of a transactivation domain, known to contribute to viral latency (34). By contrast, after stimulation with IL-1β, various peptides corresponding to different proteins belonging to the NFκB family were identified. The large number of peptides sequenced for p50 and RelA suggests that these proteins are the most abundant NFκB-family members captured by the probe, but it is not surprising to find peptides of c-Rel and p52 members, which are also expressed in HeLa cells (35). The specificity of this interaction was assessed by using a capture probe with NFκB-mutated binding sites (26). Indeed, when the mutated capture probe was incubated with NE prepared from IL-1β-stimulated HeLa cells, no peptides of any NFκB-family members were identified among proteins captured, confirming the specificity of the assay.

In-depth analysis of proteins that directly or indirectly interact with a fragment of the HIV-1 5'LTR

This analysis aims to depict the most general picture of the proteins bound to a DNA sequence of interest, in a completely unbiased approach, in order to identify putative new interacting partners. The analysis of relatively long DNA sequences implies dealing with a large number of proteins, requiring a rigorous step-by-step data

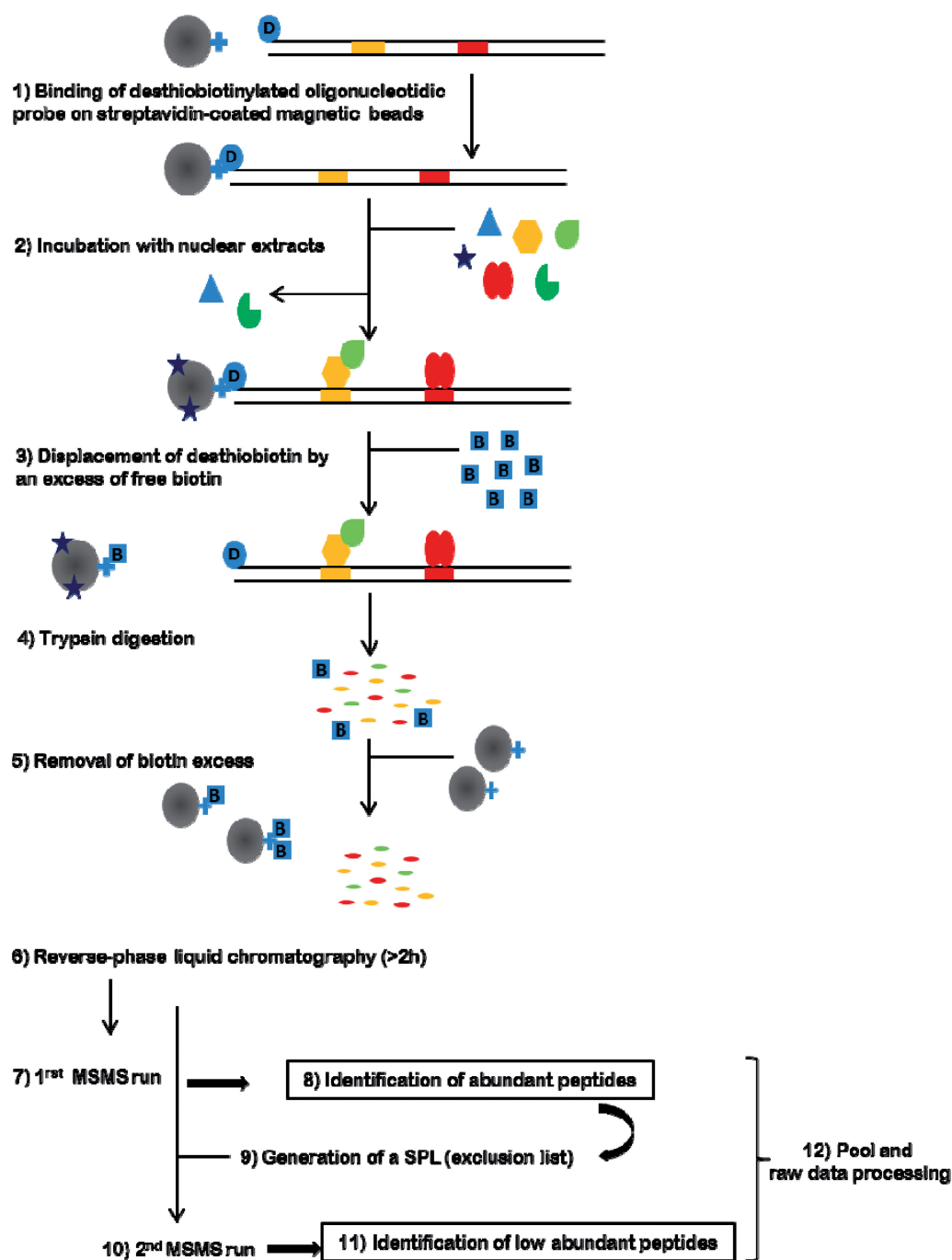


Figure 1. Schematic representation of the DNA-affinity capture method coupled with an improved mass spectrometry analysis process.

processing, detailed in Supplementary Figure S1. This analysis process takes into account the FPR, the quality of the sequencing (indicated by a minimum Mascot score), the redundantly identified proteins with several names and the technical contaminants previously defined by a “blank” experiment consisting of a MS identification of proteins captured during the complete procedure by oligonucleotide probes incubated without any NE (Supplementary Table S1). For instance, for one biological sample, 79 proteins were finally considered, out of 259 initially identified candidates listed in the raw data.

Eventually, to reinforce protein identification resulting from the purification process, sets of data were accumulated from a minimum of three independent biological replicates. In this study, all the identified proteins captured by the 226-bp-long HIV-1 5’LTR were pooled in a global list and proteins identified only once were not considered. A total protein number of 125 was generated after this global data treatment (111 identified proteins from unstimulated samples and 117 from IL-1 β -stimulated samples). According to Uniprot general protein annotations and to information from the

Table 1. NFκB family proteins identified by LC-MS/MS after DNA-affinity capture

Capture probe	HeLa nuclear extracts	Identified NFκB family proteins	Uniprot accession no.	Mascot score	Peptide number	SC (%)
WT	Unstimulated	Nuclear factor NFκB p105 subunit/p50	P19838	62.67	2	4.3
WT	IL-1β	Transcription factor RelA (p65)	Q04206	504.6	68	30.2
		Nuclear factor NFκB p105 subunit/p50	P19838	660.55	52	16.6
		Nuclear factor NFκB p100 subunit/p52	Q00653	533.79	49	34.2
		Proto-oncogene c-Rel	Q04864	263.85	22	13.7
NFκB-mutated	IL-1β	None				

A fragment of the HIV-1 5’LTR (nt 229–455), wild-type or mutated for two NFκB-binding sites, was used as a capture -probe and incubated in the presence of nuclear extracts from unstimulated or IL-1β-stimulated HeLa cells. LC-MS/MS data were compared by Mascot to a decoy database and only identified proteins with a false-positive rate <1%, otherwise a protein Mascot score >60, were accepted. SC represents the percentages of sequence coverage for each identified protein.

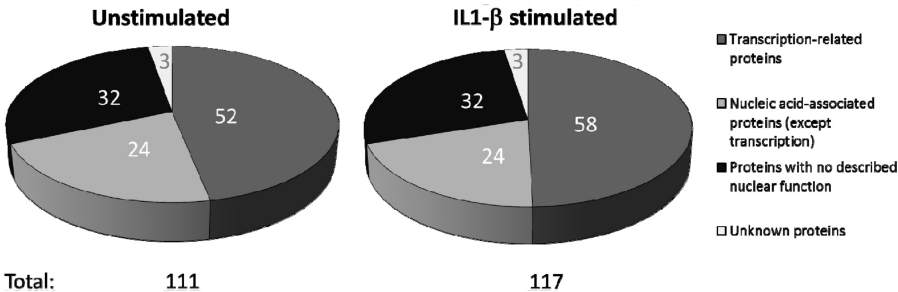


Figure 2. Functional classification of proteins identified by LC-MS/MS after DNA-affinity capture. A fragment of the HIV-1 5’LTR (nt 229–455) was used as a capture probe and incubated in the presence of nuclear extracts from unstimulated or IL-1β-stimulated HeLa cells. LC-MS/MS data were compared by Mascot to a decoy database and only identified proteins with a false-positive rate <1% otherwise a Mascot score >60, were accepted. Each experimental condition was reproduced independently three times. Identified proteins from negative control analysis (“blank”) or identified proteins present only in one of the six analyses were subtracted from the total number of identified proteins.

literature, proteins identified from both experimental conditions were further classified into four categories based on their functions, specifically, transcription-related proteins, nucleic acid-binding proteins without described implication in transcription, proteins without any reported nuclear function and proteins with unknown function. As shown in Figure 2, a large majority of identified proteins might be implicated in the regulation of transcription. Detailed list of other identified proteins is available in Supplementary Table S1.

We focused our attention on identified TF trapped by this promoter sequence (Table 2). For half of them, an interaction with the HIV-1 sequence has already been described in the literature (Figure 3A). This is the case for previously discussed NFκB family members (c-Rel, p105/p50, p100/p52 and RelA) and also for three members of the upstream stimulatory factor (USF) family (36), the activating enhancer-binding protein-4 (AP-4) (37), the constitutively activated specific protein-1 (Sp1) and -3 (Sp3) (38), the nuclease-sensitive element-binding protein 1 (YB-1) (39) and the chicken ovalbumin upstream promoter TF (COUP-TF), described to interact indirectly with this sequence through an interaction with Sp1 (40). Of note, the HIP116/helicase-like TF has also been reported to interact with HIV-1 5’LTR sequence (41) and was identified by MS in three of the six samples analysed, although the identification did not fit to the MS validation criteria adopted in this study. These results support the efficiency of this strategy.

Moreover, an *in silico* analysis of the capture probe sequence, using bioinformatics parameters minimising false negatives, reveals the presence of putative binding sites (Supplementary Table S2) for several other identified TF. That is the case for two homeodomain-containing proteins, the Meis and the pre-B-cell leukaemia TF 1 (PBX1) (Figure 3A), the oestrogen-related receptor alpha (ERR1), the MYC-associated factor (MAX) and two of its potential partners (the MAX gene-associated protein, MGA and the MAX dimerization protein 3, MAD3), the class E basic helix-loop-helix protein 40 (DEC1) and the nuclear factor 1 X-type. Furthermore, Krüppel-like factors (KLF-16 and KLF-5) and the circadian locomotor output cycles protein kaput (Clock) could, respectively, interact with GC-boxes and E-boxes present on the capture probe. Only the identification of the endothelial differentiation-related factor 1 (MBF1) could not be explained by the presence of a putative binding site in the sequence analysed. However, this TF is also known to act as a bridging factor between Tata-binding proteins (TBP) or TBP-associated proteins and several TFs belonging to the basic leucine zipper family or to the nuclear receptor family (42–44).

The identification of a relatively large number of proteins—38—known to participate in the regulation of gene expression by indirect interactions with DNA (Supplementary Table 3) suggests that this method of DNA-affinity purification of interacting proteins can not only be used to identify the TFs involved in the expression

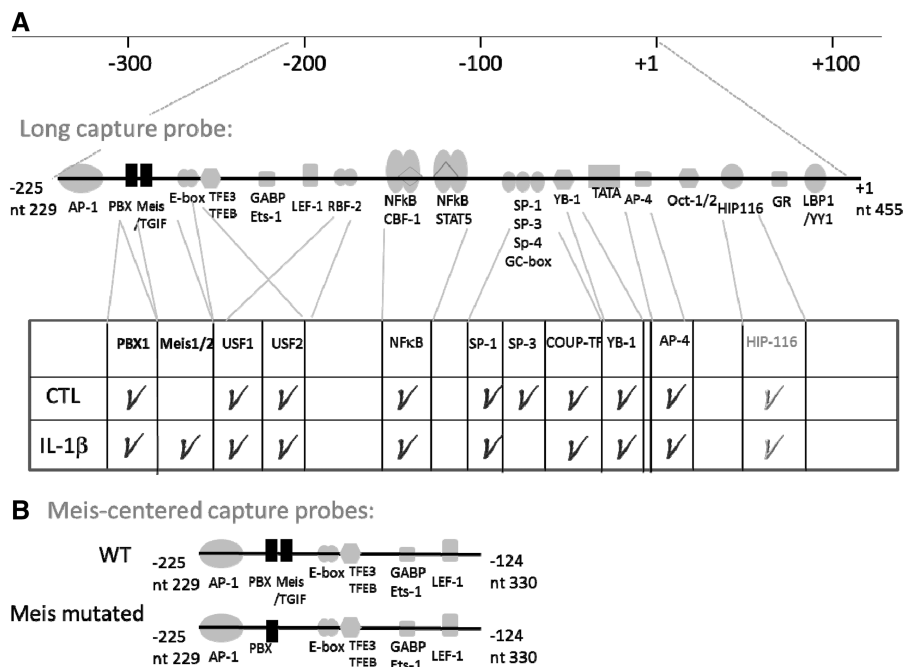


Figure 3. Schematic representation of HIV-1 LTR 5' sequences used as capture probes, TF-binding sites and corresponding interacting proteins. The 226-bp-long capture probe (**A**) and the 101-bp-long Meis-centred capture probe (**B**) are represented. Previously characterized DNA-binding sites for transcription factors present on these fragments of the HIV-1 5'LTR are indicated by grey boxes (modified from (24,95,37)). Black boxes indicate the PBX and the Meis/TGIF putative binding sites predicted by an *in silico* analysis of the sequence of interest (using TRANSFAC database). For the longer capture probe, corresponding transcription factors identified in this study and in the two experimental conditions (nuclear extract from unstimulated or IL-1β-stimulated HeLa cells) are indicated in the Table 2. HIP116: non-validated identification according MS criteria. COUP-TF interacts indirectly with this DNA-binding site as a Sp1transcriptional co-activator.

of a gene of interest but also to unravel their co-regulators. Indeed, several of the identified co-regulators listed in Supplementary Table 3 have previously been described to interact with NFκB family members such as poly(ADP-ribose) polymerase-1 (45), CAPER (RNA binding motif protein 39) (46), DNA-PK (47), the high mobility group protein B1 (HMGB1) involved in the repression of HIV-1 transcription (48), the NFκB reducing enzyme APEX/Ref-1 (the reduction of NFκB increases its binding capacity, notably on the HIV-1 5'LTR) (49,50), or some members of the DEAD-box RNA helicases, enzymes involved in RNA metabolism but recently identified as RelA co-activators (51,52). However, while RelA has only been identified in the IL1β-stimulated condition, as expected, these potential co-activators for RelA were identified in both stimulated and unstimulated conditions. This might be explained by the fact that most co-regulators, if not all, are able to interact with multiple TF, some of them being listed in Table 2. This is the case for Ref-1, which can also interact with YB-1 (53) and for DNA-PK, able to transactivate USF proteins. What is known and already reported about the role/function of these 38 proteins in the HIV-1 transcriptional regulation and/or their interactions with other identified proteins (co-regulators or other proteins in regulatory complexes) has been non-exhaustively synthesized (Supplementary Table S4). Interestingly, several members of the Sin3A co-repressor complex have been identified: the paired amphipathic helix protein Sin3A,

the histone-binding protein RBBP7 and the histone deacetylase 1 (HDAC1) involved in the epigenetic control of HIV-1 latency (54). This suggests that the method used might allow the formation of protein complexes on the capture probe, as well as their identification.

As shown by the identification of a remarkable number of transcription-related proteins, this newly developed method can be used as a starting point in the study of an unknown promoter sequence in order to identify new putative transcriptional regulators. However, these candidates must be validated to confirm the direct or indirect DNA interaction and to address their biological effect on promoter activity.

Functional analysis of a Meis putative binding site in the HIV-1 5'LTR-dependent transcriptional activity

The identification of two members of the homeodomain-containing protein family, PBX1 and Meis, as HIV-1 5'LTR interacting proteins, has attracted our attention for several reasons. First, both TFs could heterodimerize together (55). Second, an *in silico* analysis of the capture probe sequence, performed with the TRANSFAC database (5), revealed the presence of two Meis putative binding sites (with a core match of 1.0). Both sites are localized close to one of the several degenerated PBX putative binding sites identified on this sequence, but only one (in position nt 266-271 of the entire HIV-1 5'LTR sequence) displays the perfect consensus core 5'-T

GACAG-3' (56,57). Finally, as shown in Table 2, if PBX1 was identified in both stimulated and unstimulated conditions, the identification of Meis seemed to be limited to the IL-1 β treatment. Even if based on only few sequenced peptides, this observation suggested a differential Meis recruitment after IL-1 β stimulation.

To functionally characterize the TGACAG Meis putative binding site on the viral promoter activity, trans-activation assays were performed in HeLa cells, stimulated or not with IL-1 β , to compare the luciferase activities of a Luc-reporter gene driven by either the wild-type HIV-1 5'LTR sequence or a variant containing a critical point mutation in the Meis putative binding site (Figure 4A). This point mutation was previously described to significantly decrease the DNA binding of Meis on the Pax6 pancreatic enhancer sequence (58). As shown in Figure 4B, a significant up-regulation of the promoter activity was observed in the presence of the mutated Meis-binding site, in both unstimulated and IL-1 β -stimulated HeLa cells, suggesting a repressive influence of this site even in basal conditions.

To confirm the role of this Meis-binding site in a more physiologically relevant cell type, transcriptional activities of both constructs were measured in T-lymphoid Jurkat cells treated with IL-1 β , TNF- α or PMA/ionomycin. As shown in Figure 4C, although IL- β was without any effect in Jurkat cells on the activation of the pLTRwt, TNF- α and PMA/ionomycin were potent HIV-1 transcription activators in this cell line. As observed in HeLa cells, the presence of the mutation in the Meis putative binding site on the LTR sequence systematically induced an approximated twofold increase in the promoter activity when compared with the wild-type sequence in the corresponding condition.

These results strongly suggest that this sequence, predicted as a putative Meis-binding site by TRANSFAC analysis, could interact with cellular factor(s) acting as transcriptional repressor(s) on the HIV-1 5'LTR, even in basal conditions. Although the results obtained by DNA-affinity capture suggest that these TFs could be members of the homeodomain-containing protein family, the exact nature of this trans-repressor complex needed to be confirmed.

Identification of members of the trans-repressor complex that specifically interact with the Meis putative binding site

To further identify proteins involved in the trans-repressor complex interacting with the Meis putative binding site in basal conditions, our analysis was completed by a more classic comparative DNA-affinity approach carried out on shorter capture probes. In this case, a 101 bp-long capture probe centred on the Meis putative binding site and a mutated version were designed (Figure 3B) and incubated with NE prepared from unstimulated HeLa cells. Specific interactions with the TGACAG site could then be highlighted by a comparative analysis of results obtained with the wild-type versus the mutated capture probe. Table 3 contains the transcription-related proteins identified in both conditions for three independent biological replicates

(in this analysis, proteins identified by only one peptide in only one replicate were not considered). Twenty-seven TF and 35 potential co-regulators, among which, respectively, 17 and 32 candidates, were validated by adopted MS criteria. Confirming the expected homeodomain-containing proteins implication, PBX1 was exclusively captured by the probe containing the wild-type TGACA G site. It was also the case for Meis1, highlighting the capture of this protein even in basal conditions. This experiment was also performed in the presence of NE prepared from IL-1 β -stimulated HeLa cells (data not shown) and, in this case, Meis1 was captured and detected to the same extent than in basal conditions (with an average of 7.6 and 6 sequenced peptides, respectively, for unstimulated and IL-1 β -stimulated conditions) suggesting that the Meis differential recruitment observed during the DNA-affinity capture performed with the longer capture probe was probably due to a weak sensitivity.

Very interestingly, it was also the case for several proteins belonging to the Sin3A-HDAC co-repressor complex, specifically the scaffold protein Sin3A, the histone deacetylase HDAC1, the histone deacetylase complex subunit SAP130 and two histone-binding proteins, RBBP4 and RBBP7 (59,60). The histone deacetylase complex subunit SAP30 was also identified in two replicates, but this identification was not validated according to MS criteria. Moreover, the TF Rox, also known as the Mnt MAX-binding protein, was also specifically captured by the wild-type probe. This TF, known to act as a key transcriptional regulator of the Myc/Max network, represses transcription by competing with c-Myc for Max-binding and through a well-described direct interaction with Sin3A complex (61,62). These results are consistent with the repressive activity associated with this site.

Beyond these interactions specific to the Meis-binding site, several protein identifications (common to both capture probes) complete the results obtained with the long capture probe, confirming expected and also new putative protein interactions with this fragment of the HIV-1 5'LTR. For instance, this second DNA-affinity-based analysis confirms Clock and BMAL1 as TF potentially interacting with the HIV-1 5'LTR. The BMAL1-Clock heterodimer, generally associated with histone acetyltransferase activity, is a transcriptional activator that can be repressed by DEC1 through competition for the E-box-binding site (63). DEC1 is also known to repress retinoid X receptor (RXR)-mediated transactivation through direct interaction (64). Although these TFs are mainly known to be involved in the circadian rhythm control, BMAL1-Clock heterodimers have recently been associated with the transcriptional regulation of the Herpes Simplex Virus 1 (65,66).

Meis contributes to repress HIV-1 5'LTR-dependent transcription

As our results suggest that Meis1 takes place in a trans-repressor complex interacting with a specific binding site localized in nt 266–271 of the HIV-1 5'LTR

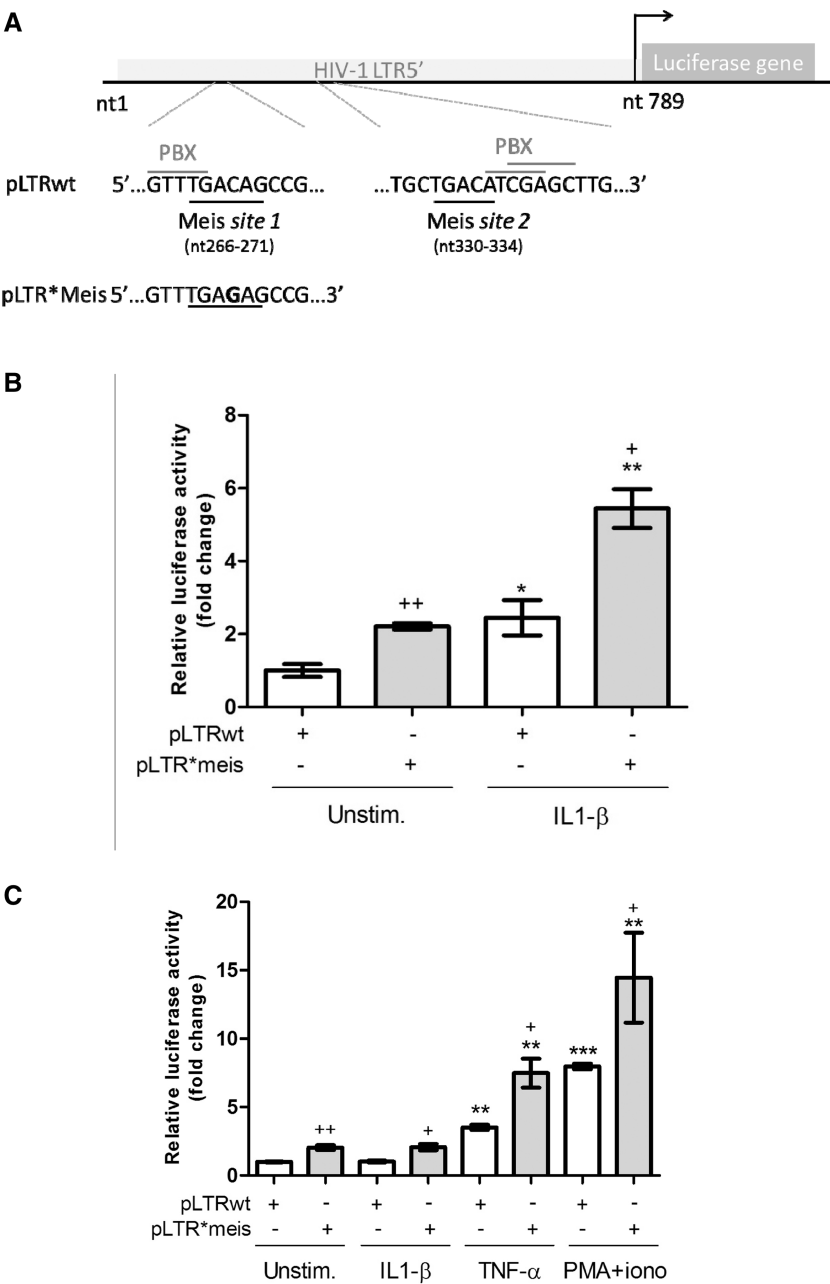


Figure 4. Effect of a mutation in the Meis putative binding site on the HIV-1 LTR5'-dependent transcriptional activity. (A) Schematic representation of the single-nucleotide mutation introduced in the Meis-putative binding site 1 of the HIV-1 5'LTR containing reporter construct. The two TRANSFAC-proposed Meis putative binding sites are underlined in black and the closest PBX putative binding sites are underlined in grey. The mutation is highlighted in boldface. (B) HeLa cells were transiently co-transfected with wild-type or with a Meis mutant version of a luciferase-linked HIV-1 5'LTR promoter construct together with a pCMV-βgal plasmid. At 24-h post-transfection, cells were stimulated or not with IL-1β (5 ng/ml). Luciferase and β-galactosidase activities were measured at 48-h post-transfection. The fold change values represent the ratio Luc/β-gal compared with the ratio obtained for unstimulated cells transfected with the wild-type pLTR construct. This assay was performed in three biologically independent experiments and results were presented as means ± S.D. *, **: significantly different from corresponding unstimulated cells as determined by a *t*-test (unpaired test) with, respectively, *P* < 0.05 and *P* < 0.01. +, ++: significantly different from corresponding wild-type reporter construct transfected cells as determined by a *t*-test (unpaired test) with, respectively, *P* < 0.05 and *P* < 0.01. (C) Jurkat cells were transiently co-transfected with wild-type or with a Meis mutant version of a luciferase-linked HIV-1 5'LTR promoter construct together with a pRL-TK plasmid. At 24-h post-transfection, cells were stimulated or not with IL-1β (5 ng/ml), TNF-α (10 ng/ml) or PMA (25 ng/ml) + ionomycin (500 nM). Luciferase activities were measured at 48 h post-transfection. The fold change values represent the ratio Luc_{Firefly}/Luc_{Renilla} compared with the ratio obtained for unstimulated cells transfected with the wild-type pLTR construct. This assay was performed in three biologically independent experiments and results were presented as means ± S.D. *, **, ***: significantly different as determined by a *t*-test (unpaired test) with, respectively, *P* < 0.05, *P* < 0.01 and *P* < *, **, ***: significantly different from corresponding unstimulated cells as determined by a *t*-test (unpaired test) with, respectively, *P* < 0.05 and *P* < 0.01. +, ++: significantly different from corresponding wild-type reporter construct transfected cells as determined by a *t*-test (unpaired test) with, respectively, *P* < 0.05 and *P* < 0.01.

Table 3. List of transcription-related proteins identified by LC-MS/MS after differential DNA-affinity capture on Meis-centred capture probe

Uniprot accession no.	Protein name	WT probe						Meis mutated probe					
		Exp1			Exp2			Exp3			Exp1		
		Mascot score	Peptide number	Peptide number	Mascot score	Peptide number	Peptide number	Mascot score	Peptide number	Peptide number	Mascot score	Peptide number	Peptide number
O00470	Meis1, myeloid ecotropic viral integration site 1	106.94	9	128.38	8	123.11	6	0	0	0	0	0	0
P40424	Pre-B-cell leukemia transcription factor 1	123.88	8	119.86	9	71.34	5	0	0	0	0	0	0
Q15583	Homeobox protein TGIF1	40.5	1	0	0	15.8	1	0	0	0	0	0	0
Q99583	Max-binding protein MNT(Rox)	81.02	4	29.6	1	119.55	9	0	0	0	0	0	0
P50539	Max interacting protein 1 (MXI1)	16.1	1	0	0	44.1	1	0	0	0	0	0	0
P22415	Upstream transcription factor 1	223.72	20	144.48	13	250.09	18	215.03	15	146.07	12	209.55	13
P15853	Upstream stimulatory factor 2	289.13	42	160.23	16	267.77	22	257.98	24	209.01	24	226.05	19
P78347	General transcription factor II-1	281.3	13	0	0	212.16	17	281	11	68.37	4	193.25	10
P19484	Transcription factor EB	0	0	0	0	0	0	56.73	4	0	0	0	0
P19532	Transcription factor E3	96.36	10	0	0	41.5	1	0	0	36	1	64.45	3
O75030	Microphthalmia-associated transcription factor (MITF)	0	0	52.3	2	46	2	0	0	0	0	0	0
Q15723	ETS-related transcription factor Elf-2	21.1	1	0	0	33.4	1	0	0	0	0	0	0
Q06546	GA-binding protein alpha chain	28.9	1	0	0	0	0	0	0	0	0	0	0
O15516	Circadian locomotor output cycles protein kaput (Clock)	75.84	5	37.9	1	89.35	4	56.56	3	41.6	1	55	1
Q8WYA1	Aryl hydrocarbon receptor nuclear translocator-like protein 2 (BMAL1)	50.4	2	0	0	84.67	5	74.17	3	103.69	6	0	0
O14503	DEC1	0	0	0	0	84.06	3	46.7	2	19.4	1	69.03	3
P19793	Retinoic acid receptor RXR-alpha	0	0	172.69	9	0	0	0	0	145.04	10	0	0
P11474	Steroid hormone receptor ERR1	17.2	1	0	0	0	0	46	2	0	0	42.3	1
P24468	COUP transcription factor 2	0	0	234.35	18	87.4	4	50.72	3	258	22	135.84	7
P61244	Protein max	207.68	11	88.02	6	188.69	21	89.57	4	75.65	3	0	0
P08651	Nuclear factor 1 C-type	0	0	60.5	6	0	0	0	0	63.07	5	0	0
Q14938	Nuclear factor 1 X-type	0	0	66.47	4	0	0	0	0	42.9	1	0	0
O00712	Nuclear factor 1 B-type	0	0	31.2	1	0	0	0	0	0	0	0	0
P16220	Cyclic AMP-responsive element-binding protein 1	89.19	6	40.2	1	59.75	4	50.1	4	19.3	1	45.1	1
O60869	Endothelial differentiation-related factor 1	0	0	22	1	0	0	0	0	19.9	1	20.6	1
P17676	CCAAT/enhancer-binding protein beta (C/EBP beta) (NF-IL6)	0	0	18	1	0	0	0	0	41.8	1	0	0
Q9UMR3	T-box transcription factor TBX20	0	0	0	0	24.7	1	38.2	1	0	0	19.2	1
P78527	DNA dependent protein kinase catalytic subunit	308.14	20	1493.81	102	356.61	20	498.86	28	1241.18	82	453.6	24
P12936	ATP-dependent DNA helicase II 70 kDa subunit(Ku 7)	579.65	51	495.84	35	488.59	40	499.5	52	340.1	21	495.14	54
P13010	ATP-dependent DNA helicase II 86 kDa subunit 2 (Ku 80)	287.15	26	256.85	13	316.05	29	333.65	28	111.23	8	442.28	39
P46063	ATP-dependent DNA helicase Q1	57	2	31.3	1	31.4	1	31.4	1	0	0	48.2	1
P04406	Glyceroldehyde-3-phosphate dehydrogenase	92.34	9	154.44	13	115.64	6	99.23	7	151.29	11	118.63	7
P19338	Nucleolin	297.57	20	397.66	23	345.84	21	405.96	27	332.38	26	422.26	24
P09874	Poly (ADP-ribose) polymerase 1	1201.74	128	822.7	61	1080.32	96	1170.12	101	654.32	50	1195.25	126
P27695	DNA(apurinic or apyrimidinesite) lyase (APEX nuclease)/Redox factor-1	218.78	25	163.56	11	219.67	18	291.35	24	168.01	12	338.68	25
PSO723	Brain acid soluble protein 1	130.46	13	170.8&11	11	219.3	19	243.91	27	160.07	14	191.19	22
P09429	High-mobility group box 1	25.3	1	0	0	35.1	1	29.3	1	0	0	33.2	1
P16401	Histone H1.5	65.58	3	16.9	1	60.36	7	96.4	5	0	0	145.33	11
Q86V81	THO complex subunit 4-A/Transcriptional coactivator Aly/REF	52.2	2	86.29	5	28.9	1	60.92	2	38.9	1	68.95	5
P82979	SAP domain-containing ribonucleoprotein (CIP29 protein)	41.5	1	20.7	1	42.5	1	69.59	3	17.2	1	35.5	1
P17844	ATP-dependent RNA helicase DDX5	31	1	60.54	4	44.9	1	67.6	3	16.9	1	84.05	5
Q5TC19	Lamin A/C	0	0	0	0	50.1	1	69.3	3	24.4	1	80.54	4
Q96AE4	Far upstream element (FUSE) binding protein 1	159.94	8	31.6	1	84.19	6	85.1	4	16.9	1	123.67	4
P11387	DNA topoisomerase 1	103.11	6	0	0	181.77	13	147.11	9	0	0	203.99	15
Q07666	Sam68deltaKH	38.4	1	26.2	1	59.1	1	67.87	4	36.8	1	47.8	2
P53999	Activated RNA polymerase II transcriptional coactivator p15	126.83	7	53.8	1	145	9	132.78	7	75.04	6	133.69	8
O75475	PC4 and SFRS1-interacting protein	74.75	5	0	0	116.08	10	121.78	8	0	0	39.9	1
Q7KZF4	Staphylococcal nuclease domain-containing protein 1 (100 kDa coactivator)	130.86	10	0	0	73.17	5	58.35	4	0	0	65.41	5
Q96PK6	RNA binding motif protein 14	0	0	304.31	17	0	0	0	0	352.35	21	0	0

(continued)

Table 3. Continued

Uniprot accession no.	Protein name	WT probe						Meis mutated probe					
		Exp1		Exp2		Exp3		Exp1		Exp2		Exp3	
		Mascot score	Peptide number	Mascot score	Peptide number	Mascot score	Peptide number	Mascot score	Peptide number	Mascot score	Peptide number	Mascot score	Peptide number
P23246	Splicing factor, proline- and glutamine-rich (p52/p100 complex)	41.9	1	58.8	3	59.39	3	19.3	1	62.27	3	92.51	5
P14866	Heterogeneous nuclear ribonucleoprotein L	407.79	42	175.25	13	318.87	30	267.81	18	154.25	10	323.6	22
Q15365	Heterogeneous nuclear ribonucleoprotein E1//Poly(rC)-binding protein 1	93.02	5	46.9	1	121.39	4	80.07	4	65.25	4	0	0
Q15366	Heterogeneous nuclear ribonucleoprotein E2//Poly(rC)-binding protein2	0	0	0	0	91.3	4	55.17	2	0	0	52.3	2
P61978	Heterogeneous nuclear ribonucleoprotein K	145.31	11	262.56	12	189.31	11	218.94	13	206.09	14	192.2	9
O43390	Heterogeneous nuclear ribonucleoprotein R	45.4	1	32.6	1	0	0	0	0	0	0	0	0
Q00839	Heterogeneous nuclear ribonucleoprotein U	89.09	5	93.91	7	80.87	6	138.19	8	66.02	3	47.9	1
Q965T3	Paired amphipathic helix protein Sin3a	347.84	24	0	0	522.1	40	0	0	0	0	0	0
Q9H0E3	HiStone deacetylase complex subunit SAP 130	54.5	2	0	0	150.28	8	0	0	0	0	0	0
Q13547	Histone deacetylase 1	57	2	0	0	65.72	6	0	0	0	0	0	0
Q16576	Histone-binding protein RBBP7	151.96	9	0	0	117.33	9	0	0	0	0	0	0
Q09028	HiStone-binding protein RBBP4	0	0	0	0	91.87	7	0	0	0	0	0	0
O75446	mSin3A associated polypeptide p30	53.8	1	0	0	49	2	0	0	0	0	0	0

Nuclear extracts from unstimulated HeLa cells were incubated in the presence of a capture probe corresponding to the 101 bp-long sequence around the Meis putative binding site contained in the HIV-1 5'LTR (nt 229-330). Two different versions of the capture probe (wild-type or Meis-mutated) were used. Three independent DNA-affinity captures were performed for each experimental condition (Exp 1 to 3). Identified proteins from negative control analysis were subtracted from the list. Blank boxes indicate validated protein identification (false positive rate <1% otherwise a Mascot score >60). Grey boxes indicate non-validated proteins identification according to MS validation criteria. Transcription factor identifications are represented in bold.

sequence, the effect of a selective knockdown of Meis in HeLa cells was evaluated on the HIV-1 5'LTR-driven reporter gene activity. Cells were first transfected with a pool of siRNA targeting Meis1, leading to an efficient depletion of both Meis1 and Meis2 isoforms (Figure 5A). As shown in Figure 5B, in both basal and IL-1β-stimulated HeLa cells, the Meis knock-down induces a significant up-regulation of the promoter activity, to the same extent than with a mutated Meis1-binding site vector. Even if these results cannot distinguish the specific impact of different Meis1 or Meis2 isoforms, they confirm the central role of Meis in HIV-1 transcriptional regulation through direct binding on a specific binding site on 5'LTR and recruitment of different members of the Sin3A repressor complex.

DISCUSSION

Deciphering the transcriptional regulation of a promoter of interest is a tough task considering the long lists of putative TF-binding sites generated by *in silico* analyses and the absolute necessity to validate each ensued interesting candidate. Although an interesting starting point, *in silico* analyses are nevertheless restrictive as they are limited to already characterized TF-binding sites and cannot specify the TF family members involved. In addition, they do not take into account the accessibility, or not, of the suggested TF and never give access to the layer of transcriptional regulators indirectly interacting with DNA through the first layer of TF (67). These drawbacks of *in silico* analysis have highlighted the crucial need for developing methods to identify proteins that physically interact with DNA—no matter if this interaction is direct or not. Just like ChIP-Seq provides the identity of all the DNA sequences bound by an immunoprecipitated TF, methods combining DNA-affinity protein capture with MS-based protein identification should ideally provide the identity of all the proteins bound to a DNA sequence of interest. However, a major hindrance comes from the fact that DNA can interact with an intricate protein network presenting a high dynamic range between some very low abundant transcriptional regulators and high abundant unspecific DNA-binding proteins. Therefore, most current techniques are focused on a precise binding site present on a short oligonucleotide, for which unspecific background is subtracted thanks to a comparative analysis of the proteins captured by the mutated DNA sequence (20,21).

On the contrary, the goal of the method described in this study is to rapidly provide the most complete picture of the endogenous proteins interacting with a relatively long (200–300 bp) DNA sequence, starting from reasonable amounts of material (<2 mg of nuclear proteins). These criteria make this procedure different from other recently published DNA-affinity methods. Indeed, most of them require large amounts of starting material (from 40 mg (13,20) to 700 mg of nuclear proteins (22)), narrowing the range of biological samples that can be studied by these techniques. In addition, we have preferred a one-step 1D-LC separation of the peptide mixture instead of

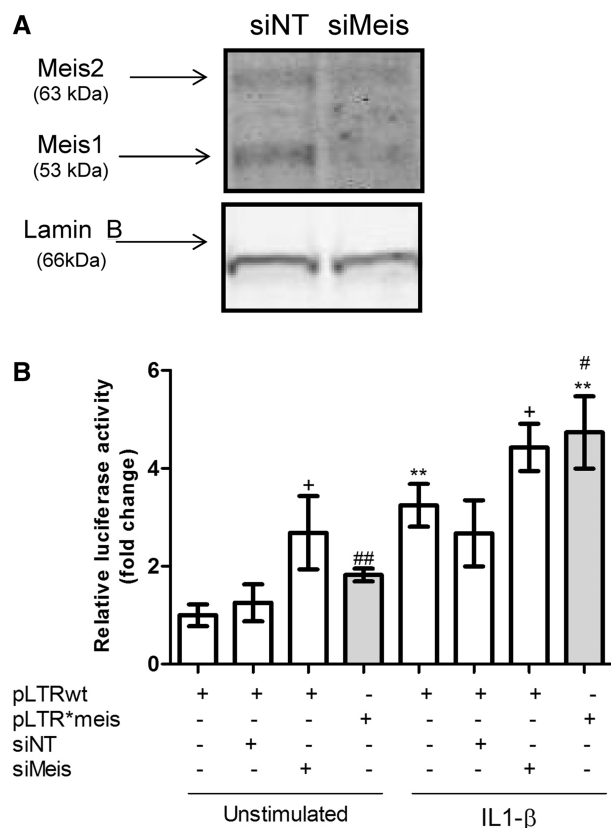


Figure 5. Impact of Meis-1 knockdown on the HIV-1 5'LTR transcriptional activity. HeLa cells were transfected with 50 nM of Meis-1 targeting siRNA (siMeis) or with 50 nM of siRNA non-targeting (siNT). (A) Western blot analysis performed on HeLa proteins extracted 24 h after siRNA transfection. Lamin B signal was used as a loading control. (B) At 24-h post-siRNA transfection, cells were co-transfected with wild-type or a Meis mutant version of a luciferase-linked HIV-1 5'LTR promoter construct together with a pCMV-βgal plasmid. At 24 h post-DNA transfection, HeLa were stimulated or not with IL-1β (5 ng/ml). Luciferase and β-galactosidase activities were measured at 48 h post-DNA transfection. The fold change values represent the ratio Luc/β-gal compared with the ratio obtained for unstimulated cells transfected with the wild-type pLTR construct in the absence of siRNA. This assay was performed in three biologically independent experiments and results were presented as means ± S.D. **: significantly different from corresponding unstimulated cells as determined by a *t*-test (unpaired test) with, respectively, $P < 0.01$. #, ##: significantly different from corresponding wild-type reporter construct transfected cells as determined by a *t*-test (unpaired test) with, respectively, $P < 0.05$ and $P < 0.01$. +: significantly different from corresponding siNT-transfected cells as determined by a *t*-test (unpaired test) with $P < 0.05$.

“multidimensional protein identification technology”-based strategies (68) requiring laborious and time-consuming analyses of all generated fractions (19) and increasing the risk to lose low abundant proteins. Finally, this method is completely unbiased as it does not require pre-existing knowledge of the sequence, like the mapping of regulatory binding sites. Indeed, most DNA-affinity-based methods currently used are centred on one binding site, taking advantage of the comparative analysis of the proteins captured by the mutated counterpart of the bait, either in non-quantitative (18,69) or in quantitative analyses based on metabolic (SILAC) (23) or

chemical labelling (ICAT) (70,22). Although quantitative data are impressive by their specificity and undoubtedly useful to depict the fine tuning of transcriptional complex regulation, such questions are centred on pre-defined regulatory binding sites, pursuing a different purpose than a large-scale unbiased analysis to identify new putative interactants of a relatively long sequence.

To identify a maximum number of proteins physically interacting with a long DNA sequence, without having the benefit of a comparison with an inactive bait, the main difficulty was inherent to the abundant unspecific protein binding and to the high dynamic range of the analysed peptide mixture. These obstacles were circumvented using an efficient separation step of the DNA–protein complexes from the solid support, avoiding irrelevant support interacting protein identification, and an adapted gel-free LC-MS/MS analysis process that improves the sensitivity of the complex peptide mixture analysis using peptide exclusion lists.

Applying this DNA-affinity method to a well-described 226-bp-long sequence of the HIV-1 promoter region, we identified over a hundred different captured proteins among which >50% have a demonstrated link with transcription. Among the other half of identified proteins, one can notice the presence of many nucleic-acid-binding proteins, including RNA-binding proteins or proteins involved in DNA repair, which might represent contaminant proteins or potential uncharacterized transcriptional regulators (71). Focusing on transcription-related identified proteins, the unbiased identification of a large number of TFs described to interact with the HIV-1 5'LTR sequence gives weight to this approach (Figure 3), although not all the expected TFs were identified, like the lymphoid enhancer-binding factor 1 (LEF-1) (72) or the ubiquitous Ying Yang 1 factor and the late SV40 factor (32). The reasons for this might be either biological or technical. From a biological point of view, beside cell type specificities (for instance, LEF-1 has not been described to bind the HIV-1 5'LTR in HeLa cells), we have to point out that the capture probe used is a PCR-produced double stranded oligonucleotide, different from an *in vivo* chromatin environment. An elegant PICh method (for proteomics of isolated chromatin segments), based on the isolation of genomic DNA with its associated proteins, has recently been developed to characterize the *in vivo* telomere interactome (73). However, the limited sensitivity of protein identification by MS represents the major restriction of this technique, which is currently limited to study repetitive DNA sequences and requires huge amounts of starting material. In addition, we cannot exclude some technical limitations of the method we developed dealing with an inefficient protein capture possibly due to an incomplete nuclear protein extraction or an unadapted binding or washing conditions (74), a non-specific binding of some peptides on beads during the biotin removal step or a lack of MS sensitivity for detecting very low abundant peptides in spite of the use of exclusion peptide lists. Of note, some proteins such as KLF-5 were identified on the basis of relatively small number of sequenced peptides, and not in each replicate experiment. As this method involves

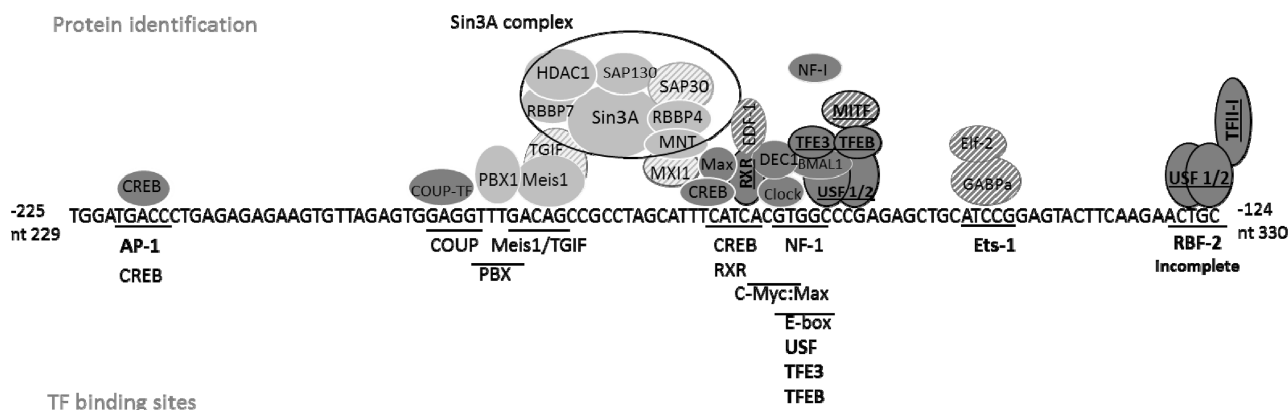


Figure 6. Proposed model for the organization of the proteins identified in this DNA-affinity study interacting—directly or not—with a fragment of the HIV-1 5'LTR (from nt 229 to 330). The binding sites indicated in this Figure are either experimentally demonstrated in the literature (bold) or predicted by *in silico* analysis of this sequence. All the indicated proteins have been captured by this DNA sequence and identified in this study. Proteins previously described to interact with this sequence are underlined. The proteins represented by dashed ovals are indicative as the MS validation criteria for identification were not fulfilled. Proteins in light grey are specifically recruited by the Meis-binding site, as they were not captured by this DNA sequence containing the mutated Meis-binding site.

sequential steps of cell culture, nuclear extract preparation, DNA-affinity purification, trypsin digest and reverse-phase chromatography upstream the MS sequencing, it is essential to repeat the analysis out of several independent biological replicates. Moreover, the candidates that will be selected by the researcher for further functional studies should have been identified on the basis of several peptides sequencing.

Beyond the identification of several TFs known to regulate HIV-1, the most interesting data are probably the identification of 12 TFs that are potential new candidates for the regulation of HIV-1 transcription. This was unexpected, considering the large number of studies devoted to this regulatory DNA sequence of the highest pathophysiological interest. An *in silico* analysis of the sequence indicates the presence of at least one putative binding site for 11 of them. Such results might constitute an interesting starting point for future studies in the field of HIV-1 regulation, like in the case of ERR1, for which this putative novel interaction might explain the reduced HIV-1 promoter activity due to estradiol treatment in glial cells (75).

We focused our attention on two other interesting candidates, the homeodomain-containing proteins, PBX1 and Meis. Mainly described as homeobox protein (HOX) co-regulators, either co-activators or co-repressors (76), these TFs can also, in a HOX-independent manner, act as homo- (77) or hetero-dimers together (78,79) or with other homeodomain- and non-homeodomain-containing TFs (80–83). Interestingly, some members of the homeodomain-containing family among which PBX1 were already documented as transcriptional regulators of different viruses such as the herpes simplex virus (83), the human papillomavirus (84,85), the murine Leukemia virus (86) and the human Cytomegalovirus (87). After an *in silico* analysis of the sequence (supplementary data), we studied the functionality of the 5'-TGACAG-3' sequence (nt 266–271), corresponding to a perfect

consensus core for Meis, located nearby a degenerated PBX-binding site (Figure 4A), that could correspond to a PBX-Meis heterodimer-binding criteria (88). A point mutation in this binding site increases the HIV-1 5'LTR promoter activity, as shown by luciferase reporter gene assay in both HeLa and Jurkat cells, even in basal conditions (Figure 4B and C). The silencing of Meis by RNA interference also provokes an increase in the HIV-1 5'LTR promoter activity (Figure 5B), suggesting that this binding site recruits a Meis-containing transcriptional repressor complex, possibly involved in the preservation of the viral latency.

In an attempt to define more precisely the composition of this transcriptional repressor complex, we applied the DNA-affinity technique to a shorter fragment of the HIV-1 5'LTR promoter centred on the newly validated Meis-binding site, either wild-type or mutated. The results we obtained confirm the specific capture of Meis-1 by the wild-type probe. In addition, seven proteins that could take place in the mSin3A repressive complex were also identified as specific interactants of this Meis putative-binding site. While, to our knowledge, it is the first time that a Meis-mSin3A interaction is documented, other members of homeodomain-containing TF family like the transcriptional repressor transforming growth-interacting factor (TGIF) can recruit mSin3 and HDAC (89). Collectively, these results led us to propose in Figure 6 a partial transcriptional interactome surrounding this precise region.

Although the method developed here was initially devoted to conduct a totally unbiased analysis of proteins interacting with a relatively long DNA sequence, it was also successfully applied to a shorter DNA sequence centred on a binding site of interest. This functional “zoom” on the promoter interactome, through the comparative analysis of proteins captured by the wild-type versus a mutant sequence, makes the investigation of transcriptional co-regulators more potent, as

illustrated by the Meis-dependent recruitment of the Sin3A repressing complex—identified by seven different components. This underlines that this DNA-affinity method can be used to display not only direct DNA-interacting proteins such as TF but also proteins that indirectly interact with DNA. In addition, we have to note that the analyses of a 226-bp sequence or of a 101-bp capture probe (centred on the Meis-binding site) generated lists of, respectively, 52/58 (CTL/IL-1 β conditions) and 62 putative interacting transcriptional regulators, pinpointing that this method is still limited to a maximum number of proteins that can be identified, the LC/MSMS analysis being probably the limiting step of the whole process.

In summary, we have described a performant DNA-affinity procedure followed by a gel-free proteomic analysis of the promoter interactome. As a proof of concept, this procedure was applied to a very well-described regulatory sequence, the HIV-1 5'LTR, providing a list of >50 recognized transcriptional regulators (24 TFs and 38 transcriptional regulators not reported to bind DNA), including expected and unexpected TFs which represent new candidates for the transcriptional regulation of HIV-1. This highlights the strength of a completely unbiased analysis of regulatory DNA sequences. When deciphering the mechanism underlying the transcriptional regulation of a gene of interest, we believe that having a list of putative candidates that physically interact with the DNA sequence, although this list is probably not exhaustive, as for any proteomic analysis, constitutes a great advantage over *in silico* analyses. Indeed, the candidates that will further be selected for functional validations will have much more chance to be confirmed.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1, Supplementary Figures 1–4 and Supplementary References [32,46–50,53,96–132].

FUNDING

Funding for open access charge: University of Namur (FUNDP), Belgium.

Conflict of interest statement. None declared.

REFERENCES

- James, T.K. (2004) Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell*, **116**, 247–257.
- Dori-Bachash, M., Shema, E. and Tirosh, I. (2011) Coupled evolution of transcription and mRNA degradation. *PLoS Biol.*, **9**, e1001106.
- Wong, E. and Wei, C.-L. (2009) ChIP'ing the mammalian genome: technical advances and insights into functional elements. *Genome Med.*, **1**, 89.
- Kim, T.-M. and Park, P.J. (2011) Advances in analysis of transcriptional regulatory networks. *Wiley Interdiscip. Rev. Syst. Biol. Med.*, **3**, 21–35.
- Matys, V. (2006) TRANSFAC(R) and its module TRANSCOMP(R): transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.*, **34**, D108–D110.
- Sandelin, A. (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.*, **32**, 91D–94D.
- Westholm, J.O., Xu, F., Ronne, H. and Komorowski, J. (2008) Genome-scale study of the importance of binding site context for transcription factor binding and gene regulation. *BMC Bioinformatics*, **9**, 484.
- Bulyk, M.L. (2003) Computational prediction of transcription-factor binding site locations. *Genome Biol.*, **5**, 201.
- Nordhoff, E., Krogsdam, A.M., Jorgensen, H.F., Kallipolitis, B.H., Clark, B.F., Roepstorff, P. and Kristiansen, K. (1999) Rapid identification of DNA-binding proteins by mass spectrometry. *Nat. Biotechnol.*, **17**, 884–888.
- Drewett, V., Molina, H., Millar, A., Muller, S., von Hesler, F. and Shaw, P.E. (2001) DNA-bound transcription factor complexes analysed by mass-spectrometry: binding of novel proteins to the human c-fos SRE and related sequences. *Nucleic Acids Res.*, **29**, 479–487.
- Samuel, S., Twizere, J.-C. and Bernstein, L.R. (2005) YB-1 represses API-dependent gene transactivation and interacts with an AP-1 DNA sequence. *Biochem. J.*, **388**, 921–928.
- Yaneva, M. and Tempst, P. (2003) Affinity capture of specific DNA-binding proteins for mass spectrometric identification. *Anal. Chem.*, **75**, 6437–6448.
- Dobretsova, A., Johnson, J.W., Jones, R.C., Edmondson, R.D. and Wight, P.A. (2008) Proteomic analysis of nuclear factors binding to an intronic enhancer in the myelin proteolipid protein gene. *J. Neurochem.*, **105**, 1979–1995.
- Jiang, D., Zhou, Y., Moxley, R.A. and Jarrett, H.W. (2008) Purification and identification of positive regulators binding to a novel element in the c-Jun promoter. *Biochemistry*, **47**, 9318–9334.
- Jiang, D., Moxley, R.A. and Jarrett, H.W. (2006) Promoter trapping of c-jun promoter-binding transcription factors. *J. Chromatogr. A*, **1133**, 83–94.
- Park, S.-S., Ko, B.J. and Kim, B.-G. (2005) Mass spectrometric screening of transcriptional regulators using DNA affinity capture assay. *Anal. Biochem.*, **344**, 152–154.
- Hégarat, N., Cardoso, G.M., Rusconi, F., Francois, J.-C. and Praseuth, D. (2007) Analytical biochemistry of DNA–protein assemblies from crude cell extracts. *Nucleic Acids Res.*, **35**, e92.
- Meng, Z., Camalier, C.E., Lucas, D.A., Veenstra, T.D., Beck, G.R. Jr and Conrads, T.P. (2006) Probing early growth response 1 interacting proteins at the active promoter in osteoblast cells using oligoprecipitation and mass spectrometry. *J. Proteome Res.*, **5**, 1931–1939.
- Himeda, C.L., Ranish, J.A., Angello, J.C., Maire, P., Aebersold, R. and Hauschka, S.D. (2004) Quantitative proteomic identification of six4 as the trex-binding factor in the muscle creatine kinase enhancer. *Mol. Cell. Biol.*, **24**, 2132–2143.
- Himeda, C.L., Ranish, J.A. and Hauschka, S.D. (2008) Quantitative proteomic identification of MAZ as a transcriptional regulator of muscle-specific genes in skeletal and cardiac myocytes. *Mol. Cell Biol.*, **28**, 6521–6535.
- Himeda, C.L., Ranish, J.A., Pearson, R.C.M., Crossley, M. and Hauschka, S.D. (2010) KLF3 regulates muscle-specific gene expression and synergizes with serum response factor on KLF binding sites. *Mol. Cell. Biol.*, **30**, 3430–3443.
- Qi, Y., Ranish, J.A., Zhu, X., Krones, A., Zhang, J., Aebersold, R., Rose, D.W., Rosenfeld, M.G. and Carrière, C. (2008) Atf1 is required for the Pit1 gene early activation. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 2481–2486.
- Mittler, G., Butter, F. and Mann, M. (2009) A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. *Genome Res.*, **19**, 284–293.
- Pereira, L.A., Bentley, K., Peeters, A., Churchill, M.J. and Deacon, N.J. (2000) A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. *Nucleic Acids Res.*, **28**, 663–668.
- Chen, Y., Takeshita, A., Ozaki, K., Kitano, S. and Hanazawa, S. (1996) Transcriptional regulation by transforming growth factor beta of the expression of retinoic acid and retinoid X receptor

- genes in osteoblastic cells is mediated through AP-1. *J. Biol. Chem.*, **271**, 31602–31606.
26. Quivy, V., Adam, E., Collette, Y., Demonte, D., Chariot, A., Vanhulle, C., Berkhout, B., Castellano, R., de Launoit, Y., Burny, A. *et al.* (2002) Synergistic activation of human immunodeficiency virus type 1 promoter activity by NF-kappaB and inhibitors of deacetylases: potential perspectives for the development of therapeutic strategies. *J. Virol.*, **76**, 11091–11103.
 27. Kilareski, E.M., Shah, S., Nonnemacher, M.R. and Wigdahl, B. (2009) Regulation of HIV-1 transcription in cells of the monocyte-macrophage lineage. *Retrovirology*, **6**, 118.
 28. Hirsch, J.D., Eslamizar, L., Filanoski, B.J., Malekzadeh, N., Haugland, R.P., Beechem, J.M. and Haugland, R.P. (2002) Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: uses for protein labeling, detection, and isolation. *Anal. Biochem.*, **308**, 343–357.
 29. Van Cutsem, E., Simonart, G., Degand, H., Faber, A., Morsomme, P. and Boutry, M. (2011) Gel-based and gel-free proteomic analysis of *Nicotiana tabacum* identifies proteins involved in secondary metabolism and in the (a)biotic stress response. *Proteomics*, **11**, 440–454.
 30. Nabel, G. and Baltimore, D. (1987) An inducible transcription factor activates expression of human immunodeficiency virus in T cells. *Nature*, **326**, 711–713.
 31. Hayden, M.S. and Ghosh, S. (2008) Shared principles in NF-kappaB signaling. *Cell*, **132**, 344–362.
 32. Coull, J.J., Romero, F., Sun, J.M., Volker, J.L., Galvin, K.M., Davie, J.R., Shi, Y., Hansen, U. and Margolis, D.M. (2000) The human factors YY1 and LSF repress the human immunodeficiency virus type 1 long terminal repeat via recruitment of histone deacetylase 1. *J. Virol.*, **74**, 6790–6799.
 33. Hayden, M.S. and Ghosh, S. (2004) Signaling to NF-kappaB. *Genes Dev.*, **18**, 2195–2224.
 34. Williams, S.A., Chen, L.-F., Kwon, H., Ruiz-Jarabo, C.M., Verdin, E. and Greene, W.C. (2006) NF-kB p50 promotes HIV latency through HDAC recruitment and repression of transcriptional initiation. *EMBO J.*, **25**, 139–149.
 35. Naumann, M. and Scheidereit, C. (1994) Activation of NF-kappa B in vivo is regulated by multiple phosphorylations. *EMBO J.*, **13**, 4597–4607.
 36. Lu, Y.C., Touzjian, N., Stenzel, M., Dorfman, T., Sodroski, J.G. and Haseltine, W.A. (1990) Identification of cis-acting repressive sequences within the negative regulatory element of human immunodeficiency virus type 1. *J. Virol.*, **64**, 5226–5229.
 37. Imai, K. and Okamoto, T. (2006) Transcriptional repression of human immunodeficiency virus type 1 by AP-4. *J. Biol. Chem.*, **281**, 12495–12505.
 38. Majello, B., De Luca, P., Hagen, G., Suske, G. and Lania, L. (1994) Different members of the Sp1 multigene family exert opposite transcriptional regulation of the long terminal repeat of HIV-1. *Nucleic Acids Res.*, **22**, 4914–4921.
 39. Sawaya, B.E., Khalili, K. and Amini, S. (1998) Transcription of the human immunodeficiency virus type 1 (HIV-1) promoter in central nervous system cells: effect of YB-1 on expression of the HIV-1 long terminal repeat. *J. Gen. Virol.*, **79**(Pt 2), 239–246.
 40. Rohr, O., Aunis, D. and Schaeffer, E. (1997) COUP-TF and Sp1 interact and cooperate in the transcriptional activation of the human immunodeficiency virus type 1 long terminal repeat in human microglial cells. *J. Biol. Chem.*, **272**, 31149–31155.
 41. Sheridan, P.L., Schorpp, M., Voz, M.L. and Jones, K.A. (1995) Cloning of an SNF2/SWI2-related protein that binds specifically to the SPH motifs of the SV40 enhancer and to the HIV-1 promoter. *J. Biol. Chem.*, **270**, 4575–4587.
 42. Millership, J.J., Waghela, P., Cai, X., Cockerham, A. and Zhu, G. (2004) Differential expression and interaction of transcription co-activator MBF1 with TATA-binding protein (TBP) in the apicomplexan *Cryptosporidium parvum*. *Microbiology (Reading, Engl.)*, **150**, 1207–1213.
 43. Brendel, C., Gelman, L. and Auwerx, J. (2002) Multiprotein bridging factor-1 (MBF-1) is a cofactor for nuclear receptors that regulate lipid metabolism. *Mol. Endocrinol.*, **16**, 1367–1377.
 44. Jindra, M., Gaziova, I., Uhlirova, M., Okabe, M., Hiromi, Y. and Hirose, S. (2004) Coactivator MBF1 preserves the redox-dependent AP-1 activity during oxidative stress in *Drosophila*. *EMBO J.*, **23**, 3538–3547.
 45. Hassa, P.O. and Hottiger, M.O. (2002) The functional role of poly(ADP-ribose)polymerase 1 as novel coactivator of NF-kappaB in inflammatory disorders. *Cell. Mol. Life Sci.*, **59**, 1534–1553.
 46. Dutta, J., Fan, G. and Gélinas, C. (2008) CAPERalpha is a novel Rel-TAD-interacting factor that inhibits lymphocyte transformation by the potent Rel/NF-kappaB oncoprotein v-Rel. *J. Virol.*, **82**, 10792–10802.
 47. Ju, J., Naura, A.S., Errami, Y., Zerfaoui, M., Kim, H., Kim, J.G., Abd Elmageed, Z.Y., Abdel-Mageed, A.B., Giardina, C., Beg, A.A. *et al.* (2010) Phosphorylation of p50 NF-kappaB at a single serine residue by DNA-dependent protein kinase is critical for VCAM-1 expression upon TNF treatment. *J. Biol. Chem.*, **285**, 41152–41160.
 48. Naghavi, M.H., Nowak, P., Andersson, J., Sönnnerborg, A., Yang, H., Tracey, K.J. and Vahlne, A. (2003) Intracellular high mobility group B1 protein (HMGB1) represses HIV-1 LTR-directed transcription in a promoter- and cell-specific manner. *Virology*, **314**, 179–189.
 49. Ando, K., Hirao, S., Kabe, Y., Ogura, Y., Sato, I., Yamaguchi, Y., Wada, T. and Handa, H. (2008) A new APE1/Ref-1-dependent pathway leading to reduction of NF-kappaB and AP-1, and activation of their DNA-binding activity. *Nucleic Acids Res.*, **36**, 4327–4336.
 50. Stevens, M., De Clercq, E. and Balzarini, J. (2006) The regulation of HIV-1 transcription: molecular targets for chemotherapeutic intervention. *Med. Res. Rev.*, **26**, 595–625.
 51. Ishaq, M., Ma, L., Wu, X., Mu, Y., Pan, J., Hu, J., Fu, Q. and Guo, D. (2009) The DEAD-box RNA helicase DDX1 interacts with RelA and enhances nuclear factor kappaB-mediated transcription. *J. Cell. Biochem.*, **106**, 296–305.
 52. Tetsuka, T., Uranishi, H., Sanda, T., Asamitsu, K., Yang, J.-P., Wong-Staal, F. and Okamoto, T. (2004) RNA helicase A interacts with nuclear factor kappaB p65 and functions as a transcriptional coactivator. *Eur. J. Biochem.*, **271**, 3741–3751.
 53. Bhakat, K.K., Mantha, A.K. and Mitra, S. (2009) Transcriptional regulatory functions of mammalian AP-endonuclease (APE1/Ref-1), an essential multifunctional protein. *Antioxid. Redox Signal.*, **11**, 621–638.
 54. Williams, S.A., Chen, L.-F., Kwon, H., Ruiz-Jarabo, C.M., Verdin, E. and Greene, W.C. (2006) NF-kB p50 promotes HIV latency through HDAC recruitment and repression of transcriptional initiation. *EMBO J.*, **25**, 139–149.
 55. Bischof, L.J., Kagawa, N., Moskow, J.J., Takahashi, Y., Iwamatsu, A., Buchberg, A.M. and Waterman, M.R. (1998) Members of the meis1 and pbx homeodomain protein families cooperatively bind a cAMP-responsive sequence (CRS1) from bovine CYP17. *J. Biol. Chem.*, **273**, 7941–7948.
 56. Chang, C.P., Jacobs, Y., Nakamura, T., Jenkins, N.A., Copeland, N.G. and Cleary, M.L. (1997) Meis proteins are major in vivo DNA binding partners for wild-type but not chimeric Pbx proteins. *Mol. Cell. Biol.*, **17**, 5679–5687.
 57. Shen, W.F., Montgomery, J.C., Rozenfeld, S., Moskow, J.J., Lawrence, H.J., Buchberg, A.M. and Largman, C. (1997) AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins. *Mol. Cell. Biol.*, **17**, 6448–6458.
 58. Zhang, X., Rowan, S., Yue, Y., Heaney, S., Pan, Y., Brendolan, A., Selleri, L. and Maas, R.L. (2006) Pax6 is regulated by Meis and Pbx homeoproteins during pancreatic development. *Dev. Biol.*, **300**, 748–757.
 59. Fleischer, T.C., Yun, U.J. and Ayer, D.E. (2003) Identification and characterization of three new components of the mSin3A corepressor complex. *Mol. Cell. Biol.*, **23**, 3456–3467.
 60. Grzenda, A., Lomberg, G., Zhang, J.-S. and Urrutia, R. (2009) Sin3: master scaffold and transcriptional corepressor. *Biochim. Biophys. Acta*, **1789**, 443–450.
 61. Wahlström, T. and Henriksson, M. (2007) Mnt takes control as key regulator of the Myc/Max/Mxd network. *Advances in Cancer Research*, **97**, 61–80.
 62. Terragni, J., Nayak, G., Banerjee, S., Medrano, J.-L., Graham, J.R., Brennan, J.F., Sepulveda, S. and Cooper, G.M. (2011) The E-box binding factors Max/Mnt, MITF, and USF1 act coordinately

- with FoxO to regulate expression of proapoptotic and cell cycle control genes by phosphatidylinositol 3-kinase/akt/glycogen synthase kinase 3 signaling. *J. Biol. Chem.*, **286**, 36215–36227.
63. Sato, F., Kawamoto, T., Fujimoto, K., Noshiro, M., Honda, K.K., Honma, S., Honma, K. and Kato, Y. (2004) Functional analysis of the basic helix-loop-helix transcription factor DEC1 in circadian regulation. Interaction with BMAL1. *Eur. J. Biochem.*, **271**, 4409–4419.
 64. Cho, Y., Noshiro, M., Choi, M., Morita, K., Kawamoto, T., Fujimoto, K., Kato, Y. and Makishima, M. (2009) The basic helix-loop-helix proteins differentiated embryo chondrocyte (DEC) 1 and DEC2 function as corepressors of retinoid X receptors. *Mol. Pharmacol.*, **76**, 1360–1369.
 65. Kalamvoki, M. and Roizman, B. (2010) Circadian CLOCK histone acetyl transferase localizes at ND10 nuclear bodies and enables herpes simplex virus gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 17721–17726.
 66. Kalamvoki, M. and Roizman, B. (2011) The histone acetyltransferase CLOCK is an essential component of the herpes simplex virus 1 transcriptome that includes TFIID, ICP4, ICP27, and ICP22. *J. Virol.*, **85**, 9472–9477.
 67. Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X. *et al.* (2009) Diversity and complexity in DNA recognition by transcription factors. *Science*, **324**, 1720–1723.
 68. Yates, J.R., Ruse, C.I. and Nakorchevsky, A. (2009) Proteomics by mass spectrometry: approaches, advances, and applications. *Annu. Rev. Biomed. Eng.*, **11**, 49–79.
 69. Lu, X., Beck, G.R. Jr, Gilbert, L.C., Camalier, C.E., Bateman, N.W., Hood, B.L., Conrads, T.P., Kern, M.J., You, S., Chen, H. *et al.* (2011) Identification of the homeobox protein Prx1 (MHox, Prx-1) as a regulator of osterix expression and mediator of tumor necrosis factor α action in osteoblast differentiation. *J. Bone Miner. Res.*, **26**, 209–219.
 70. Ranish, J.A., Yi, E.C., Leslie, D.M., Purvine, S.O., Goodlett, D.R., Eng, J. and Aebersold, R. (2003) The study of macromolecular complexes by quantitative proteomics. *Nat. Genet.*, **33**, 349–355.
 71. Hu, S., Xie, Z., Onishi, A., Yu, X., Jiang, L., Lin, J., Rho, H., Woodard, C., Wang, H., Jeong, J.-S. *et al.* (2009) Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell*, **139**, 610–622.
 72. Waterman, M.L. and Jones, K.A. (1990) Purification of TCF-1 alpha, a T-cell-specific transcription factor that activates the T-cell receptor C alpha gene enhancer in a context-dependent manner. *New Biol.*, **2**, 621–636.
 73. Déjardin, J. and Kingston, R.E. (2009) Purification of proteins associated with specific genomic loci. *Cell*, **136**, 175–186.
 74. Gadgil, H. and Jarrett, H.W. (2002) Oligonucleotide trapping method for purification of transcription factors. *J. Chromatogr. A*, **966**, 99–110.
 75. Wilson, M.E., Allred, K.F., Bisotti, A.J., Bruce-Keller, A., Chuahan, A. and Nath, A. (2006) Estradiol negatively regulates HIV-LTR promoter activity in glial cells. *AIDS Res. Hum. Retroviruses*, **22**, 350–356.
 76. Moens, C.B. and Selleri, L. (2006) Hox cofactors in vertebrate development. *Dev. Biol.*, **291**, 193–206.
 77. Featherstone, M. (2003) HOX proteins and their co-factors in transcriptional regulation. *Advances in Developmental Biology*, **13**, 1–42.
 78. Mojsin, M. and Stevanovic, M. (2010) PBX1 and MEIS1 up-regulate SOX3 gene expression by direct interaction with a consensus binding site within the basal promoter region. *Biochem. J.*, **425**, 107–116.
 79. Pillay, L.M., Forrester, A.M., Erickson, T., Berman, J.N. and Waskiewicz, A.J. (2010) The Hox cofactors Meis1 and Pbx act upstream of gata1 to regulate primitive hematopoiesis. *Dev. Biol.*, **340**, 306–317.
 80. Mann, R.S. and Affolter, M. (1998) Hox proteins meet more partners. *Curr. Opin. Genet. Dev.*, **8**, 423–429.
 81. Laurent, A., Bihan, R., Omilli, F., Deschamps, S. and Pellerin, I. (2008) PBX proteins: much more than Hox cofactors. *Int. J. Dev. Biol.*, **52**, 9–20.
 82. Zhang, X., Friedman, A., Heaney, S., Purcell, P. and Maas, R.L. (2002) Meis homeoproteins directly regulate Pax6 during vertebrate lens morphogenesis. *Genes Dev.*, **16**, 2097–2107.
 83. Pomerantz, J.L., Kristie, T.M. and Sharp, P.A. (1992) Recognition of the surface of a homeo domain protein. *Genes Dev.*, **6**, 2047–2057.
 84. Ndisang, D., Morris, P.J., Chapman, C., Ho, L., Singer, A. and Latchman, D.S. (1998) The HPV-activating cellular transcription factor Brn-3a is overexpressed in CIN3 cervical lesions. *J. Clin. Invest.*, **101**, 1687–1692.
 85. O'Connor, M.J., Stünkel, W., Koh, C.H., Zimmermann, H. and Bernard, H.U. (2000) The differentiation-specific factor CDP/Cut represses transcription and replication of human papillomaviruses through a conserved silencing element. *J. Virol.*, **74**, 401–410.
 86. Chao, S.-H., Walker, J.R., Chanda, S.K., Gray, N.S. and Caldwell, J.S. (2003) Identification of homeodomain proteins, PBX1 and PREP1, involved in the transcription of murine leukemia virus. *Mol. Cell. Biol.*, **23**, 831–841.
 87. Chao, S.-H., Harada, J.N., Hyndman, F., Gao, X., Nelson, C.G., Chanda, S.K. and Caldwell, J.S. (2004) PDX1, a cellular homeoprotein, binds to and regulates the activity of human cytomegalovirus immediate early promoter. *J. Biol. Chem.*, **279**, 16111–16120.
 88. Jacobs, Y., Schnabel, C.A. and Cleary, M.L. (1999) Trimeric association of Hox and TALE homeodomain proteins mediates Hoxb2 hindbrain enhancer activity. *Mol. Cell. Biol.*, **19**, 5134–5142.
 89. Wotton, D., Knoepfler, P.S., Laherty, C.D., Eisenman, R.N. and Massagué, J. (2001) The Smad transcriptional corepressor TGIF recruits mSin3. *Cell Growth Differ.*, **12**, 457–463.
 90. Hansen, S.K., Guerrini, L. and Blasi, F. (1994) Differential DNA sequence specificity and regulation of HIV-1 enhancer activity by cRel-RelA transcription factor. *J. Biol. Chem.*, **269**, 22230–22237.
 91. Sadowski, I. and Mitchell, D.A. (2005) TFII-I and UF (RBF-2) regulate Ras/MAPK-responsive HIV-1 transcription in T cells. *Eur. J. Cancer*, **41**, 2528–2536.
 92. Dahabieh, M.S., Ooms, M., Malcolm, T., Simon, V. and Sadowski, I. (2011) Identification and functional analysis of a second RBF-2 binding site within the HIV-1 promoter. *Virology*, **418**, 57–66.
 93. Harada, N., Yasunaga, R., Higashimura, Y., Yamaji, R., Fujimoto, K., Moss, J., Inui, H. and Nakano, Y. (2007) Glyceraldehyde-3-phosphate dehydrogenase enhances transcriptional activity of androgen receptor in prostate cancer cells. *J. Biol. Chem.*, **282**, 22651–22661.
 94. Sheng, W.-Y. and Wang, T.-C.V. (2009) Proteomic analysis of the differential protein expression reveals nuclear GAPDH in activated T lymphocytes. *PLoS One*, **4**, e6322.
 95. Van Maele, B., Busschots, K., Vandekerckhove, L., Christ, F. and Debysier, Z. (2006) Cellular co-factors of HIV-1 integration. *Trends Biochem. Sci.*, **31**, 98–105.
 96. Montes de Oca, R., Shoemaker, C.J., Gucek, M., Cole, R.N. and Wilson, K.L. (2009) Barrier-to-autointegration factor proteome reveals chromatin-regulatory partners. *PLoS One*, **4**, e7050.
 97. Gautier, V.W., Gu, L., O'Donoghue, N., Pennington, S., Sheehy, N. and Hall, W.W. (2009) In vitro nuclear interactome of the HIV-1 Tat protein. *Retrovirology*, **6**, 47.
 98. Tsai, S.C., Valkov, N., Yang, W.M., Gump, J., Sullivan, D. and Seto, E. (2000) Histone deacetylase interacts directly with DNA topoisomerase II. *Nat. Genet.*, **26**, 349–353.
 99. Sperling, A.S., Jeong, K.S., Kitada, T. and Grunstein, M. (2011) Topoisomerase II binds nucleosome-free DNA and acts redundantly with topoisomerase I to enhance recruitment of RNA Pol II in budding yeast. *Proc. Natl. Acad. Sci. USA*, **108**, 12693–12698.
 100. Yung, T.M.C., Sato, S. and Satoh, M.S. (2004) Poly(ADP-ribosylation) as a DNA damage-induced post-translational modification regulating poly(ADP-ribose) polymerase-1-topoisomerase I interaction. *J. Biol. Chem.*, **279**, 39686–39696.
 101. Edwards, T.K., Saleem, A., Shaman, J.A., Dennis, T., Gerigk, C., Oliveros, E., Gartenberg, M.R. and Rubin, E.H. (2000) Role for

- nucleolin/Nsr1 in the cellular localization of topoisomerase I. *J. Biol. Chem.*, **275**, 36181–36188.
102. Takahashi, H., Sawa, H., Hasegawa, H., Shoya, Y., Sata, T., Hall, W.W., Nagashima, K. and Kurata, T. (2002) Topoisomerase I and ATP activate cDNA synthesis of human immunodeficiency virus type 1. *Biochem. Biophys. Res. Commun.*, **294**, 509–517.
 103. Tyagi, S., Ochem, A. and Tyagi, M. (2011) DNA-dependent protein kinase interacts functionally with the RNA polymerase II complex recruited at the human immunodeficiency virus (HIV) long terminal repeat and plays an important role in HIV gene expression. *J. Gen. Virol.*, **92**, 1710–1720.
 104. Kraus, W.L. (2008) Transcriptional control by PARP-1: chromatin modulation, enhancer-binding, coregulation, and insulation. *Curr. Opin. Cell Biol.*, **20**, 294–302.
 105. Krishnakumar, R. and Kraus, W.L. (2010) The PARP side of the nucleus: molecular actions, physiological outcomes, and clinical targets. *Mol. Cell*, **39**, 8–24.
 106. Lin, Y., Tang, X., Zhu, Y., Shu, T. and Han, X. (2011) Identification of PARP-1 as one of the transcription factors binding to the repressor element in the promoter region of COX-2. *Arch. Biochem. Biophys.*, **505**, 123–129.
 107. Oei, S.L., Griesenbeck, J., Schweiger, M. and Ziegler, M. (1998) Regulation of RNA polymerase II-dependent transcription by poly(ADP-ribosylation) of transcription factors. *J. Biol. Chem.*, **273**, 31644–31647.
 108. Parent, M., Yung, T.M.C., Rancourt, A., Ho, E.L.Y., Vispé, S., Suzuki-Matsuda, F., Uehara, A., Wada, T., Handa, H. and Satoh, M.S. (2005) Poly(ADP-ribose) polymerase-1 is a negative regulator of HIV-1 transcription through competitive binding to TAR RNA with Tat-positive transcription elongation factor b (p-TEFb) complex. *J. Biol. Chem.*, **280**, 448–457.
 109. Uribe, D.J., Guo, K., Shin, Y.-J. and Sun, D. (2011) Heterogeneous nuclear ribonucleoprotein K and nucleolin as transcriptional activators of the vascular endothelial growth factor promoter through interaction with secondary DNA structures. *Biochemistry*, **50**, 3796–3806.
 110. Ogawa, H., Ishiguro, K.-I., Gaubatz, S., Livingston, D.M. and Nakatani, Y. (2002) A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science*, **296**, 1132–1136.
 111. Yoo, J.-Y., Choi, K.-C., Kang, H., Kim, Y.J., Lee, J., Jun, W.J., Kim, M.-J., Lee, Y.-H., Lee, O.-H. and Yoon, H.-G. (2010) Histone deacetylase 3 is selectively involved in L3MBTL2-mediated transcriptional repression. *FEBS Lett.*, **584**, 2225–2230.
 112. Ishaq, M., Hu, J., Wu, X., Fu, Q., Yang, Y., Liu, Q. and Guo, D. (2008) Knockdown of cellular RNA helicase DDX3 by short hairpin RNAs suppresses HIV-1 viral replication without inducing apoptosis. *Mol. Biotechnol.*, **39**, 231–238.
 113. Guo, J., Hong, F., Loke, J., Yea, S., Lim, C.L., Lee, U., Mann, D.A., Walsh, M.J., Sninsky, J.J. and Friedman, S.L. (2010) A DDX5 S480A polymorphism is associated with increased transcription of fibrogenic genes in hepatic stellate cells. *J. Biol. Chem.*, **285**, 5428–5437.
 114. Fuller-Pace, F.V. and Ali, S. (2008) The DEAD box RNA helicases p68 (Ddx5) and p72 (Ddx17): novel transcriptional co-regulators. *Biochem. Soc. Trans.*, **36**, 609–612.
 115. Warner, D.R., Bhattacharjee, V., Yin, X., Singh, S., Mukhopadhyay, P., Pisano, M.M. and Greene, R.M. (2004) Functional interaction between Smad, CREB binding protein, and p68 RNA helicase. *Biochem. Biophys. Res. Commun.*, **324**, 70–76.
 116. Rossow, K.L. and Janknecht, R. Synergism between p68 RNA helicase and the transcriptional coactivators CBP and p300. *Oncogene*, **22**, 151–156.
 117. Williams, S.A., Chen, L.-F., Kwon, H., Ruiz-Jarabo, C.M., Verdin, E. and Greene, W.C. (2006) NF-kappaB p50 promotes HIV latency through HDAC recruitment and repression of transcriptional initiation. *EMBO J.*, **25**, 139–149.
 118. Liao, M., Zhang, Y., Kang, J.-H. and Dufau, M.L. (2011) Coactivator function of positive cofactor 4 (PC4) in Sp1-directed luteinizing hormone receptor (LHR) gene transcription. *J. Biol. Chem.*, **286**, 7681–7691.
 119. Dechat, T., Adam, S.A., Taimen, P., Shimi, T. and Goldman, R.D. (2010) Nuclear lamins. *Cold Spring Harb. Perspect. Biol.*, **2**, a000547.
 120. Agresti, A., Lupo, R., Bianchi, M.E. and Müller, S. (2003) HMGB1 interacts differentially with members of the Rel family of transcription factors. *Biochem. Biophys. Res. Commun.*, **302**, 421–426.
 121. Dobi, A., Szemes, M., Lee, C., Palkovits, M., Lim, F., Gyorgy, A., Mahan, M.A. and Agoston, D.V. (2006) AUF1 is expressed in the developing brain, binds to AT-rich double-stranded DNA, and regulates enkephalin gene expression. *J. Biol. Chem.*, **281**, 28889–28900.
 122. Lee, C., Gyorgy, A., Maric, D., Sadri, N., Schneider, R.J., Barker, J.L., Lawson, M. and Agoston, D.V. (2008) Members of the NuRD chromatin remodeling complex interact with AUF1 in developing cortical neurons. *Cereb. Cortex*, **18**, 2909–2919.
 123. Wolf, D., Witte, V., Clark, P., Blume, K., Lichtenheld, M.G. and Baur, A.S. (2008) HIV Nef enhances Tat-mediated viral transcription through a hnRNP-K-nucleated signaling complex. *Cell Host Microbe*, **4**, 398–408.
 124. Küniger, D.T., Izumi, T., Papaconstantinou, J. and Mitra, S. (2002) Human AP-endonuclease 1 and hnRNP-L interact with a nCaRE-like repressor element in the AP-endonuclease 1 promoter. *Nucleic Acids Res.*, **30**, 823–829.
 125. Obrdlik, A., Kukalev, A., Louvet, E., Farrants, A.-K.O., Caputo, L. and Percipalle, P. (2008) The histone acetyltransferase PCAF associates with actin and hnRNP U for RNA polymerase II transcription. *Mol. Cell Biol.*, **28**, 6342–6357.
 126. Choi, H.S., Hwang, C.K., Song, K.Y., Law, P.-Y., Wei, L.-N. and Loh, H.H. (2009) Poly(C)-binding proteins as transcriptional regulators of gene expression. *Biochem. Biophys. Res. Commun.*, **380**, 431–436.
 127. Fukuda, A., Nakadai, T., Shimada, M. and Hisatake, K. (2009) Heterogeneous nuclear ribonucleoprotein R enhances transcription from the naturally configured c-fos promoter in vitro. *J. Biol. Chem.*, **284**, 23472–23480.
 128. Miralles, F. and Visa, N. (2006) Actin in transcription and transcription regulation. *Curr. Opin. Cell Biol.*, **18**, 261–266.
 129. Zheng, B., Han, M., Bernier, M. and Wen, J. (2009) Nuclear actin and actin-binding proteins in the regulation of transcription and gene expression. *FEBS J.*, **276**, 2669–2685.