CpG binding protein is a nuclear matrix- and euchromatin-associated protein localized to nuclear speckles containing human trithorax:
Identification of nuclear matrix targeting signals.

Jeong-Heon Lee and David G. Skalnik*

Herman B Wells Center for Pediatric Research, Section of Pediatric
Hematology/Oncology, Department of Pediatrics, and Department of Biochemistry and
Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana 46202

*Corresponding author to whom inquiries should be addressed:
David G. Skalnik, Ph.D
Herman B Wells Center for Pediatric Research
Cancer Research Building, Room 472
Indiana University School of Medicine
1044 West Walnut Street
Indianapolis, Indiana 46202.
Phone: (317) 274-8977
Fax: (317) 274-8928
E-mail: dskalnik@iupui.edu

Running Title: Sub-cellular localization of CpG binding protein
SUMMARY

CpG binding protein (CGBP) binds unmethylated CpG dinucleotides and is essential for mammalian development. CGBP exhibits a punctate nuclear localization correlated with DAPI-light regions and is excluded from metaphase chromosomes. The distribution of CGBP is distinct from the heterochromatin-associated proteins MBD1, MeCP2, and HP1α. Some CGBP-containing nuclear speckles co-localize with splicing factor SC-35 and actively transcribed regions of the genome, while most CGBP co-localizes with acetylated histones, indicating that CGBP is localized to active chromatin. CGBP contains two nuclear localization signals that are insufficient to direct punctate sub-nuclear distribution. Instead, localization of CGBP to nuclear speckles requires signals within the acidic, basic, and coiled-coil domains. CGBP associates with the nuclear matrix, and fragments of CGBP that fail to associate with the nuclear matrix fail to localize to nuclear speckles and exhibit reduced transcriptional activation activity. Mutated versions of CGBP that lack DNA-binding activity exhibit a normal nuclear distribution, suggesting that CGBP accumulates at nuclear speckles as a result of protein/protein interactions. Importantly, the sub-cellular distribution of CGBP is identical to human trithorax, suggesting that these proteins may be components of a multimeric complex analogous to the histone-methylating Set1 complex of Saccharomyces cerevisiae that contains CGBP and trithorax homologues.
INTRODUCTION

Cytosine methylation of CpG motifs is an important epigenetic modification in higher eukaryotes. Methylated DNA is generally associated with transcriptionally inactive heterochromatin (1). In contrast, actively expressed genes are generally hypo-methylated and found in an open euchromatin configuration. Appropriate cytosine methylation is essential for normal mammalian development. For example, individual disruption of the genes encoding the methyltransferases Dnmt1, Dnmt3a, or Dnmt3b result in an embryonic-lethal phenotype in mice (2,3), and overexpression of Dnmt1 leads to hyper-methylation of DNA, loss of genomic imprinting, and embryonic lethality (4). More subtle mutations within human Dnmt3b result in ICF (immunodeficiency, centromere instability, and facial anomalies) syndrome (5), and mutations of methyl-CpG binding protein 2 (MeCP2) lead to Rett syndrome, a progressive neurodegenerative disorder (6). In addition, hyper-methylation of tumor suppressor genes is commonly observed in cancer cells (7). Despite the importance of appropriate cytosine methylation patterns for normal mammalian development, little is known regarding the regulation of this process.

Several DNA-binding proteins have been described that interact specifically with methylated CpG motifs. These include MeCP2, methyl CpG-binding domain (MBD) protein 1, MBD2, and MBD4 (8). In addition, MBD3 is a component of the Mi-2 histone deacetylase and nucleosome remodeling complex (9). Similarly, MBD2 is a component of the MeCP1 histone deacetylase complex (10), thus linking cytosine methylation with histone acetylation and providing a unifying framework for the control of chromatin structure and gene regulation.

CpG binding protein (CGBP) is a recently described DNA-binding protein that exhibits a novel binding affinity for DNA sequences containing unmethylated CpG motifs (11). Consistent with this binding specificity, CGBP acts as a trans-activator of transcription in co-transfection experiments. Interestingly, the single DNA-binding domain of CGBP is comprised of a cysteine-rich CXXC domain (12). The CXXC domain is only found in a few other proteins, including Dnmt1, MBD1, human trithorax (HRX/MLL/ALL-1), and MLL-2 (11,13,14). Reciprocal chromosomal translocations involving the HRX gene are
commonly observed in acute leukemia. Little is known regarding the function of the related MLL-2 gene, although it is amplified in some cancer cell lines (15,16). In contrast to CGBP, these other CXXC domain proteins contain additional distinct DNA-binding domains, and the role of the CXXC domain within these factors is unclear. However, recent studies demonstrate that similar to CGBP, CXXC domains within MBD1 and HRX exhibit binding affinity for unmethylated CpG motifs (17,18).

Ligand-selection of high-affinity binding sites from a pool of degenerate double-stranded oligonucleotides reveals a consensus binding site for CGBP of (A/C)CpG(A/C) (12). Individual mutation of highly conserved cysteine residues within the CXXC domain of CGBP completely abrogates DNA-binding activity, and this domain requires the presence of zinc for efficient DNA-binding activity (12). In addition, CGBP contains two copies of the plant homeodomain (PHD), a motif found in several dozen proteins implicated in modulation of chromatin structure and gene regulation (11,19). Recently, the PHD finger of the polycomb-like protein was found to mediate direct interaction with the protein "enhancer of zeste" (20), and a PHD domain within the chromatin-associated KAP-1 protein cooperates with a bromodomain to recruit a chromatin-remodeling complex (21). Furthermore, the PHD finger within CREB binding protein contributes to this factor’s acetyltransferase activity (22).

Importantly, we have recently reported that CGBP-null mice exhibit a peri-implantation lethal phenotype (23), thus demonstrating the requirement of CGBP for normal mammalian development. Although the molecular basis for this phenotype is still under investigation, it is intriguing that CGBP-null embryos die at a developmental stage during which global changes of cytosine methylation occurs (24,25). Given the DNA-binding specificity of CGBP and its structural similarity to other proteins involved in the regulation and function of cytosine methylation and chromatin structure, it is tempting to speculate that CGBP-null embryos die from dysregulated gene expression, at least partially as a consequence of altered chromatin structure. In an effort to better understand the global function of CGBP, and its possible role in modulating chromatin
structure, experiments were performed to determine the sub-cellular localization of CGBP and define the protein domains required to direct this distribution.
EXPERIMENTAL PROCEDURES

Cell culture and transfection

HEK-293 (human embryonic kidney) and NIH-3T3 (mouse embryonic fibroblast) cells were grown as monolayers in Dulbecco’s modified eagle’s medium (DMEM) supplemented with 10% fetal calf serum (Hyclone, Logan, UT) at 37°C in a humidified atmosphere of 5% CO₂. For sub-cellular localization and immunofluorescence experiments, cells were cultured on a coverslip in a 24-well dish and transfected with 1-2 µg of DNA by calcium-phosphate co-precipitation (26). Following transfection, cells were grown in fresh media for 24 h, and fixed with cold methanol or 4% paraformaldehyde. For co-transfection assays, cells at ~60% confluency were transfected with 10-20 µg of DNA by calcium phosphate co-precipitation. Following transfection, cells were grown in fresh media for 24-48 hours prior to analysis. Trans-activation assays were performed by co-transfecting HEK-293 cells with a fixed amount (2.5 µg) of CMV-based pEGFP reporter vector and various amounts of pcDNA3 FLAG-CGBP expression vector. The total amount of DNA in each sample was equalized by adding pcDNA3 plasmid. Cells were harvested and analyzed by western blot as described above.

Sub-nuclear biochemical fractionation and western blot analysis

Sequential nuclear extraction was performed as described (27). Cells were transfected with 5–10 µg of DNA by calcium-phosphate co-precipitation. Following transfection, cells were grown in fresh media for 48 h and used for analysis. Cells were washed with cold phosphate buffered saline (PBS) and extracted with cytoskeleton buffer (CSK) containing 10 mM Pipes (pH 6.8), 100 mM NaCl, 300 mM sucrose, 3 mM MgCl₂, 1 mM EGTA, supplemented with the protease inhibitors leupeptin, aprotinin, and pepstatin (1 µg/ml each), 1 mM PMSF, 1 mM DTT, and 0.5% (vol/vol) Triton X-100. After extraction for 5 min on ice, the proteins solubilized in CSK buffer were recovered following separation from the insoluble fraction by centrifugation at 5,000 x g for 3 min. Chromatin was solubilized by digestion with 30 units of DNase I (RNase-free) in CSK buffer containing protease inhibitors for 20 min at 37°C. The chromatin fraction was extracted for 5 min on ice by adding 1 M ammonium sulfate in CSK buffer to a final concentration of 0.25 M and centrifuging as above. The pellet was further extracted with
2 M NaCl in CSK buffer for 5 min on ice. The remaining pellet was solubilized with sodium phosphate buffer containing 8 M urea, which was considered the nuclear matrix fraction. Protein concentration was determined by the Bradford assay. For western blot analysis, an equivalent proportion of each fraction was solubilized in Laemmli sample buffer (28). Following electrophoresis on a 6-12% SDS polyacrylamide gel, proteins were transferred onto nitrocellulose membrane (MSI, Westborough, MA). The membrane was then incubated with either anti-CGBP antibody (11), anti-acetyl histone H3 antibody (Upstate Biotech, Lake Placid, NY), anti-NuMA antibody (Oncogene, Cambridge, MA), anti-FLAG antibody (Sigma), or anti-GFP antibody (Clontech, Palo Alto, CA) followed by HRP-labeled secondary antibody, and detected by using an ECL detection kit (Amersham Pharmacia Biotech, Piscataway, NJ) according to the manufacturer’s instructions.

Construction of truncated CGBP expression vectors

The cDNA of human CGBP (11) was subcloned into the pEGFP-C2 (Clontech) and pcDNA3-FLAG vectors (29). The cDNA of human MBD1 was amplified from a cDNA clone (IMAGE:361404) (ResGen, Huntsville, AL) and subcloned into the pEGFP vector. PCR amplification was performed using the primer sequences 5’-aagcttgaattcatggctgaggactggctggactgc-3’ and 5’-tcaaggatcctgctttctagctccaggttttttaag-3’. pSVK-FLAG-MeCP2 and pSVK-FLAG-HP1α were described previously (30), and were generously provided by Dr. Steven Baylin (Johns Hopkins University School of Medicine). pF-MLLΔT, encoding amino acids 1-1436 of human HRX was described previously (31), and was generously provided by Dr. Jay Hess (Washington University School of Medicine).

Deletion mutations of CGBP were prepared using a combination of restriction enzyme digestions and PCR amplification using Pfu polymerase (Stratagene, La Jolla, CA), and subcloned into pEGFP and/or pcDNA3-FLAG vectors. Site-directed mutagenesis was performed on the CXXC domain of CGBP using primers that mutate cysteine residues to alanine using the QuickChange site directed mutagenesis kit (Stratagene, CA) in accordance with the protocol provided by the manufacturer.
oligonucleotides include C169A (mutates cysteine 169 to alanine), 5’-cggctacgcccgcgtgc
ctggtgtctgtgcagcag-3’; and C208A (mutates cysteine 208 to alanine), 5’-gcggctgcccgcgtgc
gccagctgcggccc-3’. Each mutation was subcloned into the pFLAG-CMV (Sigma Chem. Co., St. Louis, MO) and pEGFP vectors. The nucleotide sequences of truncated and mutated CGBP constructs were confirmed by automated DNA sequencing.

Indirect immunofluorescence and confocal microscopy

Cells were seeded onto a cover glass at 2–5 x 10⁴ cells/well in a 24-well dish and transfected as described above. Cells were then washed twice with cold PBS and fixed with 4% (vol/vol) paraformaldehyde in PBS for 20 min at room temperature, and then washed with PBS. Cells were permeabilized with 0.2% Triton X-100 in PBS for 10 min at room temperature, and a blocking solution (PBS containing 2.5% normal serum (Santa Cruz Biotech, Santa Cruz, CA) and 0.2% Tween 20) was added and incubated for 1 h at room temperature. Anti-CGBP rabbit IgG (1:500) (11), anti-acetyl histone 3 rabbit IgG (1:100, 5 µg/ml) (di-acetylated Lys9, Lys14; catalog #06-599 Upstate Biotech.) and anti-acetyl-histone 4-rabbit IgG (tetra-acetylated Lys4, Lys7, Lys11, Lys15; Upstate Biotech., catalog #06-598) (1:100, 5 µg/ml), (Upstate Biotech.), anti-SC-35 (1:500, 9.2 µg/ml) (Sigma Chem. Co.), or anti-FLAG-mouse IgG (1:1000, 3.5 µg/ml)(Sigma Chem. Co.) were added and incubated for 2 h at room temperature. Cells were then washed three times with PBS containing 0.2% Tween 20 for 5 min. Appropriate secondary antibody labeled with Texas Red (2 µg/ml in blocking solution) (Santa Cruz Biotech.) was added and incubated for 1 h at room temperature. Cells were washed three times with PBS containing 0.2% Tween 20 for 5 min. Nuclear counterstaining was performed with 0.1 µg/ml DAPI in PBS for 5 min followed by washing with PBS. Cells were mounted with 10 µl of Fluoromount G (Southern Biotechnology Associates, Birmingham, AL) and observed using a fluorescence microscope or were scanned with a Zeiss LSM 510 laser scanning confocal microscope. Bromouridine triphosphate (BrU) was used for in situ detection of RNA synthesis in permeabilized cells using immunodetection with antibody directed against bromouridine (Harlan Sera-lab, Indianapolis, IN) as previously described (32).
RESULTS

Determination of the sub-cellular localization of CGBP

Confocal immuno-fluorescence was performed on HEK-293 and NIH-3T3 cells using antisera directed against CGBP. The endogenous CGBP protein is localized to the nucleus in both cell lines, and exhibits a punctate or speckled distribution (Fig. 1A). Transfection of these cells with a vector expressing GFP-tagged full-length human CGBP results in a similarly speckled nuclear distribution of the GFP-CGBP fusion protein (Fig. 1A). These speckles are concentrated in areas of DAPI-light staining, consistent with localization of CGBP to euchromatic regions of the nucleus (Fig. 1B).

Co-localization experiments were conducted to better characterize the nature of the nuclear speckles containing CGBP. Cells were co-transfected with vectors expressing epitope-tagged (GFP or FLAG) CGBP or an epitope-tagged marker protein. The results presented in figure 2 indicate that CGBP exhibits a sub-nuclear distribution that is entirely distinct from that of MBD1, MeCP2, and HP1α, each of which has been found to localize within regions of heterochromatin (8,30,33,34). In contrast, a small fraction of CGBP-containing speckles co-localize with the RNA splicing factor SC-35 and with actively transcribed regions of the genome. However, CGBP exhibits significant co-localization with acetylated histone H3 and acetylated histone H4, which are markers for areas of euchromatin. Importantly, the set of nuclear speckles that contain CGBP is identical to that detected as containing the amino terminus (amino acids 1-1436) of HRX, which has previously been reported to localize within a novel class of nuclear speckles and to associate with the nuclear matrix (35-37).

Further microscopy studies were performed to examine the localization of CGBP in metaphase cells. As previously reported (18), MBD1 is found tightly associated with heterochromatic condensed chromosomes at metaphase (Fig. 3, bottom row). In contrast, GFP-CGBP exhibits a highly diffuse distribution during this phase of the cell cycle (Fig. 3, top row). It is excluded from the condensed chromosomes, and does not exhibit a punctate distribution during metaphase.
Lastly, biochemical fractionation experiments were utilized as an independent method to further assess the sub-cellular localization of CGBP. HEK-293 and NIH-3T3 cells were successively extracted to recover soluble, chromatin-associated, and nuclear matrix-associated protein fractions. Western blot analyses demonstrate that endogenous CGBP is nearly exclusively associated with the nuclear matrix fraction (Fig. 4). Western blots were also performed with antisera directed against NuMA and acetylated histone H3 as controls for the integrity and purity of the nuclear matrix and chromatin-associated protein fractions, respectively (38,39). We conclude that CGBP associates with the nuclear matrix, consistent with the previously reported nuclear matrix association of the co-localizing protein HRX (35).

Identification of protein domains responsible for nuclear localization of CGBP and for targeting to nuclear speckles and association with the nuclear matrix.

Experiments were performed to identify the signals required for CGBP association with the nuclear matrix and nuclear speckles. For example, the DNA-binding activity of Ikaros has been found to be essential for its localization to heterochromatin speckles (40). Experiments were conducted to determine if the DNA-binding activity of CGBP is similarly required for the observed sub-cellular distribution. Mutations of conserved cysteine residues within the CXXC domain that ablate DNA-binding activity (12) were introduced into GFP-CGBP fusion proteins. The sub-cellular localization of these DNA-binding deficient forms of CGBP was assessed in both NIH-3T3 and HEK-293 cells. Confocal microscopy and biochemical fractionation reveals that both of these mutated constructs co-localize with wild type full-length CGBP FLAG-tagged protein (Fig. 5A), and associate with the nuclear matrix (Fig. 5B). Hence, DNA-binding activity is not required for targeting of CGBP to nuclear speckles. Rather, this localization is presumably mediated via protein/protein interactions.

We have previously demonstrated that CGBP strongly trans-activates the cytomegalovirus (CMV) promoter, which contains numerous CpG motifs (11). As expected, the DNA-binding deficient CGBP constructs fail to trans-activate a co-transfected reporter gene vector comprised of the CMV promoter/enhancer driving
expression of GFP (Fig. 5C). Note that wild type CGBP additionally auto-activates expression of the CGBP expression vector, which also contains the CMV promoter (Fig. 5C, top panel).

In order to define the nuclear localization signals within CGBP, a series of truncated versions of CGBP (Fig. 6A) were generated and expressed in NIH-3T3 cells as GFP fusion proteins (Fig. 6B). Successive truncations from the carboxyl terminus of CGBP reveals that a fragment as short as amino acids 1-122 is localized to the nucleus (Fig. 6). However, truncation to amino acids 1-103 results in a cytoplasmic localization. Hence, a nuclear localization signal resides within amino acids 104-122 of CGBP. Inspection of the amino acid sequence of this domain (DEGGGRKRPRPVPDPDLQRRRA) (11) reveals a cluster of basic residues (in bold) consistent with a consensus bipartite nuclear localization signal (RKR X₈ RR) (41).

Importantly, the 302-656 amino acid fragment of CGBP is also partially localized to the nucleus, indicating the presence of at least one additional nuclear localization signal in the carboxyl half of the CGBP protein. Expression of the basic domain of CGBP (amino acids 318-367) as a GFP fusion protein is sufficient to direct efficient nuclear localization, thus defining a second nuclear localization signal. This forty amino acid fragment contains 65% histidine, arginine, or lysine residues. The GFP fusion protein containing CGBP amino acids 361-481 is evenly distributed between the nucleus and cytoplasm, suggesting a non-specific diffusion of this small peptide throughout the cell. The downstream fragment of CGBP containing amino acids 483-656 is excluded from the nucleus, indicating the absence of nuclear localization signals downstream of the coiled-coil domain. It is surprising that the isolated basic domain localizes to the nucleus more efficiently than a longer fragment (amino acids 302-656), which also contains the basic domain. Presumably elements within the downstream sequence act to mask the nuclear localization signal located within the basic domain when expressed in this fashion. Addition of the upstream CGBP nuclear localization signal (amino acids 109-121) to the 302-656 construct results in efficient nuclear localization (Fig. 6B).
Additional truncated versions of GFP-CGBP fusion proteins were expressed in NIH-3T3 cells to identify protein domains required to target CGBP to nuclear speckles. A fragment of CGBP (amino acids 1-481) lacking the PHD2 domain exhibits a speckled nuclear distribution (Fig. 6B). However, additional truncation to amino acid 367 (amino acids 1-367), which deletes the coiled-coil domain, results in a partially speckled pattern with much of the GFP-fusion protein distributed diffusely throughout the nucleus. Further truncation to amino acid 320 (amino acids 1-320), which additionally deletes the basic domain, results in a complete loss of nuclear speckling, and a diffuse nuclear distribution. Hence, signals encoded within the basic and coiled-coil domains are required for directing a punctate distribution of CGBP.

Truncation of the amino terminus of the protein reveals that amino acids 1-301, including the PHD1 domain, CXXC DNA-binding domain, and a portion of the acidic domain, are not required for directing a punctate distribution, as the CGBP fragment containing amino acids 302-656 exhibits a speckled nuclear distribution (Fig. 6B). The CGBP fragment containing amino acids 103-481 also exhibits a punctate nuclear distribution, similar to the pattern produced by expression of full-length GFP- or FLAG-CGBP (Fig. 6B and C). Similar to the results described above, removal of the coiled-coil domain from the carboxyl end of this fragment of CGBP (leaving amino acids 103-367) leads to a partially speckled nuclear distribution, while additional removal of the basic domain (leaving amino acids 103-320) leads to a diffuse nuclear distribution. Step-wise truncations from the amino terminus of the 103-481 amino acid CGBP fragment produce similar results. Truncation that leaves amino acids 213-481, which removes the CXXC domain, leads to a speckled sub-nuclear distribution. Additional truncation to position 317 (leaving amino acids 318-481), which removes the acidic domain, leads to a partially speckled distribution with a background of diffuse localization. This indicates that the acidic domain, which contains 29% glutamate and aspartate residues, participates in the appropriate localization of CGBP to nuclear speckles.

Importantly, however, none of the individual domains located in the central region of CGBP are sufficient to direct a speckled nuclear distribution (Fig. 6B). Expression of the
coiled-coil domain (amino acids 361-481), acidic domain (amino acids 213-320), or basic domain (amino acids 318-367) as individual GFP fusion proteins results in a diffuse distribution throughout the nucleus. The coiled-coil and acidic domains were linked to the amino terminus CGBP nuclear localization signal (amino acids 109-121), as these fragments lack both of the CGBP nuclear localization signals. Hence, multiple domains distributed throughout the central region of CGBP contribute cooperative signals that lead to a speckled nuclear distribution.

Biochemical fractionation experiments were also performed on cells expressing truncated GFP-CGBP fusion proteins to compare nuclear matrix association with the sub-nuclear distribution observed by confocal microscopy. These two parameters are found to be highly correlated. Without exception, CGBP constructs that exhibit a predominant punctate nuclear distribution, such as fragments containing amino acids 1-481, 302-656, 103-481, and 213-481 are also nearly exclusively associated with the nuclear matrix (Fig. 7). However, constructs that exhibit a partially speckled distribution, such as fragments containing amino acids 1-367, 103-367, 318-367, or 318-481, are partitioned between the chromatin-associated fraction and the nuclear matrix fraction. Constructs that exhibit a diffuse nuclear staining, such as fragments containing amino acids 1-320, 103-320, or 361-481 are found predominantly in the soluble fraction.

We conclude that multiple signals present within the central portion of CGBP are responsible for association with the nuclear matrix and generation of a punctate sub-nuclear distribution. Signals within the acidic, basic, and coiled-coil domains all contribute to this localization, yet no individual domain is sufficient to direct this sub-cellular targeting. Interestingly, in addition to its activity as a nuclear localization signal, the basic domain appears essential for proper sub-nuclear localization of CGBP, and is sufficient for directing at least a partially speckled distribution when linked to either the coiled-coil or acidic domains. Although not sufficient for normal CGBP localization, expression of the isolated basic domain as a GFP-fusion protein permits partial association with the chromatin fraction.
The functional significance of the association of CGBP with the nuclear matrix and nuclear speckles was examined in co-transfection assays. As demonstrated in figure 5C, CGBP trans-activates expression of a GFP reporter gene under the control of the CMV promoter, which contains several dozen CpG motifs. A similar analysis was performed with several truncated CGBP constructs that exhibit various degrees of nuclear speckling and association with the nuclear matrix. All of the constructs tested contain both the acidic domain that is sufficient for trans-activation activity of CGBP (42), and the CXXC DNA-binding domain (11,12). A construct encoding amino acids 1-481, which exhibits a sub-nuclear distribution identical to that of the full-length CGBP protein, exhibits a trans-activation activity similar to the wild type CGBP construct (Fig. 8). The construct containing amino acids 1-367, which exhibits a partially speckled distribution and partial association with the nuclear matrix, exhibits a reduced trans-activation activity (~50% of wild type when normalized for the level of FLAG-CGBP expression). Importantly, the CGBP construct containing amino acids 1-320, which exhibits a diffuse nuclear distribution and fails to associate with the nuclear matrix, exhibits a dramatically reduced trans-activation activity (~15% of wild type) despite containing DNA-binding and trans-activation domains. Hence, appropriate sub-nuclear targeting appears to be critical for normal CGBP function.
DISCUSSION

These studies were conducted to determine the physical distribution of CGBP within a cell in the hopes that such information will offer insight into the global function and mechanism of action of CGBP. CGBP localizes nearly exclusively to euchromatic nuclear speckles. The distribution of CGBP is distinct from that exhibited by the heterochromatin-associated proteins MBD1, MeCP2, and HP1α. In contrast, the distribution of CGBP slightly overlaps with the RNA splicing factor SC-35 and with actively transcribed regions of the genome, and exhibits significant overlap with a subset of regions containing acetylated histones. Hence, CGBP localizes to areas of euchromatin rather than sites of active transcription, suggesting that it plays a role in modulation of chromatin structure.

The CGBP protein contains two nuclear localization signals, located upstream of the CXXC DNA-binding domain and within the basic domain. Neither of these nuclear localization signals is capable of directing a speckled nuclear distribution. Rather, the signals that direct CGBP to active chromatin nuclear speckles are distributed throughout an extended central region of the protein that includes acidic, basic, and coiled-coil domains. However, none of these domains is individually able to comprise a nuclear speckle targeting signal, although the basic domain associates with the chromatin fraction and directs partial targeting to nuclear speckles when linked to either the adjacent acidic or coiled-coil domains. In addition, biochemical fractionation experiments demonstrate that CGBP associates nearly exclusively with the nuclear matrix, an interaction that is independent of DNA-binding activity, and is presumably a consequence of protein/protein interactions. Association of CGBP with the nuclear matrix appears to be causally related to localization to nuclear speckles, as these two characteristics coincide for all mutated versions of CGBP analyzed. Importantly, association of CGBP with nuclear speckles and the nuclear matrix is functionally important for the trans-activation activity of this factor.

Interestingly, CGBP uniquely localizes to a set of nuclear speckles that also contains HRX (MLL/ALL-1). HRX is the human homologue of Drosophila trithorax, which is a
regulator of chromatin structure and homeobox gene expression. HRX is at least partially associated with the nuclear matrix (35), and HRX-containing nuclear speckles have been shown to be distinct from nuclear speckles that contain PML or TAL-1 (36). Consistent with the results reported here, HRX-containing speckles were previously found to be distinct from sites of active transcription, as determined by the localization of the SC-35 splicing factor (43). Interestingly, in contrast to HRX that associates with condensed chromosomes during mitosis, CGBP fails to interact with metaphase chromosomes. Hence, these proteins co-localize to a unique class of nuclear speckles during interphase, yet exhibit disparate sub-cellular localizations during cell division, indicating that CGBP-containing nuclear speckles are dynamic structures that undergo remodeling throughout the cell cycle.

Although CGBP and HRX both contain CXXC and PHD domains, analysis of truncation mutations indicates that neither of these domains is responsible for directing CGBP to HRX-containing nuclear speckles. Similar to CGBP, the signals directing HRX to nuclear speckles are complex. At least two domains (SNL-1 and SNL-2) within the amino terminus of HRX are capable of directing HRX to nuclear speckles (37). However, a version of HRX lacking these two domains still localizes to nuclear speckles, indicating that at least one additional domain is involved in this sub-cellular targeting. Comparison of the SNL-1 and SNL-2 sequences of HRX with the central region of CGBP reveals intriguing pockets of sequence homology. A highly basic region within SNL-2 (amino acids 1067-1074; RIKHVCRR) shares 63% identity and 88% similarity to sequence present within the basic domain of CGBP (amino acids 326-333; KVHKVKRR). Mutation of residues 1065-1073 of HRX results in the loss of ~80% of nuclear speckles (37). Similarly, mutation of amino acids 418-423 (SSRIIK) of HRX results in the loss of 60% of nuclear speckles. Sequence comparison reveals that this region is imbedded within a short domain that exhibits 40% identity and 60% similarity with a region of the CGBP acidic domain, amino acids 260-269 (AVASSTVKEP). Whether these short regions of similarity will prove to be critical for directing association of CGBP with the nuclear matrix and localization to nuclear speckles is currently under investigation.
Co-localization of HRX and CGBP to an identical set of nuclear speckles suggests that they may directly interact or may be components of a common multimeric protein complex. A direct interaction of HRX with CGBP has not been reported, despite several yeast two-hybrid screens performed with the amino terminal portion of HRX that co-localizes with CGBP to nuclear speckles (44-50). The alternative hypothesis that CGBP and HRX are components of a common multimeric complex gains additional support from studies in *Saccharomyces cerevisiae*, in which the yeast homologue of CGBP (Spp1) and a trithorax family member (Bre2) are both components of a complex denoted COMPASS or Set1 (51-54). This complex contains 7-8 components, and is required for methylation of the K4 residue of histone H3, which is associated with transcriptionally competent chromatin (55). The Set1 complex is also involved in gene silencing at telomeres and mating type loci (56), and activation of DNA repair genes (57). The results presented here provide evidence that an analogous complex may exist in higher eukaryotes. Yeast two-hybrid screening failed to demonstrate a direct interaction between Spp1 and Bre2 (51), consistent with the failure to detect an interaction with CGBP using HRX as the bait in two-hybrid screens.

Direct analysis of the protein interactions of CGBP in vivo are complicated by the finding that CGBP is nearly exclusively associated with the nuclear matrix, thus making co-immunoprecipitation and affinity pull-down assays difficult. However, the results described here predict that CGBP -/- cells may exhibit abnormal patterns of histone modification and chromatin structure, consistent with peri-implantation embryonic death of CGBP-null embryos, a time during which the genome undergoes global remodeling of chromatin structure and cytosine methylation (24,25,58).
ACKNOWLEDGMENTS

We thank Dr. Steven Baylin for generously providing MeCP2 and HP1α expression vectors, and Dr. Jay Hess for generously providing the HRX expression vector. We also thank Dr. Joe Bidwell for helpful discussions regarding the nuclear matrix.

This work was supported by National Institutes of Health Grant CA58947 (to D.G.S.) and by the Riley Memorial Association.
REFERENCES
1. Bird, A. P. (1987) *Trends Genet.* **3**, 342-347
2. Li, E., Bestor, T. H., and Jaenisch, R. (1992) *Cell* **69**, 915-926
3. Okano, M., Bell, D. W., Haber, D. A., and Li, E. (1999) *Cell* **99**, 247-257
4. Biniszkiewicz, D., Gribnau, J., Ramsahoye, B., Gaudet, F., Eggan, K., Humpherys, D., Mastrangelo, M.-A., Jun, Z., Walter, J., and Jaenisch, R. (2002) *Mol. Cell. Biol.* **22**, 2124-2135
5. Xu, G.-L., Bestor, T. H., Bourc'his, D., Hsieh, C.-L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russso, J. J., and Viegas-Pequignot, E. (1999) *Nature* **402**, 187-191
6. Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., and Zoghbi, H. Y. (1999) *Nature Genet.* **23**, 185-188
7. Baylin, S. B., and Herman, J. G. (2000) *Trends Genet.* **16**, 168-173
8. Hendrich, B., and Bird, A. (1998) *Mol. Cell. Biol.* **18**, 6538-6547
9. Wade, P. A., Gegonne, A., Jones, P. L., Ballestar, E., Aubry, F., and Wolffe, A. P. (1999) *Nature Genet.* **23**, 62-66
10. Ng, H.-H., Zhang, Y., Hendrich, B., Johnson, C. A., Turner, B. M., Erdjument-Bromage, H., Tempst, P., Teinberg, D., and Bird, A. (1999) *Nature Genet.* **23**, 58-61
11. Voo, K. S., Carlone, D. L., Jacobsen, B. M., Flodin, A., and Skalnik, D. G. (2000) *Mol. Cell. Biol.* **20**, 2108-2121
12. Lee, J.-H., Voo, K. S., and Skalnik, D. G. (2001) *J. Biol. Chem.* **276**, 44669-44676
13. Huntsman, D. G., Chin, S.-F., Muleris, M., Batley, S. J., Collins, V. P., Wiedemann, L. M., Aparicio, S., and Caldas, C. (1999) *Oncogene* **18**, 7975-7984
14. Cross, S. H., Meehan, R. R., Nan, X., and Bird, A. (1997) *Nature Genet.* **16**, 256-259
15. Tkachuk, D. C., Kohler, S., and Cleary, M. L. (1992) *Cell* **71**, 691-700
16. Gu, Y., Nakamura, T., Alder, H., Prasad, R., Canaani, O., Cimino, G., Croce, C. M., and Canaani, E. (1992) *Cell* **71**, 701-708
17. Birke, M., Schreiner, S., Garcia-Cuellar, M.-P., Mahr, K., Titgemeyer, F., and Slany, R. K. (2002) *Nucleic Acids Res.* **30**, 958-965
18. Fujita, N., Takebayashi, S.-I., Okumura, K., Kudo, S., Chiba, T., Saya, H., and Nakao, M. (1999) *Mol. Cell. Biol.* **19**, 6415-6426
19. Aasland, R., Gibson, T. J., and Stewart, F. (1995) *Trends Biochem. Sci.* **20**, 56-59
20. O'Connell, S., Wang, L., Robert, S., Jones, C. A., Saint, R., and Jones, R. S. (2001) *J. Biol. Chem.* **276**, 43065-43073
21. Schultz, D. C., Friedman, J. R., and Raushcer, F. J. (2001) *Genes Dev.* **15**, 428-443
22. Kalkhoven, E., Teunissen, H., Houweling, A., Verrijzer, C. P., and Zantema, A. (2002) *Mol. Cell. Biol.* **22**, 1961-1970
23. Carlone, D. L., and Skalnik, D. G. (2001) *Mol. Cell. Biol.* **21**, 7601-7606
24. Monk, M., Boubelik, M., and Lehnert, S. (1987) *Develop.* **99**, 371-382
25. Kafri, T., Ariel, M., Brandeis, M., Shemer, R., Urven, L., McCarrey, J., Cedar, H., and Razin, A. (1992) *Genes Dev.* **6**, 705-714
26. Chen, C., and Okayama, H. (1987) *Mol. Cell. Biol.* **7**, 274-2752
27. Reyes, J. C., Muchardt, C., and Yaniv, M. (1997) *J. Cell Biol.* **137**, 263-274
28. Laemmli, U. K. (1970) *Nature* **227**, 680-685
29. Lee, J.-H., Shimojo, M., Chai, Y.-G., and Hersh, L. B. (2000) *Mol. Brain Res.* **80**, 88-98
30. Bachman, K. E., Rountree, M. R., and Baylin, S. B. (2001) *J. Biol. Chem.* **276**, 32282-32287
31. Caslini, C., Shilatifard, A., Yang, L., and Hess, J. L. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 2797-2802
32. Zhang, S., Herrmann, C., and Grosse, F. (1999) *J. Cell. Sci.* **112**, 2693-2703
33. Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992) *Cell* **69**, 905-914
34. Murzina, N., Verreault, A., Laue, E., and Stillman, B. (1999) *Mol. Cell.* **4**, 529-540
35. Caslini, C., Serna Alarcon, A., Hess, J. L., Tanaka, R., Murti, K. G., and Biondi, A. (2000) *Leukemia* **14**, 1898-1908
36. Butler, L. H., Slany, R., Cui, X., Cleary, M. L., and Mason, D. Y. (1997) Blood 89, 3361-3370
37. Yano, T., Nakamura, T., Blechman, J., Sorio, C., Dang, C. V., Geiger, B., and Canaani, E. (1997) Proc. Natl. Acad. Sci. USA 94, 7286-7291
38. Zeng, C., He, D., and Brinkley, B. R. (1994) Cell Motil. Cytoskeleton 29, 167-176
39. Berube, N. G., Smeenk, C. A., and Picketts, D. J. (2000) Hum. Mol. Genet. 9, 539-547
40. Cobb, B. S., Morales-Alcegay, S., Kleiger, G., Brown, K. E., Fisher, A. G., and Smale, S. T. (2000) Genes Dev. 14, 2146-2160
41. Jans, D. A., and Hubner, S. (1996) Physiol. Rev. 76, 651-685
42. Fujino, T., Hasegawa, M., Shibata, S., Kisimoto, T., Imai, S.-I., and Takano, T. (2000) Biochem. Biophys. Res. Comm. 271, 305-310
43. Gordon, J. A., Pockwinse, S. M., Stewart, F. M., Quesenberry, P. J., Nakamura, T., Croce, C. M., Lian, J. B., Stein, J. L., Van Wijnen, A. J., and Stein, G. S. (2000) J. Cell. Biochem. 77, 30-43
44. Adler, H. T., Nallaseth, F. S., Walter, G., and Tkachuk, D. C. (1997) J. Biol. Chem. 272, 28407-28414
45. Fair, K., Anderson, M., Bulanova, E., Mi, H., Tropschug, M., and Diaz, M. O. (2001) Mol. Cell. Biol. 21, 3589-3597
46. Leshkowitz, D., Rozenblatt, O., Nakamura, T., Yano, T., Dautry, F., Croce, C. M., and Canaani, E. (1996) Oncogene 13, 2027-2031
47. Ernst, P., Wang, J., Huang, M., Goodman, R. H., and Korsmeyer, S. J. (2001) Mol. Cell. Biol. 21, 2249-2258
48. Adler, H. T., Chinery, R., Wu, D. Y., Kussick, S. J., Payne, J. M., Fornace, A. J., Jr., and Tkachuk, D. C. (1999) Mol. Cell. Biol. 19, 7050-7060
49. Rozenblatt-Rosen, O., Rozovskaia, T., Burakov, D., Sedkov, Y., Tillib, S., Blechman, J., Nakamura, T., Croce, C. M., Mazo, A., and Canaani, E. (1998) Proc. Natl. Acad. Sci. USA 95, 4152-4157
50. Katsani, K. R., Arredondo, J. J., Kal, A. J., and Verrijzer, C. P. (2001) Genes Dev. 15, 2197-2202
51. Miller, T., Krogan, N. J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J. F., and Shilatifard, A. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 12902-12907

52. Nagy, P. L., Griesenbeck, J., Kornberg, R. D., and Cleary, M. L. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 90-94

53. Roguev, A., Schaft, D., Shevchenko, A., Pim Pijnappel, W., Wilm, M., Aasland, R., and Stewart, A. F. (2001) *EMBO J.* **20**, 7137-7148

54. Krogan, N. J., Dover, J., Khorrami, S., Greenblatt, J. F., Schneider, J., Johnston, M., and Shilatifard, A. (2002) *J. Biol. Chem.* **277**, 10753-10755

55. Strahl, B. D., Ohba, R., Cook, R. G., and Allis, C. D. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 14967-14972

56. Nislow, C., Ray, E., and Pillus, L. (1997) *Mol. Biol. Cell.* **8**, 2421-2436

57. Schramke, V., Neecke, H., Brevet, V., Corda, Y., Lucchini, G., Pia Longhese, M., Gilson, E., and Géli, V. (2001) *Genes Dev.* **15**, 1845-1858

58. Razin, A. D., and Shemer, R. (1995) *Hum. Mol. Genet.* **4**, 1751-1755
FIGURE LEGENDS

Figure 1. CGBP localizes to nuclear speckles in DAPI-light regions.
A. Nuclear distribution pattern of endogenous CGBP and GFP-CGBP fusion proteins. Endogenous CGBP protein was detected in NIH-3T3 and HEK-293 cells using anti-CGBP antibody and Texas red-conjugated secondary antibody. Nuclei were counterstained with DAPI and observed by confocal microscopy. Vectors expressing GFP-tagged CGBP were transiently transfected into NIH-3T3 and HEK-293 cells as described in Experimental Procedures.
B. Nuclear distribution of speckles containing GFP-CGBP was compared with DAPI staining in NIH-3T3 and HEK-293 cells.

Figure 2. CGBP is localized to euchromatin and co-localizes with HRX.
Nuclear distribution of CGBP was compared with the methyl-CpG specific proteins MBD1 and MeCP2, heterochromatin marker protein HP1α, RNA splicing factor 35, actively transcribed regions, markers of active chromatin region acetylated histone H3 and acetylated histone H4, and HRX. FLAG-tagged or GFP-tagged CGBP were co-transfected with GFP-tagged MBD1 or FLAG-tagged MeCP2, HP1α, or amino terminus (amino acids 1-1436) HRX. The FLAG epitope was detected with anti-FLAG antibody and Texas red conjugated secondary antibody. GFP-tagged CGBP was transiently expressed and compared with endogenous SC-35, acetylated histone H3, and acetylated histone H4 using specific antisera and Texas red-conjugated secondary antibody as described in Experimental Procedures. Actively transcribed regions were labeled with bromouridine and detected as described in Experimental Procedures. All experiments were performed in NIH-3T3 cells except for SC-35, which was examined in HEK-293 cells.
Figure 3. CGBP is excluded from chromosomes during mitosis.
GFP-tagged CGBP or GFP-tagged MBD1 was transiently expressed in HEK-293 cells, which were then grown in medium containing 0.1 µg/ml colcemid to block the cells in mitosis. Nuclei were counterstained with DAPI and observed with confocal microscopy.

Figure 4. CGBP is associated with the nuclear matrix.
NIH-3T3 and HEK-293 cells were fractionated by sequential nuclear extraction as described in Experimental Procedures. An equal proportion of each fraction was analyzed by western blotting using anti-CGBP, anti-acetyl histone H3, or anti-NuMA antisera.

Figure 5. Association of CGBP with nuclear speckles and the nuclear matrix is independent of DNA-binding activity.
A. Targeting of CGBP to nuclear speckles does not require DNA-binding activity. GFP-tagged CGBP mutants (C169A and C208A) that fail to bind to DNA (12) and FLAG-tagged CGBP were co-transfected into NIH-3T3 and HEK-293 cells and detected with confocal microscopy using anti-FLAG antibody and Texas red conjugated secondary antibody as described in Experimental Procedures.
B. DNA-binding domain mutants of CGBP are associated with the nuclear matrix. FLAG-tagged CGBP expression vectors were transiently expressed in HEK-293 cells and fractionated as described in Experimental Procedures. An equal proportion of each fraction was analyzed by western blotting using anti-FLAG antibody.
C. DNA-binding domain mutants of CGBP do not trans-activate reporter genes. Increasing amounts (1 µg, 2.5 µg, 5 µg, and 10 µg) of FLAG-tagged CGBP expression vectors were co-transfected with fixed amounts of CMV-based pEGFP-C2 vector into HEK-293 cells. Total amounts of transfected DNA were normalized using pcDNA3 vector DNA. Cells were harvested two days following transfection, washed with PBS, and lysed with 8 M urea buffer. Equal proportions of each fraction were analyzed by western blot using anti-FLAG or anti-GFP antisera as described in Experimental Procedures.
Figure 6. Analysis of protein domains that direct nuclear localization and target CGBP to nuclear speckles.

A. Schematic diagram of CGBP and various truncated fragments. Protein regions such as the PHD, CXXC, acidic, basic, and coiled-coil domains are indicated, along with corresponding amino acid positions. The amino terminus bipartite nuclear localization signal of CGBP (amino acids 109-121) was inserted between the GFP epitope and CGBP sequence for constructs containing amino acids 361-481, 302-656, and 213-320.

B. Sub-cellular distribution of GFP-CGBP fusion proteins. GFP-tagged constructs were transiently expressed in NIH-3T3 cells, and observed with confocal microscopy as described in Experimental Procedures.

C. The co-localization of a truncated mutant (amino acids 103-481) with wild type CGBP. GFP-tagged mutant and FLAG-tagged CGBP were co-transfected and detected using anti-FLAG antibody and Texas red-conjugated secondary antibody with confocal microscopy.

Figure 7. Identification of CGBP domains that mediate association with the nuclear matrix.

A. Schematic diagram of CGBP and truncated fragments. The amino terminus bipartite nuclear localization signal of CGBP (amino acids 109-121) was inserted between the epitope (GFP or FLAG) and CGBP sequence for constructs containing amino acids 302-656 and 361-481. Numbers indicate amino acid residues of CGBP.

B. Biochemical fractionation of GFP- and FLAG-CGBP fusion proteins. FLAG or GFP-tagged constructs were transiently expressed in HEK-293 cells and fractionated as described in Experimental Procedures. An equal proportion of each fraction was analyzed by western blotting using anti-FLAG or anti-GFP antisera. S, soluble fraction; C, chromatin fraction; M, nuclear matrix fraction.
Figure 8. Localization of CGBP to nuclear speckles and association with the nuclear matrix is required for trans-activation activity.

FLAG-tagged full length CGBP (5 µg), FLAG-tagged CGBP amino acids 1-481 (5 µg), FLAG-tagged CGBP amino acids 1-367 (7.5 µg) or FLAG-tagged CGBP amino acids 1-320 (20 µg) were co-transfected with 2.5 µg of CMV-based pEGFP-C2 reporter vector into HEK-293 cells as described in Experimental Procedures. Variable amounts of CGBP expression vector plasmid were utilized in an effort to produce comparable levels of FLAG-CGBP expression. Western blot analysis was performed to detect the expression levels of FLAG-CGBP and GFP.
Figure 1. Lee & Skalnik

A

NIH-3T3
Endogenous CGBP  GFP-CGBP

HEK-293
Endogenous CGBP  GFP-CGBP

B

GFP-CGBP  DAPI  Merge

NIH-3T3

HEK-293
Figure 2. Lee & Skalnik
Figure 3. Lee & Skalnik

GFP-tagged protein  
CGBP  

GFP-tagged protein  
MBD1  

DAPI
Figure 4. Lee & Skalnik

| HEK-293 | NIH-3T3 |
|---------|---------|
| Soluble | Chromatin | 2M NaCl | Matrix | Soluble | Chromatin | 2M NaCl | Matrix |
| CGBP    |          |        |        |        |          |        |        |
| NuMA    |          |        |        |        |          |        |        |
| Acetylated Histone H3 |          |        |        |        |          |        |        |
Figure 5. Lee & Skalnik

A

NIH-3T3
HEK-293

GFP-CGBP (C109A)
FLAG-CGBP
Merge

B

Soluble Chromatin Soluble Chromatin Soluble Chromatin
CGBP CGBP(C169A) CGBP(C208A)

Anti-FLAG

C

Vector CGBP CGBP(C169A) CGBP(C208A)

Anti-FLAG

Anti-GFP
