Journal of Cellular Biochemistry 98:1106–1122 (2006)

SP100B, a Repressor of Gene Expression Preferentially Binds to DNA With Unmethylated CpGs

Anne Isaac, Kent W. Wilcox, and Jerry L. Taylor*

Department of Microbiology and Molecular Genetics, Medical College of Wisconsin, Milwaukee, Wisconsin 53226

Abstract SP100A and SP100B are mammalian nuclear proteins encoded by alternatively-spliced transcripts from the *SP100* gene. The N-terminal portion of SP100B (aa 1–476) is identical to SP100A and contains an HP1 interaction domain. The C-terminal portion of SP100B (aa 477–688) contains an HMG2 interaction domain and a SAND domain. The SAND domain is a DNA-binding domain identified in several nuclear proteins involved in transcriptional regulation. We have previously reported that SP100B represses expression of genes present on transfected DNA in a SAND domain-dependent manner. The goal of the present study was to characterize the DNA binding properties of full-length SP100B expressed in mammalian cells. SP100B associated with DNA whereas SP100A did not. The SP100B SAND domain was essential for DNA binding. Deletion of the HP1- or HMG2-binding domain had no effect on DNA binding. SP100B preferentially associated with sequences containing CpG dinucleotides. Our results did not reveal any preference of SP100B for bases flanking CpG dinucleotides. The number of CpGs in a DNA sequence and spacing between CpGs influenced SP100B binding, suggesting that multimers of SP100B bind DNA in a cooperative manner. Binding of SP100B was abrogated by methylation of the cytosine residue within the context of the CpG dinucleotide. We propose that the preference of SP100B for non-methylated CpGs provides a mechanism to target SP100B to foreign DNA, including plasmid DNA or viral DNA genomes, most of which are hypomethylated. J. Cell. Biochem. 98: 1106–1122, 2006.

Key words: SP100; SAND domain; CpG methylation

SP100 was first identified as a protein immunoreactive with sera from patients suffering from the autoimmune disease, primary biliary cirrhosis [Szostecki et al., 1987, 1990, 1992]. To further characterize the SP100 autoantigen, Szostecki et al. [1990] isolated the constitutively expressed SP100A cDNA, which encodes a 480-amino acid (aa) protein. Guldner et al. [1999] observed that treatment of cells with interferon induces the expression of several splice variants from the *SP100* gene, including a transcript that codes for SP100B, a 688 aa protein. SP100B is essentially equivalent to SP100A (the first 476 aa are identical) plus a C-terminal extension containing a high mobility group protein (HMG2) interaction domain (aa 476–528) [Lehming et al., 1998], and a SAND domain (aa 595–688) [Bottomley et al., 2001] (Fig. 2A).

The SAND domain is a DNA binding domain found in the *Drosophila melanogaster* protein, deformed epidermal autoregulatory factor-1 (DEAF-1) [Gross and McGinnis, 1996] and in several mammalian nuclear proteins including the glucocorticoid modulatory element binding proteins GMEB1 (PIF96) and GMEB2 (PIF97) [Christensen et al., 1999; Jimenez-Lara et al., 2000; Kaul et al., 2000], some members of the SP100 family of proteins [Bloch et al., 1996, 2000; Bottomley et al., 2001], autoimmune regulator 1 (AIRE-1) [Gibson et al., 1998; Purohit et al., 2005], the nuclear DEAF-1-related protein NUDR [Huggenvik et al., 1998; Michelson et al., 1999], and two nuclear phosphoproteins, NucP41 and NucP75. Solution of the NMR structure of the SP100B SAND domain (aa 595–688) revealed an α/β fold

Grant sponsor: NIH to JLT; Grant numbers: DE14137, EY13546.

*Correspondence to: Jerry L. Taylor, Department of Microbiology and Molecular Genetics, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226. E-mail: jtaylor@mcw.edu

Received 8 December 2005; Accepted 4 January 2006

DOI 10.1002/jcb.20841

containing four α-helices that pack against a five-stranded antiparallel β-sheet [Bottomley et al., 2001]. The conserved KNWK (KDWK in DEAF-1, NUDR, GMEB1, and GMEB2) motif within the SP100B SAND domain is part of a positively charged surface formed by the α-helical region. Chemical shift perturbation experiments performed with the SP100B SAND domain and an oligonucleotide resembling the GMEB target sequence suggested that the DNA binding site maps to the positively charged region encompassing the KNWK motif. Site directed mutagenesis and in vitro binding studies indicate that the KDWK motif is essential for the ability of NUDR and GMEB1 to bind DNA [Bottomley et al., 2001; Surdo et al., 2003]. Most SAND domain proteins that have been examined do not show high sequence-specificity and appear to bind DNA with a relatively low affinity.

We have previously characterized SP100B as a repressor of viral and cellular promoters that are introduced into cells via transfection [Wilcox et al., 2005]. We have also observed that SP100B represses the expression of nuclear DNA viruses such as herpes simplex virus 1 (HSV1) and adenovirus 5 during an active infection (data not shown). The goal of the current study was to extend the characterization of SP100B by investigating the DNA binding properties of full-length SP100B expressed in mammalian cells. In this report, we have established that SP100B associates with plasmid DNA in a SAND domain-dependent manner. Point mutations within the KNWK domain abrogate DNA binding. SP100B exhibits a binding preference for sequences that contain CpG dinucleotides. Methylation of the cytosines in the CpG dinucleotides abrogates the binding of SP100B.

MATERIALS AND METHODS

Transfections

U2OS cells (ATCC HTB-96), a human osteosarcoma cell line, were grown as monolayers in Dulbecco's minimal essential medium (DMEM) supplemented with 10% donor bovine calf serum containing iron (Invitrogen, Carlsbad, CA) and 25 μg/ml gentamicin (Invitrogen). U2OS cells (1×10^6 cells per 60 mm dish) were transfected with DNA (up to 1 μg) and NovaFECTOR (8 μl per μg of DNA; VennNova, LLC, Pompano Beach, FL) in 2 ml of DMEM. The cells were incubated at 37°C for 6 h and thereafter supplemented with 2 ml of DMEM containing 20% bovine calf serum and 50 μg/ml gentamicin.

Plasmids and Oligonucleotides

Plasmids that use the CMV immediate early promoter to drive expression of proteins with an N-terminal 3XFLAG epitope were constructed from derivatives of the p3XFLAG-CMV-7.1 vector (Sigma, St. Louis, MO). Construction of plasmids that express 3XFLAG-SP100A (pXK1385), -SP100B (pXK1383), -SP100BΔ295-332 (pXK1368; ΔHP1), -SP100BΔ477-514 (pXK1465; ΔHMG2), -SP100BΔ594-688 (pXK1472), and -SP100BΔ 631-659 (pXK1384, ΔSAND) has been described previously [Wilcox et al., 2005]. The plasmid (pXK1919) expressing 3XFLAG-SP100B (ANWA) was constructed by deleting codons for SP100B residues 649–688 from pXK1383 and inserting two oligonucleotides in tandem, coding for residues 649–688 and containing the K653A and K656A mutations. The plasmid (pXK1577) expressing histone H2B with a C-terminal 3XFLAG epitope was constructed by conventional techniques using the histone H2B cDNA (pH2B-CYFP) that was a kind gift from Jan Ellenberg (Heidelberg, Germany). Construction of plasmid pXK859 that contains the HSV-1 UL26.5 promoter (−463 to +99) upstream of the luciferase gene has been described [Bruce and Wilcox, 2002].

Derivatives of pUC13 were constructed by insertion of defined oligonucleotides into the multiple cloning site. All oligonucleotides were purchased from Operon (Qiagen, Alameda, CA). Oligonucleotides L1, L2, and L3 were inserted in tandem into a derivative of pUC13 to construct the plasmid pXK1295. The insert, designated AA100, is similar to the sequence of the 100- bp ApaLI–AlwNI fragment from pUC13, with the exception of several base changes introduced to create a *Bam*HI site at the junction of L1 and L2 and an *Xma*I site at the junction of L2 and L3. Plasmids pXK1296 and pXK1297 containing two (2XAA100) and four copies (4XAA100), respectively, of the AA100 base pair sequence were constructed from pXK1295 using conventional cloning procedures.

The Hi-Lo DNA markers were obtained from Minnesota Molecular (Minneapolis, MN) as a mixture of linear *Eco*RI DNA fragments ranging in size from 50 to 10,000 bp. The 100-, 200-, 400-, 500-, 1,000-, and 2,000-bp markers were

cloned into the unique *Eco*RI site in a derivative of pUC13. Sequence analysis of the plasmids containing the Hi-Lo markers was performed followed by a BLAST search to identify the source of the sequences. Plasmid pHL101 containing the Hi-Lo 100-bp marker (HL100) was constructed as described above. Plasmids pHL102 and pHL104 containing two and four copies of HL100, respectively, were constructed by conventional cloning techniques. Plasmid DNA was purified from *E.coli* by the Qiagen Qiafilter (Valencia, CA) protocol and then centrifuged at 80,000*g* for 45 min at 14°C to remove insoluble material.

Oligonucleotides used in the DNA binding assays were designed to contain double-stranded molecules with either 5′ overhangs or blunt ends. Double-stranded methylated oligonucleotide, B7C was created by annealing two complementary single-stranded oligonucleotides in which 5-methyl cytosine (5 mC) was substituted for the cytosine in the context of the CpG dinucleotides. Hemimethylated oligonucleotides were created by annealing methylated and complementary, unmethylated oligonucleotides.

Radiolabeling Linear DNA With [^{32}P]-dATP

Double-stranded oligonucleotides containing 5′ overhangs were radiolabeled by filling in the 3′ ends with α^{32}P-dATP (3,000 Ci/mM; Perkin Elmer, Boston, MA) using the large fragment of DNA polymerase (5 units per 100 ng of DNA; Invitrogen) in a reaction buffer containing dNTPs (30 μM) lacking dATP, and 1× REact 2 buffer (Invitrogen). The reaction was carried out for 15 min at room temperature. The labeled probes were purified through a Probe Quant G50 column (Amersham Biosciences, UK). Plasmids containing the *Eco*RI inserts for DNA binding assays were digested with *Eco*RI. The cleaved vector and insert were radiolabeled as described above with the exception that the labeling reactions were not subjected to purification to remove the unincorporated dNTPs.

Radiolabeling Supercoiled Plasmid DNA With [^{3}H] (methyl) Thymidine

Supercoiled plasmid pXK859 labeled with [^{3}H] (methyl) thymidine was isolated from the temperature-sensitive thyA *E.coli* strain BW190 after growth at 40°C for 16 h in M9 minimal medium supplemented with 8 μM thymidine and 10 μCi/ml of [^{3}H] (methyl) thymidine (90 Ci/mmol; Perkin Elmer). Plasmid DNA was purified using the Biorad Midiprep kit (Hercules, CA) followed by phenol-chloroform extraction and ethanol precipitation. Plasmid DNAs with specific activities ranging from 10^5 to 10^6 dpm/μg were obtained.

Preparation of Soluble Nuclear Extracts

U2OS cells were transfected with a plasmid expressing the required FLAG-tagged protein. All subsequent manipulations were carried out at 4°C or on ice in solutions (volumes indicated were used for cells from a 60 mm dish) supplemented immediately before use with protease inhibitors PMSF (500 μM; Sigma), TPCK (500 μM; Sigma), leupeptin (1 μM; Boehringer Mannheim), and pepstatin A (1 μM; Sigma). Cells (60 mm dishes) were harvested at 24 h after transfection in phosphate-buffered saline (1 ml containing 0.54 mM EDTA) and pelleted (2,000 rpm, 5 min) in a microcentrifuge. The cells were suspended in hypotonic buffer (1 ml; 10 mM Tris-Cl, 10 mM NaCl, 1% Triton X-100, 10 mM β-mercaptoethanol) and incubated for 10 min, after which sucrose (200 μl of a 50% solution) was added. Nuclei were pelleted at 2,500 rpm for 5 min. The pelleted nuclei were suspended in IP buffer (0.5 ml; 50 mM Tris-Cl, 150 mM NaCl, 1% CHAPS, 10% glycerol, 10 mM β-mercaptoethanol) and lysed by sonication (5 s). Nuclear lysates were clarified by centrifugation at 12,000 rpm for 5 min. The supernatant (hereafter designated as the soluble nuclear extract) was used for co-immunoprecipitation assays.

Co-Immunoprecipitation Assays

Soluble nuclear extracts (0.5 ml) were supplemented with radiolabeled DNA or a mixture of radiolabeled DNA and competitor DNA as indicated in the figure legends. Anti-FLAG beads (60 μl of a 20% slurry; EZview Anti-FLAG M2 Affinity gel; Sigma) or protein G beads (no antibody control; 60 μl of a 20% slurry of Protein G Sepharose 4 Fast Flow; Amersham Pharmacia, Uppsala, Sweden) were added immediately and the mixture was gently rocked at 4°C for 24 h. The beads were pelleted by centrifugation at 2,000 rpm for 2 min and washed three times with ice-cold IP buffer (1 ml/wash). Exceptions to this procedure are noted in the figure legends. Radiolabeled DNA was recovered from the pellet and analyzed as described in the figure legends.

RESULTS

SP100B Associates With Transfected Plasmid DNA

To investigate the hypothesis that SP100B represses gene expression by binding to transfected plasmid DNA, we have characterized the DNA binding properties of SP100B. Bottomley et al. [2001] expressed the SP100B SAND domain (aa 595–688) in *E.coli* and reported that this domain could bind sequences in DNA that resemble the GMEB target sequence. Because Bottomley et al. [2001] did not exclude the possibility that other domains in SP100B contribute to DNA binding specificity or affinity, it was important to investigate the DNA binding properties of full-length SP100B. However, we were unable to express full length SP100B (aa 1–688) in *E.coli*. Therefore, we decided to investigate the DNA binding properties of full-length SP100B expressed in transfected mammalian cells. Potential advantages of this approach are that (1) post-translational modifications of SP100B that may affect its DNA binding efficiency and/or specificity will be present and (2) cellular co-factors that may be required for the efficient association of SP100B with DNA will be present. For each of the DNA binding assays described in this report, variants of SP100B tagged with the 3XFLAG epitope were expressed from plasmids transfected into U2OS cells. Soluble nuclear extracts prepared from these transfected cells were used as a source of SP100B in the DNA binding assays.

Among the plasmids demonstrated to be repressed by SP100B is a plasmid (pXK859) that expresses luciferase from the HSV-1 UL26.5 promoter [Wilcox et al., 2005]. To test whether SP100B binds pXK859, we co-transfected U2OS cells with [^{3}H]-pXK859 and a plasmid expressing either FLAG-SP100B, FLAG-SP100A or histone H2B-FLAG. Soluble nuclear extracts were prepared 24 h after transfection. The FLAG-tagged proteins were immunoprecipitated with anti-FLAG antibody using a modified version of the McKay immunobinding assay [McKay, 1981] and the amount of [^{3}H]-pXK859 that co-immunoprecipitated was determined by scintillation counting. The results (Fig. 1) revealed that 23% of the [^{3}H]-pXK859 in the extract co-immunoprecipitated with histone H2B-FLAG, a well-characterized protein that binds DNA non-specifically. In comparison, 15% of the [^{3}H]-pXK859 co-immunoprecipitated with FLAG-SP100B. FLAG-SP100A, which does not contain the SAND domain, did not associate with [^{3}H]-pXK859 in this assay (Fig. 1). These results demonstrate that full-length SP100B expressed in mammalian cells associates with transfected plasmid DNA whereas SP100A does not.

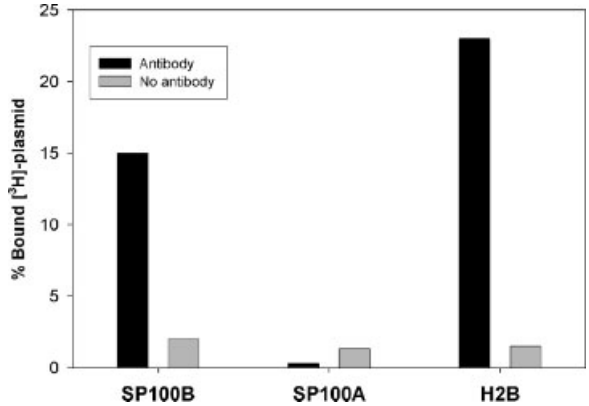

Fig. 1. Co-immunoprecipitation of transfected plasmid DNA with FLAG-tagged proteins. U2OS cells were co-transfected with [^{3}H]-pXK859 (0.5 μg) and a plasmid (1 μg) encoding FLAG-SP100B, FLAG-SP100A, or histone H2B-FLAG. Soluble nuclear extracts (0.5 ml) were prepared as described in Materials and Methods. The amount of [^{3}H]-pXK859 in an aliquot (0.1 ml) of the soluble nuclear extract was determined by scintillation counting. Immunoprecipitations were carried out using anti-FLAG beads (antibody) or protein G beads (no antibody) as described in Materials and Methods. The beads were suspended in liquid scintillation cocktail (Beckman Coulter, Fullerton, CA) and the [^{3}H]-plasmid was quantified by scintillation counting. The amount of co-immunoprecipitated [^{3}H]-plasmid is plotted as a percentage of the total [^{3}H]-plasmid in the soluble nuclear extract.

The HP1 and HMG2 Binding Domains do not Contribute to the Association of SP100B With DNA

Lehming et al. [1998] reported that the non-histone chromatin associated proteins HP1 and HMG2 interact with specific domains within a variant of SP100B (aa 5–528). HMG2 binds DNA in a sequence-independent manner and induces structural distortions in the DNA that facilitate the formation of nucleoprotein complexes [Thomas and Travers, 2001]. Heterochromatin protein 1 (HP1) plays a central role in creating a stable heterochromatic network [Eissenberg and Elgin, 2000]. Lehming et al. [1998] hypothesized that HMG2 recruits SP100B to DNA, which, in turn, recruits HP1 to form a complex on the DNA. To determine if the HP1 and HMG2 interaction domains in SP100B (Fig. 2A) influence the DNA binding

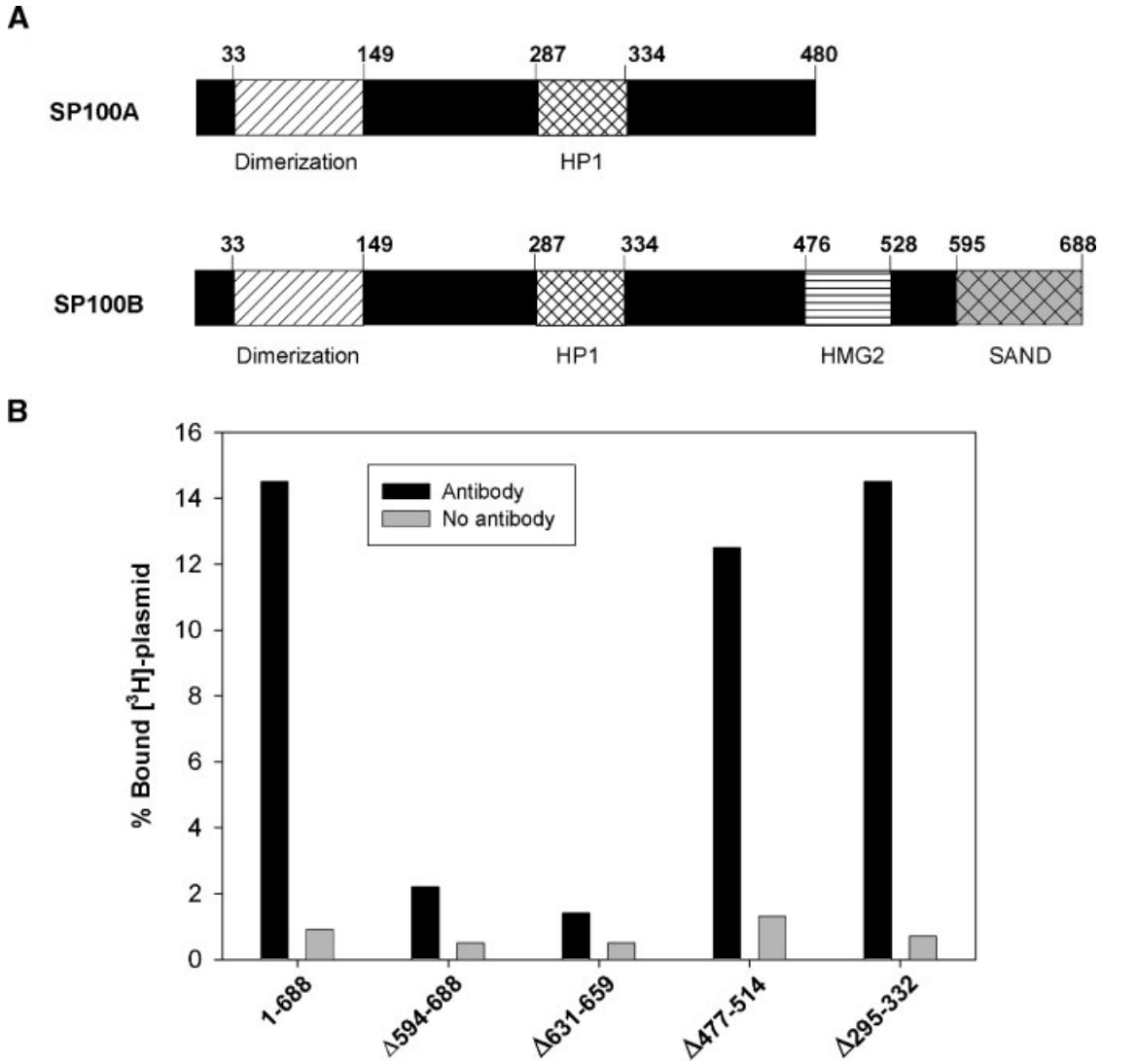

Fig. 2. Association of SP100B deletion variants with transfected plasmid DNA. **A**: Schematic representation of domains in SP100A and SP100B. The numbers correspond to amino acid positions. HMG2, high mobility group 2 binding domain; HP1, heterochromatin protein 1 binding domain; SAND, DNA binding domain. **B**: Soluble nuclear extracts were prepared from U2OS cells that were co-transfected with [^{3}H]-pXK859 (0.5 μg) and a plasmid (1 μg) encoding FLAG-SP100B (1–688), -SP100BΔSAND (-SP100BΔ594-688, -SP100BΔ631-659), -SP100BΔHMG2(Δ477-514) or -SP100BΔHP1(Δ295-332). Immunoprecipitations were carried out using anti-FLAG beads (antibody) or protein G beads (no antibody) as described in Materials and Methods. The beads were suspended in liquid scintillation cocktail and the [^{3}H]-plasmid was quantified by scintillation counting. The amount of co-immunoprecipitated [^{3}H]-plasmid is plotted as a percentage of the total [^{3}H]-plasmid in the soluble nuclear extract.

activity of SP100B, we tested the ability of variants lacking the HP1 or HMG2 binding domains to bind DNA. Immunofluoresence and immunoblot assays revealed that each of the SP100B variants localized to the nucleus and was expressed at a similar or higher level than full-length SP100B (data not shown). Both the ΔHMG2 (SP100BΔ477-514) and the ΔHP1 variants (SP100BΔ295-332) associated with [^{3}H]-pXK859 as efficiently as full-length SP100B (Fig. 2B). The ΔHP1 variant also lacks the sumoylation site (K297) identified in SP100A [Sternsdorf et al., 1999], indicating that sumoylation at residue K297 is not required for DNA binding. The variant with a deletion within the SAND domain (SP100BΔ631-659) or the variant lacking the SAND domain (SP100BΔ594-688) did not associate with [^{3}H]-pXK859 (Fig. 2B). These results indicate that neither the HP1 nor the HMG binding domain is

required for the association of SP100B with DNA, whereas the SP100B SAND domain is essential.

SP100B Associates With Bacterial Plasmid DNA, but does not Associate With Calf Thymus DNA

Given that SP100B binds to pXK859 and represses expression of luciferase driven by the UL26.5 promoter in pXK859, we hypothesized that SP100B binds to a specific sequence within the UL26.5 promoter or the luciferase gene. To test this hypothesis, we compared the ability of pXK859 or pUC13 (the vector backbone of pXK859 lacking both the UL26.5 promoter and the luciferase gene) to compete with [^{3}H]-pXK859 for binding to SP100B. The competition assays were performed by adding a mixture of [^{3}H]-pXK859 (0.05 μg) and cold competitor DNA (1–16 μg) to soluble nuclear extracts prepared from cells transfected with a plasmid that expresses FLAG-SP100B. The co-immunoprecipitation results revealed that pXK859 and pUC13 were equally effective as competitors (Fig. 3A). This result suggests that either both pXK859 and pUC13 contain binding sites for SP100B or that SP100B binds non-specifically to both plasmids.

To test the hypothesis that SP100B binds non-specifically to double-stranded DNA, we conducted competition assays with [^{3}H]-pXK859 as the target and calf thymus DNA (Worthington Biochemical Co. Lakewood, NJ) as the competitor. We were surprised to observe that the addition of calf thymus DNA resulted in a 1.5-fold increase in the amount of [^{3}H]-pXK859 bound to FLAG-SP100B (Fig. 3B). Even with a 300-fold excess of calf thymus DNA, no competition was observed. One explanation for the observation that linear calf thymus DNA does not compete with supercoiled [^{3}H]-pXK859 for binding to SP100B is that SP100B does not bind linear DNA. A direct test of the binding of SP100B to linear DNA is presented in Figure 4, showing that SP100B can associate with linear DNA. A second explanation is that calf thymus DNA lacks binding sites for SP100B. Given the sequence complexity of calf thymus DNA, we considered this unlikely, but could not rule out the possibility that the SP100B binding sequence is either present in low abundance or is modified in some way that prevents binding by SP100B. Data that support the latter interpretation are presented in Figure 8.

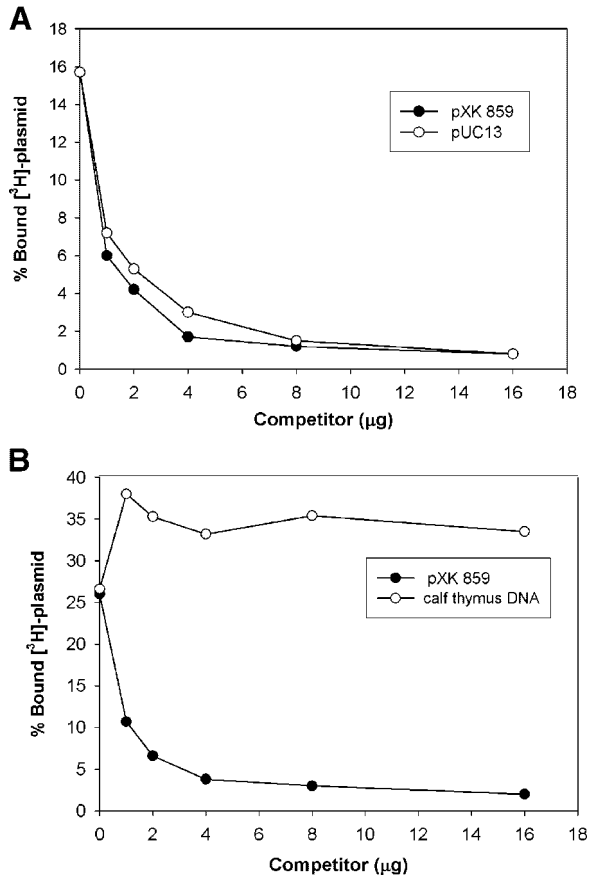

Fig. 3. Association of SP100B with plasmid DNA and calf thymus DNA. Soluble nuclear extracts (0.5 ml) prepared from U2OS cells transfected with a plasmid (0.25 μg) encoding FLAG-SP100B were supplemented with a mixture containing [^{3}H]-pXK859 (0.05 μg) and the indicated amount (1–16 μg) of unlabeled competitor DNA. **A**: Supercoiled plasmid pXK859 or pUC13; **B**: Supercoiled plasmid pXK859 or linear calf thymus DNA. Immunoprecipitations were carried out using anti-FLAG beads as described in Materials and Methods. The beads were suspended in liquid scintillation cocktail. The [^{3}H]-plasmid was quantified by scintillation counting. The amount of co-immunoprecipitated [^{3}H]-plasmid is plotted as a percentage of the total [^{3}H]-plasmid in the soluble nuclear extract.

SP100B Binds Linear DNA Fragments Selectively

To further investigate the DNA binding properties of SP100B, we used a modified co-immunoprecipitation assay in which linear ^{32}P-DNA (either restriction-enzyme fragments or synthetic double-stranded oligonucleotides) was added as the target DNA to soluble nuclear extracts prepared from cells transfected with a plasmid that expresses FLAG-SP100B. In order to determine whether SP100B associates with linear DNA, and if so, to determine whether SP100B exhibits a preference for specific sequences, we conducted co-immunoprecipitation

assays with a mixture of 15 linear DNA fragments ranging in size from 50 to 10,000 bp. This mixture of DNA fragments is commercially available as Hi-Lo markers that are derived from a variety of sources including a bacterial virus (phage fd), a plasmid and an insect virus (baculovirus) (Table I). The markers are distributed by the supplier as a mixture of *Eco*RI restriction fragments excised from a vector after propagation in *E. coli*. The supplier adjusts the mixture so that in general, the molar amount of each marker decreases with increasing size (Table I).

For the co-immunoprecipitation assays with these linear DNAs, the Hi-Lo markers were end-labeled with [^{32}P] and added to soluble nuclear extracts from cells transfected with plasmids that express FLAG-SP100B variants.

TABLE I. The Size, Source, and Molar Amounts of Hi-Lo Markers Used in the DNA Binding Assay

Marker (bp)	Source	*pmoles/250 ng of marker mix
6,000	NS	0.003
4,000	NS	0.005
3,000	NS	0.01
2,000	pUC	0.03
1,550	NS	0.025
1,400	NS	0.03
1,000	Baculovirus	0.05
750	NS	0.015
500	Baculovirus	0.05
400	Baculovirus + pUC	0.02
300	*E.coli* LacZ	0.05
200	pUC + Phage fd	0.06
100	pUC	0.075
50	NS	0.125

NS, not sequenced.
*Hi-Lo marker mix (250 ng) was used in the assays in Figure 4A.

Radiolabeled DNAs recovered from the immunoprecipitate were analyzed by gel electrophoresis and detected by autoradiography. Since each fragment was end labeled, the intensity of the bands in the autoradiogram is proportional to moles of DNA rather than mass. Thus, in the input DNAs, the intensities of the 50-, 100-, and 200- bp markers were much greater than the intensities of the larger markers (Fig. 4A). The autoradiogram revealed that the linear markers co-immunoprecipitated with FLAG-SP100B compared to the negative controls (FLAG-

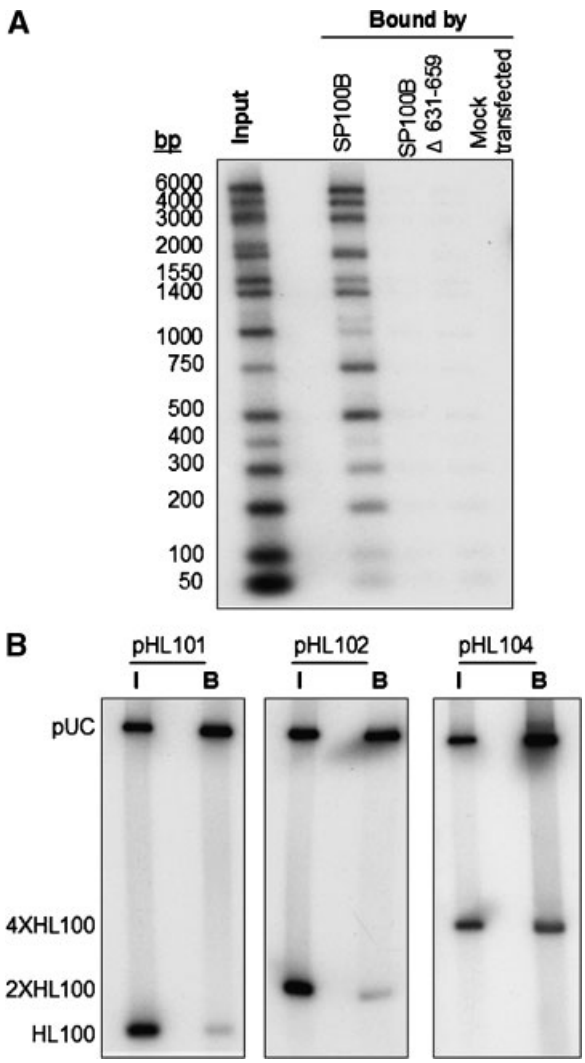

Fig. 4. Association of SP100B with linear DNA fragments (Hi-Lo markers). **A**: Soluble nuclear extracts (0.5 ml) prepared from U2OS cells that were either mock transfected or transfected with a plasmid (0.5 μg) expressing the indicated FLAG-SP100B variant were supplemented with the radiolabeled Hi-Lo marker DNAs (0.25 μg). Immunoprecipitations were carried out using anti-FLAG beads as described in Materials and Methods. The bound [^{32}P]-DNA was recovered from the beads in 30 μl of elution buffer (50 mM Tris-Cl, 10 mM EDTA, 1% SDS). A sample (2.5 ng) of the input [^{32}P]-DNA and a fraction (50%) of the bound DNA were electrophoresed on a 1.5% agarose gel. The agarose gel was fixed in 10% trichloroacetic acid for 30 min at 4°C and then dried at 50°C on Whatmann filter paper. The [^{32}P]-DNA fragments were detected by autoradiography. **B**: Plasmids containing one (pHL101), two (pHL102), or four (pHL104) copies of HL100 flanked by *Eco*RI sites were digested with *Eco*RI. The cleaved vector and insert were radiolabeled with [^{32}P]-dATP. The radiolabeled fragments (0.25 μg) were added to soluble nuclear extracts (0.5 ml) prepared from U2OS cells transfected with a plasmid (0.5 μg) expressing FLAG-SP100B. Immunoprecipitations were carried out using anti-FLAG beads as described in Materials and Methods. The bound [^{32}P]-DNA was recovered from the beads in 30 μl of elution buffer. A sample (1.6 ng) of the input (I) [^{32}P]-DNA and a fraction (50%) of the bound DNA (B) were electrophoresed on a 1.5% agarose gel. The agarose gel was analyzed as described in the legend (A).

SP100BΔSAND or untransfected cells) (Fig. 4A). We conclude that SP100B binds to linear DNA from different sources. The results also revealed that the DNA markers co-immunoprecipitated with SP100B with varying efficiencies. Each marker in the immunoprecipitate was classified as equally-represented, over-represented, or under-represented, based on its intensity relative to the intensity of the corresponding marker in the input. The 500-bp marker and most of the markers greater than 2,000 bp in size were equally-represented in the input and the immunoprecipitate. Relative to the intensities in the input DNA, the 750-bp marker was over-represented in the immunoprecipitate, whereas the 1,000-bp marker and markers smaller than 500 bp were under-represented in the immunoprecipitate (Fig. 4A). These results suggest that SP100B associates with DNA with some degree of sequence specificity.

One explanation for the under-representation of the smaller fragments compared to the larger fragments in the immunoprecipitate is that the larger fragments have several SP100B binding sites and, hence, bind to several molecules of SP100B, whereas SP100B binding sites are under-represented in the smaller fragments such as the 100- bp marker (designated here as HL100). If SP100B binding sites are under-represented in HL100, then a fragment containing multiple copies of HL100 should contain more binding sites and bind more efficiently to SP100B than a single copy. To compare the binding of SP100B to fragments containing one or more copies of HL100, we constructed plasmids containing one (pHL101), two (pHL102), or four (pHL104) copies of HL100 inserted in a head-to-tail fashion between *Eco*RI sites in a derivative of pUC13. The linear pUC vector and excised *Eco*RI fragments were end-labeled with [^{32}P]-dATP and added to nuclear lysates from cells expressing FLAG-SP100B. The results revealed that the fragments containing four copies of HL100 co-immunoprecipitated with SP100B more efficiently than fragments with one or two copies of HL100 (Fig. 4B). The results suggest that DNAs with multiple SP100B binding sites bind SP100B more efficiently.

SP100B Associates With Multiple Sites on pUC13

The simplest explanation for the efficient co-immunoprecipitation of pUC13 with SP100B compared to DNA fragments with a single or multiple copies of HL100 (Fig. 4B), is that there are multiple binding sites for SP100B distributed throughout pUC13. To test this hypothesis, we conducted co-immunoprecipitation assays with fragments derived from pUC13 by digestion with a variety of restriction endonucleases. The results (data not shown) revealed that SP100B binding sites are distributed throughout pUC13. In order to characterize the sequences that SP100B preferentially binds, we focused our attention on a 100-bp ApaLI–AlwNI fragment from pUC13 that co-immunoprecipitated with SP100B relatively well, given its size. Three oligonucleotides (L1, L2, and L3) spanning the 100-bp ApaLI–AlwNI region (Fig. 5A) were inserted in tandem between two *Eco*RI sites in a derivative of pUC13. The insert, designated AA100, is similar to the sequence of the 100-bp ApaLI–AlwNI fragment from pUC13, with the exception of several base changes introduced to create a *Bam*HI site at the junction of L1 and L2 and an *Xma*I site at the junction of L2 and L3. In order to increase the size and potential binding efficiency of the target DNA, a plasmid containing four head-to-tail copies of AA100 (4XAA100) flanked by *Eco*RI sites was constructed (Fig. 5B). The pUC vector and excised *Eco*RI fragments were end-labeled with [^{32}P]-dATP and added to soluble nuclear extracts from cells expressing FLAG-SP100B. As expected, immunoprecipitation of FLAG-SP100B resulted in the co-immunoprecipitation of both the pUC vector and AA100 (Fig. 5C). Immunoprecipitation of FLAG-SP100B also resulted in the co-immunoprecipitation of 4XAA100 (Fig. 5D). To determine whether the conserved KNWK motif within the SP100B SAND domain is required for binding to 4XAA100, we constructed an SP100B variant in which the lysines in the conserved KNWK motif were mutated to alanines (ANWA). The amount of 4XAA100 that co-immunoprecipitated with the FLAG-SP100 BΔSAND variants (SP100BΔ631-659 and SP100BΔ594-688) or FLAG-SP100B (ANWA) was much less than the amount that co-immunoprecipitated with wild type FLAG-SP100B and was similar to the background amount obtained with cells that express FLAG-SP100A (Fig. 5D). These results revealed that FLAG-SP100B binds 4XAA100 efficiently and enabled us to further investigate the preference of SP100B for sequences within this defined target.

SP100B Requires a CpG Dinucleotide for Binding

In the next series of experiments, we sought to identify the sequence(s) in AA100 to which SP100B preferentially binds. The fact that the SAND domain proteins NUDR, DEAF-1, and GMEB bind DNA sequences that contain CpG dinucleotides [Gross and McGinnis, 1996; Christensen et al., 1999; Michelson et al., 1999; Bottomley et al., 2001] prompted us to investigate if SP100B exhibits a similar DNA binding specificity. AA100 contains 10 CpGs (Fig. 5A). To explore the possibility that CpGs are required for binding, we compared the abilities of oligonucleotides L1, L2, or L3 and the

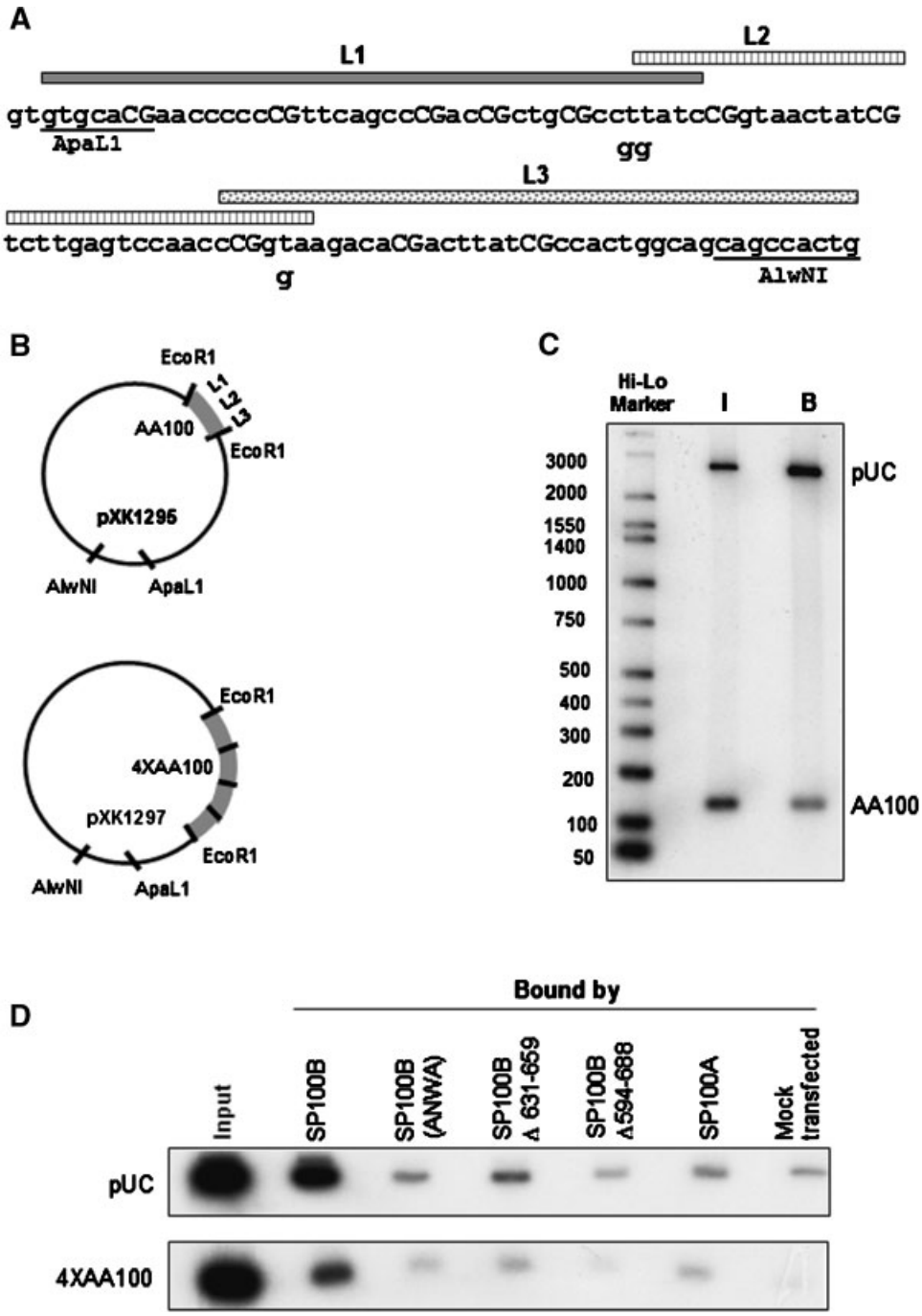

Fig. 5.

corresponding ΔCpG variants (CpGs were replaced with different nucleotides) to compete with [^{32}P]-4XAA100 for binding to SP100B (Fig. 6A). The amounts of co-immunoprecipitated [^{32}P]-4XAA100 and pUC were determined by Phosphoimager analysis. The co-immunoprecipitation assays revealed that there were 10- and 30-fold decreases in the amounts of [^{32}P]-4XAA100 and [^{32}P]-pUC, respectively, bound to FLAG-SP100B in the presence of L1 (7 CpGs) (Fig. 6B). There was approximately a 10-fold decrease in the amounts of each fragment that co-immunoprecipitated with FLAG-SP100B in the presence of L2 (4 CpGs) and L3 (3 CpGs). In contrast, there was approximately twofold less [^{32}P]-4XAA100 and [^{32}P]-pUC bound to FLAG-SP100B in the presence of the ΔCpG oligonucleotides (Fig. 6B). The results indicate that CpG-containing oligonucleotides compete efficiently with AA100 for binding to FLAG-SP100B, whereas the ΔCpG oligonucleotides do not, suggesting that SP100B binds preferentially to CpG containing DNA sequences.

Binding of SP100B to DNA Is Dependent on CpG Number and Spacing

Through competition assays, SP100B was demonstrated to associate with oligonucleotide L1, which contains five CpGs (designated A, B, C, D, and E) present within the 100-bp ApaL1-AlwNI pUC13 fragment and two additional CpGs within restriction sites introduced to facilitate cloning (Fig. 7A). To examine through direct binding assays whether SP100B preferentially binds to DNAs that contain CpG dinucleotides, we compared the binding of SP100B to [^{32}P]-labeled double-stranded oligonucleotides. Oligonucleotides used were derivatives of oligonucleotide L1 that contain CpGs at positions B and C (B7C), one CpG at position B only (B), one CpG at position C only (C), and a derivative of B7C in which both CpGs were changed to ApGs (ΔB7C) (Fig. 7A). Control immunoprecipitations were performed using soluble nuclear extracts from cells transfected with an empty FLAG-vector to establish the amount of oligonucleotide that co-immunoprecipitated non-specifically. The results revealed that SP100B bound most efficiently to oligonucleotide B7C (Fig. 7B). The amounts of oligonucleotides B and C recovered from the immunoprecipitates were twofold above the non-specific amount, suggesting that SP100B binds inefficiently to target sites that contain a single CpG. As expected, SP100B did not bind detectably to ΔB7C (Fig. 7B). We conclude that SP100B preferentially binds to sequences containing more than one CpG dinucleotide.

Two simple explanations for the efficient recovery of oligonucleotide B7C from the immunoprecipitates, compared to oligonucleotides B or C, are (1) SP100B binds independently to each CpG, thereby increasing the probability that an oligonucleotide containing two CpGs will be co-immunoprecipitated compared to an oligonucleotide containing only one CpG; or (2) SP100B binds as a homodimer or in a cooperative fashion to two properly-spaced CpGs, thereby increasing the stability of the complex. If explanation (1) is correct, then increasing the spacing between the two CpGs should have no affect on the amount of oligonucleotide that is recovered from the immunoprecipitate. In contrast, if explanation (2) is correct, then increasing the CpG spacing beyond the span of the

Fig. 5. Binding of SP100B to sequences from the pUC13 plasmid. **A**: The sequence of the 100- bp ApaLI–AlwNI region from pUC13 is shown. The ApaLI and AlwNI cleavage sites are underlined. The boxes represent the oligonucleotides L1, L2, and L3 that span the ApaLI–AlwNI region and were used to construct AA100. The CpG dinucleotides are in uppercase. The bases (bold) below the sequence indicate changes in the sequence that were introduced to facilitate linking of L1 to L2 and L2 to L3. **B**: Schematic representation of plasmids pXK1295 and pXK1297 containing one and four head-to-tail copies of AA100 (L1, L2, and L3 in tandem), respectively, flanked by *Eco*RI sites in a derivative of pUC13. **C**: Plasmid pXK1295 was digested with *Eco*RI. The cleaved vector and insert (AA100 plus approximately 35 flanking base pairs introduced to facilitate cloning) were radiolabeled with [^{32}P]-dATP. The radiolabeled fragments (0.25 μg) were added to soluble nuclear extracts (0.5 ml) prepared from U2OS cells transfected with a plasmid (0.5 μg) expressing FLAG-SP100B variants. Immunoprecipitations were carried out using anti-FLAG beads as described in Materials and Methods. The bound [^{32}P]-DNA was recovered from the beads in 30 μl of elution buffer. A sample (0.80 ng) of the input (I) and a fraction (50%) of the bound DNA (B) were electrophoresed on a 1.5% agarose gel. The agarose gel was analyzed as described in the legend for Figure 4A. **D**: Plasmid pXK1297 was digested with *Eco*RI. The cleaved vector and insert (4XAA100) were radiolabeled with [^{32}P]-dATP. The radiolabeled fragments (0.25 μg) were added to soluble nuclear extracts (0.5 ml) prepared from U2OS cells transfected with a plasmid (0.5 μg) expressing FLAG-SP100B variants. Immunoprecipitations were carried out using anti-FLAG beads as described in Materials and Methods. The bound [^{32}P]-DNA was recovered from the beads in 30 μl of elution buffer. A sample (3.7 ng) of the input and a fraction (50%) of the bound DNA were electrophoresed on a 1.5% agarose gel. The agarose gel was analyzed as described in the legend for Figure 4A.

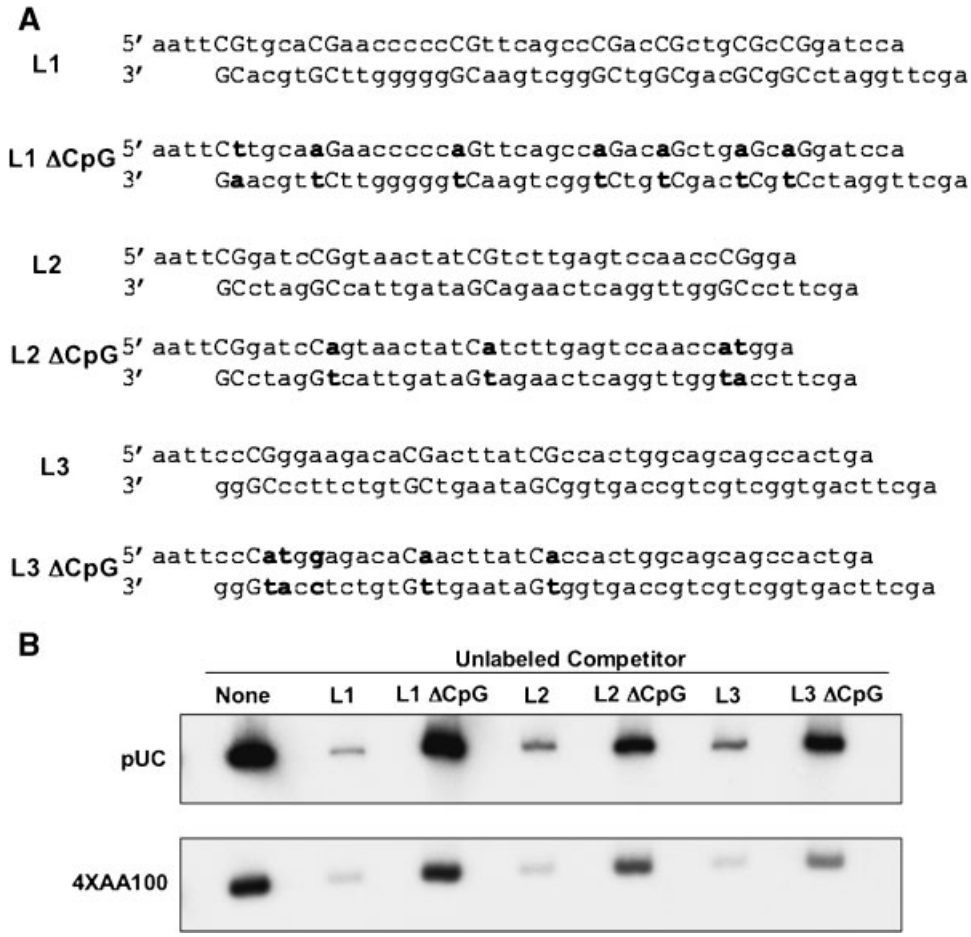

Fig. 6. SP100B associates with CpG dinucleotides. **A**: The sequences of oligonucleotides L1, L2, and L3 and the corresponding ΔCpG variants that lack CpG dinucleotides are shown. L2ΔCpG retained one CpG dinucleotide within a *Bam*HI restriction site used for cloning. **B**: Soluble nuclear extracts (0.5 ml) prepared from U2OS cells transfected with a plasmid (0.5 μg) expressing FLAG-SP100B were supplemented with a mixture containing [^{32}P]-DNA prepared from an *Eco*RI digest of plasmid pXK1297 (0.12 pmoles) and the indicated unlabeled competitor oligonucleotide (975 pmoles). Immunoprecipitations were carried out using anti-FLAG beads as described in Materials and Methods. The bound [^{32}P]-DNA was recovered from the pellet in 30 μl of elution buffer. A fraction (50%) of the bound DNA was electrophoresed on a 1.5% agarose gel. The agarose gel was analyzed as described in the legend for Figure 4A.

homodimer will result in a decrease in the stability of the complex. To distinguish between these two possibilities, we compared the ability of SP100B to bind to ^{32}P-oligonucleotides containing two CpGs separated by 4 (B4C), 7 (B7C), 11 (B11C), or 20 (B20C) base pairs. The results of the co-immunoprecipitation clearly revealed that ^{32}P-B11C and B20C bound less efficiently to SP100B than B4C or B7C (Fig. 7C). The binding observed with B11C and B20C was equivalent to binding with oligonucleotides containing only one CpG (data not shown). These results indicate that spacing between the CpGs plays an important role in the binding efficiency of SP100B to the DNA. This result suggests that SP100B binds properly-spaced CpGs either as a homodimer or in a co-operative fashion. The observation that binding of SP100B to oligonucleotide B4C was slightly less efficient than to oligonucleotide B7C, but more efficient than binding to oligonucleotides B11C and B20C, suggests that SP100B is somewhat flexible with respect to CpG spacing.

CpG Methylation Suppresses Binding of SP100B to DNA

DNA binding by many proteins that associate with CpG dinucleotide-containing sequences is blocked by methylation of the cytosine residue. To determine if methylation of cytosine influences binding of SP100B to DNA, we compared the abilities of unmethylated, methylated, or

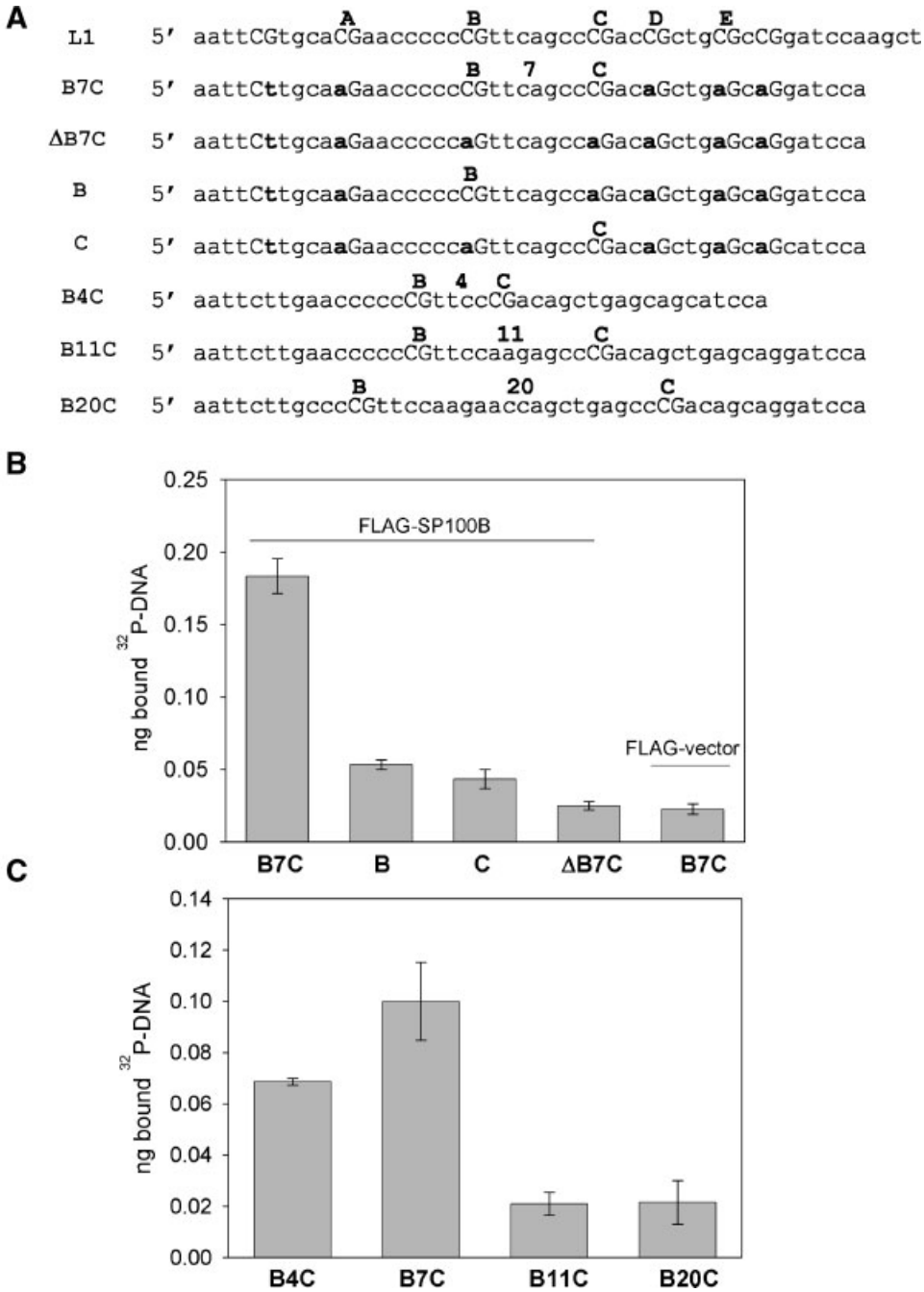

Fig. 7. Binding of SP100B to DNA is dependent on CpG density and spacing. **A**: Sequences of one strand of the double-stranded oligonucleotides used as radiolabeled targets in the DNA binding assays are shown. A, B, C, D, and E denote the CpG dinucleotides (uppercase) present within the 100- bp ApaLI–AlwNI fragment from pUC13. Sequences in bold show the changes made to eliminate CpG dinucleotides. The numbers in bold indicate the number of base pairs between two CpGs. **B**: Soluble nuclear extracts (0.5 ml) prepared from U2OS cells transfected with the 3XFLAG-vector (1 μg; p3XFLAG-CMV-7.1), or a plasmid (1 μg) encoding FLAG-SP100B were supplemented with a mixture of the indicated [^{32}P]-oligonucleotide (10 ng) and calf thymus DNA (2 μg). The extracts were mixed for 1 h at 4°C. Anti-FLAG beads (60 μl of a 20% slurry) were added and the mixture was incubated for an additional hour at 4°C. The beads were pelleted by centrifugation at 2,000 rpm for 2 min and washed three times with 1 ml of IP buffer. The beads were suspended in liquid scintillation cocktail and the [^{32}P]-oligonucleotides were quantified by scintillation counting. Immunoprecipitations were performed in triplicate and the error bars represent the standard error of the mean. **C**: Soluble nuclear extracts (0.5 ml) prepared from U2OS cells transfected with a plasmid (1 μg) expressing FLAG-SP100B were supplemented with a mixture of the indicated [^{32}P]-oligonucleotide (10 ng) and calf thymus DNA (2 μg). Subsequent steps were conducted as described in the legend for panel B. Immunoprecipitations were performed in triplicate and the error bars represent the standard error of the mean.

hemi-methylated versions of oligonucleotide B7C to compete with unmethylated [^{32}P]-B7C for binding to SP100B. The cytosines at positions B and C on both strands (1, 2, 3, and 4—Fig. 8A) were replaced with 5 mC in methylated B7C. Hemi-methylated versions of B7C were synthesized in which the cytosines at position 1 and 2 or 3 and 4 were replaced with 5 mC (Fig. 8A). The co-immunoprecipitation results (Fig. 8B) indicated that methylated B7C did not compete efficiently with [^{32}P]-B7C for binding to SP100B. The hemi-methylated versions of B7C competed with [^{32}P]-B7C for binding to SP100B, but not as efficiently as unmethylated B7C. From these results, we conclude that methylation of the cytosines in the context of the CpG dinucleotide inhibits binding of SP100B.

We propose that the inability of SP100B to bind methylated CpGs provides one explanation for our observation in Figure 3 that calf thymus DNA does not compete with plasmid DNA for binding to SP100B. Approximately 70–80% of the CpGs in calf thymus DNA are methylated [Gruenbaum et al., 1981] and, hence, these sites are inaccessible for binding by SP100B. In contrast, the CpGs within plasmid DNA isolated from *E.coli* are unmethylated,

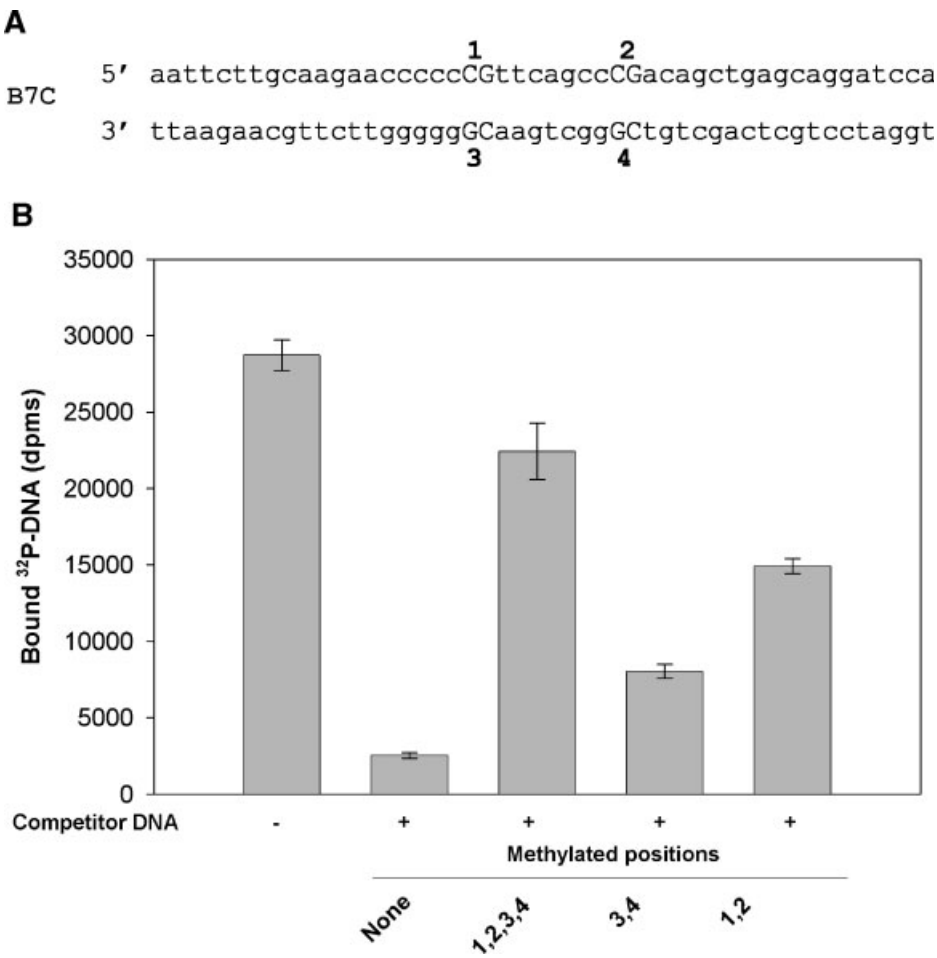

Fig. 8. Binding of SP100B to DNA is inhibited by CpG methylation. **A**: The sequence of the oligonucleotide B7C is shown. The numbers above and below the sequence indicate the positions where the cytosines were replaced with 5 mC. Double-stranded methylated oligonucleotide B7C was prepared by annealing complementary oligonucleotides synthesized with 5 mC at positions 1, 2, 3, and 4. Double-stranded hemimethylated oligonucleotides were prepared by annealing an oligonucleotide with 5 mC at positions 1 and 2 or positions 3 and 4 with the complementary, unmethylated oligonucleotide. **B**: Soluble nuclear extracts (0.5 ml) prepared from U2OS cells transfected with a plasmid (0.5 μg) expressing FLAG-SP100B were supplemented with a mixture containing unmethylated double-stranded [^{32}P]-B7C (10 ng, 0.32 pmoles) and calf thymus DNA (2 μg). Where indicated, the DNA mixture also contained 6 μg (200 pmoles) of unlabeled double-stranded B7C methylated at the indicated positions. Subsequent steps were conducted as described in the legend for Figure 7B. Immunoprecipitations were performed in triplicate and the error bars represent the standard error of the mean.

serving as potential binding sites for SP100B. However, there maybe additional factors involved, beside the methylation of CpGs that prevents binding of SP100B to calf thymus DNA.

DISCUSSION

We have previously demonstrated that SP100B represses the expression of genes associated with transfected plasmid DNA in a SAND domain-dependent manner [Wilcox et al., 2005]. The goal of the current study was to determine if full-length SP100B expressed in mammalian cells associates with DNA in a SAND domain-dependent manner and if so, to identify the SP100B target sequence within the DNA. Although it has been shown that the SAND domain is sufficient for DNA binding [Bottomley et al., 2001], it has also been proposed that other SP100B domains can recruit SP100B to DNA in the absence of the SAND domain. Specifically, Lehming et al. [1998] proposed that an SP100B variant (aa 5–528) that lacks the SAND domain is recruited to DNA through an interaction with DNA-bound HMG2 and further suggested that HP1 is recruited to this complex through an interaction with SP100B. Our results demonstrate that SP100B binds DNA in a SAND-domain-dependent manner and, thus, do not support the proposal by Lehming et al. In particular, our observation that SP100BΔ594-688 does not bind DNA when tested in the context of a nuclear extract reveals that the HMG2 and HP1 interaction domains are not sufficient to recruit SP100B to DNA. Furthermore, our data reveal that neither the HMG2 nor HP1 interaction domains in SP100B contribute to the ability of SP100B to bind DNA within the context of these assays.

The evidence presented here reveals that SP100B binds preferentially to sequences in double-stranded DNA that contain CpGs. To determine if SP100B exhibits any preference for sequences flanking CpGs, binding assays were conducted with oligonucleotides containing a single CpG at positions A, B, C, or D (Fig. 7 and data not shown). Less than a twofold difference was observed in the amount of each oligonucleotide bound by SP100B. These results indicate that SP100B binds with similar affinity to CpGs with different flanking sequences. However, the amounts bound were less than twofold above background levels and, thus, the assays were not sufficiently sensitive to quantify the preference of SP100B for sequences flanking CpGs.

There are at least three models that fit our observations that SP100B, in a SAND domain-dependent manner, bound oligonucleotides containing two CpGs separated by seven base pairs more efficiently than oligonucleotides containing a single CpG or oligonucleotides containing CpGs separated by 4, 11, or 20 base pairs. These models are: (1) the SP100:DNA complex is more stable when SP100B binds as a monomer to two properly spaced CpGs than to a single CpG, (2) the SP100B:DNA complex is more stable when SP100B binds as a homodimer to a pair of properly-spaced CpGs than to a single CpG, or (3) SP100B binds weakly to DNA as a monomer that recognizes a single CpG and interacts with co-factors that bridge monomers of SP100B to form a stable complex. Model 1 is not consistent with the proposal that the single KDWK motif within the SAND domain provides the primary surface for recognition of the DNA structure [Bottomley et al., 2001]. This model is also not consistent with our data that mutation of two amino acids within the KNWK sequence of the SAND domain abrogates DNA binding. The second model is supported by the observation that SP100A forms homodimers through a dimerization domain mapped within residues 33–149 [Sternsdorf et al., 1999] that are identical in SP100B. Furthermore, a dimerization domain in SP100B was mapped within residues 1–189 by yeast two hybrid analysis [Lehming et al., 1998]. Though, the dimerization domain is not required for DNA binding, given that the SAND domain independently binds DNA [Bottomley et al., 2001], the possibility that homodimerization contributes to increased stability of the SP100B:DNA complexes remains to be tested. The third model is potentially the most interesting. Our data do not support an essential role for HMG2 or HP1 as stabilizing co-factors. However, because our assays were conducted in the context of soluble nuclear extracts, there are numerous other proteins that could have contributed to the stability of the SP100B:DNA complexes that we observed. SP100B has been reported to localize to discrete bodies within the nucleus [Guldner et al., 1999]. We speculate that other proteins within these nuclear bodies stabilize the SP100B:DNA complexes.

The DNA binding properties of SP100B closely resemble those of most other SAND domain proteins in that these proteins bind bipartite CpG-containing sequences and the DNA binding activity requires the KDWK motif [Huggenvik et al., 1998; Michelson et al., 1999; Bottomley et al., 2001]. Homodimers or heterodimers of the glucocorticoid modulatory element binding proteins GMEB1 and GMEB2 bind to two variably-spaced half sites, favoring an ACGPy element at one site and a PuCGPy sequence at the other half site [Burnett et al., 2001]. The GMEB proteins associate with other CpG-containing sequences with lower efficiency. DEAF binds to TTCG nucleotide sequences in the drosophila deformed response element [Gross and McGinnis, 1996]. NUDR, the human ortholog of DEAF-1, forms dimers that associate with a direct repeat of TTC (G/C)GG within the heterogeneous nuclear ribonucleoprotein (hnRNP) A2/B1 promoter [Michelson et al., 1999]. AIRE is a SAND domain protein that does not share these features. AIRE binds tandem repeats of ATTGGTTA or TTATTA. In addition, the AIRE SAND domain does not contain the KNWK/KDWK motif, which may explain the difference in its target specificity [Kumar et al., 2001; Purohit et al., 2005].

Methylation of cytosine residues in the context of the CpG dinucleotide is a common DNA modification that occurs in the vertebrate genome. In this report, we have demonstrated that methylation of the cytosine residue in the context of the CpG dinucleotide inhibits binding of SP100B to DNA, suggesting that binding of SP100B to its target sequence is regulated via CpG methylation. DNA methylation regulates the ability of certain DNA binding proteins to bind to their target sequences. For instance, transcription factors such as E2F and cAMP-responsive element-binding protein (CREB) do not bind DNA when the CpG in their recognition sequence is methylated [Iguchi-Ariga and Schaffner, 1989; Campanero et al., 2000]. On the other hand, proteins such as MeCP1, MeCP2, and methyl binding domain proteins MBD1, MBD2, MBD3, and MBD4 specifically associate with methylated CpGs [Meehan et al., 1989; Lewis et al., 1992; Hendrich and Bird, 1998]. In contrast to the above mentioned CpG methylation sensitive proteins, there are proteins that are indifferent to the methylation status of CpG dinucleotides within their target sequence. These include the transcription factors SP1 and CTF [Harrington et al., 1988; Holler et al., 1988; Ben-Hattar et al., 1989] and an alternatively spliced form of MBD1 (MBDIvI) [Fujita et al., 1999]. Association of these proteins with their target sequence is not regulated by CpG methylation.

We propose that the specificity of SP100B DNA binding and repression is regulated by two mechanisms. First, SP100B specificity for unmethylated CpG dinucleotide-containing sequences contributes to the identification of DNAs that SP100B will target for repression. SP100B non-specifically represses the expression from a number of viral and cellular promoters that are associated with transfected plasmid DNA [Wilcox et al., 2005]. Preliminary data from our lab suggests that SP100B also represses the expression of nuclear DNA viruses such as HSV1 and adenovirus 5 (data not shown). Plasmid DNA isolated from *E.coli* and certain viral DNAs (HSV-1 and adenovirus 5) are hypomethylated [Lundberg et al., 2003], and thus, are potential targets for SP100B binding. We hypothesize that the observed SP100B-mediated repression of transfected plasmid DNA and viral DNA is a consequence of SP100B binding to unmethylated CpG dinucleotides present within these DNAs.

There are approximately 30,000 CpG islands in the promoters for housekeeping genes in the human genome [Antequera and Bird, 1993; Antequera, 2003]. In actively transcribed genes, these CpG islands are unmethylated [Caiafa and Zampieri, 2005], serving as potential binding sites for SP100B. However, our observation that cells that overexpress SP100B do not exhibit any defects in cell proliferation suggests that SP100B does not function as a global repressor of housekeeping genes (data not shown). We hypothesize that a second feature of SP100B that contributes to its selective repression is the localization of SP100B within discrete nuclear domains [Guldner et al., 1999], thereby limiting the access of SP100B to cellular DNAs. DNA viruses have been reported to initiate gene expression and replication from input DNA that localizes adjacent to discrete nuclear structures [Everett, 2001]. We are currently investigating the possibility that SP100B colocalizes with viral DNA during an active infection.

ACKNOWLEDGMENTS

We thank Heather Janscha, Kay Nicholson, Meiying Yang, and Scott Sheriff for their excellent assistance in all phases of this investigation. We would also like to thank James Bruce for very helpful discussions in the course of this study. This research would not have been possible without the generous contributions of plasmids from A. DeJean and Jan Ellenberg.

REFERENCES

Antequera F. 2003. Structure, function and evolution of CpG island promoters. Cell Mol Life Sci 60:1647–1658.

Antequera F, Bird A. 1993. Number of CpG islands and genes in human and mouse. Proc Natl Acad Sci USA 90:11995–11999.

Ben-Hattar J, Beard P, Jiricny J. 1989. Cytosine methylation in CTF and Sp1 recognition sites of an HSV tk promoter: Effects on transcription in vivo and on factor binding in vitro. Nucleic Acid Res 17:10179–10190.

Bloch DB, de la Monte SM, Guigaouri P, Filippov A, Bloch KD. 1996. Identification and characterization of a leukocyte-specific component of the nuclear body. J Biol Chem 271:29198–29204.

Bloch DB, Nakajima A, Gulick T, Chiche JD, Orth D, de La Monte SM, Bloch KD. 2000. Sp110 localizes to the PML-Sp100 nuclear body and may function as a nuclear hormone receptor transcriptional coactivator. Mol Cell Biol 20:6138–6146.

Bottomley MJ, Collard MW, Huggenvik JI, Liu Z, Gibson TJ, Sattler M. 2001. The SAND domain structure defines a novel DNA-binding fold in transcriptional regulation. Nat Struct Biol 8:626–633.

Bruce JW, Wilcox KW. 2002. Identification of a motif in the C terminus of herpes simplex virus regulatory protein ICP4 that contributes to activation of transcription. J Virol 76:195–207.

Burnett E, Christensen J, Tattersall P. 2001. A consensus DNA recognition motif for two KDWK transcription factors identifies flexible-length, CpG-methylation sensitive cognate binding sites in the majority of human promoters. J Mol Biol 314:1029–1039.

Caiafa P, Zampieri M. 2005. DNA methylation and chromatin structure: The puzzling CpG islands. J Cell Biochem 94:257–265.

Campanero MR, Armstrong MI, Flemington EK. 2000. CpG methylation as a mechanism for the regulation of E2F activity. Proc Natl Acad Sci USA 97:6481–6486.

Christensen J, Cotmore SF, Tattersall P. 1999. Two new members of the emerging KDWK family of combinatorial transcription modulators bind as a heterodimer to flexibly spaced PuCGPy half-sites. Mol Cell Biol 19: 7741–7750.

Eissenberg JC, Elgin SC. 2000. The HP1 protein family: Getting a grip on chromatin. Curr Opin Genet Dev 10: 204–210.

Everett RD. 2001. DNA viruses and viral proteins that interact with PML nuclear bodies. Oncogene 20:7266–7273.

Fujita N, Takebayashi S, Okumura K, Kudo S, Chiba T, Saya H, Nakao M. 1999. Methylation-mediated transcriptional silencing in euchromatin by methyl-CpG binding protein MBD1 isoforms. Mol Cell Biol 19:6415–6426.

Gibson TJ, Ramu C, Gemund C, Aasland R. 1998. The APECED polyglandular autoimmune syndrome protein, AIRE-1, contains the SAND domain and is probably a transcription factor. Trends Biochem Sci 23:242–244.

Gross CT, McGinnis W. 1996. DEAF-1, a novel protein that binds an essential region in a deformed response element. EMBO J 15:1961–1970.

Gruenbaum Y, Stein R, Cedar H, Razin A. 1981. Methylation of CpG sequences in eukaryotic DNA. FEBS Lett 124:67–71.

Guldner HH, Szostecki C, Schroder P, Matschl U, Jensen K, Luders C, Will H, Sternsdorf T. 1999. Splice variants of the nuclear dot-associated Sp100 protein contain homologies to HMG-1 and a human nuclear phosphoprotein-box motif. J Cell Sci 112(Pt 5):733–747.

Harrington MA, Jones PA, Imagawa M, Karin M. 1988. Cytosine methylation does not affect binding of transcription factor Sp1. Proc Natl Acad Sci USA 85:2066–2070.

Hendrich B, Bird A. 1998. Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol 18:6538–6547.

Holler M, Westin G, Jiricny J, Schaffner W. 1988. Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. Genes Dev 2:1127–1135.

Huggenvik JI, Michelson RJ, Collard MW, Ziemba AJ, Gurley P, Mowen KA. 1998. Characterization of a nuclear deformed epidermal autoregulatory factor-1 (DEAF-1)-related (NUDR) transcriptional regulator protein. Mol Endocrinol 12:1619–1639.

Iguchi-Ariga SM, Schaffner W. 1989. CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. Genes Dev 3:612–619.

Jimenez-Lara AM, Heine MJ, Gronemeyer H. 2000. Cloning of a mouse glucocorticoid modulatory element binding protein, a new member of the KDWK family. FEBS Lett 468:203–210.

Kaul S, Blackford JA Jr, Chen J, Ogryzko VV, Simons SS, Jr. 2000. Properties of the glucocorticoid modulatory element binding proteins GMEB-1 and -2: Potential new modifiers of glucocorticoid receptor transactivation and members of the family of KDWK proteins. Mol Endocrinol 14:1010–1027.

Kumar PG, Laloraya M, Wang CY, Ruan QG, Davoodi-Semiromi A, Kao KJ, She JX. 2001. The autoimmune regulator (AIRE) is a DNA-binding protein. J Biol Chem 276:41357–41364.

Lehming N, Le Saux A, Schuller J, Ptashne M. 1998. Chromatin components as part of a putative transcriptional repressing complex. Proc Natl Acad Sci USA 95: 7322–7326.

Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, Bird A. 1992. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell 69:905–914.

Lundberg P, Welander P, Han X, Cantin E. 2003. Herpes simplex virus type 1 DNA is immunostimulatory in vitro and in vivo. J Virol 77:11158–11169.

McKay RD. 1981. Binding of a simian virus 40 T antigen-related protein to DNA. J Mol Biol 145:471–488.

Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP. 1989. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell 58:499–507.

Michelson RJ, Collard MW, Ziemba AJ, Persinger J, Bartholomew B, Huggenvik JI. 1999. Nuclear DEAF-1-related (NUDR) protein contains a novel DNA binding domain and represses transcription of the heterogeneous nuclear ribonucleoprotein A2/B1 promoter. J Biol Chem 274:30510–30519.

Purohit S, Kumar PG, Laloraya M, She JX. 2005. Mapping DNA-binding domains of the autoimmune regulator protein. Biochem Biophys Res Commun 327:939–944.

Sternsdorf T, Jensen K, Reich B, Will H. 1999. The nuclear dot protein sp100, characterization of domains necessary for dimerization, subcellular localization, and modification by small ubiquitin-like modifiers. J Biol Chem 274: 12555–12566.

Surdo PL, Bottomley MJ, Sattler M, Scheffzek K. 2003. Crystal structure and nuclear magnetic resonance analyses of the SAND domain from glucocorticoid modulatory element binding protein-1 reveals deoxyribonucleic acid and zinc binding regions. Mol Endocrinol 17:1283–1295.

Szostecki C, Krippner H, Penner E, Bautz FA. 1987. Autoimmune sera recognize a 100 kD nuclear protein antigen (sp-100). Clin Exp Immunol 68:108–116.

Szostecki C, Guldner HH, Netter HJ, Will H. 1990. Isolation and characterization of cDNA encoding a human nuclear antigen predominantly recognized by autoantibodies from patients with primary biliary cirrhosis. J Immunol 145:4338–4347.

Szostecki C, Will H, Netter HJ, Guldner HH. 1992. Autoantibodies to the nuclear Sp100 protein in primary biliary cirrhosis and associated diseases: Epitope specificity and immunoglobulin class distribution. Scand J Immunol 36:555–564.

Thomas JO, Travers AA. 2001. HMG1 and 2, and related 'architectural' DNA-binding proteins. Trends Biochem Sci 26:167–174.

Wilcox KW, Sheriff S, Isaac A, Taylor JL. 2005. SP100B is a repressor of gene expression. J Cell Biochem 95:352–365.