MOLECULAR AND CELLULAR BIOLOGY, Jan. 1996, p. 414–421 Vol. 16, No. 1
0270-7306/96/$04.00+0

DNA Methylation Specifies Chromosomal Localization of MeCP2

XINSHENG NAN,[1] PERI TATE,[1] EN LI,[2] AND ADRIAN BIRD[1]*

Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, United Kingdom,[1] and Cardiovascular Research Center, Massachusetts General Hospital, Charlestown, Massachusetts 02129[2]

Received 5 September 1995/Returned for modification 17 October 1995/Accepted 26 October 1995

MeCP2 is a chromosomal protein that is concentrated in the centromeric heterochromatin of mouse cells. In vitro, the protein binds preferentially to DNA containing a single symmetrically methylated CpG. To find out whether the heterochromatic localization of MeCP2 depended on DNA methylation, we transiently expressed MeCP2-LacZ fusion proteins in cultured cells. Intact protein was targeted to heterochromatin in wild-type cells but was inefficiently localized in mutant cells with low levels of genomic DNA methylation. Deletions within MeCP2 showed that localization to heterochromatin required the 85-amino-acid methyl-CpG binding domain but not the remainder of the protein. Thus MeCP2 is a methyl-CpG-binding protein in vivo and is likely to be a major mediator of downstream consequences of DNA methylation.

Widespread methylation of genomic DNA is a characteristic of the vertebrates. The target of methylation is the dinucleotide CpG, which acquires a methyl group at the 5 position on the cytosine ring. Functions of methylation have long been the subject of experiment and speculation, but recent experiments provide evidence for an essential role in development. Li et al. (16) disrupted the gene that encodes the DNA methyltransferase (MTase) enzyme in mice and found that homozygous mutant embryos had greatly reduced levels of methylation, were developmentally retarded, and died at midgestation. Undifferentiated embryonic stem (ES) cells, on the other hand, behaved normally in culture, despite very low levels of methylation, indicating that methylation may be less important at this totipotent stage. Why methylation is essential for development is not yet completely clear. One possibility is that methylation-mediated repression of genes is somehow built into the developmental program (discussed in reference 1).

To understand the biology of DNA methylation, it is necessary to identify all components of the system. Several proteins that bind specifically to methylated DNA are known (reviewed in reference 31). Two of these, MeCP1 and MeCP2, bind to symmetrical methyl-CpG pairs in any sequence context and may therefore be of general significance (14, 17, 18). Studies of MeCP1 have implicated it in methylation-mediated gene inactivation (2, 3). The second protein, MeCP2, consists of a single chain of 492 amino acids in the rat, including an 85-amino-acid domain near the N terminus that encodes all of the specificity required for binding to methylated DNA (22). The mouse protein (484 amino acids) is very similar to its rat homolog, showing an overall identity of 95% at the amino acid level and absolute conservation of the methylated DNA binding domain (29). This methyl-CpG binding domain (MBD) can bind as a monomer to a single symmetrically methylated CpG pair. Immunofluorescence analysis of mouse chromosomes shows that MeCP2 is preferentially localized in pericentromeric heterochromatin (14), which is also the region of highest 5-methylcytosine concentration (19). The predominant DNA sequence in mouse heterochromatin is the major satellite that accounts for about 7% of genomic DNA and between 25 and 50% of genomic 5-methylcytosine. Purified MeCP2 can bind to naturally or artificially methylated satellite DNA in vitro but not to satellite that has been demethylated by cloning (14). These results suggest that MeCP2 may be attracted to heterochromatin by the high concentration of methyl-CpG therein. In this study, we have tested this hypothesis by transfection of cloned constructs expressing MeCP2 into cultured cells. We find that localization depends on the presence of methyl-CpG and upon the MBD.

* Corresponding author. Mailing address: Institute of Cell and Molecular Biology, University of Edinburgh, Kings Buildings, Edinburgh EH9 3JR, United Kingdom. Phone: 0131-650-5670. Fax: 0131-650-5379.

MATERIALS AND METHODS

Construction of expression vectors. The expression vector pββgeopA (a gift from W. Skarnes) consists of the human β-actin promoter driving transcription of a fusion between *lacZ* and the neomycin resistance gene. This vector was modified by insertion of 24-bp blunt-ended double-stranded oligonucleotides, AGCTAGCAAGCTTGGTCACCGTCG, into the *Nru*I site of pββgeopA to introduce two unique sites for *Nhe*I and *Bst*EII between the *Nco*I and *Nru*I sites to generate the expression vector pββgeoNheBst. The cDNA of MeCP2 was inserted into either the *Nco*I site or the *Nru*I site of pββgeoNheBst for nested deletion from the 3′ or 5′ end, respectively. Exonuclease III-nuclease S1 deletions were performed with a deletion kit (Pharmacia). Deletion sites from the 5′ end were determined by sequencing with a 24-mer primer, ACAGAGCCTCGC CTTTGCCGAT, which is located in the human β-actin promoter region. Deletion sites from the 3′ end were determined by sequencing with appropriate primers from the sense strand of MeCP2 cDNA.

The MeCP2-βgeo fusion tested in ES cells (pββgeoMeCP2) consisted of amino acids 1 to 437 of the rat protein, introduced as a 1.3-kb *Nco*I fragment into the pββgeopA polylinker, as an in-frame carboxy-terminal fusion with *lacZ*. An expression vector consisting of the entire rat MeCP2 open reading frame under control of the β-actin promoter was constructed from pββgeoMeCP2 by deletion of a 4.5-kb *Xba*I-*Nru*I fragment containing βgeo followed by insertion of a 500-bp *Sma*I-*Cla*I fragment containing the remainder of the MeCP2 open reading frame upstream of the polyadenylation site.

Western-Southwestern (DNA-protein) assays. Nuclei were isolated from control L cells or cells stably transfected with MeCP2 fusion constructs. Nuclear proteins were separated on a 7% denaturing polyacrylamide gel and electroblotted onto a nitrocellulose filter. Southwestern blotting with poly(GAM) and poly(GAC) probes was as described previously (14, 22). For Western blots (immunoblots), filters were blocked with 2% nonfat dried milk in TBST buffer (50 mM Tris [pH 8], 150 mM NaCl, 0.1% Triton X-100) and then incubated with a 1:5,000 dilution of anti-β-galactosidase (β-Gal) antibody (Promega) in 5% fetal calf serum–TBST. The secondary anti-mouse immunoglobulin G (IgG) antibody, conjugated with peroxidase (Sigma), was used in a 1:5,000 dilution in 5% fetal calf serum–TBST. After extensive washing, filters were processed with the enhanced chemiluminescence system (Amersham).

Cell culture and transfection. L929 cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum at 37°C and 5% CO_2. Transient transfections were performed by the DEAE-dextran method (27) with slight modification. L cells at confluence were split 1:10 and seeded on coverslips (22 by 22 mm). After 24 h, a DNA-DEAE-dextran mixture was added in Dulbecco's modified Eagle's medium with 1% fetal calf serum.

ES cells were maintained in GMEM supplemented with 1× modified Eagle's

medium nonessential amino acids–1 mM sodium pyruvate–2 mM glutamine–50 μM β-mercaptoethanol–15% fetal calf serum (Globepharm)–100 U of soluble differentiation-inhibitory activity/leukemia-inhibitory factor per ml (28) in a 5% CO_2 atmosphere. DNA constructs were introduced into ES cells by electroporation. A total of 10^7 cells were mixed with 30 μg of plasmid DNA and subjected to a 250-V pulse (500 μF) with a Bio-Rad Gene Pulser apparatus. Following electroporation, aliquots of 10^6 cells were plated on gelatinized coverslips and maintained in ES cell medium for 48 h prior to antibody staining. The embryonal carcinoma cell line 28/10 was derived from differentiated embryonic tumor cells of mice (21) and consists of a mixture of fibroblastoid and epithelial morphologies.

To establish stable L-cell lines expressing fusion proteins, cells were electroporated in the presence of linearized DNA (10 μg) at 1,000 V, 250 μF. Cells were plated at a density of 10^6/10-cm plate. Selection with G418 (Gibco) at 200 μg/ml was begun 24 h after plating. G418-resistant colonies were picked, trypsinized, and split into two wells of a 24-well plate about 10 days after selection. Expression of fusion proteins was tested by X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) staining in one of the two wells. X-Gal-positive cells were grown in Dulbecco's modified Eagle's medium–10% fetal calf serum and 200 μg of G418 per ml.

Immunofluorescence staining. For immunofluorescence staining of transiently transfected L cells on coverslips, all solutions were prepared in phosphate-buffered saline with 1.5 mM $MgCl_2$ and 1 mM $CaCl_2$ unless specified otherwise. Three days following transfection, cells were fixed in 3.7% formaldehyde for 10 min and permeabilized with 0.2% Triton X-100 for 12 min. Coverslips were incubated with blocking buffer (5% goat serum) for 10 min and then with monoclonal anti-β-galactosidase antibody (Promega) in blocking buffer for 45 min. Fluorescein isothiocyanate (FITC)-conjugated anti-mouse IgG antibody (Sigma) was used to detect location of β-galactosidase fusion proteins. Finally, cells on coverslips were counterstained with Hoechst 33258. Metaphase chromosome spreads are prepared as reported previously (14) with minor modifications. Briefly, 24 h after passaging confluent cells at a split ratio of 1:5, Colcemid (Sigma) was added to the culture to a final concentration of 0.1 μg/ml. After 1 h, the cells were trypsinized, washed with supernatant medium, and recovered by centrifugation at 800 × *g* for 3 min. Cell pellets were resuspended in 0.075 M KCl hypotonic solution to a final density of approximately 2×10^5 cells per ml and swollen at 37°C for 10 min. Cell suspensions were dropped on glass slides. Slides were immersed in KCM buffer (120 mM KCl, 20 mM NaCl, 10 mM Tris HCl [pH 8.0], 0.5 mM EDTA, and 0.1% Triton X-100) for at least 10 min to permeabilize cells and blocked in KCM plus 5% goat serum for 5 min. Cell preparations were incubated in either anti-MeCP2 antiserum diluted 1:200 or anti-β-galactosidase antibody diluted 1:200 in KCM plus 5% goat serum for 1 h at room temperature. The anti-MeCP2, Ab674, was raised against a bacterially expressed glutathione *S*-transferase fusion protein containing amino acids 1 to 392 of the rat protein and was a gift from R. Meehan. After three washes, slides were incubated with secondary antibody of FITC-conjugated anti-rabbit (for anti-MeCP2 antiserum) or anti-mouse (for anti-β-Gal antibody) IgG (Sigma) at a 1:200 dilution for 45 min.

Immunofluorescence staining of ES cells was performed on cells fixed in 3% paraformaldehyde for 20 min. Preparations were treated with 50 mM ammonium chloride prior to permeabilization with Triton X-100. Cells were blocked with 5% sheep serum (Sigma) for 20 min prior to incubation with antibodies. MeCP2 proteins were detected with a 1:400 dilution of polyclonal rabbit antisera either against β-galactosidase (gift of Jack Price) or against amino acids 1 to 392 of MeCP2 (see above). A 1:400 dilution of a goat anti-rabbit IgG-FITC conjugate (Jackson Immunologicals) was used as the secondary antibody in both cases. Following counterstaining with Hoechst 33258, coverslips were mounted on microscope slides in Moviol mountant. Photomicrographs were taken on an Olympus Vanox microscope with 800/1600 ASA film. For anti-m^5C antibody staining, acetic acid-methanol fixed metaphase chromosome spreads were prepared from ES cells (26), treated overnight with UV light, and stained with a monoclonal antibody against m^5C as described previously (20). In all experiments, FITC exposures were made before Hoechst exposures to eliminate bleed-through of Hoechst signal into the FITC channel.

RESULTS

An MeCP2 fusion protein is correctly localized. In order to distinguish between endogenous MeCP2 and exogenously introduced forms, expression constructs were generated by fusing sequences from the MeCP2 open reading frame with the bacterial *lacZ* gene. The *lacZ* gene product, β-galactosidase (β-Gal), is known to tolerate large fusions while retaining its enzymatic activity (4). Furthermore, fusions between portions of many mammalian proteins and β-Gal often become localized to sites normally occupied by the unfused endogenous protein (10, 13). We fused the entire open reading frame of the rat MeCP2 gene in frame with the 5′ end of a preexisting fusion (known as βgeo) between the *lacZ* gene and the neomycin resistance (*neo*) gene (see Fig. 5D; see reference 9; the version used here had been modified to correct a mutation in *neo* [27a]). Transcription of βgeo was driven by a 450-bp fragment derived from the promoter of the human β-actin gene (8) and was terminated by a polyadenylation signal from simian virus 40 (5).

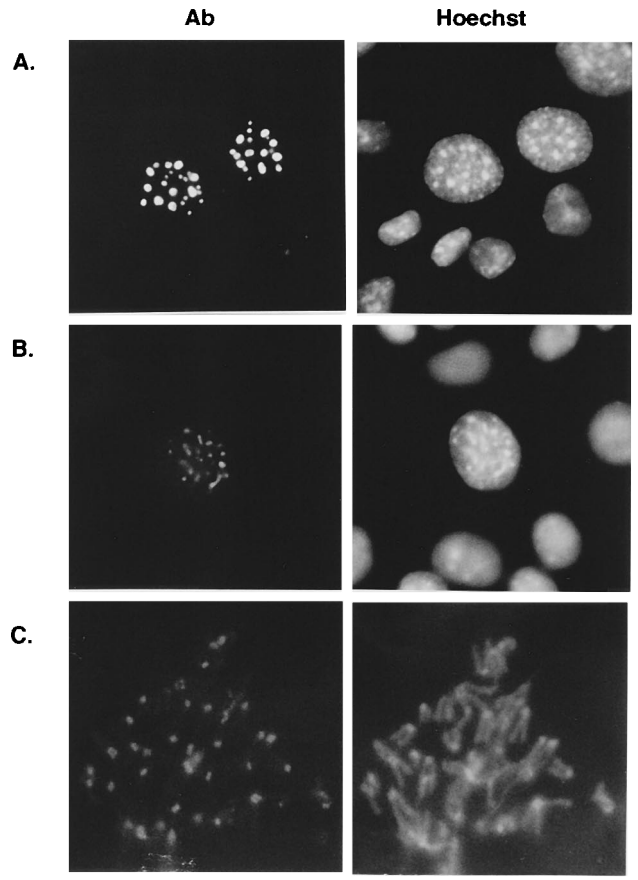

FIG. 1. Correct localization of MeCP2-βgeo fusion protein in transiently transfected cells. Mouse EB28/10 cells or L929 cells were transfected with constructs encoding an MeCP2-βgeo fusion, and localization of the fusion protein was compared with that of endogenous MeCP2 by indirect immunofluorescence. (A) Anti-β-Gal antibody staining of EB28/10 cells expressing the MeCP2-βgeo fusion protein; (B) anti-MeCP2 antibody staining of untransfected EB28/10 cells (only a minority of cells of this line express the protein at stainable levels); (C) anti-β-Gal antibody staining of metaphase chromosomes from L cells expressing the MeCP2-βgeo fusion protein. In each case, left-hand panels show immunofluorescence, and right-hand panels show counterstaining of the same cells with Hoechst 33258 to visualize heterochromatic foci.

Initially, we asked whether the MeCP2-βgeo fusion protein was targeted to heterochromatin-like endogenous MeCP2 when transiently expressed in mouse cells. The expression construct was transfected into the EB28/10 mouse cell line that was established from differentiated embryonal carcinoma cells (21). The location of the fusion protein was determined by indirect immunofluorescence staining with an anti-β-Gal antibody 48 h later. Figure 1A shows that the MeCP2-βgeo fusion protein is distributed in distinct nuclear foci that match regions of intense Hoechst 33258 staining in the same cell. The staining pattern is indistinguishable from that of untransfected cells with an anti-MeCP2 antibody (Fig. 1B). Expression of exogenous MeCP2 was further characterized in stable cell lines. Analysis of metaphase spreads in mouse L929 fibroblasts that were stably transfected with the MeCP2-βgeo construct

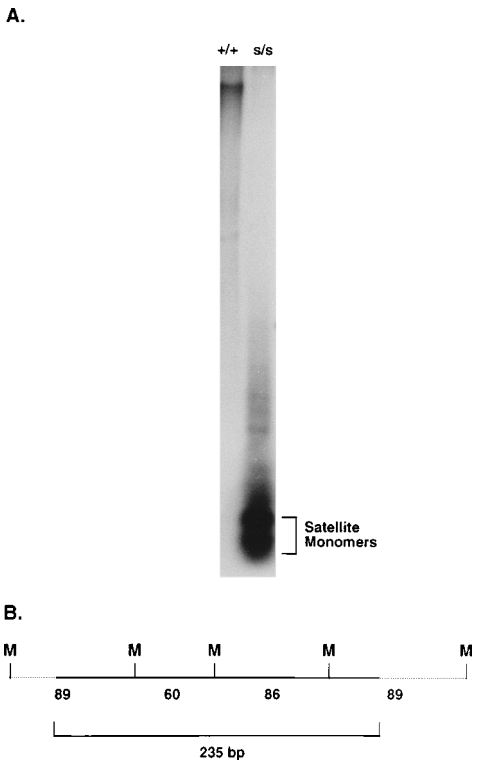

FIG. 2. Low levels of methylation at satellite DNA sequences in mouse ES cells that are homozygous for disruption of the MTase gene (15). Control cells (+/+) are from the same line that was used to generate the MTase^s mice. Genomic DNA from each cell line was digested with the methylation-sensitive enzyme *Mae*II prior to Southern blotting and probing with the fragment of the mouse major satellite. (A) Extensive cleavage of satellite from the mutant cells (s/s) indicating extreme hypomethylation at ACGT sites. The resistance of satellite from the control cells (+/+) indicates high levels of methylation at these sites. (B) Map of a satellite repeat unit showing *Mae*II sites (M) and fragment sizes within the consensus sequence.

showed that the fusion protein was associated with metaphase chromosomes, being concentrated in regions of pericentromeric heterochromatin (Fig. 1C). This distribution mirrors that of endogenous MeCP2 in untransfected L cells (see reference 14). We conclude that, in spite of the large size of the βgeo moiety, localization of the MeCP2 fusion protein in transiently and stably transfected cells is indistinguishable from that of native MeCP2.

Localization of MeCP2 depends upon the presence of CpG methylation. Does the preferential localization of MeCP2 within heterochromatin depend on the presence of methyl-CpG? To answer this question, we transfected expression constructs for the MeCP2-βgeo fusion into an ES cell line homozygous for a targeted mutation in the MTase gene, MTase^s (15). The MTase^s cell line B was isolated from a homozygous mutant embryo and contains approximately 5% of the wild-type levels of m^{5}C (14a). As a control, the wild-type ES cell line J1, which was used to create the original MTase^s mutation, was also transfected. The degree of DNA methylation in both the MTase mutant cells and the wild-type cell line was checked by digesting genomic DNA with the methyl-sensitive restriction enzyme *Mae*II (Fig. 2) and probing the resulting Southern blots with a fragment of the major mouse satellite. There are three sites for *Mae*II (recognition sequence ACGT) within the satellite consensus sequence (bottom of Fig. 2) which can only be cleaved when the site is unmethylated at cytosine. Most of the hybridization signal from mutant ES cells was in fragments of the size expected for fully digested satellite monomers. In contrast, satellite sequences in the parental cell line were highly resistant to *Mae*II digestion, indicating the high levels of methylation that are typical of most mouse somatic cell types. This result was confirmed at the cytological level by staining metaphase chromosomes with a monoclonal antibody raised against m^{5}C (24) (Fig. 3). The antibody preferentially stained the pericentromeric heterochromatin of control chromosomes and also stained heterochromatic foci within control interphase nuclei (Fig. 3A) (19). MTase mutant cells, on the other hand, exhibited general background staining with the antibody (Fig. 3B), although heterochromatic staining was apparent on a small number of mutant metaphase spreads. This suggests that a small minority of mutant cells express significant levels of MTase.

MeCP2 was transiently expressed in the J1 (wild type) and B (homozygous MTase^s) cell lines both as a βgeo fusion and in the unfused state. The latter experiment was possible because ES cells do not produce immunologically detectable amounts of endogenous MeCP2 protein. A construct expressing a fusion between the first 437 amino acids of the rat MeCP2 protein and βgeo (pββgeoMeCP2) and a construct expressing the entire (unfused) rat MeCP2 open reading frame (pβMeCP2) were each introduced by electroporation. Immunofluorescence staining of fixed interphase cells with either anti-β-Gal or anti-

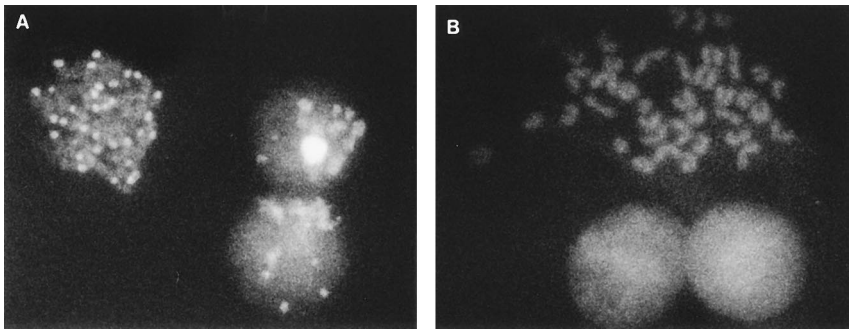

FIG. 3. Loss of methylation within the centromeric heterochromatin of homozygous MTase mutant ES cells. Fixed metaphase chromosomes from both MTase^s mutant cells (line B) and wild-type cells (J1) were stained with a monoclonal antibody against m^{5}C (20). (A) Intense staining of centromeric heterochromatin apparent on wild-type ES cell chromosomes characteristic of mouse cells. Heterochromatic staining is also observed as foci of immunofluorescence in interphase nuclei. (B) Background levels of staining on chromosomes and a lack of punctate staining in interphase nuclei of mutant cells, indicative of a reduced m^{5}C content within constitutive heterochromatin.

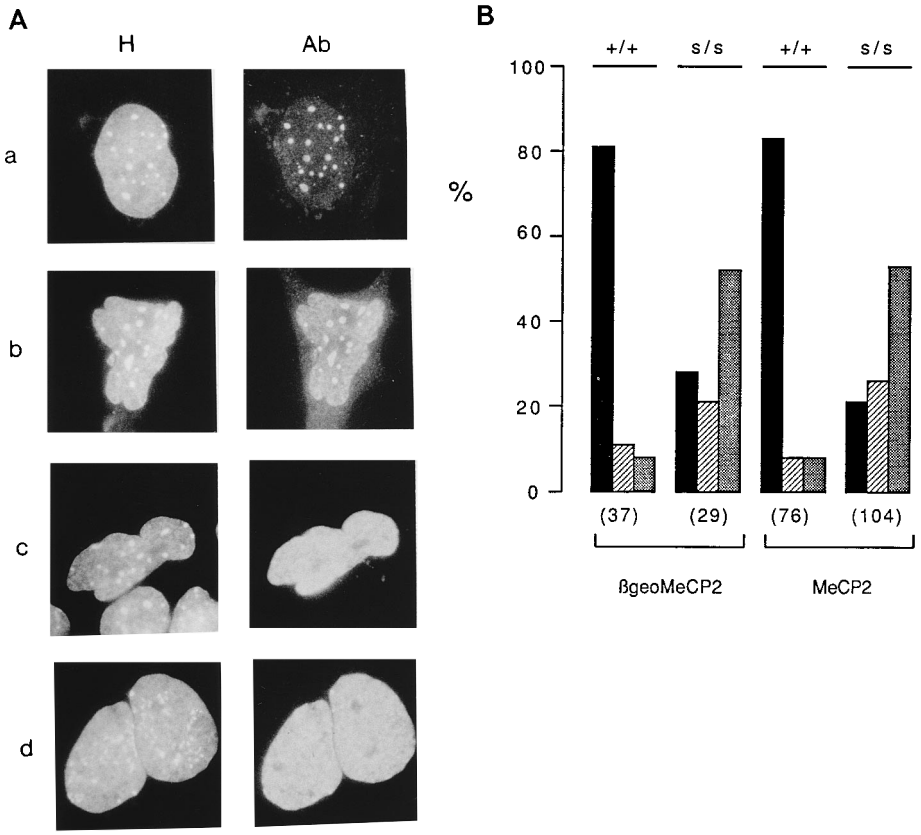

FIG. 4. Reduced levels of DNA methylation prevent efficient localization of MeCP2 to heterochromatic sites. Mouse ES cells (line B) that are homozygous for disruption of the MTase gene (15) or control J1 cells from which the original mutation was derived were transiently transfected with a construct that expressed the βgeo-MeCP2 fusion (pββgeoMeCP2) or nonfusion MeCP2 (pβMeCP2). (A) The categories of localization observed with anti-β-Gal or anti-MeCP2 antibodies. Panels on the left (H) show Hoechst staining and on the right (Ab) show staining of the same nucleus with anti-β-Gal antibody. (a) A J1 (wild type) nucleus showing signal concentrated in heterochromatic foci (localized); (b) a J1 nucleus showing some signal in heterochromatin and some dispersed throughout the nucleus (semilocalized); (c and d) MTase-mutant nuclei showing all signal dispersed throughout nucleus (unlocalized). (B) Histogram showing the percentage of transfected nuclei exhibiting each category of localization (localized [solid bars], semilocalized [hatched bars], and unlocalized [checkered bars]). The number of nuclei that were analyzed is shown in parentheses. Most mutant ES cells (s/s) were unable to localize MeCP2, whereas most control cells (+/+) localized the protein efficiently.

MeCP2 antisera gave similar results (Fig. 4A). The results are quantified in the histogram in Fig. 4B. Over 80% of wild-type J1 cells displayed the punctate nuclear staining of MeCP2 characteristic of other mouse cell lines (Fig. 4). MTase mutant B cells, on the other hand, showed a more diffuse distribution of MeCP2 within the nucleus, with only about 20 to 30% of the transfected mutant cells exhibiting correctly localized nuclear staining (Fig. 4B). The inability of mutant cells to localize MeCP2 efficiently indicates that DNA methylation is a major determinant of the subcellular localization of MeCP2.

MBD is necessary and may be sufficient for localization to heterochromatin. Previous work established that the specificity for binding to methylated DNA resides in an 85-amino-acid methyl-CpG binding domain, MBD (22). To test whether MBD is necessary for the targeting of MeCP2 to heterochromatin, a deletion of 27 amino acids within the MBD region of the MeCP2 gene was constructed (see Fig. 7) and fused with the βgeo reporter as before. That the deletion abolished binding to methylated DNA was shown by Southwestern analysis of stably transfected mouse L cells (Fig. 5B and C). An antibody against β-Gal detected a large fusion protein in extracts from cells containing the deletion construct Δ99-126 (Fig. 5A). This protein was indistinguishable in size from the undeleted fusion protein, as a difference of 27 amino acids in proteins of this size would not be detectable by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis. The proteins differed, however, in their binding to methylated DNA. Unlike the intact MeCP2 fusion protein, the Δ99-126 protein failed to bind poly(GAM), as expected for a deletion that abolishes the specificity of MBD for methylated DNA (Fig. 5B). The localization of Δ99-126 was assayed by immunofluorescence staining of transiently and stably transfected cells. Unlike intact MeCP2, this protein was dispersed in nuclei and in about 80% of nuclei did not localize to heterochromatin (Fig. 6). Approximately 20% of nuclei showed weak heterochromatic staining with the anti-β-Gal antibody, superimposed on a uniform distribution (see arrowed nucleus, Fig. 6). Since the mutated protein shows no specificity for methylated DNA in vitro, it is likely that MeCP2 interacts weakly with components of heterochromatin other than methyl-CpG. The distribution of Δ99-126 protein was also assayed on metaphase chromosomes in stably transfected lines. Instead of the predominantly heterochromatic pattern of wild-type MeCP2, the deleted protein appeared to be associated with chromosomes in granular aggregates (data not shown). The primary determinant of MeCP2 localization, therefore, appears to be the interaction with methyl-CpG.

To further define the domains required for the subnuclear localization of MeCP2, nested deletions of the protein from either the amino or the carboxy termini were fused with the β*geo* gene. Expression of these fusion proteins was confirmed

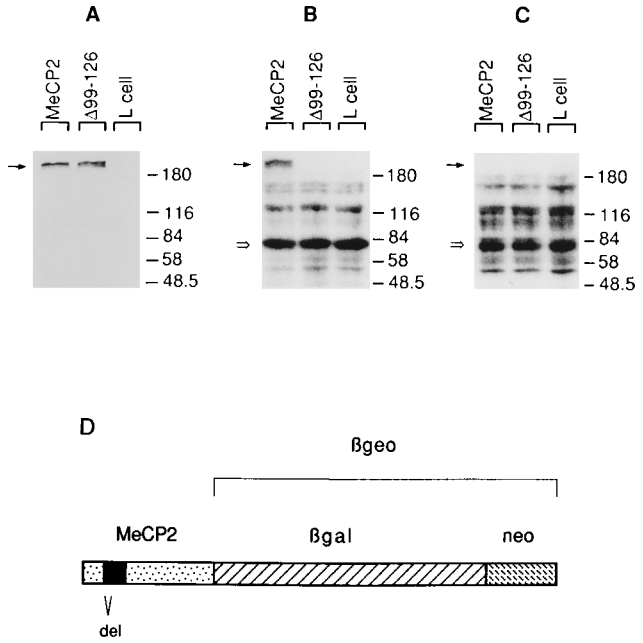

FIG. 5. Expression of MeCP2-βgeo fusion proteins in stable cell lines assayed by Western and Southwestern blots. Nuclear proteins were separated on SDS–7% acrylamide gels and transferred to nitrocellulose membranes. Lanes contain extracts from cells transfected with the intact MeCP2 fusion construct (MeCP2), the Δ99-126 deletion construct (Δ99-126), or untransfected L cells (L cells). Protein size markers (in kilodaltons) are indicated on the right of each panel. Solid arrows indicate the position of the MeCP2-βgeo fusion protein. Open arrows show the position of endogenous MeCP2. This protein has a molecular mass of 53 kDa but migrates aberrantly at about 80 kDa. Several bands not attributable to MeCP2 are seen as the experiment was carried out in the absence of nonradioactive competitor DNA to maximize signal from the fusion protein; under these conditions, some proteins bind the DNA probe nonspecifically. (A) Western blot using monoclonal anti-β-Gal antibody. (B) Southwestern blot using methylated DNA probe poly(GAM). (C) Southwestern blot using unmethylated DNA poly(GAC). The widespread binding of nonmethylated DNA in panels B and C reflects absence of competitor DNA in the incubation mix; these conditions were found to enhance detection of the MeCP2-*lacZ* fusion product. (D) Diagrammatic representation of the MeCP2-βgeo fusion protein, showing MeCP2 containing the methyl-CpG binding domain (black box) fused to βgeo, comprising the β-galactosidase protein (β-Gal) and the neomycin resistance protein (neo). The deletion in MBD (Δ99-126) is marked "del."

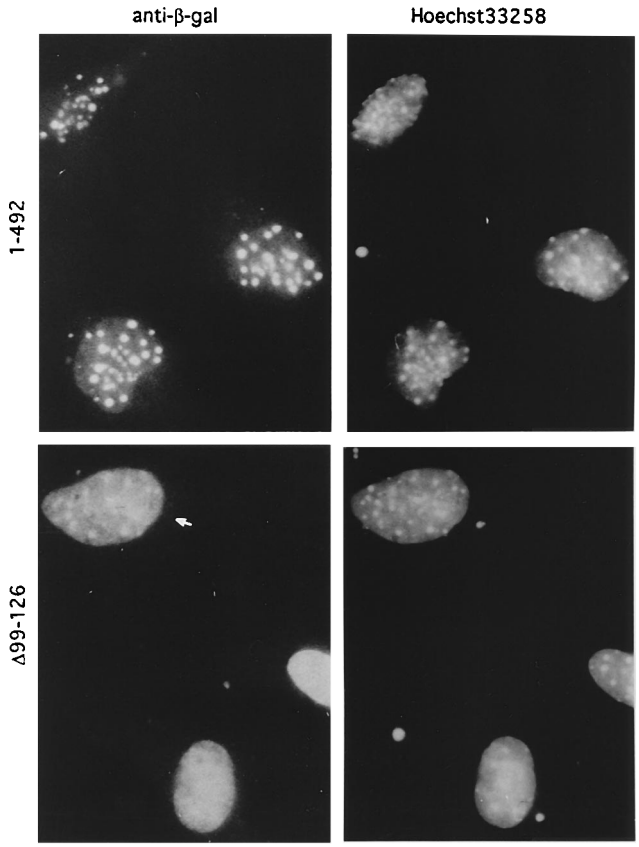

FIG. 6. An internal deletion in the methyl-CpG binding domain of MeCP2 prevents localization to heterochromatin. Transient transfections with either the full-length MeCP2 construct (1-492) or the deletion construct (Δ99-126) were stained with anti-β-Gal antibody (left panels) or Hoechst 33258 (right panels). The arrow shows a nucleus that exhibits weak heterochromatic staining superimposed on dispersed staining. Foci of this kind were detectable in about 20% of nuclei, the remainder being homogeneously stained (nonarrowed nuclei).

by Western blots with anti-β-Gal antibody as above. All constructs produced fusion proteins of the predicted sizes (data not shown). Considering first the C-terminal deletions, loss of amino acid sequences beyond position 308 did not affect localization. Deletion of sequences beyond amino acid 261, however, gave a major change in distribution, as, instead of being nuclear, most of the fusion protein remained in the cytoplasm. Nevertheless, some protein entered the nucleus, perhaps by leakage, and was localized to heterochromatin (see Fig. 8B). The region between amino acids 255 and 287 contains a consensus nuclear localization signal whose loss probably prevents efficient entry into the nucleus (see below). Since truncated MeCP2 molecules that did get into the nucleus were able to target to heterochromatin, we conclude that loss of the nuclear localization signal does not compromise this aspect of the protein's specificity. This pattern of part-cytoplasmic, part-heterochromatic localization was maintained in successive deletions from the C terminus until amino acid 178 (see Fig. 7). Further deletion to amino acid 108 caused preferential cytoplasmic location and failure to stain heterochromatin (see Fig. 8C). The C-terminal deletions show that the entire region from positions 176 to 492 (nearly two-thirds of the protein) is dispensable with respect to heterochromatic localization of MeCP2. Loss of amino acids at the N terminus is also compatible with correct subnuclear localization of truncated MeCP2. Heterochromatic staining was preserved in deletions up to position 99 (Fig. 7 and 8D), but beyond amino acid 117, all deletions caused fusion proteins to be dispersed homogeneously in nuclei or cytoplasm (Fig. 7 and 8E and F). In all of the transfections with truncated MeCP2, the vast majority of nuclei showed the same distribution, typified by the examples in Fig. 8.

NLS of MeCP2. As a nuclear protein, MeCP2 is expected to contain a nuclear localization signal (NLS). Typically, the NLS comprises a bipartite motif of two basic amino acids, a spacer region of any 10 amino acids, and a basic cluster in which three of the next five amino acids must be basic (25). Three regions in MeCP2 match the NLS motif. These candidate NLSs are located at positions 174 to 190 (KKPKSPKAPGTGRGRGR), 255 to 271 (RKAEADPQAIPKKRGRK), and 270 to 286 (RK PGSVVAAAAAEAKKK). Deletions showed that amino acids 174 to 190 do not function as an NLS, since loss of the other two sequences in deletion 1-261 (Fig. 7 and 8B) gave most protein in the cytoplasm with a small amount in the nucleus. Even when amino acids 174 to 190 were deleted (as in construct 1-178), a small fraction of protein entered the nucleus, suggesting that this is due to leakage rather than the influence

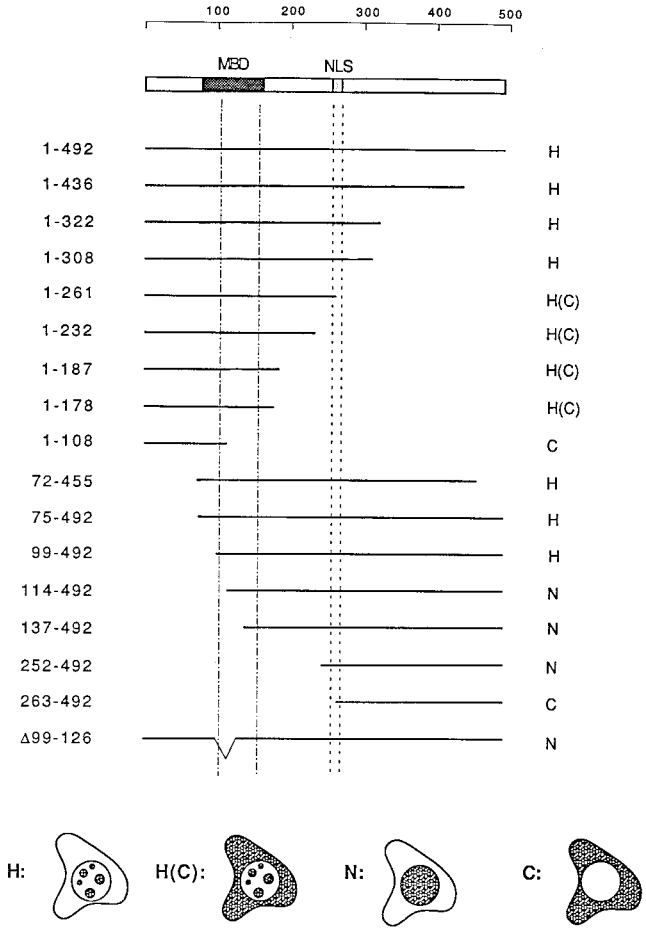

FIG. 7. Summary of the effects of nested deletions of MeCP2 on intracellular localization in L cells. MeCP2 is diagrammatically represented at the top, showing the methylated DNA binding domain (MBD) and nuclear localization sequence at positions 255 to 271 (NLS; see text). The scale at the top of the diagram corresponds to amino acid positions numbered from the N terminus. The heavy broken lines extending vertically from the MBD show the 5′ and 3′ boundaries of the region responsible for methyl-CpG binding as determined by deletion analysis of bacterially expressed MeCP2 (22). The boundaries differ from those of MBD itself as MBD was defined as the smallest free polypeptide capable of showing a binding specificity for methylated DNA. Each deletion mutant is represented by a horizontal line and is named according to the amino acids that remain (except for the internal deletion Δ99-126). All deletions were fused with βgeo as shown in Fig. 5D. Localizations of fusion proteins were grouped into four categories indicated on the right and diagrammed below the figure. H denotes a heterochromatic localization; H(C) denotes a predominantly cytoplasmic localization with some nuclear protein localized in heterochromatin; N denotes homogeneous distribution in the nucleus; C denotes exclusively cytoplasmic localization.

of this sequence. Similarly, the presence of amino acids 270 to 286 in construct 263-492 was not sufficient for nuclear localization (Fig. 7 and 8F). However, the presence of amino acids 255 to 271 in fragment 252-492 was sufficient for transportation into the nucleus (Fig. 8E), suggesting that this is the NLS for MeCP2.

DISCUSSION

The number of molecules of MeCP2 per nucleus in mouse brain and kidney has been estimated by quantitative Western blotting as 6×10^6 and 1.5×10^6, respectively (21a). This would correspond to one molecule every 5 to 20 nucleosomes, and it indicates that MeCP2 is an abundant structural component of chromosomes. Here, we have used the fact that mouse MeCP2 is concentrated in pericentromeric heterochromatin to find out which domains of the protein and which features of the DNA determine its chromosome binding. Since heterochromatin in the mouse is visible in interphase nuclei as discrete foci of dense DNA staining, we were able to assay the effects of modifying MeCP2 or DNA methylation on localization in transfected cells by a simple immunofluorescence assay. The results show that DNA methylation is the primary determinant of localization via its interaction with a methyl-CpG binding domain of the protein.

Although most MTase-negative nuclei failed to localize MeCP2, correct localization was seen in a minor proportion (21 to 28%) of nuclei. The MTase mutants are found to retain some residual methylation despite the fact that they are homozygous for a disruption of the MTase gene (14a). It is possible that cells showing localized MeCP2 may be expressing significantly more residual MTase than their neighbors. If this were the explanation, we might also expect to see punctate staining of some mutant nuclei with the anti-m^5C antibody. In fact, heterochromatic staining of nuclei and chromosomes was seen in a small percentage of nuclei, but the frequency was lower than that of nuclei localizing MeCP2. Another possible explanation for localization of MeCP2 in mutant cells is that MeCP2 might interact with other macromolecules that are also associated with pericentromeric heterochromatin. This would allow a proportion of MeCP2 molecules to find their normal location within the nucleus in the absence of methylation through protein-protein interactions alone. Indeed, we observed weak heterochromatic localization of a mutant version of MeCP2 that carried a deletion in the methyl-CpG binding domain in about 20% of nuclei. In vitro experiments indicated that the mutant protein lacks all specificity for binding to methylated DNA. Thus, it seems likely that other interactions can attract MeCP2 to heterochromatin, albeit weakly. In summary, we suggest that localization of MeCP2 in $MTase^s$ cells can be accounted for by a combination of the presence of m^5C in some mutant cells and the binding of MeCP2 to other components of heterochromatin. The major determinant of MeCP2 localization, however, is methyl-CpG.

Studies of isolated MBD have shown that it binds a single methyl-CpG pair (that is, two methyl-CpGs base paired to one another) as a monomer, covering 12 to 14 bp of DNA surrounding the methylated site (22). Our results suggest that this interaction is the primary cause of chromosome binding in vivo. Loss of heterochromatic localization in the N- and C-terminal deletions coincided precisely with previously defined boundaries of the MBD (22). In other words, all truncated proteins that could bind specifically to methylated CpG in an in vitro assay were also able to target to heterochromatin in vivo. Proteins that had lost the specificity for methylated DNA could not target to heterochromatin. These results, together with the internal deletion, suggest that the methyl-CpG binding domain is necessary and may be sufficient for specific localization.

MBD accounts for a relatively small proportion of the MeCP2 protein (85 amino acids of nearly 500), yet is sufficient for heterochromatic localization, as the regions C-terminal and N-terminal to MBD can be deleted without effect. The dispensable domains of the protein are basic and contain SPKK and GRPK motifs that have been implicated in binding to the minor groove of AT-rich duplex DNA (reviewed in reference 6). The data do not imply that these parts of the protein are functionally neutral, as only localization was assayed here. Indeed, there is evidence for substantial conservation of the amino acid sequence of MeCP2 among humans, mice, and rats, both inside and outside MBD (29a). Conservation suggests

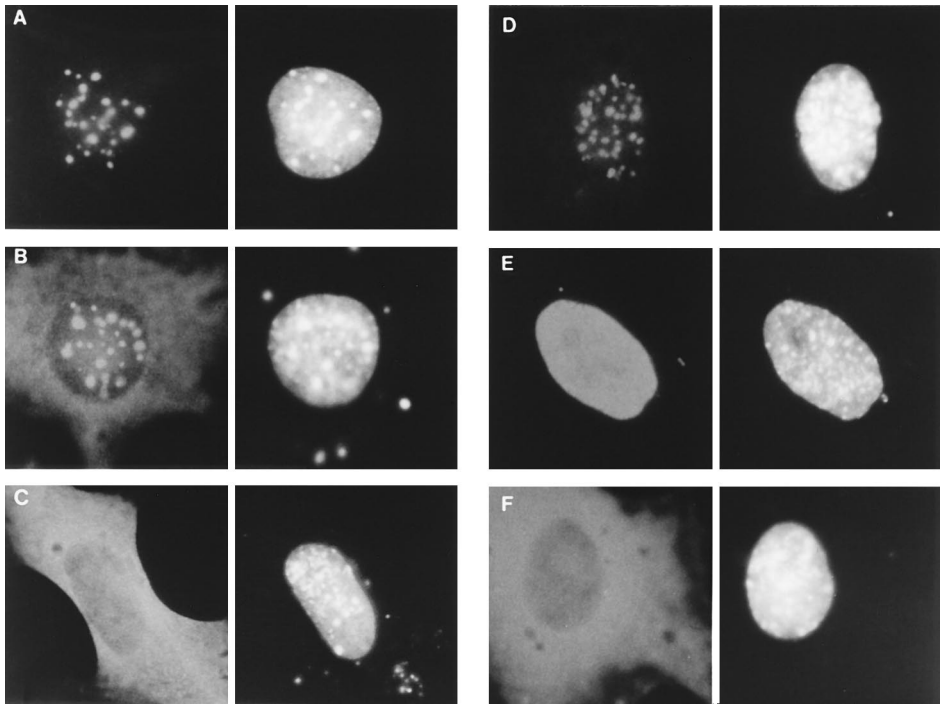

FIG. 8. Typical examples of localization of MeCP2-βgeo fusion proteins that carry deletions of parts of MeCP2. L cells were transfected with fusion constructs carrying deletion mutations of the MeCP2 gene, and transiently expressed fusion proteins were localized by immunofluorescent staining with anti-β-Gal antibody (A to F). The left panel of each pair shows antibody staining, and the right panel shows counterstaining with Hoechst 33258 of the same field. Deletions are named according to Fig. 4. (A) 1 to 492 (full length); (B) 1 to 261; (C) 1 to 108; (D) 99 to 492; (E) 252 to 492; (F) 263 to 492.

that regions beyond MBD are important for the wider function of MeCP2.

MeCP2 has been found in all somatic cell types of the mouse that have been tested but is absent or rare in germ cells of the testis and ES cells (female germ cells have not been studied). What might be the function of an abundant chromosomal protein that interacts with methyl-CpG in somatic cells? The best-known effects of DNA methylation concern transcriptional repression, and therefore MeCP2 is a candidate repressor. The preferential localization of mouse MeCP2 in transcriptionally inert heterochromatin is in line with this expectation. So far, however, we have not observed specific effects of MeCP2 on transcription from methylated templates by using naked DNA in in vitro transcription extracts, as addition of MeCP2 suppressed all transcription, regardless of the methylation status of the DNA (17). One possible explanation for this is that MeCP2 can only interact appropriately with methylated DNA in the context of chromatin.

Understanding of heterochromatin has recently been improved by the isolation of some of its constituent proteins (7, 11, 23, 32). It is possible that MeCP2 represents part of a complex of proteins involved in the formation and maintenance of heterochromatin in mouse cells, particularly in differentiated cell types. While MeCP2 may participate in heterochromatin formation in some cells, it is unlikely to be an obligatory component, as DNA MTase-deficient cells, which have very little genomic methylation, and ES cells, which have negligible levels of MeCP2, have heterochromatin of apparently normal appearance. Moreover, it is clear that MeCP2 is not confined to heterochromatin; it is also present in the euchromatic chromosome arms. This is particularly apparent in the rat, in which the amount of satellite and constitutive heterochromatin is low, and MeCP2 localizes all over the metaphase chromosome arms without appearing to concentrate in discrete regions of the genome (14). It may be that MeCP2 exerts its primary function in euchromatin rather than heterochromatin.

Recent results from this laboratory show that MeCP2, like DNA MTase, is essential during mouse development, as mouse embryos lacking the MeCP2 gene exhibit anomalies and fail to develop beyond midgestation (30). Both MeCP2 and DNA MTase, however, are dispensable in ES cells, suggesting that DNA methylation does not exert an essential function in these pluripotential stem cells. Indeed, cells of the inner cell mass, from which ES cells are derived, are thought to have very low levels of DNA methylation (12), although ES cells themselves often have heavily methylated genomes (unpublished observations). In view of the overlapping phenotypic effects of mutations in the DNA MTase and MeCP2, and of the dependence of MeCP2 on DNA methylation for its localization to chromosomes, it is reasonable to hypothesize that MeCP2 is a major mediator of the effects of DNA methylation during development. In this case, further studies of this protein should lead to a molecular explanation for at least some of the biological consequences of DNA methylation.

ACKNOWLEDGMENTS

We thank Bill Skarnes (BBSRC Centre for Genome Research, Edinburgh, United Kingdom) for the pββgeopA vector, Martin Hooper (Department of Pathology, University of Edinburgh) for the EB28/10 cell line, Richard Meehan for anti-MeCP2 antibodies and a critical reading of the manuscript, Jack Price for the polyclonal anti-β-Gal antiserum, and Evani Viegas-Pequignot (INSERM, Paris, France) for the anti-m^5C monoclonal antibody. We are also grateful to Joan Davidson and Aileen Greig for technical help and Christine Struthers for secretarial support.

This work was supported by the Wellcome Trust, the Imperial Cancer Research Fund, the Howard Hughes Medical Institute, and a grant from Bristol-Myers/Squibb (E.L.).

REFERENCES

1. **Bird, A. P.** 1993. Functions for DNA methylation in vertebrates. Cold Spring Harbor Symp. Quant. Biol. **LVIII:**281–285.
2. **Boyes, J., and A. Bird.** 1991. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell **64:**1123–1134.
3. **Boyes, J., and A. Bird.** 1992. Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. EMBO J. **11:**327–333.
4. **Casadaban, M. J., J. Chou, and S. N. Cohen.** 1980. In vitro gene fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detection and cloning of translational initiation signals. J. Bacteriol. **143:**971–980.
5. **Chu, G., and P. A. Sharp.** 1981. A gene chimaera of SV40 and mouse β-globin is transcribed and properly spliced. Nature (London) **289:**378–382.
6. **Churchill, M. E. A., and A. A. Travers.** 1991. Protein motifs that recognize structural features of DNA. Trends Biochem. Sci. **16:**92–97.
7. **Eissenberg, J. C., T. C. James, D. M. Foster-Hartnett, T. Hartnett, and S. Elgin.** 1990. Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **87:**9923–9927.
8. **Frederickson, R. M., M. R. Micheau, A. Iwamoto, and N. Miyamoto.** 1989. 5′ flanking and first intron sequences of the human β-actin gene required for efficient promoter activity. Nucleic Acids Res. **17:**253–270.
9. **Friedrich, G., and P. Soriano.** 1991. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. Genes Dev. **5:**1513–1523.
10. **Gossler, A., A. L. Joyner, J. Rossant, and W. C. Skarnes.** 1989. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. Science **244:**463–465.
11. **James, T. C., and S. C. R. Elgin.** 1986. Identification of a nonhistone chromosomal protein associated with heterochromatin in *Drosophila melanogaster* and its gene. Mol. Cell. Biol. **6:**3862–3872.
12. **Kafri, T., M. Ariel, M. Brandeis, R. Shemer, L. Urven, J. McCarrey, H. Cedar, and A. Razin.** 1992. Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. Genes Dev. **6:**705–714.
13. **Leonhardt, H., A. W. Page, H.-U. Weier, and T. H. Bestor.** 1992. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell **71:**865–873.
14. **Lewis, J. D., R. R. Meehan, W. J. Henzel, I. Maurer-Fogy, P. Jeppesen, F. Klein, and A. Bird.** 1992. Purification, sequence and cellular localisation of a novel chromosomal protein that binds to methylated DNA. Cell **69:**905–914.

14a. **Li, E.** Unpublished data.

15. **Li, E., C. Beard, A. Forster, T. Bestor, and R. Jaenisch.** 1993. DNA methylation, genomic imprinting and mammalian development. Cold Spring Harbor Symp. Quant. Biol. **LVIII:**297–305.
16. **Li, E., T. H. Bestor, and R. Jaenisch.** 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell **69:**915–926.
17. **Meehan, R. R., J. D. Lewis, and A. P. Bird.** 1992. Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. Nucleic Acids Res. **20:**5085–5092.
18. **Meehan, R. R., J. D. Lewis, S. McKay, E. L. Kleiner, and A. P. Bird.** 1989. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell **58:**499–507.
19. **Miller, O. L., W. Schnedl, J. Allen, and B. F. Erlanger.** 1974. Immunofluorescent localization of 5MC. Nature (London) **251:**636–637.
20. **Miniou, P., M. Jeanpierre, V. Blanquet, V. Sibella, D. Bonneau, C. Herbelin, A. Fischer, A. Niveleau, and E. Viegas-Pequignot.** 1994. Abnormal methylation pattern in constitutive and facultative (X inactive chromosome) heterochromatin of ICF patients. Hum. Mol. Genet. **3:**2093–2102.
21. **Morgan, R. H. M., J. A. Henry, and M. L. Hooper.** 1983. Isolation of cell lines from differentiating embryonal carcinoma cultures. Exp. Cell Res. **148:**461–473.

21a. **Nan, X.** Unpublished results.

22. **Nan, X., R. R. Meehan, and A. Bird.** 1993. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nucleic Acids Res. **21:**4886–4892.
23. **Nicol, L., and P. Jeppesen.** 1994. Human autoimmune sera recognize a conserved 26 kD protein associated with mammalian heterochromatin that is homologous to heterochromatin protein of *Drosophila*. Chromosome Res. **2:**245–253.
24. **Reynaud, C., C. Bruno, P. Boullanger, J. Grange, S. Barbesti, and A. Niveleau.** 1991. Monitoring of urinary excretion of modified nucleosides in cancer patients using a set of six monoclonal antibodies. Cancer Lett. **61:**255–262.
25. **Robbins, J., S. Dilworth, R. Laskey, and C. Dingwall.** 1991. Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. Cell **64:**615–623.
26. **Robertson, E. J.** 1987. Embryo-derived stem cell lines, p. 19–49. *In* E. J. Robertson (ed.), Teratocarcinomas and embryonic stem cells, a practical approach. IRL Press, Oxford.
27. **Selden, R. F.** 1991. Transfection using DEAE dextran, p. 9.2.1–9.2.6. *In* F. M. Ausubel et al. (ed.), Current protocols in molecular biology, vol. 1. Wiley Interscience, New York.

27a. **Skarnes, W.** Personal communication.

28. **Smith, A. G., J. K. Heath, D. D. Donaldson, G. G. Wong, J. Moreau, M. Stahl, and D. Rogers.** 1988. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature (London) **336:**688–690.
29. **Tate, P.** 1994. MeCP2: a methyl-CpG binding protein essential for embryonic development in the mouse. Ph.D. thesis. University of Edinburgh, Edinburgh, United Kingdom.

29a. **Tate, P.** Unpublished observations.

30. **Tate, P., W. Skarnes, and A. Bird.** The methyl-CpG binding protein MeCP2 is essential for embryonic development in the mouse. Submitted for publication.
31. **Tate, P. H., and A. Bird.** 1993. Effects of DNA methylation on DNA-binding proteins and gene expression. Curr. Opin. Genet. Dev. **3:**226–231.
32. **Wreggett, K. A., F. Hill, P. S. James, A. Hutchings, G. W. Butcher, and P. B. Singh.** 1994. A mammalian homologue of *Drosophila* heterochromatin protein 1 (HP 1) is a component of constitutive heterochromatin. Cytogenet. Cell Genet. **66:**99–103.