Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription

Elena Tosoni[1], Ilaria Frasson[1], Matteo Scalabrin[1,2], Rosalba Perrone[1], Elena Butovskaya[1], Matteo Nadai[1], Giorgio Palù[1], Dan Fabris[2] and Sara N. Richter[1,*]

[1]Department of Molecular Medicine, University of Padua, via Gabelli 63, 35121 Padua, Italy and [2]The RNA Institute, University at Albany-SUNY, Albany, NY 12222, USA

ABSTRACT

Folding of the LTR promoter into dynamic G-quadruplex conformations has been shown to suppress its transcriptional activity in HIV-1. Here we sought to identify the proteins that control the folding of this region of proviral genome by inducing/stabilizing G-quadruplex structures. The implementation of electrophorethic mobility shift assay and pull-down experiments coupled with mass spectrometric analysis revealed that the cellular protein nucleolin is able to specifically recognize G-quadruplex structures present in the LTR promoter. Nucleolin recognized with high affinity and specificity the majority, but not all the possible G-quadruplexes folded by this sequence. In addition, it displayed greater binding preference towards DNA than RNA G-quadruplexes, thus indicating two levels of selectivity based on the sequence and nature of the target. The interaction translated into stabilization of the LTR G-quadruplexes and increased promoter silencing activity; in contrast, disruption of nucleolin binding in cells by both siRNAs and a nucleolin binding aptamer greatly increased LTR promoter activity. These data indicate that nucleolin possesses a specific and regulated activity toward the HIV-1 LTR promoter, which is mediated by G-quadruplexes. These observations provide new essential insights into viral transcription and a possible low mutagenic target for antiretroviral therapy.

INTRODUCTION

G-quadruplexes (G4s) are nucleic acids secondary structures that may form in single-stranded G-rich DNAs and RNAs under physiological conditions (1–3). Four Gs bind *via* Hoogsteen-type hydrogen bonds to yield G-quartets that in turn stack on top of each other to form the G4. G4s are highly polymorphic, both in terms of strand stoichiometry (forming both inter- and intramolecular structures) and strand orientation/topology. The presence of K^+ cations specifically supports G4 formation and stability (4–6). In eukaryotes and prokaryotes, G4 DNA motifs have been found in telomeres, G-rich micro- and mini-satellites, near promoters, and within the ribosomal DNA (rDNA) (7–9). In the human genome, genes that are near G4 DNA motifs fall into specific functional classes; for example, promoters of oncogenes and tumor suppressor genes have particularly high and low G4-forming potential, respectively (10–12). Human G4 DNA motifs have been reported to be associated with recombination prone regions (13) and to show mutational patterns that preserved the potential to form G4 DNA structures (9). RNA G4s have been detected in the 5′ and 3′-UTR and coding regions, in which they act as important regulators of pre-mRNA processing (splicing and polyadenylation), RNA turnover, mRNA targeting and translation (14,15). Regulatory mechanisms controlled by G4s involve the binding of protein factors that modulate G4 conformation and/or serve as a bridge to recruit additional protein regulators. Indeed, G4 binding proteins can be classified into three functional groups: telomere-related proteins, such as the shelterin complex; proteins that unfold the G4 structure, such as the helicase and heterogeneous nuclear ribonucleoprotein families; proteins that stabilize G4s, a large group which includes nucleolin, MAZ and nucleophosmin (3,16–18).

G4 structures and their cognate proteins are key players in numerous essential processes in eukaryotic cells. Their misregulation has been associated with a number of relevant human diseases, such as the amyotrophic lateral sclerosis (19–21), Alzheimer (22) and fragile X syndrome (23), in which expansion of G4-forming regions has been reported. Moreover, mutations in G4-interacting proteins have been linked to genetic diseases, such as the Werner syndrome and Fanconi anemia (24,25). In recent years, new studies have

---

contributed to increase our knowledge of the biological significance of G4s in prokaryotes (26,27) and viruses (28). We and other groups have identified functionally significant G4s in the Nef coding region (29) and the unique LTR promoter (30–32) of the human immunodeficiency virus (HIV), the etiologic agent of the acquired immune deficiency syndrome (AIDS). These studies have shown that G4 folding at the LTR promoter decreased viral transcription with an effect that was augmented by G4 ligands (30,33). In this direction, the significance of these structures as focal points of interactions with host and viral factors is supported also by the observation that G4-folded sequences are specifically recognized by various viral proteins, such as the Epstein Barr Virus Nuclear Antigen 1 (34,35) and the SARS coronavirus unique domain (SUD), which occurs exclusively in highly pathogenic strains (36). For this reason, we decided to pursue the investigation of putative cellular/viral proteins that may be involved in the regulation of the G4 LTR promoter activity in HIV. We employed a concerted approach combining electrophorethic mobility shift assay (EMSA) and analysis by electrospray ionization mass spectrometric (ESI-MS) to identify possible factors capable of binding the LTR G4 structure. In order to validate the findings, we then tested their stabilizing activity on the G4 fold and evaluated their ability to inhibit LTR-driven transcription in cells. The results provided new insights into the role of the LTR G4 in the viral life cycle, which could pave the way for the possible development of novel therapeutic strategies.

MATERIALS AND METHODS

Oligonucleotides, plasmids and aptamers

All desalted oligonucleotides and aptamers were purchased from Sigma-Aldrich, Milan, Italy (Supplementary Table S1). The HIV-1 LTR region was inserted into the promoterless luciferase reporter vector pGL4.10-Luc2 (Promega Italia, Milan, Italy) to form the pGL4.10-LTR-Luc2 vector, as previously reported (30). The Renilla plasmid (p4.74, Promega Italia, Milan, Italy) was used as an internal control. The human enhanced green fluorescent protein-nucleolin plasmid (GFP-Nucleolin) was purchased from Addgene (Addgene, Cambridge, MA, USA). The pEGFP empty vector was used as control (Clontech, Takara Bio, Otsu, Japan).

Cell culturing

Human embryonic kidney (HEK) 293T cells (ATCC # CRL-3216) were grown in DMEM (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% heat-inactivated fetal bovine serum (FBS, Gibco, Thermo Fisher Scientific, Waltham, MA, USA). Jurkat T-lymphocytes cells (ATCC # TIB-152) were grown in RPMI 1640 (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% heat-inactivated FBS. MCF-7 human breast cancer cells (ATCC # HTB-22) were grown in RPMI 1640 supplemented with 10% heat-inactivated FBS. MCF10A normal human mammary epithelial cells (ATCC # CRL-10317) were grown in DMEM/F12 (Gibco,

Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% heat-inactivated FBS and EGF (0.02 μg/ml), hydrocortisone (0.5 μg/ml), cholera toxin (0.1 μg/ml), insulin (10 μg/ml) (purchased all from Sigma-Aldrich, Milan, Italy). All cultures were grown in a humidified incubator maintained at 37°C with 5% CO_2.

Protein nuclear extraction and electrophoretic mobility shift assay (EMSA)

Oligonucleotides were 5′-end labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase at 37°C for 30 min. After DNA precipitation, labeled species were resuspended in lithium cacodylate buffer (10 mM, pH 7.4) and KCl 100 mM. The oligonucleotides were denatured for 5 min at 95°C and gradually cooled to room temperature to achieve proper folding of G-quadruplex structures. Protein nuclear extracts of HEK 293T and Jurkat cells were obtained by using NX-TRACT kit (Sigma-Aldrich, Milan, Italy). Recombinant full-length human nucleolin was expressed in HEK293 cells and purified as described by the manufacturer (Origene Technologies, Rockville, USA). Labeled oligonucleotides (40 nM) were incubated in 20 μl of reaction in EMSA binding buffer and nuclear extract (0.5 μg/μl) or purified nucleolin (300 ng) for 2 h at 37°C. EMSA binding buffer composition was: Tris–HCl 20 mM, pH 8, KCl 30 mM, $MgCl_2$ 1.5 nM, DTT 1 mM, glycerol 8%, protease inhibitor cocktail (Sigma-Aldrich, Milan, Italy) 1%, NaF 5 mM, Na_3VO_4 1 mM, poly [dI-dC] (Sigma-Aldrich, Milan, Italy) 1.25 ng/μl.

In competition experiments, an excess of cold oligonucleotides was added to the samples and their ability to disrupt the G4 structures was monitored to evaluate binding specificity. After incubation, reaction solutions were loaded onto 5% native polyacrylamide gel in 1× TBE buffer and KCl 100 mM. DNA–protein complexes were resolved by running the gel overnight at 27 V at 4°C. EMSA gels were dried using a gel dryer (Bio-Rad Laboratories, Milan, Italy), free and bound DNA molecules were visualized by phosphorimaging (Typhoon FLA9000, GE Healthcare Europe, Milan, Italy) and quantified by ImageQuant TL Software (GE Healthcare Europe, Milan, Italy). After the desired complex was located on the gel, the corresponding band was cut and either directly in-gel digested for mass spectrometric (MS) analysis, or further purified by SDS-PAGE. Extraction was performed in SDS-PAGE sample buffer. After 5 min in boiling water, samples were incubated overnight at 37°C. Finally, supernatant was loaded on 12% SDS-PAGE. The band of interest was excised after coomassie staining.

Protein extraction from HIV-1 infected cells

HEK 293T cells were seeded in a 10-cm dish in DMEM supplemented with 10% heat-inactivated FBS and incubated overnight. Cells were next either mock-transfected or transfected with of pNL4-3 (the reagent was obtained throughout the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH) (37) using CalPhos^{TM} Mammalian Transfection Kit (Clontech, Otsu, Japan) according to the manufacturer's protocol. After 8 h, cells were washed with PBS 1× and fresh growth medium was added. Forty eight hours post-transfection, cells were washed twice with cold

PBS and scraped off. After a short centrifugation, pellet was resuspended in total protein extraction buffer (KCl 600 mM, Tris–HCl 20 mM, pH 7.8, glycerol 20%, DTT 2 mM, protease inhibitor cocktail). Cells were then lysed with three repeated freeze/thaw cycles and supernatant cleared by centrifugation, stored at −80°C, and subsequently used in EMSA assays.

Mass spectrometric (MS) protein identification

Bands were treated according to established in-gel digestion protocols. Briefly, they were first washed with 50% CH_3OH and 2.5% acetic acid, dehydrated with CH_3CN, and then reduced with 30 μl of DTT (10 mM in 100 mM NH_4HCO_3) for 30 min at room temperature. The excess of DTT was eliminated before treating the bands with 30 μl of iodoacetamide (50 mM in 100 mM NH_4HCO_3) for 30 min at room temperature in order to alkylate cysteine residues. Bands were washed with 100 mM NH_4HCO_3, dehydrated with CH_3CN twice, and then digested. A 1 μg aliquot of MS-grade trypsin (ThermoFisher Scientific, Waltham, MA, USA) in 50 μl of 50 mM NH_4HCO_3 was added to the dehydrated bands, followed by incubation on ice for 30 min. The excess of trypsin was eliminated and substituted with 20 μl of 50 mM NH_4HCO_3 and the sample was incubated overnight at 37°C. Peptides were extracted twice with 5% formic acid and two more times with 50% CH_3CN, 5% formic acid. The peptide mixture was further desalted in a silica nanocolumn (Polymicro Technologies, Phoenix, AZ, USA) packed in house with pinnacle C18 pack material (Thermo Fisher Scientific, Waltham, MA, USA). All materials were MS grade purchased from Sigma Aldrich, St. Louis, MO, US except where otherwise indicated.

The desalted mixture was finally analyzed by direct infusion electrospray ionization (ESI) on a Thermo Fisher Scientific (Waltham, MA, USA) LTQ-Orbitrap Velos mass spectrometer. The instrument was calibrated by using a 0.5 mg/ml solution of CsI in 50% CH_3OH, which provided a typical <2 ppm mass accuracy. All analyses were performed in nanoflow mode by utilizing quartz emitters produced in house by using a P2000 laser pipette puller (Sutter Instruments Co., Novato, CA, USA). Up to 5 μl samples were typically loaded onto each emitter by using a gel-loader pipette tip. A stainless steel wire was inserted through the back-end of the emitter to supply an ionizing voltage that ranged between 0.8 and 1.2 kV. Bands containing bovine serum albumin (BSA) and empty gel bands were used as positive and negative control, respectively. Putative peptides that were not present in blank samples were submitted to tandem mass spectrometric (MS/MS) analysis. These determinations involved isolating the precursor ion of interest in the LTQ element of the instrument, activating fragmentation in either the LTQ or the C-trap, and performing fragment detection in the Orbitrap. The masses of the 50 more intense fragments were employed to perform a Mascot Database Search (38) to identify their parent protein. The matched protein was deemed as being positively identified when two or more peptides provided a mascot score greater than 22.

Pull-down assay

HEK 293T protein nuclear extract (0.6 μg/μl) was incubated with biotinylated LTR oligo G4 folded (600 nM) in 250 μl of reaction containing Tris–HCl 20 mM, pH 8, KCl 30 mM, $MgCl_2$ 1.5 nM, protease inhibitor cocktail 1%, NaF 5 mM, Na_3VO_4 1 mM, poly [dI-dC] 1.25 ng/μl for 2 h at 37°C. The binding reaction was followed by incubation (2 h at 37°C) with 30 μl of streptavidin-agarose beads (Sigma-Aldrich, Milan, Italy). After PBS washes, proteins were eluted with increasing amount of NaCl (0.2 and 1M), and concentrated with Amicon Ultra 0.5 (Merck Millipore, Germany). Beads were collected by brief centrifugation, resuspended in 50 μl of Laemmli buffer, and finally incubated at 95°C for 5 min. Supernatants were separated on SDS-PAGE and analyzed by western blot.

FRET-melting assay

Oligonucleotides were diluted to 0.1 μM in lithium cacodylate buffer (10 mM, pH 7.4) and KCl 100 mM heat denatured for 5 min at 95°C, and folded in G4 structure at room temperature for 16 h. Samples were incubated alone, with purified human nucleolin (530 ng) or bovine serum albumin (BSA, negative control) for 1h at 37°C. Fluorescence melting curves were determined by using a LightCycler II (Roche, Milan, Italy). After a first equilibration step at 30°C for 2 min, a stepwise increase of 1°C every minute for 65 cycles was performed to reach 95°C. A measurement was completed after each cycle by using 470 nm excitation and 530 nm detection. Oligonucleotide melting was monitored by observing 6-carboxyfluorescein (6-FAM) emission, which was normalized between 0 and 1. T_m was defined as the temperature for which the normalized emission was 0.5.

siRNA and luciferase reporter assay

Gene-specific pooled siRNA trilencer targeting human NCL and a scrambled negative control duplex were purchased from Origene (NCL Trilencer-27 Human siRNA, OriGene Technologies, Rockville, MD, USA). MCF7 cells were transfected with 1, 2 and 4 nM aliquots of human NCL siRNA and control siRNA by using Lipofectamine RNAiMAX (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer's instructions. pLTR luciferase plasmid and Renilla construct were transfected into the same cells 24 h later by using Lipofectamine 3000 (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA). In the double transfected cells, LTR promoter activity was assessed as firefly luciferase signal, normalized to Renilla luciferase activity, by using Dual-Glo® Luciferase Assay System (Promega Italia, Milan, Italy), according to the manufacturer's directions (30). Depending on the transfected cell line (i.e. either MCF7 or MCF10A), the pGL4.10-LTR-Luc2 vector provided signals ranging form 10^3 to 2×10^6 luciferase units, as measured by a Victor X2 multilabel plate reader (Perkin Elmer Italia, Milan, Italy). In contrast, the promoterless pGL4.10-Luc2 and untransfected cells displayed a background signal lower than 10 luciferase units. All data were acquired in medium-free PBS. DNA aptamers AS1411 and the control CRO26 were added to cell medium at the time of transfection of

pGL4.10-LTR-Luc2 and p4.74-Renilla plasmids and the luciferase signal read 24 h after transfection. Each assay was performed in duplicate and each set of experiments was repeated at least three times.

Immunoblot analysis

Immunoblot analysis was performed on cell protein extracts obtained as previously described (39). Protein concentrations were quantified by using the Pierce® BCA Protein Assay Kit (Thermo Scientific, Rockford, IL, USA) and the samples stored at −80°C. Each sample was electrophoresed on 12% SDS-PAGE and transferred to a nitrocellulose blotting membrane (Amersham TM Protan TM, GE Healtcare Life science, Milan, Italy) by using trans-blot SD semi-dry transfer cell (Bio-Rad Laboratories, Milan, Italy). The membranes were blocked with 5% skim milk in PBST (0.05% Tween 20 in PBS). Membranes were incubated with the respective primary antibody directed against NCL (rabbit polyclonal C23 (H-250); Santa Cruz Biotechnology, Dallas, TX, USA), p24 (rabbit polyclonal; Abcam, Cambridge, UK), and β-actin (mouse monoclonal; Sigma-Aldrich, Milan, Italy). After three washes in PBST, membranes were incubated with ECL Plex Goat-α-Rabbit IgG-Cy5 or ECL Plex Goat-α-Mouse IgG-Cy5 (GE Healthcare Life sciences, Milan, Italy). Images were captured on the Typhoon FLA 9000, and quantified by ImageQuant TL software.

Taq polymerase stop assay

Taq polymerase stop assay was carried out as previously described (30). Briefly, the 5′-end labeled primer (5′-GGCAAAAAGCAGCTGCTTATATGCAG-3′) was annealed to the template (Supplementary Table S1) in lithium cacodylate buffer in the presence or absence of KCl 100 mM by heating at 95°C for 5 min and gradually cooling to room temperature. Where specified, samples were incubated with purified human nucleolin (61 ng) at 37°C for 2 h. Primer extension was then conducted by using 2 U of Ampli*Taq* Gold DNA polymerase (Life Technologies, Thermo Fisher Scientific) for 30 min at 37°C or 47°C. Reactions were stopped by ethanol precipitation. Primer extension products were separated on a 15% denaturing gel, and finally visualized by phosphorimaging (Typhoon FLA 9000).

Dimethylsulfate footprinting

The DNA substrate of interest was gel-purified before use and prepared in desalted/lyophilised form. The oligonucleotide was 5′-end-labeled with [γ-^{32}P]ATP by T4 polynucleotide kinase, purified by using MicroSpin G-25 columns (Amersham Biosciences, Europe), resuspended in lithium cacodylate buffer 10 mM, pH 7.4, KCl 100 mM, heat-denatured and folded. The oligonucleotide (1.23 μM) was incubated either alone or with purified human nucleolin (300 ng) in EMSA binding buffer for 2 h at 37°C. Sample solutions were then treated with dimethylsulfate (DMS, 0.5% in ethanol) for 5 min and stopped by addition of gel loading buffer containing 10% glycerol and β-mercaptoethanol. Samples were loaded onto 16% native polyacrylamide gels

and run until the desired resolution was obtained. DNA bands were localized via autoradiography, excised and eluted overnight. The supernatants were recovered, ethanol-precipitated and treated with piperidine 1M for 30 min at 90°C. Samples were dried in a speed-vac, washed with water, dried again and resuspended in formamide gel loading buffer. Reaction products were analyzed on 20% denaturing polyacrylamide gels, visualized by phosphorimaging analysis, and quantified by ImageQuant TL software.

Surface plasmon resonance (SPR) analysis

SPR was performed on the Biacore T100 platform (GE Healthcare, Life Science, Milan, Italy). Purified human NCL was immobilized on Serie S sensor chip CM5 by amine coupling. Immobilization was performed in HEPES–NaCl running buffer (HEPES pH 7.4 0.01 M, NaCl 0.15 M, EDTA 3 mM). The protein diluted in sodium acetate buffer pH 4.0 at the concentration of 15 ng/μl was injected to reach the response of around 1000 RU. Blank immobilization was performed in the flow cell 1 to permit reference subtraction. LTR-II+III+IV wt and scrambled sequences binding analysis was performed at a flow rate of 25 μl/min, with contact time of 180 s and dissociation time of 240 s in HEPES-KCl buffer (HEPES pH 7.4 15 mM, KCl 0.25 M, EDTA 3 mM). Oligonucleotides were diluted from stock to the concentration of 2 μM in HEPES-KCl buffer, denaturated at 95°C for 5 min and allowed to cool at room temperature to permit G4 formation. Sensograms were obtained in the concentration range of 31.25 nM–2 μM. After each oligonucleotide injection the chip surface was regenerated with KCl 1M solution. All sensograms were corrected by reference subtraction of blank flow cell response and buffer injection response. Data were fitted to a 1:1 binding model with Rmax initial parameter set to theoretical calculated Rmax of 157 using BIAevaluation software (GE Healthcare).

RESULTS

A cellular protein binds specific G-quadruplexes of the HIV-1 LTR promoter

In previous work, we demonstrated that the HIV-1 LTR region can fold at least three different G-quadruplex (G4) structures at positions −92/−48 with respect to the transcription initiation site of the representative HXB2_LAI (NC_001802) strain (30). Disruption of G4 by point mutations increased the transcript levels, thus indicating that G4 formation may contribute to the modulation of viral transcription. We reasoned that stabilization and unfolding of G4 at the LTR level are likely regulated by interactions with viral/cellular proteins. In order to test this hypothesis, LTR G4-forming sequences were incubated with nuclear protein extracts. LTR sequences of different lengths were employed to check whether individual G4 structures identified in the LTR promoter displayed different protein binding capabilities. In particular, we assayed sequences that possessed the minimal requirements to fold into an individual G-quadruplex (i.e. only 4 G-tracts: LTR-II, LTR-III and LTR-IV) and sequences capable of providing multiple G4s (i.e. more than 4 G-tracts: LTR-II+III+IV and
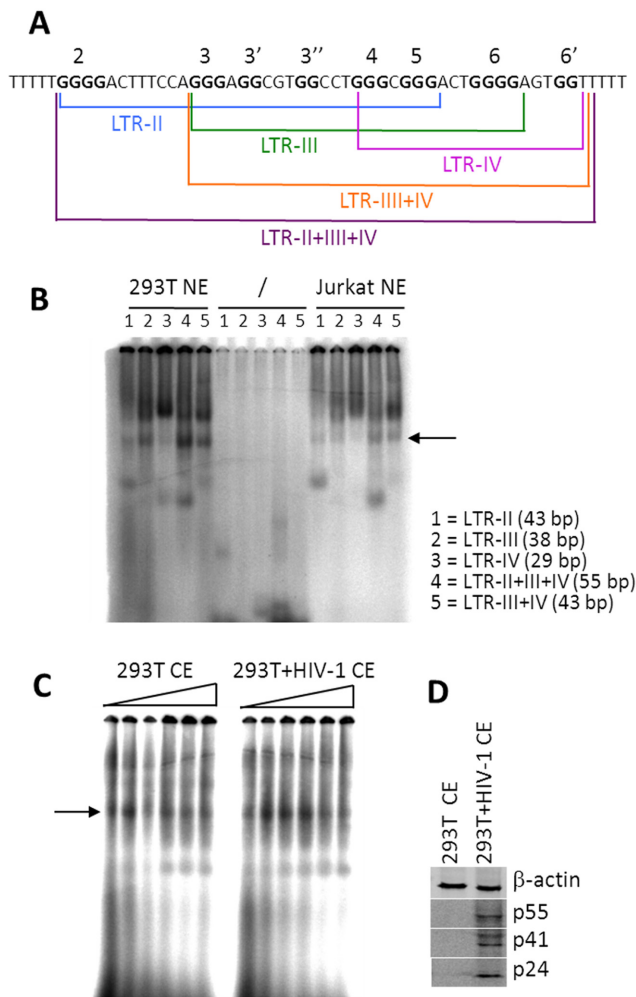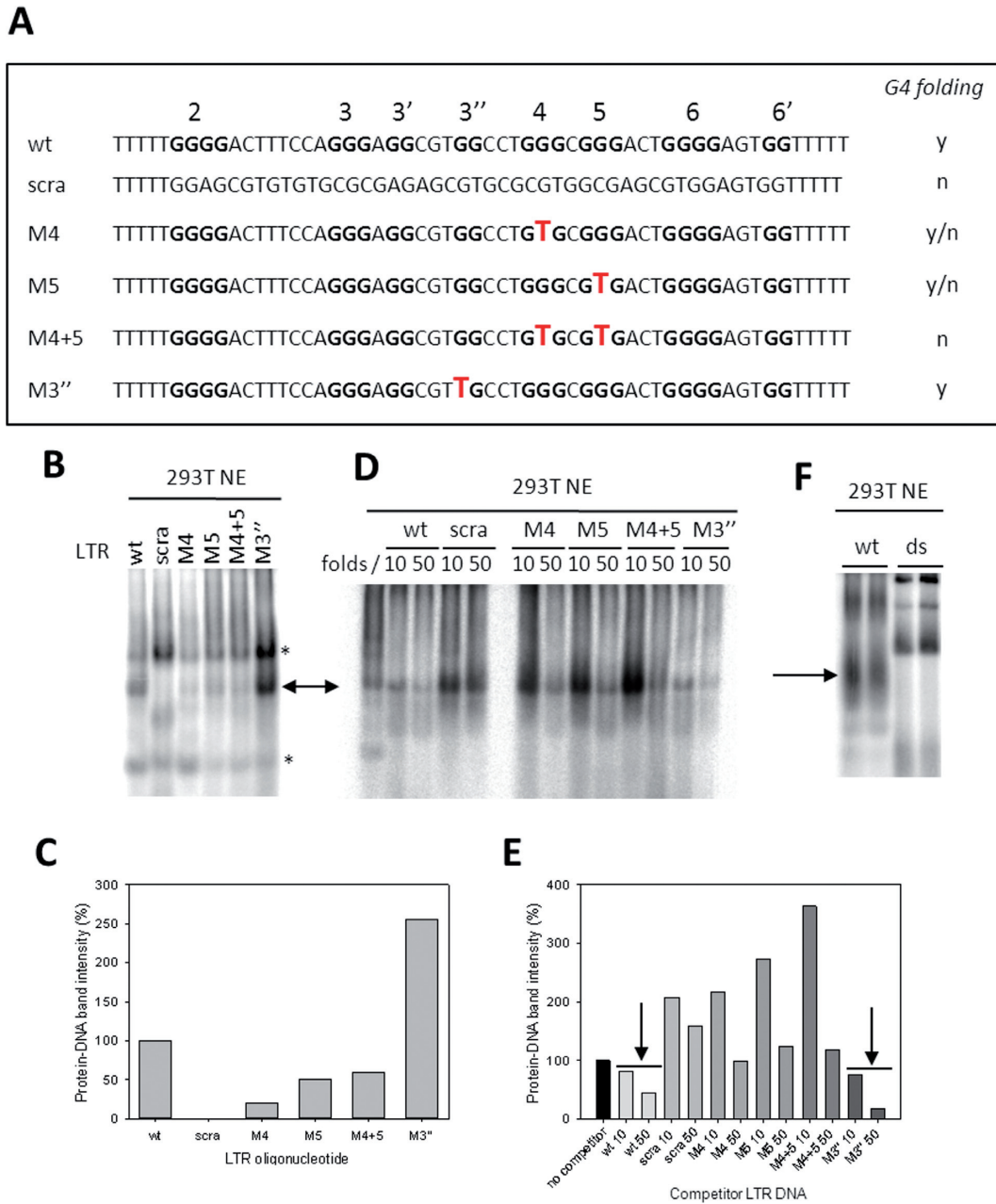
Figure 1. EMSA analysis of cellular and viral proteins binding to the G4 LTR sequences. (**A**) Sequence of the G-rich LTR region spanning −92/−48 nts. G bases in G-tracts or involved in G-quadruplexes are shown in bold. G-tracts are numbered. Brackets indicate sequences used in this study that form G-quadruplexes with three stacked tetrads. (**B**) Analysis of cellular proteins that bind to the LTR sequences. Protein nuclear extract (NE) of 293T and Jurkat cells were incubated with the different length G4 LTR sequences. Binding was analysed by native polyacrylamide gel. Lane numbers correspond to the different LTR sequences, as shown in the legend at the right bottom of the gel image. Length of the LTR oligonucleotides is shown in brackets (see also Supplementary Table S1 for reference). The relevant protein/DNA complex band present in all G4 LTR sequences is indicated by the arrow. (**C**) Analysis of viral proteins that bind to the LTR-II+III+IV sequence. Cell extracts (CE) of 293T cells either non-producing or producing HIV-1 were incubated with the LTR-II+III+IV G4 and analysed by native polyacrylamide gel. The relevant protein/DNA complex band present in all G4 LTR sequences is indicated by the arrow. D) Western blot analysis of cell extracts (CE) of 293T cells either non-producing or producing HIV-1. The effective presence of the virus in transfected cells was evaluated with a HIV-1 p24 antibody.

LTR-III+IV) (Figure 1A). The corresponding ^{32}P-labeled oligonucleotides were incubated with nuclear extracts derived from two cell lines: Jurkat T-lymphocytes that are a model for the natural HIV-1 targets *in vivo*; human embryonic kidney 293T cells that lack HIV-1 cell receptors and sustain all viral steps with the exception of virion attachment and entry. The latter can however be transfected

with the HIV-1 proviral genome to produce fully competent and infectious viral particles, which indicates that their cytoplasmic/nuclear protein makeup is competent to sustain viral replication. Samples were analysed onto native polyacrylamide gels to monitor the formation of slower running bands corresponding to oligonucleotide-protein complexes. As shown in Figure 1B, all G4 LTR oligonucleotides formed slower running bands. The fact that the patterns obtained from the two nuclear extracts were essentially identical indicated that the selected cell lines contained very similar sets of G4-binding proteins. In particular, a very specific band migrated with the same rate in all G4 oligonucleotide samples (arrow in Figure 1B), thus suggesting that the same protein was able to bind all LTR G4 sequences considered. The intensity of the observed bands indicated that the longer sequence (i.e. LTR-II+III+IV) was bound most efficiently, whereas the shorter sequence (i.e. LTR-IV) supported only very modest complex formation. The LTR-II+III+IV oligonucleotide was incubated with extracts of HIV-1 producing and non-producing 293T cells to test whether the presence of viral proteins affected in any detectable way the observed EMSA profiles (Figure 1C). The actual presence of viral proteins in the transfected cells was assessed by western blot analysis (Figure 1D). While viral proteins were well represented in the transfected extract, the EMSA determinations revealed no major difference and confirmed that cellular proteins must constitute the major players in LTR G4 binding (Figure 1C).

In previous work, we demonstrated that one- or two-nucleotide point mutations in the G-tracts involved in G4 pairing were able to partially or totally disrupt, respectively, G4 folding in the LTR sequence (LTR-M4, LTR-M5, LTR-M4+5), whereas a mutation in the loop was not (LTR-M3", Figure 2A) (30). Based on these observations, we used mutant LTR-II+III+IV oligonucleotides to assess the specificity of the observed protein for the G4 structure (Figure 2B). When incubated with the nuclear extracts, only the sequences that retained G4-folding capabilities (i.e. wt and M3") were able to form bands corresponding to the desired complexes (arrow in Figure 2B), whereas the mutants that partially fold or cannot form stable G4s were much less efficiently bound. A scrambled sequence matching the wt base composition did not bind at all (Figure 2B and C). It should be noted that both a slower and a faster running bands were present in all samples (see asterisk in Figure 2B), indicative of nucleic acid binding proteins that may not be selective for G4. A competition experiment was performed to confirm the binding specificity of the selected protein, which involved mixing the labeled wt sequence with increasing amounts of unlabeled wt, mutant LTRs or scrambled sequence. As shown in Figure 2D and E, only the wt and M3" sequences were able to effectively compete for protein binding, whereas the other oligonucleotides did not decrease the amount of protein bound to wt. Interestingly, however, 10-fold addition of M4, M5 and M4+5 oligonucleotides increased the amount of protein bound to the wt sequence. In this case, the mutant sequences may compete for binding to proteins that are sequence- but not structure-specific, thus leaving more G4-specific protein available to bind to the G4-folded wt sequence. Furthermore, the fact that wt sequences folding G4 or ds structures bound different pro-

A

		2	3	3'	3''	4	5	6	6'	G4 folding
wt		TTTTT**GGGG**ACTTTCCA**GGG**A**GG**CGT**GG**CCT**GGG**C**GGG**ACT**GGGG**AGT**GG**TTTTT								y
scra		TTTTTGGAGCGTGTGTGCGCGAGAGCGTGCGCGTGGCGAGCGTGGAGTGGTTTTT								n
M4		TTTTT**GGGG**ACTTTCCA**GGG**A**GG**CGT**GG**CCTG**T**GC**GGG**ACT**GGGG**AGT**GG**TTTTT								y/n
M5		TTTTT**GGGG**ACTTTCCA**GGG**A**GG**CGT**GG**CCT**GGG**C**G**T**GACT**GGGG**AGT**GG**TTTTT								y/n
M4+5		TTTTT**GGGG**ACTTTCCA**GGG**A**GG**CGT**GG**CCTG**T**GCG**T**GACT**GGGG**AGT**GG**TTTTT								n
M3''		TTTTT**GGGG**ACTTTCCA**GGG**A**GG**CGT**T**GCCT**GGG**C**GGG**ACT**GGGG**AGT**GG**TTTTT								y

B 293T NE

D 293T NE

F 293T NE

C

E

Figure 2. Binding of nuclear extract proteins to the wild-type (wt) and mutant LTR-II+III+IV sequences. (**A**) Sequences of the wt and mutant LTRs. G-tracts are shown in bold and are numbered (2–6') above the sequence. In the mutant sequences, the mutated bases are shown in red. The names of the mutant sequences correspond to the G-tracts where G bases have been mutated. Scra stands for a sequence where Gs have been scrambled to such an extent that no G4 can form. The ability of the mutant sequences to fold into G4 is reported on the right: y = folding, y/n = partial folding, n = no folding (measured as reported in (30)). (**B**) Binding of nuclear extract proteins to the wt and mutant labeled G4 LTR sequences. The G4-specific protein/DNA complex is indicated by the arrow, while the asterisks indicate non-specific protein/DNA complexes. (**C**) Quantification of G4-specific protein/DNA complex reported in (B). (**D**) Competition of the G4-specific protein/DNA complex with unlabelled wt and mutant LTR sequences. The molar excess of the competitor sequences is shown above each lane. (**E**) Quantification of G4-specific protein/DNA complex reported in (D). The arrows indicate the only two unlabeled sequences able to compete the wt LTR/protein complex. (**F**) Binding of different amounts of nuclear extract proteins to the wt G4 folded and ds LTR-II+III-IV sequences.

teins confirmed that the observed protein was specific for the G4 conformation (Figure 2F).

## Nucleolin specifically recognizes and stabilizes the LTR G-quadruplexes

With the goal of identifying the protein of interest, the relevant EMSA band was excised from the gel and submitted directly to trypsin digestion (see Materials and Methods). Alternatively, the material extracted from the EMSA band, which could possibly include multiple co-migrating species, was further purified on a SDS gel before trypsin digestion. In either case, the digestion products were subjected to MS/MS analysis and database searching (Figure 3A) (40,41), which provided an excellent match with human nucleolin (NCL) (Table 1). The experiment was separately repeated three times to corroborate the results. Positive identification was also confirmed by performing EMSA analysis of samples that included the G4-folded wt and mutant LTR-II+III+IV sequences with either nuclear extracts or purified human NCL. NCL displayed the same binding activity towards the wt and mutant LTR sequences manifested by the unknown protein (Figure 3B). NCL binding to the wt LTR-II+III+IV G4 was concentration dependent (Figure 3C and D). Surprisingly, however, two NCL/LTR complex bands were identified (Figure 3B and C). It has been reported that both native and purified NCL display self-cleaving activity and indeed preparations of NCL usually exhibit multiple bands (42). Our purified NCL also showed multiple bands on SDS gel, as detected by both Coomassie staining and western blot analysis (Supplementary Figure S1A). MS analysis of the four bands detected by Coomassie staining showed full-length peptide coverage only for the upper major band, whereas coverage of the N-terminal portion was missing in the lower three bands (Supplementary Figure S1B). We therefore ascribed the upper and lower bands to the full-length and cleaved forms of NCL, respectively. Interestingly, the band obtained from nuclear extracts migrated the fastest and, indeed, the N-terminal portion was not observed by MS peptide analysis (Table 1 and Supplementary Figure S1B). In addition, when the protein bound to LTR G4 in the EMSA gel was extracted and analyzed separately on SDS gel, it migrated at around 75 kDa, corresponding to one of the cleaved forms of NCL (data not shown). Therefore, we conclude that the NCL form present in the nuclear extracts corresponds to its cleaved portion lacking the N-terminus. In addition, NCL formed a complex with a scrambled sequence, which displayed a slightly slower migration rate compared to the G4-bound NCL (Figure 3B). The amount of complex was similar to that afforded by M4 and M5 LTR sequences. Binding of NCL to G-rich oligonucleotides has been previously reported (43) and provides an excellent indication that the RNA binding domains (RBD) of the purified NCL are indeed active. In analogous fashion, biotinylated wt, M4+5 LTR-II+III+IV, and scrambled-sequence oligonucleotides were incubated with nuclear extracts and then added to streptavidin-functionalized agarose beads to facilitate removal of unbound proteins. Immobilized proteins were eluted with buffers of increasing ionic strength and identified by western blot analysis (Figure 3E). NCL

was released from the wt sequence when treated with increasing concentrations of NaCl or heated to 95°C in denaturing buffer, whereas it was released from the M4+5 and scrambled sequences at the lowest NaCl concentration, consistent with a lower affinity for these non-G4-forming sequences. These data confirm that NCL specifically binds the G4 folded conformation of the LTR-II+III+IV sequence.

Taq polymerase stop assays were performed to assess the stabilization imparted by NCL to the various LTR G4 structures. The wt LTR-II+III+IV and M4+5 mutant sequences were extended to include a primer-annealing region and used as templates for a single-cycle Taq reaction (Supplementary Table S1). Elongation of the wt template was performed at 37°C or 47°C in the presence/absence of NCL. In samples containing 100 mM KCl and no NCL, a pausing site corresponding to the most 3′-end G-tract was observed only in the reaction elongated at 37°C. The pause was not observed at 47°C, consistent with possible destabilizing effects of temperature on the G4 structure (compare lanes 2 and 5 of Figure 4A). Addition of NCL induced more evident pauses at both elongation temperatures (Figure 4A, lanes 3 and 6), which clearly highlighted the stabilizing properties of NCL on G4 conformation. As expected from the inability of the M4+5 mutant to fold G4 structures, no pausing sites were observed regardless the presence of protein (Figure 4A, lanes 7–9), thus confirming the specificity of NCL binding for full-fledged G4s. At the same time, FRET melting assays were also carried out to study the stabilizing effects of NCL on the G4 structures. The assays involved the synthesis of constructs that combined selected G4-forming sequences, such as LTR-II+III+IV and the shorter LTR-II, LTR-III, LTR-IV, and LTR-III+IV (Figure 2A), with *FAM* and *TAMRA* moieties placed at their 5′- and 3′-ends, respectively. The results showed that NCL conferred the highest stabilization in the series to the LTR-II+III+IV construct, followed by LTR-III+IV. Progressively lower stabilization was observed for LTR-III and LTR-II, whereas LTR-IV was the least affected in the series (Table 2). The negative control bovine serum albumin (BSA) did not afford any detectable stabilization to the selected sequences.

To identify the position of putative NCL binding sites onto the DNA G4 structures, we performed dimethylsulfate (DMS) footprinting. When a complex of NCL with the LTR-II+III+IV DNA was assessed (Figure 4B), the methylation pattern revealed that the 3′ region of the LTR sequence was protected by protein binding with unique specificity for two G bases in G-tract 4, as highlighted in Figure 4C.

Based on the facts that NCL has been described as a RNA-binding protein (44) and that the LTR sequence is present in the U3 region of the HIV-1 genome during the initial infection steps, we tested whether NCL could also bind the RNA version of the LTR G4. In this case, NCL was incubated with labeled RNA or DNA oligonucleotides capable of folding LTR G4s. Increasing concentrations of unlabelled RNA and DNA counterparts were employed to compete for NCL binding. The results clearly showed that the protein was able to bind both types of G4 oligonucleotides (lanes 1 and 6, Figure 5A). However, the DNA G4 was consistently able to outcompete the RNA version
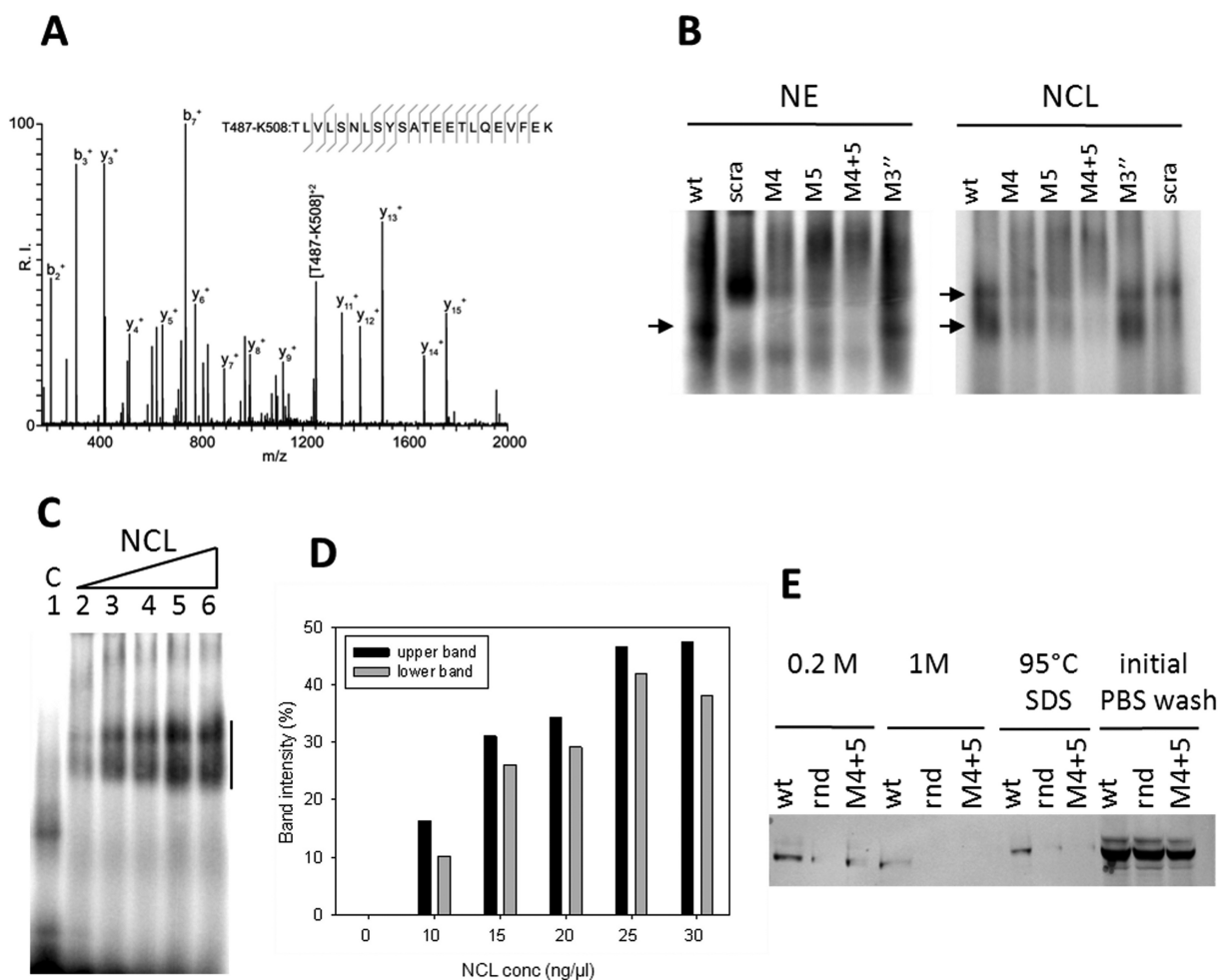
Figure 3. Identification of the G4-specific LTR binding protein. (**A**) MS/MS spectrum of the precursor ion observed at *m/z* 1251.14 in the sample mixture from native gel 1 (see Table 1). Only characteristic b and y ions are indicated (40,41). The data matched the sequence of peptide T487-K508 of NCL, which is reported on top with the observed fragments. (**B**) EMSA analysis of the binding of nuclear extract (NE) proteins and purified NCL to the wt and mutant LTR sequences. Arrows indicate relevant protein/LTR G4 complex bands. (**C**) EMSA analysis of the binding of increasing amounts of purified NCL to the wt LTR-II+III+IV G4. The vertical bar highlights the portion of the gel where the two NCL/LTR G4 complex bands are observed. (**D**) Quantification of the upper and lower NCL/LTR G4 complex bands obtained in the EMSA in panel (C). (**E**) Pull-down assay of nuclear extract proteins with wt, mutant G4 LTR-II+III+IV and random (rnd) sequences, immobilized on agarose beads. Shown is the western blot analysis with an NCL antibody. Proteins complexed to the beads-bound LTRs were washed with augmented stringency by increasing the ionic strength of the wash buffer (0.2 and 1 M). The final elution was obtained in denaturing buffer at 95°C.

Table 1. MS analysis of protein content of three gel bands collected from one SDS gel and two native gels

Sample	Experimental monoisotopic mass	Expected monoisotopic mass	Protein match	Peptide match	Score
SDS gel	2199.02	2199.02	Nucleolin	(G578-R597)	22
	1647.73	1647.73	Nucleolin	(F349-K362)	63
	1560.67	1560.67	Nucleolin	(G611-K624)	68
Native gel 1	4856.25	4856.26	Translationally-controlled tumor protein	(T39-K85)	22
	2500.26	2500.26	Nucleolin	(T487-K508)	102
	2311.14	2311.15	Nucleolin	(V298-K318)	84
	1647.73	1647.73	Nucleolin	(F349-K362)	71
Native gel 2	2500.26	2500.26	Nucleolin	(T487-K508)	93
	2311.14	2311.15	Nucleolin	(V298-K318)	62
	2199.02	2199.02	Nucleolin	(G578-R597)	30
	1560.67	1560.67	Nucleolin	(G611-K624)	60

Table 2. FRET analysis of the stabilization of NCL on the different length G4 LTR sequences

LTR sequence	T_m LTR (°C)	T_m LTR + NCL (°C)	ΔT_m (°C)
LTR-II	58.1 ± 0.2	76.2 ± 1.4	18.1
LTR-III	63.1 ± 0.1	75.1 ± 0.7	12.0
LTR-IV	62.1 ± 0.6	69.0 ± 1.4	6.9
LTR-III+IV	54.0 ± 0.6	78.5 ± 2.5	24.5
LTR-II+III+IV	48.0 ± 0.6	86.9 ± 1.4	38.9
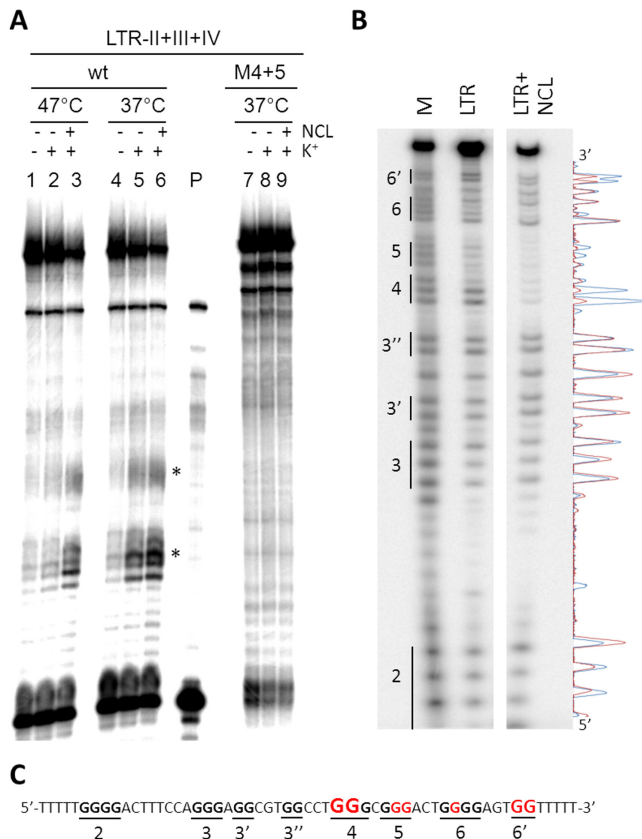
Figure 4. Characterization of NCL binding to the G4 LTR-II+III+IV. (**A**) Taq polymerase stop assay. Taq polymerization was performed in the presence/absence of K+ and NCL, as indicated, on the wt and M4+5 LTR-II+III+IV sequences. Amplification of the wt template was performed at 37°C and 47°C; on the mutant template elongation was obtained at 37°C. Stop regions are highlighted by the * symbol. (**B**) DMS protection analysis of the NCL/LTR complex and the G4 LTR-II+III+IV oligonucleotide. Band cleavage intensity is quantified by a densitogram shown on the right: red and blue lines correspond to the LTR and NCL/LTR complex, respectively. G-tracts and their numbering are indicated on the left. The two gel portions derive from a single gel run. (**C**) The LTR-II+III+IV sequence: bases protected by the DMS alkylation/cleavage are shown in red. The font of the red letters is proportional to the protection. G-tracts are numbered.

for protein binding (lanes 7–10, Figure 5A and B), whereas the RNA was incapable of outcompeting the DNA (lanes 2–5, Figure 5A and B). We had previously shown that the LTR RNA G4s fold into parallel structures (33), whereas the DNA counterparts display rather hybrid-like conformations (30); therefore, the preferential binding toward the DNA G4 is likely caused by these substantial structural differences between the DNA and RNA G4 conformations.

Surface plasmon resonance (SPR) analysis was next used to assess the affinity of the purified NCL for the wt LTR-

II+III+IV G4: a K_D of 2.5 ± 0.1 nM was obtained, which indicates an extremely high affinity of the protein for this LTR G4 (Figure 5C). In contrast, the scrambled oligonucleotide showed no affinity for NCL (Supplementary Figure S1).

## Binding of nucleolin to the LTR G-quadruplexes increases repression of viral transcription

A luciferase reporter assay was established to explore the downstream biological effects of NCL binding to the LTR promoter. Two epithelial breast cell lines were selected for the assay: MCF-7 breast cancer cells and MCF-10A normal breast epithelial cells. The latter inherently express lower amounts of NCL compared to tumor cells (45,46). For this reason, the MCF-10A cells were transfected with wt or M4+5 LTR luciferase reporter plasmids, either alone or in the presence of increasing amounts of NCL expression vector. The luciferase signal was measured to determine the level of activation of the LTR promoter. The results showed that the activity of wt LTR decreased to 65% of the control, while that of the M4+5 promoter remained unvaried (Figure 6A). The MCF7 cell line that overexpresses NCL was employed to perform additional activity assays. In this case, cells were treated with increasing amounts of siRNAs designed to target NCL mRNA (47), and then transfected with wt or mutant M4+5 LTR luciferase reporter plasmids. Analysis of NCL content showed that the protein was effectively depleted at 1–4 nM siRNA, reaching 8% of the initial amount at 4 nM (Figure 6B). Measured by luciferase activity, the effect of NCL depletion on LTR promoter activity was quite astonishing: at 4 nM of siRNA, the promoter activity of the wt sequence was 37 folds that of the control while the mutant M4+5 sequence was only 1.5 folds (Figure 6C). As a complementary approach, we tested the effect of the NCL-targeted DNA aptamer AS1411, which has been reported to bind with high affinity NCL in cells (48). At 1 µM of AS1411, we again observed a significant increment of the wt LTR promoter activity to reach 11 times that of the non-treated control while the mutant M4+5 sequence increased of only 2.5 times (Figure 6D). In contrast, the control sequence CRO26 that is complementary to AS1411 did not modify LTR promoter activity. These data indicate that the specific binding of NCL to G4 structures in the LTR promoter exerts significant repressive effects on HIV-1 promoter activity.

DISCUSSION

We have identified NCL as a prominent host factor capable of binding with high affinity to the G4 structures present in the LTR promoter of HIV-1. We observed that the specific interaction leads to G4 stabilization and contributes
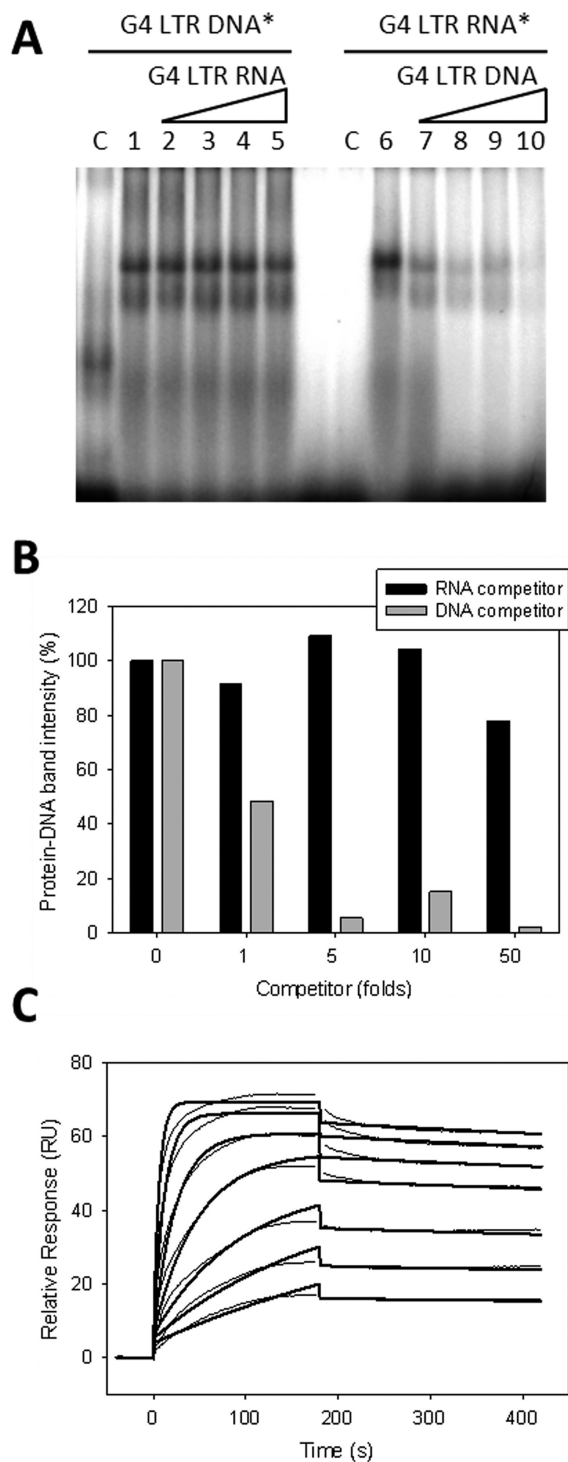
Figure 5. Properties of NCL binding to the G4 LTR-II+III+IV sequence. (**A**) EMSA competition analysis of NCL binding towards the G4 LTR-II+III+IV DNA and RNA sequences. On the left: the labeled G4 LTR DNA* was bound to NCL and competed with increasing folds of unlabelled G4 LTR RNA (1–50 folds); on the right, the labeled G4 LTR RNA* was bound to NCL and competed with increasing folds of unlabelled G4 LTR DNA (1–50 folds). C is a control lane for the oligonucleotides before binding. (**B**) NCL/LTR complex band quantification of the experiment presented in (A). (**C**) SPR binding analysis of wt LTR-II+III+IV to immobilized NCL. Oligonucleotide concentration range was 31.25 nM-2000 nM. Sensograms are shown as gray lines and their respective fits as black lines.
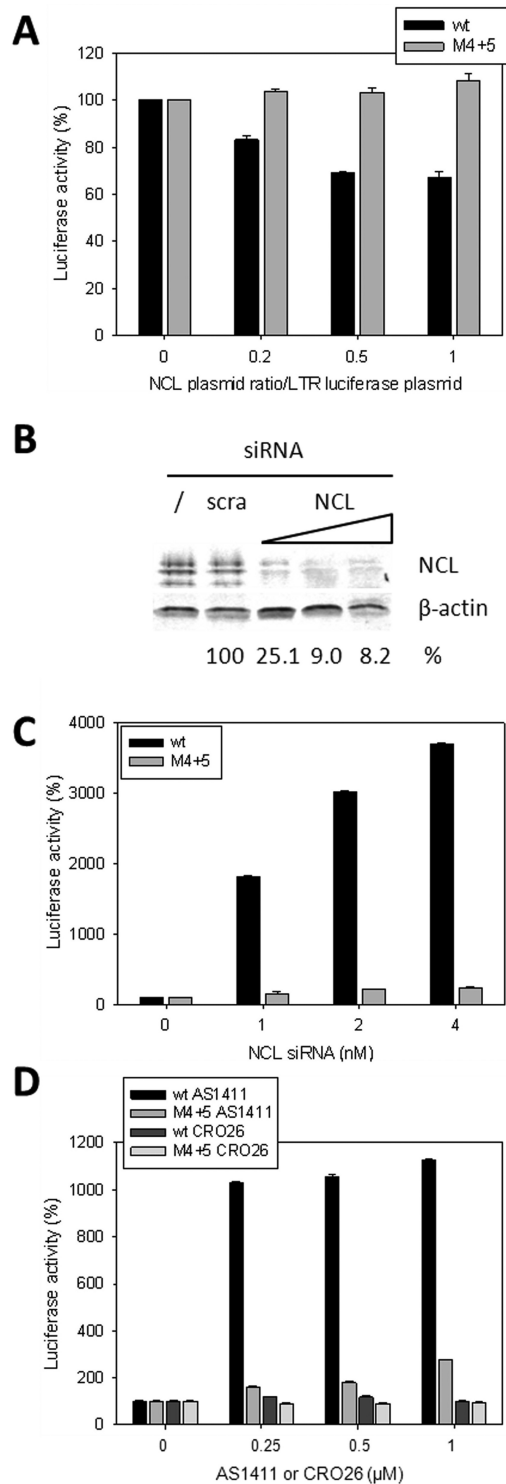
Figure 6. Activity of NCL on the HIV-1 LTR promoter in cells. (**A**) Analysis of the luciferase activity of the wt and M4+5 LTR promoters in MCF-10A cells transiently transfected with the LTR luciferase plasmids and a NCL expression plasmid. (**B**) NCL depletion in MCF7 cells by NCL siR-NAs analysed by western blot with NCL antibody. Scra indicates scrambled siRNAs. Detection of β-actin was used as control. The % of NCL expression in each sample is given below the gel. (**C**) Analysis of the luciferase activity of the wt and M4+5 LTR promoters in MCF7 cells treated with NCL siRNAs. (**D**) Analysis of the luciferase activity of the wt and M4+5 LTR promoters in MCF7 cells treated with the NCL-binding aptamer AS1411 and the negative control CRO26.

to silence viral transcription. Conversely, we also demonstrated that NCL depletion produces extraordinary enhancing effects on LTR promoter activity. These observations are consistent with the multifaceted nucleic acid binding and chaperoning activities attributed to this protein. Indeed, NCL is most abundant in the nucleolus, but can be found also in cell membranes and, upon stress stimuli, in the nucleoplasm and cytoplasm, to some extent (49,50). Among other functions, it is involved in transcription (50) by specific interactions with sequences that can adopt complex secondary structures (44,51–52). It is widely believed that NCL plays a chaperone role by helping the correct folding of complex nucleic acids structures. Indeed, NCL has been shown to display a marked preference for both endogenous and exogenous G-rich sequences that can fold into G4 (53). It has been recently reported that binding of NCL to the endogenous $(GGGGCC)_n$ hexanucleotide repeat expansion (HRE) in *C9orf72* is responsible for the initiation of molecular cascades that lead to neurodegenerative diseases (19). At the promoter level, binding of NCL to G4 structures augments the basal effect of the folded conformation (45,54–56). One of the best documented example of G4-mediated regulation among G4 promoters (57) is that of c-myc, which shows striking similarities with the G4-mediated regulation of the HIV-1 LTR promoter reported here and previously by us (30). Both cases involve multiple G-tracts that enable folding into alternative G4 conformations; G4 parallel-like topology (58); at least one $G_3N_1G_3$ motif; binding sites for SP1; silencing effect on promoter activity (59); NCL-binding activity and higher affinity of NCL towards the DNA G4 compared to the RNA G4 counterpart (45). In the case of c-myc, it has been shown that the N-terminal of NCL is dispensable for its G4 binding activity (56). Here, we show that NCL can naturally produce cleaved forms that lack the N-terminal but retain full binding capabilities. These observations demonstrate that cells allow the formation of NCL cleaved species that maintain their G4/nucleic acid binding activity. On the other, they suggest that the HIV-1 virus and human host cells have likely evolved identical mechanisms to control transcription at the DNA promoter level.

The results of our experiments provided valuable insights into the determinants of NCL binding to the various LTR G4-forming structures. The lower binding of NCL to LTR-IV compared to the other G4 structures suggests that the interaction has both conformation- and sequence-dependent characteristics, which in turn implies the fascinating possibility that different G4s in the HIV-1 LTR promoter may exert different functions based on their binding partners. Moreover, the greater affinity demonstrated for DNA than RNA constructs indicates that the interaction is deeply influenced by conformational differences between G4 structures folded by the different types of biopolymers. The facts that the HIV-1 genome consists of RNA, that this G4-forming sequence is also present in the U3 region of the genome (33), and that viral RNA during the first steps of infection is still present in the cell cytoplasm where NCL levels are low, suggest that the specific binding of NCL to the DNA version may represent an essential mechanism for regulating viral transcription. In this direction, it has been shown that NCL is involved in different steps of

the HIV-1 life cycle. Inhibition of surface NCL by different cellular and synthetic compounds (60–62) affects cell attachment/entry by the virus (63). In addition, NCL can bind HIV-1 Gag protein to promote viral budding (64), or to enhance Gag release (65). Further, NCL involvement in the viral life cycle has been corroborated also by evidence that HIV infection modifies the protein's cellular distribution (66,67). The activities performed by NCL in other viruses have also been described in a number of recent papers (68–72). For example, binding of NCL to Epstein–Barr virus (EBV) nuclear antigen 1 (EBNA1) modulates viral replication and transcription (73), and virus-induced relocalization of NCL has been observed in some instances (74,75). In this broader context, these reports support our new findings that point toward a significant role played by this protein in HIV-1 replication.

The observation that NCL interaction with LTR G4 silences viral transcription is in apparent contrast with the well-known ability of NCL to interact with histone H1, which induces chromatin decondensation (76) that in turn facilitates the passage of the DNA polymerases (77). However, the opposite effect consisting of NCL-mediated repression of DNA replication has been also reported (49,78–80), and has been attributed to the interaction of NCL with the DNA processing enzyme Replication Protein A (RPA). In addition, NCL has been reported to recruit a DNA helicase that unwinds G4 structures (81). In both cases, NCL binding to cellular proteins is a transient response to stress stimuli. These observations prompt the intriguing possibility that when NCL is redistributed by the cellular stress imposed by HIV infection (66,67), it may be then recruited by the G4 structure of viral promoter to transiently downregulate HIV transcription and enable the virus to prepare for subsequent efficient transcription, when viral proteins, such as Tat, take over. Alternatively, this interaction may be required as a first switch to viral latency and to recruit proteins that further consolidate latency. The signals that trigger latency are not known at this point and are thus the object of intense studies. However, factors that repress viral transcription at the LTR promoter have been proposed to play a determinant role in latency mechanisms (82,83). Finally, the very specific nature of nucleolin binding indicates that its viral target must be less prone to mutations. This observation makes the viral G4/nucleolin complex into a very appealing target for the development of antiviral strategies that may afford a different mechanism of action, the possibility of targeting viral latency, and a lower probability of incurring into drug resistance. In conclusion, we have shown that the specific binding of the cellular protein NCL to the LTR promoter regulates viral transcription. This result alone paves the way for the investigation of different regulation mechanisms of HIV-1 transcription/latency, which may lead to new possible targets for the design of specific inhibitors.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Lipps,H.J. and Rhodes,D. (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.*, **19**, 414–422.
2. Sen,D. and Gilbert,W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364–366.
3. Wu,Y. and Brosh,R.M. Jr (2010) G-quadruplex nucleic acids and human disease. *FEBS J.*, **277**, 3470–3488.
4. Sen,D. and Gilbert,W. (1990) A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature*, **344**, 410–414.
5. Taylor,J.P. (2014) Neurodegenerative diseases: G-quadruplex poses quadruple threat. *Nature*, **507**, 175–177.
6. Campbell,N.H. and Neidle,S. (2012) G-quadruplexes and metal ions. *Met. Ions Life Sci.*, **10**, 119–134.
7. Eddy,J. and Maizels,N. (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.*, **34**, 3887–3896.
8. Huppert,J.L. and Balasubramanian,S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406–413.
9. Nakken,S., Rognes,T. and Hovig,E. (2009) The disruptive positions in human G-quadruplex motifs are less polymorphic and more conserved than their neutral counterparts. *Nucleic Acids Res.*, **37**, 5749–5756.
10. Cogoi,S. and Xodo,L.E. (2006) G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.*, **34**, 2536–2549.
11. Eddy,J. and Maizels,N. (2008) Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic acids Res.*, **36**, 1321–1333.
12. Verma,A., Yadav,V.K., Basundra,R., Kumar,A. and Chowdhury,S. (2009) Evidence of genome-wide G4 DNA-mediated gene expression in human cancer cells. *Nucleic Acids Res.*, **37**, 4194–4204.
13. Mani,P., Yadav,V.K., Das,S.K. and Chowdhury,S. (2009) Genome-wide analyses of recombination prone regions predict role of DNA structural motif in recombination. *PLoS One*, **4**, e4399.
14. Beaudoin,J.D. and Perreault,J.P. (2013) Exploring mRNA 3′-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. *Nucleic Acids Res.*, **41**, 5898–5911.
15. Bugaut,A. and Balasubramanian,S. (2012) 5′-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.*, **40**, 4727–4741.
16. Sissi,C., Gatto,B. and Palumbo,M. (2011) The evolving world of protein-G-quadruplex recognition: a medicinal chemist's perspective. *Biochimie*, **93**, 1219–1230.
17. Cogoi,S., Paramasivam,M., Membrino,A., Yokoyama,K.K. and Xodo,L.E. (2010) The KRAS promoter responds to Myc-associated zinc finger and poly(ADP-ribose) polymerase 1 proteins, which recognize a critical quadruplex-forming GA-element. *J. Biol. Chem.*, **285**, 22003–22016.
18. Cogoi,S., Shchekotikhin,A.E. and Xodo,L.E. (2014) HRAS is silenced by two neighboring G-quadruplexes and activated by MAZ, a zinc-finger transcription factor with DNA unfolding property. *Nucleic Acids Res.*, **42**, 8379–8388.
19. Haeusler,A.R., Donnelly,C.J., Periz,G., Simko,E.A., Shaw,P.G., Kim,M.S., Maragakis,N.J., Troncoso,J.C., Pandey,A., Sattler,R. *et al.* (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature*, **507**, 195–200.
20. Ivanov,P., O'Day,E., Emara,M.M., Wagner,G., Lieberman,J. and Anderson,P. (2014) G-quadruplex structures contribute to the neuroprotective effects of angiogenin-induced tRNA fragments. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 18201–18206.
21. Sket,P., Pohleven,J., Kovanda,A., Stalekar,M., Zupunski,V., Zalar,M., Plavec,J. and Rogelj,B. (2015) Characterization of DNA G-quadruplex species forming from C9ORF72 GC-expanded repeats associated with amyotrophic lateral sclerosis and frontotemporal lobar degeneration. *Neurobiol. Aging*, **36**, 191–196.
22. Fisette,J.F., Montagna,D.R., Mihailescu,M.R. and Wolfe,M.S. (2012) A G-rich element forms a G-quadruplex and regulates BACE1 mRNA alternative splicing. *J. Neurochem.*, **121**, 763–773.
23. Fry,M. and Loeb,L.A. (1994) The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 4950–4954.
24. Fry,M. and Loeb,L.A. (1999) Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n. *J. Biol. Chem.*, **274**, 12797–12802.
25. Wu,Y., Shin-ya,K. and Brosh,R.M. Jr (2008) FANCJ helicase defective in Fanconia anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Mol. Cell. Biol.*, **28**, 4116–4128.
26. Beaume,N., Pathak,R., Yadav,V.K., Kota,S., Misra,H.S., Gautam,H.K. and Chowdhury,S. (2013) Genome-wide study predicts promoter-G4 DNA motifs regulate selective functions in bacteria: radioresistance of D. radiodurans involves G4 DNA-mediated regulation. *Nucleic Acids Res.*, **41**, 76–89.
27. Du,X., Wojtowicz,D., Bowers,A.A., Levens,D., Benham,C.J. and Przytycka,T.M. (2013) The genome-wide distribution of non-B DNA motifs is shaped by operon structure and suggests the transcriptional importance of non-B DNA structures in Escherichia coli. *Nucleic Acids Res.*, **41**, 5965–5977.
28. Metifiot,M., Amrane,S., Litvak,S. and Andreola,M.L. (2014) G-quadruplexes in viruses: function and potential therapeutic applications. *Nucleic Acids Res.*, **42**, 12352–12366.
29. Perrone,R., Nadai,M., Poe,J.A., Frasson,I., Palumbo,M., Palu,G., Smithgall,T.E. and Richter,S.N. (2013) Formation of a unique cluster of G-quadruplex structures in the HIV-1 Nef coding region: implications for antiviral activity. *PLoS One*, **8**, e73121.
30. Perrone,R., Nadai,M., Frasson,I., Poe,J.A., Butovskaya,E., Smithgall,T.E., Palumbo,M., Palu,G. and Richter,S.N. (2013) A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter. *J. Med. Chem.*, **56**, 6521–6530.
31. Amrane,S., Kerkour,A., Bedrat,A., Vialet,B., Andreola,M.L. and Mergny,J.L. (2014) Topology of a DNA G-quadruplex structure formed in the HIV-1 promoter: a potential target for anti-HIV drug development. *J. Am. Chem. Soc.*, **136**, 5249–5252.
32. Piekna-Przybylska,D., Sullivan,M.A., Sharma,G. and Bambara,R.A. (2014) U3 region in the HIV-1 genome adopts a G-quadruplex structure in its RNA and DNA sequence. *Biochemistry*, **53**, 2581–2593.
33. Perrone,R., Butovskaya,E., Daelemans,D., Palu,G., Pannecouque,C. and Richter,S.N. (2014) Anti-HIV-1 activity of the G-quadruplex ligand BRACO-19. *J. Antimicrob. Chemother.*, **69**, 3248–3258.
34. Murat,P., Zhong,J., Lekieffre,L., Cowieson,N.P., Clancy,J.L., Preiss,T., Balasubramanian,S., Khanna,R. and Tellam,J. (2014) G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. *Nat. Chem. Biol.*, **10**, 358–364.
35. Norseen,J., Johnson,F.B. and Lieberman,P.M. (2009) Role for G-quadruplex RNA binding by Epstein-Barr virus nuclear antigen 1 in DNA replication and metaphase chromosome attachment. *J. Virol.*, **83**, 10336–10346.
36. Tan,J., Vonrhein,C., Smart,O.S., Bricogne,G., Bollati,M., Kusov,Y., Hansen,G., Mesters,J.R., Schmidt,C.L. and Hilgenfeld,R. (2009) The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes. *PLoS Pathogens*, **5**, e1000428.
37. Adachi,A., Gendelman,H.E., Koenig,S., Folks,T., Willey,R., Rabson,A. and Martin,M.A. (1986) Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. *J. Virol.*, **59**, 284–291.
38. Perkins,D.N., Pappin,D.J., Creasy,D.M. and Cottrell,J.S. (1999) Probability-based protein identification by searching sequence

databases using mass spectrometry data. *Electrophoresis*, **20**, 3551–3567.

39. Andrews,N.C. and Faller,D.V. (1991) A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucleic Acids Res.*, **19**, 2499.

40. Roepstorff,P. and Fohlman,J. (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.*, **11**, 601.

41. Biemann,K. (1988) Contributions of mass spectrometry to peptide and protein structure. *Biomed. Environ. Mass Spectrom.*, **16**, 99–111.

42. Chen,C.M., Chiang,S.Y. and Yeh,N.H. (1991) Increased stability of nucleolin in proliferating cells by inhibition of its self-cleaving activity. *J. Biol. Chem.*, **266**, 7754–7758.

43. Sapp,M., Knippers,R. and Richter,A. (1986) DNA binding properties of a 110 kDa nucleolar protein. *Nucleic Acids Res.*, **14**, 6803–6820.

44. Serin,G., Joseph,G., Faucher,C., Ghisolfi,L., Bouche,G., Amalric,F. and Bouvet,P. (1996) Localization of nucleolin binding sites on human and mouse pre-ribosomal RNA. *Biochimie*, **78**, 530–538.

45. Gonzalez,V., Guo,K., Hurley,L. and Sun,D. (2009) Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J. Biol. Chem.*, **284**, 23622–23635.

46. Soundararajan,S., Chen,W., Spicer,E.K., Courtenay-Luck,N. and Fernandes,D.J. (2008) The nucleolin targeting aptamer AS1411 destabilizes Bcl-2 messenger RNA in human breast cancer cells. *Cancer Res.*, **68**, 2358–2365.

47. Watanabe,T., Takahashi,A., Suzuki,K., Kurusu-Kanno,M., Yamaguchi,K., Fujiki,H. and Suganuma,M. (2014) Epithelial-mesenchymal transition in human gastric cancer cell lines induced by TNF-alpha-inducing protein of Helicobacter pylori. *Int. J. Cancer*, **134**, 2373–2382.

48. Bates,P.J., Kahlon,J.B., Thomas,S.D., Trent,J.O. and Miller,D.M. (1999) Antiproliferative activity of G-rich oligonucleotides correlates with protein binding. *J. Biol. Chem.*, **274**, 26369–26377.

49. Daniely,Y., Dimitrova,D.D. and Borowiec,J.A. (2002) Stress-dependent nucleolin mobilization mediated by p53-nucleolin complex formation. *Mol. Cell. Biol.*, **22**, 6014–6022.

50. Storck,S., Shukla,M., Dimitrov,S. and Bouvet,P. (2007) Functions of the histone chaperone nucleolin in diseases. *Sub-cell. Biochem.*, **41**, 125–144.

51. Serin,G., Joseph,G., Ghisolfi,L., Bauzan,M., Erard,M., Amalric,F. and Bouvet,P. (1997) Two RNA-binding domains determine the RNA-binding specificity of nucleolin. *J. Biol. Chem.*, **272**, 13109–13116.

52. Allain,F.H., Bouvet,P., Dieckmann,T. and Feigon,J. (2000) Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J.*, **19**, 6870–6881.

53. Girvan,A.C., Teng,Y., Casson,L.K., Thomas,S.D., Juliger,S., Ball,M.W., Klein,J.B., Pierce,W.M. Jr, Barve,S.S. and Bates,P.J. (2006) AGRO100 inhibits activation of nuclear factor-kappaB (NF-kappaB) by forming a complex with NF-kappaB essential modulator (NEMO) and nucleolin. *Mol. Cancer Ther.*, **5**, 1790–1799.

54. 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.

55. Wei,P.C., Wang,Z.F., Lo,W.T., Su,M.I., Shew,J.Y., Chang,T.C. and Lee,W.H. (2013) A cis-element with mixed G-quadruplex structure of NPGPx promoter is essential for nucleolin-mediated transactivation on non-targeting siRNA stress. *Nucleic Acids Res.*, **41**, 1533–1543.

56. Gonzalez,V. and Hurley,L.H. (2010) The C-terminus of nucleolin promotes the formation of the c-MYC G-quadruplex and inhibits c-MYC promoter activity. *Biochemistry*, **49**, 9706–9714.

57. Brooks,T.A. and Hurley,L.H. (2010) Targeting MYC Expression through G-Quadruplexes. *Genes Cancer*, **1**, 641–649.

58. Qin,Y. and Hurley,L.H. (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie*, **90**, 1149–1171.

59. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 11593–11598.

60. Said,E.A., Courty,J., Svab,J., Delbe,J., Krust,B. and Hovanessian,A.G. (2005) Pleiotrophin inhibits HIV infection by binding the cell surface-expressed nucleolin. *FEBS J.*, **272**, 4646–4659.

61. Hovanessian,A.G. (2006) Midkine, a cytokine that inhibits HIV infection by binding to the cell surface expressed nucleolin. *Cell Res.*, **16**, 174–181.

62. Nisole,S., Said,E.A., Mische,C., Prevost,M.C., Krust,B., Bouvet,P., Bianco,A., Briand,J.P. and Hovanessian,A.G. (2002) The anti-HIV pentameric pseudopeptide HB-19 binds the C-terminal end of nucleolin and prevents anchorage of virus particles in the plasma membrane of target cells. *J. Biol. Chem.*, **277**, 20877–20886.

63. Callebaut,C., Blanco,J., Benkirane,N., Krust,B., Jacotot,E., Guichard,G., Seddiki,N., Svab,J., Dam,E., Muller,S. *et al.* (1998) Identification of V3 loop-binding proteins as potential receptors implicated in the binding of HIV particles to CD4(+) cells. *J. Biol. Chem.*, **273**, 21988–21997.

64. Ueno,T., Tokunaga,K., Sawa,H., Maeda,M., Chiba,J., Kojima,A., Hasegawa,H., Shoya,Y., Sata,T., Kurata,T. *et al.* (2004) Nucleolin and the packaging signal, psi, promote the budding of human immunodeficiency virus type-1 (HIV-1). *Microbiol. Immunol.*, **48**, 111–118.

65. Gao,W., Li,M. and Zhang,J. (2014) Tandem immunoprecipitation approach to identify HIV-1 Gag associated host factors. *J. Virol. Methods*, **203**, 116–119.

66. Galati,D., Paiardini,M., Cervasi,B., Albrecht,H., Bocchino,M., Costantini,A., Montroni,M., Magnani,G. and Silvestri,G. (2003) Specific changes in the posttranslational regulation of nucleolin in lymphocytes from patients infected with human immunodeficiency virus. *J. Infect. Dis.*, **188**, 1483–1491.

67. Visalli,G., Paiardini,M., Chirico,C., Cervasi,B., Curro,M., Ferlazzo,N., Bertuccio,M.P., Favaloro,A., Pellicano,G., Sturniolo,G. *et al.* (2010) Intracellular accumulation of cell cycle regulatory proteins and nucleolin re-localization are associated with pre-lethal ultrastructural lesions in circulating T lymphocytes: the HIV-induced cell cycle dysregulation revisited. *Cell Cycle*, **9**, 2130–2140.

68. Tayyari,F., Marchant,D., Moraes,T.J., Duan,W., Mastrangelo,P. and Hegele,R.G. (2011) Identification of nucleolin as a cellular receptor for human respiratory syncytial virus. *Nat. Med.*, **17**, 1132–1135.

69. Su,P.Y., Wang,Y.F., Huang,S.W., Lo,Y.C., Wang,Y.H., Wu,S.R., Shieh,D.B., Chen,S.H., Wang,J.R., Lai,M. *et al.* (2015) Cell surface nucleolin facilitates enterovirus 71 binding and infection. *J. Virol.*, **89**, 4527–4538.

70. Strang,B.L., Boulant,S., Kirchhausen,T. and Coen,D.M. (2012) Host cell nucleolin is required to maintain the architecture of human cytomegalovirus replication compartments. *mBio*, **3**, doi:10.1128/mBio.00301-11.

71. Bender,B.J., Coen,D.M. and Strang,B.L. (2014) Dynamic and nucleolin-dependent localization of human cytomegalovirus UL84 to the periphery of viral replication compartments and nucleoli. *J. Virol.*, **88**, 11738–11747.

72. Balinsky,C.A., Schmeisser,H., Ganesan,S., Singh,K., Pierson,T.C. and Zoon,K.C. (2013) Nucleolin interacts with the dengue virus capsid protein and plays a role in formation of infectious virus particles. *J. Virol.*, **87**, 13094–13106.

73. Chen,Y.L., Liu,C.D., Cheng,C.P., Zhao,B., Hsu,H.J., Shen,C.L., Chiu,S.J., Kieff,E. and Peng,C.W. (2014) Nucleolin is important for Epstein-Barr virus nuclear antigen 1-mediated episome binding, maintenance, and transcription. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 243–248.

74. Cancio-Lonches,C., Yocupicio-Monroy,M., Sandoval-Jaime,C., Galvan-Mendoza,I., Urena,L., Vashist,S., Goodfellow,I., Salas-Benito,J. and Gutierrez-Escolano,A.L. (2011) Nucleolin interacts with the feline calicivirus 3′ untranslated region and the protease-polymerase NS6 and NS7 proteins, playing a role in virus replication. *J. Virol.*, **85**, 8056–8068.

75. Greco,A., Arata,L., Soler,E., Gaume,X., Coute,Y., Hacot,S., Calle,A., Monier,K., Epstein,A.L., Sanchez,J.C. *et al.* (2012) Nucleolin interacts with US11 protein of herpes simplex virus 1 and is involved in its trafficking. *J. Virol.*, **86**, 1449–1457.

76. Erard,M.S., Belenguer,P., Caizergues-Ferrer,M., Pantaloni,A. and Amalric,F. (1988) A major nucleolar protein, nucleolin, induces chromatin decondensation by binding to histone H1. *Eur. J. Biochem./FEBS*, **175**, 525–530.

77. Angelov,D., Bondarenko,V.A., Almagro,S., Menoni,H., Mongelard,F., Hans,F., Mietton,F., Studitsky,V.M., Hamiche,A., Dimitrov,S. *et al.* (2006) Nucleolin is a histone chaperone with FACT-like activity and assists remodeling of nucleosomes. *EMBO J.*, **25**, 1669–1679.

78. Daniely,Y. and Borowiec,J.A. (2000) Formation of a complex between nucleolin and replication protein A after cell stress prevents initiation of DNA replication. *J. Cell Biol.*, **149**, 799–810.

79. Wang,Y., Guan,J., Wang,H., Wang,Y., Leeper,D. and Iliakis,G. (2001) Regulation of dna replication after heat shock by replication protein a-nucleolin interactions. *J. Biol. Chem.*, **276**, 20579–20588.

80. Kim,K., Dimitrova,D.D., Carta,K.M., Saxena,A., Daras,M. and Borowiec,J.A. (2005) Novel checkpoint response to genotoxic stress mediated by nucleolin-replication protein a complex formation. *Mol. Cell. Biol.*, **25**, 2463–2474.

81. Indig,F.E., Rybanska,I., Karmakar,P., Devulapalli,C., Fu,H., Carrier,F. and Bohr,V.A. (2012) Nucleolin inhibits G4 oligonucleotide unwinding by Werner helicase. *PLoS One*, **7**, e35229.

82. Duverger,A., Wolschendorf,F., Zhang,M., Wagner,F., Hatcher,B., Jones,J., Cron,R.Q., van der Sluis,R.M., Jeeninga,R.E., Berkhout,B. *et al.* (2013) An AP-1 binding site in the enhancer/core element of the HIV-1 promoter controls the ability of HIV-1 to establish latent infection. *J. Virol.*, **87**, 2264–2277.

83. van der Sluis,R.M., Jeeninga,R.E. and Berkhout,B. (2013) Establishment and molecular mechanisms of HIV-1 latency in T cells. *Curr. Opin. Virol.*, **3**, 700–706.