Published online 23 February 2011 *Nucleic Acids Research, 2011, Vol. 39, No. 12* ***5067–5081***
doi:10.1093/nar/gkr088

Nuclear protein LEDGF/p75 recognizes supercoiled DNA by a novel DNA-binding domain

Kimiko M. Tsutsui[1,*], Kuniaki Sano[1], Osamu Hosoya[1], Tadashi Miyamoto[2] and Ken Tsutsui[2,*]

[1]Department of Neurogenomics and [2]Department of Genome Dynamics, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, 2-5-1 Shikata-cho, Kita-ku, Okayama, 700-8558, Japan

Received October 15, 2010; Revised and Accepted February 3, 2011

ABSTRACT

Lens epithelium-derived growth factor (LEDGF) or p75 is a co-activator of general transcription and also involved in insertion of human immunodeficiency virus type I (HIV-1) cDNA into host cell genome, which occurs preferentially to active transcription units. These phenomena may share an underlying molecular mechanism in common. We report here that LEDGF/p75 binds negatively supercoiled DNA selectively over unconstrained DNA. We identified a novel DNA-binding domain in the protein and termed it 'supercoiled DNA-recognition domain' (SRD). Recombinant protein fragments containing SRD showed a preferential binding to supercoiled DNA *in vitro*. SRD harbors a characteristic cluster of lysine and glutamic/aspartic acid residues. A polypeptide mimicking the cluster ($K_9E_9K_9$) also showed this specificity, suggesting that the cluster is an essential element for the supercoil recognition. eGFP-tagged LEDGF/p75 expressed in the nucleus distributed partially in transcriptionally active regions that were identified by immunostaining of methylated histone H3 (H3K4me3) or incorporation of Br-UTP. This pattern of localization was observed with SRD alone but abolished if the protein lacked SRD. Thus, these results imply that LEDGF/p75 guides its binding partners, including HIV-1 integrase, to the active transcription site through recognition of negative supercoils generated around it.

INTRODUCTION

Lens epithelium-derived growth factor (LEDGF, also known as p75) had been associated with seemingly unrelated functions: a common nuclear autoantigen (1,2), a transcriptional co-activator (3) and a growth/survival factor of lens epithelial cells (4). More recently, LEDGF/p75 was shown to bind with the integrase of human immunodeficiency virus type I (HIV-1), which catalyzes the integration of reverse-transcribed HIV genome into host DNA (5). Subsequent studies have revealed that LEDGF/p75 assists productive infection of HIV-1 by targeting the integrase to chromatin (5–11). Interestingly, genome-wide identification of HIV integration hotspots in human genome indicated actively transcribed genes (12–14). Later studies showed unequivocally that LEDGF/p75 is responsible for the preferential insertion of viral DNA into active transcription units (15–17). The integrase binding domain (IBD) of LEDGF/p75 binds other partner proteins as well: Myc-interacting protein JPO2 (18), mixed-lineage leukemia (MLL)/menin complex (19), a domesticated transposase PogZ (pogo transposable element derived protein with zinc finger) (20) and Cdc7-activator of S-phase kinase (ASK) (21). Thus, IBD seems to function as an association site for divergent factors that are tethered to chromatin by LEDGF/p75. A tripartite DNA binding element, comprised of a nuclear localization signal (NLS) and two AT-hook motifs, was shown to mediate DNA binding *in vitro* (22). An evolutionally conserved N-terminal domain called PWWP is responsible for chromatin binding although it does not bind free DNA (22,23). Despite the recent progress in this field, molecular mechanisms that enable LEDGF/p75 to recognize actively transcribed genomic regions are still unknown.

In the modern view of eukaryotic transcription, template DNA is threaded through RNA polymerase molecule that is docked at active transcription sites in the nucleus. As a consequence, the template changes higher order structure: negative supercoiling in the receding upstream DNA and positive supercoiling in the approaching downsteam DNA (24). Unrestrained supercoils in the genome are drained by topoisomerases that

*To whom correspondence should be addressed. Tel: +81 86 235 7096; Fax: +81 86 235 7103; Email: otsukimi@md.okayama-u.ac.jp
Correspondence may also be addressed to Ken Tsutsui. Tel: +81 86 235 7097; Fax: +81 86 235 7103; Email: tsukken@cc.okayama-u.ac.jp

transiently break and reseal DNA strands (24,25). However, a line of evidence suggests that topoisomerases are insufficiently processive to keep pace with supercoil generation during transcription (26,27). Recently, it was clearly shown in human cells that dynamic DNA supercoiling associated with transcription is detectable even in the presence of normal concentrations of functional topoisomerases (28).

We report here that LEDGF/p75 selectively binds supercoiled DNA through a novel domain designated supercoiled DNA-recognition domain (SRD) that contains characteristic clusters of lysine and glutamic acid/aspartic acid residues. When SRD was deleted, the protein failed to show any preference toward supercoiled DNA and abolished its co-localization with nuclear domain of active transcription. The results suggest that LEDGF/p75 recruits its binding partners to active transcription units by recognizing superhelical DNA structure.

MATERIALS AND METHODS

Antibodies

Human sera from patients with Miller–Fisher syndrome were screened for autoantibodies against rat supercoiled DNA binding protein 75 (SBP75) by western blotting. IgG was purified from positive sera using MAbTrap Kit (GE Healthcare) and used at 7 μg/ml in western blot analyses. Epitope of the antibody was identified in a C-terminal region (residues 386–528) of rat SBP75/LEDGF/p75 (Supplementary Figure S1). For immunocytochemical detection, the human autoantibody was affinity-purified from the serum using His-tagged rat SBP75/LEDGF immobilized on the HiTrap NHS-activated HP column (GE Healthcare).

Anti-rat SBP75/LEDGF polyclonal antibody was raised in a rabbit against the N-terminal fragment (residues 1–197) of recombinant SBP75/LEDGF. The antibody was then affinity-purified using His-tagged rat SBP75/LEDGF, and used at 0.05 μg/ml in western blot analyses.

Anti-GST (glutathione S-transferase) polyclonal antibody was raised in a rabbit against GST that was expressed in *Escherichia coli*. Affinity-purified anti-GST antibody was prepared using His-tagged GST, and used at 0.07 μg/ml in western blot analyses.

Rabbit polyclonal antibody to human histone H3 (trimethyl-K^4) (ChIP grade) and mouse monoclonal antibody to bromodeoxyuridine (BrdU) Ab-3 (Clone BRD.3) were purchased from Abcam (Cambridge, UK) and Thermo Fisher Scientific (Fremont, USA), respectively. Alexa 488-conjugated goat anti-human IgG and Alexa 594-conjugated goat anti-rabbit IgG or goat anti-mouse IgG were purchased from Invitrogen (USA).

Preparation of nuclear extract and fractionation of supercoiled DNA-binding proteins

Nuclei were isolated from the brain of 2-week-old Sprague-Dawley rats as described (29). To prepare nuclear extract, nuclei were suspended at a final concentration of 5 mg DNA/ml in an extraction buffer containing 20 mM Tris–HCl (pH 7.5), 0.3 M NaCl, 140 mM 2-mercaptoethanol, 50 μg/ml bovine serum albumin (BSA) and 0.2 mM phenylmethylsulfonyl fluoride (PMSF) (30). For sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS–PAGE), 7.5 μl of the extract was applied unless mentioned otherwise.

Ethidium bromide (EtBr) extract was prepared as follows. Nuclei were suspended at a final concentration of 10 mg DNA/ml in a buffer containing 10 mM Tris–HCl (pH 7.0), 5 mM $MgCl_2$, 2 mM $CaCl_2$, 10 mM dithiothreitol (DTT), 0.2 mM PMSF and 20 mM EtBr. The suspension was incubated at 4°C for 30 min and centrifuged at $10\,000 \times g$ for 10 min. The supernatant was then loaded onto a Mono Q column (10/100 GL, GE Healthcare) equilibrated with 20 mM Tris–HCl (pH 8.0) and 1 mM DTT (buffer A). Proteins adsorbed on the column were eluted with a linear gradient of 0–1 M NaCl in buffer A. Fractions (1 ml) were collected and assayed for supercoiled DNA binding activity by Southwestern blotting (described below). Binding activities associated with 75 kDa (SBP75) and 48 kDa (SBP48) proteins were eluted between 0.2 and 0.3 M NaCl concentrations. Fractions containing these proteins were combined separately and subjected to $(NH_4)_2SO_4$ precipitation at 80% saturation. The resulting precipitate was collected by centrifugation and solubilized in a solution containing 20 mM Tris–HCl (pH 7.5), 0.3 M NaCl and 1 mM DTT (buffer B), and subjected to gel filtration chromatography on a Superose 12 column (10/300 GL, GE Healthcare) equilibrated with buffer B. Fractions (125 μl) were collected and assayed for supercoiled DNA binding activity by Southwestern blotting and the peak fractions for SBP75 and SBP48 were pooled. After concentration by ultrafiltration with Microcon filters (Millipore Corp.), samples were stored in 50% glycerol at −20°C.

Cloning of cDNA for SBP75 protein

The cDNA of SBP75 protein was cloned by screening a rat brain cDNA expression library (Uni-Zap XR, Stratagene) with a human autoantibody against the SBP75. Briefly, for the primary screening $\sim 2.4 \times 10^5$ plaques lifted onto 12 nitrocellulose filters (10×15 cm), pretreated with isopropyl-thio-β-D-galactopyranoside (IPTG), were incubated at 37°C for 3 h to induce the coded proteins. After blocking, filters were incubated with 10 μg/ml of human IgG (autoantibody against SBP75) that had been absorbed with boiled *E. coli*. Washed filters were incubated with the anti-human IgG conjugated with alkaline phosphatase (Promega), which was then visualized with BCIP/NBT Color Development Substrate (Promega). Six positive plaques were recovered, amplified and plated again at 500 pfu/plate for the secondary screening. Three positive clones were selected from each plate and tested for the immunogenicity by immunoblotting before and after IPTG induction. Nine clones selected from three individual plates produced the SBP75 protein-related proteins in an IPTG-dependent manner. The SBP75 protein cDNAs were recovered as pBluescript phagemid from the nine phages according to the manufacturer's instruction and

sequenced by Dye Terminator Cycle Sequencing procedure on ABI 377 automatic sequencer (deposited in DDBJ, accession #AB285525). Homologous sequences were BLAST-searched on the NCBI website. Coding sequences obtained were identical to that of rat LEDGF/p75/Psip1 (NM_175765).

To obtain a full-length cDNA of rat SBP75/LEDGF/p75 (SBP75/LEDGF), the ORF of LEDGF/p75/Psip1 was amplified by PCR from rat brain cDNA expression library (Uni-Zap XR, Stratagene) using primers containing the recognition sites for BamHI or EcoRI (listed in Supplementary Table S1). The cDNA of Psip2 (rat equivalent of human p52), a splice variant of rat LEDGF/p75/Psip1, was also obtained by PCR from the phage library with sense primer (5′-GCGGATCCCAAAC ATGACTCGCGATTTCA-3′) and antisense primer (5′-GCGAATTCTGTAGTGTGTACCTTACA-3′).

Construction of plasmids for protein expression

The multicloning sites of pFLAG-CMV-2 vector (Sigma-Aldrich) and pCold II cold-shock expression vector (Takara) were first modified so as to accept any inserts in frame that are excised out from the pGEX-3X expression constructs.

To obtain SBP75/LEDGF and its splice variant (rat p52) that are fused with GST on N-terminus, PCR-amplified cDNAs were digested with BamHI/EcoRI and inserted between the corresponding sites of pGEX-3X vector (GE Healthcare). For GST-fused deletion mutants, DNA fragments were PCR-amplified from the pGEX-3X full-length SBP75/LEDGF construct with the primers listed in Supplementary Table S1 that contained recognition sites for either BamHI or SmaI (sense primers), or for EcoRI (antisense primers), and inserted into pGEX-3X vector digested with appropriate enzymes.

GST-tagged C-terminal truncated proteins (1–336, 100–336, 200–336 and 268–336) were prepared by introducing a stop codon into the sequence that encodes K^{337} of the pGEX-3X constructs for the N-terminal truncation series (1–528, 100–528, 200–528 and 268–528). A site-directed mutagenesis kit (QuikChange II, Stratagene) was used with sense primer 5′-AGAAGCCAGAAGTTA AGTAAGTGGAGAAGAAGCGAGA-3′, and antisense primer 5′-TCTCGCTTCTTCTCCACTTACTTAACTTC TGGCTTCT-3′, where underlined nucleotides were altered to create the stop codon TAA.

To construct plasmids for GST-1–$336^{\Delta 138–205}$, 1–$336^{\Delta 134–276}$ and 1–$336^{\Delta 200–276}$, the pGEX-3X construct containing full-length SBP75/LEDGF with the stop codon at K^{337} was PCR-amplified with KOD-Plus-DNA polymerase (Toyobo Life Science) using sense primers positioned downstream at the deletion and antisense primers upstream, which contained SacII (for Δ138–205) or SalI (for Δ134–276 and Δ200–276) overhangs. PCR products were then digested with appropriate enzymes and self-ligated to restore the uninterrupted coding frame as designed.

For SBP75/LEDGF with hexa-histidine (His_6) tagged at N-terminus, PCR products were inserted between BamHI and EcoRI sites of the modified pCold II. The truncated protein (His-1–336) was prepared by introducing a stop codon at K^{337} in the modified pCold II-SBP75/LEDGF plasmid as described above.

To express FLAG-tagged proteins in mammalian cells, coding sequences for GST-fusion proteins were PCR-amplified and inserted into the BamHI/EcoRI sites of the modified pFLAG-CMV-2.

For localization studies in living cells, we first obtained EGFP cDNA from pEGFP-N1 vector (Clontech) by PCR amplification using Phusion High-Fidelity DNA Polymerase (Finnzymes) and a primer pair with HindIII and BamHI overhangs. The EGFP cDNA was then inserted into the multicloning site of modified pFLAG-CMV-2 vector, which is preceded by FLAG. The resulting plasmid (pFLAG-EGFP-CMV-2) was used as a vector for expression of proteins fused with FLAG-eGFP on N-terminus. Inserts used for the construction were PCR-amplified from cDNA constructs either in pFLAG-CMV-2 or pGEX-3X vectors. The FLAG-eGFP-fused SBP75/LEDGF (1–528, 100–528 and 200–336) was prepared by this scheme and the FLAG-eGFP-fused internal deletion mutant (1–$528^{\Delta 200–346}$) was constructed by the inverse PCR/autoligation protocol as described above for 1–$336^{\Delta 134–276}$ and 1–$336^{\Delta 200–276}$.

To obtain deletion mutants of SBP75/LEDGF fused with eGFP on the C-terminus (93–336, 135–336, 167–336), DNA fragments were PCR-amplified with primers containing restriction overhangs (EcoRI/BamHI) from the full-length SBP75/LEDGF in pGEX-3X and subcloned into pEGFP-N1 vector.

All the constructs used in this work were sequenced to confirm the absence of unintended mutations.

Expression and purification of recombinant proteins

GST-fusion proteins were expressed in JM109 cells. Bacterial pellets were sonicated in lysis buffer containing 150 mM NaCl, 1 mM EDTA, 0.5 mM PMSF, 5 mM DTT, 10 mM Tris–HCl (pH 8.0) and 1 mg/ml lysozyme. GST-proteins were then purified using glutathione–Sepharose 4B (GE Healthcare) according to the manufacturer's instruction.

His-fusion proteins were expressed in BL21 cells (Novagen) according to the manufacturer's instruction. Bacterial cells stored frozen were lysed in a buffer containing 300 mM NaCl, 0.2 mM PMSF, 10 mM imidazole, 50 mM NaH_2PO_4 (pH 8.0) and 1 mg/ml lysozyme. Sonicated lysates clarified by centrifugation were injected on to a HisTrap HP column (GE Healthcare). Bound proteins were eluted with 300 mM NaCl, 0.2 mM PMSF, 250 mM imidazole, and 50 mM NaH_2PO_4 (pH 8.0).

To obtain FLAG-tagged SBP75/LEDGF, HEK 293 cells grown to subconfluency on 100-mm dishes (2×10^6 cells/dish) were transfected with 3 μg of the pFLAG-SBP75/LEDGF plasmid using FuGENE 6 (Roche Molecular Biochemicals). After 32 h, cells were sonicated in a lysis buffer containing 420 mM NaCl and the FLAG-tagged protein was affinity-purified on anti-FLAG M2 antibody-coated magnetic beads, followed by elution with the 3× FLAG peptide (Sigma) as described (31).

Protein concentrations of purified fractions were determined by dye-binding assay kit (Bio-Rad) or densitometric scanning of CBB-stained gel bands using BSA as a standard.

Southwestern blotting, western blotting and electrophoretic mobility shift assay

DNA binding experiments with blotted proteins (Southwestern blotting) was carried out as described previously (29,32). Briefly, proteins were separated by SDS–PAGE and transferred electrophoretically onto a polyvinylidene difluoride (PVDF) membrane (Immobilon, Millipore). When a multi-channel blotter was used, the binding solution (200 µl per slot) consisted of biotin-labeled pUC18 DNA (0.1 µg) and non-specific competitor (100 µg of calf thymus DNA and 60 µg of yeast tRNA) in the standard binding buffer containing 120 mM KCl, 10 mM $MgCl_2$, 0.5 mM DTT, 0.5 mM EDTA, 250 µg/ml BSA, 0.2 mM PMSF and 50 mM Tris–HCl (pH 8.0). To label DNA photochemically with biotin, form I (negatively supercoiled), Ir (relaxed), or form III (linearized) DNA of pUC18 was incubated with Photoprobe^R (long arm) biotin (Vector Laboratories) under the conditions recommended by supplier. The biotinylated DNA bound to proteins was visualized by using alkaline phosphatase-conjugated streptavidin and BCIP/NBT color development substrate (Promega). When blotter was not used, blotted membrane of mini-gel size was incubated in 2 ml of the standard binding buffer containing biotin-labeled pUC18 DNA (1 µg), calf thymus DNA (1 mg), and yeast tRNA (0.6 mg).

Immunoblotting (western blotting) was carried out as described (33). Protein-blotted membranes were incubated with primary antibodies and then with peroxidase-conjugated secondary antibodies (Bio-Rad) that were visualized by either 4-methoxy-1-naphthol or ECL^{TM} (GE Healthcare). Protein bands were recorded on VersaDoc^{TM} imaging system (Bio-Rad) and quantified by densitometry.

For electrophoretic mobility shift assay (EMSA), purified proteins were incubated at 37°C for 30 min with 0.1 µg each of form I and III pUC18 DNA in 15 µl of reaction mixture containing 35 mM Tris–HCl (pH 8.0), 72 mM KCl, 5 mM $MgCl_2$, 5 mM DTT, 5 mM spermidine, 50 µg/ml BSA and 5 µg of linear calf thymus DNA. Samples were then subjected to electrophoresis at 1.6 V/cm on a 1% SeaKem GTG agarose gel containing 10 mM $MgCl_2$ in 0.5 × TBE (Tris–borate–EDTA) buffer. The tank buffer (0.5 × TBE) contained no $MgCl_2$. After staining with 0.5 µg/ml of EtBr, DNA bands were recorded on VersaDoc^{TM} and quantified by densitometry.

Polypeptides of 27 amino acid residues ($K_9E_9K_9$ and K_{27}) were synthesized (GenScript) and used in EMSA. Oligonucleotides 5′-ATG$_{10}$C$_{10}$G$_{10}$AT-3′, 5′-ATC$_{10}$G$_{10}$C$_{10}$AT-3′, 5′-GCA$_{10}$T$_{10}$A$_{10}$GC-3′ and 5′-GCT$_{10}$A$_{10}$T$_{10}$GC-3′ were also obtained (Sigma-Genosys) and the complementary pairs were annealed before use in EMSA.

To study the effects of chromomycin A_3 and distamycin A on supercoiled DNA binding, form I and III pUC18 DNA (0.1 µg each) were preincubated with increasing amounts of chromomycin A_3 or distamycin A hydrochloride (Sigma-Aldrich) in the EMSA reaction mixture for 5 min at room temperature before adding either 5.2 µg of GST-200-336 (protein/bp molar ratio of 0.4) or 62.5 ng of polypeptides $K_9E_9K_9$ / K_{27} (peptide/bp molar ratio of 0.063). Samples were then incubated for 30 min at 37°C and separated on 1% SeaKem GTG agarose gel as described above for EMSA procedure.

Cell transfection, Br-UTP incorporation and immunofluorescence microscopy

HeLa cells were grown on 13-mm round coverslips immersed in Dulbecco's modified Eagle's medium (DMEM) (Sigma-Aldrich) with 10% fetal calf serum. Cells in subconfluency were transfected with FuGENE 6 and 1 µg of the pFLAG-EGFP or pEGFP expression vectors harboring full-length or mutant SBP75/LEDGF. After 24 h, live cells were examined under a microscope equipped with epifluorescence and ApoTome device (Axiovert 200 M, Carl Zeiss). Images were captured by a cooled CCD camera (AxioCam MRm) using Axiovision 4.01 software (Zeiss). To observe eGFP signals in fixed cells, cells on coverslips were immersed in pure acetone cooled to −80°C for 5 h, and then incubated at −20°C for 1 h, 4°C for 20 min and finally at room temperature for 10 min. After treating the cells with 0.75% paraformaldehyde in acetone at room temperature for 10 min, cells were immersed in phosphate-buffered saline (PBS, 150 mM NaCl, 10 mM sodium phosphate, pH 7.4) containing 25% glycerol and 125 mM glycine at room temperature for 10 min, and washed with 0.1 M phosphate buffer (pH 7.4). The fixed cells were either examined with the microscope, or subjected to immunolabeling.

To incorporate Br-UTP into permeabilized HeLa cells, subconfluent cells on coverslips were treated for 3 min at room temperature with 0.05% Triton X-100 in glycerol buffer containing 20 mM Tris–HCl (pH 7.4), 5 mM $MgCl_2$, 25% glycerol, 0.5 mM PMSF and 0.5 mM EGTA. Permeabilized cells were incubated for 10 min at room temperature with 0.5 mM each of ATP, CTP, GTP and 0.2 mM Br-UTP (Sigma-Aldrich) in transcription buffer containing 100 mM KCl, 50 mM Tris–HCl (pH 7.4), 5 mM $MgCl_2$, 0.5 mM EGTA, 25% glycerol, 25 µM S-adenosyl-L-methionine, 5 U/ml placental RNase inhibitor and 1 mM PMSF (34). Cells were washed with ice-cold TBS (150 mM NaCl, 5 mM $MgCl_2$, 10 mM Tris–HCl, pH 7.4) containing 0.5% Triton X-100 and 5 U/ml RNase inhibitor and then fixed with freshly made 4% paraformaldehyde in 0.1 M phosphate buffer (pH 7.4), followed by immunolabeling.

For immunofluorescence microscopy, subconfluent HeLa cells on glass coverslips were fixed with 4% paraformaldehyde, permeabilized with 0.3% Triton X-100 in PBS, and blocked with 10% goat serum and 10% horse serum in PBS containing 0.3% Triton X-100 (TPBS). After incubation with primary antibodies diluted in TPBS containing 1% goat serum, sections were washed with PBS, followed by incubation with fluorescent conjugates of goat secondary antibodies in PBS (33,35). Cells were finally stained for DNA with 0.25 µg/ml

4′,6-diamidino-2-phenylindole (DAPI) in PBS and examined under the microscope.

RESULTS

Identity of SBP75 as LEDGF/p75

In a previous study, we detected DNA binding activities selective for supercoiled DNA in the nuclear fraction isolated from rat brain (29). To find out responsible factors for the activity, proteins were extracted from purified nuclei and subjected to Southwestern blotting analysis using pUC18 DNA labeled with biotin, either negatively supercoiled (I) or linearized (III) forms, as probes (Figure 1A). Three protein bands (75K, 120K and 180K) were detected with biotin-labeled probes (lanes 2 and 4). Unlabeled non-specific competitors, a mixture of calf thymus DNA fragments and yeast tRNA, replaced the binding of form III probe to all proteins (lane 3), whereas the competitor failed to replace the binding of form I probe to 75K and 180K proteins (lane 5). Additional unlabeled competitors, form III or form Ir (relaxed from I), also failed to replace the binding of supercoiled DNA probe (lanes 6–9). However, the binding was effectively abolished by unlabeled form I DNA (lanes 10 and 11), indicating that 75K and 180K proteins preferentially recognize supercoiled DNA (see also Supplementary Figure S2 for more evidence).

Since some DNA binding proteins that recognize DNA superstructures are selectively released from chromatin by DNA-intercalating agents such as EtBr (36), we tried to enrich these proteins by treating isolated nuclei with EtBr. As shown in Figure 1B, the EtBr extract contained proteins with apparent molecular mass of 75 and 48 kDa that showed preferential binding to supercoiled DNA. These proteins will be referred to as SBP75 and SBP48 hereafter (SBP stands for supercoiled DNA binding protein). The 180K polypeptide was not detected in the EtBr extract. The release of SBP75/48 was solely dependent on EtBr treatment since both proteins were not extracted without EtBr (data not shown). The EtBr extract was then subjected to an anion exchange chromatography (Mono Q), followed by gel filtration (Superose). The left and the middle panels of Figure 1C show that SBP75 and SBP48 were eluted together from Mono Q column at 0.2–0.3 M NaCl, and separated by gel filtration into SBP75-enriched (Superose 75K) and SBP48-enriched (Superose 48K) fractions.

In an attempt to obtain autoantibodies to nuclear proteins, we found a highly specific antibody to SBP75 in a patient serum with Miller–Fisher syndrome (Figure 1C, right panel). The syndrome is characterized as an acute autoimmune-mediated neuropathy with serum antibodies to gangliosides such as GQ1b, GM1 and GD1a (37). It is worth noting that Miller–Fisher syndrome sometimes occurs early in HIV infection before AIDS develops (38). SBP48 was not recognized by the autoantibody. Using this antibody, we screened a rat brain phagemid library for cDNA expression. Six positive clones were obtained from 2.4×10^5 plaques. After secondary screening, three positive plasmid clones were recovered and sequenced. The overlapping sequences spanning 2979 nt contained a coding sequence identical to that of rat LEDGF/p75/Psip1 (NM_175765). The resulting sequence extended further downstream, terminating at an alternative poly(A) site (deposited in DDBJ as AB285525).

Rat SBP75/LEDGF/p75 (SBP75/LEDGF) expressed as GST-fused or FLAG-tagged proteins were recognized by the human autoantibody as expected (Figure 1D). An alternatively spliced variant of p75 is expressed as p52, which shares N-terminal 325 (323 in rat) residues of p75 (3). The coding sequence of rat p52 (BC093606) was amplified from the rat cDNA library and expressed as a GST-fused protein, which failed to react with human autoantibody. Reactivity of this antibody against a series of GST-fused SBP75/LEDGF deletion mutants showed that the antibody epitope resides within the C-terminal region (amino acids 386–528) (Supplementary Figure S1). A polyclonal antibody raised against the N-terminal region (amino acids 1–197) of rat SBP75/LEDGF recognized GST-SBP75/LEDGF and GST-rat p52, as well as 75 and 48 kDa proteins in the nuclear extract, indicating that SBP75 is indeed the rat LEDGF/p75 and that SBP48 is equivalent to the recombinant rat p52. Southwestern blotting analysis of FLAG- and His-SBP75/LEDGF confirmed the preferential binding of expressed proteins to supercoiled DNA (Figure 1E and Supplementary Figure S2A). GST-tagged SBP48/p52 also showed the preference to supercoiled DNA on Southwestern blotting (Supplementary Figure S3).

The binding of SBP75/LEDGF to supercoiled DNA was further analyzed by gel retardation assay (Figure 2). We used here either His-tagged SBP75/LEDGF (His-528) or a C-terminal truncation mutant (His-336), which is nearly identical to SBP48/p52. As illustrated in Figure 2A, His-336 contains all the domains involved in chromatin tethering, PWWP (23) and DNA binding, NLS and AT-hooks (22). Both proteins bound to both DNA forms (forms I and III) in the absence of competitor DNA, but in its presence they bound selectively to supercoiled DNA, as only form I DNA retarded significantly at the protein-to-base pair molar ratio of higher than 0.02 (Figure 2B). Levels of affinity to DNA were comparable between His-528 and His-336. The supercoil-specific band shift was not observed when Mg^{2+} ion was removed from the gel buffer (data not shown). These results indicate that the N-terminal portion of SBP75/LEDGF (1–336) is sufficient for the supercoil-specific DNA binding and the C-terminal region including the IBD is dispensable.

Identification of the recognition domain for supercoiled DNA

To locate a responsible region for the supercoil specificity in the protein, GST-tagged N-terminal fragment (GST-1–336) and its deletion mutants were analyzed by Southwestern blotting (Figure 3A). Deleting amino acids 275–336 from GST-1–336 (GST-1–274) resulted in a drastic decrease of supercoiled DNA binding activity to one-third (Figure 3B and C). GST-1–206 that contains NLS and two AT-hook motifs, known elements

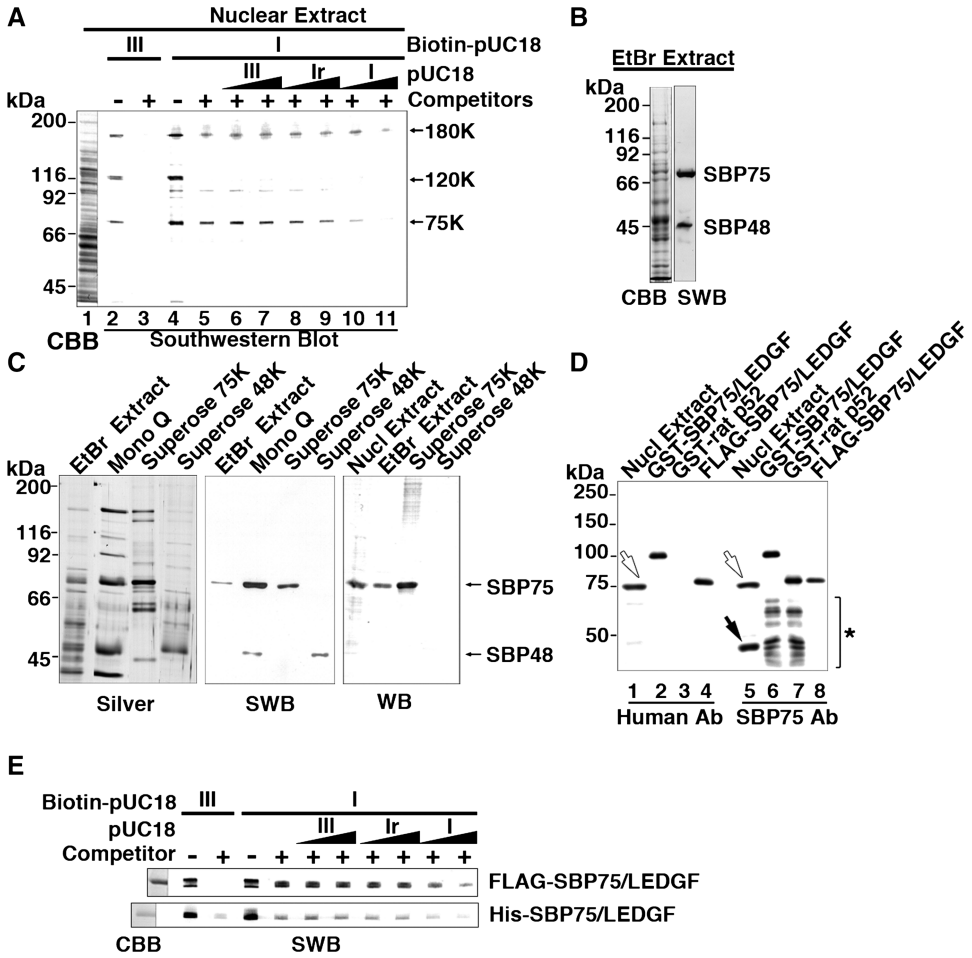

Figure 1. Supercoiled DNA-binding protein SBP75 is identical to LEDGF/p75. (**A**) Detection of DNA binding proteins in rat brain nuclear extract by Southwestern blotting. Nuclear extract (75 µl) was subjected to 7.5% SDS–PAGE and separated proteins were blotted on a membrane, which was then incubated with biotin-labeled pUC18 DNA: form III (lanes 2 and 3) or form I (lanes 4–11) in the absence (lanes 2 and 4) or the presence (lanes 3 and 5–11) of non-specific competitors (calf thymus DNA and yeast tRNA). Excess amounts of unlabeled pUC18 DNA: form III (lanes 6 and 7), form Ir (lanes 8 and 9) or form I (lanes 10 and 11) were added at 5-fold (lanes 6, 8 and 10) or 25-fold excess (lanes 7, 9 and 11). Proteins with apparent molecular mass of 180 kDa (180K), 120 kDa (120K) and 75 kDa (75K) are indicated. Total protein on the membrane was stained with CBB (lane 1). (**B**) Proteins extracted with 20 mM EtBr were subjected to Southwestern blotting (SWB). The blot was incubated with biotin-labeled form I pUC18 in the presence of non-specific competitors. Supercoiled DNA bound to the two proteins labeled SBP75 and SBP48 selectively (right lane). Total protein in the extract was stained with CBB (left lane). (**C**) Proteins in 20 mM EtBr extract were fractionated by chromatography on Mono Q and Superose columns and pooled fractions were subjected to SDS–PAGE and stained with silver reagent (left panel). The same sets of samples were blotted onto a membrane and analyzed by SWB or western blotting (WB) with human autoantibody to SBP75. (**D**) Rat nuclear extract and SBP75 recombinant proteins were subjected to western blotting analysis with human autoantibody to SBP75 (Human Ab, lanes 1–4) or with newly prepared anti-rat SBP75/LEDGF polyclonal antibody (SBP75 Ab, lanes 5–8). Open arrows indicate endogenous SBP75 and closed arrow indicates SBP48 that was reactive to polyclonal antibody but not to human autoantibody. Lower bands in GST fusion proteins that were detected by SBP75 Ab (indicated by asterisk in lanes 6 and 7) but not by human Ab are degradation products lacking C-terminal portions. For further analysis of epitope position recognized by human autoantibody, see Supplementary Figure S1. (**E**) Purified FLAG- or His-tagged SBP75/LEDGF (2.5 µg each) was subjected to Southwestern analysis as in panel A. CBB-stained proteins are shown in the left most lane.

responsible for the protein's DNA binding ability (22), did bind pUC18 DNA but without any preference to supercoiled conformation. GST-1–137 harboring PWWP domain did not bind neither form I nor form III DNAs (Figure 3B). Conversely, a mutant without PWWP domain (GST-100–336) retained the supercoil specificity (Figure 3B) although the activity reduced to almost one half (Figure 3C). To our surprise, further deletion of the canonical DNA binding domain (NLS and AT-hooks, aa 148–196) did not abolish DNA binding activity of the protein nor preference to supercoiled DNA. The supercoiled DNA binding activity of GST-200–336 was comparable to that of GST-100–336 (Figure 3B and C). We found that GST-268–336, the C-terminal half of the GST-200–336, does not show any DNA binding activity (Figure 3B), and that the N-terminal half (GST-200–274) is sufficient to recognize and bind supercoiled DNA (Figure 3B). Although the binding activity of GST-200–274 detected in Southwestern blotting was very low, the essential role of this region in recognizing supercoiled DNA was confirmed by the failure of GST-1–336 mutant that lacks aa 200–276 (Δ200–276) to recognize supercoiled DNA (Figure 3B and C). Therefore, the amino acids 200–274 region seems to be the primary site

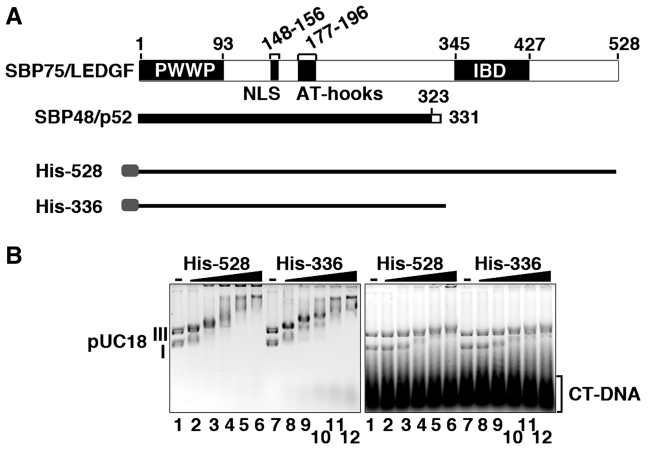

Figure 2. Preferential binding of recombinant proteins to supercoiled DNA can be detected by gel shift assay (EMSA). (**A**) The domain structure of rat SBP75/LEDGF variants and His-tagged proteins used in the binding assay. Numbers on top are amino acid residue numbers. PWWP domain, nuclear localization signal (NLS), double AT-hook motif and IBD (integrase-binding domain) are indicated. Rat SBP48/p52 shares amino acids 1–323 with SBP75/LEDGF (thick solid line) and the following eight residues are unique (hollow line). (**B**) Gel retardation patterns of pUC18 plasmid DNA (forms I and III) with His-528 or His-336 in the absence (left panel) or in the presence (right panel) of calf thymus DNA fragments (CT-DNA). The protein amounts added were expressed as protein-to-nucleotide (bp) molar ratio: 0 (lanes 1 and 7), 0.01 (lanes 2 and 8), 0.02 (lanes 3 and 9), 0.03 (lanes 4 and 10), 0.04 (lanes 5 and 11) and 0.05 (lanes 6 and 12).

for the interaction with supercoiled DNA (thus called 'core') and the contiguous C-terminal region (up to amino acid 336) appears to enhance the interaction (called 'enhancer').

Now we term the amino acids 200–336 region as SRD (illustrated in Figure 4A). The conventional DNA binding region expressed as a GST-fusion protein (GST-137–206) bound pUC18 DNA without any preference to DNA superstructure (Figure 3B). Deletion of this region (Δ138–205) did not abolish the preferential binding to supercoiled DNA (Figure 3B and C), suggesting that this domain is dispensable for the supercoiled DNA-binding activity. Therefore, the amino acids 137–206 region is termed as NRD that designates non-specific DNA-recognition domain (Figure 4A). As expected, deletion of both NRD and the core region of SRD (Δ134–276) abolished the conformation-specific as well as non-specific DNA binding activities (Figure 3B and C). It is clear now that SBP75/LEDGF contains two discrete DNA binding domains, one that binds DNA without preference to DNA superstructure (NRD) and the other that binds supercoiled DNA selectively (SRD).

SRD (amino acids 200–336) and NRD (amino acids 137–206) were further analyzed by EMSA (Figure 4A, lower panels). GST-137–206 (NRD) bound to both linear (form III) and supercoiled (form I) DNAs at a molar ratio of >0.01 as revealed by the dose-dependent band shift, which was cancelled completely by the linear competitor DNA. The result corroborates the non-specific nature of the DNA binding to NRD. In contrast, GST-200–336 (SRD) showed a preferential binding

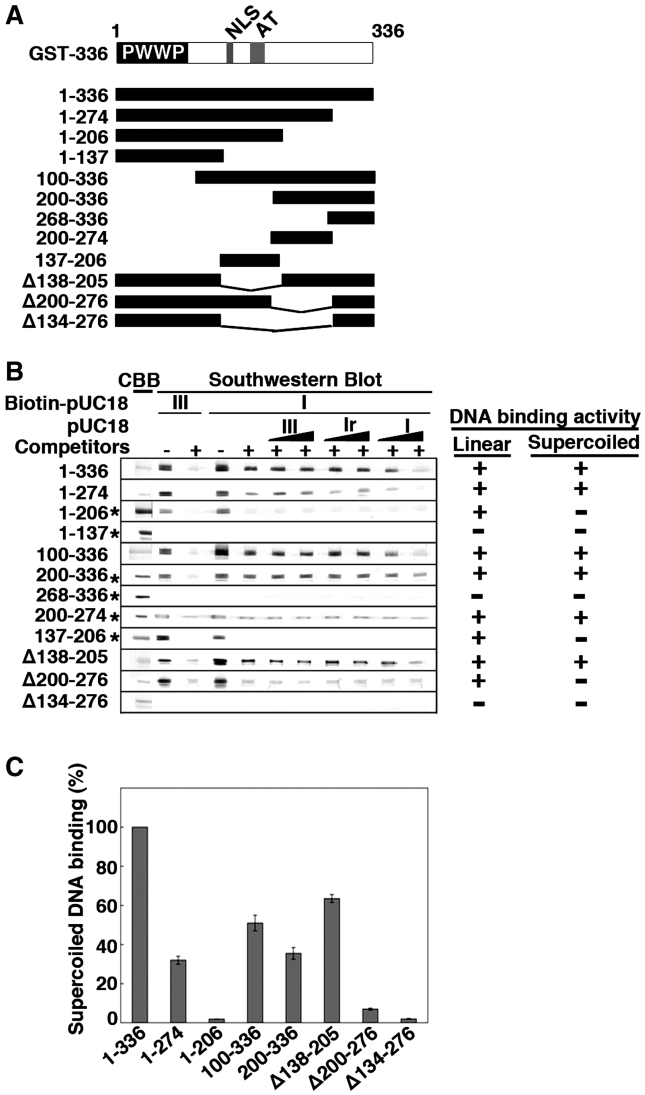

Figure 3. Nailing down the region of SBP75/LEDGF that is responsible for supercoiled DNA binding activity by Southwestern blotting. (**A**) Deletion mutants of GST-SBP75/LEDGF used for the analysis are illustrated. Domain structure of N-terminal fragment (amino acids 1–336) is shown on the top. (**B**) Protein amounts loaded to the gel were either 5 μg (indicated by asterisks on the left) or 1.5 μg (others). Blotted membranes were incubated with biotin-labeled pUC18 as in Figure 1A. DNA binding to linear DNA or preferential binding to supercoiled DNA is summarized on the right. (**C**) Relative binding activity of SBP75/LEDGF deletion mutants. GST-fusion proteins (50 ng each, except GST-1–206 and GST-200–336 that were100 ng) were loaded to 7.5% polyacrylamide gel. Binding of biotin-labeled pUC18 form I DNA in the presence of non-specific competitors was measured by scanning densitometry. To express the binding activity on a molar basis with respect to proteins, an identical blot was probed with affinity-purified anti-GST polyclonal antibody and the quantified band intensities were used to normalize the DNA binding. Data shown are relative values to GST-1–336 (average for three independent experiments with error bars).

to form I DNA both in the presence and absence of calf thymus DNA at the molar ratio of higher than 0.4. The SRD core (GST-200–274) also showed a preference to supercoiled DNA at the molar ratio of >1.0. Although the relative binding activity of SRD core was

significantly lower than that of SRD, the core consisting of 75 amino acids can be regarded as the minimal domain for supercoil recognition. The absence of any DNA binding activity associated with GST- 268–336 (Figure 4A) is consistent with the result of Southwestern blotting (Figure 3B), although the region enhanced the binding of supercoiled DNA by SRD core >2-fold when attached to it. Interestingly, the core and the enhancer portion of SRD roughly correspond to the charged region 2 (CR2) and the charged region 3–4 (CR3–4), respectively, that were reported in previous studies (23,39).

Prediction of higher order structure *in silico* revealed that the highly charged SRD is a disordered region with little folded structure (Supplementary Figure S4). The core region of SRD does not contain any known DNA-binding motifs, as well. To identify sequence motifs responsible for the recognition of DNA superstructure, SBP75/LEDGF amino acid sequences from 13 different species that are homologous to SRD were aligned (Supplementary Figure S5A). The region was enriched with lysine (K) and glutamic acid (E) residues that were highly conserved among these species (Supplementary Figure S5B). Although acidic residues, E and D, are generally used sparingly in DNA binding, both have similar preferences for amino acid–base pair hydrogen bonding (40). When E and D were considered to be equivalent, 22 out of 26 acidic residues of rat SRD are conserved. The sequence logo representation of SRD showed that conserved lysines are more or less clustered into several groups and acidic stretches composed of conserved E and D intervene the lysine clusters in SRD (Figure 4B). Five lysine clusters (K1–K5) and four acidic clusters (E/D1–E/D4) were recognizable in SRD when the criteria for being a cluster was more than three consecutive residues of K or E/D allowing the interruption by up to three unrelated residues. Four out of five lysine clusters were located in the core, K2 and K3 flanking E/D2 at the center to form a characteristic stretch of 27 amino acids composed of nine residues each of basic–acidic–basic configuration (Figure 4B).

To study whether these 27 amino acids (K2–E/D2–K3) of the core region play primary roles in the preferential binding to supercoiled DNA, we synthesized oligopeptides that mimic this stretch and used in EMSA. As expected, polypeptide $K_9E_9K_9$ preferentially bound to form I DNA at the peptide/base pair molar ratio of 0.063, about six times smaller ratio than that observed with GST-200–336. Since polypeptide K_{27} bound to both forms of DNA at the same molar ratio without showing any preference toward form I DNA, not just basic lysine stretch but alternating clusters of K–E–K are required to preferentially recognize supercoiled DNA (Figure 4C). A statistical analysis of protein–DNA contacts revealed that lysine residue can form stable hydrogen bonds with guanine base in the major groove (40). Similarly, glutamic or aspartic acid residues interact frequently with cytosine base. These facts suggest that the K–E–K cluster in SRD core may interact preferentially with DNA containing a G–C–G cluster. By assuming that the conformation of $K_9E_9K_9$ is an extended random coil and it may interact with three helical turns of DNA, we synthesized a double-stranded oligonucleotide representing a G–C–G cluster ($G_{10}C_{10}G_{10}$). When this oligonucleotide was added to the EMSA reaction with SRD (GST-200–336) and supercoiled DNA, the supercoil-specific band shift was abolished by competition in a dose-dependent manner, whereas the control oligonucleotide $A_{10}T_{10}A_{10}$ did not show any competition (Figure 4D).

Involvement of K2–E/D2–K3 region of SRD in supercoiled DNA recognition was further suggested by EMSA experiments using DNA-interacting agents with certain base preferences (chromomycin A_3 and distamycin A). Chromomycin A_3 that binds double-stranded DNA with GC base pair specificity counteracted the shift of supercoiled DNA by either GST-200–336 or polypeptide $K_9E_9K_9$ in a dose-dependent manner. Whereas, distamycin A that has AT base pair specificity showed no effect on the shift (Supplementary Figure S6A and B). Also, both agents had no effect on the binding of K_{27} with form I and III DNAs at the same peptide-to-base pair molar ratio (Supplementary Figure S6B). These results are perfectly complementary to the data shown in Figure 4C and D. Taken together, the alternating clusters of K and E/D in the SRD core are likely to be involved in the preferential binding of SBP75/LEDGF to some GC-rich segments in supercoiled DNA (see 'Discussion' section for more detail).

Colocalization of SBP75/LEDGF with active transcription sites in the nucleus

Negative supercoils generated at transcription sites were shown to propagate through chromatin fiber *in vivo* (28). Proteins capable of recognizing supercoils would be concentrated in a close proximity of actively transcribed chromatin regions. Intranuclear distribution of endogenous SBP75/LEDGF was studied in relation to transcription sites that are highlighted by immunopositive area detected with a specific antibody against trimethyl-lys^4 of histone H3 (H3K4me3). Unexpectedly, overall distribution of endogenous SBP75/LEDGF and H3K4me3 as detected by double immunostaining appeared to be mutually exclusive: SBP75/LEDGF signals tend to overlap with condensed chromatin area whereas H3K4me3 signals were observed in dispersed chromatin (Figure 5A, upper panels). However, magnified images of the H3K4me3 immunopositive area revealed a speckled distribution of SBP75/LEDGF overlapping with or juxtaposed to H3K4me3-positive area in inter-chromatin space (Figure 5A, arrows in lower panels). Additional data are shown in Supplementary Figure S7A. Sites of ongoing transcription in the nucleus (transcription foci) were visualized more directly by incorporation of Br-UTP (Figure 5B, upper panels). Closer observation at a higher magnification showed that Br-UTP signals distribute in a punctuated pattern in the area of relative DNA depletion, partially overlapping with or locating close to SBP75/LEDGF signals (Figure 5B, arrowheads in lower panels). Additional data are shown in Supplementary Figure S7B. Thus, a part of endogenous SBP75/LEDGF appears to colocalize with transcriptionally active regions in the nucleus.

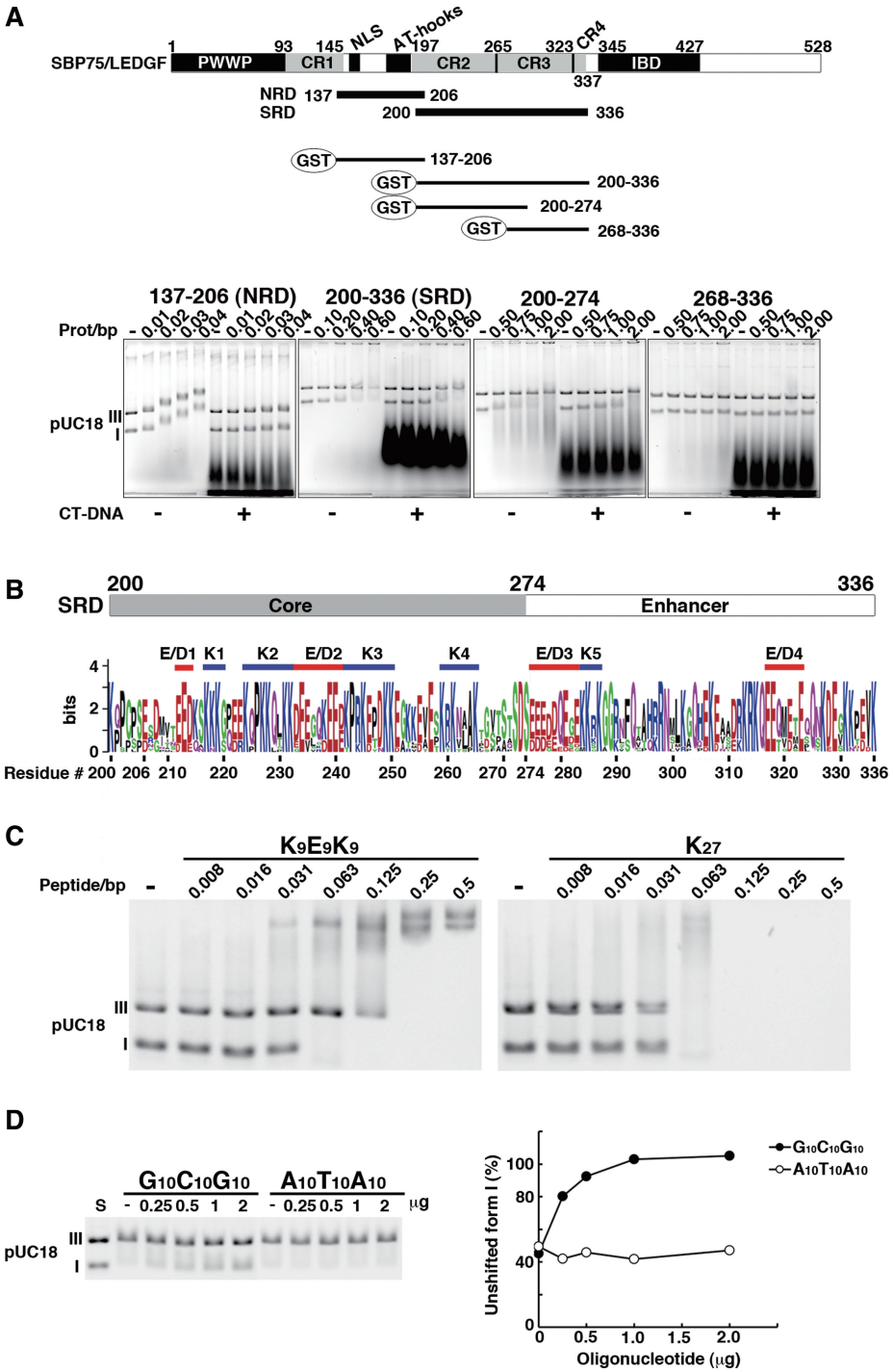

Figure 4. Further characterization of DNA-binding domains of SBP75/LEDGF. (**A**) Subregions of the central portion (137–336) were expressed as GST-fusion proteins and their DNA binding properties were analyzed by EMSA. Domain structure of SBP75/LEDGF is illustrated on the top, including the charged region (CR1–4) and newly identified DNA-binding domains: non-specific DNA-recognition domain (NRD) and supercoiled DNA-recognition domain (SRD). GST-fused fragments used in the assay are shown beneath. Numbers indicate the residue number of SBP75/LEDGF. Shown in the lower panels are mobility shift patterns of supercoiled (I) and linear (III) DNA for each protein. Increasing amounts of GST-fused proteins were incubated with DNA in the absence (−) or presence (+) of CT-DNA. Prot/bp stands for the molar ratio of protein/base pare of DNA. (**B**) Visual assessment of essential elements in SRD by using sequence logo representation. The minimal element for supercoiled DNA binding (core) and a region that enhances the supercoil-specific binding (enhancer) are indicated on the top. Homologous regions of 13 species were aligned (Supplementary Figure S5A), and each amino acid residue was scaled by the total bits of information multiplied by the relative occurrence of the amino acid residue at the position. Attached residue numbers are for rat SBP75/LEDGF. Red bars indicate acidic clusters (E/D1–E/D4) and blue bars indicate lysine clusters (K1–K5). For cluster definition see in the text. (**C**) EMSA analysis of a polypeptide $K_9E_9K_9$ that mimics the cluster structure of the core (K2–E/D2–K3). Increasing amounts of $K_9E_9K_9$ or K_{27} (control) were incubated with forms I and III pUC18 DNA. In this experiment, actual amount of added polypeptide was 0.5 μg per reaction at the polypeptide/base pair molar ratio (peptide/bp) of 0.5. (**D**) Competitive effects of double-stranded oligonucleotides $G_{10}C_{10}G_{10}$ and $A_{10}T_{10}A_{10}$ on the binding of supercoiled DNA with SRD. Increasing amounts of the oligonucleotides indicated on top were added into the reaction mixture containing 6.5 μg of GST-200–336 (equivalent to 0.5 protein/bp) (left panel). Band densities of form I DNA relative to that of DNA alone (S) are plotted (right panel).

We next examined whether or not SRD, the newly identified supercoiled DNA binding domain, is a primary determinant for the association of SBP75/LEDGF with transcription sites by using eGFP-fused rat proteins expressed in HeLa cells (Figure 6A). To identify artifacts originating from fixation, cell images were taken under different conditions of fixation and compared with live cell images (Figure 6B). The wild-type SBP75/LEDGF (aa 1–528) observed in live cells showed intranuclear speckled distribution excluded from nucleoli, which is similar to the pattern of human LEDGF/p75 expressed in CHO-K1 cells (41) or HeLa cells (6). Intranuclear distribution of the protein had been shown to be almost coincident with DNA (5,9,22). Despite the absence of NLS, SRD (eGFP-200–336) localized exclusively in nuclei like the wild-type protein (Figure 6B, left column). When cells were fixed with a standard fixative (4% paraformaldehyde), however, SRD changed its intranuclear distribution drastically to apparent association with the nucleolus-like structure that was rather avoided by the protein in live cells (Figure 6B, middle column). This relocation phenomenon was not evident with the wild-type protein. It is very likely to be caused by the absence of PWWP domain, since all the SBP75/LEDGF mutants lacking this domain showed similar relocation when fixed with paraformaldehyde (Supplementary Figure S8). We found that this artifact can be prevented by treating the cells with acetone at −80°C prior to the paraformaldehyde fixation (Figure 6B, right column). The cold-acetone pre-treatment did not affect the localization of

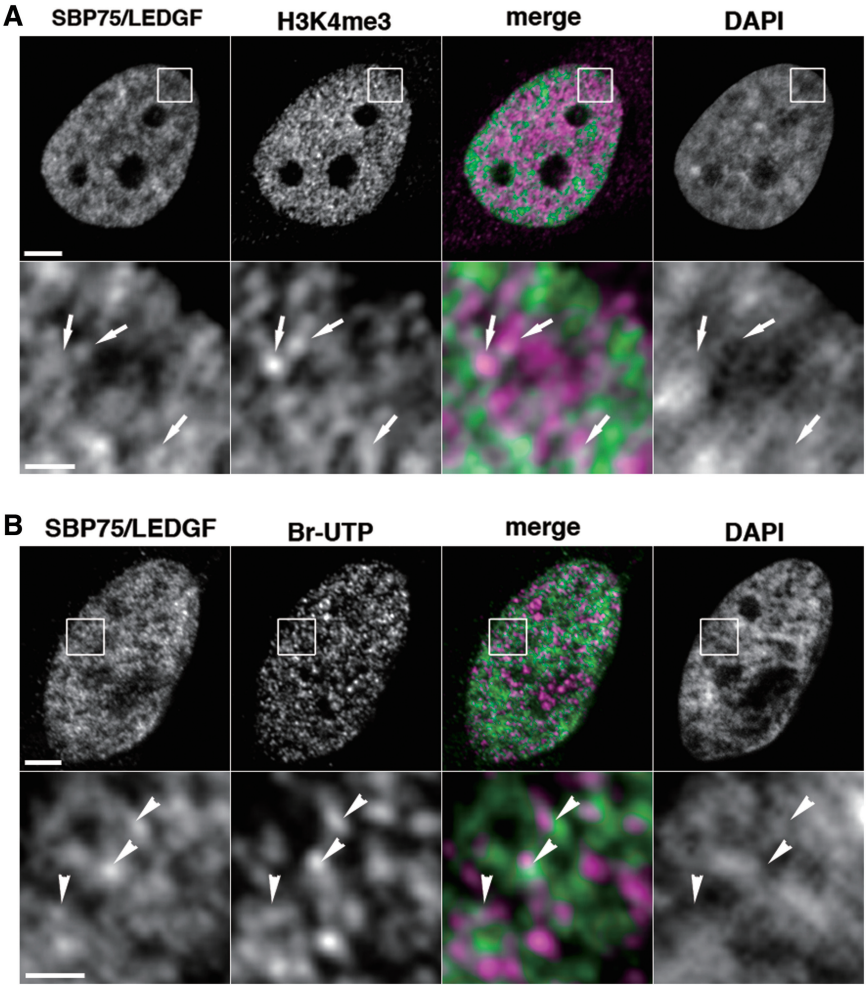

Figure 5. Intranuclear distribution of endogenous SBP75/LEDGF and active transcription sites in HeLa cells. (**A**) Cells were fixed with 4% paraformaldehyde and immunostained doubly with affinity purified human autoantibody to SBP75/LEDGF (3 μg/ml) and rabbit polyclonal antibody to human trimethyl-lysin[4] in histone H3 (H3K4me3, 1/1000×), followed by incubation with Alexa 488-conjugated goat anti-human IgG (5 μg/ml) and Alexa 594-conjugated goat anti-rabbit IgG (5 μg/ml). DNA was stained with DAPI (0.25 μg/ml). Boxed area in upper panels was magnified and shown in lower panels. Arrows in the area with weak DAPI signal indicate that Alexa 488 (green) and Alexa 594 (magenta) signals are overlapped or juxtaposed in the merged image. Scale bars designate 5 μm (upper panels) and 1 μm (lower panels), respectively. (**B**) Br-UTP was incorporated into transcription sites in the cells permeabilized with 0.05% Triton X-100 by incubation with other nucleotides and detected with anti-BrdU mouse monoclonal antibody (1 μg/ml) that was visualized with Alexa 594-labeled second antibody. Endogenous SBP75/LEDGF was also detected with affinity purified human autoantibody to SBP75/LEDGF as in panel A. Arrowheads in low DAPI area indicate that Alexa 488 (green) and Alexa 594 (magenta) signals were overlapped or juxtaposed in the merged image. Scale bars designate 2 μm (upper panels) and 0.5 μm (lower panels), respectively.

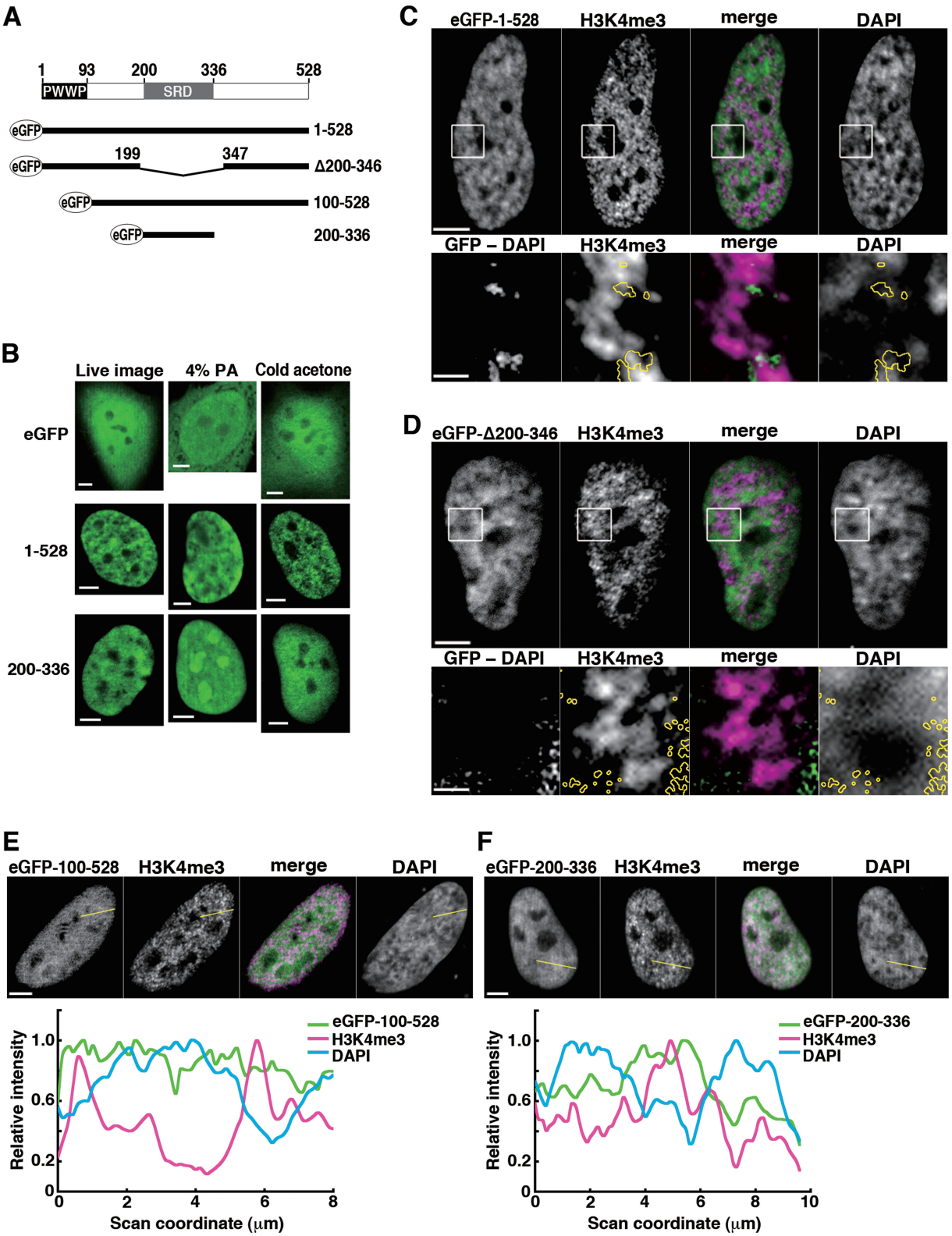

Figure 6. Analysis of domain function for intranuclear localization: eGFP-fused SBP75/LEDGF deletion mutants expressed in HeLa cells. (**A**) Schematic representation of eGFP-fused proteins in relation to domains of rat SBP75/LEDGF, PWWP and SRD. Amino acid residue numbers are indicated. (**B**) Effects of fixation conditions. eGFP-SBP75/LEDGF (1–528), eGFP-SRD (200–336) or eGFP (control) were transiently expressed in HeLa cells, and eGFP signal was detected in live cells (left panels), in cells fixed with 4% paraformaldehyde (middle panels), and in cells pretreated with cold acetone before fixation with 0.75% paraformaldehyde (right panels). Scale bars, 5 μm. (**C–F**) Distribution of transcription sites (H3K4me3) in relation to locations of eGFP-fused SBP75/LEDGF deletion mutants expressed in HeLa cells. Cells were treated with cold acetone prior to fixation with paraformaldehyde. The fixed cells were immunolabeled with anti-H3K4me3 antibody (1/1000×) and Alexa 594-conjugated goat anti-rabbit IgG (5 μg/ml). eGFP was visualized directly as fluorescence. Boxed area in upper panels of (C) and (D) was magnified and shown in lower

(Continued)

full-length SBP75/LEDGF. In the following experiments, therefore, cells were pretreated with cold acetone before aldehyde fixation.

Being consistent with the results of endogenous SBP75/LEDGF, general distribution of eGFP-SBP75/LEDGF and H3K4me3 were mutually exclusive (Figure 6C, upper panels). Like the endogenous protein, a portion of eGFP-SBP75/LEDGF colocalized with H3K4me3 in dispersed DNA area as shown in the enlarged images (Figure 6C, yellow circles in lower panels). In these images, the signal intensity of GFP overlapping with DAPI was reduced by digital subtraction (GFP-DAPI) to highlight the protein distributed in dispersed chromatin. Deletion of SRD from the SBP75/LEDGF (eGFP-Δ200–346) did not apparently changed the overall distribution of the protein, most of the signals being excluded from H3K4me3 signals (Figure 6D, upper panels). When observed more closely, however, the SRD-deleted protein was absent from the H3K4me3 positive region and rather present in DNA-rich area (Figure 6D, yellow circles in lower panels), which is clearly different from the situation with the wild-type protein (Figure 6C, lower panels). These observations are supported by additional data shown in Supplementary Figure S9. The results suggest that SRD is indeed a primary determinant for targeting the protein to transcription sites.

Then, why does most of SBP75/LEDGF locate in condensed chromatin regions regardless of SRD presence? The N-terminal PWWP domain was shown to play a dominant role in tethering the protein to chromatin (22,23,42). Therefore, we expressed a PWWP deletion mutant (eGFP-100–528) in HeLa cells and eGFP signals were looked at in relation to H3K4me3 signals. As expected, the deletion mutant lost the ability to colocalize with condensed chromatin and distributed diffusely in both H3K4me3-positive and negative regions (Figure 6E). Finally, we assessed the localization of SRD (eGFP-200–336) that lacks both PWWP and NRD. Distribution of this construct correlated well with that of H3K4me3 signals and rather excluded from condensed chromatin regions as shown more clearly by the densitometric profiles (Figure 6F). These results suggest that nuclear SBP75/LEDGF is mainly tethered to chromatin through the PWWP domain but a small portion of the protein is recruited to transcriptionally active region by SRD that recognizes negative supercoils generated at transcription sites (illustrated in Figure 7).

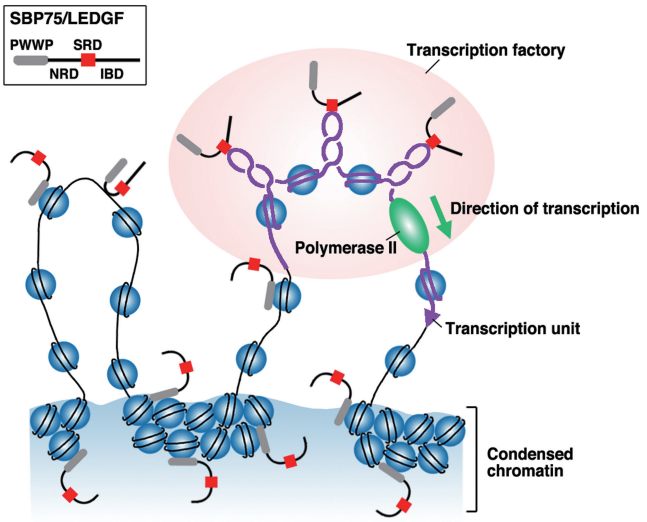

Figure 7. A model for domain-guided localization of SBP75/LEDGF in nuclear compartments. Three domains in SBP75/LEDGF (PWWP, NRD and SRD) are likely to be involved in the localization. PWWP exerts a strong tethering effect that directs the protein, for the most part, to condensed chromatin territory, probably through interaction with nucleosomes. NRD may bind to any DNA in a structure-independent manner, whereas SRD binds selectively to negative supercoils generated around the transcription factory, thus recruiting IBD-bound proteins (data not shown) for preferential targeting toward active transcription units. Although it is not indicated in the scheme, PWWP in the targeted complex may also interact with nearby nucleosomes.

DISCUSSION

We show here for the first time that LEDGF/p75, a protein known to interact with HIV-1 integrase and plays an important role in the targeting of viral genome to host DNA, recognizes and binds negatively supercoiled DNA preferentially. We had been characterizing this protein for some time as a supercoiled DNA binding protein with apparent molecular mass of 75 kDa (thus termed SBP75). Canonical DNA binding domain of LEDGF/p75 was mapped to residues #93–226 harboring NLS (amino acids 148–156) and two AT-hook motifs (amino acids 177–196) (22). We confirmed that the tagged-fragment of aa137–206 containing NLS and AT-hooks bound to DNA *in vitro*. However, this region did not show any preference to supercoiled DNA. We found instead that a novel domain spanning residues #200–336 was capable of binding to DNA with preference to supercoils. Thus, amino acids 137–206 region was termed as NRD and the region of amino acids 200–336 as SRD.

The SRD does not contain any known motif for DNA binding but this domain is highly conserved among species and had been referred to as 'charged regions' in human LEDGF/p75 (23,39). Previous studies indicated that although the charged regions were not necessary for chromatin tethering of LEDGF/p75, CR2 (amino acids 198–266, amino acids 197–265 in rat sequence) did show

Figure 6. Continued
panels. In the panels designated 'GFP-DAPI', DAPI signal was subtracted pixel-by-pixel from the eGFP signal and the remaining signal is shown here after background subtraction (see Supplementary Figure S9 for detailed description). The region is outlined by yellow circles in other panels for reference. In panels E and F, relative signal intensities were scanned along the yellow lines drawn and plotted in the graphs shown on the bottom. All the merged images were made from SBP75/LEDGF (green) and H3K4me3 (magenta). Scale bars designate 5 μm except in the magnified images in panels C and D, which represent 1 μm.

some contribution to chromatin binding (23,39), which may reflect our finding that SRD core (amino acids 200–274) possesses a weak but supercoil-specific DNA binding ability. The C-terminal end of SRD core contains an evolutionally conserved serine cluster (residues 270, 272 and 274) that was reported to be involved in HIV-1 DNA integration, although mutations of these serine residues to alanine did not alter chromatin binding, nor integrase-to-chromatin-tethering capability of LEDGF/p75 (39).

The core region of SRD has a characteristic feature of alternating clusters of acidic (E/D) and basic (K) amino acids that are evolutionally conserved (Figure 4B). The basic–acidic–basic cluster in the center (K2–E/D2–K3) was suggested to be a primary element for the recognition and binding of supercoiled DNA since a polypeptide $K_9E_9K_9$ that mimics the cluster showed a DNA binding activity as potent as GST-SRD. Furthermore, a duplex oligonucleotide $G_{10}C_{10}G_{10}$ was inferred to be an effective binder of $K_9E_9K_9$ when interactions through hydrogen bonding in the major groove were postulated (40). As expected, $G_{10}C_{10}G_{10}$ duplex interfered with the preferential binding of GST-SRD to supercoiled DNA.

Involvement of clustered lysine residues in supercoiled DNA binding was suggested previously (43,44). A lysine-rich sequence (TKKKFKD) associated with the B domain of HMG-1 was a prerequisite for supercoiled DNA binding (43). Also, lysine residues immobilized on sepharose beads through the α-amino group was shown to recognize supercoiled plasmid DNA specifically and thus could be used as an affinity matrix to remove linear DNA (44). The basic lysine stretch in SRD, however, was not the only element for supercoiled DNA recognition because the polypeptide K_{27} did not show any preference to DNA superstructure. The acidic residues flanked by lysine clusters seem to play an important role in the recognitions of supercoils although acidic amino acid side chains are rather avoided in DNA binding domains, presumably because of unfavorable electrostatic interactions between the side chains and phosphate groups of DNA. In one report, however, glutamic acid/aspartic acid clusters in chromatin protein Sin1p/Spt2p were shown to bind supercoiled DNA in a preferential manner (45).

In linear form duplex DNA whose ends are unconstrained, GC-rich tetranucleotide segments show a strong tendency to have wider minor groove than AT-rich counterparts (46). When DNA is unwound by negative supercoiling, minor-groove width is increased (47). Thus, in the negatively supercoiled circular DNA, negative writhe could be accumulated locally in GC-rich regions that can accept the wide minor groove. In other words, these segments of DNA have narrower major groove as compared to linear DNA, which can be used as a characteristic signature for negative superhelical configuration. If this is the case, the lysine–guanine and glutamic/aspartic acid–cytosine interactions through hydrogen bonding in the major groove (40) may favor the major groove with narrower-than-normal width, which may be ubiquitous in negative supercoils. Use of minor groove binders with different base preference, GC-selective chromomycin A_3 and AT-selective distamycin A, gave clearly contrasting results. The preferential binding of SRD to supercoiled DNA was strongly interfered by chromomycin A_3 but not by distamycin A. These results not only support the involvement of SRD core in the supercoil recognition but also imply that SRD has a direct contact in the major groove of GC-rich regions that are scattered in negative supercoils. The reason why the binding of chromomycin to minor groove interferes with the binding of SRD in major groove (opposite side of the duplex) could be ascribed to a propagation effect: changes in the major groove size induced by the widening of minor groove by chromomycin binding (48) or induced perturbation of electron density around the ring system of bases, which may disfavor the hydrogen bonding with SRD in the major groove. There is no evidence so far suggesting that the interaction between SRD and supercoiled DNA is based on some unpaired bases or single-stranded regions in non-B DNA structures like cruciform, triplex, or quadruplex that are known to be stabilized under negative superhelicity.

We have shown in the present study that SRD appears to be a primary determinant for SBP75/LEDGF to locate at active transcription sites in nuclei, probably by recognizing superhelical DNA regions generated at transcription sites. A recent report confirmed that negative supercoils remain intact *in vivo* for substantial period of time even in the presence of DNA relaxing enzymes such as DNA topoisomerases (28). Both endogenous and transiently expressed SBP75/LEDGF were detected in H3K4me3-positive or Br-UTP-incorporated regions where transcription was supposed to be active. However, SBP75/LEDGF protein associated with H3K4me3-positive region was not relatively abundant and large population of the protein located in condensed chromatin regions, being consistent with previous studies (5,9,22). This is due to the fact that the N-terminal PWWP domain exerts a dominating effect on the protein's intranuclear localization (22,23). We confirmed this by using PWWP deletion mutants that diffusely distributed in the nucleus and were quite mobile upon fixation (Figure 6 and Supplementary Figure S8). Recently, the PWWP domain identified in several proteins was shown to be involved in recognition of methylated histones. The PWWP domain of Pdp1 (*Schizosaccharomyces pombe* protein PWWP domain protein 1) binds to histone H4 methylated on Lys20 (49). Histone H3 trimethylated on Lys36 is recognized by human BRPF1 (boromo and plant homeodomain finger-containing protein 1) (50) and by murine Dnmt3a DNA methyltransferase (51) through the interaction with their PWWP domains. Therefore, this domain seems to target either transcriptionally active (H3K36me3) or inactive (H4K20me) signatures, probably depending on its structural difference among each protein. It is not yet clear whether the PWWP domain of SBP75/LEDGF recognizes methylated histones or other unknown modifications enriched in heterochromatic regions.

The apparent discrepancy between the subnuclear distribution and the DNA-binding profiles of SBP75/LEDGF *in vivo* was also reported recently. When LEDGF fused to *E. coli* Dam methylase was expressed in HeLa cells, the protein showed a typical distribution pattern in the nucleus, being mostly overlapped with condensed chromatin regions that are supposed to have

a low transcriptional level. However, determination of DNA sites bound by the protein, as mapped in the ENCODE region by the DamID technique, revealed that most frequent sites are transcription units of active genes (52). The LEDGF-bound chromatin islands also correlated well with the HIV-1 integration sites (52). It is very likely that only a small fraction of SBP75/LEDGF is targeted to transcription active chromatin sites and the protein's ability to recognize superhelical structure would be a driving force for targeting (schematically shown in Figure 7). It is not clear at present whether the PWWP domain has a specific interaction partner at the tethered chromatin sites and why so much proportion of the protein is tethered there. If the affinity of SRD to supercoils were not strong enough, the protein would have to be untethered before interacting with active chromatin. The view is consistent with the recent report with a combination of real-time quantitative fluorescence techniques demonstrating that SBP75/LEDGF dynamically moves between scanning and locked states and PWWP domain is crucial for locking the partner-loaded protein on chromatin (53). After all the preferential recognition of and binding to supercoiled DNA by SBP75/LEDGF could be an underlying molecular mechanism of preferential integration of HIV-1 cDNA into active transcription units.

ACCESSION NUMBERS

DDBJ, AB285525.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Prof. Takamichi Hattori of Chiba University, Japan for providing sera from the patients with Miller–Fisher syndrome. Technical assistance of Kazuko Kiyama in some experiments is highly appreciated.

FUNDING

Grant-in-Aid for Scientific Research funding (grant number 16084201 to K.T.) and for Science and Technology funding (to K.M.T.) from the Ministry of Education, Science, Sports and Culture of Japan. Funding for open access charge: Commission funds to Okayama University.

Conflict of interest statement. None declared.

REFERENCES

1. Ochs,R.L., Muro,Y., Si,Y., Ge,H., Chan,E.K. and Tan,E.M. (2000) Autoantibodies to DFS 70 kd/transcription coactivator p75 in atopic dermatitis and other conditions. *J. Allergy Clin. Immunol.*, **105**, 1211–1220.
2. Ganapathy,V. and Casiano,C.A. (2004) Autoimmunity to the nuclear autoantigen DFS70 (LEDGF): what exactly are the autoantibodies trying to tell us? *Arthritis Rheum.*, **50**, 684–688.
3. Ge,H., Si,Y. and Roeder,R.G. (1998) Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation. *EMBO J.*, **17**, 6723–6729.
4. Singh,D., Ohguro,N., Kikuchi,T., Sueno,T., Reddy,V., Yuge,K., Chylack,L. Jr and Shinohara,T. (2000) Lens epithelium-derived growth factor: effects on growth and survival of lens epithelial cells, keratinocytes, and fibroblasts. *Biochem. Biophys. Res. Commun.*, **267**, 373–381.
5. Cherepanov,P., Maertens,G., Proost,P., Devreese,B., Van Beeumen,J., Engelborghs,Y., De Clercq,E. and Debyser,Z. (2003) HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J. Biol. Chem.*, **278**, 372–381.
6. Maertens,G., Cherepanov,P., Pluymers,W., Busschots,K., De Clercq,E., Debyser,Z. and Engelborghs,Y. (2003) LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *J. Biol. Chem.*, **278**, 33528–33539.
7. Llano,M., Vanegas,M., Fregoso,O., Saenz,D., Chung,S., Peretz,M. and Poeschla,E.M. (2004) LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. *J. Virol.*, **78**, 9524–9537.
8. Cherepanov,P., Devroe,E., Silver,P.A. and Engelman,A. (2004) Identification of an evolutionarily conserved domain in human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase. *J. Biol. Chem.*, **279**, 48883–48892.
9. Vanegas,M., Llano,M., Delgado,S., Thompson,D., Peretz,M. and Poeschla,E. (2005) Identification of the LEDGF/p75 HIV-1 integrase-interaction domain and NLS reveals NLS-independent chromatin tethering. *J. Cell Sci.*, **118**, 1733–1743.
10. Llano,M., Saenz,D.T., Meehan,A., Wongthida,P., Peretz,M., Walker,W.H., Teo,W. and Poeschla,E.M. (2006) An essential role for LEDGF/p75 in HIV integration. *Science*, **314**, 461–464.
11. Van Maele,B., Busschots,K., Vandekerckhove,L., Christ,F. and Debyser,Z. (2006) Cellular co-factors of HIV-1 integration. *Trends Biochem. Sci.*, **31**, 98–105.
12. Schroder,A.R., Shinn,P., Chen,H., Berry,C., Ecker,J.R. and Bushman,F. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. *Cell*, **110**, 521–529.
13. Wu,X., Li,Y., Crise,B. and Burgess,S.M. (2003) Transcription start regions in the human genome are favored targets for MLV integration. *Science*, **300**, 1749–1751.
14. Mitchell,R.S., Beitzel,B.F., Schroder,A.R., Shinn,P., Chen,H., Berry,C.C., Ecker,J.R. and Bushman,F.D. (2004) Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.*, **2**, e234.
15. Ciuffi,A., Llano,M., Poeschla,E., Hoffmann,C., Leipzig,J., Shinn,P., Ecker,J.R. and Bushman,F. (2005) A role for LEDGF/p75 in targeting HIV DNA integration. *Nat. Med.*, **11**, 1287–1289.
16. Shun,M.C., Raghavendra,N.K., Vandegraaff,N., Daigle,J.E., Hughes,S., Kellam,P., Cherepanov,P. and Engelman,A. (2007) LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev.*, **21**, 1767–1778.
17. Marshall,H.M., Ronen,K., Berry,C., Llano,M., Sutherland,H., Saenz,D., Bickmore,W., Poeschla,E. and Bushman,F.D. (2007) Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. *PLoS One*, **2**, e1340.
18. Maertens,G.N., Cherepanov,P. and Engelman,A. (2006) Transcriptional co-activator p75 binds and tethers the Myc-interacting protein JPO2 to chromatin. *J. Cell Sci.*, **119**, 2563–2571.
19. Yokoyama,A. and Cleary,M.L. (2008) Menin critically links MLL proteins with LEDGF on cancer-associated target genes. *Cancer Cell*, **14**, 36–46.
20. Bartholomeeusen,K., Christ,F., Hendrix,J., Rain,J.C., Emiliani,S., Benarous,R., Debyser,Z., Gijsbers,R. and De Rijck,J. (2009) Lens epithelium-derived growth factor/p75 interacts with the

transposase-derived DDE domain of PogZ. *J. Biol. Chem.*, **284**, 11467–11477.
21. Hughes,S., Jenkins,V., Dar,M.J., Engelman,A. and Cherepanov,P. (2010) Transcriptional co-activator LEDGF interacts with Cdc7-activator of S-phase kinase (ASK) and stimulates its enzymatic activity. *J. Biol. Chem.*, **285**, 541–554.
22. Turlure,F., Maertens,G., Rahman,S., Cherepanov,P. and Engelman,A. (2006) A tripartite DNA-binding element, comprised of the nuclear localization signal and two AT-hook motifs, mediates the association of LEDGF/p75 with chromatin in vivo. *Nucleic Acids Res.*, **34**, 1663–1675.
23. Llano,M., Vanegas,M., Hutchins,N., Thompson,D., Delgado,S. and Poeschla,E.M. (2006) Identification and characterization of the chromatin-binding domains of the HIV-1 integrase interactor LEDGF/p75. *J. Mol. Biol.*, **360**, 760–773.
24. Lavelle,C. (2007) Transcription elongation through a chromatin template. *Biochimie*, **89**, 516–527.
25. Champoux,J.J. (2001) DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem.*, **70**, 369–413.
26. Koster,D.A., Croquette,V., Dekker,C., Shuman,S. and Dekker,N.H. (2005) Friction and torque govern the relaxation of DNA supercoils by eukaryotic topoisomerase IB. *Nature*, **434**, 671–674.
27. Darzacq,X., Shav-Tal,Y., de Turris,V., Brody,Y., Shenoy,S.M., Phair,R.D. and Singer,R.H. (2007) In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.*, **14**, 796–806.
28. Kouzine,F., Sanford,S., Elisha-Feil,Z. and Levens,D. (2008) The functional response of upstream DNA to dynamic supercoiling in vivo. *Nat. Struct. Mol. Biol.*, **15**, 146–154.
29. Tsutsui,K., Tsutsui,K. and Muller,M.T. (1988) The nuclear scaffold exhibits DNA-binding sites selective for supercoiled DNA. *J. Biol. Chem.*, **263**, 7235–7241.
30. Tsutsui,K., Tsutsui,K., Sakurai,H., Shohmori,T. and Oda,T. (1986) Levels of topoisomerase II and DNA polymerase alpha are regulated independently in developing neuronal nuclei. *Biochem. Biophys. Res. Commun.*, **138**, 1116–1122.
31. Kawano,S., Miyaji,M., Ichiyasu,S., Tsutsui,K.M. and Tsutsui,K. (2010) Regulation of DNA Topoisomerase II beta through RNA-dependent association with heterogeneous nuclear ribonucleoprotein U (hnRNP U). *J. Biol. Chem.*, **285**, 26451–26460.
32. Tsutsui,K., Tsutsui,K., Okada,S., Watarai,S., Seki,S., Yasuda,T. and Shohmori,T. (1993) Identification and characterization of a nuclear scaffold protein that binds the matrix attachment region DNA. *J. Biol. Chem.*, **268**, 12886–12894.
33. Tsutsui,K., Tsutsui,K., Sano,K., Kikuchi,A. and Tokunaga,A. (2001) Involvement of DNA topoisomerase II beta in neuronal differentiation. *J. Biol. Chem.*, **276**, 5769–5778.
34. Wansink,D.G., Schul,W., van der Kraan,I., van Steensel,B., van Driel,R. and de Jong,L. (1993) Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. *J. Cell Biol.*, **122**, 283–293.
35. Tsutsui,K., Tsutsui,K., Hosoya,O., Sano,K. and Tokunaga,A. (2001) Immunohistochemical analyses of DNA topoisomerase II isoforms in developing rat cerebellum. *J. Comp. Neurol.*, **431**, 228–239.
36. Schroter,H., Maier,G., Ponstingl,H. and Nordheim,A. (1985) DNA intercalators induce specific release of HMG 14, HMG 17 and other DNA-binding proteins from chicken erythrocyte chromatin. *EMBO J.*, **4**, 3867–3872.
37. Quarles,R.H. and Weiss,M.D. (1999) Autoantibodies associated with peripheral neuropathy. *Muscle Nerve*, **22**, 800–822.
38. Hiraga,A., Kuwabara,S., Nakamura,A., Yuki,N., Hattori,T. and Matsunaga,T. (2007) Fisher/Gullain-Barre overlap syndrome in advanced AIDS. *J. Neurol. Sci.*, **258**, 148–150.
39. Garcia-Rivera,J.A., Bueno,M.T., Morales,E., Kugelman,J.R., Rodriguez,D.F. and Llano,M. (2010) Implication of serine residues 271, 273, and 275 in the human immunodeficiency virus type 1 cofactor activity of lens epithelium-derived growth factor/p75. *J. Virol.*, **84**, 740–752.
40. Luscombe,N.M., Laskowski,R.A. and Thornton,J.M. (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.*, **29**, 2860–2874.
41. Nishizawa,Y., Usukura,J., Singh,D.P., Chylack,L.T. Jr and Shinohara,T. (2001) Spatial and temporal dynamics of two alternatively spliced regulatory factors, lens epithelium-derived growth factor (ledgf/p75) and p52, in the nucleus. *Cell Tissue Res.*, **305**, 107–114.
42. Shun,M.C., Botbol,Y., Li,X., Di Nunzio,F., Daigle,J.E., Yan,N., Lieberman,J., Lavigne,M. and Engelman,A. (2008) Identification and characterization of PWWP domain residues critical for LEDGF/p75 chromatin binding and human immunodeficiency virus type 1 infectivity. *J. Virol.*, **82**, 11555–11567.
43. Stros,M. (2001) Two mutations of basic residues within the N-terminus of HMG-1 B domain with different effects on DNA supercoiling and binding to bent DNA. *Biochemistry*, **40**, 4769–4779.
44. Sousa,A., Sousa,F. and Queiroz,J.A. (2009) Biorecognition of supercoiled plasmid DNA isoform in lysine-affinity chromatography. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.*, **877**, 3257–3260.
45. Novoseler,M., Hershkovits,G. and Katcoff,D.J. (2005) Functional domains of the yeast chromatin protein Sin1p/Spt2p can bind four-way junction and crossing DNA structures. *J. Biol. Chem.*, **280**, 5169–5177.
46. Rohs,R., West,S.M., Sosinsky,A., Liu,P., Mann,R.S. and Honig,B. (2009) The role of DNA shape in protein-DNA recognition. *Nature*, **461**, 1248–1253.
47. Vologodskii,A.V. and Cozzarelli,N.R. (1994) Conformational and thermodynamic properties of supercoiled DNA. *Annu. Rev. Biophys. Biomol. Struct.*, **23**, 609–643.
48. Hou,M.H., Robinson,H., Gao,Y.G. and Wang,A.H. (2004) Crystal structure of the [Mg2+-(chromomycin A3)2]-d(TTGGCCA A)2 complex reveals GGCC binding specificity of the drug dimer chelated by a metal ion. *Nucleic Acids Res.*, **32**, 2214–2222.
49. Wang,Y., Reddy,B., Thompson,J., Wang,H., Noma,K., Yates,J.R. 3rd and Jia,S. (2009) Regulation of Set9-mediated H4K20 methylation by a PWWP domain protein. *Mol. Cell*, **33**, 428–437.
50. Vezzoli,A., Bonadies,N., Allen,M.D., Freund,S.M., Santiveri,C.M., Kvinlaug,B.T., Huntly,B.J., Gottgens,B. and Bycroft,M. (2010) Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. *Nat. Struct. Mol. Biol.*, **17**, 617–619.
51. Dhayalan,A., Rajavelu,A., Rathert,P., Tamas,R., Jurkowska,R.Z., Ragozin,S. and Jeltsch,A. (2010) The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J. Biol. Chem.*, **285**, 26114–26120.
52. De Rijck,J., Bartholomeeusen,K., Ceulemans,H., Debyser,Z. and Gijsbers,R. (2010) High-resolution profiling of the LEDGF/p75 chromatin interaction in the ENCODE region. *Nucleic Acids Res.*, **38**, 6135–6147.
53. Hendrix,J., Gijsbers,R., De Rijck,J., Voet,A., Hotta,J.I., McNeely,M., Hofkens,J., Debyser,Z. and Engelborghs,Y. (2010) The transcriptional co-activator LEDGF/p75 displays a dynamic scan-and-lock mechanism for chromatin tethering. *Nucleic Acids Res.*, doi:10.1093/nar/gkq933 (25 October 2010, date last accessed).